

Wir alle sind mit Vorurteilen vertraut. Sie aufzuheben ist ein langsamer Prozess, denn **Schubladendenken** ist ein wichtiger **Teil unserer Evolutionsgeschichte**. Selbst in neuesten Technologien sind sie zu finden:

Algorithmen, die unser aller Leben beeinflussen, sind ein großer Stolperstein auf dem Weg in eine **vorurteilsfrei(er)e Zukunft**.

Von Sabine Gromer*

Die unsichtbare Gefahr



Schubladendenken vereinfacht Entscheidungsprozesse und spart uns viel Energie. Doch sie hat einen großen Nachteil: **sie ist immer unfair.** Bis vor einigen Jahren haben wir nur zwischen strukturellen und unbewussten Vorurteilen unterschieden. Beide sind Einstellungen und Überzeugungen, die außerhalb unseres Bewusstseins liegen. Strukturelle Vorurteile sind kollektive, systemdefinierte Vorurteile, die wir meist nicht hinterfragen, unser aller Leben aber täglich beeinflussen. Die unbewussten Vorurteile sind unsere individuellen, persönlichen, die wir uns im Lauf unseres Lebens aneignen, quasi der Fußabdruck des strukturellen Bias auf ein Individuum. Während wir strukturelle Vorurteile nicht allein lösen können, können wir in Bezug auf unseren individuellen Unconscious Bias schnell konkrete Schritte setzen. Etwas schwieriger, weil noch wenig bekannt und auch schwerer fassbar, ist es da mit einer neuen und dritten Kategorie: dem AI Bias.

Die neue Gefahr: AI Bias

Wir leben in einer neuen Ära: dem Zeitalter der künstlichen Intelligenz (AI). Viele Unternehmen und Organisationen setzen auf Algorithmen. Es gibt kaum Zweifel an der Leistungsfähigkeit von AI-Systemen. Die durch Deep Learning ermöglichte Datenanalyse und Mustererkennung versetzt AI in die Lage, Krebs im Frühstadium mit größerer Genauigkeit zu diagnostizieren als menschliche Ärztinnen. AI kann Leben retten, Leben verändern, aber auch – und genau darin liegt das Problem – negativ beeinflussen oder im schlimmsten Fall zerstören. AI bewirkt vor allem eins: sie lässt confirmation bias unkontrolliert und ungebremst wachsen und zeigt uns die Welt so, wie sie uns gefällt. Dadurch werden unsere Vorurteile täglich aufs Neue bestätigt.

Was bedeutet es, wenn Algorithmen zunehmend unser Leben bestimmen? Was sind die Konsequenzen für Menschen, gegen die ein Algorithmus voreingenommen ist?

In vielen großen Unternehmen werden grundlegend wichtige Entscheidungen durch AI unterstützt. ZB trifft AI eine Vorauswahl, wer zu Vorstellungsgesprächen eingeladen wird. Das Wissen der Algorithmen, die derartige Entscheidungen tragen, speist sich aus mangelhaften, nicht repräsentativen und vergangenen, besonders vorurteilsbehafteten Erfahrungen und hemmt damit eine positive Weiterentwicklung hin zu einer unbiased Zukunft. Weniger abstrakt erklärt: Entscheidet ein Algorithmus darüber, wer für die ausgeschriebene Stelle in einer Führungsposition in die engere Auswahl kommen soll, scheidet er vorweg Menschen mit Migrationshintergrund, Menschen mit Behinderungen und Frauen aus, weil er auf datenbasierte Erfahrungswerte zurückgreift. Für einen Algorithmus sind weiße Männer mittleren Alters eine sichere Wahl. AI-Systeme können ethische und soziale Belange nicht berücksichtigen, wenn sie nicht entsprechend programmiert werden.

AI kann nichts dafür

Maschinelles Lernen, ein Teilbereich der künstlichen Intelligenz, ist abhängig von der Qualität, Objektivität und Größe der Trainingsdaten, die zum Lernen herangezogen werden. Fehlerhafte, schlechte, unvollständige und vor allem veraltete Daten führen zu ungenauen Vorhersagen. Vorurteile im maschinellen Lernen von AI sind zum einen auf Probleme zurückzuführen, die von – nicht vorurteilsfreien – Personen verursacht werden, die diese maschinellen Lernsysteme entwerfen und/oder trainieren. Zum anderen werden unvollständige, fehlerhafte, voreingenommene und veraltete Datensätze zum Trainieren und/oder Validieren der maschinellen Lernsysteme verwendet.

AI Bias ist im Allgemeinen unbeabsichtigt. AI ist nicht böse. Doch die Folgen ihrer Entscheidungen können erheblich sein: schlechtes Kundenservice, geringere Umsätze und Erträge, unfaire und auch illegale Handlungen. Und: sie können zu potenziell gefährlichen Bedingungen führen.

In einem Gespräch mit dem Regisseur Werner Herzog über die Gefahren von AI nannte Tesla-Gründer Elon Musk ein überzeugendes Beispiel: Angenommen, AI wurde damit beauftragt, den Wert eines Anlageportfolios zu maximieren. Nehmen wir außerdem an, dass die Schöpfer des Systems nicht eindeutig spezifiziert haben, wie dieses Ziel zu erreichen ist. Theoretisch könnte AI in diesem Fall verstärkt in Rüstungsaktien investieren und so im schlimmsten Fall einen Krieg auslösen. Umso wichtiger ist es, die Daten, die zum Trainieren maschineller Lernmodelle verwendet werden, auf mangelnde Vollständigkeit und kognitive Verzerrungen zu überprüfen.

Aber Achtung! Das Blackbox-Problem

Die größte Gefahr von AI-Systemen ist das sogenannte Blackbox-Problem: wir sind inzwischen unfähig vollständig zu verstehen, warum die Algorithmen hinter AI so funktionieren, wie sie es tun.

Unser Gehirn ist das lebenswichtigste, das komplexeste und rätselhafteste unserer Organe. AI, die den Höhepunkt der technologischen Entwicklung darstellt, ist – trotzdem wir sie geschaffen haben – ebenfalls ein Rätsel für uns. Wir wissen zwar eine Menge über unser Gehirn und können mit einiger Sicherheit vorhersagen, wie es auf verschiedene Reize reagiert. Ebenso können wir mit einiger Gewissheit voraussagen, welche Ergebnisse ein AI-Algorithmus bei bestimmten Eingaben liefern wird. Das Mysterium vor dem Wissenschaftler:innen stehen, ist nicht das des Outputs, sondern wie dieser erzeugt wird. Dieses Unwissen über das Innenleben dieser AI-Blackbox ist die größte Hürde in der AI-Entwicklung und wird uns noch lange, wenn nicht für immer, begleiten.

Der breite Einsatz von AI wird eine besondere Herausforderung sein. Organisationen, die sie nutzen, werden offenlegen müssen, welche menschlichen Entscheidungen hinter dem Design ihrer AI-Systeme stehen, welche ethischen und sozialen Belange sie berücksichtigt haben und wie gut sie die Ergebnisse dieser Systeme auf Spuren von Voreingenommenheit oder Diskriminierung überwacht haben. Wir brauchen Modelle, denen wir auch vertrauen können. Das Erreichen von Transparenz bei AI-Systemen ist von entscheidender Bedeutung. Und wir müssen vor allem eines: **Frau der Lage bleiben.**

Was bedeutet das konkret? Was kann ich tun? Erstens: Wir müssen digital literate werden, uns intensiv mit den Technologien auseinandersetzen und sie so gut wie möglich verstehen. Wir dürfen das Feld nicht den Männern überlassen. Zweitens: Es ist unser Auftrag und unsere Aufgabe AI-Modelle zu hinterfragen, wir müssen darauf bestehen, dass Stellenausschreibungen genderneutral verfasst werden und dass Kolleg:innen die Auswahl der AI überprüfen und diverse Daten als Grundlage herangezogen werden. Und drittens –... unglaublich wichtig!: AI lernt von unserer Sprache. Daher ist es umso wichtiger, zu gendern.

Confirmation bias ist die Tendenz, Informationen auf eine Weise zu suchen, zu interpretieren, zu bevorzugen und abzurufen, die die eigenen Überzeugungen oder Werte bestätigt oder unterstützt.



***Sabine Gromer** hat 20 Jahre ihrer Karriere in der Finanzwelt verbracht, in dieser Zeit verschiedenste Führungsrollen übernommen und in London, New York, Hong Kong und Tokio gelebt. Zuletzt war sie als

Managing Director die Global Head of Organisational Effectiveness bei der Rating-Agentur S&P Global Ratings (ehemals Standard & Poor's Ratings) in London. Vor 2,5 Jahren ist sie nach Wien gezogen ist und hat sich mit der Gründung ihres Unternehmens „MagnoliaTree“ einen Lebensraum erfüllt. Heute wirkt sie als Executive Coach für StrategieentscheiderInnen und Transformationsarchitektin für Change Leadership und erforscht die Essenz von Würde.

Der Artikel entstand in Zusammenarbeit mit Daniela Luschin.