**Coded Bias: Können Maschinen diskriminieren?**

**Können Gesichter wie Fingerabdrücke funktionieren?**

Entsperrst du dein Smartphone auch mit deinem Gesicht? Du blickst einfach auf den Screen deines Smartphones und die Nutzeroberfläche schaltet sich automatisch frei. Dein Gesicht funktioniert wie eine PIN und du wirst als Besitzer:in deines Smartphones identifiziert.

Systeme der Künstlichen Intelligenz (KI), die Gesichter verarbeiten und zuordnen können, sind in unserem Alltag angekommen und das nicht nur auf unseren Smartphones. Sie sorgen beispielsweise auch dafür, dass sich in modernen Unternehmen die Eingangstüren für Mitarbeiter:innen öffnen oder ordnen Fotos für uns auf unserem Computer. Doch es gibt immer wieder Meldungen, dass diese KI-Technologie nicht für alle Menschen gleichermaßen funktioniert.

In der Netflix-Dokumentation „Coded Bias“ (deutscher Titel: Vorprogrammierte Diskriminierung) erzählt die ghanaisch-amerikanisch-kanadische Informatikerin Joy Buolamwini von ihren persönlichen Erfahrungen mit KI-Software zur Gesichtserkennung. Im Rahmen ihres Informatikstudiums war sie in einem Projekt tätig, das Gesichtserkennungssoftware einsetzen wollte. Schnell stellte sie fest, dass ihr Gesicht von der KI-Software nicht als Gesicht erkannt wurde. Daraufhin setzte sie sich eine weiße Maske auf und probierte es erneut. Nun identifizierte das Programm die weiße Maske als Gesicht. Das gab der Forscherin zu denken. Sie fragte sich, warum ihr Schwarzes[[1]](#footnote-2) Gesicht von der Gesichtserkennungssoftware nicht als Gesicht eingeordnet wurde, eine weiße Maske allerdings schon.

**Programmieren Entwickler:innen ihre Vorurteile in KI-Systeme ein?**

Joy Buolamwini ging diesem fehlerhaften und zugleich diskriminierenden Ergebnis der KI-Software nach. Dafür überprüfte sie die Trainingsdaten, aus denen die Gesichtserkennungssoftware ihr Vorgehen ableitete. Sie entdeckte, dass diese Trainingsdaten mehrheitlich aus Bildern bestanden, die die Gesichter von Menschen – vor allem Männern – mit hellerer Haut zeigten.

**Begriffsdefinition: Was bedeutet Diskriminierung?**

Diskriminierung bedeutet, dass Menschen aufgrund ihrer Hautfarbe, ihrer religiösen Zugehörigkeit, ihrer sexuellen Orientierung, ihres Geschlechts, ihres Alters oder anderer Merkmale benachteiligt werden. Diese Ungleichheit zwischen Menschen infolge von tatsächlichen oder zugeschriebenen Merkmalen erfolgt nicht zufällig, sondern ist historisch in unserer Gesellschaft gewachsen. Es gibt verschiedene Diskriminierungsformen wie Rassismus, Sexismus, Antisemitismus, Diskriminierung gegen Menschen mit Behinderung oder mit chronischen Krankheiten, die sich über lange Zeiträume entwickelt haben und in unserer Gesellschaft wirken. Wenn Menschen diskriminiert werden, dann muss das nicht mit Absicht geschehen und kann auch unbewusst erfolgen. Daher ist es umso wichtiger, bestehende Ungleichheiten zwischen Menschen aufzudecken und sich dafür einzusetzen, dass alle Menschen gleichberechtigt sind. Diskriminierung ist gesetzlich verboten: Das Grundgesetz und das Allgemeine Gleichbehandlungsgesetz schützen Menschen vor Diskriminierung.

Die Entwickler:innen dieser Gesichtserkennungssoftware trainierten das KI-System also mit einer Auswahl an Bildern, die die verschiedenen Hauttöne von Menschen nicht widerspiegelten. Daher wurden Gesichter von weißen Männern wesentlich häufiger vom KI-System als Gesichter eingeordnet, während die Gesichter von Schwarzen Frauen mit großem Abstand am seltensten erkannt wurden. In der Trainingsphase hatten die Entwickler:innen der KI-Software schlichtweg zu wenig Bilder von Schwarzen Frauen vorgegeben. Deshalb konnte die KI-Software das Gesicht von Joy Boulamwini nicht erkennen und einordnen, die weiße Maske allerdings schon. Wenn eine solche KI-Software nun Menschen Türen öffnen sollte, würden Schwarze Frauen häufig vor verschlossenen Türen stehen. Sie würden demnach anders behandelt als Menschen mit weißer Hautfarbe. Das ist diskriminierend.

Diese Erkenntnis veranlasste die Forscherin dazu, der Frage nachzugehen, in welchem Ausmaß technologische Entwicklungen Ungleichheiten zwischen Menschen verstärken können. Grundsätzlich ist das immer dann der Fall, wenn bewusst oder unbewusst Vorurteile von Menschen in technologische Entwicklungen einfließen und deren Funktionsweise beeinflussen. Im Fall der Gesichtserkennungssoftware, die Joy Boulamwini testete, zeigen sich die Vorurteile in den ausgewählten Daten: Die Trainingsdaten waren unvollständig. Es fehlten Fotos von Menschen mit anderen Hautfarben. Zudem waren weniger Fotos von Frauen enthalten und somit kaum Fotos von Schwarzen Frauen. In der IT-Branche sind weiße Männer deutlich präsenter als Schwarze Frauen. Unter anderem deshalb kommt es häufiger vor, dass Schwarze Menschen nicht mitgedacht werden. Dies betrifft zum einen die Daten, die vorhanden sind: Es gibt grundsätzlich weniger Bilder und andere Daten von Schwarzen Frauen als von weißen Männern. Zum anderen kann das Vorgehen bei der Software-Entwicklung dazu führen, dass ein unfaires Ergebnis entsteht: Die Entwickler:innen der Gesichtserkennungssoftware berücksichtigen nicht, dem System auch Fotos Schwarzer Frauen vorzugeben.

**Voreingenommenheit erkennen und Diskriminierung verhindern**

Diskriminierung ist in den Strukturen unserer Gesellschaft verankert. KI-Technologien werden von Menschen entwickelt, die in dieser Gesellschaft leben. Deshalb spiegeln sich diese diskriminierenden Strukturen beispielsweise auch in Daten und darauf aufbauenden Technologien wider. In vielen Fällen erfolgt dies jedoch nicht bewusst.

Um diskriminierende Ergebnisse zu verhindern und frühzeitig zu erkennen, ist es wichtig, den Einsatz von KI-Software immer wieder auf auftretende Benachteiligungen zu überprüfen. Werden welche entdeckt, muss ihnen schnell begegnet werden. Das kann zum Beispiel bedeuten, dass die KI-Anwendung vorübergehend nicht benutzt wird, bis die Benachteiligung behoben wurde. Im Fall der Gesichtserkennungssoftware hieße das, dass die Software mithilfe zahlreicher Abbildungen von Menschen mit unterschiedlichen Hautfarben und unterschiedlichen Geschlechtern verbessert wird.

Die Erfahrung von Joy Buolamwini veranlasste sie, die Organisation „Algorithmic Justice League“ (zu Deutsch etwa: Bündnis für algorithmische Gerechtigkeit) zu gründen, um auf diskriminierende Auswirkungen von Künstlicher Intelligenz aufmerksam zu machen. Die Entwicklung von KI-Technologien kann demnach auch dazu beitragen, bestehende Diskriminierungen aufzuzeigen. Zugleich wird deutlich, welche Verantwortung wir bei der Entwicklung von KI tragen. Das Ziel sollte sein, dass KI-Systeme für alle Menschen ein Werkzeug sind, das ihnen das Leben erleichtert.

1. Das Wort Schwarz wird hier mit einem großen „S“ geschrieben, weil es sich nicht um die Farbe „schwarz“ handelt. „Schwarz“ mit großem „S“ ist eine politische Selbstbezeichnung für Menschen, die aufgrund ihres Aussehens sowie tatsächlicher oder zugeschriebener Merkmale Rassismus erfahren. [↑](#footnote-ref-2)