

## KI in den Nachrichten

Lies dir die folgenden Szenarien aufmerksam durch.

### Szenario 1: Bewerberauswahl mit KI

Ein Unternehmen sucht neue Mitarbeiter und Mitarbeiterinnen. Dazu soll möglichst wenig Zeit investiert werden. Dennoch sollte die Auswahl nahezu perfekt laufen. Daher trainieren sie ein maschinelles Modell. Als Trainingsdaten werden die Einstellungen der letzten 10 Jahre genutzt. Die eingestellten Mitarbeiterinnen und Mitarbeiter werden u.a. mit Anschrift, Alter, Informationen aus dem Lebenslauf, Geschlecht und Foto erfasst.

Das KI-System sollte somit lernen, welche Eigenschaften das Unternehmen bei seinen Mitarbeiterinnen und Mitarbeitern bevorzugt und damit anschließend selbst die Bewerberauswahl vornehmen.

Das Unternehmen setzte das entstandene Modell für einige Zeit ein. Dabei bemerkte das Management, dass, obwohl viele Bewerber Frauen waren, sie dennoch eher selten vom KI-System eingestellt wurden und häufig direkt eine Ablehnung erfuhren. Weil das in der IT-Branche tätige Unternehmen in der Vergangenheit vorwiegend Männer eingestellt hatte, lernte das Modell, dass Bewerbungen von Frauen schlechter zu bewerten seien. Dies war auch dann der Fall, wenn die Bewerbung keinerlei Angabe des Geschlechts enthielt, da diese Information auch indirekt bspw. über Interessen erschlossen werden konnte.

### Szenario 2: Gesichtserkennungssysteme

Die Amerikanerin Joy Buolamwini, Wissenschaftlerin am MIT in Boston, untersuchte 2017 in einem Forschungsprojekt die damals aktuelle Gesichtserkennungssoftware großer Anbieter wie Microsoft, IBM oder Face++. Dabei stellte sie fest, dass die Software der Hersteller zwar insgesamt eine sehr hohe Genauigkeit bei der Erkennung des Geschlechts von über 90% aufwiesen, aber deutlich schlechter darin sind, das Geschlecht einer dunkelhäutigen Frau zu bestimmen. So wiesen die Systeme dann bspw. nur noch eine Genauigkeit von 65,3% (IBM), 65,5 % (Face++) und 79,3% auf.

Der Softwareentwickler Jacky Alciné entdeckte im Jahr 2015 in seiner Foto-Sammlung ein Album mit dem Titel „Gorillas“, das Googles Gesichtserkennungsalgorithmus automatisch für ihn erstellt hatte. Darin ein Foto von ihm und einer ebenfalls schwarzen Freundin. Zwei Jahre später hatte Google es immer noch nicht geschafft, das Problem zu beheben und behalf sich stattdessen damit, Primaten aus dem Lexikon der Suchbegriffe zu tilgen: eine Verlegenheitslösung. Aber das ist nur ein Beispiel dafür, wie Menschen im Alltag von schlecht funktionierenden Systemen diskriminiert werden. Gesichtserkennung wird inzwischen nicht nur verwendet, um Fotos zu sortieren. Auch die Polizei setzt sie bei der Strafverfolgung ein, auf einigen Flughäfen beschleunigt sie die Passkontrolle (Derivat von netzpolitik.org, BY-NC-SA).

### Szenario 3: Notenvergabe in Großbritannien

Wegen der Corona-Pandemie und den Schulausfällen konnten in Großbritannien Prüfungen zum Abitur und zur mittleren Reife nicht abgelegt werden. Zuerst plante das Bildungsministerium, dass die Lehrkräfte die Noten auf Grundlage früherer Bewertungen vergaben. Das führte dazu, dass die Noten im Durchschnitt besser waren als in den Vorjahren. Deswegen ließ das Bildungsministerium die Noten noch einmal mit einem Algorithmus korrigieren, der die Durchschnittsnoten der jeweiligen Schulen der Vorjahre mit einbezog. Die Folge: In 280.000 Fällen wurden Schülerinnen und Schüler schlechter bewertet, das sind fast 40 Prozent aller Schulabschlüsse des Jahres. Bei den Protesten machten die Schülerinnen und Schüler auf Schildern darauf aufmerksam, dass sie den Algorithmus als diskriminierend wahrnehmen. So war auf Plakaten zu lesen, dass Lehrerinnen und Lehrer und nicht die Postleitzahlen über Noten entscheiden sollten. Die Maßnahme wurde mit Schildern wie „Classroom not Class War“ als sozial ungerecht bezeichnet. Hintergrund ist, dass Schülerinnen und Schüler aus privaten, privilegierten und guten Schulen nicht von der Abwertung betroffen sein könnten, während „Problemschulen“ zur Abwertung führen können – ungeachtet der individuellen Schulleistungen der Einzelnen. (Derivat von netzpolitik.org, BY-NC-SA).

Diskutiere mit deinem Nachbarn folgende Fragestellungen:

- Welches Problem wird in den jeweiligen Szenarien beschrieben?
- Welche Ursache(n) könnten die problematische Entscheidungen der KI-Systeme in den beschriebenen Szenarien haben? / Erläutere unter Einbeziehung der verwendeten Daten, warum es zu diesen Problemen kommen könnte?
- Basierend auf deinen Erfahrungen mit dem Training von Modellen: Wie könnte man diese Probleme lösen?
- Welche Regeln sollten für einen fairen Einsatz von KI-Systemen gelten?