



Bias-Fehler in der künstlichen Intelligenz

Stell dir folgendes Szenario vor...

... eine künstliche Intelligenz entscheidet für dich nach der Volksschule welche weiterführende Schule du besuchen sollst (Und in Folge auch welchen Berufsweg du daraufhin einschlagen wirst)!

Diese künstliche Intelligenz versucht dabei anhand deiner bisher festgestellten Talente, Eigenschaften wie Geschlecht und Herkunft, Interessen und deiner Intelligenz die richtige Schule in deinem Umfeld zu finden.

Dafür müssen sehr viele dieser genannten Faktoren als Daten von möglichst vielen Schülerinnen und Schülern in ein Programm eingespeist werden. So können diese in Verbindung zueinander gestellt werden und dadurch kann ein möglichst genaues Ergebnis für dich erreicht werden.

Doch plötzlich entscheidet dieses KI-System, dass du in eine technische Schule gehen sollst, obwohl du das nie wolltest! Was ist passiert?



Aufgabe 1: Stell dir vor einer guten Freundin von dir ist genau das passiert und ein KI-Algorithmus möchte sie irrtümlich in die falsche Schule schicken.

- Sammelt in Gruppen möglichst viele Eigenschaften, die diese Freundin haben könnte (wo wohnt sie, was macht sie gern, welche Charaktereigenschaften hat sie...)
- Schreibt nun auf, welche Vorurteile mit diesen Eigenschaften zusammenhängen können (z.B. Frauen im Technikbereich, wohnt in einer armen/reichen Nachbarschaft, ist chaotisch, trägt eine Brille, ist blond...)
- An welche Schule würde eurer Meinung nach der schlecht programmierte Algorithmus eure Freundin schicken? An welche Schule würdet ihr sie schicken wollen?
- Was müsste passieren, damit das KI-Programm den Fehler nicht nochmal macht?

Es ist möglich, dass sich beim Sammeln der vielen Daten Fehler einschleichen. Vielleicht gibt es bereits Schülerinnen und Schüler, die dir sehr ähnlich sind, die zum Beispiel ähnlich klug sind, gerne logische Aufgaben lösen und sehr sozial sind. Nachdem diese Kinder alle an eine technische Schule gehen wollten, entscheidet die KI, dass du genauso an eine technische Schule gehen möchtest.

Solche Fehler werden Bias-Fehler genannt. Sie passieren, wenn sich Vorurteile in die Datensätze von Programmen einschleichen.



Aufgabe 2: Überlegt euch in Kleingruppen, was man machen könnte, damit keine Kinder mehr in die falsche Schule geschickt werden. Wie können sogenannte Bias-Fehler vermieden werden?

Bedenkt dabei: Woher kommen all diese Daten? Wer entscheidet welche Daten verwendet werden?

Solche Algorithmen gibt es bereits in der realen Welt. Dabei wird zum Beispiel mithilfe einer künstlichen Intelligenz darüber entschieden, ob eine Person einen Kredit erhält. Das Programm bewertet diese Menschen aufgrund ihres Alters, Verdienstes und Wohnorts. Wenn aber in einer Nachbarschaft viele Menschen bereits hohe Schulden haben, könnte es passieren, dass das Programm auch für diese Person, die um einen Kredit anfragt, entscheidet, dass sie weniger Geld bekommt, obwohl sie gar keine Schulden hat.

Du siehst also: es ist unglaublich schwierig, einen Entscheidungsalgorithmus zu entwickeln, der alle Wahrscheinlichkeiten miteinbezieht und möglichst fair durch die eingespeisten Daten entscheidet.

Diese Programme „denken“ also nur so gerecht, wie jene Menschen, die ihnen diese Entscheidungen beibringen. Dazu fällt auch häufig das Sprichwort: **„Garbage in, garbage out“**.



Was sind Bias-Fehler?

Welche Beispiele fallen euch ein?

Welche Arten von Bias-Fehler gibt es?
