4 WIE KANN FAIRE KI ENTWICKELT WERDEN?

Erinnerst du dich an den KI-Entwicklungsprozess aus dem ersten Zine? Fairness muss bei jedem Schritt berücksichtigt werden. Auf den nächsten Seiten geht es darum, wie!

DATEN-SAMMLUNG

Datensätze sind nicht vielfältig genug. Das Sammeln von mehr und den richtigen Daten – zusammen mit von der Anwendung Betroffenen – ist der beste Weg. Außerdem gibt es technische Mittel zur Änderung von Datensätzen: Zum Beispiel Reweighting (Vergrößerung unterrepräsentierter Gruppen im Datensatz) oder synthetische Daten (berechnet Schätzungen für fehlende Daten).

PROBLEM-DEFINITION

Wer hat das Sagen? Wer definiert die Herausforderungen? Das Streben nach Gerechtigkeit bedeutet, die gelebten Erfahrungen der Gruppen zu berücksichtigen, deren Leben von KI betroffen ist. Der Verzicht auf den Einsatz von KI sollte immer eine Option sein, um Schaden zu verhindern.

TRAINING

Modelle werden normalerweise so gebaut, dass sie möglichst genau sind. **Was wäre, wenn wir dem Modell sagen würden, dass auch Fairness wichtig ist?**

Dies könnte durch die Einbeziehung von Fairnesskriterien in das Trainingsziel erreicht werden, so dass das Modell sowohl Genauigkeit als auch Gleichbehandlung beim Lernen optimiert.

Ein Schritt zur Entwicklung gerechterer KI-Modelle besteht darin, das KI-Modell in einem **breiten Spektrum von Szenarien und demografischen Merkmalen zu testen** und **vielfältige Testprotokolle einzuführen**, um Multiplikatoren von Benachteiligung zu erkennen.

TESTING

Es reicht nicht aus, die Genauigkeit eines Modells zu testen, auch Fairness muss berücksichtigt werden.

Lösungen für gesellschaftliche Probleme entstehen nicht dadurch, dass unvollständige Metriken als technische Kontrollinstrumente genutzt werden. Die Prüfung der Fairness eines KI-Modells mit Hilfe von Fairness-Metriken sollte ein selbstkritischer Prozess sein.

Wichtiger als die Verwendung einer bestimmten Fairness-Metrik, um ein KI-Modell als fair zu deklarieren, ist es, die vielen Menschen, die von dem KI-Modell betroffen sind, im Auge zu behalten und ihre Erfahrungen zu berücksichtigen.

ANWENDUNG & EVALUATION

Wenn ein KI-Modell unerklärbar bleibt, hat die Gesellschaft kein Mitspracherecht. Aber bei Fairness geht es darum, Macht zu teilen!

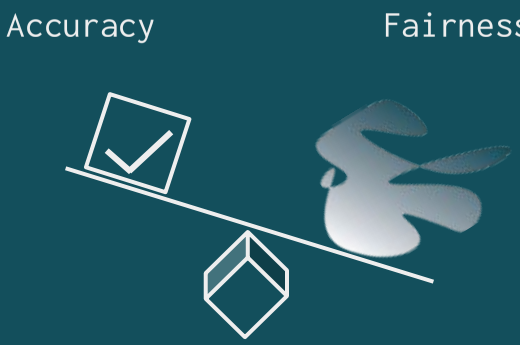
Erklärbare KI stellt sicher, dass Bewerber*innen eine Erklärung erhalten. **Open Sourcing** des Algorithmus für Crowdsourcing-Tests kann helfen, eine Vielzahl von Perspektiven einzubringen. Beides sind Ansätze zur Verbesserung der Fairness von KI-Systemen.

FAIRNESS MESSEN?

Alle Fairness-Metriken basieren auf binären Kategorien und erschweren die Berücksichtigung der Nuancen verschiedener Lebensrealitäten. Es gibt zwei Ansätze für Kritik an Fairness-Metriken:

GENAUIGKEITSBEZOGENE KRITIK

Der Versuch, Bias zu verringern, macht KI-Modelle weniger genau.



Existierenden Bias zu reduzieren, bedeutet, Datensätze zu verändern. Diese Veränderungen könnten zwar zu mehr Fairness, aber auch zu geringerer Genauigkeit (engl. accuracy) führen. Es muss einen Abwägungsprozess geben, welche Kriterien für KI-Modelle angestrebt werden.

INTERSEKTIONAL BEGRÜNDETE KRITIK

Die Verwendung von Metriken zur Messung von Fairness ist zu eindimensional.

Fairness entzieht sich Metriken

Klasse, Geschlecht, Ethnie und andere individuelle Merkmale sind nicht binär, sie überschneiden und überlappen sich. Fairness – die Gleichbehandlung verschiedener Gruppen – ist da nicht immer der richtige Ansatz. Wenn es um Gerechtigkeit geht, dann muss auch Gerechtigkeit im Sinne gleicher Zugangsmöglichkeiten (Equity) eine Rolle spielen. Diese Konzepte sind fließend und lassen sich nicht in Formeln fassen. Wichtiger als Metriken ist Anerkennung: Haben die betroffenen Menschen ein Mitspracherecht in den Prozessen?

Es reicht nicht aus, Fairness-Metriken zu verwenden, um Fairness zu messen. Um zu einer gerechteren KI zu gelangen, ist es unumgänglich, Fairness in jedem Schritt der KI-Entwicklung zu berücksichtigen. Erfahre auf den Zine-Seiten mehr dazu, wie das geht.

IDEE · UMSETZUNG · I. Mastenbroek, I. Kühnlein, B. Lübbert DESIGN · I. Kühnlein, I. Mastenbroek, L. Wunderlich

NERD NOTIZEN:

Das Goodhart'sche Gesetz wird oft so zitiert: **„Wenn eine Metrik zu einem Ziel wird, hört sie auf, eine gute Metrik zu sein."**
Was könnte das in unserem Zusammenhang mit KI und Fairness-Metriken bedeuten?

Es unterstreicht die Gefahr, dass eine bestimmte Metrik missbraucht werden kann, sobald sie zu einem Kontrollinstrument wird. Wenn ein KI-Modell für einen bestimmten quantifizierbaren Begriff von Fairness optimiert wird, bedeutet das dann wirklich, dass das KI-System fairer wird?

Es stellt auch unsere Motivation für faire KI-Systeme infrage: Verstehen wir Fairness als ein messbares, überschaubares Konzept – oder sind wir offen dafür, die komplexen Prozesse hinter Diskriminierung zu verstehen, zu überdenken und wirklich zu verändern?