

Fairness und Diskriminierungsfreiheit aus Sicht von Ethik und Informatik

Seminar im Sommersemester 2020

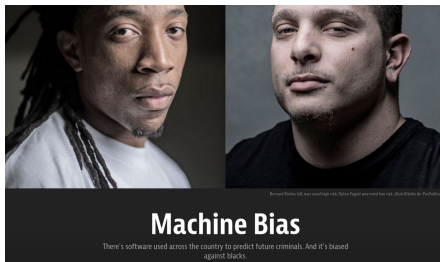
Auftaktveranstaltung | 20. April 2020

INSTITUT FÜR THEORETISCHE INFORMATIK & INSTITUT FÜR PHILOSOPHIE



This image is a modified version of "Scales of Justice - Frankfurt Version" by Michael Gouffier (CC BY-NC 3.0 from 22 Sep 2012 via Flickr).

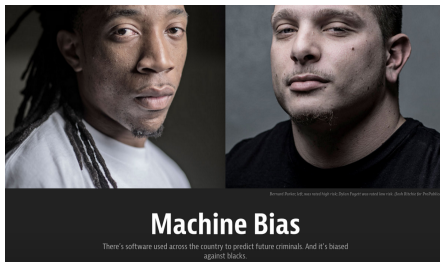
Motivation: Das COMPAS-Beispiel¹



- Prognosesoftware zur Rückfallwahrscheinlichkeit von Straftätern
- Basiert auf Antworten zu 137 Fragen durch Angeklagte und Strafregister
- Fand in vielen Gerichtsurteilen der USA für mehrere Jahre Verwendung

¹ <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>

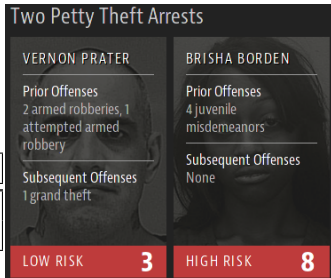
Motivation: Das COMPAS-Beispiel¹



- Prognosesoftware zur Rückfallwahrscheinlichkeit von Straftätern
- Basiert auf Antworten zu 137 Fragen durch Angeklagte und Strafregister
- Fand in vielen Gerichtsurteilen der USA für mehrere Jahre Verwendung

- Attestierte Präzision von 68 Prozent (69% für Weiße, 67% für Afro-Amerik.)
- Jedoch ...

	Weiß	Afro-Am.
„Hohes Risiko“, kein Rückfall	23,5 %	44,9 %
„Geringes Risiko“, Rückfall	47,7 %	28,0 %



¹ <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>

Diskriminierung durch (maschinell-gelernte) Entscheidungsverfahren

Diskriminierung durch (maschinell-gelernte) Entscheidungsverfahren

Hierzu Fragestellungen an der Verbindung zwischen . . .

- **Praktischer Philosophie:** Ethik
- und **Theoretischer Informatik:** Formale Logik & Kryptographie

Diskriminierung durch (maschinell-gelernte) Entscheidungsverfahren

Hierzu Fragestellungen an der Verbindung zwischen . . .

- **Praktischer Philosophie:** Ethik
- und **Theoretischer Informatik:** Formale Logik & Kryptographie

Beispielhafte Fragestellungen:

- Unter welchen Bedingungen spricht man von Diskriminierung?
- Sind faire Entscheidungen überhaupt möglich bzw. nützlich?
- Wie kann man sich sicher sein, dass ein Algorithmus fair handelt?
- Was bedeutet Fairness genau? Ist das eindeutig?
- Wie kann man Unfairness abmildern?

- Konkreter Ablauf und Form werden aktuell noch diskutiert
- Heute Test der Technik und Kennenlernrunde

- Konkreter Ablauf und Form werden aktuell noch diskutiert
- Heute Test der Technik und Kennenlernrunde
 - Wer sind Sie und was studieren Sie?
 - Was ist Ihre Motivation für das Seminar?
 - Wie möchten Sie das Seminar anrechnen? (Inf./Phil./SQ)

- Konkreter Ablauf und Form werden aktuell noch diskutiert
- Heute Test der Technik und Kennenlernrunde
 - Wer sind Sie und was studieren Sie?
 - Was ist Ihre Motivation für das Seminar?
 - Wie möchten Sie das Seminar anrechnen? (Inf./Phil./SQ)
- Bisher angedachtes Seminarformat ist Mischform:
 - 1 Erarbeitung und ausführl. Diskussion eines Textes (Phil.-Format)
 - 2 Präs., Einordnung & Diskussion verschiedener Papiere (Inf.-Format)
- In jedem Fall selbstst. Literaturrecherche und anschl. Ausarbeitung
- Außerdem voraussichtlich Mid-Term-Treffen mit Kurzpräsentationen
- Vorber. jeweiliger Beiträge unter regelm. Besprechung mit Betreuer

Zwei Terminvorschläge:

- Donnerstag, 23.04., 11.30 - 13.00 Uhr
- Donnerstag, 23.04., 13.00 - 14.00 Uhr

Wann haben Sie Zeit?

Zwei Terminvorschläge:

- Donnerstag, 23.04., 11.30 - 13.00 Uhr
- Donnerstag, 23.04., 13.00 - 14.00 Uhr

Wann haben Sie Zeit?

Donnerstag, 23.04., 11.30 - 13.00 Uhr