Ethische KI

KI-Forum, HTW Berlin

Linda Fernsel / 15. November 2023

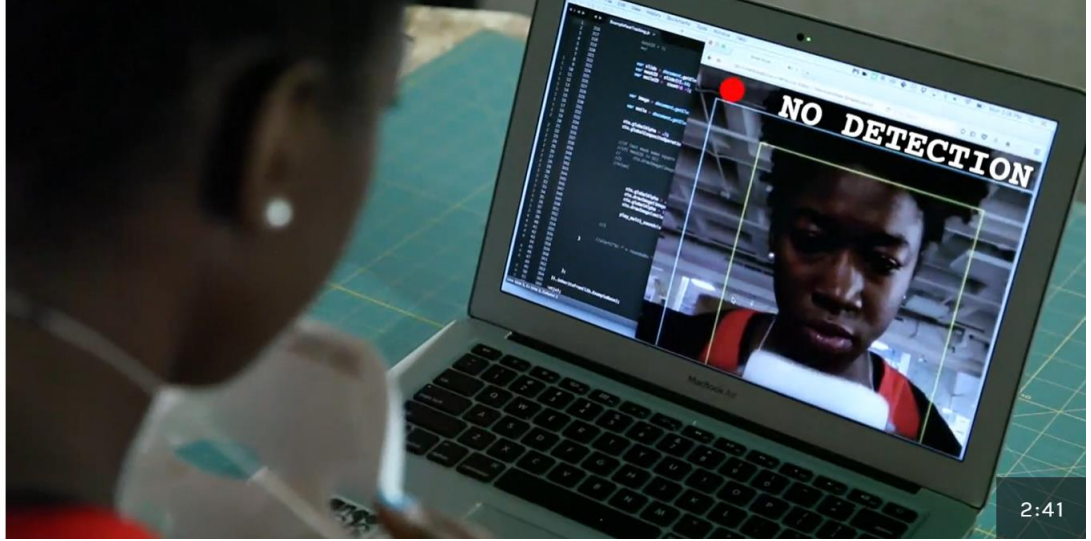
iug.htw-berlin.de/projekte/fair-enough/

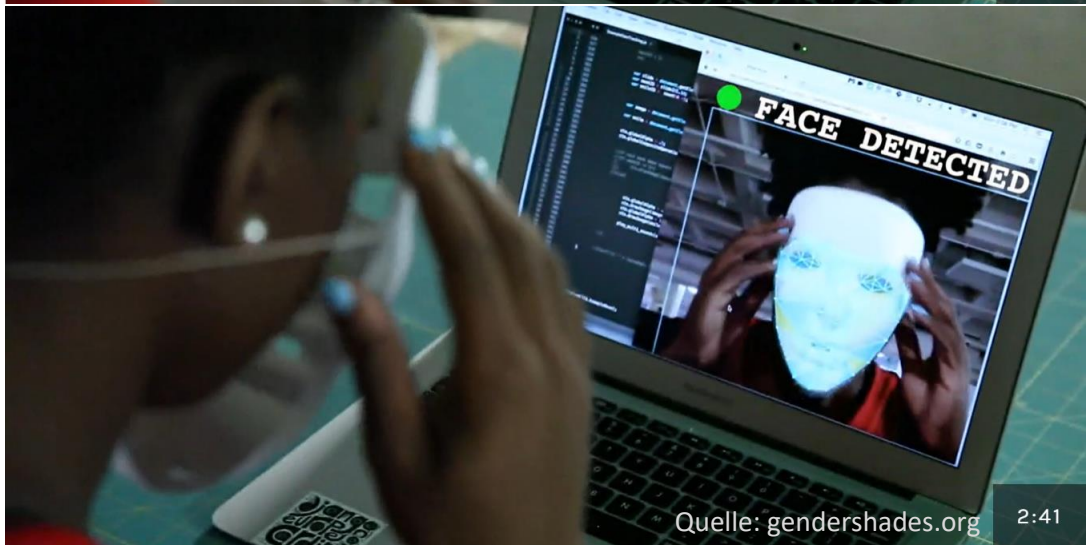
Hochschule für Technik
und Wirtschaft Berlin
University of Applied Sciences

Gender Shades

Buolamwini, Joy and Gebru, Timnit (2018). *Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification.*

gendershades.org





KI für Gesichts-klassifizierung getestet

Gender Classifier	Darker Male	Darker Female	Lighter Male	Lighter Female	Largest Gap
Microsoft	94.0%	79.2%	100%	98.3%	20.8%
FACE++	99.3%	65.5%	99.2%	94.0%	33.8%
IBM	88.0%	65.3%	99.7%	92.9%	34.4%

Ausbalanciertes Benchmark-Datenset erstellt

Quelle: gendershades.org

Wenn eine KI als Lösung verwendet wird, soll sie ethisch sein.

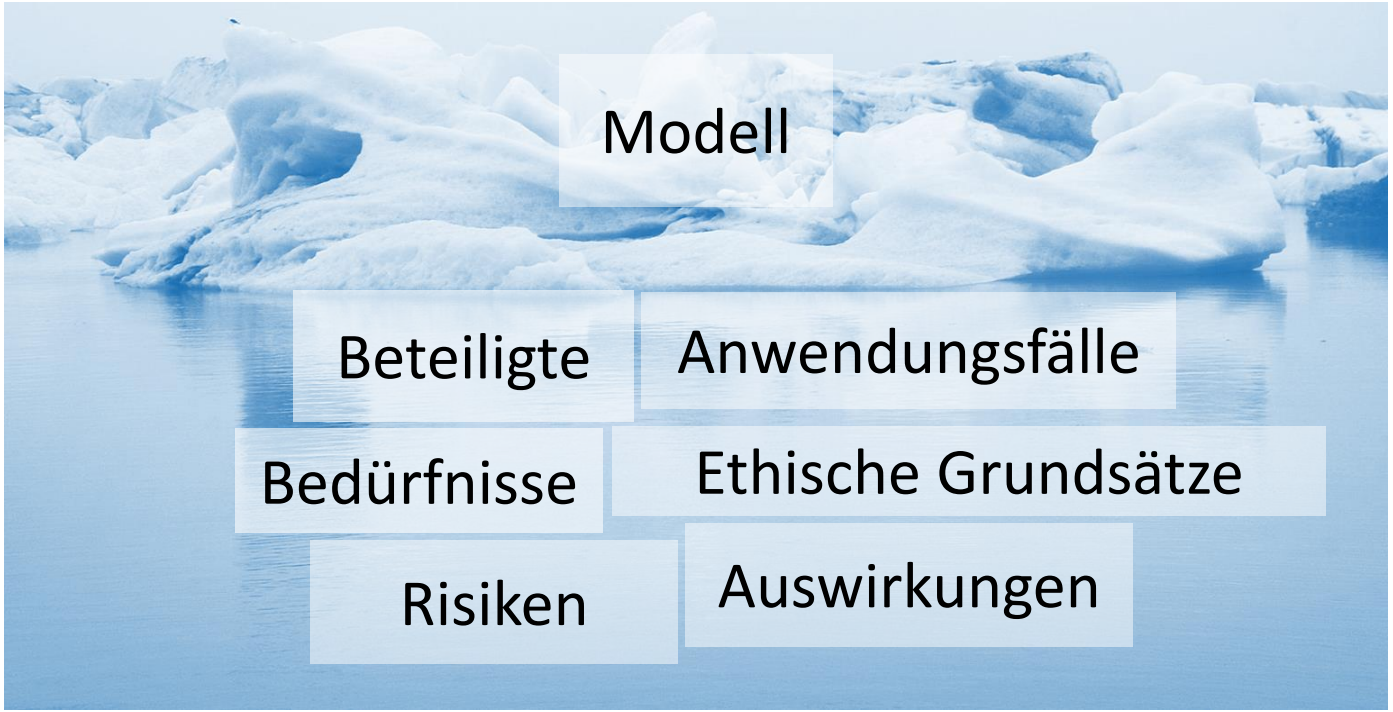
Wie erhalten wir eine ethische KI?

1. Kontext mitdenken

2. Anforderungen erweitern

3. KI auditieren

1. Kontext mitdenken



2. Anforderungen erweitern

Accuracy
Precision

Robustheit
Transparenz
Fairness

Robustheit

- Die Annahmen, die dem Modell zu Grunde liegen, sind sinnvoll.
- Die Evaluation des Modells ist valide.
- Das Modell kann mit ungewöhnlichen Fällen umgehen.
- Das Modell wird regelmäßig aktualisiert um auch mit unvorhergesehene Änderungen der Anwendungsfälle umgehen zu können.
- Es ist schwierig, das Modell auszutricksen.

Transparenz

- Die Evaluation des Modells ist nachvollziehbar.
- Die Beteiligten können nachvollziehen, welche Faktoren die Entscheidung des Modells beeinflusst haben.
- Die Beteiligten können erkennen, wie sicher sich das Modell in der Vorhersage ist.
- Alle Beteiligten wissen, dass das Modell verwendet wird.
- Alle Beteiligten wissen, welche ihrer und wie ihre Daten verwendet werden.

Fairness

- Der Einsatz des Modells ist mit geltenden ethischen Grundsätzen (z.B. Gesetze, Hochschulrichtlinien) vereinbar.
- Die Verwendung des Modells ist angemessen in Anbetracht seiner Risiken für die Beteiligten, die Gesellschaft und die Umwelt.
- Die Risiken werden bei der Entwicklung und dem Einsatz beachtet.
- Beteiligte mit unterschiedlichen Bedürfnissen werden bei der Entwicklung miteinbezogen.
- Das Modell benachteiligt keine Personengruppen.
- Die Entscheidungen des Modells können von Hand korrigiert werden.

3. KI auditieren

Anforderungen

Welche gelten?

Wie überprüfen?

Werden erfüllt?

Üben wir das.

Aufgabe: Audit

Bildet Zweierteams.

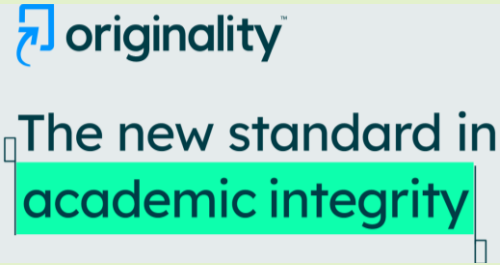
Wählt je einen Use Case und ein Anforderungspaket aus.

Überlegt, wie ihr die Anforderungen prüfen könnt.

Diskutiert, ob die Anforderungen erfüllt werden.

Use Cases

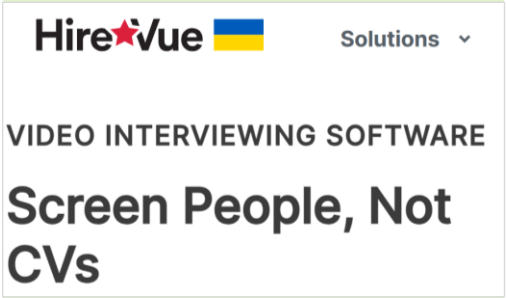
A



Turnitin Originality

bit.ly/usecaseto

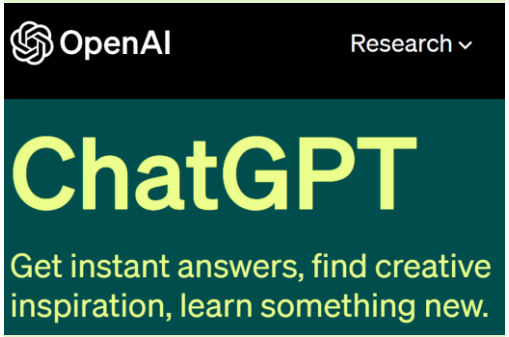
B



HireVue Video Interviewing Software

bit.ly/usecasehv

C



OpenAI ChatGPT

bit.ly/usecasegpt

Anforderungen

Konzept 1

1. Vereinbar mit Grundsätzen
2. Beachtet unterschiedliche Bedürfnisse
3. Sinnvolle Annahmen

Entwicklung 2

1. Entwicklung beachtet Risiken
2. Regelmäßig aktualisiert

Evaluation 3

1. Nachvollziehbare Evaluation
2. Valide Evaluation
3. Beachtet ungewöhnliche Fälle
4. Schwer auszutricksen

Einsatz 4

1. Keine Benachteiligung
2. Verwendung bekannt
3. Datenverwendung bekannt

Mensch-KI-Interaktion 5

1. Nachvollziehbare Ergebnisse
2. Unsicherheit transparent
3. Überschreibbar

Risiken 6

1. Angemessen in Anbetracht der Risiken

Audit-Resultate

Fazit

KIs sollen ethisch – also robust, transparent und fair – sein.

Dafür müssen wir (1) KIs in ihren Kontexten denken, (2) um ethische Aspekte erweiterte Anforderungen stellen und (3) KIs auditieren.

Voraussetzung: Die KI muss auditierbar sein.

Vielen Dank.

Linda Fernsel
Projekt „Fair Enough?“
Forschungsgruppe Informatik und Gesellschaft
iug.htw-berlin.de/projekte/fair-enough/
fernsel@htw-berlin.de

www.htw-berlin.de

Hochschule für Technik
und Wirtschaft Berlin
University of Applied Sciences