

>>> Smart Borders?

>>> Wie die EU versucht Grenzübergänge mit einem
diskriminierenden KI-Lügendetektor zu regulieren.

Name: verschiedene[†]

Date: September 1, 2021

[†]AG Link

>>> Inhalt

1. Was ist iBorderCtrl? (15min)

Akteure und Organisationsstruktur
Silent Talker
Geschichte des Lügendetektor

2. Grundlagen KI (15min)

Algorithmen, KI und Neuronale Netze
Overfitting

3. Grenzen von KI / Grenzen und Risiken von KI (20min)

KI und Bias
KI und Interpretierbarkeit

4. Bias Bingo mit / Bias in iBorderCtrl (20min)(interaktiv?)

5. Ausblick und Diskussion(20min)

Politische Einordnung
Wie und wofür forschen wir?

>>> Akteure und Organisationsstruktur

- * Horizon 2020 (auch Roborder)
- * Tresspass etc.
- * Finanzierung
- * Beteiligte Forschungseinrichtungen, beteiligte Unternehmen?
- * Aktueller Entwicklungsstand

>>> Silent Talker



Quelle: iborderctrl.eu

"The avatar is presented in a uniform to convey an air of authority." (K.Crockett et.al.)

>>> Geschichte des Lügendetektor

- * Polygraph (Genauigkeit: beinahe zufällig (Saxe, Ben-Shakhar, 1999))
- * Mirco Expressions etc.

Infos in: The politics of deceptive borders: biomarkers of deceit and the case of iBorderCtrl

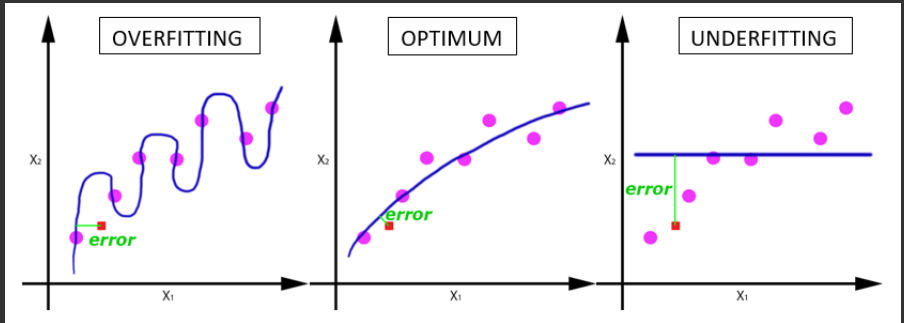
>>> Algorithmen, KI und Neuronale Netze

- * Algorithmus
- * KI / Machine Learning
- * Neuronale Netze und Deep Learning

>>> Aufbau der Datensätze

- * Supervised learning, Datenpaare
- * Training, Validation and Test Data

>>> Overfitting



Quelle: Sagar Sharma / Towards Data Science

>>> Arten von Bias

Bias = Verzerrung

- * Bias in den Daten
- * Bias durch Design des Algorithmus
- * Bias durch Rückkopplung im Gebrauch

Konsequenz: Diskriminierende Algorithmen (auch Gender Bias, Racial Bias, Neurodiversity Bias etc. genannt)

>>> Bias in den Daten

- * Measurement Bias: Wie werden bestimmte Eigenschaften gemessen/bestimmt?
 - * COMPAS: Verhaftungen auch von Familie etc. wurden genutzt um Risiko zu bewerten
- * Omitted Variable Bias: Wichtige Daten werden nicht im Modell berücksichtigt
 - * Ein Programm das die Kündigung von Abos (zB Netflix) prognostiziert, aber neue Konkurrenzangebote nicht beachtet
- * Representation Bias: Fehlende Diversität in den verfügbaren Daten
 - * Coded Gaze, Joy Buolamwini
- * Aggregation Bias: Spezifische Eigenschaften von Minderheiten gehen im gesamten Datensatz unter/Ableiten von Aussagen über Individuen aus Minderheit aus allgemeinem Datensatz
 - * Diskriminierungsklage gegen die Universität Berkeley

>>> Bias durch Design

- * Evaluation Bias: Ergebnis wird an unrealistischen Kriterien gemessen
 - * Bias in Gesichtserkennung fällt nicht auf, weil Benchmarkdatenset gebiast ist

>>> Bias durch Rückkopplung

* Beispiel Profiling

>>> Interpretierbarkeit

- * Was heißt interpretierbar?
- * Warum will mensch das?
- * Blackbox-Argument
- * Betriebsgeheimnis-Argument
- * Missklassifikationshack

>>> Bias in iBorderCtrl

- * Measurement Bias: Schauspieler*innen statt echte Situationen, Stress beim Grenzübergang löst u.U. gleiche Symptome wie Lügen aus (?)
- * Omitted Variable Bias: Lügen zeigt sich in nichterfassten Phänomenen
- * Representation Bias: fehlende Diversität race, gender, neurdiversity, disability, health, scars
- * Aggregation Bias: s. representation bias, teilweiser Versuch der Gegensteuerung
- * Overfitting: 73 Prozent in Testdaten vs 93 Prozent in Trainingsdaten, das impliziert Overfitting ist wahrscheinlich

>>> Bias in iBorderCtrl

- * Evaluation Bias: Testbedingungen entsprechen nicht den Einsatzbedingungen, zB Licht, Diversität
- * Bias durch Rückkopplung: Wenn Einsatzdaten wieder eingespeist werden, dann: LD klassifiziert mehr Benachteiligte als lügend, Überprüfungsbeamt*in ebenso und gibt Daten zurück

>>> Testergebnisse

- * Anzahl verschiedener Personen in der Testdatenmenge: 1
- * Emprische Varianz der Tests:

Table IV: Classification Outcomes using Unseen Participants

Test No	Participant				Accuracy (%)	
	Truthful		Deceptive		Truthful	Deceptive
	Gender	Ethnicity	Gender	Ethnicity		
1	M	EU	M	A/A	100	57
2	M	A/A	F	EU	50	36
3	M	A/A	F	EU	50	100
4	M	EU	F	EU	90	100
5	M	A/A	M	EU	100	10
6	M	EU	M	EU	72	100
7	M	A/A	F	EU	100	100
8	F	EU	F	A/A	38	100
9	M	EU	M	EU	80	60
Overall Accuracy (%)					75.55	73.66

>>> Politische Einordnung

- * Unfreiwillige Datenerhebung zur "Verbesserung" des Algorithmus
- * Entwicklung Ethischer Normen für KI (EU KI Standards, Gesellschaft für Informatik)

>>> Wie und wofür forschen wir?

- * Instrumentalisierung von Wissenschaft
- * Was ist mein wissenschaftlicher Standard und worauf gründet er?
- * Gibt es Ziele und Werte die nicht von der Wissenschaft vorgegeben werden, sondern die wir uns selbst setzen müssen?