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

Name: verschiedene[†]
Date: August 19, 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)**
   - Beispiele diskriminierender KI
   - 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

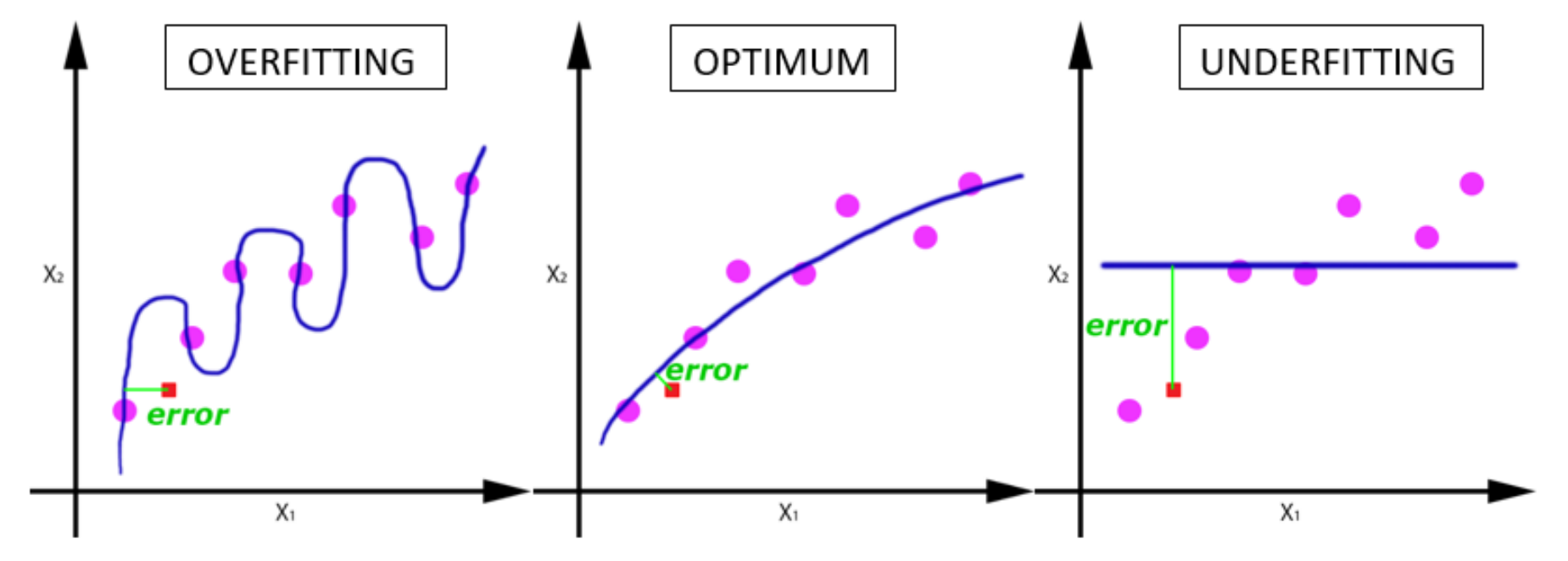
>>> Algorithmen, KI und Neuronale Netze

* 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

>>> Beispiele diskriminierender KI

* COMPAS - Risikoabschätzung Gefangener in Florida
  * Schwarze Gefangene wurden doppelt so häufig falsch als gewalttätig klassifiziert
* Vermittlung von Stellenanzeigen im MINT-Bereich
  * Stellenanzeigen wurden häufiger Männern (cis?) angezeigt.
* Einstellungsverfahren (konkreter Fall amazon?)
* Gesichtserkennung, Stimmerkennung, etc. (konkrete Fälle)

>>> 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

- Wie werden bestimmte Eigenschaften gemessen/bestimmt? (Measurement Bias)
  - COMPAS: Verhaftungen auch von Familie etc. wurden genutzt um Risiko zu bewerten
- Wichtige Daten werden nicht im Modell berücksichtigt (Omitted Variable Bias)
  - Beispiel
- Fehlende Diversität in den Verfügbaren Daten (Represenation Bias)
  - Beispiel
- Spezifische Eigenschaften von Minderheiten gehen im gesamten Datensatz unter/Ableiten von Aussagen über Individuen aus Minderheit aus allgemeinem Datensatz (Aggregation Bias)
  - Beispiel
- Ungleiche Datenmenge verschiedener Untergruppen (Sampling Bias)
  - Beispiel

>>> Bias in den Daten

* Historical Bias (Erklärung, Beispiel)
* Population Bias (Erklärung, Beispiel)
* etc.

>>> Bias durch Design

* Algorithmischer Bias (Erklärung, Beispiel)
* Evaluations Bias (Erklärung, Beispiel)
* etc.

>>> Bias durch Rückkopplung

Beispiel (Profiling?)

>>> Interpretierbarkeit

* Was heißt interpretierbar?
* Warum will mensch das?
* Blackbox-Argument
* Betriebsgeheminis-Argument
* Beispiel: Hack zum Missklassifizieren

>>> Bias in iBorderCtrl

* Measurement Bias: Schauspieler*innen statt echte Situationen
* Omitted Variable Bias: Nervosität durch Stress beim Grenzübergang
* Representation Bias: fehlende Diversität race, gender, neurdiversity, disability, health, scars
* Aggregation Bias: s. representation bias, teilweiser Versuch der Gegensteuerung
* Sampling Bias: ungleiche Datenmengen
* Overfitting: 73 Prozent in Testdaten vs 93 Prozent in Trainingsdaten, das impliziert Overfitting ist wahrscheinlich
* Historical Bias: höhere Wahrscheinlichkeit zB des Drogenschmuggels bestimmter Gruppen (trifft erst bei Benutzung zu, weil Testdaten zu klein)
* Algorythmischer Bias: zu wenig Einblick
* Evaluations Bias: Testbedingungen entsprechen nicht den Einsatzbedingungen, zB Licht, Diversität
* Bias durch Rückkopplung:

>>> 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, The Ethical Algorithm von Kearns, Roth?)

>>> Wie und wofür forschen wir?

* Instrumentalisierung von Wissenschaft
* https://youtu.be/f9WkKKZXvgA?t=3411
* Was ist mein wissenschaftlicher Standard und worauf gründet er?
* Gibt es Ziele und Werte die nicht von der Wissenschaft vorgegen werden, sondern die wir uns selbst setzen müssen?
* Hinweis auf Transparenzinitiative der Senatorin Johanna