>>> Smart Borders?

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

Name: AG Link

Date: September 13, 2021

>>> Inhalt

1. Was ist iBorderCtrl?
   - Akteure und Organisationsstruktur
   - Silent Talker
   - Geschichte des Lügendetektor
2. Grundlagen KI
   - KI, Algorithmen und maschinelles Lernen
   - Overfitting
3. Grenzen und Risiken von KI
   - KI und Interpretierbarkeit
   - KI und Bias
4. Bias in iBorderCtrl
5. Ausblick und Diskussion
   - 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.)

>>> Die Geschichte des Lügendetektors - Der Polygraph

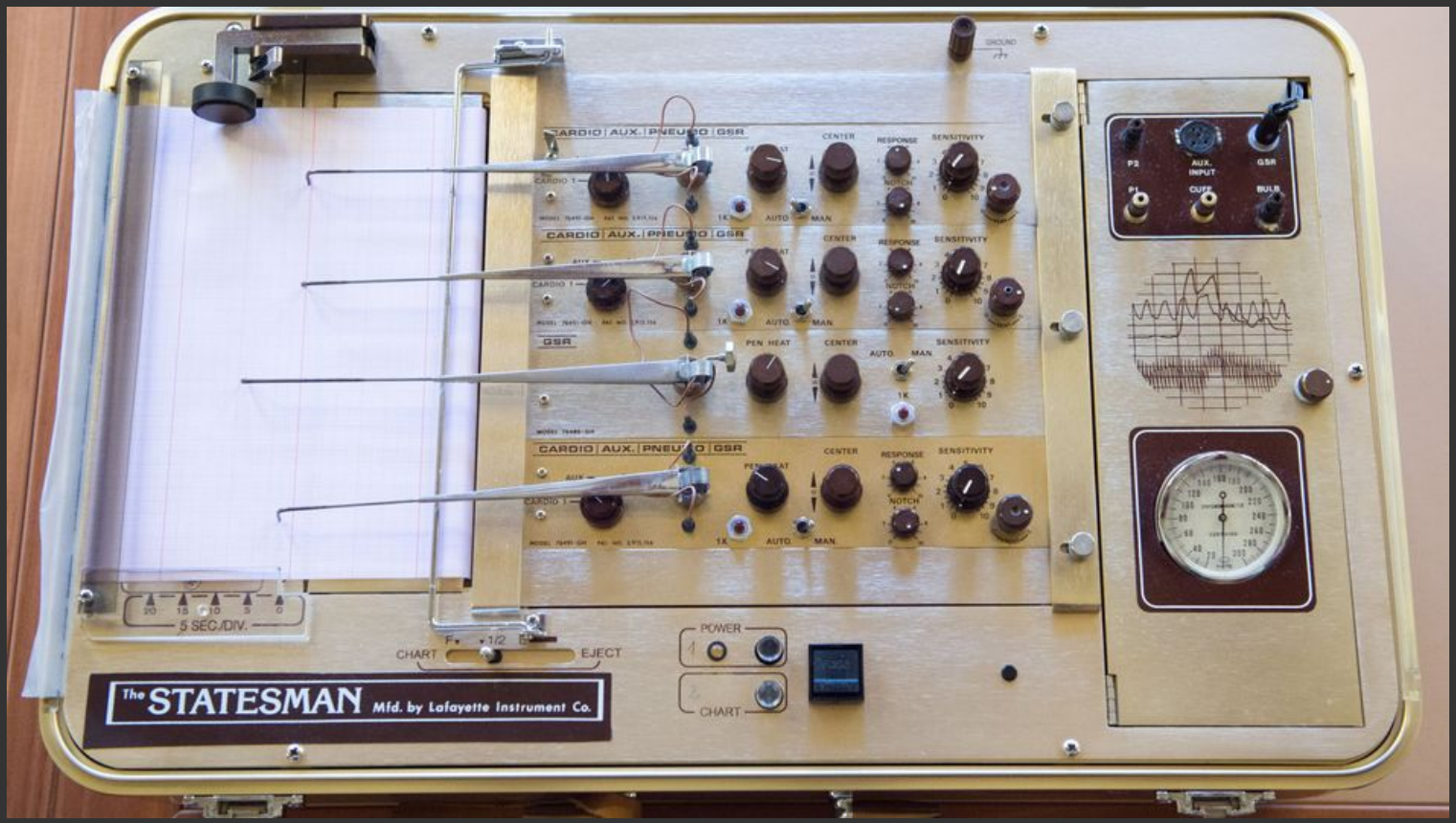


Foto: Sebastian Kahnert/ dpa

>>> Die Geschichte des Lügendetektors - Der Polygraph

* Entwicklung in der ersten Hälfte des 20. Jahrhunderts
* Maßgeblich von John A. Larson (Polizist und Physiologe) geprägt
* Später patentiert und vermarktet
* Forschung umstritten
* Einige Studien legen nahe, dass Ergebnisse beinahe zufällig sind (Saxe, Ben-Shakhar, 1999)
* Andere Systeme, die physiologische Reaktionen testen ähnlich umstritten

>>> Die Geschichte des Lügendetektors - Mikroexpressionen

Paul Ekman's Pictures of Facial Affect

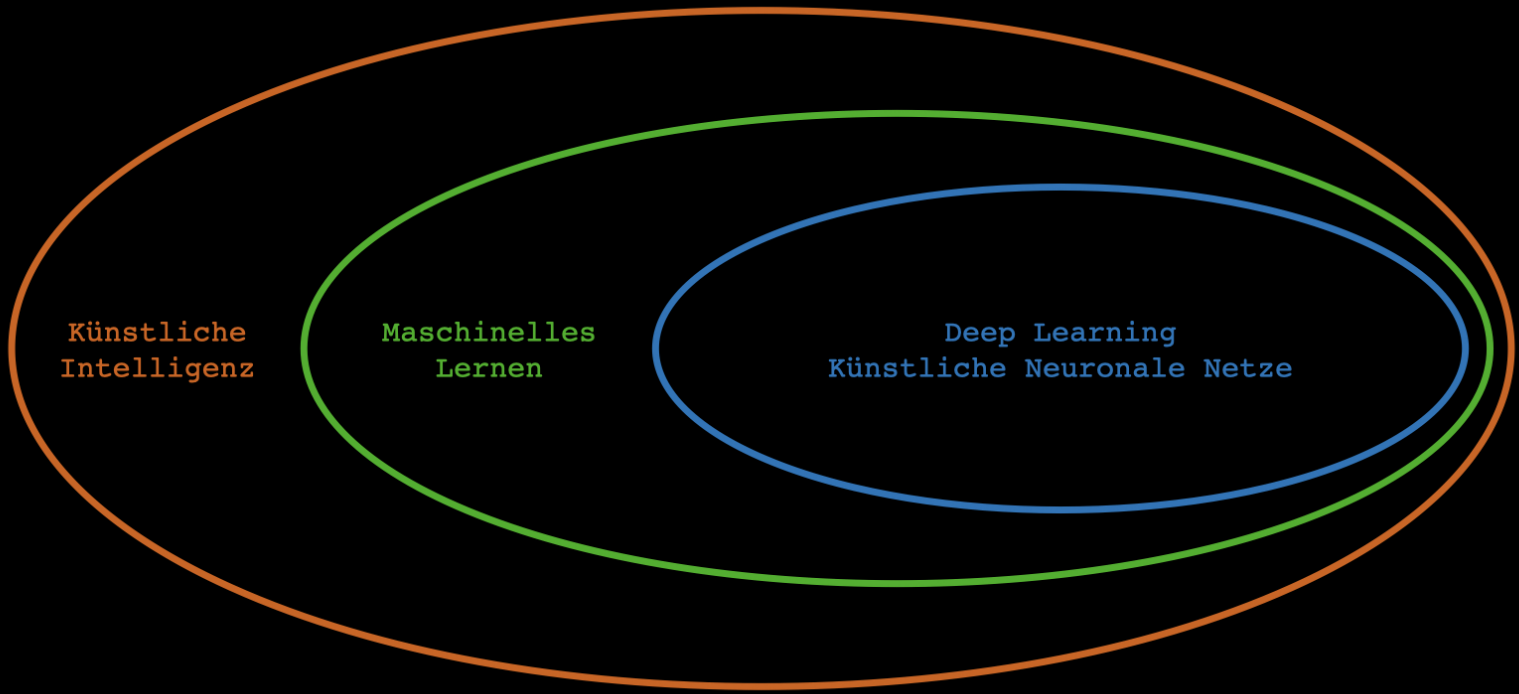
>>> Die Geschichte des Lügendetektors - Mikroexpressionen

* Facial Action Coding System (FACS) von Ekman und Friesen entwickelt,
* 'Mikroexpressionen' codieren Gefühle des Menschen,
* Annahme: Lügen ist emotionaler Akt und kann mit Mikroexpressionen entschlüsselt werden,
* Studien zeigen: Vorgehen ist so exakt wie zufällige Vermutung,
* Forschung ist uneinig, ob Mikroexpressionen bei jedem Menschen vorliegen und eindeutig sind.

>>> Die Geschichte des Lügendetektors - Die Rolle von KI

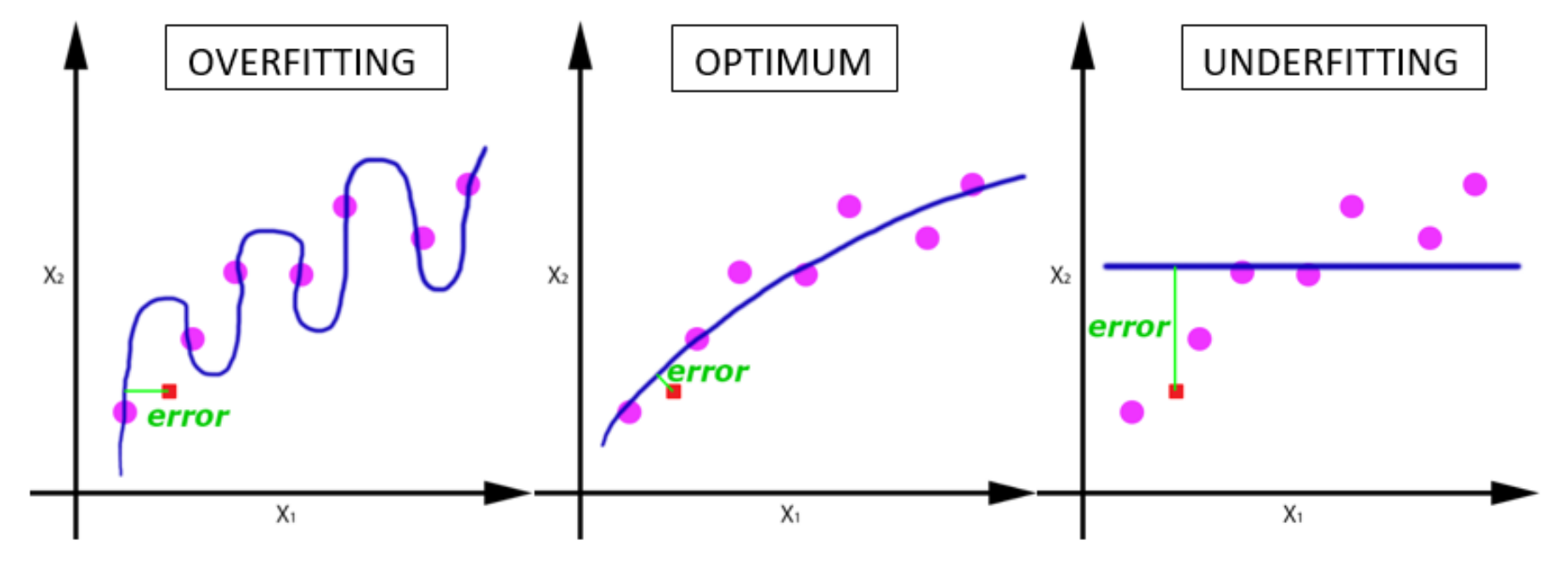
* Versuche Mikroexpressionen und ähnliches mit KI zu klassifizieren,
* Nutzung für die Optimierung von Werbeanzeigen (Facebook),
* Projekt AVATAR (US-Projekt für Grenzkontrollen)

>>> KI, Algorithmen und maschinelles Lernen



Algorithmus: Rechenvorgang nach einem bestimmten Schema

>>> Overfitting



Quelle: Sagar Sharma / Towards Data Science

>>> Was heißt interpretierbar?

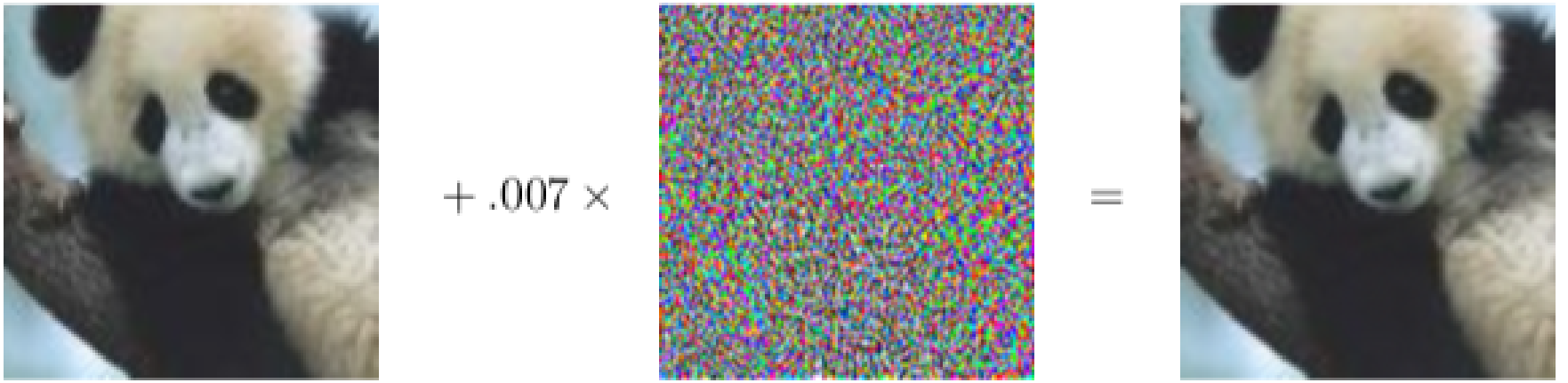
* Keine einheitliche Definition
* Interpretierbarkeit als "Verstehen des Algorithmus":
  * Bedeutung der Parameter kann verstanden werden
  * Funktionsweise des Algorithmus kann für jeden Input nachvollzogen werden
  * Kann bei KI oft nicht angewendet werden
* Interpretierbarkeit durch methodische Analyse:
  * Verschiedene Techniken
  * Benutzung weiterer KI
  * Analyse durch Manipulation von Datensätzen
  * Konzepte sind neu
  * Bedeutung und Korrektheit nicht geklärt

>>> Wofür braucht mensch Interpretierbarkeit?

* Kausalität vs. Korrelation, was lernt die KI?
* Bsp.: KI soll Basketbälle erkennen und wird mit Bildern unterschiedlicher Bälle trainiert. Erkennt die KI tatsächlich den Basketball oder einfach nur die Farbe Orange?
* Wenn KI in gesellschaftlich relevanten Bereichen eingesetzt wird, kann es entscheidend sein Ergebnisse interpretieren zu können

* Ergebnisse der KI hängen stark vom Datensatz ab
* Daraus ergibt sich die Möglichkeit eines Angriffs auf KIs
* Die Kriterien für die Entscheidung der KI stimmen nicht mit menschlichen Kriterien überein

>>> Panda oder Gibbon?



Quelle: J. Goodfellow et. al.

Quelle:

>>> Arten von Bias

Bias = Verzerrung

* Bias in den Daten
* Bias durch Design des Algorithmus
* Bias durch Rückkopplung

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

>>> Bias in den Daten

* **Measurement Bias:** Die Auswahl der Merkmale zur Darstellung des Sachverhalts kann zu Verzerrung führen

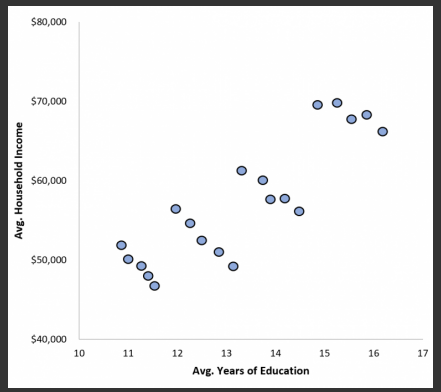
>>> Bias in den Daten

* Measurement Bias: Die Auswahl der Merkmale zur Darstellung des Sachverhalts kann zu Verzerrung führen
* Omitted Variable Bias: Wichtige Merkmale werden nicht im Modell berücksichtigt

>>> Bias in den Daten

* Measurement Bias: Die Auswahl der Merkmale zur Darstellung des Sachverhalts kann zu Verzerrung führen
* Omitted Variable Bias: Wichtige Merkmale werden nicht im Modell berücksichtigt
* Represenation Bias: Fehlende Diversität in den verfügbaren bzw. genutzten Daten

* **Aggregation Bias:** Spezifische Eigenschaften von Untergruppen gehen im gesamten Datensatz unter, Überlagerung von Trends in verschiedenen Untergruppe führt zu falschen Aussagen



Quelle: Statology.org What is Aggregation Bias?

>>> Bias durch Design

* **Evaluation Bias:** Performance des Algorithmus wird an unrealistischen Kriterien gemessen

>>> Bias durch Rückkopplung

Verstärkung der Diskriminierung durch Rückkopplungsschleife bei Predictive Policing

Quelle: Algorithmic Bias and Fairness: Crash Course AI by CrashCourse

>>> Bias in iBorderCtrl

* **Measurement Bias:** Stress beim Grenzübergang löst u.U. gleiche Symptome wie Lügen aus

>>> Bias in iBorderCtrl

* Measurement Bias: Stress beim Grenzübergang löst u.U. gleiche Symptome wie Lügen aus
* Omitted Variable Bias: Lügen/nicht Lügen zeigt sich in nicht erfassten Phänomenen.Es gibt keine zuverlässigen Theorien dazu, welche physiologischen Prozesse Lügen eindeutig identifizierbar machen.

>>> Bias in iBorderCtrl

* Measurement Bias: Stress beim Grenzübergang löst u.U. gleiche Symptome wie Lügen aus
* Omitted Variable Bias: Lügen/nicht Lügen zeigt sich in nicht erfassten Phänomenen.Es gibt keine zuverlässigen Theorien dazu, welche physiologischen Prozesse Lügen eindeutig identifizierbar machen.
* **Representation Bias:** fehlende Diversität in Hautfarbe, Geschlecht, Neurodiversität, Behinderungen, gesundheitlichen Faktoren, Narben etc.

>>> Bias in iBorderCtrl

* Measurement Bias: Stress beim Grenzübergang löst u.U. gleiche Symptome wie Lügen aus
* Omitted Variable Bias: Lügen/nicht Lügen zeigt sich in nicht erfassten Phänomenen.Es gibt keine zuverlässigen Theorien dazu, welche physiologischen Prozesse Lügen eindeutig identifizierbar machen.
* Representation Bias: fehlende Diversität in Hautfarbe, Geschlecht, Neurodiversität, Behinderungen, gesundheitlichen Faktoren, Narben etc.
* Aggregation Bias: Reaktionen (z.B. Mikroexpressionen) nicht konsistent. Gleiches Verhalten bei unterschiedlichen Individuen kann Unterschiedliches bedeuten.

>>> Bias in iBorderCtrl

* Measurement Bias: Stress beim Grenzübergang löst u.U. gleiche Symptome wie Lügen aus
* Omitted Variable Bias: Lügen/nicht Lügen zeigt sich in nicht erfassten Phänomenen.Es gibt keine zuverlässigen Theorien dazu, welche physiologischen Prozesse Lügen eindeutig identifizierbar machen.
* Representation Bias: fehlende Diversität in Hautfarbe, Geschlecht, Neurodiversität, Behinderungen, gesundheitlichen Faktoren, Narben etc.
* Aggregation Bias: Reaktionen (z.B. Mikroexpressionen) nicht konsistent. Gleiches Verhalten bei unterschiedlichen Individuen kann Unterschiedliches bedeuten.
* Overfitting: Trefferquote von 73 Prozent in Testdaten vs 93 Prozent in Trainingsdaten deutet auf Overfitting hin

>>> Bias in iBorderCtrl

* **Evaluation Bias:** Testbedingungen entsprechen nicht den Einsatzbedingungen, zB Licht, Diversität

>>> Bias in iBorderCtrl

* Evaluation Bias: Testbedingungen entsprechen nicht den Einsatzbedingungen, zB Licht, Diversität
* **Bias durch Rückkopplung:** Wenn Einsatzdaten wieder eingespeist werden, dann klassifiziert der Lügendetektor mehr Benachteiligte als lügend, Überprüfungsbeamt*innen bestätigen die Entscheidung und geben die Daten zurück an das System.

>>> Testergebnisse

* Anzahl verschiedener Personen in der Testdatenmenge: 1
* Stichprobenstreuung: 24.37 (Truthful) / 34.29 (Deceptive)

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?

* Wissenschaft kann instrumentalisiert werden
* Wir brauchen mehr kognitive Diversität in Forschungseinrichtungen
* Was ist dein wissenschaftlicher Standard und worauf gründet er?