KI-FACHKONFERENZ

Datenanforderungen an den Artificial Intelligence Act

David Berend

Frankfurt,
22.11.2021

Der AI Act gibt fundierte Erwartungen an „Trustworthy AI“ Prinzipien

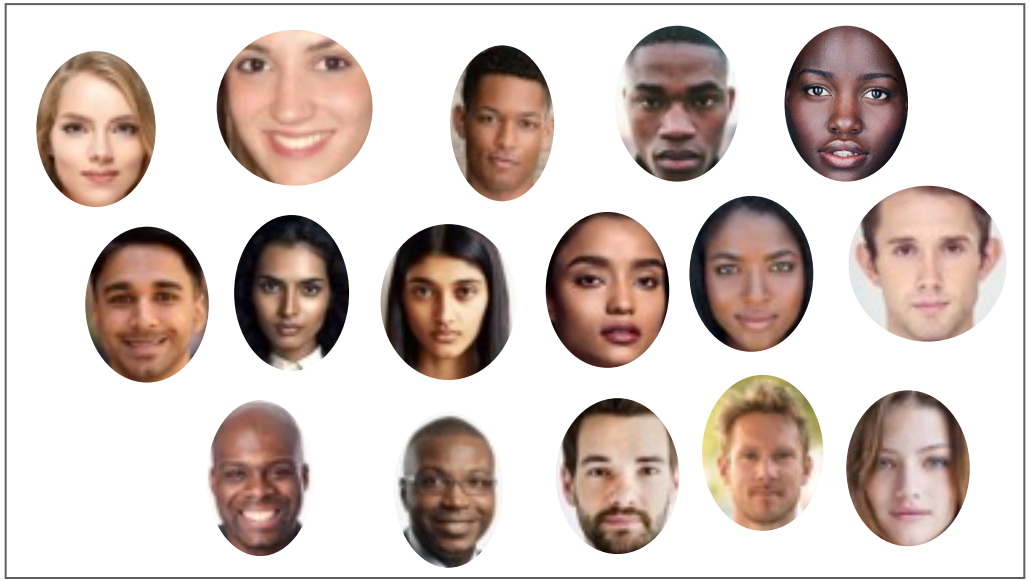
Datengetriebe Prinzipien

- Technical robustness and Safety
- Privacy & Data Governance
- Transparency
- Diversity, Non-Discrimination and Fairness
- Societal and environmental wellbeing

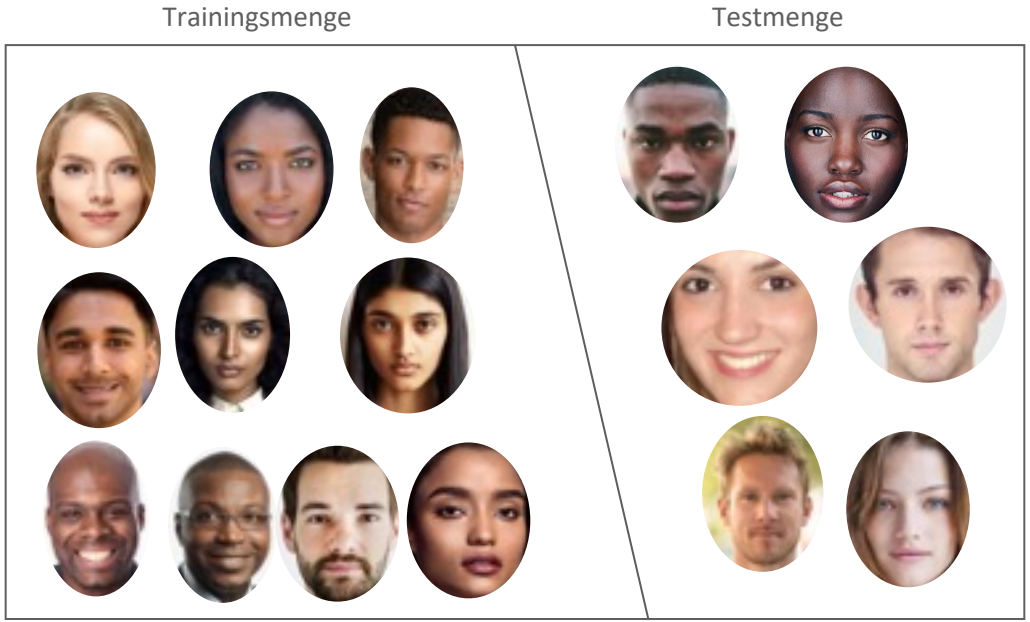
Source: European Commission, Requirements of Trustworthy AI

Jedoch ist unklar welche Daten die Grundlage bieten, um das Erfüllen der Erwartungen zu testen

Beispiel Datensatz eines Gesichtserkennungssystems

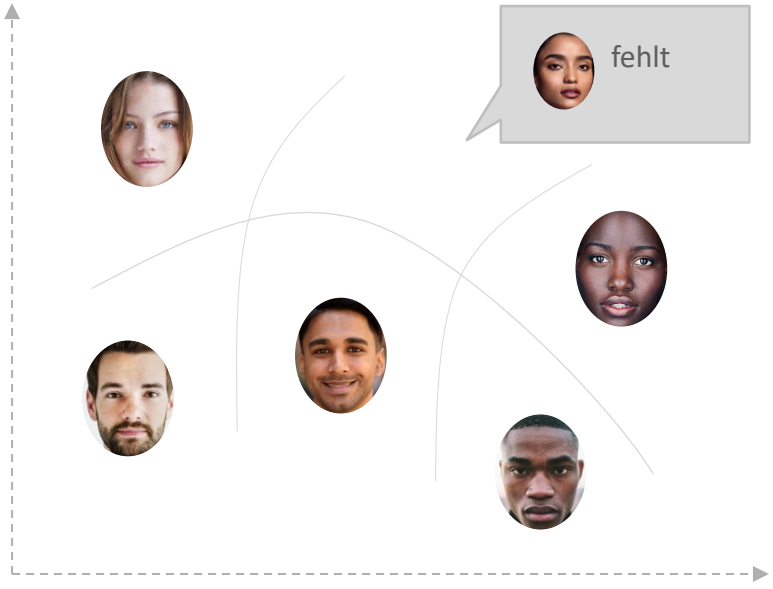


Traditionelles Testen mit einer Teilmenge des Gesamtdatensatzes ist meist nicht genug



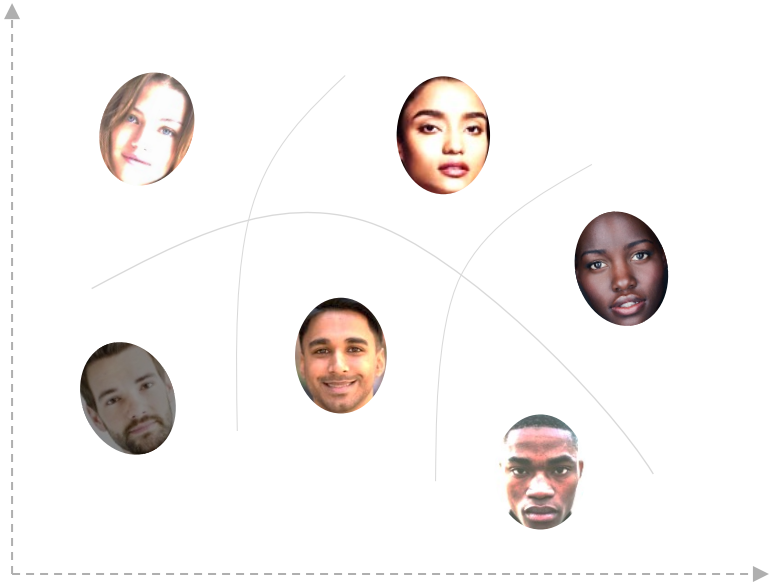
Traditionelles Testen mit einer Teilmenge des Gesamtdatensatzes ist meist nicht genug, es bedarf fundierter Methodik der Testdatenanalyse

- Testdaten sollten sämtliche Ausprägungen an Daten beinhalten, die nach Deployment relevant werden



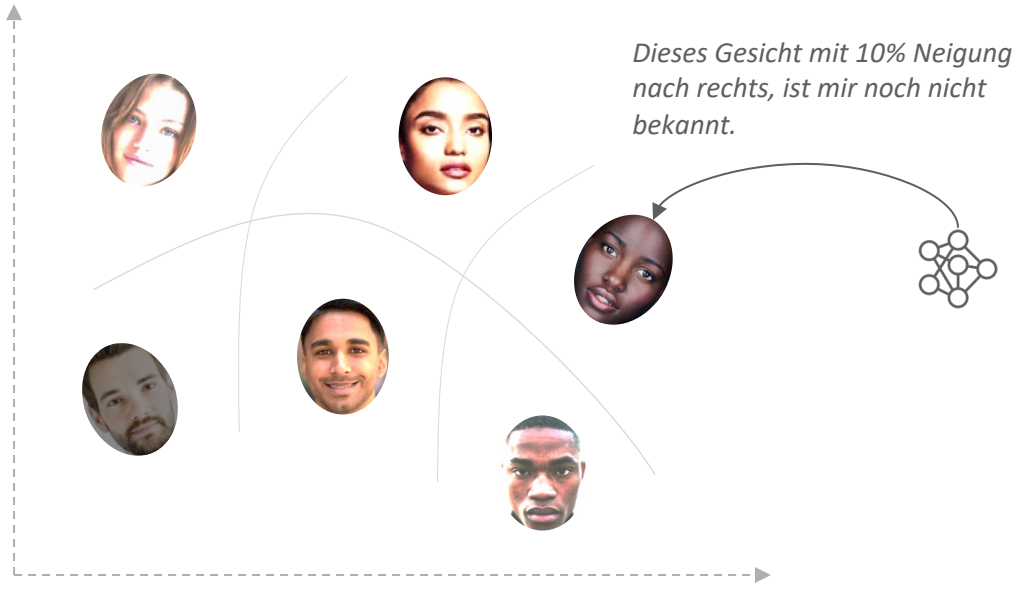
Traditionelles Testen mit einer Teilmenge des Gesamtdatensatzes ist meist nicht genug, es bedarf fundierter Methodik der Testdatenanalyse

- Testdaten sollten sämtliche Ausprägungen an Daten beinhalten, die nach Deployment relevant warden
- Diese Basis an Daten kann daraufhin vervielfacht werden, um verschiedene Situationen abzudecken



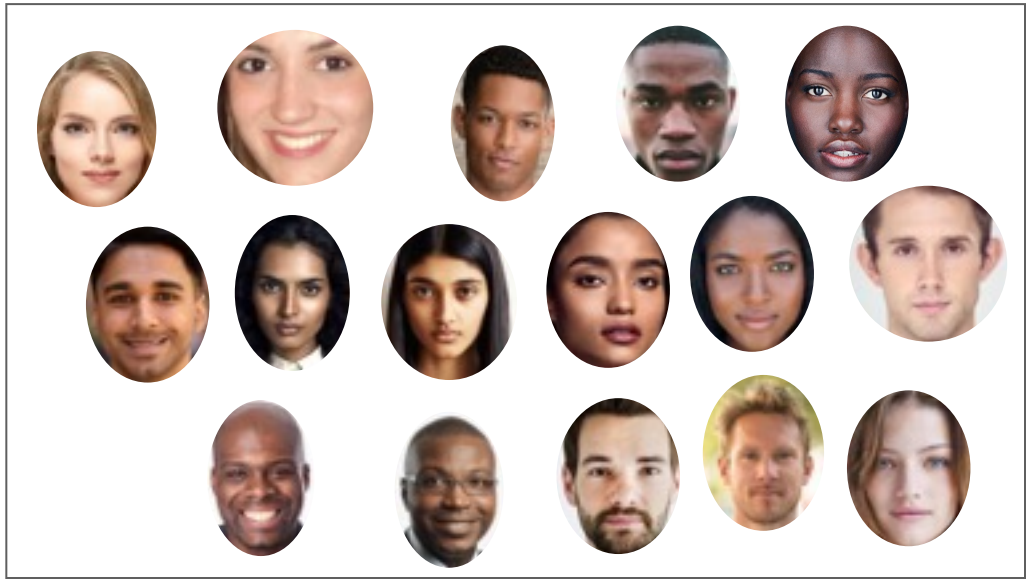
Traditionelles Testen mit einer Teilmenge des Gesamtdatensatzes ist meist nicht genug, es bedarf fundierter Methodik der Testdatenanalyse

- Testdaten sollten sämtliche Ausprägungen an Daten beinhalten, die nach Deployment relevant warden
- Diese Basis an Daten kann daraufhin vervielfacht werden, um verschiedene Situationen abzudecken
- Nicht bei Zufall, sondern mit Feedback des statistischen Modells
- Dadurch kann Inputcompleteness approximiert werden



Einmal trainiert ist nur der Anfang. Es bedarf kontinuierliches Monitoring der Input-Daten, um Änderungen in der Distribution festzustellen

Beispieldaten für das Training



Beispielinputs nach Deployment

Key takeaways

- Es Bedarf Anforderungen an
  - Statistische Analyse für ausreichende, diverse und balancierte Trainings- & Testdaten
  - Monitoring und Grenzwerte, die ein Re-Training und Re-Testen in Gang setzen

- Dabei hilfreich sind
  - Clustering Methoden
  - Datenmutation in Kombination mit…
  - Verhaltensverständnis des statistischen Modells

Vielen Dank für Ihre Aufmerksamkeit!

Ihr Ansprechpartner:

David Berend
CEO & Co-Founder VAISION AI
Co-Lead Standardisierung KI Sicherheit, Singapur

Tel.: +65 8779 5456
E-Mail: david.berend@vaision.ai