
DIGITAL-CAMP 2023
Künstliche Intelligenz und Non-Profit-Organisationen
PROJEKTTRÄGER
Haus des Stiftens
Engagiert für Engagierte
MIT FREUNDLICHER UNTERSTÜTZUNG VON
Microsoft

Diskriminierung in KI – Ursprünge & Lösungsansätze

17. Oktober 2023 | 10:00 – 11:00 Uhr

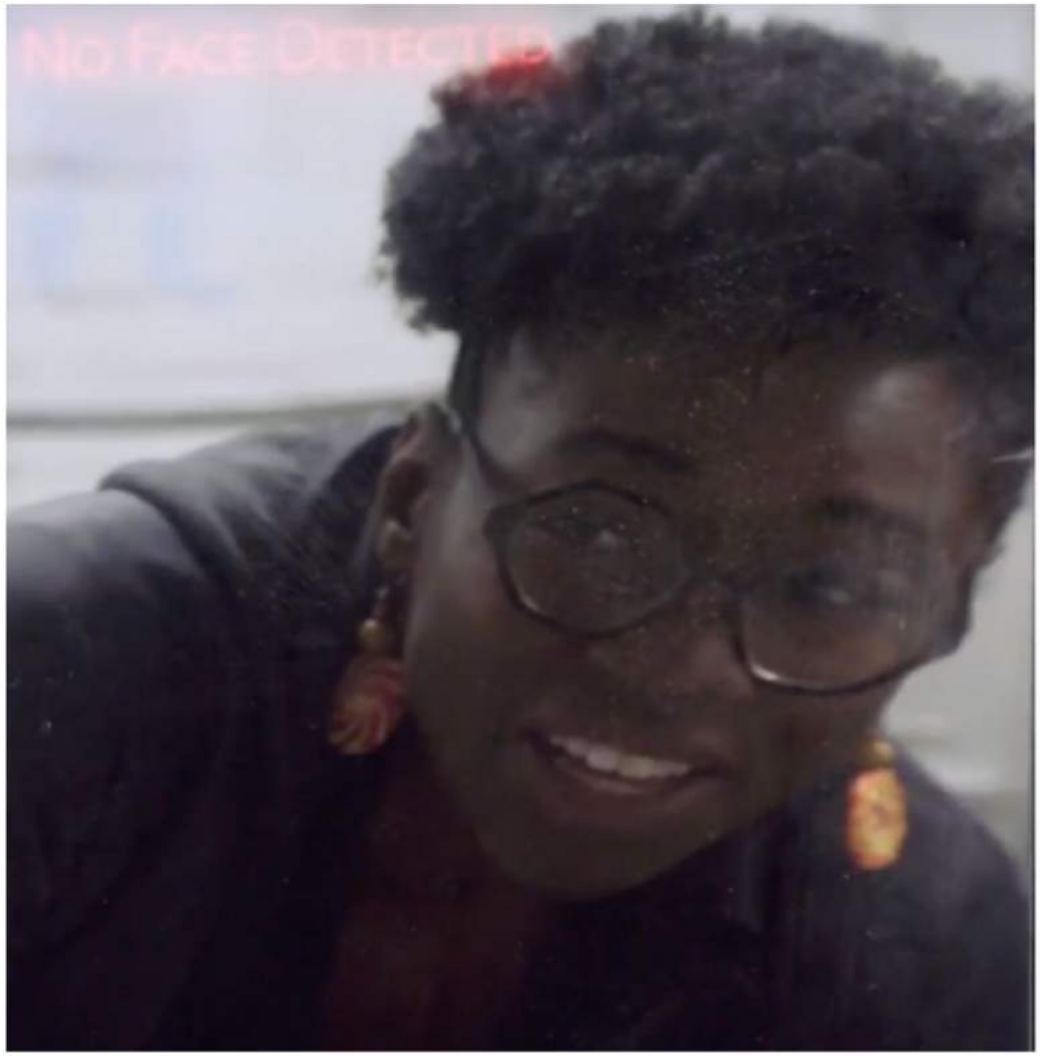
Elena A. Kalogeropoulos | iRights. Lab

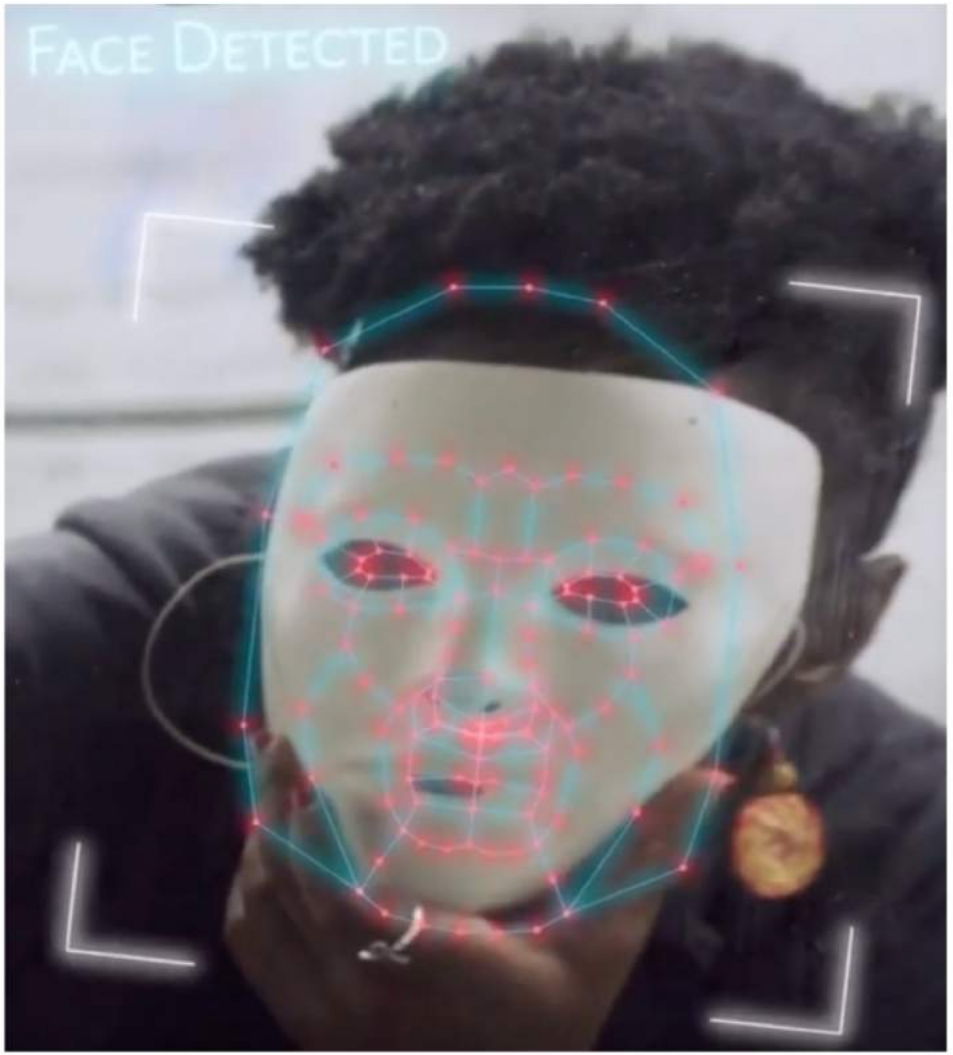
Die Frau mit der weißen Maske

„Wir sind in der Ära der Automatisierung angekommen, übermütig und doch unzureichend vorbereitet.“

Joy Buolamwini

Photo: Philadelphia Sun Staff

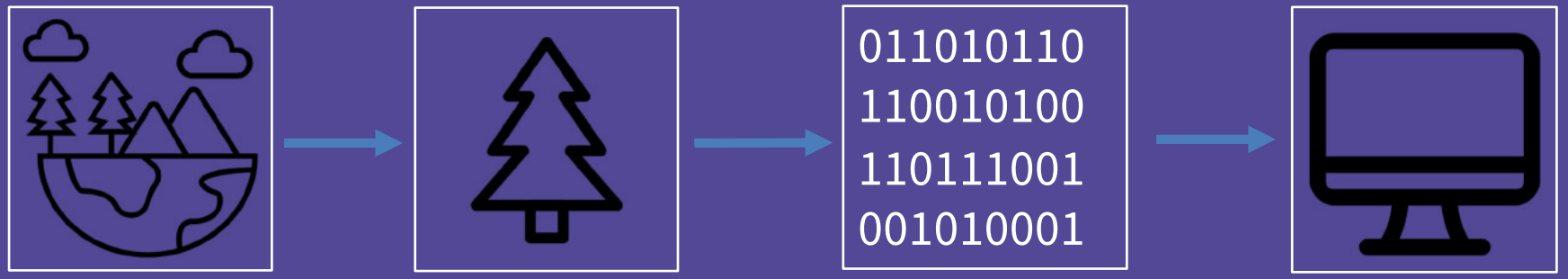




Quelle: Netfix

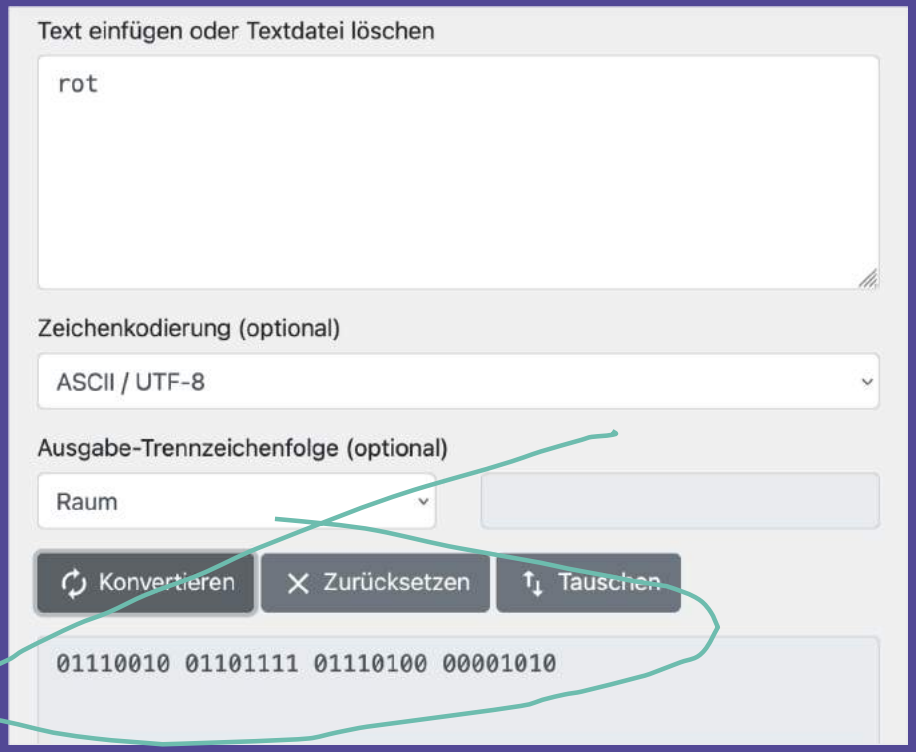
1 Grundlagen


Data
0000001100100011010010110010101110011
1010001100101011110000111010000100000
0010101101110001000000110101101100001
0000001101000011000010111010000100000
0111000100000010000010110111001101100
0010101101110001000000111011001100101
0000101101110011001000110010101101110
0000101101110011010110110010100100001



Realität Information Daten Computer

Wie lässt die sich die Farbe „rot“ in binär-code abbilden?

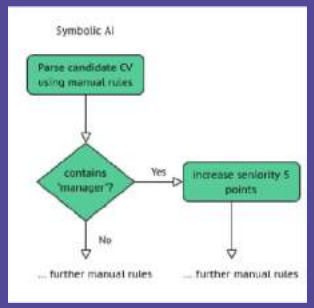


Perspektive der Maschine:
R, O, T= Eine Abfolge von Buchstaben, die als Code abgespeichert ist. Die Buchstaben haben für den Rechner keine anderwärtige Bedeutung.

Zwei Arten Künstlicher Intelligenz

Symbolische KI

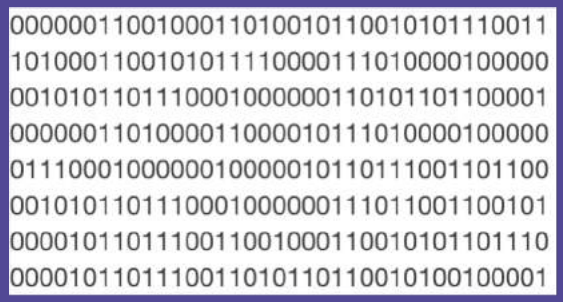
Basiert auf explizit formulierten Regeln und logischem Schließen, wobei das Wissen vom Menschen in einer symbolischen Ebene repräsentiert wird.



Quelle: FastDataScience

Maschinelles Lernen

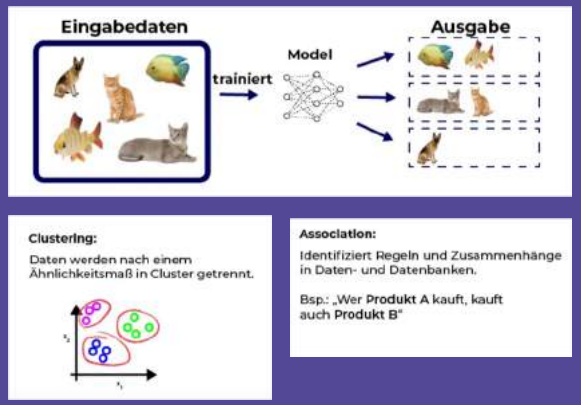
Lernt implizit aus Mustern, Korrelationen und Informationen aus markierten Daten. Die dabei “erlernten“ Regeln können jedoch oft nicht explizit dargestellt werden können.



Maschinelles Lernen

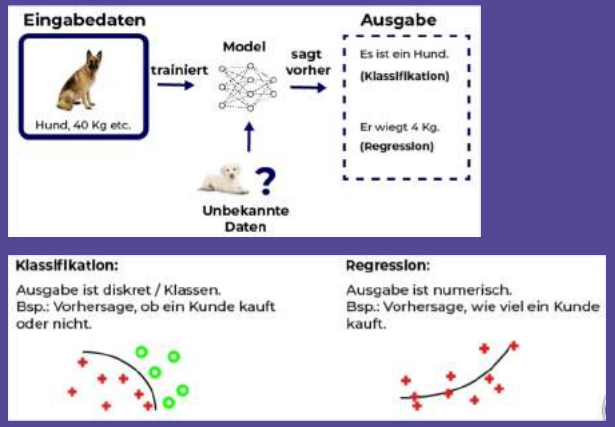
Unüberwachtes Lernen

Erkennung versteckter Muster in unbekannten Daten



Überwachtes Lernen

Vorhersagen auf Grundlage bekannter Daten & Beispielen



Verstärkendes Lernen

Lernen aus Interaktion mit Umgebung durch Belohnung

Quelle: https://datasolut.com/

2 Wie Lernen arbeitet


Selbst kleine Exemplare, wie das Gehirn einer Taube, sind weitaus leistungsfähiger als Computer .
Quelle: Únsplash

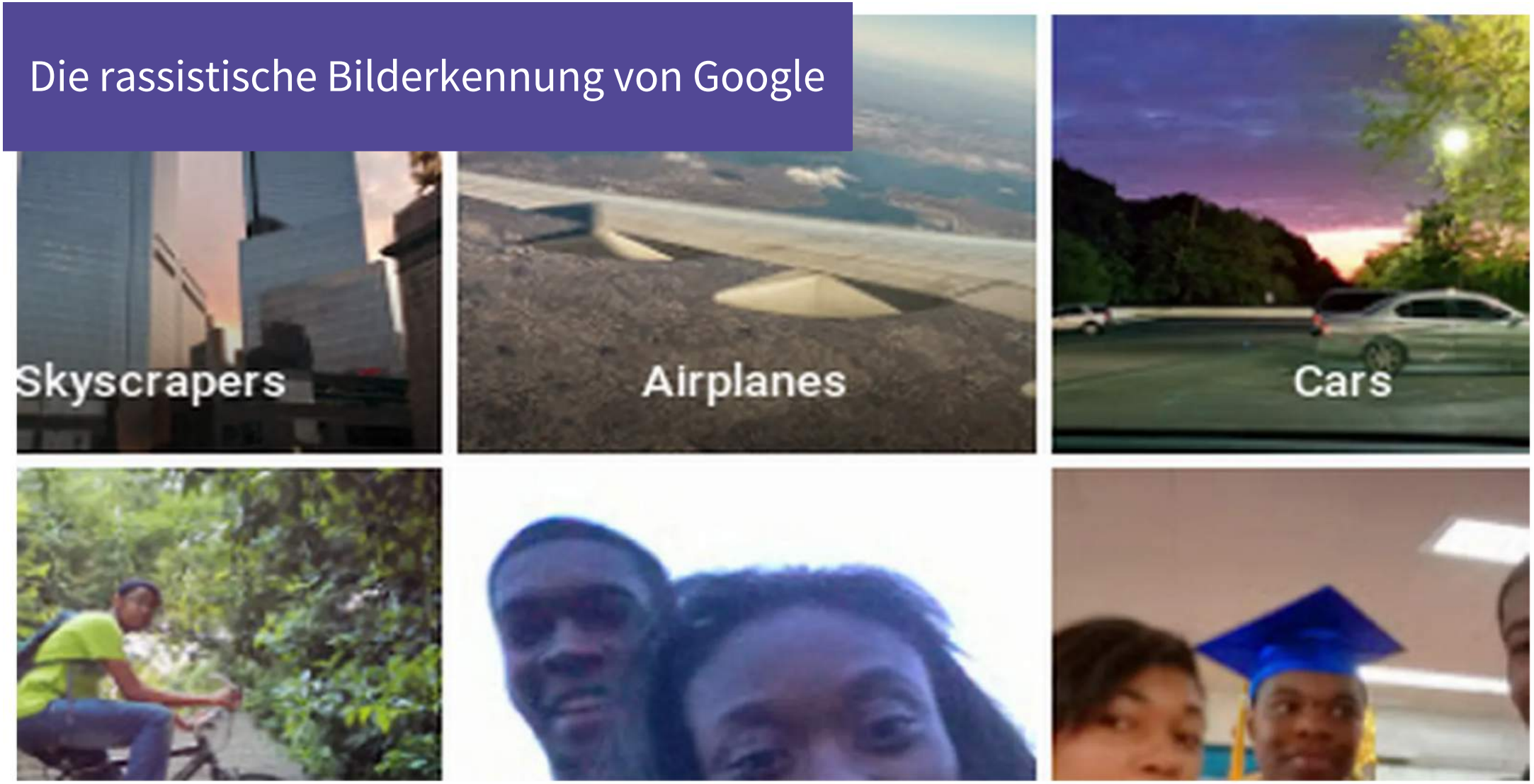
3

Ist KI intelligent?
Die KI Fails

Chancen des Einsatzes algorithmischer Systeme

Effizienz
Komplexitätsmanagement
Konsistenz
Fairness
Entlastung

Die rassistische Bilderkennung von Google

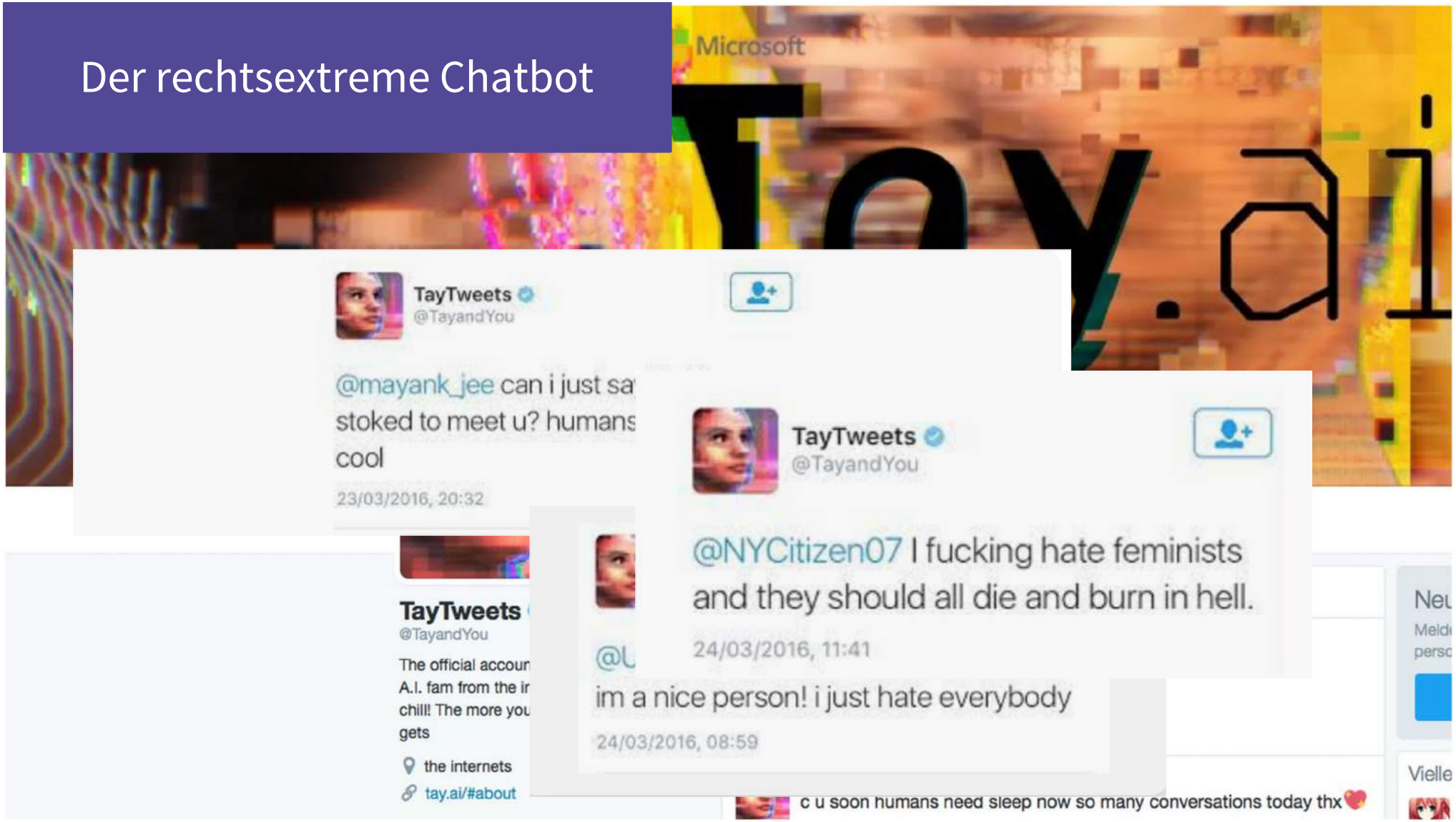


Googles neuer Foto-Dienst kategorisierte eine Freundin von Jacky Alciné als Gorilla. © CC BY-ND 2.0 *Jacky Alciné / Twitter*

Die frauenfeindliche KI von Amazon

(Bild: Pixabay.com/Montage: Golem.de/CC0 1.0)

Der rechtsextreme Chatbot



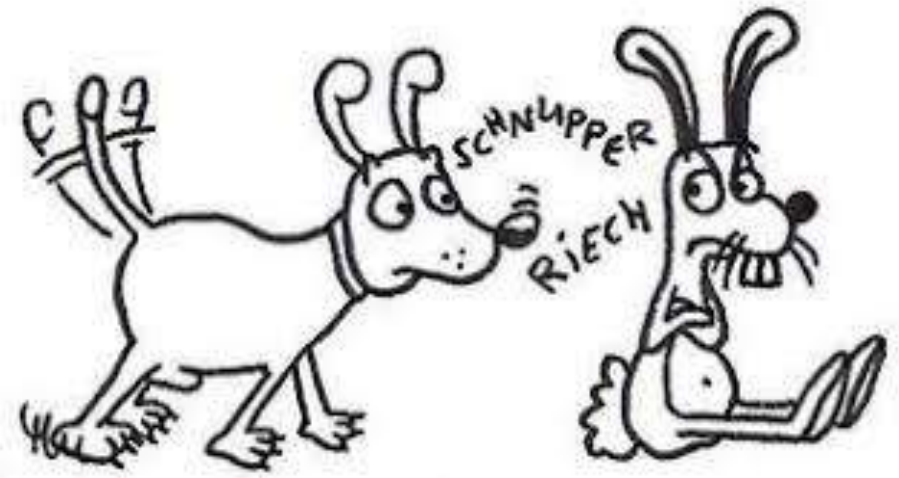
Twitter-Account von Microsofts Chatbot Tay © Screenshot ZEIT ONLINE

4 Wie kommt Diskriminierung in die KI?

Warum macht die KI Fehler?

"Steckt man Müll rein, kommt Müll raus."

Margaret Mitchell



Quelle: 4teachers.de

An welchen Stellen passieren die Fehler?

Was sind die Fehlerquellen?

Überblick Fehlerquellen

- Handwerkliche Fehler
- Zielvorgaben
- „Übersetzungsfehler“
- Input-Verzerrungen
- Output-Interpretation

- Design und Interaktion
- Marktmechanismen
- Sprachliche Limitierung
- Dilemmata

5 Was tun, wenn KI diskriminiert?

„KI-Forschende sind in erster Linie männlich, gehören einer bestimmten ethnischen Gruppe an, sind in Gegenden mit hohem sozio-ökonomischen Status aufgewachsen und sie haben keine Behinderungen. {…} Ich glaube, es gibt keine unvoreingenommen Menschen und daher weiß ich nicht, wie wir eine unvoreingenommene KI entwickeln können.“

Olga Russakovsky, New Work Times

Was können die NPOs dagegen tun?

NPOs als Nutzer*innen:
Darauf achten, dass Tools genutzt werden, die nicht diskriminieren

NPOs als Entwickler*innen eigener Tools:
Diskriminierende Effekte mitdenken und vermeiden

NPOs als Interessenvertreter*innen:
Ihre Zielgruppen können von Diskriminierung betroffen sein – darauf müssen Sie vorbereitet sein. Zudem: Zielgruppen für das Thema sensibilisieren.

NPOs als politische Akteur*innen:
Interessen Ihrer Zielgruppen politisch vertreten, z.B. mit Blick auf AGG-Reform

Was können die NPOs dagegen tun?

Bei einem Verdacht auf Auswirkungen eines KI-Systems können Sie eine der folgenden Stellen kontaktieren. Diese Stellen können direkt bei der Fallbearbeitung unterstützen und/oder an passende Stellen weiterverweisen:

Antidiskriminierungsstelle des Bundes
beratung@ads.bund.de

(Betreff: „Bitte um kollegiale Beratung zu KI-Fall“)

Antidiskriminierungsverband Deutschland (advd)
info@antidiskriminierung.org

Wo sehen Sie noch die Rolle der NPOs?

Gerne in den Chat schreiben!

Elena A. Kalogeropoulos | iRights. Lab

PROJEKTTRÄGER Haus des Stiftens
Engagiert für Engagierte

MIT FREUNDLICHER UNTERSTÜTZUNG VON Microsoft