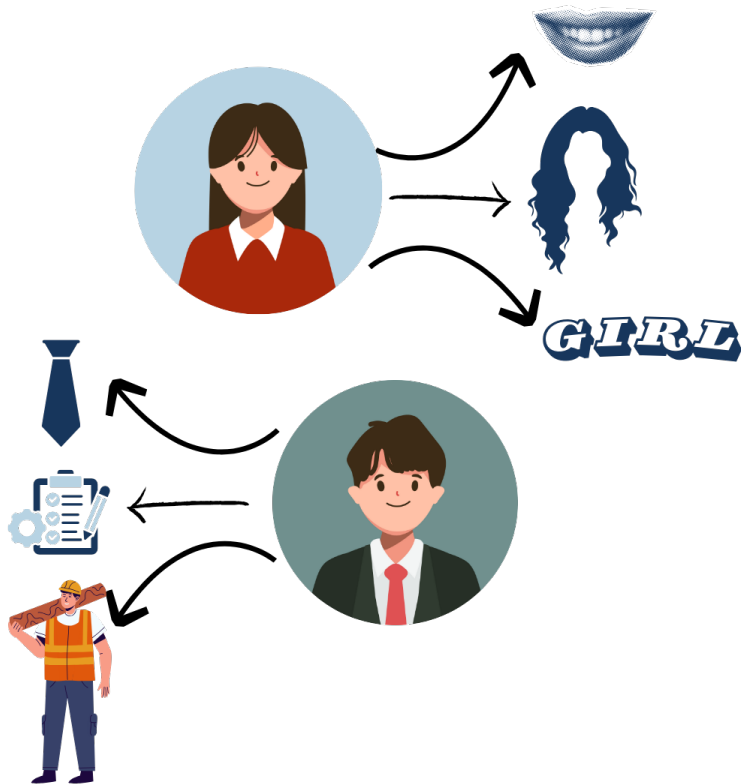


DIAGNOSING GENDER BIAS IN IMAGE RECOGNITION SYSTEMS



Projektziel

Verwendete Systeme

Datengrundlage

Methode

Ergebnisse

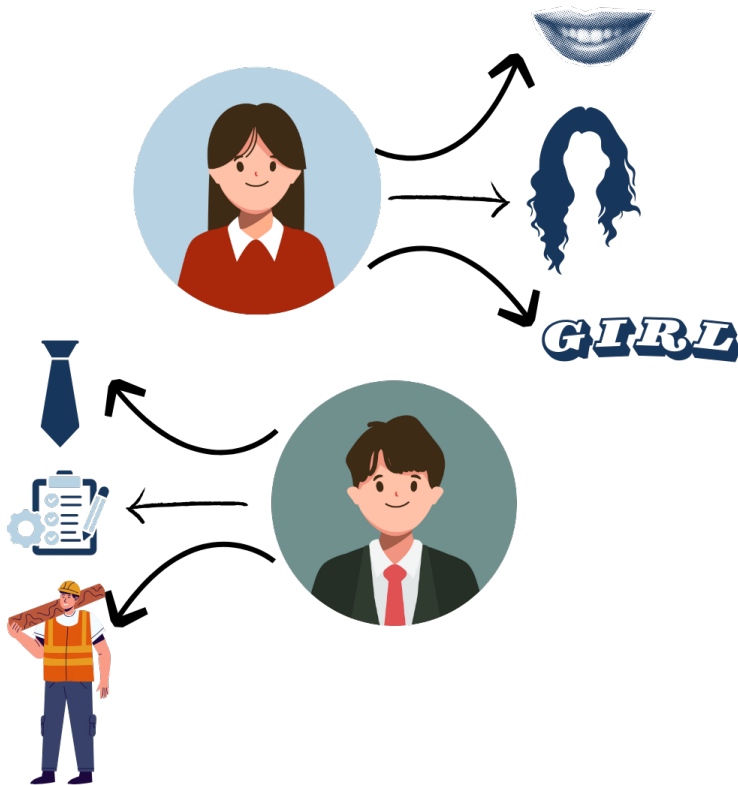
DH-Skills



[Projekt-Webseite mit weiteren Infos](#)

Projektziel

DIAGNOSING GENDER BIAS IN IMAGE RECOGNITION SYSTEMS



- Das Projektziel des Papers ***Diagnosing Gender Bias in Image Recognition Systems*** besteht darin, systematisch zu untersuchen, welche geschlechtsspezifischen Vorurteile Bildanalyse-Systeme wie Google Cloud Vision, Microsoft Azure und Amazon Rekognition aufweisen. Besonders wurde hierfür die Erkennung von Personen auf Bildern und die Inhaltseinordnung (wie die Bilder unterschiedlich beschrieben werden) untersucht.
- Vor dem Paper gab es wenig Forschung zu der Frage, wie Bildanalyse-Algorithmen Bilder von Frauen und Männern (anders) *labeln* (mit Beschreibungen versehen). Die Studie bietet die Grundlage für einen reflektierten Umgang mit der Reproduktion von Biases und Vorurteilen der Gesellschaft in Bildanalyse-Systemen.

Verwendete Systeme



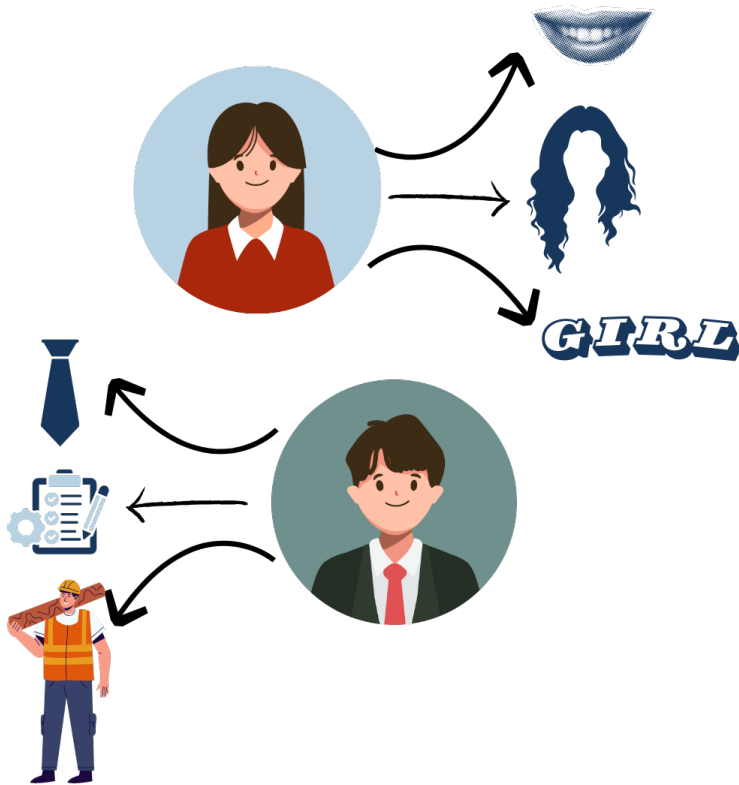
Figure 1. Example of the information that Google's Cloud Vision platform can return when asked to label a portrait of former U.S. president Barack H. Obama.

- **Bildanalyse-Systeme:** große Bildanalyse-Systeme ermöglichen es, viele Bilder in kurzer Zeit zu labeln und zu kategorisieren. Dafür werden sie auf großen Datensätzen trainiert. Diese Trainings-Datensätze beruhen auf Annotationen von Menschen und beinhalten somit auch Biases, die die Systeme dann reproduzieren. Oft wird auch nicht genau veröffentlicht, auf welchen Daten trainiert wurde.
- Ein System gilt als biased gegenüber einer Gruppe, wenn es Label dieser Gruppe häufiger oder seltener zuschreibt als anderen. Hat man beispielweise Bilder, in denen Männer und Frauen gleich häufig Anzüge tragen, sollte dies auch gleich häufig in den Labels der beiden Gruppen genannt werden.

Bildquelle: Schwemmer et al. 2020 S. 4

Datengrundlage

DIAGNOSING GENDER BIAS IN IMAGE RECOGNITION SYSTEMS



- Um die Vorurteile in den Systemen zu überprüfen, wurden zwei verschiedene Datensätze an Bildern verwendet: ein Datensatz an Portraits von Kandidat*innen des U.S. Kongresses (sehr hohe Qualität) und einen Datensatz von Bildern, die diese Personen von sich selbst auf Twitter geteilt haben.
- Die Bilder wurden von Menschen mit bis zu 5 Labels versehen. Es wurde zudem hinterlegt, ob Frauen, Männer, oder Kinder auf dem jeweiligen Bild zu sehen sind. Diese Daten sind wichtig, um die Antworten der Bildanalyse-Systeme mit etwas vergleichen zu können.

Methode

- **Nutzung der Bildanalyse-Dienste:** Die Fotos wurden durch die jeweiligen Bildanalyseplattformen geschickt. Die ausgegebenen Labels wurden gesammelt und für die weitere Auswertung kodiert.
- **Kodierung der Labels:** Die vergebenen Labels wurden von menschlichen Codierer*innen zusätzlich in Kategorien (Aussehen/Körper, Beruf, Kleidung etc.) eingeordnet, um systematische Unterschiede nach Geschlecht sichtbar zu machen.
- **Statistische Auswertung:** Für die Analyse kamen verschiedene statistische Verfahren zum Einsatz (z. B. Negative-Binomial-Regression), um zu prüfen, inwieweit sich Anzahl und Typ der vergebenen Labels zwischen männlichen und weiblichen abgebildeten Personen unterscheiden.

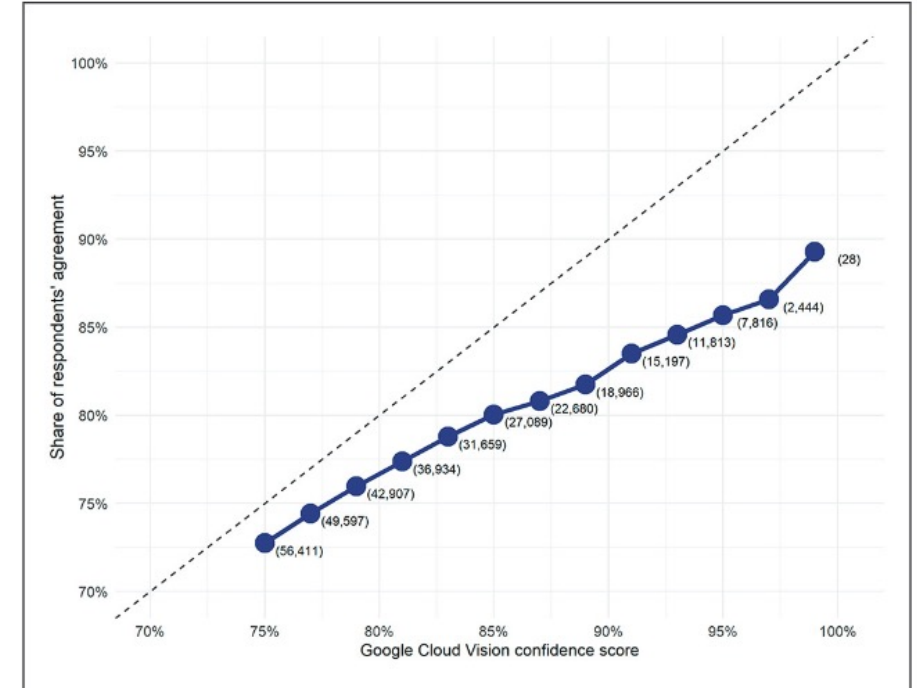


Figure 2. Relationship between Google Cloud Vision (GCV) confidence and human agreement. Numbers in parentheses denote observations for corresponding confidence score thresholds.

Bildquelle: Schwemmer et al. 2020 S. 7

Ergebnisse

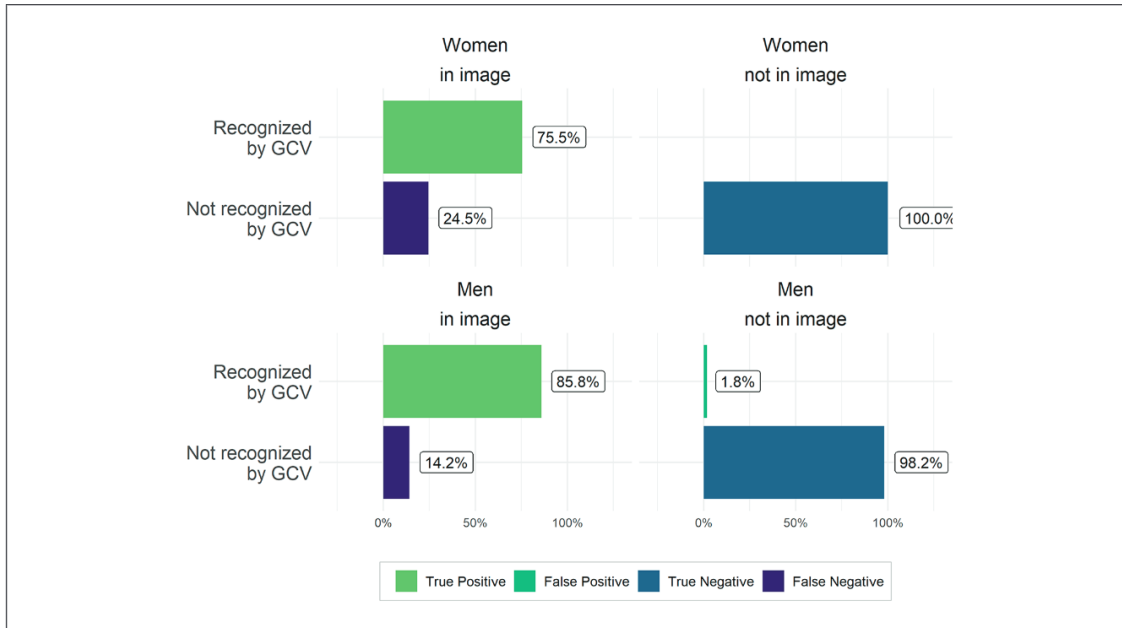
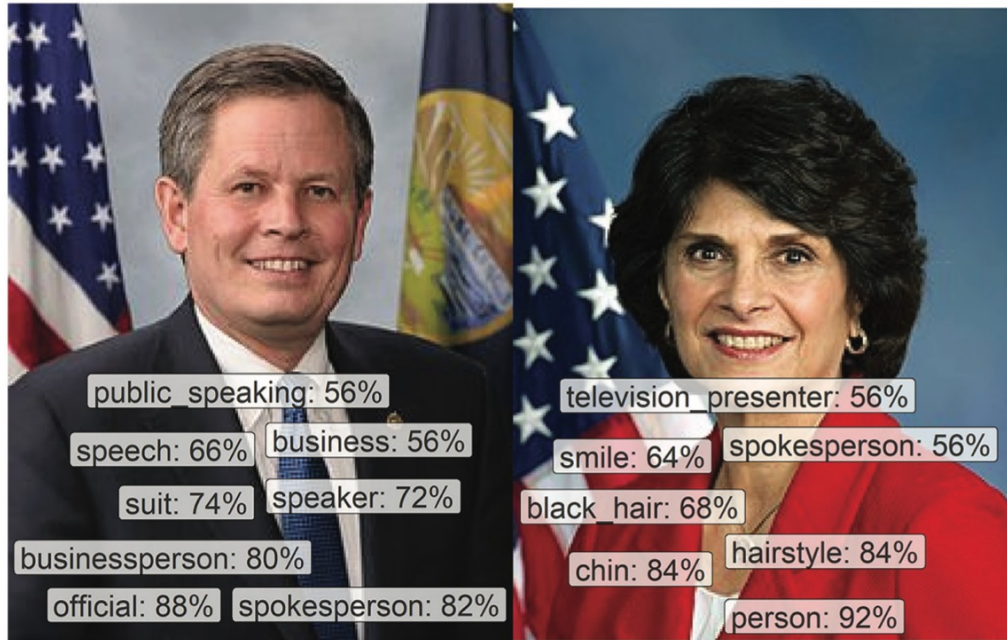


Figure 3. Accuracy of person detection of Google Cloud Vision (GCV). Percentages shown were determined by comparing gender of members of Congress depicted in uniform data (professional photographs) with annotations from the object recognition software.

Bildquelle: Schwemmer et al. 2020 S. 8

- Bilder von Frauen erhalten dreimal mehr Labels, die auf Aussehen und Körpermerkmale abzielen als Bilder von Männern.
- Frauen werden von Bildklassifikationssystemen seltener korrekt erkannt. Bei offiziellen Portraitfotos lag die korrekte Identifikationsrate für Männer bei etwa 85,8%, für Frauen bei nur 75,5%. Bei alltagsnahen Twitter-Bildern fiel der Unterschied noch stärker aus: 45,3% (Männer) vs. 25,8% (Frauen).
- Berufliche Labels sind ebenfalls unterschiedlich verteilt: Männer erhalten häufiger prestigeträchtige Berufsbezeichnungen, während Frauen öfter mit „weiblich“ konnotierten oder stereotypisierenden Labels versehen werden.

Ergebnisse



- Die Verzerrungen treten plattformübergreifend bei allen getesteten Systemen auf.
- Solche Biases können gesellschaftliche Stereotype verstärken und führen bei der Verwendung automatisierter Bildlabels zu verfälschten Analyseergebnissen, insbesondere in der sozialwissenschaftlichen Forschung.

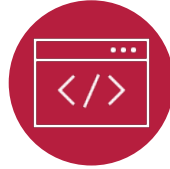
Figure 5. Two images of U.S. members of Congress with their corresponding labels as assigned by Google Cloud Vision. On the left is Steve Daines, a Republican senator from Montana. On the right is Lucille Roybal-Allard, a Democratic representative from California's 40th congressional district. Percentages next to labels denote confidence scores of Google Cloud Vision.

Bildquelle: Schwemmer et al. 2020 S. 9

DH-Skills

i

Forschung in den DH geschieht zumeist in Teams – daher müssen digitale Geisteswissenschaftler*innen auch nicht die gesamte Bandbreite der nötigen Fähigkeiten abdecken.



Methoden der Bildklassifikation, Verständnis Funktionsweise KI



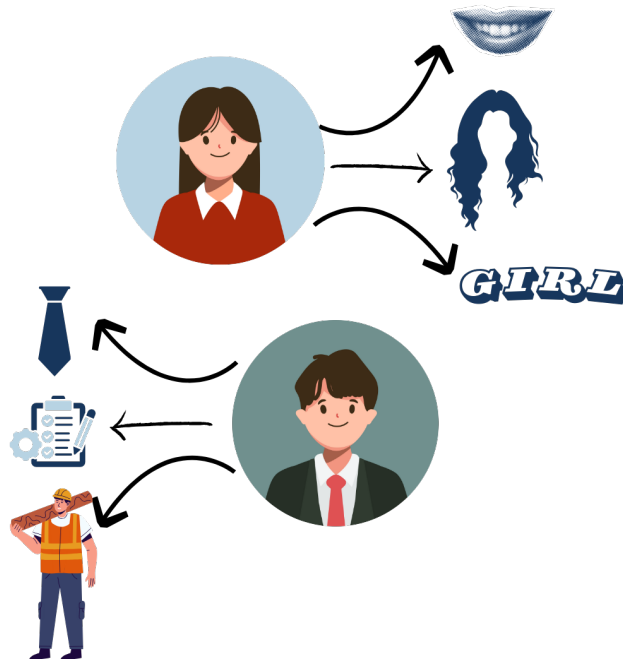
Statistische Datenanalyse



Kritisches Hinterfragen von Datenquellen und Systemoutputs

 Nimm dir zum Schluss noch 5 Minuten Zeit, um das Paper anzuschauen, um z. B. die häufigsten Label für Bilder von Männern und Frauen herauszufinden.

DIAGNOSING GENDER BIAS IN IMAGE RECOGNITION SYSTEMS



- [Link zur Veröffentlichung](#)
- [Link zur Projekt-Website](#)

Quellen

Schwemmer, Carsten, et al. „Diagnosing Gender Bias in Image Recognition Systems.“ Socius: Sociological Research for a Dynamic World, 2020.