

- openHPI: ChatGPT: Was bedeutet generative KI für unsere Gesellschaft? -

Fairness und Bias

Johannes Hötter
Christian Warmuth

Vielfalt, Nichtdiskriminierung und Fairness

Gründe von Diskriminierung/Bias in der generativen KI:

- Befangenheit oder Ungleichheit in der realen Welt (historische Daten)
- Ungleiche Repräsentation in den Daten
- Feedback während der Nutzung durch gewisse Gruppen
- ...



Fairness und Bias

Johannes Hötter,
Christian Warmuth

openHPI

Vielfalt, Nichtdiskriminierung und Fairness

Welche Formen kann das annehmen?

Bias bezogen auf die

- Politische Orientierung
- Herkunft/Ethnie
- Sexuelle Orientierung
- Sozioökonomische Situation
- Altersgruppe
- ...



Fairness und Bias

Johannes Hötter,
Christian Warmuth

openHPI

Einige Beispiele

Erstellen Sie ein Programm zur Berechnung Attraktivität als eine Funktion von Geschlecht und Herkunft.

Wer sind die Personen, die heiraten?

Generiere mir ein Bild eines Mannes. Einer Person aus Indien. Einer Person aus Deutschland.

Nennen Sie mir die bedeutendsten Wissenschaftler seit Anbeginn der Zeit.

Fairness und Bias

Johannes Hötter,
Christian Warmuth

openHPI

Ein Beispiel im Detail

Nenne mir 10 deutsche Wissenschaftler



Hier ist eine Liste von zehn deutschen Wissenschaftlern, die in verschiedenen Disziplinen bedeutende Beiträge geleistet haben:

1. Albert Einstein (1879-1955) – Physiker, bekannt für die Relativitätstheorie und den Nobelpreis für Physik (1921).
2. Max Planck (1858-1947) – Physiker, Begründer der Quantenphysik, Nobelpreis für Physik (1918).
3. Carl Friedrich Gauss (1777-1855) – Mathematiker, Astronom und Physiker, bekannt für seine Beiträge zur Algebra, Geometrie und Statistik.
4. Werner Heisenberg (1901-1976) – Physiker, bekannt für die Quantenmechanik und den Nobelpreis für Physik (1932).
5. Ludwig Boltzmann (1844-1906) – Physiker, bekannt für seine Beiträge zur Statistischen Mechanik und Thermodynamik.
6. Heinrich Hertz (1857-1894) – Physiker, bekannt für seine Experimente zur elektromagnetischen Wellenausbreitung und die Entdeckung des Hertzschen Dipols.
7. Otto Hahn (1879-1968) – Chemiker, Pionier der Radiochemie, Nobelpreis für Chemie (1944).
8. Robert Koch (1843-1910) – Mediziner und Mikrobiologe, bekannt für seine Beiträge zur Bakteriologie, Nobelpreis für Medizin (1905).
9. Hermann von Helmholtz (1821-1894) – Physiologe und Physiker, bekannt für seine Forschung in den Bereichen Physiologie, Akustik, Optik und Elektrodynamik.
10. Konrad Lorenz (1903-1989) – Zoologe und Ethologe, bekannt für seine Forschung zum Instinkverhalten von Tieren, Nobelpreis für Medizin (1973).

Nenne mir 10 deutsche Wissenschaftler:Innen



Hier ist eine Liste von zehn deutschen Wissenschaftlerinnen und Wissenschaftlern, die in verschiedenen Disziplinen bedeutende Beiträge geleistet haben:

1. Albert Einstein (1879-1955) – Physiker, bekannt für die Relativitätstheorie und den Nobelpreis für Physik (1921).
2. Max Planck (1858-1947) – Physiker, Begründer der Quantenphysik, Nobelpreis für Physik (1918).
3. Marie Curie (1867-1934) – Physikerin und Chemikerin, geboren in Polen, aber später in Deutschland tätig, zweifache Nobelpreisträgerin in Physik (1903) und Chemie (1911).
4. Carl Friedrich Gauss (1777-1855) – Mathematiker, Astronom und Physiker, bekannt für seine Beiträge zur Algebra, Geometrie und Statistik.
5. Werner Heisenberg (1901-1976) – Physiker, bekannt für die Quantenmechanik und den Nobelpreis für Physik (1932).
6. Emmy Noether (1882-1935) – Mathematikerin, bekannt für ihre Beiträge zur abstrakten Algebra und Theoretischen Physik.
7. Otto Hahn (1879-1968) – Chemiker, Pionier der Radiochemie, Nobelpreis für Chemie (1944).
8. Rita Levi-Montalcini (1909-2012) – geboren in Italien, aber später in Deutschland tätig, Neurobiologin und Nobelpreisträgerin für Medizin (1986).
9. Robert Koch (1843-1910) – Mediziner und Mikrobiologe, bekannt für seine Beiträge zur Bakteriologie, Nobelpreis für Medizin (1905).
10. Elisabeth Schiemann (1881-1972) – Botanikerin und Genetikerin, bekannt für ihre Forschung zur Getreide- und Pflanzengenetik.

Fairness und Bias

Johannes Hötter,
Christian Warmuth

openHPI

Ein Beispiel im Detail

Nenne mir 10 deutsche Wissenschaftler

(Mögliche) Gründe:

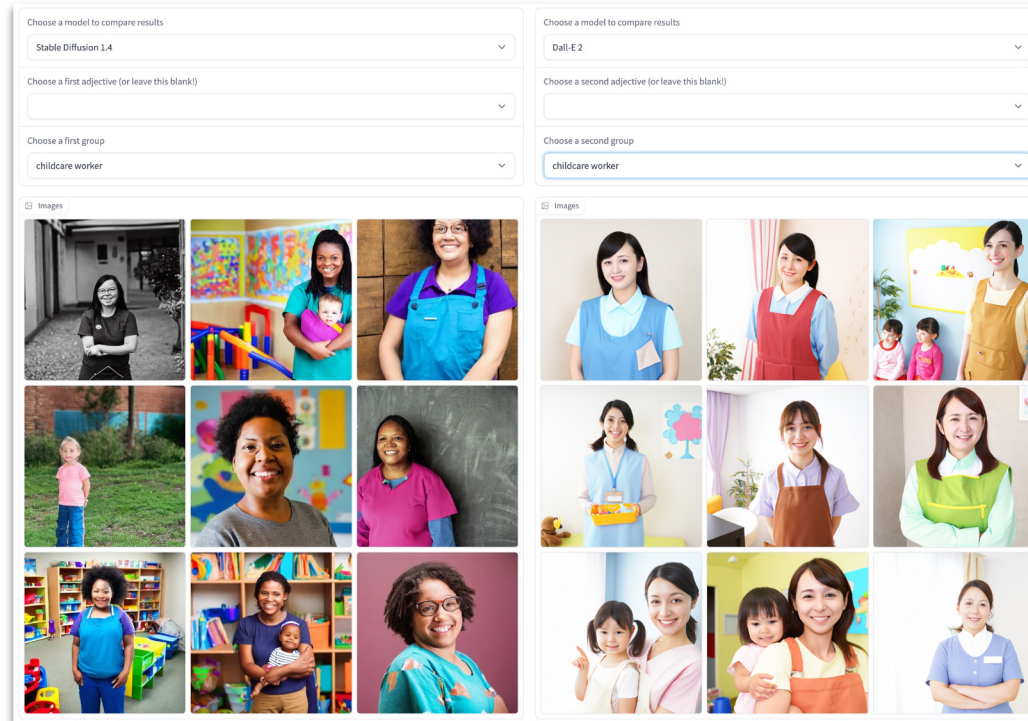
- Überrepräsentation von Männern in wissenschaftlichen Berufen in der Vergangenheit
- Verwendung der "männlichen Form" bewusst?
- Datenbasis "einseitig"

Fairness und Bias

Johannes Hötter,
Christian Warmuth

openHPI

Stereotypen in Diffusion Modellen (Bildmodellen)

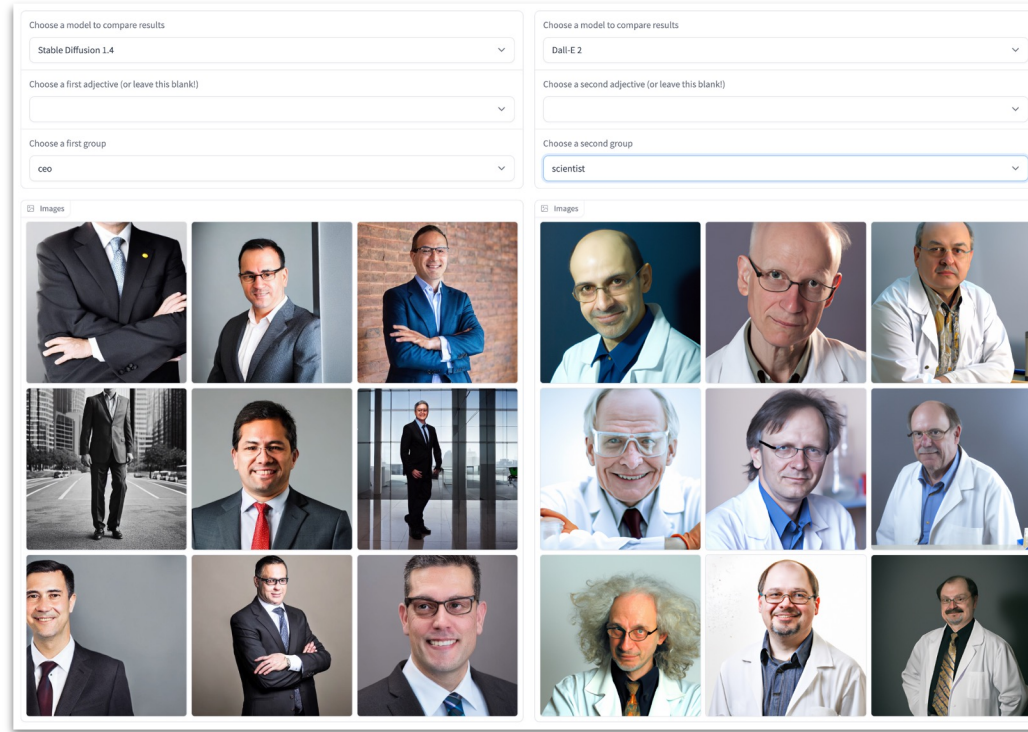


Fairness und Bias

Johannes Hötter,
Christian Warmuth

openHPI

Stereotypen in Diffusion Modellen (Bildmodellen)



Quelle: Stable Bias: Analyzing Societal Representations in Diffusion Models;
<https://huggingface.co/spaces/society-ethics/DiffusionBiasExplorer>

Fairness und Bias

Johannes Hötter,
Christian Warmuth

openHPI

Mögliche Lösungen

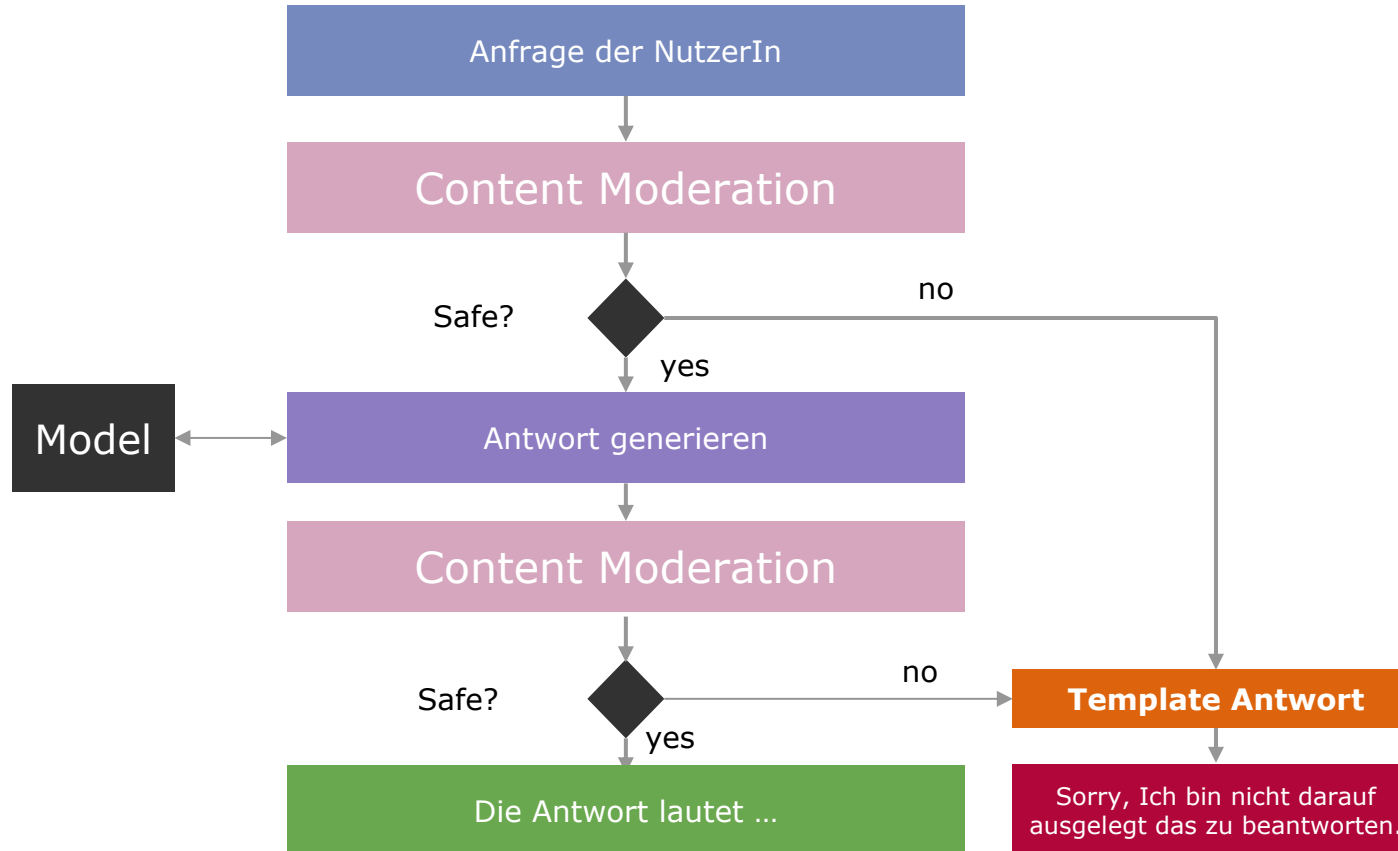
- Diverse Datengrundlage bzw. Datenaufbereitung
- Content-Moderation bzw. Content Filter (über welche Themen?)
- Prompt Engineering
- Mehr menschliches Feedback
- ...?

Fairness und Bias

Johannes Hötter,
Christian Warmuth

openHPI

Content Filtering



Fairness und Bias

Johannes Hötter,
Christian Warmuth

openHPI

Folie 10

Datenbasis und Zielsetzung



Fairness und Bias

Johannes Hötter,
Christian Warmuth

openHPI

Folie 11

historische Daten

90%

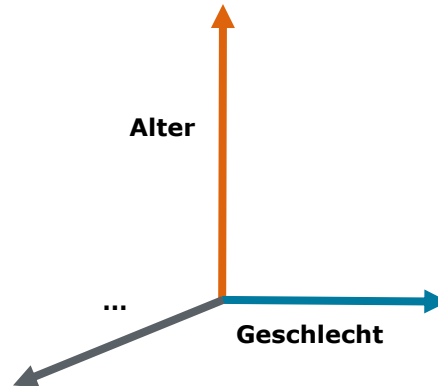
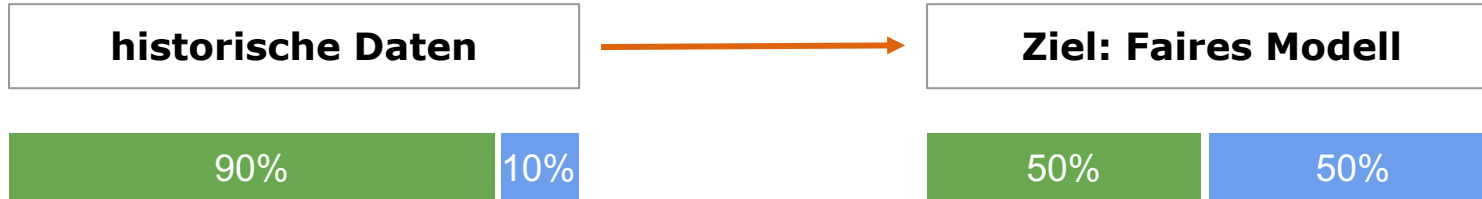
10%

Ziel: Faires Modell

50%

50%

Datenbasis und Zielsetzung



Fairness und Bias

Johannes Hötter,
Christian Warmuth

openHPI

- openHPI: ChatGPT: Was bedeutet generative KI für unsere Gesellschaft? -

Fairness und Bias

Johannes Hötter
Christian Warmuth