Exploring the Impact of Social Dominance Orientation and Ambivalent Sexism on the Perception of AI-Generated Gendered Images

Julian Bornemeier[1,2] and Prof. Dr. Jan-Philipp Stein[1]

[1] Department of Media Psychology, Chemnitz University of Technology
[2] Department of Cognitive Psychology and Human Factors, Chemnitz University of Technology

Use of Generative AI

Widespread use of AI systems in everyday life

- In recent years, AI applications have expanded significantly, influencing **sectors like data science, medical diagnosis, and marketing** (Pereyda & Holder, 2020; Wang et al., 2019; Heinz et al., 2023)

- AI's use in generative tasks, particularly in **image creation**, has grown, allowing for the production of **marketing materials, artistic content, and more** (Segato et al., 2020; Dwivedi et al., 2019)

  - e.g. DALL-E, DeepArt, and Midjourney

Generative AI and (Gender) Bias

AI systems are trained on **existing and biased data**

- Replicate and **amplify societal stereotypes** (Bolukbasi et al., 2016)
- AI-generated images tend to **underrepresent women** in traditionally male-dominated fields like law, medicine, engineering (Górska & Jemielniak, 2023; Messingschlager & Appel, 2024)

  - Women are **portrayed as smiling more often** (Sun et al., 2024; Zhou et al., 2024)

  - Women are **more sexualized** compared to men (Sandoval-Martin & Martínez-Sanzo, 2024)

Humans interacting with Generative AI

People may **assume that AI systems are neutral/objective** (e.g., Pethig & Kroenung, 2022) especially **if pre-existing beliefs are confirmed** (Jussupow et al., 2020)

> *How do people's attitudes and beliefs affect their perception/evaluation of biased AI images?*

- **Social Dominance Orientation (SDO)**: Ideological belief that describes the **preference for inequality (and hierarchy)** between social groups (Pratto et al., 1994)

- **Ambivalent Sexism:** Includes **hostile sexism** (negative views of non-traditional women) and **benevolent sexism** (patronizing views of traditional women) (Glick & Fiske, 1996)

Hypotheses

- Main effect of biased vs. unbiased images on perceived bias and fairness (H1a and H1b)
- Moderation by SDO and Ambivalent Sexism (H2a/H2b and H3a/H3b)

Perceived Bias: Do participants rate the material as being biased / not neutral

Fairness Evaluation: Do participants judge the material as fair / justified
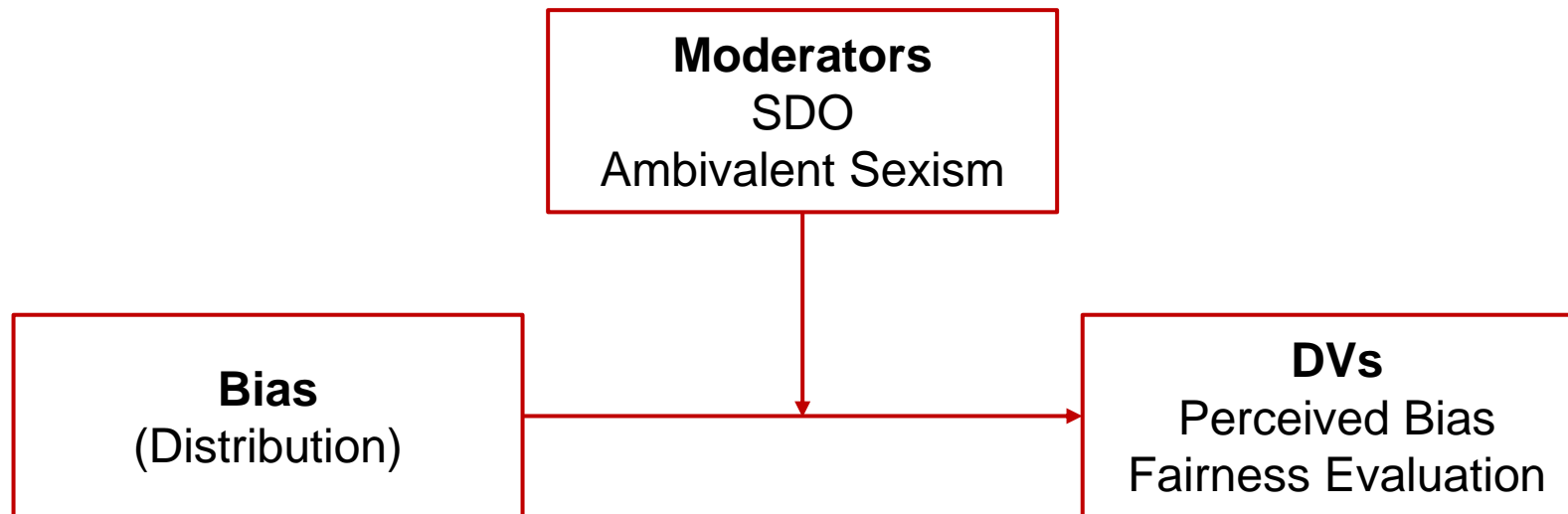
Research Questions and Exploratory Analyses

- (Interaction-) Effect of AI vs. human creator on perceived bias and fairness evaluation
- Moderation by SDO & Ambivalent Sexism and their sub-facets
- sequencing/mediation effects
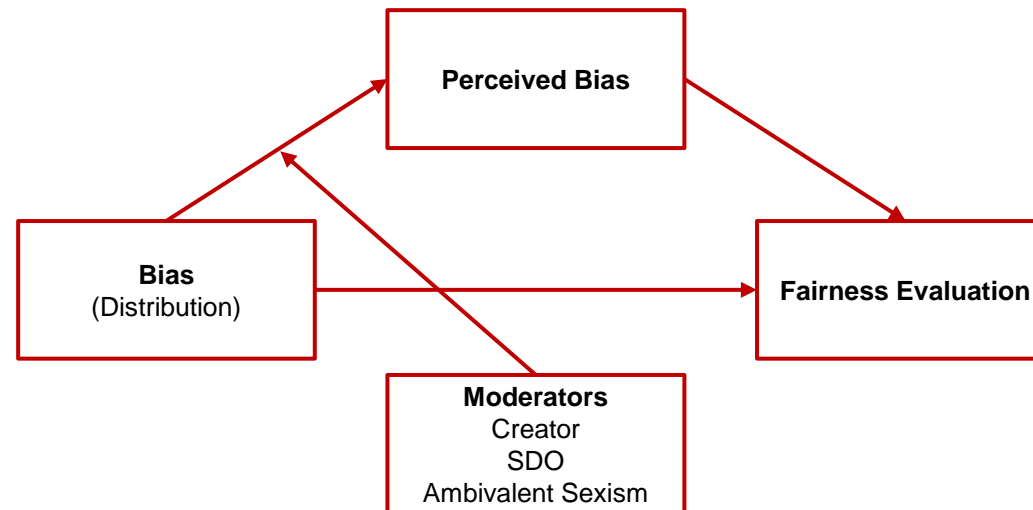
Study Design (Online-Experiment)

2x2 Design

1. Biased (2 female, 8 male)
   vs. Unbiased Images (5 each)

2. AI vs. Human Creator (Framing)

Materials

10 promotional images for STEM occupations
(2 of each: programmer, mathematician, mechanical/electrical engineer, robotics technician, biological technician)

Participants

- *N* = 545
    - Age: 32.5 (SD = 10.9)
   - Gender: 58.8% female, 2.2% non-binary

Measures

DVs

- Perceived Bias (Messingschlager & Appel, 2024) α = .86
  - *"objective"..."subjective; "neutral"..."not neutral"*
- Fairness Evaluation (derived from Marcinkowski et al., 2020; Newman et al., 2020) α = .88
 - *"The way men and women were included in the pictures seems unjust to me."*

Moderators

- SDO (SDO-7 short scale, Ho et al., 2015) α = .82
  - *"Some groups of people are simply inferior to other groups."*
- Ambivalent Sexism (ASI; Glick & Whitehead, 2010) α = .87
  - *"Women seek to gain power by getting control over men."*

Findings

H1a and H1b: **Main effects** of biased vs. unbiased images on perceived bias and fairness
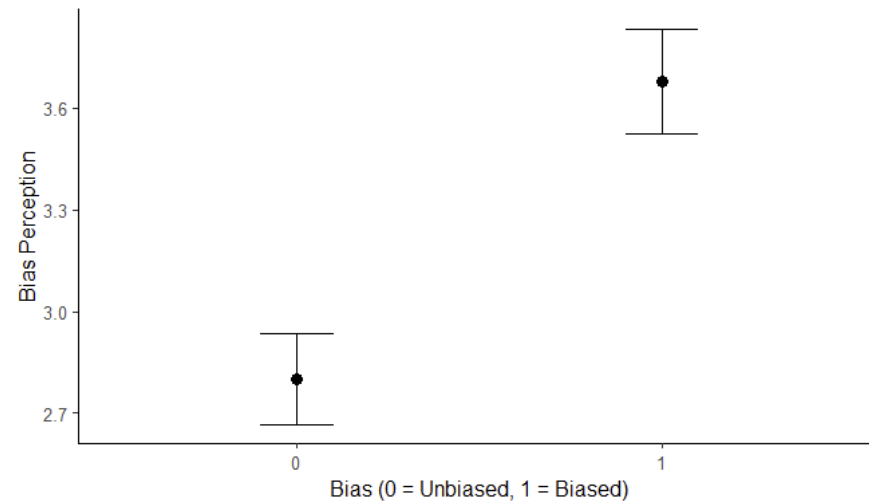
Unbiased images (M = 2.80, SD = 1.12)
Biased images (M = 3.68, SD = 1.29).
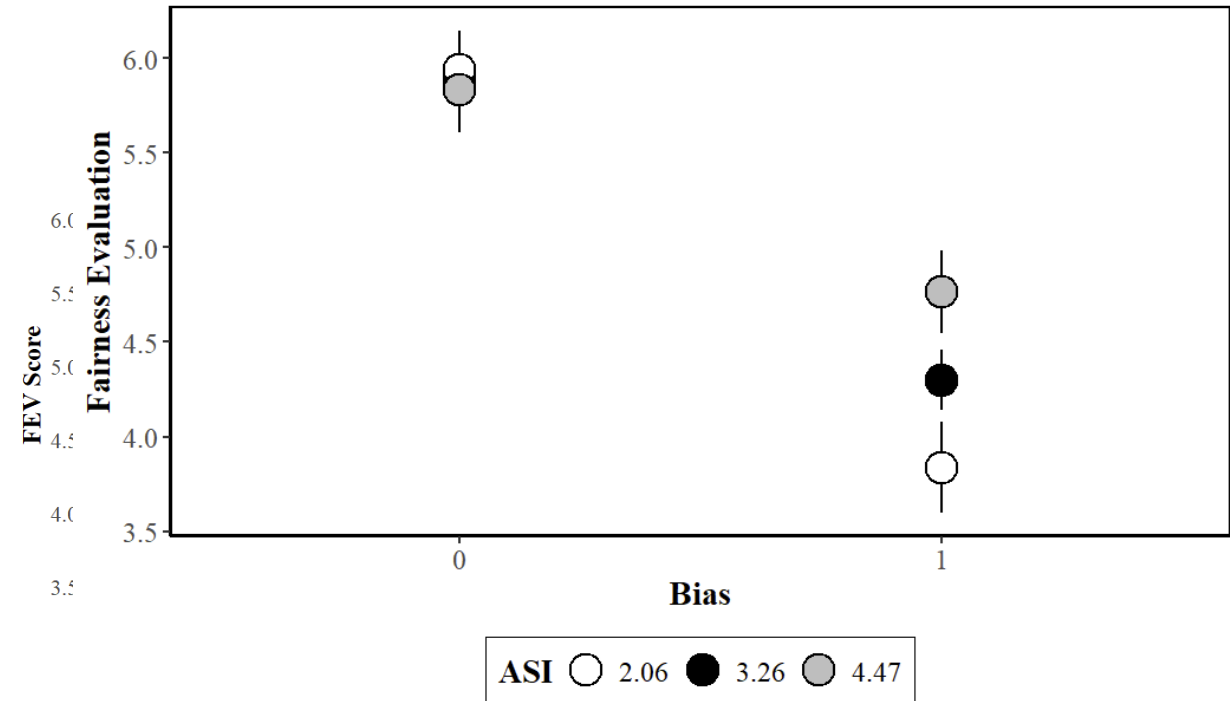t(535) = -8.50, p < .001, Cohen's d = -0.73, 95% CI [-0.90, -0.55].
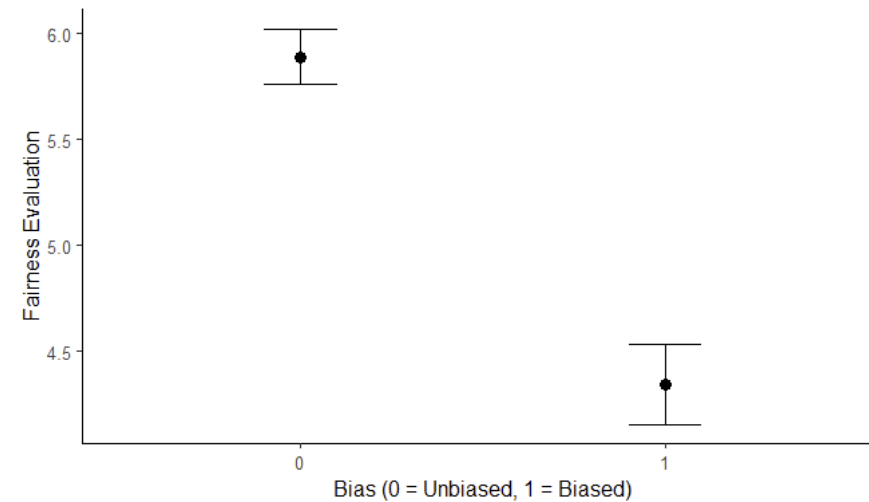
Unbiased images (M = 5.89, SD = 1.08)
Biased images (M = 4.35, SD = 1.59).
t(481) = 13.29, p < .001, Cohen's d = 1.14, 95% CI [0.95, 1.32].

Findings

H2 & H3: **Interaction effects** for SDO and ambivalent sexism

- significant moderation effects
  - e.g., people high in Ambivalent Sexism found biased images more fair compared to people with low Ambivalent Sexism
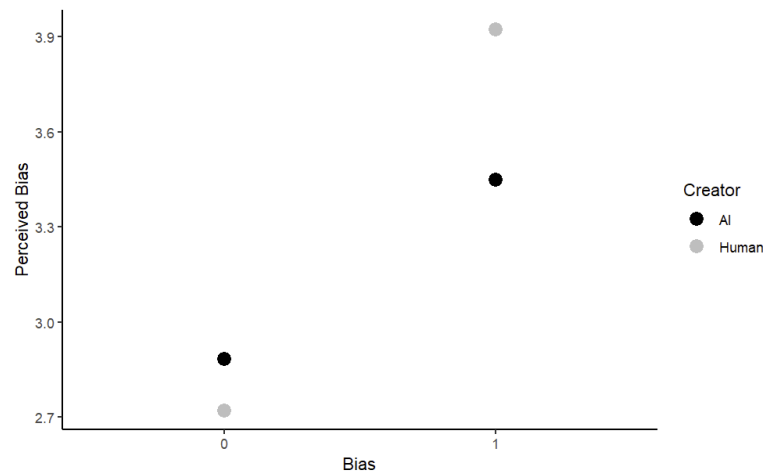

$R^2 = .30$, $F(5, 539) = 45.77$, $p < .001$
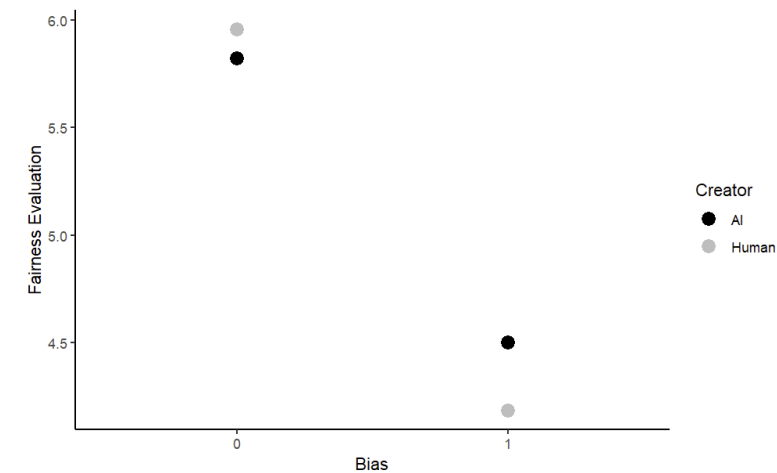Interaction effect: $b = 0.45$, $p < .001$

Findings

There was **no main effect of AI vs. human creator** on bias perception and fairness, but…

- **… Creator moderates effects on perceived bias and fairness evaluation**

Main effect of Bias: $F(1, 539) = 74.11$, $p < .001$
Interaction (Bias × Creator): $F(1, 539) = 10.50$, $p = .001$

Main effect of Bias: $F(1, 539) = 178.07$, $p < .001$
Interaction (Bias × Creator): $F(1, 539) = 4.02$, $p = .045$

Findings

Additional Analysis (SDO and Creator)

- The impact of bias on fairness is mediated by perceived bias.
- Strongest mediation occurs when Creator = Human and SDO = low (b = -1.31, p < .001).
- Weaker mediation occurs when Creator = AI and SDO = high (b = -0.07, n.s.).

Perceived Bias

Bias
(Distribution)

Fairness Evaluation

Moderators
Creator
SDO
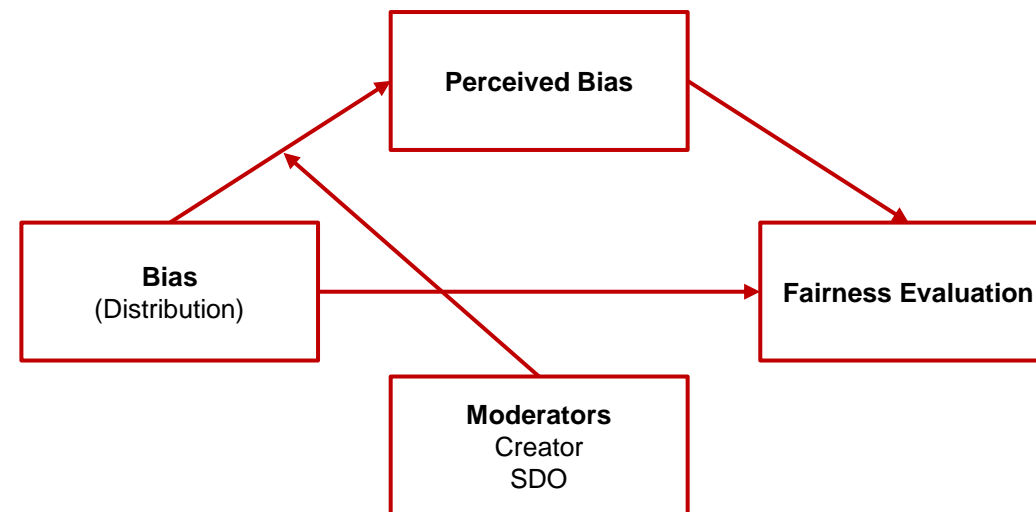
$R^2 = 0.60, p < .001$

Conclusions

- Biased AI-generated images are perceived as more biased and less fair than unbiased images.

  - Individuals with high Social Dominance Orientation (SDO) and high Ambivalent Sexism perceive biased images as less biased and more fair than individuals with lower scores on these traits.

  - The creator (AI vs. human) moderates these effects.

- AI's perceived objectivity may mask the reproduction of societal biases, especially for individuals with high levels of these attitudes.

  - Addressing biases in AI systems is essential to prevent the reinforcement of harmful societal stereotypes.

Thank you!

✈ julian.bornemeier@psychologie.tu-chemnitz.de

Mittelstand-Digital
Zentrum
Fokus Mensch