



Article

Gender bias in text-to-image generative artificial intelligence: Neglect and stereotypical presentations across three popular platforms

new media & society

1–23

© The Author(s) 2026



Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/14614448261435197

journals.sagepub.com/home/nms



Hannah Weinmann

University of Würzburg, Germany

Tanja V Messingschlager 

University of Würzburg, Germany

Markus Appel 

University of Würzburg, Germany

Abstract

Images generated by artificial intelligence (AI) were assumed to underrepresent women and to contain stereotypical portrayals. A total of 1344 images were generated at two times of measurement by DALL·E, Midjourney, and Stable Diffusion and were analyzed via a preregistered content analysis. Results revealed representational and presentational bias, varying between prompts and between AI platforms. Women were depicted less frequently than men in images generated with the prompt “competent person,” whereas women were generated more often with a neutral prompt (i.e. “person”) or with the prompt “warm person.” Regarding stereotypical presentations, images representing women (vs men) showed lower facial prominence (face-ism), higher intensity smiles, and more pronounced lateral head tilts (head canting). These biases varied significantly between AI platforms. The findings suggest that gender stereotypes

Corresponding author:

Markus Appel, Psychology of Communication and New Media, University of Würzburg, Oswald-Külpe-Weg 82, 97074 Würzburg, Germany.

Email: markus.appel@uni-wuerzburg.de

are spread by generative AI systems and highlight the need for interventions in the development and deployment of image-generating AI.

Keywords

Algorithmic bias, competence, face-ism, gender, head tilt, smiling, warmth

Artificial Intelligence (AI) has revolutionized industries over the past few years (Maslej et al., 2025), leading to the development of a wide range of applications. In 2022, some of the first realistic text-to-image (T2I) generative AI platforms became publicly available (OpenAI, 2022; Stability AI, 2022). With the introduction of newer versions of T2I AI producing outputs increasingly indistinguishable from real photographs, concerns have emerged regarding stereotype-driven results (*algorithmic bias*; Maslej et al., 2025; Nicoletti and Bass, 2023). T2I AI may generate pictures with women and minorities being underrepresented as compared to their presence in societies (*representational bias*), or T2I AI may generate pictures in which women and minorities are presented in stereotypical ways (*presentational bias*). Such biases could have downstream effects on individuals' mental representations of women and minorities, on attitudes, and on behavior (e.g. Dixon, 2019; Guilbeault et al., 2024). Because AI-generated content has rapidly become omnipresent in daily life (Maslej et al., 2025), it is crucial to examine the degree of misrepresentation in AI images.

The present study aims to quantify stereotypical gender portrayals that current T2I platforms seem to be creating. A sample of AI-generated images was systematically collected using three of the most popular T2I platforms (Midjourney, DALL.E, Stable Diffusion). The resulting images were content analyzed with regard to both representational and presentational bias. Based on the stereotype content model (Fiske et al., 2002) our focus was on frequency of generated images depicting women and men when prompts asked for a *competent* or a *warm person*. We further examined the prevalence of stereotypical presentations of women, as indicated by facial prominence (face-ism; Archer et al., 1983), the lateral tilting of a person's head (head canting; Costa et al., 2001), and smiling (Hess et al., 2005)—three visual markers which have been linked to the subordination of women.

Algorithmic bias

A subset within the larger field of generative AI, and the focus of our research, are T2I platforms, which enable users to enter a textual description of the desired image in the form of a prompt. In response, the AI platform generates an image based on the training data previously fed to it (Friedrich et al., 2023). Scholars from disciplines ranging from philosophy and critical cultural studies to computer science have outlined the perils of bias in algorithmic solutions (e.g. Crawford, 2021; Noble, 2018) and T2I platform output, specifically (e.g. Afreen et al., 2025).

Depending on the philosophical, legal, ethical, or social perspective taken to examine algorithmic bias and fairness, the concept and definition of what is considered a *bias* may

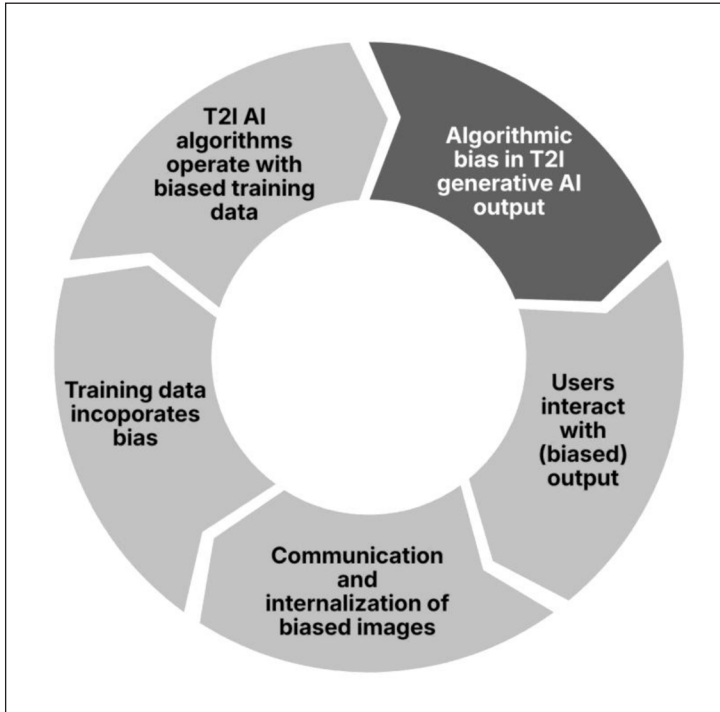


Figure 1. Algorithmic bias: antecedents and consequences of algorithmic bias in T2I generative AI output.

vary. For the present work, algorithmic bias in T2I generative AI is understood as algorithmic output that underrepresents members of groups (as compared to real world prevalences, representational bias) or displays members of groups in stereotyped ways (presentational bias; see more below). Kordzadeh and Ghasemaghaei (2022) concluded that an overarching definition of algorithmic bias needs to encompass a deviation from an equality principle, as well as systematic and repeated occurrence of said deviation. In the present study, the equality principle is defined by the deviation from an equal distribution in terms of representational bias and significant group differences in terms of presentational bias.

The focus of our study lies on the assessment of the second component of the definition, that is, if inequalities in T2I AI output occur systematically and repeatedly. Therefore, we will draw a larger sample of images and statistical inferences will be employed to detect algorithmic bias. Assessing the prevalence of algorithmic bias is part of a larger framework that can explain and predict the proliferation of inequalities through T2I generative AI (see Figure 1).

First, the main predictor of T2I AI output is the training data. Over the past decades, theory and empirical research outlined the underrepresentation and stereotypical depictions of women and minorities in our mediated environments (e.g. Appel and Gnambs,

2023; Guilbeault et al., 2025; Mulvey, 1976). As Crawford (2021: 131) noted, “patterns of inequality across history shape access to resources and opportunities, which in turn shape data.” Incomplete, unrepresentative, or poorly curated training data is a well-known source of potential bias in AI-generated output (Ferrara, 2024). If the data sets that are used to train the AI contain bias, it is likely for the algorithmic outcome to be biased (Danks and London, 2017). Algorithmic biases can often be traced to sociocultural and historical inequalities that become embedded not only in the data used to train algorithms but also in the perspectives of those who design and curate them (Akter et al., 2021). One approach to mitigating such biases involves the deliberate selection of training data that counters or contrasts with existing distributions of relevant attributes, thereby requiring developers to actively identify and understand potential sources of bias in their datasets (O’Connor and Liu, 2024).

Bias can also originate from the algorithmic architecture itself, including the focus and weighting of particular features as well as the statistical estimators on which the model relies on. Likewise, technological solutions may be applied to mitigate bias on the algorithmic level (Afreen et al., 2025). Different T2I platforms may generate images with different degrees of algorithmic bias. Such differences may emerge from different training data and different algorithmic models. Potential differences may be subject to the companies that make the platforms available (Midjourney: Midjourney Inc., DALL.E: OpenAI, Stable Diffusion: Stability AI), their financial models, or the markets they target.

When interacting with AI output more generally, and with biased T2I AI output specifically, users of T2I platforms could reflect on representational and presentational fairness and recognize a biased output. However, initial research on the perception of bias suggests that users may fail to detect it (e.g. Messingschlager and Appel, 2025). Encountering stereotypes in media—which can include using T2I platforms and seeing AI-generated content that has been distributed by media outlets and shared by other users—can reinforce existent stereotypes both through short- and long-term effects (Dixon, 2019). This proliferation of existent bias could lead to more biased images being created and shared, which can ultimately become part of future training data sets. The model outlined in Figure 1 constitutes a bias-reinforcing circle with generative AI output as a link between algorithmic and human information processing.

Our work on bias in images generated by AI is influenced by the stereotype content model (Fiske, 2018; Fiske et al., 2002). It proposes two fundamental dimensions of social perception, *warmth* (e.g. friendly, trustworthy), and *competence* (e.g. intelligent, confident). These dimensions have been used to describe intergroup perceptions and stereotypes with women stereotypically perceived as higher in warmth than men, whereas men are stereotypically perceived as higher in competence than women (Fiske et al., 2002). Based on this model, we were interested how the wording of prompts containing “warm” versus “competent” could affect T2I platform output, thereby reflecting and proliferating gender stereotypes.

Before we introduce representational and presentational algorithmic bias in greater detail, as well as our ensuing hypotheses, two caveats need to be noted upfront. First, our analysis is aimed at images generated by T2I AI, with a focus on persons depicted and their gender expressions that are often associated with women or men. We did not include

a non-binary category. Second, for the sake of brevity, these gender representations are described as *images of women* or *images of men*, or similar wordings.

Representational bias

As outlined above, *representational bias* is the over- or underrepresentation of different social groups compared to the population, such as women or ethnic minorities (Mehrabi et al., 2022). Regarding gender, bias potentially stems from various types of media representations, which have a history of underrepresenting women. For example, printed news articles and television advertisements (Coltrane and Adams, 1997; Shor et al., 2014) featured women less frequently and/or less prominently. In online news media and images from digital platforms, women were found to be widely underrepresented as well (Ash et al., 2021; Singh et al., 2020). This disparity was notably stronger when content was algorithmically curated versus curated by humans (Singh et al., 2020). Research regarding bias in search engine outputs has documented that men appear more often than women in Google Image search results as well as in datasets from IMDB and Wikipedia (Guilbeault et al., 2024). Some online search algorithms tend to underrepresent women when performing a non-gendered search query (Ulloa et al., 2024). Initial research on representational bias in T2I generative AI focused on portraits of persons in occupational roles, showing that men are displayed more frequently than women if no gender is specified (Sun et al., 2024; Zhou et al., 2024). Extending this research to persons without the specification of occupational contexts, the following hypothesis is proposed:

H1.1: AI-generated images depict men more frequently than women.

Ulloa et al. (2024) found that adding the qualifier “intelligent” to an online search query amplifies women’s underrepresentation across different search engines. This aligns with the two central dimensions of interpersonal and group judgments, identified by Fiske et al. (2002). The competence dimension contains agentic and assertive traits, while the warmth dimension encompasses more communal traits. Traditionally, competence-related traits have been ascribed to men more often, while warm traits are seen as more feminine (Eagly, 1987; Fiske et al., 2002). Initial research by Messingschlager and Appel (2025) investigated this bias with data from 2023. They examined T2I AI output using only one platform (Midjourney) and focused on images of scientist exclusively. In their research, images generated with prompts qualifying competence-related traits contained faces of men more often than images generated with warmth-related trait adjectives. Our aim was to extend these results and compare three different generative AI models using prompts that did not specify an occupational role. The following hypotheses are proposed.

H1.2: Prompts containing the qualifier “warm” generate a higher proportion of pictures depicting women.

H1.3: The overrepresentation of men in AI-generated images is stronger when using prompts containing the qualifier “competent.”

Presentational bias

Presentational bias concerns the manner in which social groups are depicted (Sun et al., 2024). Regarding images, presentational bias is reflected in common visual display of stereotypical aspects of a specific group, while another group may be displayed without these aspects or with a different intensity thereof. We examined three characteristics of images which have been examined in prior theory and research on gender stereotypes in the media: Face-ism (e.g. Archer et al., 1983), smiling (e.g. Hess et al., 2005), and head canting (e.g. Costa et al., 2001).

The face-ism effect. One common difference in how men and women are portrayed is the *face-ism* effect, which describes gender differences in facial prominence. Facial prominence is defined as the face-to-body proportion in depictions of a person. Facial prominence can shape perceptions of competence and other traits (Archer et al., 1983; Zuckerman and Kieffer, 1994), which gives reason to work toward mitigation of the effects’ occurrence in AI-generated content. Portrayals of men are likely to exhibit a higher facial prominence than portrayals of women (Archer et al., 1983). Face-ism was found in various types of images, both in traditional and online media. Classic portrait paintings and photographs (Archer et al., 1983), commercials (Hall and Crum, 1994), self-selected images by social network users (Sczesny and Kaufmann, 2018) all featured face-ism at least to some degree. The face-ism phenomenon has been found in samples from Western cultures (Konrath and Schwarz, 2007), as well as in other cultures worldwide (Archer et al., 1983; Konrath et al., 2012). Therefore, it can be expected to be present in AI training data, as the datasets often contain images from various types of media from all around the world (Thomee et al., 2016). In regards to results of search engines, as an indicator of training data used for T2I platforms, Papakyriakopoulos and Mboya (2023) found stronger relations between the qualifier “close-up” and men than with women in their analysis of an image searching algorithm. Moreover, evidence from Ulloa et al. (2024) indicates that images in different search engines are prone to face-ism. The present study aims to quantify gender differences in facial prominence that are likely to be present in AI-generated images, employing the following hypothesis.

H2.1: Depictions of men exhibit a higher facial prominence than those of women.

Gender differences in smiling. Traditionally, women were often deemed the subordinate gender in comparison to men (Sidanius and Pratto, 1999). Smiling is one of the behavioral cues related to dominance and subordination (Hess et al., 2005; Ragan, 1982), with non-smiling faces being judged as more dominant than smiling faces (Keating et al., 1981). The tendency to deem women more likely to smile is mediated by their lower perceived dominance (Hess et al., 2005). A higher rate of women smiling was documented in visual representations—online and offline (e.g. Ragan, 1982; Tifferet and Vilnai-Yavetz, 2018). For example, samples of yearbook photographs across a century

(Ginosar et al., 2015), online self-presentation via profile pictures (McAndrew and Jeong, 2012; Tifferet and Vilnai-Yavetz, 2014, 2018), and photographs in news media (Gruszczynski et al., 2023) all showed women to express emotions, especially those of positive valence, with higher intensity and frequency than men. This effect has been observed in images generated by the DALL·E 2, Midjourney, and Stable Diffusion AI systems (Sun et al., 2024; Zhou et al., 2024). The present study aims to build on this research and examine if the effect is still present in the newest versions of the three generative AI systems, and with prompts that do not specify occupational roles. Therefore, the following hypothesis is proposed:

H2.2: Women are more frequently depicted with a smiling face than men.

Gender differences in head canting. Lateral head tilts (head canting) are another sign of nonverbal behavior that have received substantial attention by researchers in recent years. A canted head is one of the nonverbal cues displayed in a low-dominance position, predominantly by women (Goffman, 1976, as cited in Costa et al., 2001). The prevalence of this effect can be traced back to its presence in paintings from as far back as the 14th to the 20th century (Costa et al., 2001). In pictures in online and offline contexts, women are depicted with their heads tilted sideways (head roll) to a higher degree and more often than men (Ragan, 1982; Tifferet and Vilnai-Yavetz, 2018). Given the tendency for a higher degree of head canting in depictions of women from the 14th century to today, we assume that it will likely be reflected in AI-generated images. Thus, the following hypothesis is proposed:

H2.3: Women are depicted with a higher degree of head canting than men.

Given that differences in training data and AI platform algorithms can lead to variations in platform output, distinct AI platforms should yield disparate effects and outcomes. A comparison of AI platform versions older than the ones employed in the present study found that, for example, Stable Diffusion was able to generate more realistic faces than Midjourney and DALL·E 2 (Borji, 2023). Concerning representational bias, prior comparative analyses of the three AI platforms revealed variability in performance across different prompts (Zhou et al., 2024). This study aims to further assess the differences among the latest versions of the three AI systems by posing the following research question.

Research Question 1: How do biases in different text-to-image AI platforms differ?

Method

Overview

The hypotheses were examined via content analysis of an AI-generated sample. Three T2I AI platforms (Stable Diffusion SDXL, Midjourney Version 6.1, OpenAI DALL·E 3) were employed. For DALL·E 3, the API embedded in Microsoft Bing was used. Data

Table 1. Detailed sample sizes by prompt and AI platform.

Prompt	DALL·E		Stable Diffusion		Midjourney		
	t1	t2	t1	t2	t1	t2	
Person	32	32	32	32	32	32	192
Competent person	32	32	32	32	32	32	192
Warm person	32	32	32	32	32	32	192
Man	64	64	64	64	64	64	384
Woman	64	64	64	64	64	64	384
Total	224	224	224	224	224	224	1344

Prompts all commenced with “color portrait photograph of a,” followed by the prompt variations.

collection took place during March 1st and March 8th, 2025, at t1 and 4 weeks later during March 29th and April 6th at t2. We conducted two time points of data collection to increase the temporal generalizability of the results. Although this procedure does not eliminate the time and context-dependency of the results (see Discussion section), two points of measurement reduce the likelihood that the results are due to a very short-lived fluctuation of the algorithmic results. The AI was prompted to generate color portrait photographs. The preregistration, data, and supplemental materials are available on the Open Science Framework (<https://osf.io/uf3sp>).¹

Design

For the assessment of representational bias, a 3×3 factorial design was used with competence (person with no further specification, competent person, and warm person) and AI platform (DALL·E, Stable Diffusion, and Midjourney) as between-subjects factors. Presentational bias was assessed using a 2×3 design with gender (man, woman) and AI platform as factors. The conditions were induced by alterations in the prompt used to generate an image. For example, the prompt “color portrait photograph of a competent person” was used to generate an image in the competent condition.

Sample

A total of 1344 images were generated (see Table 1). Supplements S1.1 and S1.8 provide detailed information on the image generation process and our sample size considerations. Using the prompts “person,” “competent person,” and “warm person,” 576 images were created. One hundred ninety-two images were generated per prompt, with groups of 32 images for each AI platform per each time of measurement. With the prompts “man” and “woman,” 384 images were generated per prompt, resulting in 64 images per AI platform and time of measurement. Two images generated with DALL·E contained two faces. These were included in the sample, resulting in a total of 770 faces in 768 images. In the man and woman conditions, images were excluded if no face was visible in the image, if the face was covered by stylistic elements, or if only one eye was visible

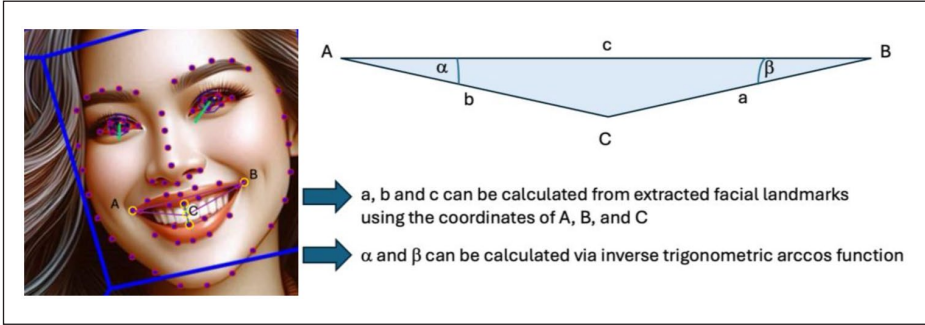


Figure 2. Triangle for lip curvature metric.

due to head turning or coverage. Images that demanded exclusion were generated anew during the same time of measurement they were originally created in. A total of 59 images generated with Midjourney needed to be excluded and were therefore generated anew. No images generated by Stable Diffusion or DALL·E demanded exclusion. Supplement S1.3 contains details on the data curation and exclusion procedure.

Measures

Representational bias. The gender of the person depicted was coded manually (see also Supplement S1.4). Each picture was coded into one of two categories, “representing a woman” or “representing a man.” The material was coded by the first author. The second author coded 20% of the material independently, which resulted in a 100% agreement rate. In the following sections, female-presenting individuals in images are referred to as “women” and male-presenting individuals in images are referred to as “men.”

Presentational bias

Smiling. Due to a lack of variance in smiling individuals in the DALL·E sample (all generated persons smiled) and the Midjourney sample (almost no generated persons smiled, see below), the preregistered method of binary coding (i.e. smiling vs not smiling) proved to be an inadequate measure to examine the hypothesized differences regarding smiling for generated images of men and women. Therefore, lip curvature as a continuous metric for smile intensity was employed, as smiling intensity was found to differ between genders in past studies (Ginosar et al., 2015; Tifferet and Vilnai-Yavetz, 2014). This method had previously been validated by Ginosar et al. (2015; see Supplement S1.5 for more information on coding and validation). In this method, the midpoint between the bottom of the top lip and the top of the bottom lip served as point C and the outer corners of the mouth as points A and B, as illustrated in Figure 2. Using the arccosine law, interior angles were calculated. Lip curvature ($M=15.33$, $SD=0.15$) was computed by taking the average of angles α and β . Smiling intensity was calculated only for smiling faces in the DALL·E and Stable Diffusion sample, because the images generated by Midjourney did not show any smiles.

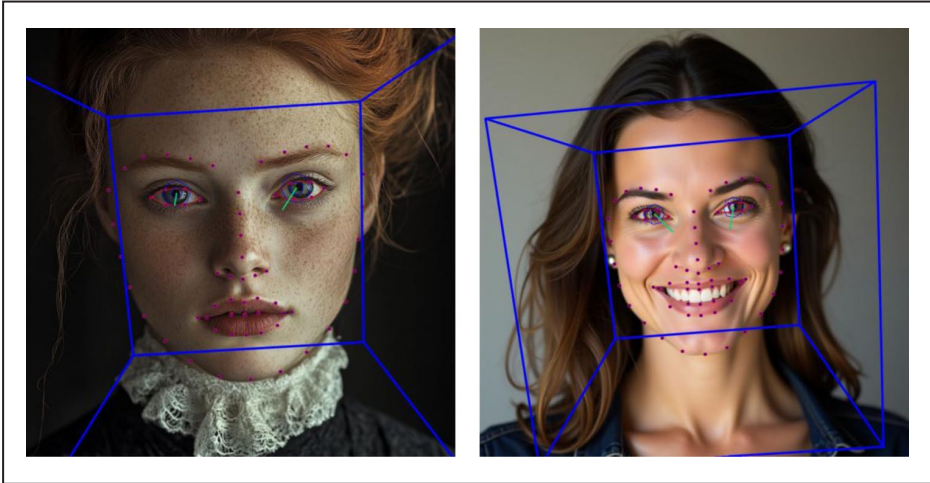


Figure 3. Visualization of detected facial landmarks.

The red dots represent the 64 facial landmarks contained in the OpenFace output. The blue box visualizes the software's internal 3D representation used to estimate head pose. Both images were generated with the prompt "color portrait photograph of a woman," the image on the left in Midjourney and the image on the right in Stable Diffusion.

Face-ism and head canting. For the dependent variables facial prominence and head canting, the facial landmark detection software *OpenFace 2.0* (Baltrušaitis et al., 2018) was employed (see Supplement S1.2). Facial prominence ($M=.12$, $SD=.05$) and head canting in degrees ($M=6.65$, $SD=6.14$) were measured in the second part of the sample generated with the prompts "man" and "woman." To ensure the quality of measurement, results were screened for errors by inspecting the landmark visualizations (see Figure 3). In 28 images generated by Midjourney, the software could not detect the face properly. In these cases, image properties were adjusted until software detection delivered accurate results (see Supplement S1.3 for more information). No images generated with DALL-E or Stable Diffusion needed adjustment. Extreme outliers in facial prominence were win-sorized. The variable head canting contained no extreme outliers.

Software used to analyze pictures of women and men can be a source of bias in itself (e.g. Lee et al., 2022; Schwemmer et al., 2020). Thus, the software results were validated by drawing a random sample of 30 images, measuring head canting and facial prominence by hand, and comparing the results (see Supplement S1.7). Facial prominence using OpenFace landmarks was measured by calculating the proportion of the face surface area to the total image size (see Supplement S1.6). This differs from previous measurements of facial prominence as introduced by Archer et al. (1983), where the proportion of face height to the total height of the person depicted was measured. The surface proportion method used in this study delivered results highly overlapping with the Archer method, $r(30)=.87$, $p<.001$, 95% CI [.75, .93]. Measures of head canting in degrees were equal to measurement by hand, $r(30)=.99$, $p<.001$, 95% CI [.97, .99].

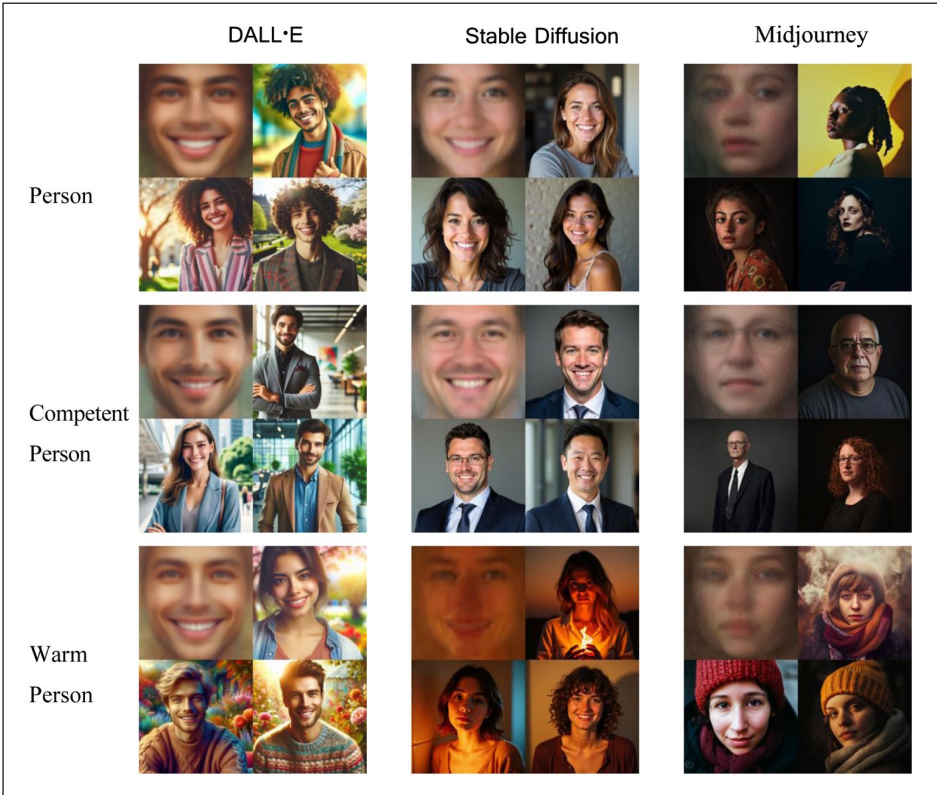


Figure 4. Average faces and example images for the prompts “person,” “competent person,” and “warm person.”

The first displayed image in each group of four shows the average face over all generated images within the specified condition. Average images were generated using the aligned images from the OpenFace output and layering them using ImageJ. Images 2–4 in each group of four are exemplary images from each respective subsample. The three example images were drawn from distinct thirds of the subsample. Clearer average images indicate more uniform output in each condition.

Results

Representational bias

Gender of the depicted person was coded as a dichotomous variable (men = 1, women = 0) for the 576 images generated with the prompt variations “person,” “competent person,” and “warm person.” Statistical significance for inference statistics was evaluated at the $p = .05$ level. In Figure 4, average faces and exemplary images are displayed. The main results are illustrated in Figure 5.

Our first analysis pertained to representational bias in the person condition. Across all three AI platforms, depictions of women ($n = 141$) appeared more often than those of men ($n = 51$). Table 2 shows the results of inference statistics (binomial tests) and

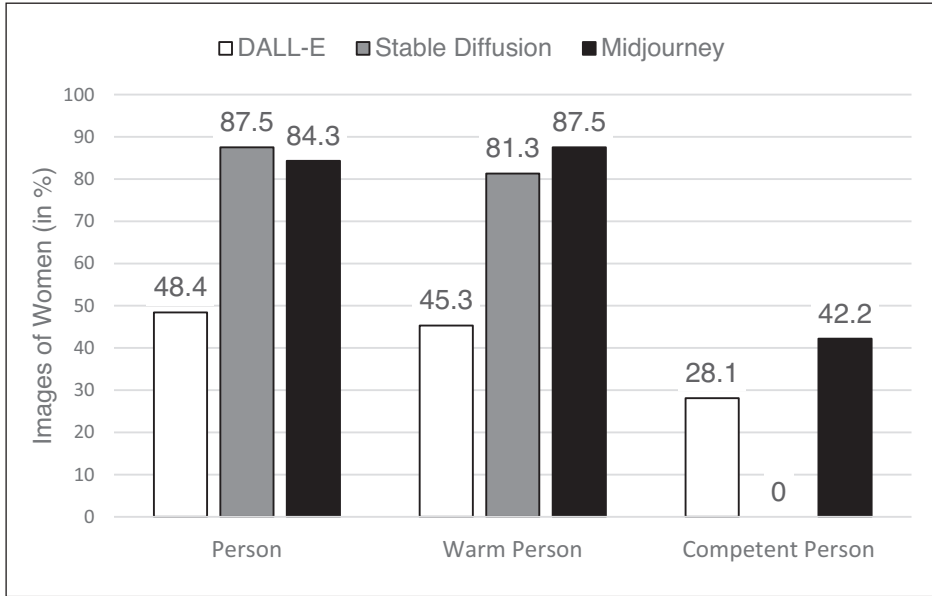


Figure 5. Percentage of images representing women.
Percentage of images representing women, depending on the prompt and the platform.

effect sizes. Testing for each individual AI platform revealed the highest deviation in the Stable Diffusion sample. The Midjourney sample showed similar tendencies. Only the DALL·E sample produced a nonsignificant deviation from the 50/50 distribution. Overall, the distribution of men and women differed from an equal distribution, but not in the expected direction. In contrast to earlier results (Sun et al., 2024; Zhou et al., 2024), and H1.1, more images of women than men were generated when prompted with “person.”

Next, we conducted a logistic regression analysis to examine whether including “warm” or “competent” to the prompt significantly changed the gender representation of the generated images. To this end, the factor prompt was dummy-coded with the person condition serving as the comparison group. Adding the qualifier “warm” did not change the gender representation, $B=0.10$, $SEB=0.23$, $Wald(1)=0.21$, $p=.648$, $OR=1.11$: Across the three AI platforms, depictions of women ($n=137$) appeared more often than those of men ($n=55$). Testing for each individual AI platform revealed the highest deviation in the Midjourney sample. The Stable Diffusion sample yielded a similarly overrepresentation of women. Again, the DALL·E sample was the only sample to show a nonsignificant deviation from the 50/50 distribution. In sum, adding the qualifier “warm” did not change the distribution of images representing women or men, respectively. Thus, H1.2 found no support.

Importantly, adding the qualifier “competent” to the prompt reversed the gender distribution as compared to the person condition, $B=2.20$, $SEB=0.24$, $Wald(1)=86.91$, $p<.001$, $OR=9.03$. The odds ratio shows that it was 9.03 times more likely that a man was depicted

Table 2. Descriptives for gender distribution by platform and prompt.

AI platform	Portrait of a . . .						Total	
	Person		Warm person		Competent person		n	%
	n	%	n	%	n	%		
DALL-E								
Images of Women	31	48.4	29	45.3	18	28.1	78	40.6
Images of Men	33	51.6	35	54.7	46	71.9	114	59.4
Binomial Test and Effect Size	p = .901 Cohen's g = .016		p = .532 Cohen's g = .047		p < .001 Cohen's g = .219		p < .001 Cohen's g = .094	
Stable Diffusion								
Images of Women	56	87.5	52	81.3	0	0	108	53.6
Images of Men	8	12.5	12	18.8	64	100	84	43.8
Binomial Test and Effect Size	p < .001 Cohen's g = -.375		p < .001 Cohen's g = -.313		p < .001 Cohen's g = .500		p < .001 Cohen's g = -.036	
MidJourney								
Images of Women	54	84.4	56	87.5	27	42.2	137	71.4
Images of Men	10	15.6	8	12.5	37	57.8	55	28.6
Binomial Test and Effect Size	p < .001 Cohen's g = -.344		p < .001 Cohen's g = -.375		p = .260 Cohen's g = .078		p < .001 Cohen's g = -.214	
Total								
Images of Women	141	73.4	137	71.4	45	23.4	323	56.1
Images of Men	51	26.6	55	28.6	147	76.6	253	43.9
Binomial Test and Effect Size	p < .001 Cohen's g = -.234		p < .001 Cohen's g = -.214		p < .001 Cohen's g = .266		p = .004 Cohen's g = -.061	

N = 576 (n = 64 per condition). This table shows how many images of women and images of men were generated by each platform with one of the three different prompts. Binomial tests were conducted regarding deviations from the expected 50/50 distributions. Effect size statistics are based on Cohen's g (not to be confused with Hedges' g), an effect size metric for binomial data (Cohen, 1988). The theoretical range of the values of Cohen's g is -.500 to .500. Interpretation guidelines are (in absolute values): negligible = 0-0.05; small: 0.05-0.15; medium: 0.15-0.25; large: 0.25 or more (Cohen, 1988: 147-149).

if the prompt included “competent person” as compared to “person.” An extreme output was observed for the Stable Diffusion platform with generated *zero* depictions of women, all images showed men in the competent condition. DALL·E generated some images of women, but less than images of men. In the Midjourney sample, no overrepresentation of men was observed. In sum, when asked to portray a competent person, the gender distribution differs significantly from the distribution observed when no qualifier was added to “person”: In line with H1.3, men were portrayed much more often than women.

A logistic regression for the sake of comparing potential differences between the prompts for the three platforms was suspended due to a zero cell count in one condition (no women as portrayals of competent persons in the Stable Diffusion subsample). A Firth penalized logistic regression yielded a matrix that was not positive definite.

Presentational bias

We examined presentational bias with the subsample of 768 images for which we specified the generated gender (i.e. 384 images were generated with the prompt “portrait of a woman,” 384 images were generated with the prompt “portrait of a man”). For average faces and example images, see Figure 6. The main results are shown in Figure 7.

Face-ism. Across all three AI platforms, images of men displayed higher facial prominence ($M=.13$, $SD=.05$) than images of women ($M=.10$, $SD=.05$), $t(768)=9.01$, $p<.001$, $d=0.66$, 95% CI [.51, .80]. To show potential interactions with the platform used, an ANOVA was conducted, yielding a significant main effect for the platform, $F(2, 764)=58.35$, $p<.001$, $\omega^2=0.13$, and a significant interaction effect with medium effect size between AI platform and prompted gender, $F(2, 764)=26.23$, $p<.001$, $\omega^2=0.06$. Gender had the strongest influence on facial prominence in the DALL·E condition, $F(1, 764)=95.12$, $p<.001$, $\omega^2=0.11$. In the Midjourney condition, the effect for gender was significant as well, $F(1, 764)=57.01$, $p<.001$, $\omega^2=0.07$, whereas the Stable Diffusion condition showed no significant effect, $F(1, 764)=0.00$, $p=.98$, $\omega^2=0.001$ (see Figure 7(a) and Online Supplement Table S4.1). In sum, H2.1 was supported for the dataset overall; the expected tendency for higher facial prominence of men was found for two of the three examined T2I platforms.

Smiling. Regarding the presence or absence of a smile, 509 (66.1%) of the displayed faces were coded as smiling in the subsample generated with the prompts “man” or “woman.” Pronounced differences between the platforms emerged: DALL·E displayed 258 (100%) smiling persons, Stable Diffusion displayed 249 (97.3%) smiling persons, whereas Midjourney displayed only two smiling persons (0.8%).

To examine potential gender differences in smiling, we examined smile intensity, using the degree of lip curvature within the subsample of smiling persons. A univariate general linear model was calculated for smiling faces in the DALL·E subsample ($n=258$) and the Stable Diffusion subsample ($n=249$). Out of all smiling faces, $n=249$ were men, and $n=258$ were women. See Figure 7(b) for an illustration of mean degrees of lip curvature and Online Supplement Table S4.2 for descriptive statistics.

An ANOVA was conducted, with the degree of lip curvature as the dependent variable, and portrait gender, platform and the interaction between both variables as

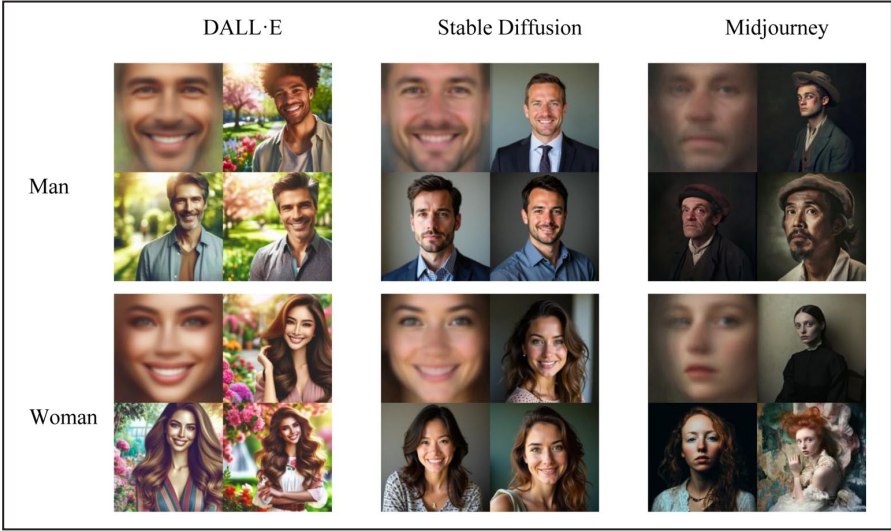


Figure 6. Average faces and example images for the prompts “man” and “woman.” The first displayed image in each group of four shows the average face over all generated images within the specified condition. Average images were generated using the aligned images from OpenFace output and layering them using ImageJ. Images 2–4 in each group of four are exemplary images from each respective subsample. The three example images were drawn from distinct thirds of the subsample. Please note that in the aligned images the differences in facial prominence and head canting are no longer distinguishable. Clearer average images indicate more uniform output in each condition.

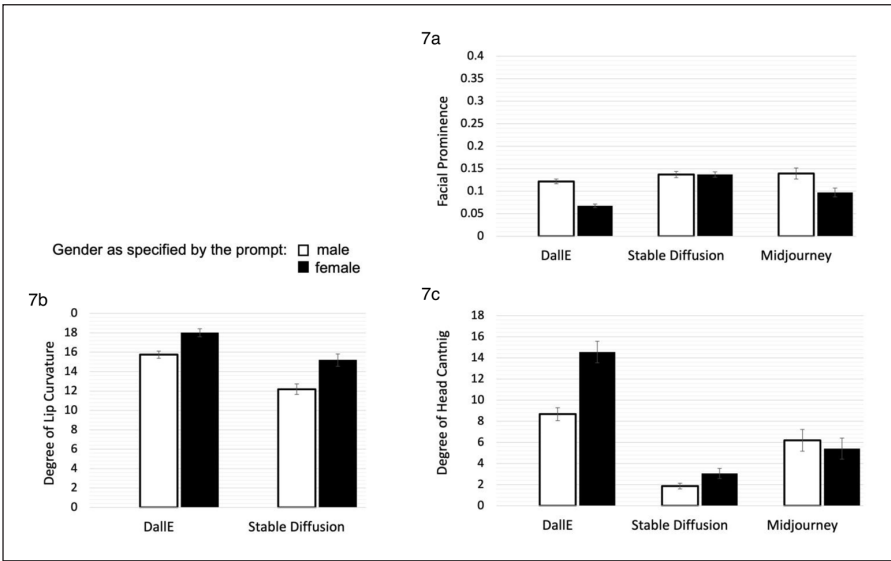


Figure 7. Presentational bias by platform (means and standard errors): (a) Facial prominence. (b) Degree of lip curvature and (c) Degree of head canting. Midjourney is omitted in Figure 7(b) due to the absence of smiling individuals in images generated by this AI platform.

predictors. It showed a significant main effect of gender, $F(1, 503)=110.46, p < .001, \omega^2=0.18$, with women showing more intense smiling ($M=16.60; SD=3.33$) than men ($M=14.02; SD=3.12$). Platform did also influence smiling intensity (main effect: $F(1, 503)=164.31, p < .001, \omega^2=0.24$). The results show no significant interaction effect between AI platform and smile intensity, $F(1, 503)=2.18, p = .140, \omega^2 = .002$.

Whereas all three platforms show strong tendencies to generate either smiling or non-smiling images of people, women are portrayed with stronger smiles than men for both T2I platforms that frequently portray smiling people. Thus, H2.2 was supported.

Head canting. Across the three AI platforms, depictions of women showed a significantly higher degree of head canting ($M=7.71, SD=7.01$) than those of men ($M=5.58, SD=4.90$), $t_w(688.92)=-4.89, p < .001, d=0.35, 95\% \text{ CI } [0.21, 0.50]$. To show potential interactions with the platform used, an ANOVA was conducted. It showed a significant main effect for the platform, $F(2, 764)=271.99, p < .001, \omega^2=0.41$, and a significant interaction effect with medium effect size between AI platform and prompted gender, $F(2, 764)=37.14, p < .001, \omega^2 = .09$. Figure 7(c) displays the mean degrees of head canting, descriptive statistics are reported in Online Supplement Table S4.3. Gender had the greatest impact on head canting in the DALL·E condition, $F(1, 764)=109.83, p < .001, \omega^2 = .12$. In the Stable Diffusion condition, a small significant effect for gender was revealed, $F(1, 764)=4.54, p = .033, \omega^2 = .01$. Midjourney showed no significant gender effect, $F(1, 764)=1.97, p = .161, \omega^2 = .001$. As expected in H2.3, head canting as a presentational gender bias could be observed in T2I platforms overall, but differences between the three tested platforms need to be noted.

Discussion

In this study, a systematic content analysis of T2I AI-generated images across three different platforms was performed, revealing systematic bias regarding stereotypical gender representation, face-ism, smiling intensity, and head canting. Representational bias was striking when images were generated with the prompt “competent person” in DALL·E and Stable Diffusion, both depicting men more often than women. The latter software exhibited the most severe imbalance in gender representation with no women generated when prompted to generate a competent person. This result is in line with the algorithmic overrepresentation of men found by Ulloa et al. (2024), who examined Google image search results for an intelligent person. Women were, however, overrepresented when prompts asked for a “person” or a “warm person” (Midjourney and Stable Diffusion). Such a pattern was expected for the “warm person” prompt but not for the “person” prompt without further specification. This gender representation differed from the findings by Zhou et al. (2024), who focused on depictions of people in occupational roles and found that in all three T2I platforms, women were underrepresented, and it differed from the results by Sun et al. (2024) who analyzed pictures generated by DALL·E.

The overrepresentation of women in these two conditions may be a manifestation of mitigation strategies against algorithmic bias already in effect. For instance, OpenAI has announced the usage of such strategies (Nichol, 2022). This highlights the chances, but also the potential ethical issues that come with adaptations made to algorithms in order

to tackle algorithmic bias. Whereas T2I platforms hold the potential to generate images that are more diverse than their training data base (e.g. more women in science or more smiling men), these adjustments can have unforeseen and unintended consequences, like an overrepresentation of social groups that have been underrepresented so far.

Presentational bias was observed for each of our three indicators, but to different degrees for the three T2I platforms. First, face-ism was found in images generated by DALL·E and Midjourney. Both platforms included classical portrait angles as well as more body-centered perspectives. Stable Diffusion appeared to follow a narrower interpretation of the “portrait” aspect of the prompt, resulting in no observable gender differences in facial prominence. Nevertheless, the prevalence of face-ism in two out of three of the examined T2I platforms points to the fundamental nature of the effect. This aligns with previous findings that also identified the effect in various types of media (e.g. Sczesny and Kaufmann, 2018; Ulloa et al., 2024). Given that lower facial prominence has been associated with less attention to mental aspects and more attention to bodily aspects of a depicted person (e.g. Archer et al., 1983), AI-generated face-ism may contribute to an objectification of women, along with negative downstream consequences (Loughnan et al., 2010; see also Mulvey, 1976).

When distinguishing between smiling and non-smiling portraits in a binary fashion, the T2I platforms exhibited either a tendency to display only smiling individuals (DALL·E and Stable Diffusion) or no smiling individuals at all (Midjourney). Consequently, lip curvature was measured, yielding continuous data for smile intensity among the smiling portraits. A post hoc analysis based on this measure showed the expected tendency for women to smile with higher intensity than men. The finding that women smiled more expressively is consistent with the results found by Ginosar et al. (2015), who found women to consistently smile with higher intensity than men. This difference being represented in AI-generated images could reinforce norms that women smile more than men. However, LaFrance et al. (2003) found gender differences in smiling to depend on the context. These differences were stronger when women were aware that they are being watched, which is the case when a portrait photograph is taken. This tendency may be reflected in AI-generated portraits.

DALL·E and Stable Diffusion showed the expected tendency for women to tilt their head to a higher degree than men. A substantial disparity was observed in the mean degree of head canting between the three platforms with Midjourney showing no gender difference. The DALL·E subsample showed the strongest gender differences, similar to those observed in paintings examined by Costa et al. (2001). Both the results for differences in smiling and head canting indicate that women continue to be displayed in a low-dominance way. Therefore, it is crucial to further examine how different genders are depicted in AI-generated images and how the depictions are perceived by the recipients.

Overall, our results illustrate gender bias in three popular T2I platforms. None of the three T2I platforms generated biased output on all variables, but all three platforms exhibited bias to some extent. The arguably most striking result was that Stable Diffusion generated 0% images of women when asked for a “competent person” whereas women were overrepresented for the prompts “person” and “warm person.” The platform further exhibited biased output for head canting and smiling. DALL·E showed significant representational bias when prompted to generate a “competent person” (vs both other

conditions), as well as the expected differences in facial prominence, head canting, and smiling. Midjourney showed an overrepresentation of women in the “warm person” and “person” conditions but not in the “competent person” condition. The platform generated images with a lower facial prominence for images of women but did not exhibit the expected bias for head canting. As an inspiration for future research, we wish to add the observation that Midjourney produced the most diverse sample, being the only T2I platform out of the three to include individuals deviating from contemporary beauty standards, thereby representing a broader range of social groups and individuals. Systematic research is encouraged to follow up on this observation.

As outlined in the introduction, we assume that the representational and presentational bias found in this study can be part of a vicious circle that ultimately reinforces inequalities in our digital societies (Figure 1). This assumption is backed by theory and societal analyses that position algorithmic bias in the bigger picture of political and economic forces (e.g. Benjamin, 2019; Noble, 2018). Empirical research showed that interactions with biased AI systems amplified biases in human judgment (Glickman and Sharot, 2025). This is particularly problematic, as, for example, the underrepresentation of women in images of competent individuals may lead to inaccurate perceptions of women’s competence in real life. The persistent portrayal of women engaging in postures associated with submissiveness may also result in normative perceptions regarding women’s social standing. Consequently, this may result in women being disadvantaged in various domains.

We wish to emphasize that measures can be taken to reduce algorithmic bias in T2I AI platforms. These platforms could function as a powerful means toward a more inclusive and equitable society because the reception of counter-stereotypical imagery can be a successful strategy to overcome gender bias (Finnegan et al., 2015). In addition, it is not only women and marginalized people who may benefit from reduced bias and a more inclusive society. Bareket and Fiske (2025) argued, that a shift away from gendered social roles may improve well-being in (white) men as well.

From our (micro-level, psychological) perspective, at least two approaches to break the vicious circle of increasing neglect and stereotyping of women and minorities through algorithmic bias appear promising. First, technological mitigation strategies (outlined above, see also Afreen et al., 2025) are warranted to reduce bias in AI output. Bias in visual datasets can be detected with software tools that highlight unrepresentative or anomalous patterns (Wang et al., 2022). As this might not be of intrinsic interest to tech companies, healthy democracies could oblige AI companies to disclose algorithms and training data and be held accountable to amend algorithms in a way that does not further disadvantage or privilege specific groups or individuals. As outlined recently for text-to-text applications (Bai et al., 2025), generative AI can appear unbiased on standard benchmarks (e.g. responses to explicit stereotypes) but, at the same time, show stereotype biases in tasks informed by psychological theory (e.g. word associations). Second, awareness of algorithmic bias among media professionals and everyday users should be increased. Users are often unaware of faulty or biased representations by AI algorithms (Glickman and Sharot, 2025). Therefore, more science communication and measures to increase AI literacy are required. Higher literacy may be accomplished by teaching algorithmic bias in schools and universities. As a case in point, research conducted with Finnish elementary and

secondary children suggests that interventions at school can increase the awareness of bias and the knowledge about its origin (Vartiainen et al., 2025).

Limitations

First, our analyses were limited to gender. Although not the focus of our analyses, other biased representations also became evident. People of Color lacked representation in almost all conditions. None of the generated images displayed overweight or visibly disabled individuals, except for some range in body type representation by Midjourney. Thus, gender bias appears to be not the only bias present in T2I generative AI. Therefore, future studies are encouraged to examine algorithmic bias with an intersectional approach, studying different areas of bias and their interactions (see Guilbeault et al., 2025, for the connection of gender and age stereotypes in different digital environments).

Relatedly, the present study focused on gender using a binary distinction between men and women. This focus was derived from literature on face-ism, head canting, and smiling, as well as from past research that found algorithms to use a binary classification (Papakyriakopoulos and Mboya, 2023). Gender coding was based on physical features, which does not necessarily determine the gender an individual identifies with. Given that AI-generated individuals are fictitious and therefore do not have identities, visual properties are the only reference point. Preliminary considerations led to the conclusion that visually classifying a person as trans* or non-binary would hardly be feasible. For this reason, the distinction was made solely for characteristics of the binary categorization. Therefore, the results of this study may apply to images of male-*presenting* and female-*presenting* individuals, but whether these presentations are inclusive of trans* or non-binary individuals remains unclear. Future research is encouraged to find ways for a more inclusive operationalization of gender.

Furthermore, the prompts used in this study specifically asked for color portrait photographs. T2I platforms can respond to even small changes in prompts. For example, colorful portraits might be associated with women more often than with men, which could explain the unexpected overrepresentation of women, when the platforms were asked for a “color portrait photograph of a person.” Future research is encouraged to test bias with a wider variety of different prompts or analyze larger data sets of user prompts and generated images to identify unexpected patterns of bias. On a related note, the prompt of “warm person” might have unintendedly elicited the generation of portraits that involve high temperature (e.g. individuals feeling warm due to wearing a wooly hat). Future research is encouraged to compare the present results with those based on prompts of adjectives that have been used to operationalize warmth as a fundamental dimension of social perception (e.g. friendly, trustworthy).

We further need to acknowledge that the aggregated results of T2I platforms are subject to change. T2I platforms providers constantly modify the algorithm and the underlying training data. Thus, the nature and extent of algorithmic bias likely vary over time (Sun et al., 2024). We gathered pictures at two points of time to reduce the likelihood that our results are due to a very short-lived fluctuation of the algorithmic results. Nonetheless, future research is encouraged to replicate our findings, possibly adding comparisons to other T2I platforms (e.g. Imagen, Firefly). As outlined in our theoretical background, we

argue that algorithmic bias is likely systematic; therefore, we expect that algorithmic bias will be a noteworthy phenomenon in the foreseeable future.

Conclusion

This preregistered study demonstrates systematically occurring algorithmic gender bias in AI images, generated with three of the most popular platforms (DALL·E, Midjourney, and Stable Diffusion). Unlike previous research on people in occupational roles, we showed that a neutral prompt asking for a portrait of a “person” yielded more images representing women than men. When the prompt asked for a “competent person,” however, we identified a stark underrepresentation of women (representational bias). Moreover, women were displayed in ways that have been associated with subordination and objectification (face-ism, smiling, head canting). Given the potential of a vicious circle that reinforces inequality in societies, including negative downstream effects of biased images on users and the general public, measures to address algorithmic bias need to be taken.

Declaration of conflicting interest

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was supported by the Bavarian Research Institute for Digital Transformation (BIDT), Project *Algorithmic Bias from a User Perspective: Assessment, Impact, Interventions (ADUBAI)* awarded to M.A.

ORCID iDs

Tanja V Messingschlager  <https://orcid.org/0000-0001-6737-8997>

Markus Appel  <https://orcid.org/0000-0003-4111-1308>

Supplemental material

Supplemental material for this article is available online.

Note

1. Two deviations from the preregistration need to be noted. First, we slightly amended the hypothesis wordings for the sake of clarity, and the order of H1.2 and H1.3 was switched. Second, we conducted ANOVAs instead of MANOVAs (see Online Supplement S2 for more details).

References

- Afreen J, Mohaghegh M and Doborjeh M (2025) Systematic literature review on bias mitigation in generative AI. *AI & Ethics* 5: 4789–4841.
- Akter S, McCarthy G, Sajib S, et al. (2021) Algorithmic bias in data-driven innovation in the age of AI. *International Journal of Information Management* 60: 102387.

- Appel M and Gnambs T (2023) Women in fiction: Bechdel-Wallace test results for the highest-grossing movies of the last four decades. *Psychology of Popular Media* 12(4): 499–504.
- Archer D, Iritani B, Kimes DD, et al. (1983) Face-ism: five studies of sex differences in facial prominence. *Journal of Personality and Social Psychology* 45(4): 725–735.
- Ash E, Durante R, Grebenshikova M, et al. (2021) Visual representation and stereotypes in news media, CEPR discussion paper series, article DP 16624. Available at: <https://ssrn.com/abstract=3960205>
- Bai X, Wang A, Sucholutsky I, et al. (2025) Explicitly unbiased large language models still form biased associations. *Proceedings of the National Academy of Sciences* 122(8): e2416228122.
- Baltrušaitis T, Zadeh A, Lim YC, et al. (2018) OpenFace 2.0: facial behavior analysis toolkit. In: *13th IEEE international conference on automatic face gesture recognition (FG2018)*, Xi'an, China, 15–19 May, pp. 59–66. New York: IEEE.
- Bareket O and Fiske ST (2025) Lost opportunities: how gendered arrangements harm men. *Proceedings of the National Academy of Sciences* 122(5): 2320788122.
- Benjamin R (2019) *Race after Technology: Abolitionist Tools for the New Jim Code*. Cambridge: Polity Press.
- Borji A (2023) Generated faces in the wild: quantitative comparison of stable diffusion, Midjourney and DALL-E 2. *arXiv[preprint]*. DOI: 10.48550/arXiv.2210.00586.
- Cohen J (1988) *Statistical Power Analysis for the Behavioral Sciences*. 2nd ed. Mahwah, NJ: Erlbaum.
- Coltrane S and Adams M (1997) Work–family imagery and gender stereotypes: television and the reproduction of difference. *Journal of Vocational Behavior* 50(2): 323–347.
- Costa M, Menzani M and Bitti PER (2001) Head canting in paintings: an historical study. *Journal of Nonverbal Behavior* 25(1): 63–73.
- Crawford K (2021) *The Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence*. New Haven, CT: Yale University Press.
- Danks D and London AJ (2017) Algorithmic bias in autonomous systems. In: Sierra C (ed.) *Proceedings of the 26th International Joint Conference on Artificial Intelligence*. Washington, DC: AAAI Press, pp. 4691–4697.
- Dixon TL (2019) Media stereotypes: content, effects, and theory. In: Oliver MB, Raney AA and Bryant J (eds) *Media Effects: Advances in Theory and Research*. New York: Routledge, pp. 243–257.
- Eagly AH (1987) *Sex Differences in Social Behavior*. East Sussex: Psychology Press.
- Ferrara E (2024) Fairness and bias in artificial intelligence: a brief survey of sources, impacts, and mitigation strategies. *Sci* 6(1): 3.
- Finnegan E, Oakhill J and Garnham A (2015) Counter-stereotypical pictures as a strategy for overcoming spontaneous gender stereotypes. *Frontiers in Psychology* 6: 1291.
- Fiske ST (2018) Stereotype content: warmth and competence endure. *Current Directions in Psychological Science* 27(2): 67–73.
- Fiske ST, Cuddy AJC, Glick P, et al. (2002) A model of (often mixed) stereotype content: competence and warmth respectively follow from perceived status and competition. *Journal of Personality and Social Psychology* 82(6): 878–902.
- Friedrich F, Brack M, Struppek L, et al. (2023) Fair diffusion: instructing text-to-image generation models on fairness. *arXiv[preprint]*. DOI: 10.48550/arXiv.2302.10893.
- Ginosar S, Rakelly K, Sachs S, et al. (2015) A century of portraits: a visual historical record of American high school yearbooks. *2015 IEEE International Conference on Computer Vision Workshop (ICCVW)*, Santiago, Chile, 7–13 December, pp. 652–658. New York: IEEE.
- Glickman M and Sharot T (2025) How human–AI feedback loops alter human perceptual, emotional and social judgements. *Nature Human Behaviour* 9(2): 345–359.

- Gruszczynski M, Brown DK, Pierce H, et al. (2023) Facing the competition: gender differences in facial emotion and prominence in visual news coverage of democratic presidential primary candidates. *Journalism & Mass Communication Quarterly* 100(3): 498–528.
- Guilbeault D, Delecourt S and Desikan BS (2025) Age and gender distortion in online media and large language models. *Nature* 646(8087): 1129–1137.
- Guilbeault D, Delecourt S, Hull T, et al. (2024) Online images amplify gender bias. *Nature* 626(8001): 1049–1055.
- Hall CCI and Crum MJ (1994) Women and “body-isms” in television beer commercials. *Sex Roles* 31(5): 329–337.
- Hess U, Reginald AJ and Kleck R (2005) Who may frown and who should smile? Dominance, affiliation, and the display of happiness and anger. *Cognition and Emotion* 19(4): 515–536.
- Keating CF, Mazur A, Segall MH, et al. (1981) Culture and the perception of social dominance from facial expression. *Journal of Personality and Social Psychology* 40(4): 615–626.
- Konrath S, Au J and Ramsey LR (2012) Cultural differences in face-ism: male politicians have bigger heads in more gender-equal cultures. *Psychology of Women Quarterly* 36(4): 476–487.
- Konrath SH and Schwarz N (2007) Do male politicians have big heads? Face-ism in online self-representations of politicians. *Media Psychology* 10(3): 436–448.
- Kordzadeh N and Ghasemaghaei M (2022) Algorithmic bias: review, synthesis, and future research directions. *European Journal of Information Systems* 31(3): 388–409.
- LaFrance M, Hecht MA and Paluck EL (2003) The contingent smile: a meta-analysis of sex differences in smiling. *Psychological Bulletin* 129(2): 305–334.
- Lee S, Oh S, Kim M, et al. (2022) Measuring embedded human like biases in face recognition models. *Computer Sciences and Mathematics Forum* 3(1): 2.
- Loughnan S, Haslam N, Murnane T, et al. (2010) Objectification leads to depersonalization: the denial of mind and moral concern to objectified others. *European Journal of Social Psychology* 40(5): 709–717.
- Maslej N, Fattorini L, Perrault R, et al. (2025) *The AI Index 2025 Annual Report*. Stanford, CA: AI Index Steering Committee, Institute for Human-Centered AI, Stanford University. Available at: <https://hai.stanford.edu/ai-index/2025-ai-index-report>
- McAndrew FT and Jeong HS (2012) Who does what on Facebook? Age, sex, and relationship status as predictors of Facebook use. *Computers in Human Behavior* 28(6): 2359–2365.
- Mehrabi N, Morstatter F, Saxena N, et al. (2022) A survey on bias and fairness in machine learning. *ACM Computing Surveys* 54(6): 115.
- Messingschlager TV and Appel M (2025) Algorithmic bias in image-generating artificial intelligence: prevalence and user perceptions. *Information, Communication & Society, Advance Online Publication*. Epubaheadofprint 14 November. DOI: 10.1080/1369118X.2025.2584146.
- Mulvey L (1976) Visual pleasure and narrative cinema. *Screen* 16(3): 6–18.
- Nichol A (2022) DALL·E 2 pre-training mitigations. *OpenAI*, 28 June. Available at: <https://openai.com/index/dall-e-2-pre-training-mitigations/>
- Nicoletti L and Bass D (2023) *Humans Are Biased. Generative AI Is Even Worse*. New York: Bloomberg. Available at: <https://www.bloomberg.com/graphics/2023-generative-ai-bias/>
- Noble SU (2018) *Algorithms of Oppression: How Search Engines Reinforce Racism*. New York: New York University Press.
- O’Connor S and Liu H (2024) Gender bias perpetuation and mitigation in AI technologies: challenges and opportunities. *AI & Society* 39: 2045–2057.
- OpenAI (2022) DALL·E now available without waitlist. *OpenAI*, 28 September. Available at: <https://openai.com/index/dall-e-now-available-without-waitlist/>
- Papakyriakopoulos O and Mboya AM (2023) Beyond algorithmic bias: a socio-computational interrogation of the Google search by image algorithm. *Social Science Computer Review* 41(4): 1100–1125.

- Ragan JM (1982) Gender displays in portrait photographs. *Sex Roles* 8(1): 33–43.
- Schwemmer C, Knight C, Bello-Pardo ED, et al. (2020) Diagnosing gender bias in image recognition systems. *Socius* 6: 2378023120967171.
- Sczesny S and Kaufmann MC (2018) Self-presentation in online professional networks: men's higher and women's lower facial prominence in self-created profile images. *Frontiers in Psychology* 8: 2295.
- Shor E, Arnout VDR, Charles W, et al. (2014) Time trends in printed news coverage of female subjects, 1880–2008. *Journalism Studies* 15(6): 759–773.
- Sidanius J and Pratto F (1999) *Social Dominance: An Intergroup Theory of Social Hierarchy and Oppression*. Cambridge: Cambridge University Press.
- Singh VK, Chayko M, Inamdar R, et al. (2020) Female librarians and male computer programmers? Gender bias in occupational images on digital media platforms. *Journal of the Association for Information Science and Technology* 71(11): 1281–1294.
- Stability AI (2022) Stable diffusion public release. *Stability AI*, 22 August. Available at: <https://stability.ai/news/stable-diffusion-public-release>
- Sun L, Wei M, Sun Y, et al. (2024) Smiling women pitching down: auditing representational and presentational gender biases in image-generative AI. *Journal of Computer-Mediated Communication* 29(1): zmad045.
- Thomee B, Shamma DA, Friedland G, et al. (2016) YFCC100M: the new data in multimedia research. *Communications of the ACM* 59(2): 64–73.
- Tifferet S and Vilnai-Yavetz I (2014) Gender differences in Facebook self-presentation: an international randomized study. *Computers in Human Behavior* 35: 388–399.
- Tifferet S and Vilnai-Yavetz I (2018) Self-presentation in LinkedIn portraits: common features, gender, and occupational differences. *Computers in Human Behavior* 80: 33–48.
- Ulloa R, Richter AC, Makhortykh M, et al. (2024) Representativeness and face-ism: gender bias in image search. *New Media & Society* 26(6): 3541–3567.
- Vartiainen H, Kahila J, Tedre M, et al. (2025) Enhancing children's understanding of algorithmic biases in and with text-to-image generative AI. *New Media & Society* 27(9): 5342–5368.
- Wang A, Liu A, Zhang R, et al. (2022) Revise: a tool for measuring and mitigating bias in visual datasets. *International Journal of Computer Vision* 130(7): 1790–1810.
- Zhou M, Abhishek V, Derdenger T, et al. (2024) Bias in generative AI. *arXiv[preprint]*. DOI: 10.48550/arXiv.2403.02726.
- Zuckerman M and Kieffer SC (1994) Race differences in face-ism: does facial prominence imply dominance? *Journal of Personality and Social Psychology* 66(1): 86–92.

Author biographies

Hannah Weinmann, M.Sc. (University of Würzburg, Germany), was a research assistant at the Human-Computer-Media Institute, University of Würzburg, Germany. Her research interests include bias in media and generative AI, as well as attitudes toward AI and AI-generated content.

Tanja V Messingschlager, Dr. rer. nat. (University of Würzburg, Germany), is a postdoctoral researcher at the Human-Computer-Media Institute, University of Würzburg, Germany. Her research focuses on psychological aspects of social media and new technologies, with a special interest in generative AI.

Markus Appel, Dr. phil. (University of Cologne, Germany), is a Professor at the Human-Computer-Media Institute, University of Würzburg, Germany, where he heads the Psychology of Communication and New Media research team. His research focuses on psychological aspects of media and technology.