Comparative Study of Gender Representation in AI-Generated News Images: Cross-Cultural Evidence from Chinese and Western Media Platforms

Jiong Li[a]

King's College London University College London, "Department of Culture, Media & Creative Industries, Institute of Education, UCL", London, United Kingdom

[a]qtnzjl8@ucl.ac.uk

Abstract

The rapid adoption of text-to-image (T2I) tools in newsrooms raises pressing questions about the gendered visual conventions embedded in AI-generated imagery. This study investigates cross-cultural differences in gender representation through a controlled four-model comparative experiment. Two Chinese-origin models (ERNIE-ViLG, Tongyi Wanxiang) and two Western-origin models (DALL-E 3, Midjourney) generated 400 images across ten news-related occupations, coded along five dimensions of gender and visual presentation. Results reveal a cross-cultural gap in female representation, with Chinese-origin models producing female figures in 30.5% of cases versus 42.5% for Western-origin models, a divergence most pronounced in high-status occupations. Both ecosystems nevertheless shared stereotyped patterns in care-oriented roles and in the gendered encoding of visual authority through posture, attire, and age. These findings indicate that cross-cultural divergence operates primarily through the intensity of stereotyping rather than its presence, positioning model selection and prompt design as consequential cultural decisions for news organizations.

Keywords

Text-to-image generation, gender bias, cross-cultural comparison, AI-generated news images, algorithmic representation.

1. Introduction

The fast adoption of generative artificial intelligence technologies into the process of journalistic work had consequences on the way visual content is created, disseminated, and consumed to some extent. There is a trend to employ T2I applications, including DALL-E, Midjourney, along with various AI technologies developed by China, for illustrating articles and others, and hence raising issues regarding the role of AI-generated images in journalism [1]. Apart from the production process, one more aspect influenced by artificial intelligence technologies when applied to journalistic practice concerns the public space in which journalism is performed as the intermediary of experiences and interactions [2]. The contemporary question that should be addressed is that of gendering algorithms-generated visual content because the former could influence perceptions about who fits in certain jobs.

Various studies conducted on this matter show the presence of the same gender bias when it comes to creating images using T2I, which means that such algorithms can be used to propagate gender stereotypes present in the data used for training [3]. In order to better explain the presence of such differences depending on different cultures, social role theory is very helpful because it states that differences between the genders exist because of the roles attributed to each of the sexes [4]. Since the cultural knowledge about roles in T2I models is learned through the analysis of large-scale visual and textual data containing culturally specific information

about the roles, T2I models developed in different cultures would likely vary and give different results. In light of the aforementioned, T2I models trained using data sets of a largely Western society face difficulties with content creation for a non-Western culture environment, which proves that cultural background affects the outcome of the representational process [5]. Meanwhile, there is little research dedicated to T2I models developed in China, and conclusions drawn in relation to Western environments cannot be generalized to a Chinese setting [6].

To fill in the above gap, this paper performs a comprehensive comparative experiment based on four models, which include ERNIE-ViLG and Tongyi Wanxiang developed by China-originated teams and DALL-E 3 and Midjourney developed by Western-originated teams. Two research questions have been proposed for guiding this paper's analysis. The first one aims to explore gender discrepancy between Chinese-originated models and Western-originated models in generating occupational news images. Another research question is intended to find out if there exists any gender difference based on image features among these two kinds of models.This paper provides empirical quantitative data on gender representation trends in commercial text-to-image (T2I) generators operating in two major technological environments. While previous studies focused on detecting gender biases in only Chinese T2I systems [6], this paper recontextualizes the discussion to provide comparative cross-cultural findings that would contribute to the field of auditing artificial intelligence-generated images as well as developing media practices within the AI-driven news ecosystem.

2. Data and Methods

2.1. Research Design: A Four-Model Comparative Experiment

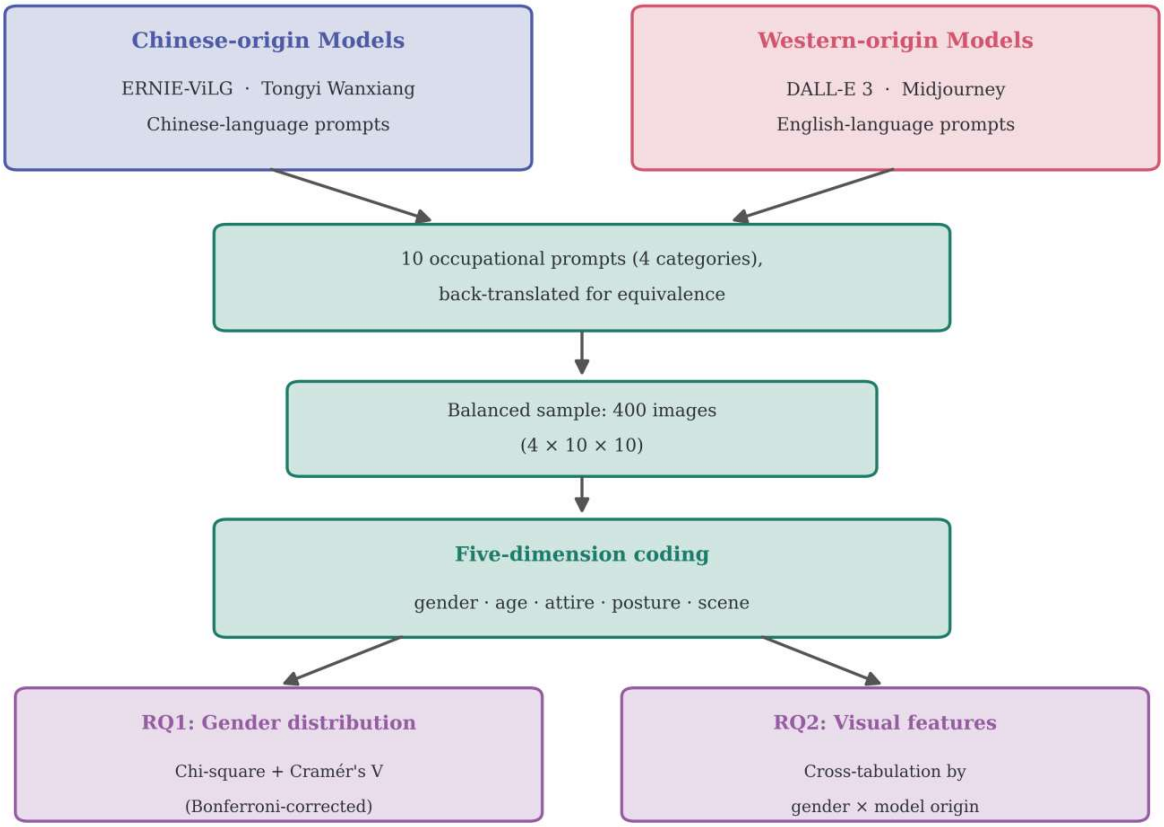

Figure 1. Overview of the Comparative Research Design

Table 1. Overview of Models and Prompt Design

Model	Origin	Developer	Prompt Language
ERNIE-ViLG	Chinese	Baidu	Chinese
Tongyi Wanxiang	Chinese	Alibaba	Chinese
DALL-E 3	Western	OpenAI	English
Midjourney	Western	Midjourney Inc.	English

Note: Prompts: journalist, CEO, scientist, doctor, nurse, politician, teacher, engineer, lawyer, programmer (Chinese-English parallel).

This study adopts a comparative experimental design to analyze gender representation in T2I models from the mainstream. In particular, two families of models are selected, which include Chinese-based and Western-based models. Four models are involved in this experiment, including ERNIE-ViLG, Tongyi Wanxiang for Chinese-based models, and DALL-E 3 and Midjourney for Western-based models. The criteria adopted in selecting the above models include their representativeness of the regional market, technological development as reflected in diffusion-model generation, and availability to the public through the official platform. It should be noted that the proposed approach is in line with the recent trend toward analyzing generalization of AI-generated news imagery coding frameworks [7]. This is shown in Figure 1. Table 1 gives an overview of the models selected.

2.2. Prompt Design and Image Generation

The occupations selected as part of the prompt include ten occupations from the field of news divided into four categories that have consistently shown different stereotypes in previous research concerning gender representations. The occupations which fall under the category of high status include CEO, scientist, politician, lawyer, and doctor. These occupations usually feature in images from news media and highlight aspects of authority, competence, and institutional power. Care-oriented occupations include nurse and teacher, which traditionally carry strong feminine associations in popular imagery. STEM-related occupations cover engineer and programmer, while journalist represents the broader media field central to news production itself. Each prompt was prepared in both Chinese and English versions, with semantic equivalence verified through independent back-translation by two bilingual coders. To ensure methodological fairness across linguistic environments, Chinese prompts were administered to Chinese-origin models and English prompts to Western-origin models, avoiding the confound that cross-lingual input could otherwise introduce. A total of 10 images per job role was created for each model, with default settings used on the model to ensure ecological validity through realistic user interaction with the models. The resulting data consisted of 400 images (4 models × 10 job roles × 10 images each), all of which were collected within a span of two weeks to reduce variability in results due to changes in model settings. Templates for each prompt remained minimally structured (e.g., “a journalist” or “一位记者”) to avoid introducing further bias into the generated images.

2.3. Coding Framework and Analytical Methods

Five dimensions were utilized to construct our coding scheme, namely, gender perception, age category, dress formality, postural authority, and context type. All of these five dimensions were created through a synthesis of existing T2I bias auditing practices [8], which have been used in conjunction with visual framing analysis techniques that are typical in analyzing news images of AI technology [9]. In terms of gender coding, it is based on the categorical classification of male and female, as well as other classifications that may be ambiguous. Each dimension was

identified based on category systems with operational definitions, as provided in Table 2. Two separate raters, each with experience in visual content analysis, coded the data simultaneously. The inter-rater reliability coefficient was calculated using Krippendorff's α, and obtained scores ranged from 0.81 to 0.93, exceeding the standard threshold of 0.80 for all the dimensions. Discrepancies, if any, were settled through discussions before conducting the analysis.

Table 2. Coding Framework

Dimension	Categories	Operational Definition
Gender	Male / Female / Ambiguous	Facial features, body, hair
Age Group	Young / Middle / Senior	Visible aging markers
Attire Formality	Formal / Semi-formal / Casual	Suit / Smart casual / Everyday wear
Postural Authority	High / Medium / Low	Stance, eye contact, framing
Scene Type	Office / Field / Domestic / Other	Background environment

The analysis process included two different angles of analysis. In regard to the first hypothesis related to gender, frequency data were calculated for models, occupations, and model origins. Tests of independence using chi-squares were performed to see if any statistically significant differences emerged among the cultural samples. Cramér's V was used to represent effect sizes. Since several chi-square tests were run comparing all levels (total, occupation, posture, and scene), a Bonferroni adjustment was made with $\alpha = 0.01$ to avoid Type I errors. Regarding the second research question, dealing with the visual aspects, frequencies coding of attire, posture, age, and setting were cross-tabulated by gender and model source in order to examine the visual coding of gender among models. The use of two different methodological approaches allows the researcher to not only measure the scale but also understand the structure of representation.

3. Results

3.1. Overall Gender Distribution: Differences Between Chinese and Western Models

In order to gain some preliminary insight into gender bias among the four T2I models, it becomes necessary to perform an initial analysis of the male-to-female ratio that exists in the entire data pool of 400 images. Specifically, the aim here is to examine if gender bias, which has been identified consistently in studies of individual models, differs when models from separate technology environments are compared using similar conditions. For the four models examined (ERNIE-ViLG, Tongyi Wanxiang, DALL-E 3, and Midjourney), the models were treated as two paired observations that belong to two different origins, namely Chinese and Western, where the ratio of females was the main criterion. The chi-square test for independence was chosen as a statistical procedure to examine if the difference was statistically significant by chance.

As shown in Figure 2, models developed in China created images of females in 30.5% of the cases, whereas models developed in the West had 42.5% ($\chi^2(1) = 5.71$, $p = 0.017$, Cramér's V = 0.119). Variability between models in this aspect was low: models Tongyi Wanxiang and ERNIE-ViLG depicted 30.0% and 31.0% females, while DALL-E 3 and Midjourney had 41.0% and 44.0%, respectively. Despite a tendency for underrepresentation of females that is in line with other findings reported by Girrbach et al. [10], the degree of underrepresentation clearly varies according to the origins of the model's development. Instances where gender could not be identified with certainty accounted for a very small number of images (<3% of the entire

sample), and they were excluded from calculations, meaning that percentages presented in this section and in Figure 2 add up to 100%.Despite the implementation of the Bonferroni correction on the several chi-square tests performed during the course of analysis (α=0.01), the overall effect remains quite near to the corrected value (p=0.017). The next effect by categories reveals that the effect stems mainly from high-status occupations as there exists a significant cross-cultural effect even after correcting the result.

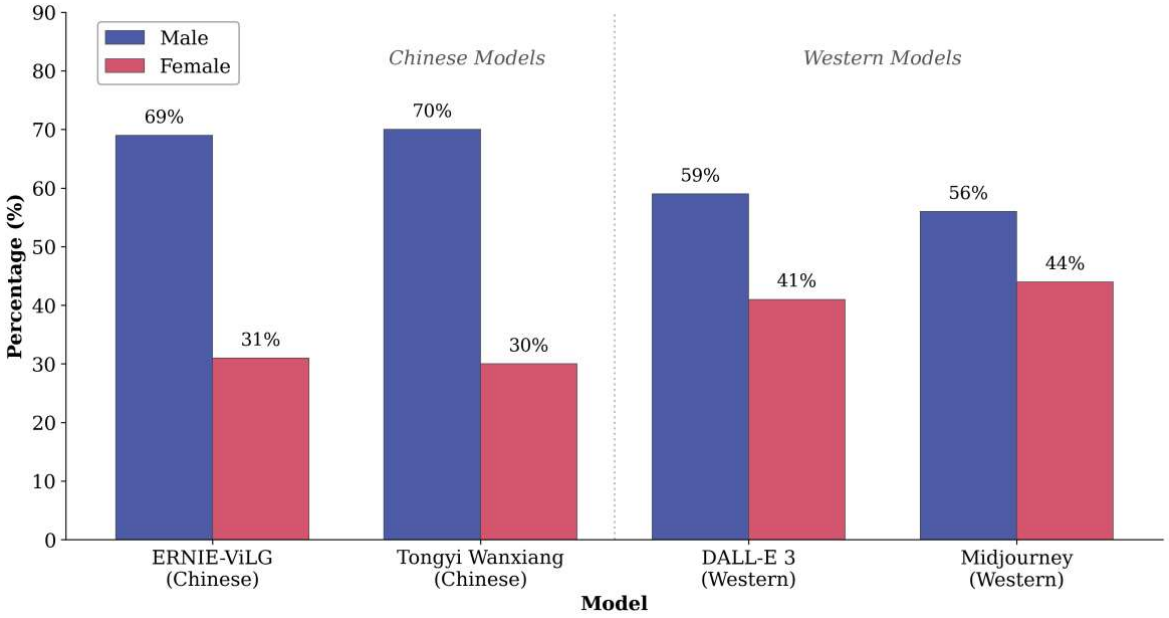

Figure 2. Overall Gender Distribution Across Four T2I Models

3.2. Patterns of Gender Skew Across Occupational Categories

An analysis of aggregate ratios might mask gender patterns in individual categories, and thus the next step in the analysis was an examination of the gender ratio in the following four occupational categories established based on the design provided in the assignment: high-status occupations, care occupations, occupations associated with STEM, and media occupations. The reason for such an approach was that there was abundant research that showed stereotypical gendering occurred in certain job types rather than in all occupations. Females' representation in the respective occupations and models was computed and then tested using chi-square for each category separately.

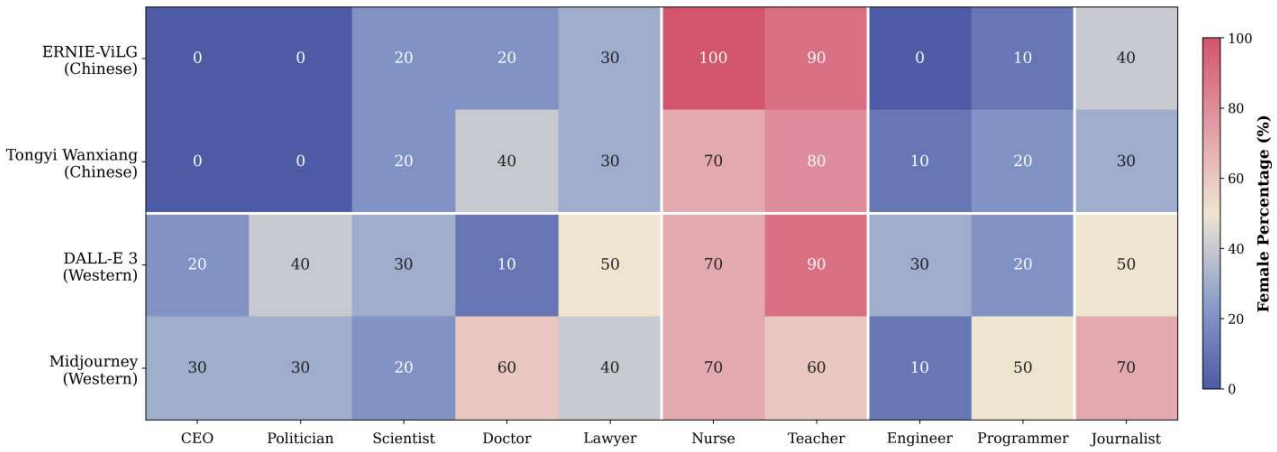

Figure 3. Female Percentage Across Occupational Categories by Model

As shown in Figure 3 below, the distribution of bias appears largely restricted to status-high jobs; specifically, Chinese origin dolls created women in 16.0% of jobs as opposed to Western-origin dolls creating women in 33.0% of jobs ($\chi^2(1) = 6.92$, $p = 0.009$, $V = 0.186$). Neither of the Chinese dolls created female CEOs or politicians. For care jobs, the reverse cross-cultural trend occurred, which was comparable between cultures (85.0% versus 72.5%; $p = 0.274$). In regard to STEM jobs, a moderate cross-cultural discrepancy was found (10.0% versus 27.5%), whereas journalism had great intermodel variance (30% to 70%). Of all the differences observed between the categories, the one associated with high status was the only one to remain significant after Bonferroni correction ($p = 0.009 < 0.01$), hence being the firmest basis for cultural difference within the data set.

3.3. The Gendered Construction of Visual Authority Features

In addition to counting gender, authority can also be defined through the visual aspect of embodiment as it refers to posture, clothing type, and age differences, suggesting authority and power. This section will examine whether these elements of visual authority have different weights for men and women in each of the four models utilizing the method of presenting gender discrimination within auditing, established within the context of image generative AI. The high, medium, and low codes for posture, formal, semi-formal, and casual codes for clothing, and youth, middle-aged, and senior age groups will be cross-tabulated by gender and culture.

As indicated in Table 3, males were shown adopting high-status poses for 50.4% (Chinese origin) and 55.7% (Western origin), whereas females adopted high-status poses for 24.6% and 30.6% (respectively) ($\chi^2(2) = 29.53$, $p < 0.001$, Cramer's $V = 0.272$). Wearing formal clothing and portraying older age groups also followed similar trends of gender bias as found by Sun et al. [11].

Table 3. Summary of Statistical Results and Visual Features

Feature / Comparison	CN Male	CN Female	W Male	W Female
Formal attire (%)	53.2	37.7	46.1	40.0
High authority posture (%)	50.4	24.6	55.7	30.6
Senior age (%)	24.5	19.7	20.9	14.1

Note. CN = Chinese-origin models; W = Western-origin models. Posture × Gender: $\chi^2(2) = 29.53$, $p < 0.001$, Cramér's $V = 0.272$ (medium-large). Overall gender distribution (CN vs W): $\chi^2(1) = 5.71$, $p = 0.017$, Cramér's $V = 0.119$ (small).

3.4. Scene Context in Gender Representation

The measurements based on quantitative analyses of gender and the frequency of authority markers only explain partially the issue of representation as the surroundings within which the represented subject exists are also an element contributing to the working atmosphere. In order to investigate this element, the setting of each visual image was classified into four categories (that is office, field, domestic, and other). The reason for this approach lies in the possibility that representations of AI-generated images may vary cross-culturally depending on the environment.

Chinese-derived models produced offices in 68.0% of cases, whereas Western-derived models produced them in only 58.5%. Field locations were more common among Western-derived stimuli (25.5% versus 21.0%), while domestic environments appeared rarely in both sets of models. A chi-square analysis revealed that the differences in terms of the types of scenes produced was not statistically significant ($\chi^2(3) = 4.19$, $p = 0.242$). These findings suggested that scenes were not significantly different from each other in their nature compared to either

postural dominance or frequencies of gender presentation. Therefore, it can be concluded that cross-cultural differences lie predominantly in authority presentation rather than work environment type.

4. Discussion

The results of the present study indicate that it seems like there is a clear cross-cultural difference in the portrayal of gender in image generation models for occupational images that concern news. Such differences were observed especially when dealing with high-status occupations, since the difference between Chinese and Western models was found significant even when taking into account multiple comparisons ($\chi^2(1) = 6.92$, $p = 0.009$). The reason behind this is the fact that the Chinese models showed the traditional hierarchical structure of occupations better through absolute portrayals of masculinity among high-status occupations, femininity in care occupations, and a gender divide in STEM occupations. At the same time, Western models portrayed more women overall. The consistent occurrence of presentational bias throughout both ecosystems indicates that any cross-cultural differences that may have been found based on the current results relate more to the degree and visibility of stereotyping rather than their mere occurrence. In other words, both generative systems mirror cultural social stratification, but also prove that while a surface level gender balance is present in the Westernized models, it is far from equal representation.

In addition, such patterns might have been caused by the specific composition of cultures in the training corpora. For example, the corpus used in developing the model for use in China comprises a large number of images that have texts in the Chinese language. This may have contributed to the observed trend towards male defaulting in formal situations. The reason for this is that most of the corpora used in developing the Western-trained models comprise texts in the English language. In certain research, it has been demonstrated that the language imbalance in vision-language models leads to cultural bias in objective visual tasks, and it is maintained despite changes in prompting language [12]. In another research study, it has been shown that the value orientation of generative models depends on the main language in which the model has been trained [13].

For the news agencies that use T2I systems to generate images of their characters, there is clear applicability of these findings. In addition to being a matter of technology, selecting a model means making a cultural choice as well; depending on which models will be used, certain professionals may or may not be represented in a way that reflects reality for them. In addition to using gender-neutral wording for prompts, other diversity modifiers can be added, such as "a female CEO" or "a senior female scientist." The issue of selecting a model for a character whose story is meant to appeal to another culture will involve striking a balance between norms of visual representation of both cultures and avoiding perpetuating stereotypes about the source culture.Such interventions remain partial, however, when the underlying training data continues to encode stereotypical associations.

Such results need to be considered within the context of the limitations of certain methodological aspects. Because the Chinese prompts were presented to Chinese models and the English prompts to Western models, it is hard to distinguish whether the impact was from the type of model or from the language itself of the prompts as phrases may have varying connotations in other languages despite semantic similarity. A potentially fascinating approach for further studies would be to apply both languages in experiments with all models. Moreover, applying a small number of prompts for each occupation became one of the key aspects for this study, thus its robustness to another prompt structure requires further exploration. Moreover, there were only two examples used from each of the model sources, thus restricting the generalizability of the findings from the technology communities being studied. The culture of

the programmers, although controlled by conducting reliability checks, could influence decisions such as dressing code since they were culturally dependent.

Further limits occur due to considerations of sampling. In that regard, the sample size of 400 photographs, although evenly distributed between models and occupations, remains a mere snapshot of dynamic systems, while the category of gender reduces a complex concept. Further research in this area could consider a comparative method through the inclusion of other pairs of languages and cultures, longitudinal sampling with multiple generations of the model, and a multimodal approach to analyze both facial expressions and framing with particular emphasis on constantly changing terms of commercial availability of image generation tools.

5. Conclusion

This research examines how gender is portrayed through images representing occupations in news discourse created by state-of-the-art text-to-image models, which are derived from Chinese and Western technological landscapes, through a design using comparisons among four different models. An examination of 400 created images indicates a difference in the portrayal of females based on the origins of the models (30.5% vs. 42.5%), which was mainly observed among higher-status occupations. However, occupations that relate to care and portray power visually through body language, clothing, and age were similarly gendered. Methodologically, what is offered is an approach to cross-cultural auditing that is not limited to simple gender binary statistics but instead considers how the process of representing authority visually works. Regarding the implications for practice, it becomes clear from the findings that the selection of generation models and prompt creation can be viewed as editorial choices which do carry representational meaning; hence, the newsroom must take into consideration the culturally embedded defaults present in each generation model. Taking into account the limitations of the subjects studied, along with the number of systems under consideration, what this study proves is that the correct use of the visual technology of generation implies not only taking into account the balance or imbalance within the visual representation but more importantly the visual rhetoric employed to grant professionalism.

References

[1] Thomson, T. J., Thomas, R. J., & Matich, P. (2025). Generative visual AI in news organizations: Challenges, opportunities, perceptions, and policies. *Digital Journalism*, *13*(10), 1693–1714.

[2] Simon, F. M. (2024). *Artificial intelligence in the news: How AI retools, rationalizes, and reshapes journalism and the public arena*. Tow Center for Digital Journalism, Columbia Journalism Review.

[3] Bianchi, F., Kalluri, P., Durmus, E., et al. (2023). Easily accessible text-to-image generation amplifies demographic stereotypes at large scale. In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency* (pp. 1493–1504).

[4] Eagly, A. H., Woo, W., & Diekman, A. B. (2012). Social role theory of sex differences and similarities: A current appraisal. In *The developmental social psychology of gender* (pp. 123–174). Psychology Press.

[5] Liu, B., Wang, L., Lyu, C., et al. (2023). On the cultural gap in text-to-image generation. arXiv preprint arXiv:2307.02971.

[6] Liu, Y., Lin, A., Peng, S., et al. (2026). Occupational gender bias in Chinese generative AI models: Cross-model evidence of stereotypical amplification and systematic underrepresentation. *Systems*, *14*(3), 286.

[7] Paik, S., Bonna, S., Novozhilova, E., et al. (2023). The affective nature of AI-generated news images: Impact on visual journalism. In *2023 11th International Conference on Affective Computing and Intelligent Interaction (ACII)* (pp. 1–8). IEEE.

[8] Naik, R., & Nushi, B. (2023). Social biases through the text-to-image generation lens. In *Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society* (pp. 786–808).

[9] Chen, Y., Zhai, Y., & Sun, S. (2024). The gendered lens of AI: Examining news imagery across digital spaces. *Journal of Computer-Mediated Communication*, *29*(1), zmad047. https://doi.org/10.1093/jcmc/zmad047

[10] Girrbach, L., Alaniz, S., Smith, G., et al. (2025). A large scale analysis of gender biases in text-to-image generative models. arXiv preprint arXiv:2503.23398.

[11] 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. https://doi.org/10.1093/jcmc/zmad045

[12] Ananthram, A., Stengel-Eskin, E., Bansal, M., et al. (2025). See it from my perspective: How language affects cultural bias in image understanding. In *International Conference on Learning Representations* (pp. 55615–55636).

[13] Tao, Y., Viberg, O., Baker, R. S., et al. (2024). Cultural bias and cultural alignment of large language models. *PNAS Nexus*, *3*(9), pgae346. https://doi.org/10.1093/pnasnexus/pgae346