Investigating the Presence of Bias and Potential Copyright Concerns in LLM Image Generation Capabilities

ANONYMOUS AUTHOR(S)

As large language models (LLMs) are being adopted across an increasing range of areas, their potential to perpetuate social biases requires rigorous, systematic evaluation. To evaluate biases in image generation, we compile a dataset of real-world movie posters and use quantitative and qualitative socio-technical evaluation methods to investigate biases in LLM-generated analogs. Our results indicated both the general presence of racial and gender bias and the exacerbation of biases found in real-world posters. We also observe that generated posters exhibited high levels of similarity with their real-world counterparts, potentially to the point of copyright infringement. This suggests that LLMs may be mirroring and exaggerating existing societal biases.

CCS Concepts: • **Applied computing** → **Sociology**; • **Computing methodologies** → **Artificial intelligence**.

Additional Key Words and Phrases: AI, LLMs, image-generation, socio-technical evaluation, bias analysis

ACM Reference Format:

Anonymous Author(s). 2026. Investigating the Presence of Bias and Potential Copyright Concerns in LLM Image Generation Capabilities. 1, 1 (January 2026), 7 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 Introduction

The rise of diffusion models in image generation has enabled text-to-image generation systems to grow in popularity, revolutionizing LLMs by extending their capabilities beyond generating text outputs [10]. However, its increasing prevalence has led to the emergence of generative AI bias issues, such as data bias and algorithmic bias [7]. Research [4, 11, 14, 18] and real-world examples [3, 5] have highlighted the presence of demographic bias in various AI systems. Although mitigation strategies [1, 15] have been developed in response to these bias incidents, their implementation has the potential to create adverse effects, such as over-correction leading to the production of inaccurate and distorted outputs [9]. Recent studies have also pinpointed LLM reproduction of copyrighted material as another growing area of bias concern [19]. These observations, alongside the growing public use of LLMs for image generation, highlight the need for evaluation processes that effectively audit LLM-generated images for demographic bias and potential copyright reproduction concerns.

In this paper, we investigate the presence of demographic bias using a socio-technical evaluation model applied to LLM-generated image outputs. We chose a domain of movie posters as one form of visual media with a real-world "ground truth" against which we could compare. We split our evaluation into quantitative and qualitative methods. For the quantitative analysis, we computed proportions to assess demographic imbalances in a movie poster. For qualitative analysis, we used a three-component rubric to assess the degree of similarity between a generated poster and its real-world counterpart. We also measured how centrally characters of different demographic groups appeared in the posters, as focal points often suggest importance in this form of media.

Manuscript submitted to ACM

We implemented our methods on 500 IMDb synopses across 5 genres. We observed indications of gender and race bias. We also found that our data tended to overemphasize real-world bias trends. These findings affirm the presence of a general bias in LLM image generation. We also observe that movie posters generated in 2025 were very likely to match real movie posters for the same movie, a trend not observed in earlier pilots, with potential copyright implications.

2 Related Work

Prior work in film and media studies has shown evidence of gender and racial bias patterns [6, 8, 12, 13]. Studies have consistently concluded that there is a prevalent male bias in movie posters, using methods such as positional analysis of agency [8] and deep learning techniques [13] to measure and identify gender inequality. These findings are further supported by bias analyses of movies as a whole [13], which found a ratio of 2.2 male characters per female character in the top 100-grossing films in 2015. Regarding race and ethnicity bias in movies, [13] found that only 26.3% of speaking characters in the sampled films were from underrepresented racial groups, and [6] showed that these trends persist in movie posters. However, despite these consistent racial and ethnic bias trends, [6] has shown that there is an improvement in the overall diversification of individuals within movie posters. The documentation of bias trends in real-world movie posters makes this domain a suitable setting for our study, in which we will determine whether these trends were preserved, mitigated, or exacerbated in AI-generated images. As AI begins to integrate into the film industry, the importance of analyzing this domain increases, as it sheds light on the implications of using these systems to generate creative content.

Bias in image generation has been well explored [7, 15, 16]. Models capable of image generation become susceptible to societal biases due to their training datasets [16], making them harmful if unchecked. Mitigation and audit strategies [1] have been developed in response to these bias incidents. Techniques such as reinforcement learning from human feedback (RLHF) [2] and adversarial learning [17], when integrated with leading image-generation models (e.g., DALL-E with GPT-4), have the potential to mitigate bias. However, sometimes these implementations fail, with Gemini's over-correction tendencies leading to inaccurate outputs [9]. Critiques of current mitigation strategies have also noted that they do not fully and holistically address bias concerns [15].

We aim to synthesize these two areas, framing our analysis as a socio-technical audit of generative AI image generation grounded in the domain of visual film representations. By doing so, we analyze whether real-world bias patterns are observable in AI outputs and whether the limitations of current mitigation strategies are evident.

3 Methods

3.1 Genre and Synopsis Selection

We utilized a dataset containing the title, genre, Motion Picture Association film rating (G, PG, PG-13, R, NC-17), and original synopsis of 6097 movies. We selected the five most frequently occurring genres this dataset: Action, Animation, Comedy, Drama, and Horror. From each of these genres, we randomly selected 100 titles, ensuring that the number of titles from each rating group (G, PG, PG-13, R, NC-17) were proportional to the number of available titles from each rating group. For example, if the Horror genre had approximately 83% of its titles in the R-rating category (432 titles out of 522), approximately 83 of the 100 titles selected by the script also had an R-rating.

3.2 Synopsis Modification

We edited each synopsis using the following guidelines:

(1) Changes to synopses are minimal to preserve plot lines and prevent confounding variables.
(2) Replace proper names of human characters and groups specifically relevant to the movie with a neutral placeholder term, such as "an individual," "another individual," "a group," or "the group" depending on the grammatical context.
(3) Remove demographic markers, such as gender pronouns, parental labels, partnership labels, and mentions of race/ethnicity. Replace these with neutral terms, such as "individual," "parent," "partner," "person," etc.
(4) Preserve geographic demographics, as they are racially ambiguous (ie. American, Canadian, etc).
(5) Several aspects that do not contribute to demographic markers should be preserved. This includes positions of power (with some exceptions), occupations, and locations.
(6) Because we are not evaluating age-related biases, we retained age identifiers.

3.3 Image Generation

We generated one image per synopsis using the ChatGPT-4o model. This model was kept consistent throughout the entire process. Using a Selenium script to automate the process, we used the following prompt for each synopsis: *"Generate a movie poster from the following synopsis. Please be as detailed as possible, showing the faces of the characters. Please also have this be as similar as possible to real-world movie posters: [insert synopsis here]."* This prompt was kept consistent across all synopses to prevent any confounding variables. Any requests from ChatGPT for more details were ignored. We recorded 47 prompt rejections, with ChatGPT providing the rationale that the prompt violated OpenAI's content policies. Image generation for the 500 selected titles spanned from June to August 2025.

3.4 Image Annotation/Evaluation

Annotation was conducted simultaneously by both the primary and secondary authors. Edge cases were discussed together until a consensus was reached, and the authors executed a validation check of the other's annotations to minimize possibilities of human error. Inter-rater reliability was calculated by Cohen's κ (all $\kappa > 0.650$). Our annotation was split into a quantitative and qualitative component. Non-human characters, such as animals and monsters, were not included in our annotation because they do not exhibit any clear gender or race. This condition meant that our total generated dataset size was 413 posters and our total real-world dataset size was 382.

3.4.1 Quantitative Proportion Analysis. We used proportions in half of our bias analysis of the generated images. Within both the generated images and their real-world counterparts, we recorded the number of male, female, gender-ambiguous, White, Black, Asian, Hispanic/Latino, and racially ambiguous characters. For each poster, we computed the proportion of characters belonging to each demographic group relative to the total number of characters present. These proportions were then averaged across posters. The proportions from generated and real-world images were compared to determine whether ChatGPT exhibited higher or lower levels of bias relative to real-world patterns.

3.4.2 Qualitative Labeling Analysis. We compared the generated poster with its real-world counterpart across three components: title, artistic style (color scheme, layout, etc.), and main characters. We recorded the degree of similarity between the generated and real-life posters across these aspects. If the posters shared similarities in two or more components, then we labeled these images as replication concerns. We also recorded the demographics of the most significant and central character(s). Some films were identified as having more than one central character. For example, if two characters were the same size and shared the poster's central space, we considered them both central characters.

Table 1. Percentages of posters with a male, female, or gender ambiguous central character.

Genre	Male (%)	Female (%)	Gender Amb. (%)
Action	87.0	11.0	0.0
Animation	40.0	17.0	0.0
Comedy	80.0	52.0	0.0
Drama	66.0	32.0	1.0
Horror	51.0	49.0	1.0

Table 2. Average proportion of female characters in generated and real-world posters, computed as women / (women + men) per poster and averaged across posters. Values are rounded to one significant figure.

Genre	Generated	Real-World
Action	0.3	0.2
Animation	0.4	0.4
Comedy	0.4	0.4
Drama	0.4	0.4
Horror	0.4	0.5

4 Results

Table 1 shows that across all genres, a higher percentage of generated movie posters feature a male as the central character. The proportions for generated images in Table 2 also indicate that across all genres, male characters constitute a larger share of characters per generated poster on average. This indicates a gender bias in favor of male characters. Action movies have the highest percentage of generated movie posters with a male central character (87%) and the lowest average proportion of women appearing in generated movie posters overall (30%).

Table 2 also shows that there is no difference in the average proportion of women appearing in generated and real-world movie posters for the Animation, Comedy, and Drama genres. Generated action movie posters had a higher average proportion of women than real-world action movie posters, whereas generated horror movie posters had a lower average proportion of women than real-world action movie posters. Both generated and real-world posters exhibit a gender bias in favor of male characters, as evidenced by average proportions below 0.5 (except for Horror real-world posters). These results suggest that ChatGPT is mirroring and emphasizing existing gender societal biases.

Table 3 shows that across all genres, a higher percentage of generated movie posters feature a white individual as the central character. Table 4 also shows that the White proportions, across all genres, are higher than the POC proportion and all other demographic proportions. Therefore, it shows that White characters constitute a larger share of characters per generated poster on average. Horror movies have the highest percentage of posters featuring a white central character (91%). Comedy, Drama, and Horror movies have the highest average proportion of White characters in generated movie posters (90%). In generated posters across all genres, Black, Asian, Hispanic/Latino, and racially ambiguous demographics had little to no representation as the central character. The only exception is the Comedy genre, where 52% of central characters were Black. The small demographic-specific average proportions also highlight their under-representation in the generated dataset, with Hispanic/Latino characters not appearing in any Horror movie posters. Combined with the large White proportions, this emphasizes racial bias in favor of White individuals.

Table 3. Percentages of posters with a white, Black, Asian, Hispanic/Latino, or racially ambiguous central character.

Genre	White (%)	Black (%)	Asian (%)	Hispanic/Latino (%)	Race Amb. (%)
Action	83.0	19.0	2.0	1.0	0.0
Animation	44.0	1.0	2.0	2.0	3.0
Comedy	95.0	52.0	0.0	3.0	0.0
Drama	81.0	14.0	0.0	1.0	0.0
Horror	91.0	6.0	0.0	0.0	0.0

Table 4. Average proportion of characters from specified demographic groups in generated (G) and real-world (R) posters, computed as demographic / (all demographics summed together). Values are rounded to one significant figure.

	POC		White		Black		Asian		Hispanic/Latino		Race Amb.	
Genre	G	R	G	R	G	R	G	R	G	R	G	R
Action	0.2	0.2	0.8	0.8	0.1	0.2	0.04	0.04	0.02	0.0	0.03	0.04
Animation	0.2	0.4	0.8	0.6	0.05	0.08	0.05	0.08	0.03	0.03	0.06	0.2
Comedy	0.1	0.2	0.9	0.8	0.06	0.1	0.008	0.01	0.02	0.02	0.005	0.04
Drama	0.1	0.2	0.9	0.8	0.1	0.1	0.006	0.01	0.008	0.004	0.009	0.05
Horror	0.1	0.3	0.9	0.7	0.07	0.1	0.008	0.002	0	0	0.02	0.2

Table 4 shows that across all genres, generated posters either mirror or exaggerate White representation found in real-world counterparts. Black proportions in generated posters were lower across all genres except Drama, where they were equivalent to the real-world proportion. Proportions for Asian and Hispanic/Latino characters in generated posters showed mixed, minimal differences compared to real-world posters, which may be due to existing low representation baselines. Generated posters also showed lower proportions of racially ambiguous characters than real-world posters. These comparisons suggest that ChatGPT is both exacerbating and mirroring existing racial societal biases.

We also observed very close matches between many generated posters and real movie posters, with a significant majority of synopsis prompts yielding images that resembled their real-world counterparts. An example appears in Figure 1, where generated posters for the movies *Armageddon* (1998) and *Eurotrip* (2004) are compared with their real-world counterparts. These two comparisons show the preservation of the real-world movie's title, artistic style, and main characters. 69% of Action, 52% of Animation, 37% of Comedy, 53% of Drama, and 33% of Horror movie posters shared similarities in two or more of our assessed components with its real-world counterpart, which we defined as grounds for potential copyright issues. However, many posters that did not pass our subjective similarity threshold still contained several key elements of the real movies from whose anonymized synopses they were created. For example, even when generated posters differed considerably from their real-world counterparts, the poster may contain illustrations that resemble the actor/actress who played the role in the real-world film. Earlier pilot testing showed no examples of these patterns, therefore providing assurance that our prompt asking ChatGPT to generate examples that are "as similar as possible to real-world posters" did not become a confounding variable in our analysis.

5 Discussion

The results highlight the over-representation of specific demographic groups in ChatGPT-generated images. Given that the generated dataset was dominated by male and white demographics, this suggests that image generation continues to reinforce predominant demographic representations in media and film. The minuscule Asian and Hispanic/Latino

Fig. 1. Comparison of generated image examples and their real-world counterparts.

representation in the dataset also perpetuates the real-world lack of representation of these demographics in media and film. Our comparisons of proportions for generated and real-world images also show that ChatGPT may be mirroring and exaggerating bias trends observed in real-world media, a concerning finding that warrants more attention. The consistent similarities between generated and real-world movie posters also support the mirroring of existing social biases that we found in our demographic bias analyses. Potentially, by using the real-world counterparts as a foundation, LLMs are following the patterns found in its training data distributions, highlighting the importance of developing debiasing mechanisms that might mitigate this.

This study underscores the necessity of approaches that ensure the fairness of AI system outputs, whether by integrating such safeguards into the system itself or by creating less biased training data. How we go about these approaches also requires the creation of robust AI governance frameworks that uphold these procedures and the proliferation of these standards.

5.1 Limitations and Future Work

One limitation is our criteria potentially not being exhaustive. Establishing a clear threshold for whether a generated image resembles its real-world counterpart is a matter of subjective judgment, and it is part of the broader discussion of how we develop technical evaluation metrics for theoretical situations.

A natural extension of this work will be to use synopses from other genres, thereby ensuring exhaustive analysis. It may also be beneficial to analyze the posters for additional ethical concerns, such as biases related to sexuality, socioeconomic status, and disability. Integrating these nuanced bias analyses into our methodologies will lay the groundwork for future debiasing methods by providing a more comprehensive understanding of which biases require the most reduction in AI systems.

References

[1] Juveria Afreen, Mahsa Mohaghegh, and Maryam Doborjeh. 2025. Systematic literature review on bias mitigation in generative AI. *AI and Ethics* 5, 5 (2025), 4789–4841.

[2] Yuntao Bai, Andy Jones, Kamal Ndousse, Amanda Askell, Anna Chen, Nova DasSarma, Dawn Drain, Stanislav Fort, Deep Ganguli, Tom Henighan, Nicholas Joseph, Saurav Kadavath, Jackson Kernion, Tom Conerly, Sheer El-Showk, Nelson Elhage, Zac Hatfield-Dodds, Danny Hernandez, Tristan Hume, Scott Johnston, Shauna Kravec, Liane Lovitt, Neel Nanda, Catherine Olsson, Dario Amodei, Tom Brown, Jack Clark, Sam McCandlish, Chris Olah, Ben Mann, and Jared Kaplan. 2022. Training a Helpful and Harmless Assistant with Reinforcement Learning from Human Feedback. arXiv:2204.05862 [cs.CL] https://arxiv.org/abs/2204.05862

[3] Jeffery Dastin. 2018. *Insight - Amazon scraps secret AI recruiting tool that showed bias against women.* Retrieved December 16, 2025 from https://www.reuters.com/article/world/insight-amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK0AG/

[4] Samuel Dooley, Ryan Downing, George Wei, Nathan Shankar, Bradon Thymes, Gudrun Thorkelsdottir, Tiye Kurtz-Miott, Rachel Mattson, Olufemi Obiwumi, Valeriia Cherepanova, Micah Goldblum, John P Dickerson, and Tom Goldstein. 2021. *Comparing Human and Machine Bias in Face Recognition.* arXiv:2110.08396 [cs.CV] https://arxiv.org/abs/2110.08396

[5] Lauren Kirchner Jeff Larson, Surya Mattu and Julia Angwin. 2016. *How We Analyzed the COMPAS Recidivism Algorithm.* Retrieved December 16, 2025 from https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm

[6] Dima Kagan, Mor Levy, Michael Fire, and Galit Fuhrmann Alpert. 2024. Ethnic representation analysis of commercial movie posters. *Humanities and Social Sciences Communications* 11, 180 (Feb. 2024). doi:10.1057/s41599-023-02040-y

[7] Xiaojian Lin and Michael Losavio. 2024. A Comprehensive Survey on Bias and Fairness in Generative AI: Legal, Ethical, and Technical Responses. In *Smart Innovation, Systems and Technologies*, Vol. 444. Springer, Singapore. doi:10.1007/978-981-96-7273-8_22

[8] Leonard AW Memon. 2025. Gender bias in movie posters through the lens of Spatial Agency Bias. *Visual Communication* 24, 2 (2025), 469–484. arXiv:https://doi.org/10.1177/14703572231206461 doi:10.1177/14703572231206461

[9] Dan Milmo. 2024. *Google pauses AI-generated images of people after ethnicity criticism.* Retrieved January 12, 2026 from https://www.theguardian.com/technology/2024/feb/22/google-pauses-ai-generated-images-of-people-after-ethnicity-criticism

[10] Jie Qin, Jie Wu, Weifeng Chen, Yuxi Ren, Huixia Li, Hefeng Wu, Xuefeng Xiao, Rui Wang, and Shilei Wen. 2024. *DiffusionGPT: LLM-Driven Text-to-Image Generation System.* arXiv:2401.10061 [cs.CV] https://arxiv.org/abs/2401.10061

[11] Julie M. Smith. 2024. "I'm Sorry, but I Can't Assist": Bias in Generative AI. In *Proceedings of the 2024 on RESPECT Annual Conference* (Atlanta, GA, USA) *(RESPECT 2024)*. Association for Computing Machinery, New York, NY, USA, 75–80. doi:10.1145/3653666.3656065

[12] Stacy L Smith, Marc Choueiti, Katherine Pieper, Traci Gillig, Carmen Lee, and Dylan DeLuca. 2016. Media, diversity, & social change initiative. *Institute for Diversity and Empowerment at Annenberg* 22 (2016), 30.

[13] Yusen Song, Andreea Pocol, and Lesley Istead. 2024. An Evaluation of Gender Bias in 167K Movie Posters. In *Intelligent Systems and Applications*, Kohei Arai (Ed.). Springer Nature Switzerland, Cham, 332–358.

[14] Lucia Vicente, Helena Matute, Caterina Fregosi, and Federico Cabitza. 2025. Machine learning systems as mentors in human learning: A user study on machine bias transmission in medical training. *International Journal of Human-Computer Studies* 198 (2025), 103474. doi:10.1016/j.ijhcs.2025.103474

[15] Yixin Wan, Arjun Subramonian, Anaelia Ovalle, Zongyu Lin, Ashima Suvarna, Christina Chance, Hritik Bansal, Rebecca Pattichis, and Kai-Wei Chang. 2024. Survey of Bias In Text-to-Image Generation: Definition, Evaluation, and Mitigation. arXiv:2404.01030 [cs.CV] https://arxiv.org/abs/2404.01030

[16] Wenxuan Wang, Haonan Bai, Jen-tse Huang, Yuxuan Wan, Youliang Yuan, Haoyi Qiu, Nanyun Peng, and Michael Lyu. 2024. New Job, New Gender? Measuring the Social Bias in Image Generation Models. In *Proceedings of the 32nd ACM International Conference on Multimedia* (Melbourne VIC, Australia) *(MM '24)*. Association for Computing Machinery, New York, NY, USA, 3781–3789. doi:10.1145/3664647.3681433

[17] Han Xu, Xiaorui Liu, Yaxin Li, Anil Jain, and Jiliang Tang. 2021. To be Robust or to be Fair: Towards Fairness in Adversarial Training. In *Proceedings of the 38th International Conference on Machine Learning (Proceedings of Machine Learning Research, Vol. 139)*, Marina Meila and Tong Zhang (Eds.). PMLR, 11492–11501. https://proceedings.mlr.press/v139/xu21b.html

[18] Travis Zack, Eric Lehman, Mirac Suzgun, Jorge A Rodriguez, Leo Anthony Celi, Judy Gichoya, Dan Jurafsky, Peter Szolovits, David W Bates, Raja-Elie E Abdulnour, et al. 2024. Assessing the potential of GPT-4 to perpetuate racial and gender biases in health care: a model evaluation study. *The Lancet Digital Health* 6, 1 (2024), e12–e22.

[19] Weijie Zhao, Huajie Shao, Zhaozhuo Xu, Suzhen Duan, and Denghui Zhang. 2024. *Measuring Copyright Risks of Large Language Model via Partial Information Probing.* arXiv:2409.13831 [cs.CL] https://arxiv.org/abs/2409.13831