Shattering Bias: A Path to Bridging the Gender Divide with Creative Machines

Juliana Shihadeh and Margareta Ackerman
Department of Computer Science and Engineering
Santa Clara University, Santa Clara, California
{jshihadeh,mackerman}@scu.edu

Abstract

The widespread emergence of creative machines poses a significant challenge, as they tend to reinforce biases, including gender bias. This paper presents a novel perspective on how creative machines can be utilized to counteract gender disparities and mitigate bias. We propose research directions that explore the potential of creative systems to empower women and promote gender equity. Our aim is to leverage computational creativity to actively contribute to fostering a more inclusive and equitable society.

Introduction

The emergence of generative AI in the world at large came with a high price: The most widespread generative technologies are riddled with bias. Multiple biases have been identified in large generative models including biases related to culture (Saharia et al. 2022) and religion (Abid, Farooqi, and Zou 2021). Gender bias, which is our focus here, has been found in such models with respect to occupations (Cheong et al. 2023), intellect (Shihadeh et al. 2022), and leadership (Lucy and Bamman 2021). Millions of people are using creative machines of the likes of Dall-E (Q.ai 2023), Midjourney (Salkowitz 2022) and ChatGPT (Milmo 2023), which perpetuate and amplify these biases. In turn, the dissemination of content created with these systems further amplifies the biases inherent in these models.

Is is well known that nearly half of the world population consists of women.[1] Gender biases have a profound impact on the potential of nearly half of the global population and limit the contributions women can make compared to if they were granted equal voice, say, and opportunity. Furthermore, eradicating gender bias in large models is proving challenging as seen in the sparsity of research on solutions to this issue, challenges in accessing datasets and limited computation power to retrain models (Berg et al. 2022). Doubtlessly, research into how to reduce bias in large language models (Liu et al. 2022) and text-to-image generators (Orgad, Kawar, and Belinkov 2023) despite these challenges is critical to mitigating the societal damage that these models are inflicting on our world. In this paper, however, we offer a complementary approach.

[1] https://data.worldbank.org/indicator/SP.POP.TOTL.FE.ZS

What if, instead of perpetuating and amplifying biases, creative machines could instead be utilized to bring about a more just world? In this paper, we propose a computational creativity research agenda to promote gender equity. Grounded in interdisciplinary research spanning psychology, arts, computer science and gender studies, we offer a roadmap for a research program that has the potential to both empower women and bring profound awareness to gender bias.

The CC community carries unparalleled expertise in creative machines. In a world where creative machines are rapidly integrated into the very fabric of society, the risk of these machines to amplify biases is substantial, if not inevitable. Our community possesses the necessary expertise to envision and pioneer novel approaches for integrating creative machines into the world at large. We hope that the ideas in this paper encourage the CC community to delve deeper into the use of creative machines for fostering gender equality.

We begin this paper by briefly discussing the presence of gender bias in creative machines and the impact of gender bias on society at large. Next, we propose future directions of research in CC to mitigate gender bias along with guidelines for maneuvering the development of creative machines that close, rather than widen, the gender gap.

Gender Bias in Creative Machines

Due to their large (and as such difficult to prune) data sets, large language models and text-to-image generators are highly problematic when it comes to bias. For instance, one study found that when Stable Diffusion was prompted with "a photo of the face of" an emotional or exotic person, more women were seen (Bianchi et al. 2022). Another study found when writing stories with GPT-3, more women were associated to family, appearance and less leadership (Lucy and Bamman 2021). Studies on brilliance bias found that GPT-3 affiliated higher intellectual abilities to men (Shihadeh et al. 2022) and associated brilliance to images of men more often than women (Shihadeh and Ackerman 2023). Multiple studies look at gender stereotypes in professions, whereby the models assume that, for example, nurses must be female and doctors must be male (Cheong et al. 2023; Bianchi et al. 2022; Kirk 2021; Caliskan, Bryson, and Narayanan 2017). In the Computational Creativity commu-

nity, there has not been much focus on discriminatory bias, with the exception of Loughran's discussion of both algorithmic and discriminatory bias in AI (Loughran 2022).

The Effects of Biases

Biases cause *stereotype threat*, feelings of anxiety and stress that result from not feeling that one belongs (Calaza 2021). Stereotype threat further causes social isolation, rejection, and reduced memory which negatively impact one's well-being, immune system, and work quality. In a related study (Calaza 2021), female participants were given one of two math tests, one stating that their performance was to study gender differences and another focusing on cognitive processes. Women's resulting scores were lower in the former setting. Additionally, biases induce stereotypical occupational fit (Kirk 2021). When repeated through multiple sources, biases further influence the world and increase inequalities (Calaza 2021). Based on cultivation theory, repeated content exposure amplifies this effect (Potter 1993).

Research Directions

In this section, we propose avenues for computational creativity research to contribute towards creating a more gender-inclusive world. Our objective is for 1) young women to explore their potential rather than succumb to society's biases on this matter, and 2) help others gain deeper insight into women's experiences around gender bias.

Helping women realize their potential

Using creative machines, we can demonstrate that the "glass ceiling" is breakable and give girls the space to explore their true potential despite societal biases that often limit their ambitions. As Melanie Perkins, CEO of Canva - a suite of design tools - emphasizes, it is important to dream, to dream ten years out even and dream of what kind of world you want to live in (Ventures 2022).

It is important to help girls and women "close their imagination gap" [2] and show them that there are no limits to what they can pursue. For example, we can start by helping girls explore career paths. It has been identified that images "put ideas into your head" (Hibbing and Rankin-Erickson 2003). Minorities are more negatively affected by media stereotypes (Appel and Weber 2021) and girls have been found to replicate stereotypical behaviour (Essig 2018). If visuals influence people's view of the world, we could instead offer up creative visuals that allow girls to consider a more inclusive world. To this end, a co-creative system can visualize a young girl as a grown women across a wide range of professions, offering stories to flesh out the kind of impact that she may have.

Role model intervention works by having a positive representation of a minority to demonstrate that the stereotype doesn't always hold (Eschenbach 2014). Role models can have a positive impact even when their influence is conveyed through text rather than in-person interaction (Eschenbach 2014). Machines that create role models could make it more accessible for girls to see, hear, or even engage with role-models. As a result, this can reach a larger audience of girls and have a bigger impact on girls' trajectory. Evidence shows that female role models in counter stereotypical roles encourage younger female students to pursue such careers including politics, science, and engineering (Olsson 2018). A CC system can be designed to help young girls create their own personalized role models, combining images and text. The role model might take on some of the values, interests, or personality traits of the girl. The added level of engagement, perhaps through a chatbot representing the generated role model, may contribute to the value of such an experience.

Additionally, other individuals, such as writers, could use a role model creator to influence and inspire their stories. Similarly, systems that help create images that promote a gender equitable world can be incorporated in children's books, as their images have been found to influence female stereotypes (Hamilton et al. 2006).

(a) (b) (c) (d)

Figure 1: These images, made with Midjourney, demonstrate the power of visuals for engaging our imagination in "what if" scenarios that can support gender equity and build empathy towards the challenges faced by women. (a) visualizes what breaking the "Glass Ceiling" could look like, in a world where even God is viewed as a woman, whereas (b), (c), (d) showcase female entrepreneurs, computer scientists and politicians, respectively. Creative machines can help us step into a world where such representation were the norm, which can open up the imagination of women towards their own futures and foster empathy through demonstrating what it is like to be a minority.

[2] https://www.careergirls.org/about/

Helping women cope with gender discrimination

Offering various therapeutic means to process one's experience with gender discrimination that they have experienced can help women reduce its impact. In collaboration between computational creative practitioners and therapists, computational creativity demonstrates potential to be a means for having positive psychological influence (Pease et al. 2022). In particular, Pease et al. (Pease et al. 2022) discuss how the arts have helped soldiers overcome war injuries, enable people to reconnect with their true self, restore their sense of expertise and self-esteem, and help people create a "tangible expression of who [they] are and what [they] do." Women can work with co-creative machines to write songs, poetry and make art about their challenges with sexism. This can even be more beneficial in a co-creative context that involves creative machines and multiple people (Pease et al. 2022).

Earlier generations saw many women with limited choices, often forced to remain at home as caretakers instead of pursuing their intellectual and personal aspirations beyond the boundaries of the home. Creative machines could enable women to explore alternative life paths, considering what their lives could have been like had they not been limited by gender stereotypes. In doing so we can capture the lost voices of what happens when women are not empowered to reach their true potential. This can provide a self-reflective opportunity to come to acceptance with what was lost and help with expressing and validating their identifies (Pease et al. 2022).

Lastly, we can imagine machines designed to collect and creatively, anonymously share the experiences of women. Women could tell their stories related to struggles with gender bias, and the machine could create representative, anonymous stories that capture common experiences shared by women. The resulting stories may be shared not only through text, but through any relevant modality, such as film. Both sharing of experiences and learning of others' struggles has the potential to offer therapeutic value.

Fostering Empathy

Empathy is the ability to put oneself in another's place, to see and feel the world as they do (Rusu and others 2017). The arts can help facilitate experiences of empathy (Rusu and others 2017; Bollmer 2017; Pozo 2018). By offering immersive, multi-modal, and interactive experiences that envision a world where women outnumber men in domains where biases persist, such as leadership positions and STEM, we can enable people of all genders to empathize with the biases women encounter. To create a more personalized experience, the systems could consider the user's point of view, such as their profession and hobbies, in the creation of the experience.

For example, a user may be presented with various generative, personalized scenes and prompted by the machine with follow up questions to reflect on what they saw. The generated scenarios may comprise, for instance, the opportunity to participate in a meeting where the majority of engineers are women, a congressional session where most of the politicians are female, or a meeting with a leading venture investment firm that primarily supports female founders.

Additionally, we could develop creative machines that help demonstrate what it would be like to speak in female rather than male. For example, what if language defaults to female pronouns, causing men to experience gender exclusion in pronouns? Through linguistic-centered creative experiences, we can help portray injustices that women regularly experience in the world. For instance, consider the statement "girls are as good as boys at math", which subtly expresses the idea that boys are typically better in math. What if this statement became "boys are as good as girls at math"? Neil Armstrong's quote "That's one small step for man, one giant leap for mankind." On the contrary it could be, "That's one small step for a woman, one giant leap for womankind." For example, a CC system could be a collaborative partner to create a gender-flipped dialogue that can then be brought to life as a movie scene created by the machine. Similarly, we can imagine a co-creative system for the creation of songs or stories where stereotypes are inverted. Stepping into this type of alternative reality can foster profound empathy for the realities that women face on a daily basis, and can help people of all genders become allies in the struggle towards closing the gender gap.

Guidelines

We propose several guidelines to consider when tackling gender bias through a computational creativity lens.

1. **Evoke personal connection and reflection.** In designing creative machines for closing the gender gap, it is worth considering how the machine might elicit personal connection and reflection. For instance, working with co-creative machines can facilitate introspection on personal experiences, interests, goals, or aspirations. Similarly, an immersive experience may be designed to encourage us to notice instances of gender bias in the future that we may have otherwise missed.

2. **Facilitate creative engagement**. Consider designing experiences where users are called to tap into their own creative capabilities as they explore what is possible (for themselves or others) outside of stereotypical gender roles. By enabling the user to be creative, they become more personally invested and engaged in the experience, as active engagement has been proven to enhance brain development and learning (Immordino-Yang, Darling-Hammond, and Krone 2019).

3. **Constructive and non-judgemental**. We recommend that creative systems that aim to counter gender bias be designed in an inherently constructive and non-judgemental manner, as to put the user at ease and reduce defensiveness. Living in a world where gender bias is ubiquitous, all of us carry implicit gender biases. Recognizing bias within oneself can be challenging and initially upsetting, be it biases against oneself (ex. stereotypical beliefs that limit one's own potential) or others. As such, systems that aim to correct gender biases should be designed in a manner that is respectful to the sensitive nature of this issue.

Conclusions

In this paper, we outline research directions that invite the Computational Creativity community to take the lead in steering creative machines away from widening the gender gap, and instead utilizing them towards reducing gender biases. We present the opportunity to create CC systems that lead users into personalized, creative, and immersive experiences that showcase a vision of the world we aspire to, and invite people of all genders to experience the current realities of living as a woman. In order to promote gender equity, creative machines can assist in widening women's insight into their potential while also aiding others in relating to gender disparities that women have endured in the past and continue to face presently. By embracing the potential of computational creativity to address gender biases, we can bring forth creative machines as a powerful means to break down barriers, promote gender equity, and advance towards a more diverse and equitable future.

While this work focuses on addressing bias against women, we hope that the ideas here may spark broader interest on how creative machines may be utilized to reduce discriminatory biases in a broader sense. In particular, the research directions we propose may be adapted across the gender spectrum. Further, while our focus here is on bias stemming from gender, our proposed research directions may inspire ideas for how to utilize creative machines to reduce bias based on race, age, sexual orientation, etc. Our aspiration is that this work stimulates research on the utilization of CC in reducing discriminatory prejudice, not only aiming to minimize bias in current creative machines but also employing new and imaginative methods to build creative machines that are purposefully designed to contribute to a more fair and just society.

References

Abid, A.; Farooqi, M.; and Zou, J. 2021. Persistent anti-muslim bias in large language models. In *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society*, 298–306.

Appel, M., and Weber, S. 2021. Do mass mediated stereotypes harm members of negatively stereotyped groups? a meta-analytical review on media-generated stereotype threat and stereotype lift. *Communication Research* 48(2):151–179.

Berg, H.; Hall, S. M.; Bhalgat, Y.; Yang, W.; Kirk, H. R.; Shtedritski, A.; and Bain, M. 2022. A prompt array keeps the bias away: Debiasing vision-language models with adversarial learning. *arXiv preprint arXiv:2203.11933*.

Bianchi, F.; Kalluri, P.; Durmus, E.; Ladhak, F.; Cheng, M.; Nozza, D.; Hashimoto, T.; Jurafsky, D.; Zou, J.; and Caliskan, A. 2022. Easily accessible text-to-image generation amplifies demographic stereotypes at large scale. *arXiv preprint arXiv:2211.03759*.

Bollmer, G. 2017. Empathy machines. *Media International Australia* 165(1):63–76.

Calaza, Karin C & Erthal, F. . P. M. G. . M. K. C. . D. V. T. . D. I. . C. H. C. . V. M. D. . M. L. B. . S. J. B. . o. 2021. Facing racism & sexism in science by fighting against social implicit bias: A latina & black woman's perspective. *Frontiers in Psychology*.

Caliskan, A.; Bryson, J. J.; and Narayanan, A. 2017. Semantics derived automatically from language corpora contain human-like biases. *Science* 356(6334):183–186.

Cheong, M.; Abedin, E.; Ferreira, M.; Reimann, R. W.; Chalson, S.; Robinson, P.; Byrne, J.; Ruppanner, L.; Alfano, M.; and Klein, C. 2023. Investigating gender and racial biases in dall-e mini images.

Eschenbach, Elizabeth A & Virnoche, M. . C. E. M. . L. S. M. . C. M. M. 2014. Proven practices that can reduce stereotype threat in engineering education: A literature review. In *2014 IEEE Frontiers in Education Conference (FIE) Proceedings*. IEEE.

Essig, L. W. 2018. *A Content-Analytic Meta-Analysis of Gender Stereotyping in Screen Media*. Brigham Young University.

Hamilton, M. C.; Anderson, D.; Broaddus, M.; and Young, K. 2006. Gender stereotyping and under-representation of female characters in 200 popular children's picture books: A twenty-first century update. *Sex roles* 55:757–765.

Hibbing, A. N., and Rankin-Erickson, J. L. 2003. A picture is worth a thousand words: Using visual images to improve comprehension for middle school struggling readers. *The reading teacher* 56(8).

Immordino-Yang, M. H.; Darling-Hammond, L.; and Krone, C. R. 2019. Nurturing nature: How brain development is inherently social and emotional, and what this means for education. *Educational Psychologist* 54:185 – 204.

Kirk, Hannah Rose & Volpin, F. . I. H. . B. E. . D. F. . S. A. . A. Y. . o. 2021. Bias out-of-the-box: An empirical analysis of intersectional occupational biases in popular generative language models. *Advances in Neural Information Processing Systems*.

Liu, R.; Jia, C.; Wei, J.; Xu, G.; and Vosoughi, S. 2022. Quantifying and alleviating political bias in language models. *Artificial Intelligence* 304:103654.

Loughran, R. 2022. Bias and creativity.

Lucy, L., and Bamman, D. 2021. Gender and representation bias in gpt-3 generated stories. In *Proceedings of the Third Workshop on Narrative Understanding*, 48–55.

Milmo, D. 2023. Chatgpt reaches 100 million users two months after launch.

Olsson, Maria & Martiny, S. E. 2018. Does exposure to counterstereotypical role models influence girls' & women's gender stereotypes & career choices? a review of social psychological research. *Frontiers in psychology*.

Orgad, H.; Kawar, B.; and Belinkov, Y. 2023. Editing implicit assumptions in text-to-image diffusion models. *arXiv preprint arXiv:2303.08084*.

Pease, A.; Ackerman, M.; Pease, N.; and McFadden, B. 2022. A roadmap for therapeutic computational creativity. In *13th International Conference on Computational Creativity*.

Potter, W. J. 1993. Cultivation theory and research. *Human Communication Research* 19(4):564–601.

Pozo, T. 2018. Queer games after empathy: Feminism and haptic game design aesthetics from consent to cuteness to the radically soft. *Game Studies* 18(3).

Q.ai. 2023. Dall·e mini and the future of artificial intelligence art.

Rusu, M., et al. 2017. Empathy and communication through art. *Review of artistic education* (13+ 14):139–146.

Saharia, C.; Chan, W.; Saxena, S.; Li, L.; Whang, J.; Denton, E. L.; Ghasemipour, K.; Gontijo Lopes, R.; Karagol Ayan, B.; Salimans, T.; et al. 2022. Photorealistic text-to-image diffusion models with deep language understanding. *Advances in Neural Information Processing Systems* 35:36479–36494.

Salkowitz, R. 2022. Midjourney founder david holz on the impact of ai on art, imagination and the creative economy.

Shihadeh, J., and Ackerman, M. 2023. What does genius look like? an analysis brilliance bias in text-to-imagine models.

Shihadeh, J.; Ackerman, M.; Troske, A.; Lawson, N.; and Gonzalez, E. 2022. Brilliance bias in gpt-3. In *2022 IEEE Global Humanitarian Technology Conference (GHTC)*, 62–69. IEEE.

Ventures, B. 2022. Sunrise australia day 1 - melanie perkins keynote. https://youtu.be/FXtGE811GCw.