What Does Genius Look Like?
An Analysis of Brilliance Bias in Text-to-Image Models

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

Abstract

As text-to-image models and the visuals that they create become increasingly integrated into society, it is imperative to develop an awareness of the inherent biases within these technologies. While earlier visual creative machines such as AARON by Harold Cohen (Cohen 1999) and The Painting Fool by Simon Colton (Colton et al. 2015) have exhibited remarkable creativity, the methodology underlying today's popular text-to-image models rely heavily on public data to produce visuals, resulting in an increased risk for bias. Further, recent image generation technologies, such as Dall-E (Q.ai 2022) and Midjourney (Salkowitz 2022) and applications such as LensaAI have attracted millions of users (Curry 2023), making it more urgent to ascertain the risks of these technologies. In this paper, we initiate an analysis of text-to-image models focusing on Brilliance Bias, a negative stereotype of women's intellectual abilities and holds back women's potential. Our findings reveal a significant presence of Brilliance Bias in Dall-E, Midjourney, and Stable Diffusion.

Figure 1: Sample output of Midjourney prompted on "Genius person." The parameters are set to generate four images per output.

"You can't be what you can't see"
-Lean In Organization and Getty Images[1]

Introduction

Creative machines have long been made and studied within academia. When it comes to machines creating art, one of the earliest examples includes Harold Cohen's AARON, which Cohen taught to draw and later paint in his own style (Cohen 1999; Sundararajan 2021). AARON has been showcased in galleries as early us 1995 (Garcia 2016). Another notable system is The Painting Fool by Simon Colton, which unlike AARON, aimed to be taken seriously as an artist in its own right (Colton et al. 2015). The Painting Fool and its work have been showcased at public venues, such as the 2013 Paris exhibit "You can't know my mind" (Shubber 2013). Industry involvement in the arena of creative machines was gradual, with systems such as Google's DeepDream entering the scene in 2015 (Rayner 2016).

Microsoft's investment in OpenAI began to change the landscape, focusing on the creation of large (and expensive) models at a magnitude that was not previously possible with respect to amounts of data used for training and the size of the models. The introduction of the text-to-image model DALL-E, leading to Stable Diffusion and proliferation of commercial apps, such as LensaAI, brought generative AI visuals to the masses. AI-generated art is now incorporated into advertising (Nestle (Kiefer 2022), future Super Bowl ads (CBInsights 2023), and Rosebud.ai (Koidan 2020)). At the same time, firms such as Microsoft (Microsoft 2023; Q.ai 2022), Canva (Adams 2022), and Shutterstock (Shutterstock 2023) have integrated image generative capabilities into their products.

Text-to-image generation through large models does not come without pitfalls. One of the main concerns with these models is their reflection, and perhaps even amplification, of biases present in the data they are trained on. Researchers study racial bias (Agarwal et al. 2021; Wiggers 2021; Srinivasan and Uchino 2021a) in these models, and gender bias is analyzed with respect to clothing and physique in text-to-image models associations to women and men (Chiriguayo and Ta 2022; Steele 2022).

A little known, but significantly impactful, bias called

[1]leanin.org/getty, Accessed: 2/27/2023

Figure 2: "Genius person" prompted to Dall-E. The top and bottom rows show examples of women and men respectively.

Figure 3: "Brilliant person" prompted to Dall-E. The top and bottom rows show examples of women and men respectively.

"Brilliance Bias," hinders some of the most high-potential members of our society. Brilliance Bias is the association of higher intellectual capabilities to males (Leslie et al. 2015), that is, the implicit belief that intellectual brilliance is more likely to be present in men than in women. It impedes women's potential through both their self-perception and the opportunities that others are willing to grant them.

We initiate the study of Brilliance Bias in large text-to-image models. Our analysis focuses on evaluating its presence in some of the most popular models, namely Dall-E, Stable Diffusion, Midjourney and Craiyon (formerly Dall-E mini). Visuals influence people's perception of the world. For example, a review on stock photos showing stereotypical depictions of women such as in supporting roles proves to negatively impact women's career potentials (Miller 2014). A developmental psychology study on media from 2000-2020 reveals that media significantly influences young people's views on gender roles (Ward and Grower 2020).

Given the rapidly growing popularity of text-to-image models and the powerful societal influences of images, it is critical to understand the biases exhibited by these models. As Sun et al. (Sun et al. 2022) points out, "social transparency – making visible the socio-organizational factors that govern the use of AI – can help users form a socially situated understanding of an AI system and take more effective actions with it." A clear insight into the presence of

Figure 4: "Brainiac person" prompted to Dall-E. The top and bottom rows show examples of women and men respectively.

the biases found in text-to-image models will help find effective solutions to mitigate those biases, and limit their impact. Our study initiates an analysis of the Brilliance Bias in these models, acting as an essential first step to mitigate the amplified impact of this bias.

Background

Brilliance Bias

Despite the numerous intellectual contributions made by women, their intellectual abilities are consistently downplayed through a pervasive bias known as "Brilliance Bias." Brilliance Bias is the implicit belief that intellectual brilliance is a male trait. This bias is found to be pervasive in the STEM and Humanities fields, and correlates to lower female to male ratios of PhD students studying Computer Science, Mathematics, Philosophy and Music Composition (Leslie et al. 2015). Studies on children show that it starts as early as 5-7 years old, as seen by children selecting boys in a game for "really, really smart" teammates (Bian, Leslie, and Cimpian 2018). When asked to pick out images associated with stories and descriptions of being "really, really smart", girls are less likely to pick from their gender and more likely to associate to being "really, really nice" starting at the age of 6 (Bian, Leslie, and Cimpian 2017). Furthermore, at the age of 6, girls' interests shift because they think of themselves as less brilliant; they are more likely to pick a game for children who try "really, really hard" and less likely to pick games for "really, really smart" children. At the age of 5, "really, really smart" children's games are more equally selected by boys and girls.

Research focusing on STEM fields shows that both women and men affiliate brilliance to STEM (Deiglmayr, Stern, and Schubert 2019) (the belief that people who are in STEM are brilliant). Furthermore, the study shows men are less likely than women to believe in the existence of Brilliance Bias and more likely to feel like they belong in STEM fields. It is further shown that women are less likely to be referred to jobs that require high levels of intellectual ability (Bian, Leslie, and Cimpian 2018).

Brilliance Bias has only recently began to be studied in the context of generative models. Last year, an adjective

Figure 5: "Brainiac person" prompted to Midjourney. Of 100 images, only two are identified as female. They are shown in the top row, left to right. The rest of the images are identified as male.

Figure 6: "Genius person" prompted to Craiyon. No images out of the 100 images we generate display a woman.

Figure 7: "Brilliant person" prompted to Craiyon. The top and bottom rows show examples of women and men respectively.

and lexicon study on Brilliance Bias in large text models, specifically OpenAI's models, reveals a significant presence of Brilliance Bias. When the OpenAI models are prompted with identical brilliance prompts other than gender, men are associated with higher levels of power, agency, valence, arousal, and dominance (Shihadeh et al. 2022).

Biases in Images

A Google Search Engine study analyzes the occupational gender biases in image search queries (Kay, Matuszek, and Munson 2015). Its results show a significant representation of stereotypical gender roles and minorities, such as women, portrayed unprofessionally in images. Furthermore, it points out people are more likely to use image results that align with their stereotypical beliefs causing a dangerous loop of increasing biases.

One paper looks at the bias of CEO genders in the Google Search Engine and finds that results are dominated with white men (Lam et al. 2018). Another study finds that even though efforts were put to mitigate the gender bias of the query "CEO", combinations of "CEO" with a country such as "United States" resurface the gender bias (Feng and Shah 2022). Thus revealing the challenges in fully mitigating a bias that is deeply embedded in a system.

Studies on facial recognition show a bias in being able to identify white men more accurately, in particular significantly misclassifying black women as male (Raji et al. 2020). An analysis of image recognition models shows that images of women are annotated more on appearance and less likely to be identified in image detection technology compared to men (Schwemmer et al. 2020). If image recognition tools are used to annotate and label images for training text-to-image models, computer labeling biases could further increase gender biases society gets exposed to.

Biases in Generative AI

While biases are studied in text-to-image models, no prior research of this kind focuses on Brilliance Bias. For example, gender bias in occupations is found in the text-to-image model CLIP (Wiggers 2021; Agarwal et al. 2021). A high correlation of stereotypical occupations is found associated to women, such as "nanny" and "housekeeper", and men, such as "prisoner" and "mobster". Furthermore, racial biases are found such as black people misclassified to be non-human, being labeled as "animal", "gorilla", and "chimpanzee". Additional racial biases are found on lightening the skin tone of a person (Srinivasan and Uchino 2021a; Mattei 2022). One study finds race and gender biases in Stable-Diffusion with descriptive phrases like "emotional" showing women and "poor" showing more dark skinned people (Bianchi et al. 2022). A study on cycleGAN examines how an art style miscaptured in generative models can cause "inaccurate information about socio-political-cultural aspects" (Srinivasan and Uchino 2021b).

Generative AI app users note seeing their race being erased (Mello-Klein 2022; Sung 2022). Others point out Asian women in particular being depicted in tears and showing more nudity (Heikkilä 2022). Some users see stereotype portrayals of women, such as slimming waists (Chiriguayo and Ta 2022) furthermore exposing women's skin and anatomy more, while men are more likely shown in professional apparel (Steele 2022; Heikkilä 2022). OpenAI attempted to add more diversity to DALL-E, particularly as it applies to occupation (OpenAI 2022), and maybe appending words like "black" and "female"[2]

[2]https://twitter.com/minimaxir/status/1549070583035416576

Figure 8: "Brainiac person" prompted to Craiyon. The top and bottom rows show examples of women and men respectively.

How Visuals Affect Society

Images are an integral part of our world. Research that looks at how images affect students' learning in middle school concludes that images influence their understanding of the world, finding that "if you look at an image, it puts more ideas in your head" (Hibbing and Rankin-Erickson 2003). Furthermore, based on the cultivation theory, repeated exposure over time alters one's perception of the world (Potter 1993; Shrum 1995). One study finds that short term exposure also affects one's views. It finds that skewing Google search results changes people's choice in selecting a woman or man to represent a job (Kay, Matuszek, and Munson 2015). Another study finds that stock photos put women in supporting roles, stereotyped roles, and sexualized their images further finding that seeing these images hurts women's career aspirations (Kay, Matuszek, and Munson 2015; Suddath 2014). This work, led by Sheryl Sandberg and Getty Images, resulted in the initiative of "You can't be what you can't see" (LeanIn.Org 2023). To mitigate visual biases, they curate a set of creative images with archetypes rather than stereotypes; these images portray diverse examples of families, women in powerful roles and men as caretakers in addition to earners.

Figure 9: "Brilliant person" prompted to Stable Diffusion. The top and bottom rows show examples of women and men respectively.

Multiple studies on media reveal it has a powerful influence on society. Stereotypes influence a person's inclination to join a field, changing how media, such as television for instance, portrays computer scientists can in turn help with demonstrating the diversity of a field (Cheryan et al. 2013). Due to the "digital generation", teens are especially prone to being influenced by media about how they see themselves and socialize (Celestin 2011). For instance, Silicon Valley and the Big Bang Theory show women as a background character, usually for the role of a love plot in a story rather than a leader (Javed 2015). Furthermore, these TV shows have a stereotypical nerd association to the male characters which can be discouraging for girls' perception of a field (Javed 2015; Welsh 2013). Supporting studies show that girls who see stereotypical portrayals or behaviors of people are more likely to demonstrate the stereotypical behavior themselves (Essig 2018). Geena Davis who is an advocate of more women in film leadership roles, shows the film industry influences women's ambitions, changes toxic relationship dynamics, and encourages success (Ford 2019; Institute 2016). The mass effect of media on people's perceptions of the world and themselves demonstrates how influencing visuals are.

Methodology

We analyze the output of four text-to-image models to determine whether these models exhibit Brilliance Bias. The models we evaluate are Dall-E, Midjourney, Craiyon and Stable Diffusion. To study the presence of this bias, we provide each model with a set of brilliance prompts, designed to elicit the creation of an image of a person the model deems "genius" or "brilliant." Furthermore, we test the models on the base case prompt "person" to compare against brilliance prompts. To analyze the results, we evaluate the differences of the number of women and men in the generated output.

Data

Figure 10: "Super Smart person" prompted to Stable Diffusion. The top and bottom rows show examples of women and men respectively.

We generate 400 images using each text-to-image model for four different brilliance traits[3]. Instead of feeding the models a single prompt, like "Brilliant person," we expand

[3]Our data can be found at https://github.com/julishi/Brilliance-Bias-in-Text-to-Image-Models/tree/main

our analysis to a set of carefully designed prompts. The reason is that words inherently have multiple meanings, and if we seek to understand how models visualize intellectual brilliance, it is best to tackle this challenge through several prompts that all aim to uncover this aspect of the models. The brilliance traits (other ways to say "brilliant") that we use are based on the ones selected in Storage et al.'s (Storage et al. 2020) study to analyze if people associate brilliance with men more than women. These are "brilliant", "genius", "brainiac", and "super smart."

Each prompt is constructed as "[trait] person," resulting in the following 4 prompts: "Brilliant person", "Genius person", "Brainiac person", and "Super Smart person." We capitalize all the traits. We use the word "person" with each trait to neutralize the gender of the trait in our prompt and guide the models toward creating a human. The aim is to determine whether the model tends to identify high intellect with men or women.

To more accurately ascertain the models' Brilliance Bias versus other forms of gender bias, we have decided to test the models' behaviour on the more basic prompt "person." This exploration is motivated by the presence of such a bias in humans, whereby people assume that gender neutral words refer to men (Bailey, Williams, and Cimpian 2022). We label it as "None" in our graph results.

We generate 100 images per prompt, totalling to 500 images per model. All together, we look at 2000 images across Dall-E, Midjourney, Craiyon and Stable Diffusion. In this set of experiments, we intentionally avoid specifying style in order to reduce the risk of additional influences.

We run: Dall-E on its website, Midjourney on its discord, Stable Diffusion on its DiffusionBee app, and Craiyon via its website too. We run each prompt on Dall-E and Midjourney 25 times, with each generating 4 images to create 100 images. Craiyon produces 9 images per prompt, so we run each prompt 12 times and take the first 100 images. We set Stable Diffusion to generate 100 images per prompt and set its Guidance Scale to the maximum 20 in order to understand its behavior when it is more strongly influenced by the prompt.

Across the gender spectrum, for the purposes of our study we focus on Brilliance Bias in the context of the binary genders male and female. We use terms representing binary genders such as "male" and "female" and "woman" and "man" in our paper as shorthand for a figure identified by our analysis as exhibiting binary male-identifying or female-identifying traits.

Once all the images are generated, we look at how many are of a woman vs a man to study if high intellect is more often associated with men or women. In order to do this, we manually count the number of women and men in these images. We count our images based on 3 categories: Male, Female, and Other. For an image we could not determine a gender or that did not have a person, we count it as "Other." Although rare, some images display multiple people (most often seen in Craiyon and Stable Diffusion). If an image includes at least one male and at least one female, we label it as "Other." Most images portray a single person, and are easily classified as showing male-identifying or female-identifying traits looking at a combination of physique, clothing and facial features. We expect negligible deviation if multiple people were to label the images. [4]

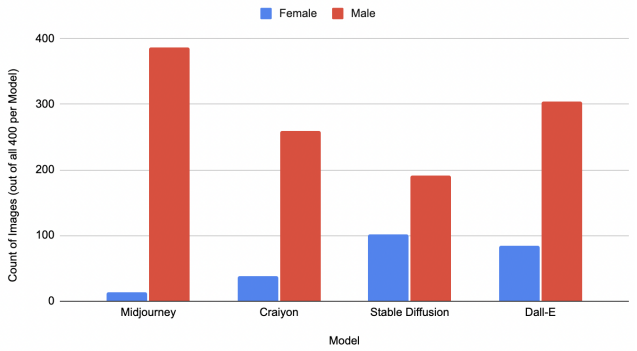

Figure 11: **Comparing models' net count of female and male images for the brilliance prompts**. We created 400 images per model, 100 images for each brilliance trait we evaluated: "genius", "brilliant", "super smart" and "brainiac".

Results

Our analysis demonstrates a clear presence of Brilliance Bias in Midjourney, Stable Diffusion, and Dall-E. The results for Craiyon are inconclusive due to gender bias seen on "person" as well. Furthermore while Stable Diffusion clearly demonstrates Brilliance Bias, it is less biased than the other models because of its performance on the prompts "Brilliant person" and "Super Smart person".

We consider the overall ratio of generated images of women to images of men across all the traits we test for each model, shown in Figure 11. In most cases, the models produce at least twice as many men as women on the brilliance prompts, often with the disparity being much greater. Midjourney shows the greatest disparity in number of images of women to men, with 3.25% women vs. 96.5% men, followed by Craiyon 9.5% women vs. 65% men, Dall-E 21% women vs. 76% men and Stable Diffusion 25.25% women vs. 47.75% men. The only exception is Stable Diffusion, which shows a ratio slightly below 2x. These results are rather unfortunate, since Midjourney is known to incorporate more art style.[5]

[4]For completeness, we look at studies of gender assigning based on facial features and clothing. Men are found to have a more prominent chin/jaw and protuberant nose/brows (Bruce et al. 1993). Women are found to have higher eyebrows while men have thicker eyebrow closer to their eyes (Brown and Perrett 1993). Women are noted to have fuller cheeks and less facial hair including around their eyebrows, while men have more facial hair or hair follicles otherwise (Burton, Bruce, and Dench 1993). A study looking at the Halloween clothing of children found female clothing are more decorative and exposing of skin, while male clothing are more functional (Murnen et al. 2016).

[5]https://simplified.com/blog/ai-text-to-image/dall-e-2-vs-midjourney/, https://startuptalky.com/dall-e-vs-midjourney/

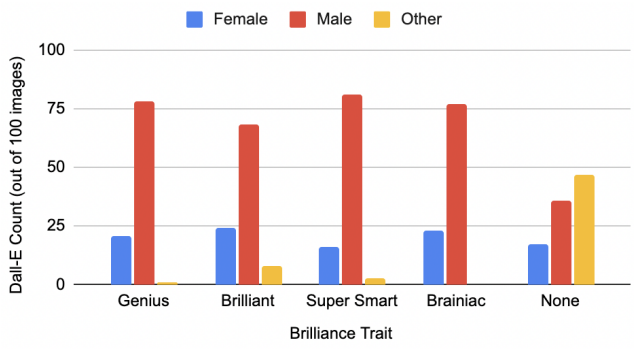

Figure 12: **Dall-E** Brilliance Bias results. The number of Female, Male, and Other count for each Brilliance Trait tested.

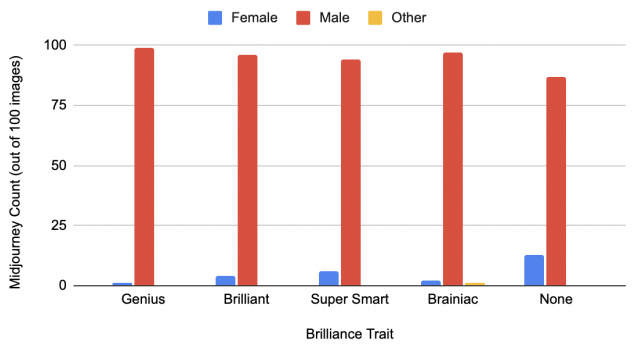

Figure 13: **Midjourney** Brilliance Bias results. The number of Female, Male, and Other count for each Brilliance Trait tested.

Across all four models, nearly all prompts result in a significantly stronger association of high levels of intellect to men. Midjourney in particular has the largest gap between men and women, as shown in Figure 11. Craiyon and Stable Diffusion, seen in Figures 14 and 15, have the highest number of images labeled as "Other" amongst the models studied. Meanwhile, Stable Diffusion is an exception on the prompt "Brilliant person" resulting in a higher female:male ratio as seen in Figure 15 compared to the other models. This could be due an alternate meaning of "brilliance," which can be defined as "full of light, shining, or bright in color".[6] Stable Diffusion's images on "Super Smart person" results in the closest count of images between all three categories: Female, Male, and Other.

Midjourney produces almost no women for the prompts "Genius person" and "Brainiac person." Craiyon generates no women out of 100 images for the prompt "Genius" and almost no women for the prompt "Super Smart person." There is no one consistent brilliance trait that shows the highest level of Brilliance Bias across all the models, as each model varies in performance on the four traits. However, all the models show a significant difference between the number of female and male images for brilliance prompts.

We compare how the ratios of male to female images

[6]https://dictionary.cambridge.org/us/dictionary/english/brilliant

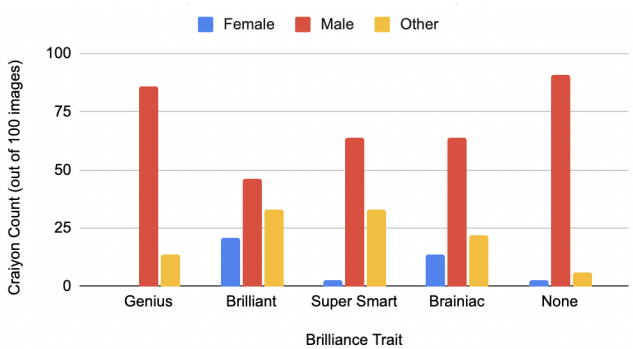

Figure 14: **Craiyon** Brilliance Bias results. The number of Female, Male, and Other count for each Brilliance Trait tested.

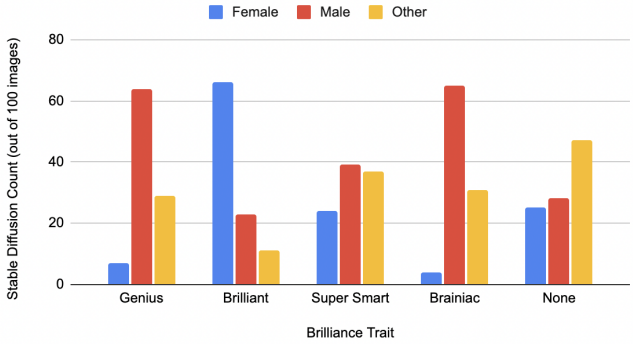

Figure 15: **Stable Diffusion** Brilliance Bias results. The number of Female, Male, and Other count for each Brilliance Trait tested.

in brilliance prompts contrast to the same ratios on the prompt "person." Stable Diffusion has the closest 50:50 female:male ratio on the non-brilliance prompt "person" as seen in Figure 15. On the other hand, the rest of the models have a more notable higher count of men to women when prompted to generate a "person". However, Midjourney's results in Figure 13 show it generates less women for brilliance invoking prompts compared to the non-brilliance prompt "person". For these three models, brilliance prompts lead to a much greater difference in the number of women vs men generated, suggesting strong evidence of Brilliance Bias.

On the contrary, Craiyon generates a higher count of men for both the brilliance prompts and non-brilliance prompt as seen in Figure 14. This makes it more challenging to separate Brilliance Bias from other forms of gender bias for this model. Future work will be needed to assess the "default male" bias and separate it from Brilliance Bias in the Craiyon model.

Discussion

A comparison of the models' performance on brilliance prompts and the non-brilliant prompt "person" indicates Stable Diffusion, Midjourney, and Dall-E are Brilliance Biased

while Craiyon's Brilliance Bias is questionable for the time being. Since these models are trained on data created by people, they are simply revealing the biases that exist in society. The Brilliance Bias we are seeing here is a mirror of the collective unconsciousness of society at large. However, these models, and biases they embody, will influence society on a large scale due to their popularity and the influence of images and media on people. Consequently, they will stand to hold back inclusivity progress. Below we discuss a comparison of how the models' portray brilliant men and women. Furthermore, we discuss how we can mitigate Brilliance Bias in text-to-image models.

How Brilliant Men and Women are Portrayed

We observe notable differences in how text-to-image models visualize intellectually brilliant men vs women. One such difference came across in the prompt "Brilliant person." In this case, we find that Dall-E visualizes the "brilliance" emanating from the men, while for female characters, the brilliance is visualized as a decorative environmental factor. See Figure 3.

The results also suggest that the term "brilliant" more often represents the non-intellectual interpretation of the term when it came to women, "full of light, shining, or bright in color." [7] This came across in higher adornment of women with fancy jewels, makeup, and radiant smiles, which was not the case for the generated images of men under the same prompt, whose visualization better align with the intellectual interpretation of the term "brilliant." See Figure 3.

Stable Diffusion makes images of brilliant men more often photorealistic, while brilliant women are visualized in a more artistic fashion, as seen in Figure 9. It is interesting to see that Stable Diffusion shows groups of women multiple times when prompted with "Brilliant person," as well, compared to more often showing a man by himself. This appears to imply that women are not individually capable of holding high-levels of intellect, rather it is through a group effort that they achieve brilliance. This may reflect unconscious biases in society, absorbed through the images that the models are trained on.

Additionally, we notice objects around women's heads more often compared to men, for example, as seen in Figure 2 with a light bulb and cloud above two women's heads. However, such illustrative elements are not as commonly seen above men's head. Why do the models end up adding these objects for brilliant women but not men? This seems to suggest that to appear convincingly brilliant, a woman needs visualizations of her thinking, while a man's intellect can be assumed without such props. In future work, it may be interesting to analyze the items (ex. swirling icons, thought bubbles, items emanating from a person's head) that tend to co-occur in generated visualizations of brilliant men vs women.

Moreover, we notice multiple images cut off women's faces in Dall-E and Stable Diffusion. This can be seen in Dall-E's images in Figures 2 and 3, and in Stable Diffusion's Figure 10 top row last image. Even more so, we find images Craiyon creates, in particular for the prompt "Brilliant person" and "Brainiac person", portray women with more exposed skin, as seen in Figures 7 and 8, and nudity. Furthermore, Midjourney more often shows men as cyberborgs as seen in Figure 5. For the "genius" prompt, Craiyon generates zero women as seen in Figure 6.

The above summarizes our observations. Further analysis would be needed to conclusively report on the above.

General Stylistic Elements

Across all the models, we note a few generic stylistic elements. For instance "brainiac" is affiliated with green colors, robotic-like figures, and persons that have a Frankenstein-like look too. These images resemble the comic book character "brainiac"[8], potentially suggesting that for this prompt the models may be more influenced by that character than the intellect-related definition of "brainiac." Stable Diffusion incorporates more colors to "brainiac" though, particularly pink and purple In addition, we notice Dall-E often times shows a brain with "brainiac" as seen in the images in Figure 4. Furthermore, the trait "super smart" results in common superman stylistic details across all models, including caps and red and blue colors. Additionally, Midjourney shows a person's face the most clearly but adds some artistic texture, with Dall-E showing a person in a photographic style more often. Craiyon least often shows a real-person. Stable-Diffusion most often adds text to images, although a majority of the time it did not make sense. Lastly, Midjourney affiliates "brilliant" and "genius" less often to younger people compared to the other models.

Mitigating Brilliance Bias

We explore purposefully altering the style specified in a prompt to see if it can help mitigate the Brilliance Bias we found. We assume adding the keyword "contemporary art style" might influence the models to generate more gender inclusive images. This is in consideration that society has progressed (to a certain degree) toward being more inclusive of women and thus we hypothesize that a contemporary style would reflect that. However, an exploratory analysis shows that just adding "contemporary art style" keeps most of the images male-dominant.

We further explore the specific contemporary art style "Feminist art", defined as a "movement [that] arose in an attempt to transform stereotypes and break the model of a male-dominated art history" (Invaluable 2021), and find it to very clearly increase the number of images that had a woman or women. This is not too surprising though given that the art style focused on enhancing the representation of women, making a point that the text-to-image models are representing society's cultures and beliefs accurately. However, it is worth exploring the variety of art styles more in depth in future work. Furthermore, the word "feminist" itself tends to be associated with women, and may prompt more images of women as these models often use words that appear in the prompts out of the context.

[7] https://dictionary.cambridge.org/us/dictionary/english/brilliant

[8] https://www.dc.com/characters/brainiac

Conclusions and Future Work

In this paper, we evaluate the presence of Brilliance Bias in four text-to-image models: Dall-E, Midjourney, Craiyon and Stable Diffusion. Our results reveal that text-to-image models show men much more often than women when asked to generate a person portraying brilliance.

There is a substantial presence of the Brilliance Bias in the Dall-E, Midjourney, and Stable Diffusion. The results are more ambiguous in the case of Craiyon as it reveals gender bias regardless of brilliance. Midjourney and Stable Diffusion generate fewer images of women on brilliance traits compared to the non-brilliant prompt "person." Dall-E often presents gender-neutral images when prompted to create a "person", while associating brilliance to men. Midjourney shows the most significant difference in ratio of women compared to men when given brilliance prompts, with women shown in only 3.25% of its images. Craiyon created 9.5% images of brilliant women, followed by Dall-E with 21%, and Stable Diffusion with 25.25%.

This analysis leads us to realize that there is another fundamental bias that needs to be studied in text-to-image models. That bias, which has been found in humans, is the tendency to assume that gender neutral terms such as "person" refer to men rather than women (Bailey, Williams, and Cimpian 2022). Craiyon generates more images of men than women for brilliance induced prompts. However, it creates even more images of men when prompted with the non-brilliance prompt "person." Thus, Craiyon seems to exhibit a more fundamental bias, the assumption that people are men, making it more challenging to ascertain the extent to which it exhibits Brilliance Bias.

We hope that this work spurs interest in further analysis as well as mitigation of biases in generative models, particularly those that are widely accessible. We have conducted an initial analysis into this foray. In particular, the bias whereby the models assume that general neutral words refer to men deserves further study. One of the greatest challenges arising from our results is the mitigation of Brilliance Bias in generative models. Solutions can come in the form of creating new models that do not exhibit this bias, or corrective tools that work in conjunction with large models. While in this initial study we focus on male vs female analysis of Brilliance Bias, it is worth expanding this analysis across the gender spectrum.

Images play a critical role in influencing people's perception of themselves, their abilities and of the potential they see in themselves and others. Given text-to-image models are rapidly growing in popularity, it is important to understand their biases to help mitigate their spread. Rather than introduce biases that set back progress society makes on inclusivity efforts, it is important to navigate these popular image generators toward a more equitable and diverse representation of society.

References

Adams, C. 2022. Turn imagination into reality with text to image in Canva. https://www.canva.com/newsroom/news/text-to-image-ai-image-generator/. (accessed: 02.27.2023).

Agarwal, S.; Krueger, G.; Clark, J.; Radford, A.; Kim, J. W.; and Brundage, M. 2021. Evaluating clip: towards characterization of broader capabilities and downstream implications.

Bailey, A. H.; Williams, A.; and Cimpian, A. 2022. Based on billions of words on the internet, people = men. *Science Advances* 8(13):eabm2463.

Bian, L.; Leslie, S.-J.; and Cimpian, A. 2017. Gender stereotypes about intellectual ability emerge early and influence children's interests. *Science* 355(6323):389–391.

Bian, L.; Leslie, S.-J.; and Cimpian, A. 2018. Evidence of bias against girls and women in contexts that emphasize intellectual ability. *American Psychologist* 73(9):1139.

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*.

Brown, E., and Perrett, D. I. 1993. What gives a face its gender? *Perception*.

Bruce, V.; Burton, A. M.; Hanna, E.; Healey, P.; Mason, O.; Coombes, A.; Fright, R.; and Linney, A. 1993. Sex discrimination: how do we tell the difference between male and female faces? *perception*.

Burton, A. M.; Bruce, V.; and Dench, N. 1993. What's the difference between men and women? Evidence from facial measurement. *Perception*.

CBInsights. 2023. The future of the Super Bowl Ad: How tech like generative AI is automating TV ad creation production. https://www.cbinsights.com/research/future-of-super-bowl-tv-advertising/. (accessed: 02.27.2023).

Celestin, M. 2011. Empowering and engaging teen girls through media from the perspective of a practitioner and producer.

Cheryan, S.; Plaut, V. C.; Handron, C.; and Hudson, L. 2013. The stereotypical computer scientist: Gendered media representations as a barrier to inclusion for women. *Sex roles*.

Chiriguayo, D., and Ta, A. 2022. Lensa AI portrait app may be using old beauty standards and male biases. https://www.kcrw.com/news/shows/press-play-with-madeleine-brand/ai-misogyny-migrants-soccer/lensa-ai. (accessed: 02.27.2023).

Cohen, H. 1999. Colouring without seeing: a problem in machine creativity. *AISB quarterly* 102:26–35.

Colton, S.; Halskov, J.; Ventura, D.; Gouldstone, I.; Cook, M.; and Ferrer, B. P. 2015. The Painting Fool sees! new projects with the automated painter. In *ICCC*, 189–196.

Curry, D. 2023. Lensa AI revenue and usage statistics. https://www.businessofapps.com/data/lensa-ai-statistics/. (accessed: 02.27.2023).

Deiglmayr, A.; Stern, E.; and Schubert, R. 2019. Beliefs in "brilliance" and belonging uncertainty in male and female stem students. *Frontiers in psychology*.

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

Feng, Y., and Shah, C. 2022. Has CEO gender bias really been fixed? adversarial attacking and improving gender fairness in image search. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36.

Ford, L. 2019. Geena Davis: damaging stereotypes on screen limit women's aspirations. https://www.theguardian.com/global-development/2019/oct/01/geena-davis-damaging-stereotypes-on-screen-limit-womens-aspirations. (accessed: 02.27.2023).

Garcia, C. 2016. Harold Cohen and AARON — a 40-year collaboration. https://computerhistory.org/blog/harold-cohen-and-aaron-a-40-year-collaboration/. (accessed: 02.27.2023).

Heikkilä, M. 2022. The viral AI avatar app Lensa undressed me—without my consent. https://www.technologyreview.com/2022/12/12/1064751/the-viral-ai-avatar-app-lensa-undressed-me-without-my-consent/. (accessed: 02.27.2023).

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).

Institute, G. D. 2016. Female characters in film and TV motivate women to be more ambitious, more successful, and have even given them the courage to break out of abusive relationships. https://seejane.org/gender-in-media-news-release/female-characters-film-tv-motivate-women-ambitious-successful-even-given-courage-break-abusive-relationships-release/. (accessed: 02.27.2023).

Invaluable. 2021. Art history timeline: Western art movements and their impact. https://www.invaluable.com/blog/art-history-timeline/. (accessed: 02.27.2023).

Javed, A. 2015. The media, the women and stem fields.

Kay, M.; Matuszek, C.; and Munson, S. A. 2015. Unequal representation and gender stereotypes in image search results for occupations. In *Proceedings of the 33rd annual acm conference on human factors in computing systems*, 3819–3828.

Kiefer, B. 2022. Nestlé brand is latest to venture into brave new world of AI art direction. https://www.adweek.com/creativity/nestle-brand-is-latest-to-venture-into-brave-new-world-of-ai-art-direction/. (accessed: 02.27.2023).

Koidan, K. 2020. 8 AI companies generating creative advertising content. https://www.topbots.com/ai-companies-generating-creative-advertising-content/. (accessed: 02.27.2023).

Lam, O.; Broderick, B.; Wojcik, S.; and Hughes, A. 2018. Gender and jobs in online image searches.

LeanIn.Org. 2023. You can't be what you can't see: The Lean In collection on Getty images. leanin.org/getty. (accessed: 02.27.2023).

Leslie, S.-J.; Cimpian, A.; Meyer, M.; and Freeland, E. 2015. Expectations of brilliance underlie gender distributions across academic disciplines. *Science* 347(6219):262–265.

Mattei, S. E.-D. 2022. Careful — Lensa is using your photos to train their AI. https://www.artnews.com/art-news/news/does-lensa-ai-use-your-face-data-for-selfies-1234649204/. (accessed: 02.27.2023).

Mello-Klein, C. 2022. The AI portrait app Lensa has gone viral, but it might be more problematic than you think. https://news.northeastern.edu/2022/12/09/portrait-ai-app/. (accessed: 02.27.2023).

Microsoft. 2023. Microsoft and OpenAI extend partnership. https://blogs.microsoft.com/blog/2023/01/23. (accessed: 02.27.2023).

Miller, C. C. 2014. LeanIn.org and Getty aim to change women's portrayal in Stock photos. https://www.nytimes.com/2014/02/10/business/leaninorg-and-getty-aim-to-change-womens-portrayal-in-stock-photos.html?smid=pl-share. (accessed: 02.27.2023).

Murnen, S. K.; Greenfield, C.; Younger, A.; and Boyd, H. 2016. Boys act and girls appear: A content analysis of gender stereotypes associated with characters in children's popular culture. *Sex roles*.

OpenAI. 2022. Reducing bias and improving safety in DALL·E 2. https://openai.com/blog/reducing-bias-and-improving-safety-in-dall-e-2. (accessed: 02.27.2023).

Potter, W. J. 1993. Cultivation theory and research: A conceptual critique. *Human communication research* 19(4):564–601.

Q.ai. 2022. Dall·E Mini and the future of artificial intelligence art. https://www.forbes.com/sites/qai/2022/10/21/dalle-mini-and-the-future-of-artificial-intelligence-art/?sh=50e8121b7d78. (accessed: 02.27.2023).

Raji, I. D.; Gebru, T.; Mitchell, M.; Buolamwini, J.; Lee, J.; and Denton, E. 2020. Saving face: Investigating the ethical concerns of facial recognition auditing. In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*.

Rayner, A. 2016. Can Google's Deep Dream become an art machine? https://www.theguardian.com/artanddesign. (accessed: 02.27.2023).

Salkowitz, R. 2022. Midjourney founder David Holz on the impact of AI on art, imagination and the creative economy. https://www.forbes.com/sites/robsalkowitz. (accessed: 02.27.2023).

Schwemmer, C.; Knight, C.; Bello-Pardo, E. D.; Oklobdzija, S.; Schoonvelde, M.; and Lockhart, J. W. 2020. Diagnosing gender bias in image recognition systems. *Socius* 6:2378023120967171.

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)*.

Shrum, L. J. 1995. Assessing the social influence of television: A social cognition perspective on cultivation effects. *Communication Research* 22(4):402–429.

Shubber, K. 2013. Artificial artists: when computers become creative. https://www.wired.co.uk/article/can-computers-be-creative. (accessed: 02.27.2023).

Shutterstock. 2023. Shutterstock introduces generative AI to its all-in-one creative platform. https://www.prnewswire.com/news-releases/shutterstock-introduces-generative-ai-to-its-all-in-one-creative-platform-301729904.html. (accessed: 02.27.2023).

Srinivasan, R., and Uchino, K. 2021a. Biases in generative art: A causal look from the lens of art history. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*.

Srinivasan, R., and Uchino, K. 2021b. Quantifying confounding bias in generative art: A case study. *arXiv preprint arXiv:2102.11957*.

Steele, C. 2022. Lensa AI is carrying gender bias into the future. https://www.pcmag.com/opinions/lensa-ai-is-carrying-gender-bias-into-the-future. (accessed: 02.27.2023).

Storage, D.; Charlesworth, T. E.; Banaji, M. R.; and Cimpian, A. 2020. Adults and children implicitly associate brilliance with men more than women. *Journal of Experimental Social Psychology*.

Suddath, C. 2014. How Sheryl Sandberg and Getty are making Stock photos less sexist. https://www.bloomberg.com/news.

Sun, J.; Liao, Q. V.; Muller, M.; Agarwal, M.; Houde, S.; Talamadupula, K.; and Weisz, J. D. 2022. Investigating explainability of generative AI for code through scenario-based design. In *27th International Conference on Intelligent User Interfaces*.

Sundararajan, L. 2021. Harold Cohen and AARON: Collaborations in the last six years (2010–2016) of a creative life. *Leonardo*.

Sung, M. 2022. The AI Renaissance portrait generator isn't great at painting people of color. https://www.artnews.com/art-news/news/does-lensa-ai-use-your-face-data-for-selfies-1234649204/. (accessed: 02.27.2023).

Ward, L. M., and Grower, P. 2020. Media and the development of gender role stereotypes. *Annual Review of Developmental Psychology* 2:177–199.

Welsh, J. 2013. These are the 7 things keeping women out of science careers. https://www.businessinsider.com/7-things-keeping-women-out-of-science-2013-10. (accessed: 02.27.2023).

Wiggers, K. 2021. Audit finds gender and age bias in OpenAI's CLIP model. https://venturebeat.com/business/audit-finds-gender-and-age-bias-in-openais-clip-model/. (accessed: 02.27.2023).