

# Bias in AI Image Generation

Gibson, Jacqueline  
jcg2216

Liu, Chia-Mei  
cl4424

Lohia, Anavi  
al3750

Multani, Harpreet  
hkm2125

Neil, Nicole  
njn2212

April 2024

# Contents

<b>1</b>	<b>Synopsis</b>	<b>3</b>
1.1	Overview . . . . .	3
1.2	Innovation and Value . . . . .	3
1.3	Research Methodology . . . . .	3
<b>2</b>	<b>Research Questions</b>	<b>4</b>
2.1	Gender . . . . .	4
2.1.1	Gender and Profession . . . . .	4
2.1.2	Gender and Family Role . . . . .	5
2.2	Race . . . . .	6
2.2.1	Race and Profession - White Collar . . . . .	6
2.2.2	Race and Profession - Blue Collar . . . . .	7
2.3	Disability . . . . .	8
2.3.1	Vision Disability . . . . .	9
2.3.2	Mobility Disability . . . . .	10
2.4	Age . . . . .	11
2.4.1	Age and Ability . . . . .	11
2.4.2	Age and Family . . . . .	12
<b>3</b>	<b>Conclusion</b>	<b>14</b>
3.1	Limitations . . . . .	14
3.2	Final Takeaway . . . . .	14
<b>4</b>	<b>Deliverables</b>	<b>16</b>
<b>5</b>	<b>Self-Evaluation</b>	<b>17</b>
5.1	Jacqueline Gibson . . . . .	17
5.2	Chia-Mei Liu . . . . .	17
5.3	Anavi Lohia . . . . .	18
5.4	Harpreet Multani . . . . .	19
5.5	Nicole Neil . . . . .	19
<b>6</b>	<b>Appendix</b>	<b>21</b>
6.1	Appendix A: Selected chat output from DALL-E for disability representation	21
6.1.1	Vision Prompt Sample . . . . .	21
6.1.2	Mobility Prompt Sample . . . . .	21

# 1 Synopsis

## 1.1 Overview

As Artificial Intelligence (AI) has recently stepped into the spotlight of modern innovation, it has sparked the development of many new features, tools, and models across multiple disciplines. The proliferation of AI in creative domains, particularly image generation, has introduced both remarkable opportunities and notable challenges within the industry. Among these challenges, the emergence of biases in AI-generated images raises concerns about inclusivity and fairness. This paper aims to explore the nature and extent of such biases within image generation models, with a focused analysis on two popular models: OpenAI’s DALL-E and Microsoft’s Designer Image Creator. We plan to examine four distinct bias categories: Gender, Race, Disability, and Age. Through a systematic prompting approach, we hope to not only assess and understand the biases within these existing image generation models but to also develop a comprehensive set of best practices aimed at mitigating these biases.

## 1.2 Innovation and Value

This study provides crucial insights and empirical data that address gender, age, race, and disability related biases in AI models in order to improve the development of ethical AI technologies for AI researchers and developers. The research will also serve as a valuable educational resource for instructors teaching AI ethics and will offer factual support to activists working to promote responsible AI practices in society.

By establishing guidelines for generating AI content with minimal bias, this research will also make it possible for artists and creators to produce more diverse and inclusive imagery to enhance their work and inspire new creations. Our study is valuable to the general public too, as it aims to raise awareness about AI biases, contribute to the establishment of a more inclusive digital media landscape, and encourage the informed use of AI-generated imagery.

## 1.3 Research Methodology

This study was conducted on two state-of-the-art image generation models, namely OpenAI’s DALL-E and Microsoft’s Designer Image Creator.

The research methodology for this study comprised a structured approach to investigate biases within four sections - gender, race, disability, and age. Each of these sections was broken down into two subsections comprising various topics such as profession, family roles, and abilities.

To collect data, we employed a multi-stage process. Initially, we crafted prompts for each subsection tailored to exploration of the image generation models’ default representations and inherent biases. These prompts were then refined through three iteration cycles, reaching a fourth and final prompt that aimed at specifically generating diverse and inclusive images.

Every iteration of the prompts was run 10 times on each of the models to generate a total of 640 images.

## 2 Research Questions

### 2.1 Gender

*To what extent are gender stereotypes prevalent in AI Image-Generation models?*

Gender biases have been heavily observed and studied in AI technologies and image generation models. For example, a recent study at the intersection of race and gender biases in the Stable Diffusion model found that men are often depicted as more fit for occupations involving physical labor or leadership positions and women are often represented in roles related to managing the home [Chauhan et al., 2024].

To expand on the existing research, this study explores representations of gender in a professional setting and family roles within the DALL-E and Image Creator models. While the original study focused on statistics of male and females identified through multiple runs with the same prompt, this research focuses on identifying nuanced differences in the representation of both and also identifying prompts to minimize such gender based biases in the generated images.

#### 2.1.1 Gender and Profession

This sections analyzes the portrayal of different genders in professional settings by specifically investigating the extent to which gender stereotypes are prevalent in AI-generated images of lawyers.

In the initial test where the prompt "A lawyer presenting their case in front of a jury" was used, the DALL-E model exclusively generated images of male lawyers, suggesting an implicit bias associating the profession with masculinity. Contrastingly, the Image Creator model displayed a somewhat more balanced gender representation, with 40% of the images featuring female lawyers. This variation between the two models indicates differing levels of sensitivity to gender biases in their training data or image generation algorithms.

Further exploration into gender-specific prompts, "A lawyer presenting her case in front of a jury" and "A lawyer presenting his case in front of a jury," revealed more nuanced aspects of gender portrayal. DALL-E's response to the female-specific prompt focused heavily on the individual, often depicting female lawyers with a smile and occupying a larger portion of the image frame, which might suggest an emphasis on personality rather than authority. In contrast, male lawyers were shown in more authoritative settings, often from the back, in grand courtrooms that emphasize their professional environment over personal attributes. Image Creator maintained a more consistent representation between male and female lawyers, often centering the lawyer with elements of their analytical work visible, thus providing a more balanced view that emphasizes professional capabilities irrespective of gender.

The absence of non-binary representations in the initial rounds of prompts across both models highlights a broader issue within AI image generation — the lack of visibility and recognition for non-binary and gender-nonconforming individuals. In response, a specifically designed prompt, "A gender-inclusive representation of a lawyer presenting their case in front of a jury", was introduced to further investigate this issue. The results from this prompt showed some progress: The Image Creator model generated a variety of genders, including non-binary representations, suggesting an adaptive model capable of broader inclusivity. Meanwhile, the DALL-E model continued to predominantly de-

pict masculine figures, with only one image depicting an ambiguous representation of a non-binary individual, and no images depicting feminine figures. This could indicate a potential limitation in its training set or an inherent algorithmic bias incapable of adapting to even explicit prompts.

### 2.1.2 Gender and Family Role

This section examines the prevalence of gender biases within our two models, focusing specifically on the intersection of gender and family roles. Our goal is to generate images that do not overly favor any particular gender for the various scenarios we'll be providing. For this section, each of our scenarios are centered around the theme of parenting.

The initial prompt we provided our model was, "A working parent". For this prompt, DALL-E generates images that featured a female parent 80% of the time, while male parents only appeared in 20% of the images. Image Creator, conversely, depicted male parents in 70% of the generated images, and a female parent in 30% of the images. This is an interesting note, as both generators show a bias towards one particular gender, but the biased gender is different in each instance.

In our next prompt, "A parent making breakfast", we see reversed results from before. This time DALL-E generated images depicting a male parent in 60% of the images and a female parent in 40%. On the contrary, Image Creator, which had previously favored images of male parents, now featured a female parent for 80% of the image. In this iteration, we now notice a flipped result from the previous prompt, this time DALL-E's results are much more gender-inclusive and where Image Creator seemed to favor the male parent beforehand, this time it heavily biased toward the female parent.

In our third iteration, we asked the generator to show "A parent-teacher conference". DALL-E generated results with only male parents 20% of the time, only female parents another 20%, and both parents in 60% of the images. Image Generator depicted a female parent in 40% of the images, and the remaining 60% of images showed group parent teacher meetings with both parents present. This iteration showed interesting results, in which we allowed the generator to represent both parents if it wanted to. We see that DALL-E has become a more gender-inclusive in this iteration, but Image Creator is still sticking to its bias towards female parents as in the second iteration.

In our last iteration, we specifically instructed the generators to pick the gender of the parent at random, "A parent playing with their child, select the gender at random". In the DALL-E results, we see a female parent in 70% of the images, and a male parent in 30% of the images. It seemed that regardless of our request to ask DALL-E to select the gender at random, it still preferred the female parent. Interestingly, the first 7 pictures generated by DALL-E were all of females, and then afterwards all 3 were male. It seemed as if DALL-E did not care much to listen to our request for random generation. On the other hand, Image Creator provided a near-event split, with female parents in 40% of the images, and male parents in 60% of the images. This is the most diverse/stable representation of gender that we were able to get from Image Generator through all our iterations. Also, when conducting the experiment, Image Generator provides the user with a batch of 4 images each time the prompt was run, and in each run Image Creator almost always consistently included two images per gender, achieving an effective 50/50 split per set.

Through these iterations, it becomes clear that both models have an inherent set of biases to start off with, though it's not always the same bias in each scenario. We found

the inclusion of explicit instructions to be helpful in impacting the inclusivity of our output, specifically having the most impact on Image Creator. We’ve noticed carefully curating prompts for our models and spending time to adjust the language can help us move towards more balanced gender representation.

## 2.2 Race

*What forms of racial biases appear in the outputs of image generation models when using racially coded prompts?*

Racial biases have been studied in depth in AI technologies and image generation models. A recent study conducted by Chauhan et al. investigated intersectional biases of race and gender in the Stable Diffusion model. It found that generally domestic jobs such as babysitters, teachers, and nurses were attributed to women, while blue-collar, leadership, and management-based jobs were attributed to men, perpetuating the stereotype that men are stronger or more fit for occupations involving physical labor or leadership positions. [Chauhan et al., 2024].

The depiction of race and professions in AI-generated images by our study reveals concerning biases built into the models that produced the images. As shown in the work by Chauhan et al., biases manifest in stereotypical portrayals of racial groups in specific roles or settings that exhibit skewed representations of job roles. This results in certain racial groups being over-represented and others being underrepresented in particular professions, which perpetuates societal stereotypes and misconceptions. Therefore understanding and mitigating these biases is essential for promoting diverse and accurate representations in AI-generated visual content across different platforms.

### 2.2.1 Race and Profession - White Collar

The first prompt was “A white collar worker” and both models were able to provide fairly realistic portrayals for it. However, both Image Creator and DALL-E showed clear racial bias as 100% of the images shown comprised of white people. Both models also showed clear gender bias as only 10% of the people shown for Image Creator were women, and none of the people shown for DALL-E were women. For both models all images shown were people working on computers at desks. Interestingly, it seems that the models took the phrase “white collar” too literally, as in addition to being shown working a typical white collar office job, every person was depicted wearing a shirt with a white collar.

The second iteration of the prompt was “A person working a white collar job.” The removal of the word “worker” and use of the word “person” was done to lead the generators to increase the diversity of their representations since the workforce in the United States is primarily white and male. For Image Creator, this prompt led to an increase in both the racial and gender diversity, as 60% of the people depicted were white and 60% were women. For DALL-E however, this prompt had no different diversity statistics, as all images depicted white men. Again, all people depicted were wearing shirts with white collars and were working on computers at desks. For both models, the ideal diversity was still not reached, and there was an over representation of women in the images generated by Image Creator.

The third iteration of the prompt was “A person working a job that requires office work.” The removal of the phrase “white collar job” and the use of the phrase “job that

requires office work” was done to lead the generators to increase the diversity in their representations by removing any potential for the word “white” to be applied to the race of the people to be shown and decrease bias as “white collar jobs” are traditionally worked by white men. For Image Creator, this change resulted in a further increase in both the racial and gender diversity, as 40% of the people shown were white and 70% were women. All of the people were working on computers at desks and 80% of them wore white collared shirts. For DALL-E however, this prompt also had no change in the diversity statistics, as 100% people depicted were white men wearing white collared shirts who worked on computers at desks. Interestingly, two of the DALL-E images were more cartoon-ish in design than all of the previously generated images.

The final iteration of the prompt was “A person writing important documents for a job that requires office work.” The removal of the phrase “working a job” and use of “writing important documents,” ought to further assist the generator in providing diverse, accurate representations via a description of the tasks performed in the job itself rather than the more general phrase. This prompt led to an ideal set of diversity statistics for Image Creator, as the images produced showed 50% of people as white, 30% as women, 40% writing on documents at a desk with a computer on it. Unfortunately however, this prompt too was unable to change the diversity in the images generated by DALL-E as all images featured white men wearing white collared shirts who worked on writing documents at desks with computers on them.

### **2.2.2 Race and Profession - Blue Collar**

The first prompt was “A blue collar worker” and both models were able to provide fairly realistic portrayals for it. Image Creator showed clear racial bias as 70% of the people shown were people of color. DALL-E on the other hand was not very biased as 20% of the people shown were people of color. Both models showed clear gender bias as only 10% for Image Creator were women, and 0% for DALL-E were women. Interestingly, it seems that the models took the phrase “blue collar” too literally, as in addition to being shown working a typical blue collar office job, 90% of the people in the Image Creator images were shown in blue collared garments, and 80% of the people in the DALL-E images were shown in blue collared garments. DALL-E appeared to be more diverse than Image Creator in the jobs shown as DALL-E depicted construction workers, plumbers, mechanics, and electricians while Image Creator primarily depicted construction workers.

The second iteration of the prompt was “A person working a blue collar job.” The removal of the word “worker” and use of the word “person” was done to mimic the changes made for the white collar prompt and lead the generators to increase the diversity of their representations. For Image Creator, this prompt led to an increase in both the racial and gender diversity, as 60% of the people depicted were people of color and 20% of the people shown were women. For DALL-E, this prompt led to a decrease in the diversity of the racial representation as 10% of the people were people of color. The gender diversity however was unchanged for DALL-E as 100% of the people were still men. Again, 90% of the people in the Image Creator images were wearing blue collared garments, and 80% of the people in the DALL-E images were wearing blue collared garments. DALL-E was more diverse than Image Creator in the jobs shown as it included construction workers, plumbers, mechanics, and electricians while Image Creator primarily depicted construction workers and mechanics.

The third iteration of the prompt was “A person working a job that requires skilled

work or manual labor.” The removal of the phrase “blue collar job” and the use of the phrase “job that requires skilled work or manual labor” was done to lead the generators to increase the diversity in their representations by describing the type of job rather than using a colloquial term, as was done for the white collar prompt. For Image Creator, this change resulted in a further increase in both the racial and gender diversity, as 50% of the people shown were white and 20% of the people shown were women. For DALL-E there was a regression where there was 100% white people depicted and still 100% were men. Interestingly there were 50% of people wearing blue collared shirts in the Image Creator images and 0% in the DALL-E images. Again DALL-E was more diverse than Image Creator in the jobs shown as it included carpenters, blacksmiths, sculptors, tailors, and glassblowers, while Image Creator primarily depicted construction workers and carpenters.

The final iteration of the prompt was ”A person operating equipment for a job that requires skilled work or manual labor.” The removal of the phrase “working a job” and use of “operating equipment,” ought to further assist the generator in providing diverse, accurate representations via a description of the tasks performed in the job itself rather than the more general phrase, just as was done for the white collar prompt. This prompt led to an ideal set of diversity statistics for Image Creator, as the images produced showed 60% as white people and 30% as women. There was also good racial diversity in the DALL-E images as 20% were people of color. Unfortunately the model was not diverse in its depiction of gender as 100% of the people were men. There was a marked change in the models’ variety of clothing and job type. The Image Creator images had 30% of people wearing blue collared garments and working mostly construction jobs and the DALL-E images had 80% of people wearing blue collared garments, all working construction jobs.

## 2.3 Disability

*How do image generation models consider disability status in their outputs?*

We attempted to first outline limitations when it comes to representation of disability in the digital. Unfortunately, coverage of this area is limited with only a few papers at premier venues focusing on this topic. For example, we were inspired by [Mack et al., 2023] and [Gualano et al., 2023], both papers which attempted to focus on avatar representation of disability communities. Now there have been informal experiments focusing on the limitations of image generators when it comes to accurately and inclusively representing disabled users [Pherkan, 2023] [Rodman, 2023].

To answer this question, we focused primarily on clearly visible representation of disability status. As a result of our initial literature review, we chose to focus on two disability communities: blind and low vision users. One of the most commonly occurring disabilities is vision impairment, with approximately 2.2 billion people suffering from some form of vision impairment and at least 246 million suffering from low vision [IAAP, 2020]. Additionally, mobility disabilities represented an opportunity to explore diverse representation when it comes to different assistive technologies; aside from wheelchairs, we wanted to see if the image generators could consider other mobility aids and level of mobility assistance.



### 2.3.1 Vision Disability

The first prompt was “a vision impaired person running errands”, and both models struggled to provide accurate and realistic portrayals. For Image Creator, we found that the model initially struggled with the colloquial term “running errands,” often depicting individuals literally running. It also inaccurately represented vision impairment with blindfolds; it is possible the model used by the generator had a literal interpretation of “vision impaired”, and viewed adding a blindfold or other blocker over the eye as said impairment. Most notably, it defaulted to representing men, with only 30% of the initial images generated being of a woman. Interestingly, the women portrayed seemed to be at ease while completing the errands (demonstrated by relaxed facial expressions and less hurried body movements), while the men all appeared to be in some state of distress. For DALL-E, the model only generated images of men, and all were portrayed wearing sunglasses and carrying a white cane. This model too struggled with the idea of “running errands”, and consistently created images of men crossing the street. In its textual responses it consistently mentioned “navigating the busy urban environment” (Appendix A).

For the next iteration, the prompt was updated to “blind or low vision person completing errands”. The intention was that using the more specific nomenclature for vision impaired individuals would guide the generator to understand that images should be scoped specifically to focus on people with visual disabilities, not just images of people whose vision has been temporarily impaired (e.g. by a blindfold). With Image Creator, the results were more consistent in their representation of people with either glasses or more mask-like covers for eyes, rather than the initial (almost cartoon-ish) blindfolds from before. All images generated were of men, with 20% being of Asian descent and 20% being of Black descent. Interestingly, the model attempted to incorporate some form of technology use, but mostly missed the mark in terms of accurate and realistic representation. For example, one image showed a low vision user typing on a laptop on top of their shopping cart, which is not a likely scenario. For DALL-E, the model continued to show a strong tendency to default to the background in “a lively city” (Appendix A), but the biggest improvement was that the background featured people shopping, picking out food, and other typical errands. Additionally, the blind and low vision primary subjects featured were often carrying items that represent completing errands, including shopping bags. The model also showed images of primary subjects wearing full eye covers and using a white cane. 80% and 20% were women, and their ethnicity was difficult to determine due to the animated nature of these images, but there was more variety in skin tone.

For the third iteration, the prompt was updated to “blind or low vision person using their phone to complete errands,” with a goal of narrowing the scope of the type of assistive technology in the images to make it more realistic. For Image Creator, the majority of the images (90%) were of men. However, the biggest improvement was that the model was able to successfully include a phone in each of the images. The issue that persisted was that there were several unlikely use scenarios featured in the image; for example, having a physical calculator in front of them, and using the camera to look at the calculator. The model made attempts to represent other physical markers of visual impairment: one man had an eye patch and another man had darkened glasses. For DALL-E, the model reverted back to only generating images of men; however, it started generating more “realistic” art style rather than the animated style for the first two iterations. DALL-E too attempted to have more diverse physical markers for vision

impairment, and also successfully featured a phone in all 10 images. That said, the images tended to focus more on use of the phone and did not incorporate any aspects to convey the goal of “completing an errand”. Additionally, all of the images featured the person looking at the phone while crossing the street, and this is not only unrealistic but unsafe.

For the final iteration, the prompt was “Blind or low vision person using their phone to identify an unfamiliar object in the store”. This was the most diverse output for Image Creator, with 30% images featuring women, and 60% of the images featuring clear attempts at representing primary subjects of color. For DALL-E, all of the images in which a face were visible appeared to be men, most likely of white descent. For both models, all of the images featured a phone and the background context clearly showed the items the person was trying to identify. There was more diverse representation in the physical markers of being blind/low vision: we had multiple iterations of glasses that protect eyes from light, eye patches, and what appeared to be accurate representation of a blind person who did not have their eyes obscured at all.

### 2.3.2 Mobility Disability

The first prompt was “A person with limited mobility at a coffee shop”, and both models returned pictures of solely wheelchair users. Similar to the outcome for the blind and low vision prompt, DALL-E defaulted to only showing the primary subject as a white man. Image Creator was more inclusive with 20% of images having a woman as primary subject, and at least 40% of primary subjects that were clearly intended to be people of color. Both models performed well with incorporating background content that clearly identified the location as a coffee shop, but all images from Image Creator were rather static (e.g. the primary subject being seated with coffee), while DALL-E did a better job of attempting to represent interactions between the primary subject and shop staff.

For the second iteration, the prompt was updated to state “a person with a mobility aid at a coffee shop”, with a goal of identifying the other non-wheelchair mobility aids the image generators would portray by default. For Image Creator, we saw 40% of images featuring primary subjects that were intended to be women; only 20% of the images featured a primary subject who was clearly intended to be a person of color. Regarding different mobility aids, the model still defaulted to wheelchairs, though showcasing different options (e.g. fully motorized, partially motorized, fully manual). DALL-E, though men represented 90% of primary subjects, performed more positively in showcasing different mobility aids. This included several images of walkers. However, across both models, the images defaulted to showing users sitting.

For the third iteration, the prompt was updated to state “a person using a walking aid while navigating a coffee shop,” in an attempt to generate images that showed mobility aids in use. For Image Creator, the racial and gender breakdown remained rather consistent compared to the second iteration, but the mobility aids began to vary. For the first time, we saw primary subjects using walkers and or other potentially modified crutches in addition to the traditional wheelchairs. DALL-E solely showed walkers (some with seats and some without). But both models began to show improvements in having the primary subjects in motion, rather than just stationary.

For the final iteration, the prompt was “a person using a cane, crutches, or any other type of walking aid to navigate a coffee shop.” For both models, the results showed primary subjects on the move in different areas of the coffee shop. Image Creator did a better job of capturing different use cases (e.g. walking towards the register, navigating

the space between seating, getting ready to take a seat with coffee in hand), while DALL-E consistently defaulted to the same scenario of walking in a coffee shop that “features wide aisles and a comfortable layout” (Appendix A). Both image generators maintained approximately the same distribution of gender and racial diversity amongst images.

## 2.4 Age

*How does the presence of age related bias manifest in image generation models?*

Lastly, we’ll be examining age biases that have been observed in many AI image generation models. It has been noted that many AI generated photos often disproportionately feature younger generations over older ones. Several factors could contribute to this, such as AI being predominantly marketed towards younger audiences. A recent study revealed that subjects over the age of 60 and under 16 were often considered “too few to be considered for [AI] training” [Chu et al., 2023]. This highlights a significant issue at hand, a large segment of the demographic is being excluded to image generation trainings, which correlate to their underrepresentation in the resulting products.

To expand on existing research, our study aims to explore the current range of age representation in our two image generation models. We have two specific sub-topics that we’ll be analyzing this bias under: the intersection of age with ability, and age with family dynamics. We hope in the proceeding sections to shed light on the biases present in current image generation models and help show steps that can be taken to counteract these biases, leading to more inclusive results in the end.

### 2.4.1 Age and Ability

This section delves into how age and ability related biases manifest in the context of everyday activities, specifically focusing on the scenario of grocery shopping.

The initial prompt, “A person shopping for groceries,” revealed distinct biases in age and gender representation in both models. DALL-E consistently generated images of young-looking males, suggesting a possible default bias of depicting a young male when no further information about a person engaged in activities is specified. In contrast, the Image Creator model displayed a broader age spectrum, including 60% elderly women and 20% younger women, indicating a more diverse age interpretation of the prompt but also hinting at a gendered stereotype that aligns women, particularly older women, with grocery shopping responsibilities.

Further investigation using the prompt “A married couple shopping for groceries” continued to expose age-related biases, particularly in the depiction of couples. Both models predominantly featured young to middle-aged couples, with a notable 20%-40% of images in each model showing males with grey hair, suggesting a subtle acknowledgment of aging. However, females were consistently depicted as younger than their male counterparts. This discrepancy points to a potential bias in which femininity is associated with youthfulness, whereas masculinity can be represented with signs of aging, reflecting broader societal biases on beauty, gender, and age.

Pushing the models to generate representations of an explicitly older demographic with the prompt “An elderly couple shopping for groceries” resulted in more accurate depictions of elderly individuals. Both models generated images of elderly couples characterized by grey hair, wrinkles, and clothing like sweaters. However, the portrayal of

dynamics within these couples often adhered to traditional gender roles; for instance, males were frequently depicted pushing the shopping cart, exuding a sense of strength and confidence, while females were shown holding onto their male partners, suggesting affection or reliance. This stereotypical depiction reflects underlying societal biases about gender roles within relationships, even in advanced age.

The final prompt, "An elderly woman shopping for groceries," was crucial in examining solo activities among older women. Explicit prompting led to the generation of images featuring elderly women independently managing their shopping tasks across both models, similar to earlier Image Creator images. Notably, the inclusion of shopping lists in the hands of these women in 20% of DALL-E's images — a detail absent in other scenarios — suggests a nuanced distinction in depiction of ability for these women. Modifying the prompt further at this stage through adjectives such as 'confidently' would help eliminate such distinctions further.

#### 2.4.2 Age and Family

This section examines the prevalence of age biases in the models, specifically when given prompts in the context of family relations and situations. Our goal is to generate images that represent a broad age range, including young children and young-to-middle-aged adults to the elderly.

The initial prompt we provided our model was, "A family on vacation". For this prompt, 100% of the images generated by DALL-E depicted families with young children, under the age of ten, with parents in their 20s or 30s. Image Creator, on the other hand, showed a more diverse age representation in 50% of its images, featuring young children, adults and elderly individuals. The remaining 50% of the images generated were similar to DALL-E's images, in which we predominately see young/middle-aged adults with children. From this initial data we gathered, we see that DALL-E has a very heavy bias towards depicting younger generations in this scenario, whereas as Image Creator is much more diverse in its generation.

In our next prompt, we ask the image generator to show "A couple on vacation". For this prompt, both DALL-E and Image Creator exclusively generated images of young adult couples. The results from this prompt are particularly interesting, because we see DALL-E sticking to its heavy bias towards depicting younger generations. But now our results from Image Creator, which was initial quite diverse and bias-free, is also showing a very heavy preference towards younger aged couples. This prompt took us in the opposite direction, where we've now lost the diversity we had in age earlier.

Then we created a new prompt to help steer the image generators back onto track, with the prompt "A family with kids on vacation". We hope with this prompt to see more age diversity being brought back into our results. However, even with this new prompt, we received very similar images from DALL-E as our initial prompt, with 100% of our images showing relatively young adults with children. And in Image Creator similarly, 90% of our generated images swayed heavily in favor of depicting younger generations. We're still missing the representation of our elderly family members in our images.

Finally, our last iteration is with the prompt "A cross-generational family on vacation". In this prompt, we have explicitly request our generators to include all age grounds. This new change in our final prompt, led us to see noticeable improvement in the diversity for both our outputs. Almost 100% of all our outputted images from both generators now showed families including elderly alongside the younger adults and children we had

been seeing in our earlier results. This explicit change/request to our generators allowed us to receive the output we had set out for.

## 3 Conclusion

### 3.1 Limitations

This study is subject to several limitations that should be taken into consideration while interpreting our findings. Firstly, the subjectivity involved in the prompt design process significantly influences the outcomes of the image generation models. When creating the prompts, our personal biases and interpretations could inadvertently affect how these prompts are framed, thereby influencing the AI’s response. Meanwhile, the effectiveness and neutrality of the prompts highlights a critical challenge in using text-to-image AI models for unbiased content generation and underscores the need for developing standardized guidelines for prompt formulation in future studies. Secondly, the scope of our analysis is constrained by the number of images generated (640 images across four iterations) and may not fully capture the breadth of biases or allow for comprehensive statistical validation. Thirdly, the subjective interpretation of prompts by different AI models, Image Creator and DALL-E, underscores the challenge in achieving standardized results, thus limiting the generalizability of our conclusions beyond the specific models tested. Finally, the black-box nature of these AI technologies limits our understanding of how input prompts are interpreted and processed, potentially affecting the consistency and reliability of our findings. These factors call for cautious interpretation of our results and suggest that further research, involving a broader array of models and more extensive iterations, is essential in reducing biases in AI-generated imagery.

### 3.2 Final Takeaway

Based on the results of our experimentation, the biggest takeaway is that by default, image generators struggle to produce inclusive and accurate representation across all the identities analyzed. Notably, we found that Image Creator and DALL-E has a tendency to initially uphold traditional gender stereotypes for professional roles. For disability, the model still seems to have an inability to understand the nuance of disabled experience. It often falls back on tropes (either those created by popular culture or based on its limited understanding of language) and has to be explicitly asked to provide a more diverse range of representation. In terms of age, we observed that Image Creator displayed age diversity from the outset, while DALL-E predominantly generated images of younger individuals until we explicitly incorporated terms like "cross-generational" or "age-diverse" into the prompts. These findings suggest that specific language can significantly mitigate age biases in image generators, leading to more diverse and accurate outcomes. In the study of race, DALL-E consistently showed bias, predominantly depicting white men in white-collar roles, while Image Creator adjusted to iterations, achieving a balanced representation of gender and race. For blue-collar jobs, DALL-E showed limited diversity, but Image Creator improved over iterations to avoid stereotyping blue-collar workers as men of color.

Across all categories, a notable finding was the models’ struggle with intersectional representation of primary subjects. Both defaulted to primarily depicting white, male primary subjects in the majority of the initial prompts. In general, there was little consideration for inclusivity across the different categories; for example, both models did not seem to prioritize racial and gender diversity when disability status was the primary identity being featured. Interestingly, as we generated the improved prompts, Image Creator began to provide more diverse representation in primary subjects across both

racial and gender identity. However, DALL-E consistently generated images of white and primarily male users. The analysis underscores a critical need for continuous evaluation and adjustment of AI models to better reflect the diversity of professional identities in real-world contexts, ensuring that these technologies promote inclusivity rather than perpetuate existing disparities.

## 4 Deliverables

In addition to this paper, our deliverables include:

- A dataset of 640 images (2 models, 4 categories, 2 prompts per category, 4 iterations per prompt, 10 images per iteration per model)
- A detailed breakdown of the prompts per category and a list of the iterations to eliminate bias
- A list of best practices that describes to users consistent actions they should take when developing prompts to generate images with more inclusive output

The deliverables are stored in this public facing GitHub repository:

[Inclusive-AI-Image-Generation](#)



## 5 Self-Evaluation

### 5.1 Jacqueline Gibson

**Accomplishments and learnings:** This project represented my first time exploring image generators. In previous courses, I studied bias in media representation and read a few papers about the experiences of disabled communities in online spaces, but this was the first opportunity I had to actually experiment with AI models in a class. I had the opportunity to work with Image Creator and learn more about how DALL-E works, which broadened my knowledge and skills in this area. Furthermore, I explored a new subtopic of Human-Computer Interaction (HCI) that I had not previously explored in research: digital representation and inclusivity. This period has been a journey of discovery. I learned that the issue of inclusivity in image generation extends beyond the commonly discussed biases in media, such as race and gender. There are significant opportunities for growth and improvement in areas related to age and disability status. I also found that current models struggle with intersectionality, highlighting the need for a more nuanced and comprehensive approach. Moving forward, I am excited to explore an intersectional approach to HCI, as I believe it represents another component of accessibility and holds the key to creating more inclusive and representative digital experiences.

**Responsibilities:** I worked with Chia-Mei and Nicole on analyzing bias based on race and disability status. I developed the rough drafts for the disability prompts, and through this process did the initial work to identify the four iteration progression we used across all prompting. Since I did not have an OpenAI subscription, I focused on generating the 80 images for Image Creator, while Chia-Mei handled the DALL-E equivalent. Additionally, I did the initial image migration and repository organization for the repo; this included defining the folder substructure for the data, creating a tracking list for repo additions, transcribing all prompts so that our work could be replicated, and moving over the images we had generated ahead of our final demo last week. I authored the sections on disability status, contributed to the Best Practices document, and contributed to the organization and edits of this paper.

**Challenges:** My most significant challenge this semester was balancing the work of this class with my other obligations. Each week I had to make sure I completed the readings for this class, meet milestones for the midterm or final project, and also complete homework for my other course (which featured bi-weekly problem sets). Additionally, I had to balance my work responsibilities, which included having to lead our team efforts in triaging weekly security issues. This semester was a demanding period, but it also taught me valuable lessons about time management and prioritization.

### 5.2 Chia-Mei Liu

This project marked my initial effort into AI-related research, specifically focusing on image generation and inclusivity, which is a subject I had seldom encountered before. Prior to this research, my understanding of AI image generation biases was primarily formed by browsing related news articles, which highlighted how these biases demonstrated broader real-world issues. My engagement with relevant research papers throughout this project not only deepened my comprehension but also significantly broadened my perspective on the complexities of AI technologies.

Throughout this semester, I worked closely with Jacqueline and Nicole on the Disability and Race components of our study. As the only member of our subgroup with

access to OpenAI’s ChatGPT 4.0, my primary responsibility was to generate images using DALL-E for these two subtopics. This task was particularly challenging as I needed to carefully evaluate each image’s suitability for our academic research, discern and mitigated my inherent biases—a process that sometimes challenged my own perceptions and experience. Working with Jacqueline, I contributed to crafting the Conclusion section of our paper. Furthermore, I was in charge of refining and submitting our research topic proposal and the progress report, after gathering input from each team member. Besides, I drafted the README for our github repository.

One of the key growth areas during this project was working within a team of native English speakers and leading some discussion calls, which was a first-time experience for me. This experience pushed me to refine my communication skills and adapt to different cultural perspectives, which was both challenging and enriching. It significantly enhanced my ability to navigate complex discussions and fostered a deeper understanding of inclusive practices in AI technologies.

### 5.3 Anavi Lohia

In conducting this research, I achieved a thorough examination of gender and age related biases in AI image-generation models. My research unveiled nuanced portrayals of gender disparities within professional settings. I delved into the responsiveness of these models to gender-inclusive prompts, revealing variations in sensitivity to such instructions and lack of representation for gender-nonconforming individuals. My work also identified age and ability related bias, specifically for elderly women, and the level of granularity required in prompts to ensure diverse and accurate representations.

I found it challenging to iterate the prompts to truly ensure a diverse representation across both models with the same prompts. I learned that while specificity in the prompt was necessary to generate diverse images, this practice merely transfers the onus of the eliminating bias from the model to the prompter. Relatedly, another challenge I encountered was tagging images based on my interpretation of the genders/ages depicted in the images. I found myself questioning my own inherent biases when looking at an image of a person from the back and deciding that it depicted a necessarily male or female figure. Further, I recognized the voluminous amount of testing and modifications that would be required to create an image generation model that does not reinforce societal gender and age biases. In my research I also found that Microsoft’s platforms often use OpenAI’s technologies in the backend but the Image Creator model is able to significantly reduce biases that are seen in the DALL-E model. This was my biggest takeaway from this exercise as it demonstrated the possibility of building an inclusive image generation model while using existing technology.

My specific contributions to this project were the gender and age sections. I formulated all prompts for the ‘gender and profession’ and ‘age and ability’ sections. I then conducted the image generation and analysis using DALL-E and Image Creator for these sections. I also contributed with initial prompt ideation for the ‘gender and family roles’ and ‘age and family’ sections. I reviewed past research for gender and race related biases. Additionally, I worked on the research methodology section, the references section, and the best practices document in the GitHub repository. I further contributed through ideation and edits across all sections of this paper.

## 5.4 Harpreet Multani

This research project provided me the opportunity to delve deeper into the topic of AI Image generation, which is a field I had been quite interested in for some time now. This project gave me the opportunity to really go into detail and explore different image generation, including some that I had never used before. It was a challenging initially to come up with prompts and coax the generators to create more diverse and representative images. After a long process of a lot of trial and error, I was able to develop prompts that helped create more inclusive image results. I'm really grateful for the experience to work on this project and learn so much and had such a great and supportive team alongside me throughout it all.

For this project, I focused on the gender and age bias categories. Specifically, the 'gender and family roles' and 'age and family' subsections. My contributions included developing various themes for exploration, creating prompts for the different iterations, and generating images using both DALL-E and Image Creator. Additionally, I wrote the synopsis overview, reviewed prior works for age related biases, my subsections under gender and age, and helped create/organize the overall structure for the paper.

## 5.5 Nicole Neil

This project serves as a sort of culmination of the research I conducted in the COMS 6156 course. I focused on the representation of different groups in different scenarios, such as old women's roles in software development, gender bias and AI ethical thinking in ChatGPT, options for disabled avatars in digital platforms, and now race in profession as shown by AI image generators. I had the opportunity to work with Microsoft's Image Creator and further my knowledge of the biases present in AI generation models. Unlike in my previous project about bias and ethics in AI, I was able to discover best practices for mitigating bias in the Image Creator, but not always for DALL-E. I am excited to continue research of bias and ethics in AI development, representation, and use.

I worked with Chia-Mei and Jacqueline on analyzing bias based on race and disability status. I developed the rough drafts for the race prompts, and with their help was able to identify the four iterations to use for the prompting. This was my most significant challenge during this project. It was very difficult to think of prompts that when tested on Image Creator would provide the desired results. Since I did not have an OpenAI subscription, I focused on generating the 80 images for Image Creator, while Chia-Mei handled the DALL-E equivalent. I also did the last steps of the image migration, included standardizing the naming convention for all the folders and images. I also completed the Innovation and Value section of the paper and assisted with the organization of the overall paper structure.

## References

- [Chauhan et al., 2024] Chauhan, A., Anand, T., Jauhari, T., Shah, A., Singh, R., Rajaram, A., and Vanga, R. (2024). Identifying Race and Gender Bias in Stable Diffusion AI Image Generation. In *2024 IEEE 3rd International Conference on AI in Cybersecurity (ICAIC)*, pages 1–6. IEEE.
- [Chu et al., 2023] Chu, C. H., Donato-Woodger, S., Khan, S. S., Nyrup, R., Leslie, K., Lyn, A., Shi, T., Bianchi, A., Rahimi, S. A., and Grenier, A. (2023). Age-related bias and artificial intelligence: a scoping review. *Humanities and Social Sciences Communications*, 10(1):510.
- [Gualano et al., 2023] Gualano, R. J., Jiang, L., Zhang, K., Won, A. S., and Azenkot, S. (2023). “Invisible Illness Is No Longer Invisible”: Making Social VR Avatars More Inclusive for Invisible Disability Representation. In *The 25th International ACM SIGACCESS Conference on Computers and Accessibility*, pages 1–4, New York, NY, USA. ACM.
- [IAAP, 2020] IAAP (2020). Body of Knowledge. Technical report, International Association of Accessibility Professionals.
- [Mack et al., 2023] Mack, K., Hsu, R. C. L., Monroy-Hernández, A., Smith, B. A., and Liu, F. (2023). Towards Inclusive Avatars: Disability Representation in Avatar Platforms. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, pages 1–13, New York, NY, USA. ACM.
- [Pherkan, 2023] Pherkan (2023). Have you ever wondered how AI, particularly Midjourney, perceives and represents people who are blind or have low vision? *Medium*.
- [Rodman, 2023] Rodman, L. (2023). Why AI-Generated Art Is Missing the Mark for People With Disabilities. *Association for Talent Development*.

## 6 Appendix

### 6.1 Appendix A: Selected chat output from DALL-E for disability representation

#### 6.1.1 Vision Prompt Sample

**Exact prompt:** please generate a picture about ‘Blind or low vision person completing errands’

**Response**

Here is the image of a blind or low vision person completing errands in a bustling urban environment. The illustration shows them using a white mobility cane and navigating confidently amidst the lively city atmosphere.

#### 6.1.2 Mobility Prompt Sample

**Exact prompt:** please generate a picture about ‘a person using a cane, crutches, or any other type of walking aid to navigate a coffee shop’

**Response**

Here is the image of a person using a walking aid, such as a cane or crutches, to navigate a coffee shop. The scene captures the individual moving through a coffee shop with a spacious and stylish interior, designed for accessibility. The coffee shop features wide aisles and a comfortable layout, with other patrons enjoying their time in a relaxed and welcoming environment.