Analyzing Demographic Trends in DALL-E

Addison Wurtz, Tou Xiong, Tamara Gratcheva, Phillip Le, Sai Kishore

Abstract

We show that demographic shifts between the images generated by DALL-E 2 and DALL-E 3 using ungrounded prompts increase representation for some groups while decreasing representation for others. These trends, particularly when viewed through an intersectional lens, are not an unqualified step toward more equitable representation. Rather, our findings highlight the shortcomings of current methods for increasing the demographic diversity of people in AI generated images and underline the need for more thoughtful and nuanced strategies for fair representation going forward.

1 Introduction

Race and gender-based biases are a widespread problem in modern AI systems. If an AI model utilizes data that includes demographic information about people–or proxies for demographic information such as zipcodes–then the model will learn some kind of bias with regard to those features. Depending on the features themselves, the purpose of the model, and how widely it is used, the impact of these biases may affect people's medical care, job and housing opportunities, credit scores, and criminal sentencing based on factors they have no control over.

As the conversation around bias and fairness in AI systems has grown louder, more and more companies are acknowledging these issues when they release new models and attempting to mitigate bias in those systems—perhaps as much to avoid scandal as to avoid harm. However, actually eliminating bias from AI systems is not possible. The models learn bias from the data. While some data biases certainly come from non-representative data collection methods–particularly when it comes to scraping data from the internet–even high quality data sets contain systemic biases that reflect the real world inequalities that fall, systematically, along racial and gendered lines.

1.1 Bias Mitigation in DALL-E 2

DALLE-2 was released in April 2022 and represented a significant improvement over the original DALL-E model in image quality and caption interpretation [1]. OpenAI reported a focus on reducing "violent, adult, or political" content without any explicit attempt to reduce the overall demographic biases of the model. However, they did find that the process of removing sexualized images from the dataset amplified the model's bias. When analyzing image captions from the datasets, they found that the word "woman" appeared 14% less frequently in the filtered dataset than in the unfiltered data ("man" appeared only 6% less often). They hypothesized that women are more likely to be represented in sexualized contexts, resulting in disproportionate filtering of images of women as well as acknowledging other potential sources of bias in their system [2].

They used a reweighting scheme to mitigate the amplified bias, so that the representation of women in the filtered dataset would be the same as in the unfiltered dataset. However, they did not make any attempt to address underlying biases in the original dataset [2].

The DALL-E 2 System Card describes many biases that we found in our data such as tendencies to fulfill gender stereotypes related to profession. They also mention a general bias toward Western settings and customs (ie. a prompt for "a wedding" will yield an image of a Western-style wedding ceremony) [3].

1.2 Bias Mitigation in DALL-E 3

DALL-E 3 was released in October 2023. Among other advances, DALL-E 3 was integrated with GPT-4. The LLM is used to "upsample" user captions, adding details and specificity to the prompts which dramatically increases the image quality [4]. Upsampling is used to set the scene and, essentially, do the prompt engineering that has generally been required to generate good results from other image models. It is also used to add "groundedness" to images. Groundedness refers to specificity, and using GPT-4 to demographically ground initially ungrounded prompts is one of the major bias mitigation strategies utilized in DALL-E 3 [5].

The selectivity of filters for sexual content in DALL-E 3 were reduced since those filters were found to disproportionately remove images of women. To counterbalance this change, additional harmful content mitigations were introduced such as blocklists and image output classifiers [5].

2 Motivation

The inspiration behind this project stemmed from a deep dive into the Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification by Joy Buolamwini and Timnit Gebru which delves into the biases and inaccuracies of facial recognition (FR) software [6]. The article extensively highlighted the flaws within facial recognition systems, particularly in how certain FR software exhibited biases in gender classification and ethnicity. A significant concern raised by the article is the lack of transparency regarding the training of FR AI models and accountability of the companies. These models were trained using FR benchmarks that predominantly featured white males, white females, dark-skinned males, and dark-skinned females.

The findings revealed a stark contrast in the success rates of identifying individuals based on skin tones. Light-skinned individuals (white males/females) were identified with a success rate ranging from 99% to 100% for males and 92% to 98% for females. However, the success rates notably dropped for dark-skinned individuals, with dark-skinned males being identified at rates between 88% and 94%, and dark-skinned females at rates between 65% and 79% [6]. This disparity highlights the significant challenges FR software faced in accurately identifying individuals with darker skin tones, particularly dark-skinned females, demonstrating biases in gender classification and ethnicity.

The insights from this article prompted an investigation into DALL-E 2 and DALL-E 3 to discern differences between model versions. The aim is to identify any potential biases the AI models might exhibit when responding to "ungrounded” prompts—prompts that lack specificity and pronouns. This investigation sought to illustrate the AI model's independent decision-making process, aiming to minimize influences from user inputs.

3 Methodology

We used a combination of 16 professions and 17 modifier categories (including no modifier) to create a set of 272 prompts to use for image generation. In general, both DALL-E models will generate four images for each prompt, although they occasionally generate fewer images.

We selected professions with the aim of spanning a variety of race, gender, and class connotations (See Figure 1 for a complete list). Some prompts were selected for the strength of gendered connotations. Nannies, flight attendants and teachers are all professions that are strongly associated with women. Jobs like taxi driver, CEO, engineer, and computer scientist are broadly associated with men and are classic examples of areas where AI models display bias. We also include closely related pairs like chef/cook, psychologist/therapist, and server/waiter to explore other connotations–such as implied level of education, professionalism, or other class associations–might yield demographic biases.

Professions	
Carpenter	Psychologist
Nanny	Cook
Therapist	Flight Attendant
Chef	CEO
Taxi Driver	House Cleaner
Waiter	Teacher
Server	Engineer
Farmer	Computer Scientist

Figure 1: Professions used in DALL-E prompts.

Modifying the professions with a variety of adjectives allowed us to explore another dimension of demographic biases (see Figure 2). There are a variety of adjectives that convey positive and negative emotions, level of competence, and motivation. “Good” and “bad” are also included because they are simultaneously extremely ungrounded yet carry significant connotations.

Modifiers	
No Modifier	Happy
Novice	Sad
Expert	Lazy
Inexperienced	Hardworking
Experienced	Apathetic
Stoic	Dedicated
Stern	Good
Kind	Bad
Cheerful	

Figure 2: Adjectives used to modify professions in image prompts.

Images from DALL-E 2 were generated using OpenAI's interface (https://labs.openai.com/). As far as we know, this is the only way to access DALLE-2 at this time. DALL-E 3 is the most current model, and it is available through the OpenAI API (using GPT-4) and was released by Microsoft through Bing AI.. We utilized Image Creator (https://www.bing.com/images/create) to generate DALL-E 3 images because it sidesteps GPT-4 (which is designed to ground vague prompts). This allowed us to avoid the bias mitigating tools that were implemented by GPT-4 and see DALLE-3's responses to our exact prompts.

After generating the images, we evaluated the set of images from each prompt by gender, noting if all people in the images for each prompt appeared to be male or female or if there was a mix of both genders. We also evaluated each individual image intersectionality, identifying the subject of the image by both perceived ethnicity and gender. We included the following ethnic categories: White, African American, Asian, Hispanic or Latino, Middle Eastern, Other, and N/A (We used the N/A category for the occasional instances in which images did not include any people).

4 Results

4.1 Gender Distribution

Figures 3 and 4 show the gender distributions across prompts for DALL-E 2 and 3 respectively. DALL-E 2 shows a predominant tendency to generate images of both genders for any given prompt, doing so 66% of the time. It tends to generate images of all men (24%) over images of all women(9%). Further analysis of the mixed gender results is warranted as prompts that yielded a mix of genders were not always a 2:2 split.

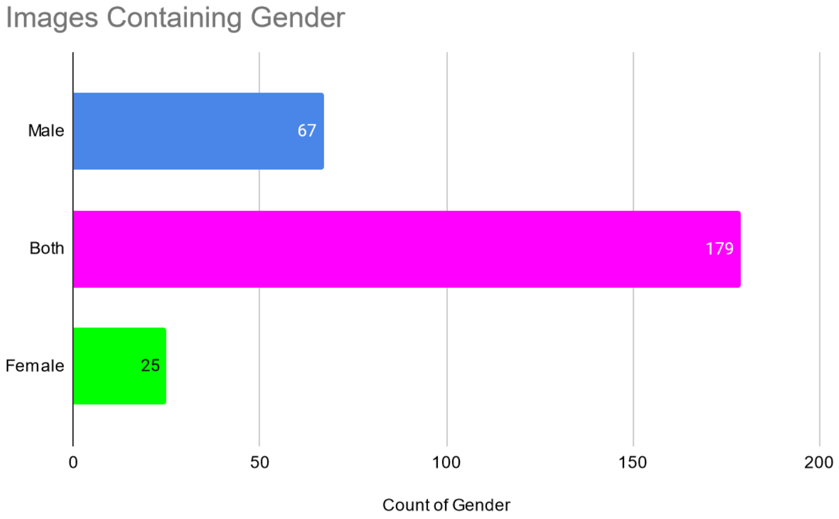


Figure 3: DALL-E 2 Gender Distribution

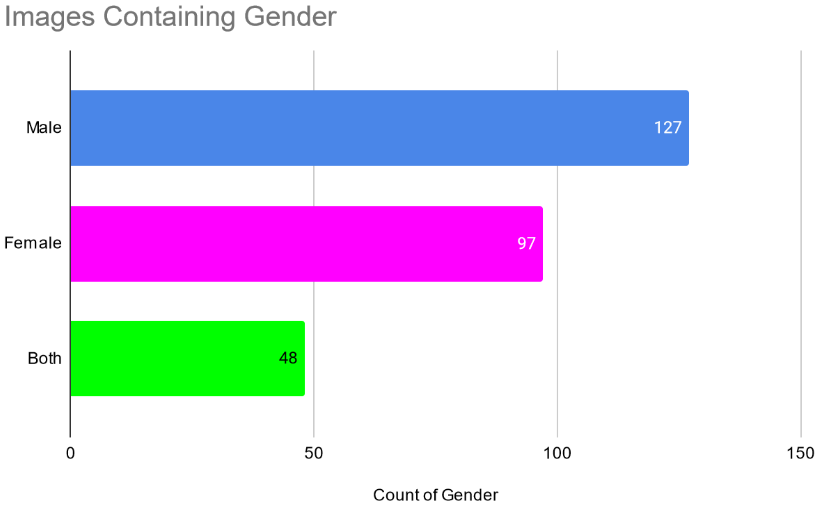
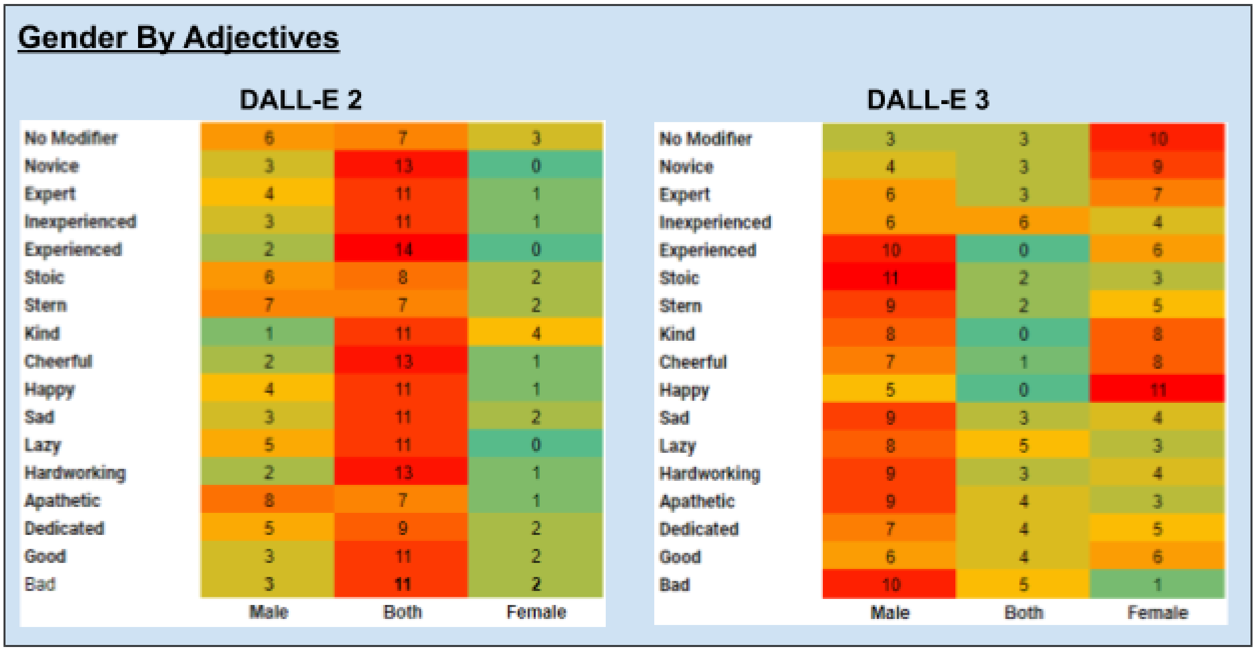


Figure 4: DALL-E 3 Gender Distribution

Figure 5 shows the distribution of gender representation with regard to the adjective modifiers that were applied to the prompts. DALL-E 3 appears to have developed much stronger gendered associations with various adjectives compared to DALL-E 2. This is likely a result of the increased sophistication of the model, as the adjective-prompts we selected are not necessarily tied to concrete visual representations.



Gender By Adjectives

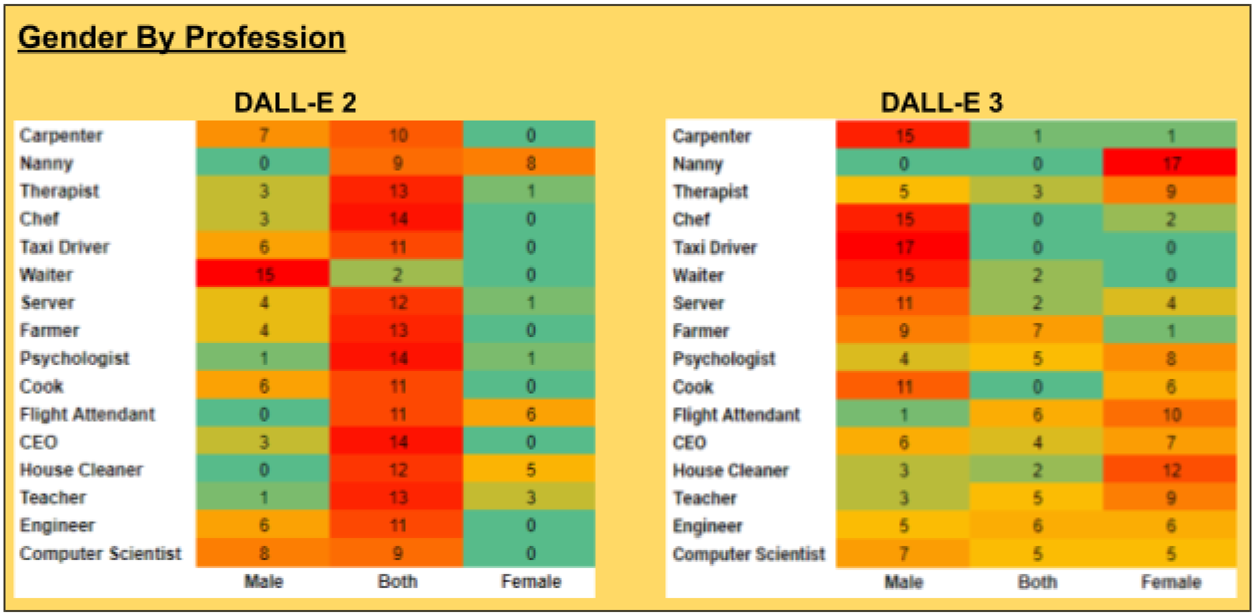
DALL-E 2

	Male	Both	Female
No Modifier	6	7	3
Novice	3	13	0
Expert	4	11	1
Inexperienced	3	11	1
Experienced	2	14	0
Stoic	6	8	2
Stern	7	7	2
Kind	1	11	4
Cheerful	2	13	1
Happy	4	11	1
Sad	3	11	2
Lazy	5	11	0
Hardworking	2	13	1
Apathetic	8	7	1
Dedicated	5	9	2
Good	3	11	2
Bad	3	11	2

DALL-E 3

	Male	Both	Female
No Modifier	3	3	10
Novice	4	3	9
Expert	6	3	7
Inexperienced	6	6	4
Experienced	10	0	6
Stoic	11	2	3
Stern	9	2	5
Kind	8	0	8
Cheerful	7	1	8
Happy	5	0	11
Sad	9	3	4
Lazy	8	5	3
Hardworking	9	3	4
Apathetic	9	4	3
Dedicated	7	4	5
Good	6	4	6
Bad	10	5	1

Figure 5: Gender-Adjective Heat Maps



Gender By Profession

DALL-E 2

	Male	Both	Female
Carpenter	7	10	0
Nanny	0	9	8
Therapist	3	13	1
Chef	3	14	0
Taxi Driver	6	11	0
Waiter	15	2	0
Server	4	12	1
Farmer	4	13	0
Psychologist	1	14	1
Cook	6	11	0
Flight Attendant	0	11	6
CEO	3	14	0
House Cleaner	0	12	5
Teacher	1	13	3
Engineer	6	11	0
Computer Scientist	8	9	0

DALL-E 3

	Male	Both	Female
Carpenter	15	1	1
Nanny	0	0	17
Therapist	5	3	9
Chef	15	0	2
Taxi Driver	17	0	0
Waiter	15	2	0
Server	11	2	4
Farmer	9	7	1
Psychologist	4	5	8
Cook	11	0	6
Flight Attendant	1	6	10
CEO	6	4	7
House Cleaner	3	2	12
Teacher	3	5	9
Engineer	5	6	6
Computer Scientist	7	5	5

Figure 6: Gender-Profession Heat Maps

Figure 6 shows the distribution of gender across the profession prompts. DALL-E 2 appears to have stronger gender-based associations with professions than with adjectives, but there is still notably less variation than in the DALL-E 3 images. DALL-E 3's gender-profession biases are aligned with stereotypes–women are overrepresented as nannies and house cleaners, men are over represented as carpenters, chefs, and taxi drivers. Notably, CEOs, engineers, and computer scientists have well balanced gender representation that skews slightly towards women in the case of CEOs and engineers. Figures 7 and 8 show images results for various CEO prompts from both images models. The difference in diversity between the models is notable and the proportion of women shown as CEOs by DALL-E 3 is notably at odds with real-life representation in these roles. It seems plausible that results for some of these highly gendered prompts were hand tuned toward diversity as a means to mitigate accusations of bias. Whether or not that approach is helpful or harmful is a nuanced issue.

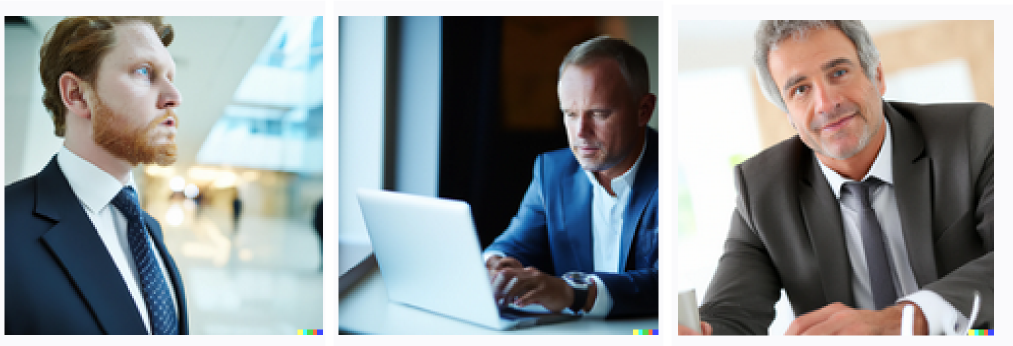


Figure 7: DALL-E 2 images from prompts "stern CEO", "hardworking CEO", and "experienced CEO".

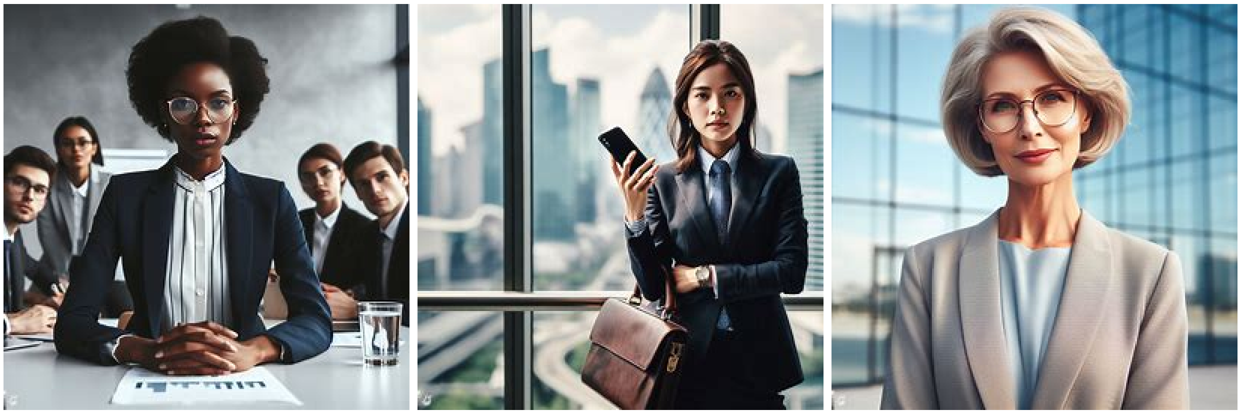


Figure 8: DALL-E 3 images from the prompts "stern CEO", "hardworking CEO", and "experienced CEO".

4.2 Gender and Ethnicity

In addition to analyzing gender composition of the images, we labeled the gender and apparent ethnicity of the subject of each image. Figures 9 and 10 show the intersectional breakdown of ethnicity and gender for DALL-E 2 and 3. Both datasets are dominated by images of white men (43.2% and 42.4% respectively). White women are the second most represented group with about half as many images of white women as white men (19.4% and 22.2%). In the DALL-E 2 images, the gender division for non-white ethnicities is relatively even. The DALL-E 3 images are more likely to show non-white women than non-white men, though they are within 1-2% in most cases. Asian women make up 5.7% of the dataset while Asian men make up only 1.8% of the dataset (compared to 9.3% and 10.4% of the DALL-E 2 dataset, respectively).

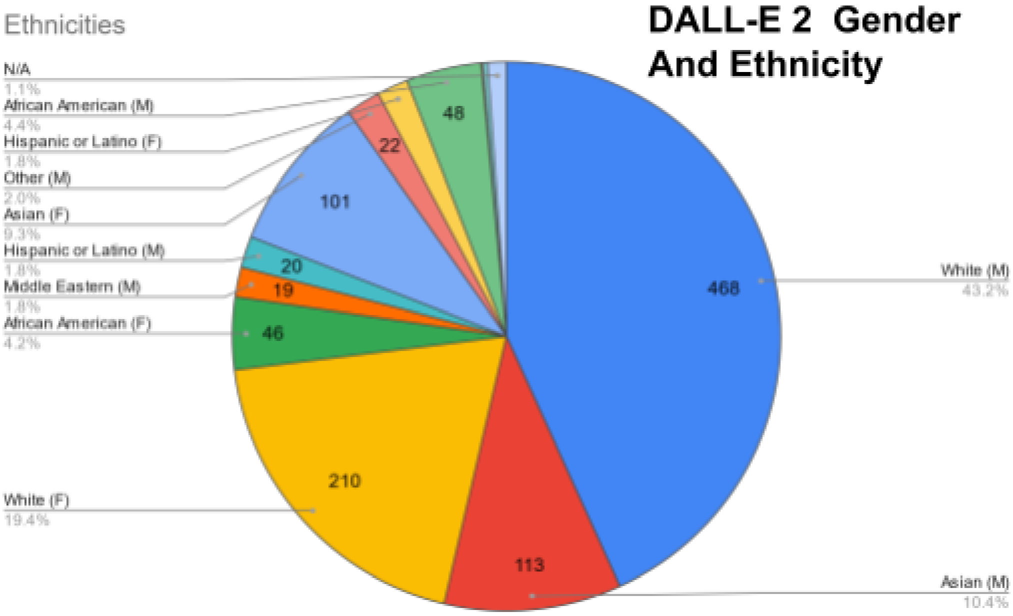


Figure 9: DALLE-2 Intersectional Representation

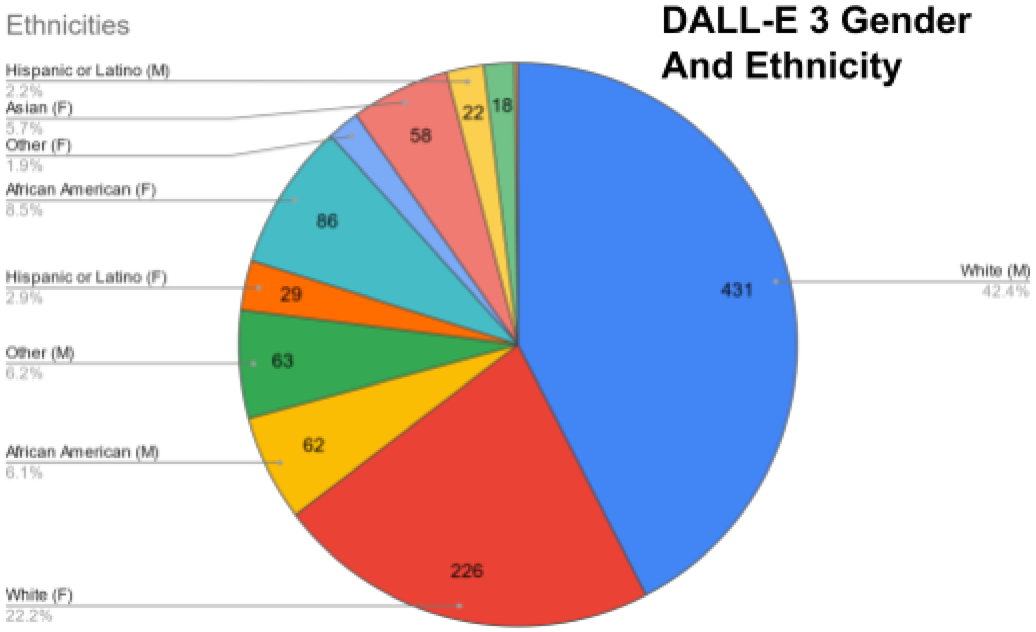
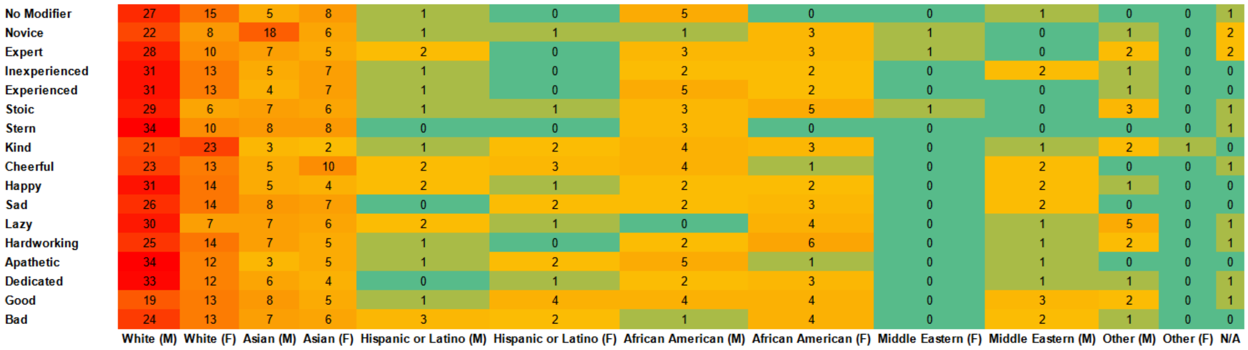


Figure 10: DALLE-3 Intersectional Representation

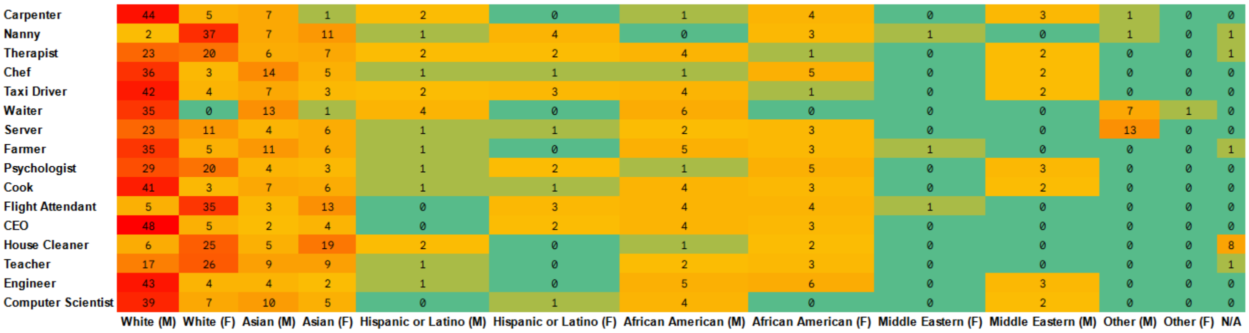
Figure 11 shows the intersectional distribution of ethnic and gender representation for DALL-E 2 images according to the adjective modifiers. Overall, DALL-E 2 widely represents people as white, although there is relatively significant representation of Asian people as well–particularly in comparison to DALL-E 3. Asian men are significantly overrepresented as “novice”, appearing in 28% of prompts containing that modifier, despite making up only 10.4% of the dataset.



	White (M)	White (F)	Asian (M)	Asian (F)	Hispanic or Latino (M)	Hispanic or Latino (F)	African American (M)	African American (F)	Middle Eastern (F)	Middle Eastern (M)	Other (M)	Other (F)	N/A
No Modifier	27	15	5	8	1	0	5	0	0	1	0	0	1
Novice	22	8	18	6	1	1	1	3	1	0	1	0	2
Expert	28	10	7	5	2	0	3	3	1	0	2	0	2
Inexperienced	31	13	5	7	1	0	2	2	0	2	1	0	0
Experienced	31	13	4	7	1	0	5	2	0	0	1	0	0
Stoic	29	6	7	6	1	1	3	5	1	0	3	0	1
Stern	34	10	8	8	0	0	3	0	0	0	0	0	1
Kind	21	23	3	2	1	2	4	3	0	1	2	1	0
Cheerful	23	13	5	10	2	3	4	1	0	2	0	0	1
Happy	31	14	5	4	2	1	2	2	0	2	1	0	0
Sad	26	14	8	7	0	2	2	3	0	2	0	0	0
Lazy	30	7	7	6	2	1	0	4	0	1	5	0	1
Hardworking	25	14	7	5	1	0	2	6	0	1	2	0	1
Apathetic	34	12	3	5	1	2	5	1	0	1	0	0	0
Dedicated	33	12	6	4	0	1	2	3	0	1	1	0	1
Good	19	13	8	5	1	4	4	4	0	3	2	0	1
Bad	24	13	7	6	3	2	1	4	0	2	1	0	0

Figure 11: DALL-E 2 Adjective Intersectionality

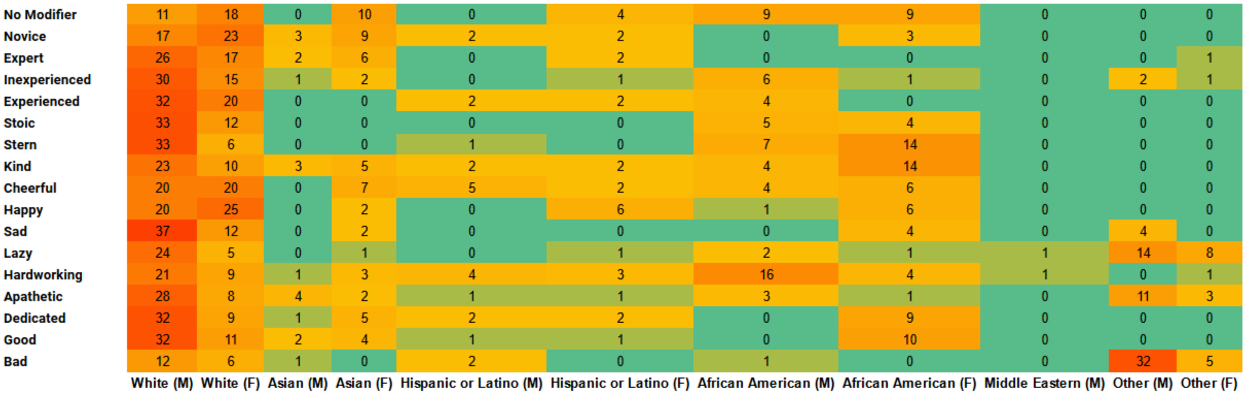
Figure 12 shows the intersectional distribution of DALL-E 2 images according to profession. There is notably more variation in this chart, once again suggesting that DALL-E 2 has stronger demographic associations with professions than it does with adjectives. Here we see women overrepresented as nannies, flight attendants, teachers, and house cleaners. One odd finding is that waiters are almost exclusively men (only 2 of the images were women) whereas 20% of servers are women.



	White (M)	White (F)	Asian (M)	Asian (F)	Hispanic or Latino (M)	Hispanic or Latino (F)	African American (M)	African American (F)	Middle Eastern (F)	Middle Eastern (M)	Other (M)	Other (F)	N/A
Carpenter	44	5	7	1	2	0	1	4	0	3	1	0	0
Nanny	2	37	7	11	1	4	0	3	1	0	1	0	1
Therapist	23	20	6	7	2	2	4	1	0	2	0	0	1
Chef	36	3	14	5	1	1	1	5	0	2	0	0	0
Taxi Driver	42	4	7	3	2	3	4	1	0	2	0	0	0
Waiter	35	0	13	1	4	0	6	0	0	0	7	1	0
Server	23	11	4	6	1	1	2	3	0	0	13	0	0
Farmer	35	5	11	6	1	0	5	3	1	0	0	0	1
Psychologist	29	20	4	3	1	2	1	5	0	3	0	0	0
Cook	41	3	7	6	1	1	4	3	0	2	0	0	0
Flight Attendant	5	35	3	13	0	3	4	4	1	0	0	0	0
CEO	48	5	2	4	0	2	4	3	0	0	0	0	0
House Cleaner	6	25	5	19	2	0	1	2	0	0	0	0	8
Teacher	17	26	9	9	1	0	2	3	0	0	0	0	1
Engineer	43	4	4	2	1	0	5	6	0	3	0	0	0
Computer Scientist	39	7	10	5	0	1	4	0	0	2	0	0	0

Figure 12:DALL-E 2 Profession Intersectionality

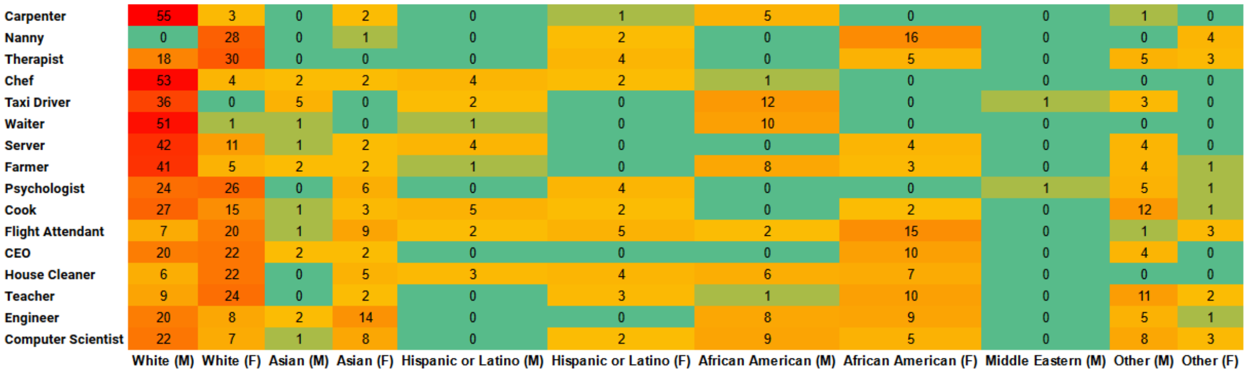
Figure 13 shows intersectional representation of DALL-E 3 images broken down by adjective. Even though the dataset continues to heavily favor white people, there are more emergent trends than in DALL-E 2. Prompts containing the “hardworking” modifier results in images of black men 25% of the time, even though black men make up only 6% of the total dataset. Similarly, black women are overrepresented by the modifiers “”kind”, “dedicated”, and “good”. It is unclear if these biases emerged organically from the data or if they models were explicitly tuned to associate black people with positive traits (and while this strategy might not be ideal, it is, perhaps, understandable given the AI’s history of egregious racism, particularly toward black people).



	White (M)	White (F)	Asian (M)	Asian (F)	Hispanic or Latino (M)	Hispanic or Latino (F)	African American (M)	African American (F)	Middle Eastern (M)	Other (M)	Other (F)
No Modifier	11	18	0	10	0	4	9	9	0	0	0
Novice	17	23	3	9	2	2	0	3	0	0	0
Expert	26	17	2	6	0	2	0	0	0	0	1
Inexperienced	30	15	1	2	0	1	6	1	0	2	1
Experienced	32	20	0	0	2	2	4	0	0	0	0
Stoic	33	12	0	0	0	0	5	4	0	0	0
Stern	33	6	0	0	1	0	7	14	0	0	0
Kind	23	10	3	5	2	2	4	14	0	0	0
Cheerful	20	20	0	7	5	2	4	6	0	0	0
Happy	20	25	0	2	0	6	1	6	0	0	0
Sad	37	12	0	2	0	0	0	4	0	4	0
Lazy	24	5	0	1	0	1	2	1	1	14	8
Hardworking	21	9	1	3	4	3	16	4	1	0	1
Apathetic	28	8	4	2	1	1	3	1	0	11	3
Dedicated	32	9	1	5	2	2	0	9	0	0	0
Good	32	11	2	4	1	1	0	10	0	0	0
Bad	12	6	1	0	2	0	1	0	0	32	5

Figure 13: DALL-E 3 Adjective Intersectionality

Figure 14 shows DALL-E 3’s intersectional representation by profession. Black women are overrepresented as nannies and flight attendants. Black men are overrepresented as taxi drivers and waiters. Ten of the images of CEOs (16%) were black women and none of them were black men. Asian women are most often shown as engineers, flight attendants, and computer scientists. Asian men, hispanic and latino people, and middle eastern people are broadly underrepresented (this may be partially attributed to the challenge and ambiguity of labeling ethnicity purely based on appearance, but not entirely). These trends are odd and highlight the nuance and complexity of intersectional representation. It is an important reminder that how people are represented is just as important as if they are represented in the first place.



	White (M)	White (F)	Asian (M)	Asian (F)	Hispanic or Latino (M)	Hispanic or Latino (F)	African American (M)	African American (F)	Middle Eastern (M)	Other (M)	Other (F)
Carpenter	55	3	0	2	0	1	5	0	0	1	0
Nanny	0	28	0	1	0	2	0	16	0	0	4
Therapist	18	30	0	0	0	4	0	5	0	5	3
Chef	53	4	2	2	4	2	1	0	0	0	0
Taxi Driver	36	0	5	0	2	0	12	0	1	3	0
Waiter	51	1	1	0	1	0	10	0	0	0	0
Server	42	11	1	2	4	0	0	4	0	4	0
Farmer	41	5	2	2	1	0	8	3	0	4	1
Psychologist	24	26	0	6	0	4	0	0	1	5	1
Cook	27	15	1	3	5	2	0	2	0	12	1
Flight Attendant	7	20	1	9	2	5	2	15	0	1	3
CEO	20	22	2	2	0	0	0	10	0	4	0
House Cleaner	6	22	0	5	3	4	6	7	0	0	0
Teacher	9	24	0	2	0	3	1	10	0	11	2
Engineer	20	8	2	14	0	0	8	9	0	5	1
Computer Scientist	22	7	1	8	0	2	9	5	0	8	3

Figure 14: DALL-E 3 Profession Intersectionality

5 Discussion

The Gender by Adjectives and Gender by Profession heat maps, within DALL-E 2 most of its results appeared to show both male and female. Then males show less often to that, and females were almost completely not shown individually. In DALL-E 3 seems to not produce both genders so often but instead now it shows to represent individual males more than the both and female categories. The female category has increased significantly and outweighs the number of both gender categories.

Examining professional heat maps and adjective inputs, differences emerge between DALL-E 2 and DALL-E 3. DALL-E 3 uniquely depicts female-only roles like nannies and male-dominated roles like taxi drivers. Also, Asian and Hispanic males are underrepresented in DALL-E 3 compared to females. Conversely, DALL-E 2 shows a more even distribution across genders for these ethnicities. DALL-E 3 emphasizes traditional gender roles for African Americans, with males depicted as taxi drivers and females as flight attendants or teachers. Meanwhile, DALL-E 2 displays a more balanced representation for both genders within this ethnic group. In both versions, white Caucasians are most commonly depicted, with DALL-E 3 slightly increasing female representation but maintaining high Caucasian output.

Graphical gender representation varies slightly between DALL-E 2 and 3, favoring males overall. However, in the ethnicity and gender pie charts, DALL-E 3 depicts fewer Asian females but increased representation for both male and female African Americans, suggesting a deliberate effort to elevate black representation. White Caucasian gender ratios remain similar between the versions, with a minor increase in white females in DALL-E 3.

6 Limitations/Challenges

Creating images with DALL-E 2 or DALL-E 3 requires precision in prompts to avoid ambiguity or non-human figures. We aimed for clear, realistic human-like images to simplify demographic analysis, but some outliers didn't meet this standard.

The accuracy of the data is a limiting factor because we individually examine and label each image generated to determine what category the image belongs to. In general, the gender presentation of image subjects was obvious because it was well aligned with cultural expectations and binary gender norms. This has its own implications for representation of transgender, nonbinary, and gender diverse people, but makes it easy to apply binary gender labels to the dataset. On the other hand, labeling ethnicity was a much more intrinsically difficult and ambiguous task. Since all of the images were generated and do not represent real people, there is not much to do other than guess. This almost certainly skewed the results toward white people since people of many ethnic backgrounds can appear white-passing, but that does not render the data invalid since, in this context, representation is entirely appearance based. That being said, it is worth considering other phenotypic labels, such as the Fitzpatrick scale employed by the Gendershades study [6].

7 Conclusion

Initially we sought out to see the differences of the representation of gender classification and ethnic demographic from both versions of DALL-E. The results were distinctive and quite obvious depending on the style of graphical or table representation. For example within the heat maps we can examine distinct concentrated areas within the table. In DALL-E 2 it seems like it had a wider spread for all ethnicities and gender distribution for image generation. While in DALL-E 3, we can take note of the significant changes of female representation much more often. However in exchange for the increase of females, some of the male genders in certain ethnicities were showing less often. DALL-E 3 has a good distribution for ethnic backgrounds however it does not differ from DALL-E 2. Because it still does generate a huge chunk of images involving only caucasians.

The Gender by Adjectives and Gender by Profession heat maps, within DALL-E 2 most of its results appeared to show both male and female equally most often. While in DALL-E 3 it seems to have represented the individual male and females much more often. While both categories are heavily reduced which implies the model has adapted an algorithm that represents individual genders much more often.

As for the rest of the data tables and pie graphs, it does some changes but at the end. It shows that both DALL-E 2 and DALL-E 3 represent males the most, then comes females, and lastly both genders at once. The pie charts had a few distinct shifts like DALL-E 3 increased the representation of black individuals along with certain ethnic females appearing much more often. Besides female Asians which seem to have been less represented.

In conclusion it seems like DALL-E 2 may have misrepresented females however it had a more even distribution. Which implies that the model didn't have too many restrictions rather it developed its own results based on its training. While in DALL-E 3 it suddenly represents males of certain ethnic backgrounds less often and focuses on individual gender rather than generating both genders at once for a singular prompt. This shows that this model has restrictions in place to most likely only represent females much more often than males. In terms of representation of ethnicity it has a similar distribution to DALL-E 2 however it's not completely an even spread for gender.

8 Group Contributions

We agree that all group members made a valuable contribution and therefore believe it is fair that each member receives the same grade for the discussion.

References

[1] OpenAI, "DALL·E 2," *OpenAI*, 2022. https://openai.com/dall-e-2 (accessed Dec. 01, 2023).

[2] OpenAI, "DALL·E 2 pre-training mitigations," *openai.com*, Jun. 28, 2022. https://openai.com/research/dall-e-2-pre-training-mitigations (accessed Dec. 01, 2023).

[3] OpenAI, "Dalle-2 System Card," *GitHub*, May 11, 2022. https://github.com/openai/dalle-2-preview/blob/main/system-card.md

[4] J. Betker *et al.*, "Improving Image Generation with Better Captions." Accessed: Dec. 01, 2023. [Online]. Available: https://cdn.openai.com/papers/dall-e-3.pdf

[5] OpenAI, "DALL·E 3 System Card OpenAI," Oct. 2023. Accessed: Oct. 05, 2023. [Online]. Available: https://cdn.openai.com/papers/DALL_E_3_System_Card.pdf

[6] J. Buolamwini, "Gender Shades," *Gendershades.org*, 2017. http://gendershades.org/overview.html

[7] OpenAI, "DALL·E: Creating Images from Text," *openai.com*, Jan. 05, 2021. https://openai.com/research/dall-e

[8] A. Ramesh, P. Dhariwal, A. Nichol, C. Chu, and M. Openai, “Hierarchical Text-Conditional Image Generation with CLIP Latents,” Apr. 2022. Available: https://cdn.openai.com/papers/dall-e-2.pdf

Links to Images and Data:

Tables and Graphs of DALL-E 2 Images
https://docs.google.com/spreadsheets/d/1b-OnL0GB4b5Jf9zC1rNWs8uN3mGkCYvJ63GhM2es-eI/edit?usp=sharing

Tables and Graphs of DALL-E 3 Images
https://docs.google.com/spreadsheets/d/1CM6CSeIFwvWhH7wZ7y0lN7EkDMYEMFkv9O1WgO4F3bw/edit#gid=545735274

All Images Generated From DALL-E 2
https://docs.google.com/document/d/1MiZDGE4lM9KNYQvfQz2E_nTuspsHjMz6J1U7q4qIpNk/edit

All Images Generated From DALL-E 3
https://docs.google.com/document/d/12weRQFxCkvn8nZ5_BwtOdmndyhbztAszu9RuKVgwiSo/edit#heading=h.g7b6lrsl0eeu