Viccini Fleurival, Alyssa Fernandez, Aboubacar Sidik D

Professor Stemberg

ENG 21007 Writing for Engineers

22nd October 2023

Lab Report Paper (Racial and Gender Disparities in AI Image Generation)

Artificial Intelligence (AI) bias has gained significant attention for the reason that AI potentially perpetuates and amplifies biases and stereotypes in our society. This research aims to analyze gender and race biases from images of workers from a variety of professions, which has been generated by an AI image generation platform with a focus on comparing the findings to the true datasets of the job title demographics from the Department of Labor. Aiming to show that AI image generation exhibits stereotypes and bias. With a comprehensive approach to the study, the research refers to other existing studies on gender and race biases in artificial intelligence systems for more information to better understand the ethical issues and the consequences AI biases and stereotypes may have in our society. The approach will follow such that images of professionals collected from the AI platform are Analyzed and observed to identify which characteristics are noticed the most and the least, and what patterns the AI followed regarding the distribution of these professions among people based on their gender and race. Then based on the true datasets of job title demographics from zippia or the Department of Labor, we will compare our findings and conclude whether the study explicitly exhibits biases and stereotypes in AI image generation.

Artificial Intelligence (AI) software such as OpenAI and DALL-E2, (OpenAI's image-generating system) can produce images within seconds. Images are spawned from natural text-based language descriptions and inputs, outputs allowing for images that resemble realistic

photos, drawings, paintings, and other forms of art and styles. Drawing from a data set of over 650 million images across the internet, learning relationships between images and filtering out data sets that weren't appropriate, such as sexual, violent, and hateful content. Despite the prowess that OpenAI possesses, there are ethical concerns that come to mind. As with any technological advancement, there are still limitations that arise, OpenAI is designed to not generate images it's filtered out however there are limits, and content that could be deemed hateful or negative in any manner may still seep out. Given the possibilities, OpenAI was not released to the public and was only made available to a select few hundred people, AI researchers, and artists. OpenAI's DALL-E2 is regarded as a multimodal system as it was trained on text and images, for this reason, DALL-E2 displays some sort of multimodal bias. Regarding professions, when a user inputs words that pertain to a profession such as CEO, technology journalist, or builder, the outputs heavily feature men representing these professions. In a sense, DALL-E2 replicates societal stereotypes regarding most professions, not only replicating stereotypes but also heavily portraying white people and Western-like styles and settings. Racial bias in AI image generation isn't the only issue at hand, gender bias is also another major ethical concern. AI language models have revolutionized various fields, including natural language processing and sociological research. However, concerns have emerged regarding the presence of gender bias in these models, which can perpetuate stereotypes and reinforce societal biases. Our results indicate that text corpora contain recoverable and accurate imprints of our historic biases, whether morally neutral as toward insects or flowers, problematic as toward race or gender, or even simply veridical, reflecting the status quo distribution of gender concerning careers or first names. Our methods hold promise for identifying and addressing sources of bias in culture, including technology. In this Lab Report paper, my group mates and I look to answer

the question of Does AI image generation exhibits stereotypes and bias? We believe If we use AI to generate images of professions then AI will show white males.

In our research study, we selected three careers with neutral titles software engineer, nurse, and cosmetologist. Utilizing the website craiyon.io to artificially generate 100 diverse images of each career using exact phrasing that would not implicitly create biased images. Downloading the images and compiling them into OneDrive allowed the researchers to further examine and cross-reference them in tables. Creating tables with the provided tools in Word documents that accurately sorted the images based on societal perceptions such as gender, race, and age. Analyzed the table keeping an open mind but also observing which characteristic appeared the most, appeared the least, and what patterns the AI followed when a certain gender race, or age appeared and recording these findings on a PowerPoint. Compared the statistics generated from AI to statistics on the website Zippia and the Department of labor.

Graph of Races
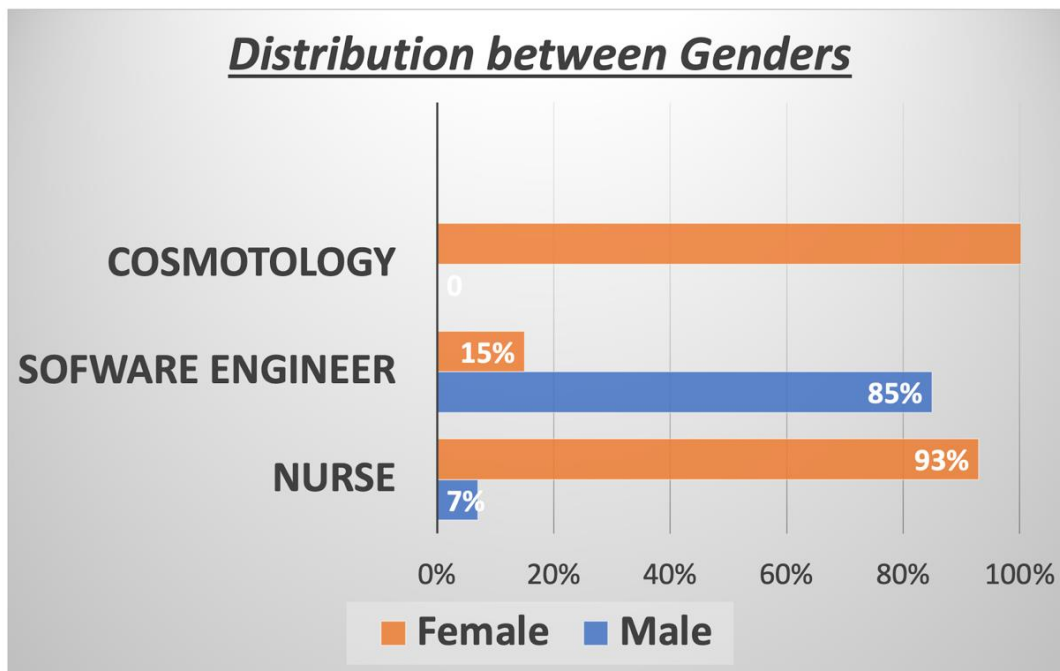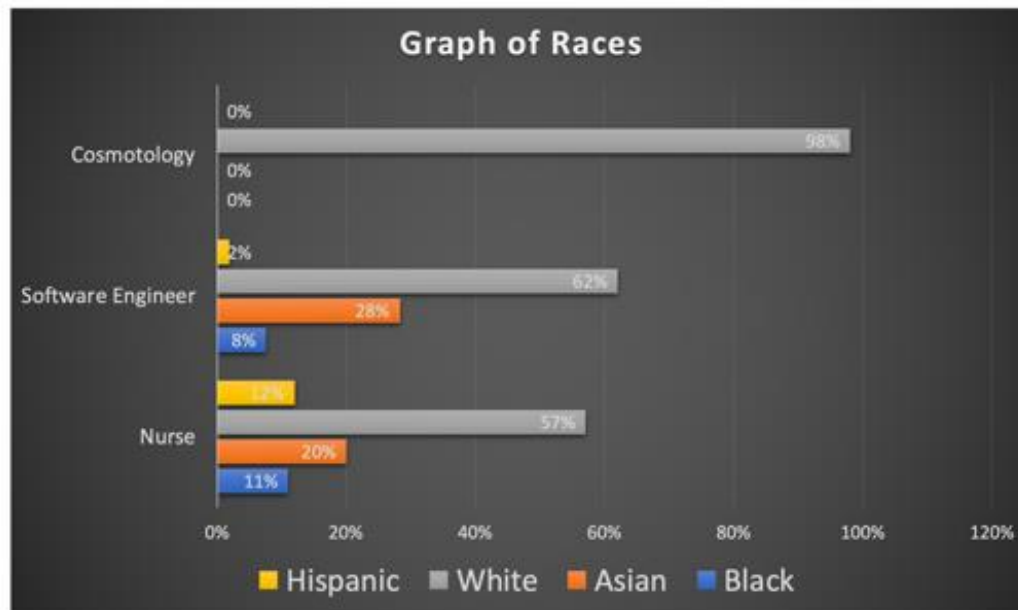
After generating 100 images for the search term "Software Engineer", 55 images were generated that contained perceptible faces. Out of those 55 images, 53 of them contained perceptible races that could be interpreted. White people were represented in 33 out of 53 images (62.3 percent). Asian people were represented in 15 out of 53 images (28.3 percent). Black people were represented in only 4 out of 53 images (7.5 percent). Hispanic people were represented in only 1 out of 53 images (1.9 percent). Males overall were represented in 45 out of 53 images (84.9 percent) and females were represented in 8 out of 53 images (15.1 percent). All of the white people that were represented appeared to be in their 20s to 40s. All of the Asian people that were represented appeared to also be in their 20s to 40s. All of the black people that were represented appeared to be in their 20s to 30s. The only Hispanic person represented appeared to be in their 20s. Regarding bias, the images generated by Craiyon featured white people being overly

represented. Asians were represented while blacks and Hispanics were extremely underrepresented.

In the case of nurse profession, contrary to software engineering, female nurses have the highest percentage of workers among the 100 images of nurses collected from craiyon.io. About 93% of the nurses were females and the remaining 7% were males. Based on race, 50% of the 93% females were white females, 20% were Asian women, 12 percent of them were Hispanic and the remaining 11% were black women. For male nurses, all 7% of male nurses collected during the data collection process from craiyon.io were white males. Based on these findings, we can observe that the bias in the algorithm regarding the profession of nurse is such that the algorithm assumes that this profession is a female-dominant profession. Based on race, white people are the dominant race in the nursing profession, followed by Asians, Hispanics, and black race.

Using AI to generate one hundred images of what it depicts as a cosmetologist resulted in 100 percent of the images being women, 100 percent of them being young seeming to be 30 or younger, and 98 percent being white, while two can be ambiguous enough where they might be perceived as Asian. In most of the images it appears that rather than the subject being a cosmetologist they were a patient of a cosmetologist., this can be observed as the instruments from a usual cosmetologist were being used to poke prod the women in these images. All the women had flawless, fair skin, skinny straight noses, sharp defined jawlines, high cheekbones, wide eyes, and hair.

The purpose of this experiment was to identify whether or not AI would perpetuate racial and gender biases upon image generation regarding certain professions, biases specifically leaning towards white males. Upon executing the experiment, we noticed that for the Software

engineering profession, AI generated a majority of males representing and a minority of females representing the profession. Out of 100 images generated from Craiyon, only 55 images had faces, and 53 with faces where race could be identified. The total is 53 images where race is perceptible. Of those 53 images, males were represented in 45 images (84.9 percent). Females were only represented in 8 images (15.1 percent). The massive margin between males being overly represented when it comes to the generation of images based on Software Engineers as opposed to women is already an indicator of gender bias within AI image generation. In terms of racial bias, out of 53 images, 33 images represented white people (62.3 percent). Asian people were represented in 15 out of 53 images (28.3 percent). Black people were only represented in 4 out of 53 images (7.5 percent) and only one person represented Hispanic people. 1 out of 53 images (1.9 percent) represented a Hispanic person. With white people being the clear majority in the representation of Software Engineers and black and Hispanic people specifically being the minorities in the representation of Software Engineers, there is a clear racial bias within AI image generation for this profession. When looking at statistics regarding Software Engineers in Zippia, results appeared to be similar. 78 percent of males represented Software Engineers across the United States while 22 percent of females represented Software Engineers across the United States. Once again, the majority of males represented the field, and a minority of females represented the field. When looking at racial statistics, 48.4 percent of Software Engineers across the United States were white, 34.0 percent were Asian, 8.1 percent were Hispanic or Latino and 4.8 were regarded as "unknown". Similar to the images that we generated, white people were represented as the majority for Software Engineers, with Asians having a considerable representation percentage as well. However, there was an abundance of Hispanic Software

Engineers and black people weren't even identified in the field. A clear gender gap and racial gap within the field can be seen in reality, the photos Craiyon generated only follow suit.

Regarding the nursing profession, after collection of data from generated images, we observed that the AI algorithm was generated by far more female nurses than male nurses. about 86% higher than the percentage of male nurses. This somehow collaborate with the true findings from the datasets of the job title demographics found at Zippia, which is about 14% of male nurses compared to 86% of a female nurse. Also, However, most images generated by the AI image used in the study more often generated images of the same individual in a different uniform or different position as a nurse. Which more often consisted of the same female nurses appearing in several images generated by the algorithm. Which to some extent shows the existence of gender biases in the algorithm. On the other hand, our collected data reflected a distribution of nurse professionals that shows biases against some races. The reason is that, from the datasets of the job title demographics found at Zippia based on races, The most common ethnicity of nurses is White (65.3%) which is close to our 57% result for white nurses. However, zippia data showed that Black or African Americans followed the lead with (11.5%), then Hispanic or Latino (9.6%), and Asian (8.9%). The AI image generation suggests that Asian nurses come second behind white nurses with 20% out of the 100% data collected. Followed by Hispanics with 12% and then Black or African American with 11%. Although the percentage of black nurses reflects the true data from zippia 11% against 11.5%, the order of the distribution among races based on the rank of races does not. Although with 11.5% black nurses, zippia data showed that there are more black nurses in the US than any other race except the white race.

Cosmetologists generated by AI were 98 percent white women presenting images. While the other 2 were still women while seemingly Asian. Observing the images all the women are fair-

skinned with slim noses full lips sharp features and high cheekbones. Using Zippia to compare the male and female ratio AI isn't completely inaccurate just slightly more skewed results. cosmetologists were shown to have a gender ratio of 93 percent women to seven percent men. Ai did accurately depict the most dominant gender in that field, however, when it comes to race it completely eradicated any semblance of diversity. Cosmetology is a field dominated by 60.8 percent white people, 14.6 percent Hispanic or Latino, 11.5 percent black or African American, and 7.5 Asian. This is an erasure of 39.2 percent of any race other than white in the field of cosmetology. AI completely disregarded the possibility of a different age because all the images presented displayed younger women looking below thirty while the average of a cosmetologist according to Zippita is 40. Most of the images were positioned as women getting some kind of medical or surgical work done. implying AI associates cosmetology with the action of the work rather than the profession. AI hyper-focusing on women attempting to look better shows what society has fed into AI especially the beauty standard in media and society being associated with young white women who have a set of certain features. The bias of AI for white women in cosmetology exposes the beauty standard in media ironically the women who seem to be the standard are the ones getting the work done.

If we use Ai to generate images of the profession's software engineer, nurse and cosmetologist it will show stereotypes and bias.

References:

Zippia. (2023, July 21). *Software Engineer Demographics and Statistics in the US*. Zippia.

https://www.zippia.com/software-engineer-jobs/demographics/

Strickland, E., Patel, P., Moore, S. K., & Choi, C. Q. (2022, August 8). *DALL-E2's Failures*

Reveal the Limits of AI > OpenAI's Text-to-Image Generator Struggles with Text, Science, and

Bias. IEEE Xplore. https://ieeexplore-ieee-org.ccny-

proxy1.libr.ccny.cuny.edu/document/9852405

Vishesh Thakur (Submitted on 18 Jul 2023) "Unveiling Gender Bias in Terms of Profession
Across LLMs: Analyzing and Addressing Sociological Implications"

https://arxiv.org/abs/2307.09162

AYLIN CALISKAN, JOANNA J. BRYSON, AND ARVIND NARAYANAN (14 APR 2017) "Semantics derived

automatically from language corpora contain human-like biases"

https://www-science-org.ccny-proxy1.libr.ccny.cuny.edu/doi/10.1126/science.aal4230