Assessing AI Model Bias in Depictions of Gender and Racial Diversity in the Workforce

Joshua Immordino, Al Jabar, Justin Mckenzie

1. Abstract

This study investigates bias in AI-generated images, focusing on how models like Craiyon.ai represent demographic diversity in occupations. By comparing AI-generated images for professions such as “Mathematician,” “Construction Worker,” and “Accountant” to real-world workforce statistics from the Bureau of Labor Statistics, we found notable discrepancies in gender and racial representation. Our results show an overrepresentation of white and male subjects, even in fields with substantial diversity. This phenomenon suggests that biases in training data are being reproduced in the AI's output.

The study explores potential sources of this bias, particularly in the selection and availability of training datasets. For example, recent industry partnerships, like Lionsgate's collaboration with Runway, indicate that media content is increasingly used in model training, potentially embedding existing societal biases from underrepresented groups in entertainment into AI. This trend underscores the importance of carefully curated, diverse training data to avoid reinforcing demographic stereotypes.

Our findings emphasize that AI models are not inherently neutral and require intentional design and oversight to reflect accurate and inclusive demographic representations across professional fields. We recommend a comprehensive approach to data selection and ongoing monitoring to reduce bias, aiming for more balanced and equitable AI applications in modeling diverse populations.

2. Introduction

Creating large-scale machine learning models is a dynamic and cyclical process that requires collaboration among many individuals. This process begins with collecting and labeling data for the model for training and testing. This data is then analyzed by Data scientists and then organized into frameworks called schemas. Hypotheses are then made and tested to evaluate their performance. During the assessment phase, Data scientists determine whether additional data augmentation or modification of inputs is necessary to improve results (Zhang et al., 2020; Schmidhuber, 2015). The cyclical nature of model training allows each stage to build upon the insights gathered from previous phases. It is designed like this to allow Data scientists and Engineers to make continuous adjustments to both the data frameworks and the hypotheses.

The utilization of AI has changed the game for 2 -3 years in many ways that we didn't imagine would happen with generating ideas helping with solving problems even speaking in similar accents and languages. Ultimately, has it really pulled us into the right direction in giving us the proper information, or does it have a mind of its own. Our research aims to show explicit bias in an Image Generation model (Craiyon.Ai) and to provide a focused analysis of the potential sources of this bias. In our study we have purposefully selected gender-neutral professions with diverse backgrounds, as documented by the U.S. Bureau of Labor Statistics (2023); and have generated over 300 images, utilizing synthetic images generated through Craiyon.ai.

3. Hypothesis

When generating images of gender-neutral professions with Crayion.ai, biases will become apparent in the representations of race and gender. Additionally, we intend to evaluate sources of potential bias, we anticipate that these images may reflect stereotypical or biased patterns. This study aims to demonstrate that bias can arise from both the dataset and the decisions made by data scientists and labelers, which ultimately affect the interpretations and outputs of the model.

4. Materials and Method

4.1 Data Selection and Professions Chosen

Since the objective of our study is to assess potential bias in AI-generated images for gender-neutral professions, we started with finding and selecting professions that we considered diverse. This diversity can be presented in two categories; race or gender. We referenced the U.S. Department of Labor (U.S. Department of Labor, 2023) to assist with the selection of diverse professions. The US Department of Labor diversity data lists the gender and race diversity of surveyed workers in their corresponding fields. This categorization provided by the Department of Labor was later used in our survey to identify a generated image's perceived labeling.

4.2 Data Collection and Generation

We ultimately decided to use Crayion.AI to synthesize our collection of images because we wanted a larger selection of images to survey researchers. Crayon.ai is a free-to-use Image generation model without a daily generation limit, this allowed us to quickly and easily produce a large data set. Craiyon.ai is known to rely on large datasets, but details about these datasets are not fully transparent to the end user. Without clear and ethical guidelines models act as a black box, which may lead to unintentional biases in the model's performance. In contrast, DALL-E; a popular Image generation model from OpenAI, integrates data curation efforts to minimize bias in generated images .”We are continuing to research how AI systems, like DALL·E, might reflect biases in its training data and different ways we can address them”(OpenAI,2022).

Once 300 images were generated, they were organized into collections of equal sizes. Each collection was based on the keywords used to originally generate the data. Placing them into these collections helps us study individual collections of images and compare data with the

Department of Labor data (Department of Labor,2023). This will ultimately share insight into potential biases that the image generation might hold in these categories.

4.3 Survey

A survey was distributed to our research group, a simple Excel spreadsheet. Each researcher was asked to evaluate every image in every collection, their responses were collected in a sheet dedicated to their responses. To regulate users' responses, 300 rows were generated each with a corresponding photo, and two drop-down inputs for race and gender. Each drop-down had the same selections:

Race	**Gender**
White	Male
Black	Female
Hispanic/Latino	Non-Binary
Asian	Other
None	None

Gender: Male, Female, Non-Binary, Other and None.
Race: Black, White, Asian, Hispanic/Latino or None.

It is important to note that if a user selects None in a category their selections would be omitted from the final statistics of that category. Additionally, researchers were not told about this feature but instead instructed to complete the survey with certainty. As the researchers completed their surveys, statistics about every image categorization are automatically updated on a sheet dedicated to reporting.

4.4 Analysis of AI Bias

Once our data was collected and organized we compared it to the Department of Labor statistics (Department of Labor,2023). We compared both race and gender diversity within the profession and hypothesized on possible causation of the results we received. We decided to compare our data to other sources to explain patterns we saw within our surveyed data. We started by examining the race and gender distribution in Film and TV media and comparing any patterns that are similar to our collected data.

4.5 Limitations and Controls

Since the craiyon.ai algorithm is beyond our control, we tried to make our keywords as gender-neutral as possible. Additionally, when selecting professions we purposefully found careers that were somewhat diverse according to the Department of Labor diversity statistics (Department of Labor,2023). Finally, while performing the survey portion of this experiment images were cross-examined by multiple researchers to ensure consistency in identifying visual patterns and reduce subjective interpretation.

5. Results

5.1 Mathematicians

The "Mathematician" keyword was initially chosen due to the remarkable diversity statistics in the mathematical sciences field, both by gender and racial composition. According to the U.S. Census Bureau (2023), Asians make up approximately 6% of the total U.S. population; however, they represent a significant 23.4% of the workforce within mathematical science occupations (Bureau of Labor Statistics, 2023). Additionally, gender representation in this field shows that 50.7% of mathematicians identify as female, a higher-than-average percentage compared to many STEM-related careers.

Figure 1.

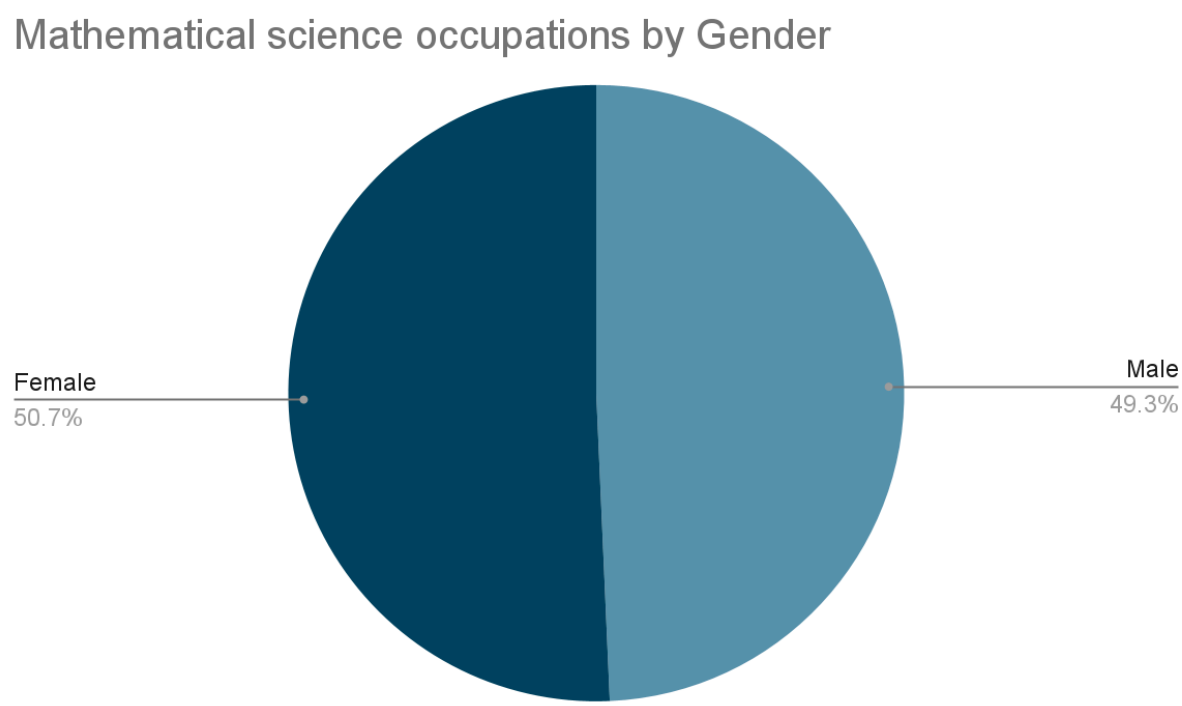
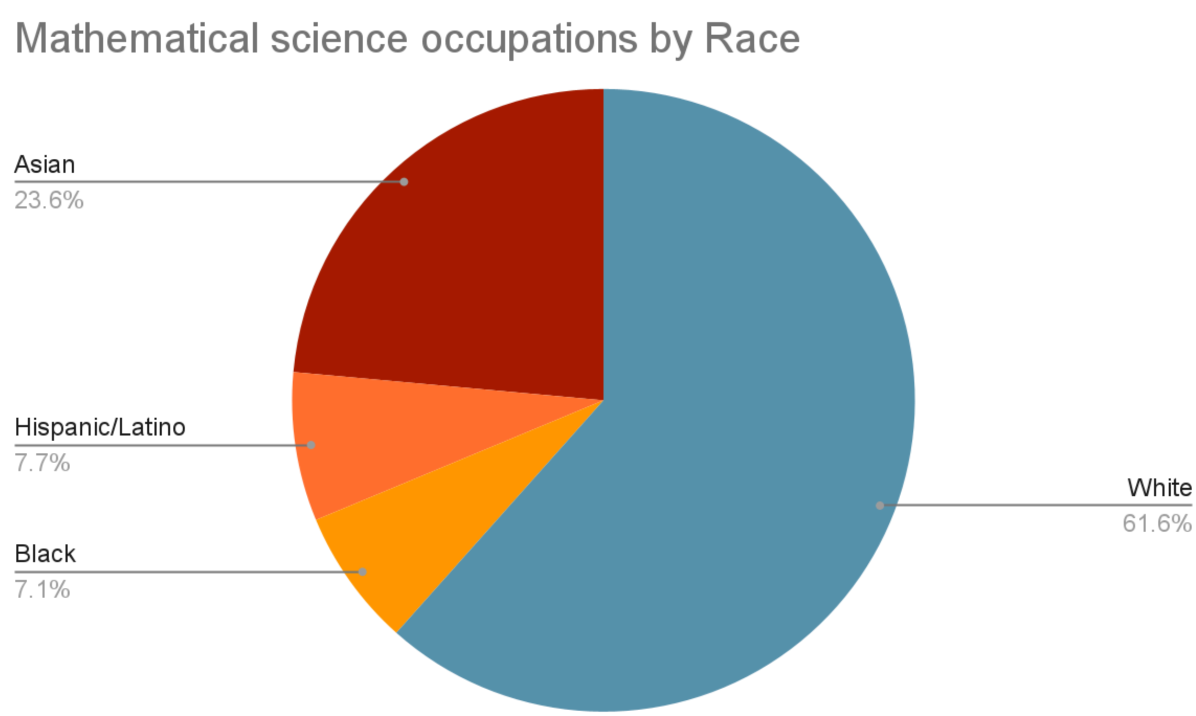


Figure 2.



U.S. Bureau of Labor Statistics. (n.d.). *Employed persons by detailed occupation, sex, race, and Hispanic or Latino ethnicity*. U.S. Bureau of Labor Statistics. Retrieved November 6, 2024, from https://www.bls.gov/cps/cpsaat11.htm

In contrast, our analysis of images generated by Craiyon.ai using the "Mathematician" keyword revealed notable disparities in both gender and racial representation. When researchers assessed the synthesized images for perceived demographic characteristics, the results highlighted a skewed portrayal:

- Gender Representation: Approximately 81% of the images were perceived as male, with only 7% perceived as female and 10.7% perceived as non-binary. This stark contrast against real-world statistics suggests an underrepresentation of female-presenting individuals, despite the BLS data indicating that women make up over half of the mathematical science profession.

- Racial Representation: The racial diversity in the generated images was even more limited. Of the images, 96.6% of the subjects were perceived as white, with only 1.8% perceived as Asian, 1.1% as Hispanic or Latino, and a mere 0.03% as Black. These percentages contrast sharply with the BLS data, where a substantial portion of the mathematical workforce includes Asian individuals. The model's output suggests a strong inclination toward white representation, which does not align with actual workforce demographics.

Figure 3.

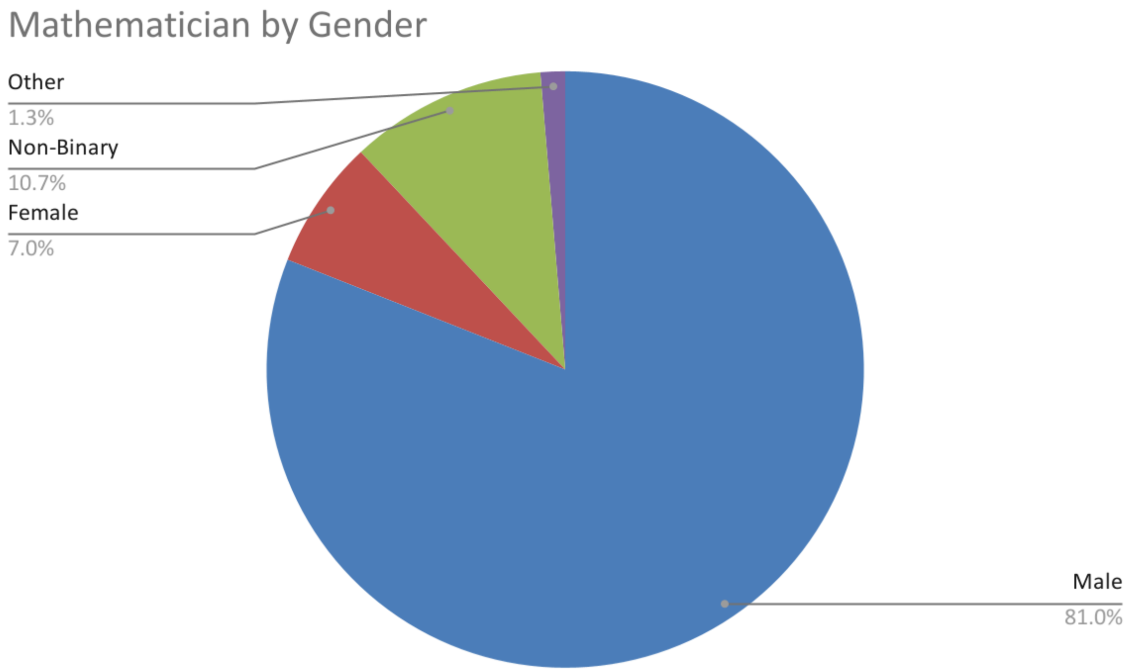
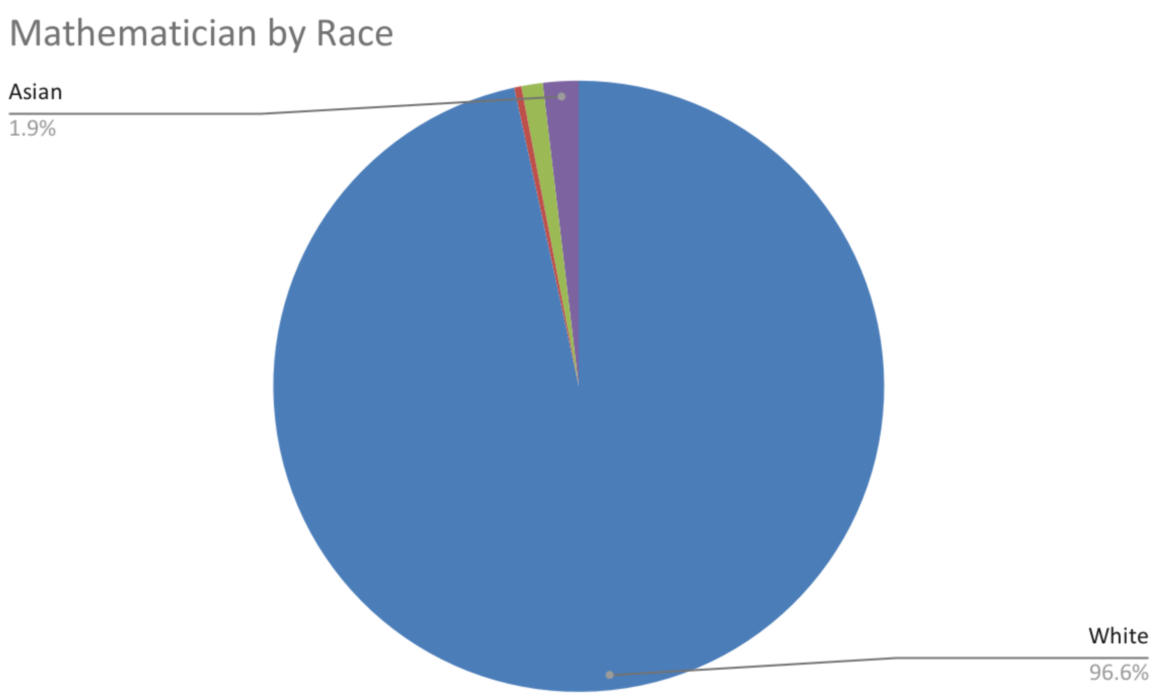


Figure 4.



These findings suggest that the Craiyon.ai model may contain biases that impact the portrayal of diversity within specific professions. The model's training data, likely shaped by existing media or image datasets, may not sufficiently capture the diversity present in real-world professions, particularly in fields where demographic variety is well-documented. This lack of alignment highlights the potential for AI-generated images to perpetuate stereotypes, limiting the accurate portrayal of professionals across different demographic backgrounds.

5.2 Construction and Excavation occupations

The "Construction Worker" keyword was chosen due to its broad application across both skilled and unskilled labor roles. This category encompasses a range of trades within construction and excavation, fields that reflect distinct demographic trends. According to the Bureau of Labor Statistics (2023), this workforce is overwhelmingly male, with 95.7% identifying as male. Racially, the industry has more diversity: approximately 69% of construction workers identify as white, 29.9% as Hispanic or Latino, 5.3% as Black or African American, and only 0.9% as Asian.

Figure 5.

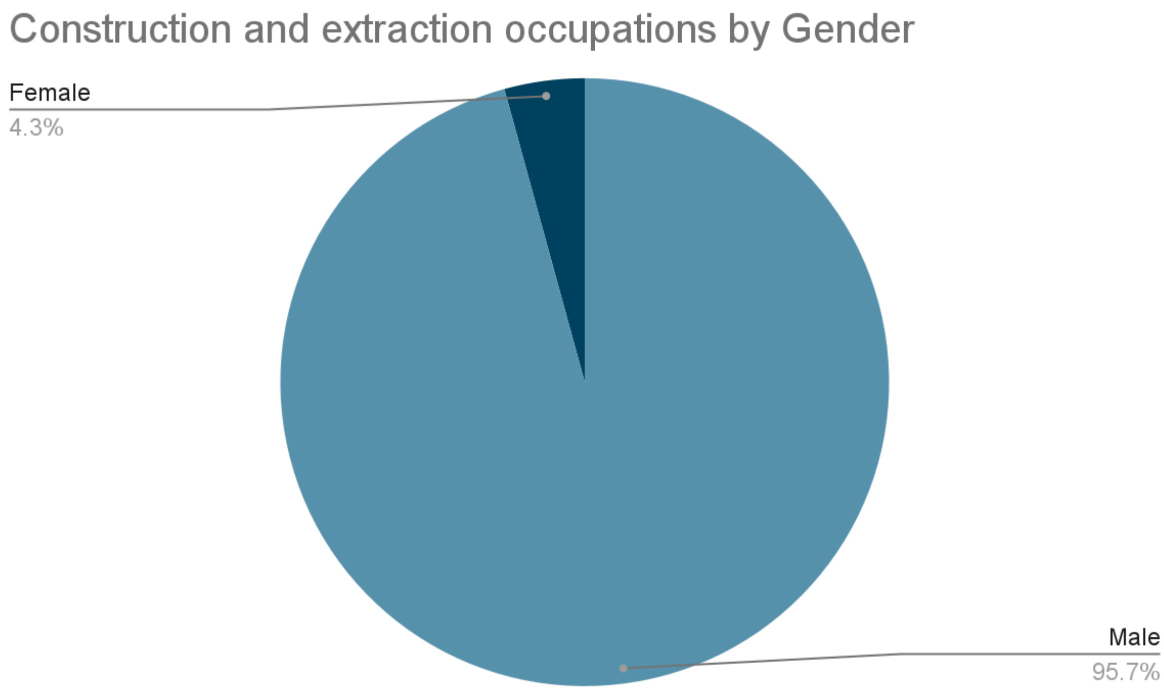
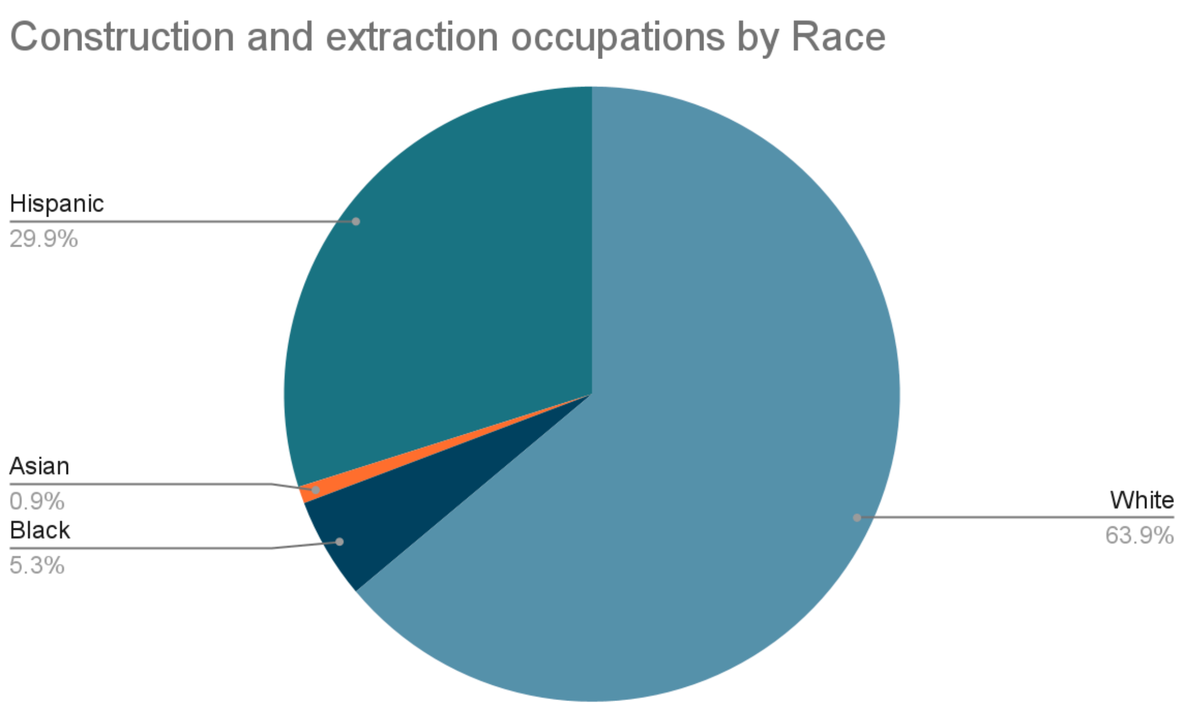


Figure 6.



U.S. Bureau of Labor Statistics. (n.d.). *Employed persons by detailed occupation, sex, race, and Hispanic or Latino ethnicity*. U.S. Bureau of Labor Statistics. Retrieved November 6, 2024,

from https://www.bls.gov/cps/cpsaat11.htm

In our analysis, images generated by Craiyon.ai using the "Construction Worker" keyword displayed a mixed representation of the demographic patterns documented by the BLS:

- Gender Representation: The model produced 92.3% male images, closely aligning with real-world statistics. This suggests that the AI model is effective at representing the gender composition of the construction industry.

- Racial Representation: However, racial representation in the generated images revealed a different trend. Among the synthesized images, 82% of subjects were perceived as white, 5.1% as Hispanic or Latino, 5.7% as Black or African American, and 6.4% as Asian. These results show an overrepresentation of white individuals and Asian representation compared to real-world demographics, while Hispanic and Black/African American representation was notably lower than actual workforce statistics.

Figure 7.

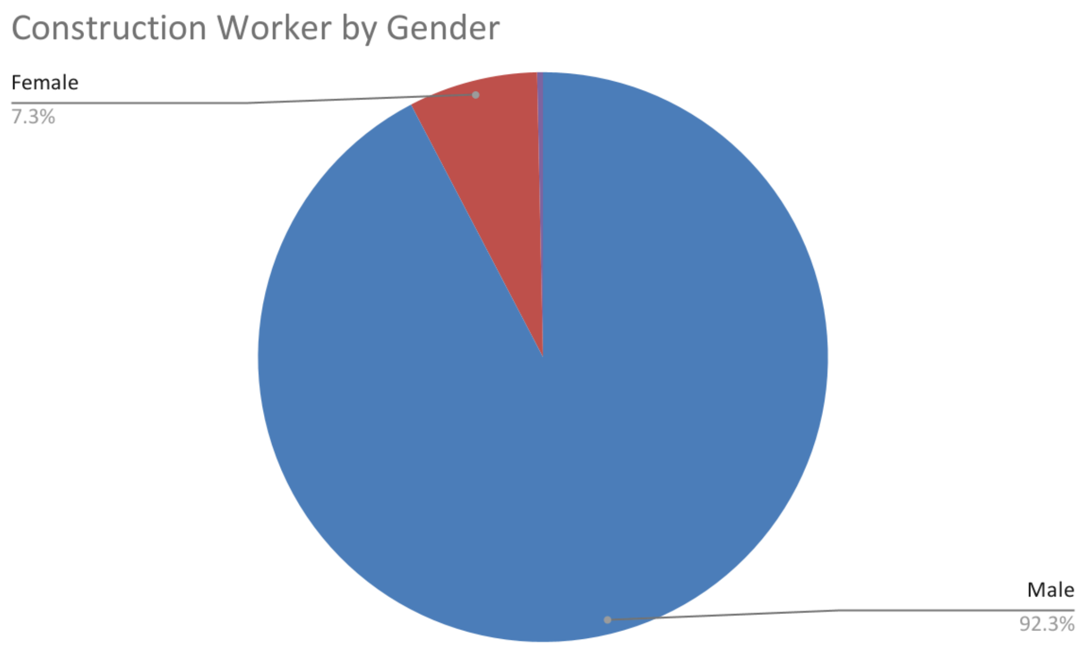
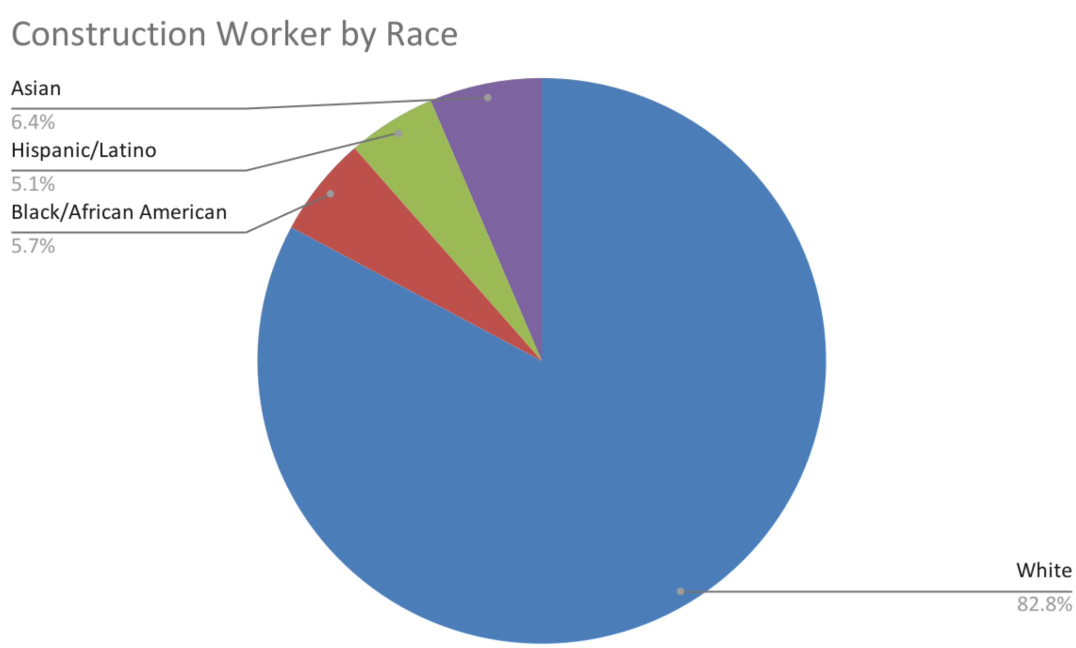


Figure 8.



These results suggest that while the Craiyon.ai model aligns with gender expectations in this field, it exhibits potential racial bias, favoring certain groups and underrepresenting others commonly found in the workforce. This outcome may result from biases in the training data, where imagery of construction workers may have skewed representations of race, thereby impacting the generated images. This discrepancy underscores the need for careful evaluation and selection of data sources to ensure a balanced representation across both gender and racial demographics.

5.3 Accountant

The "Accountant" keyword was chosen due to its notable gender diversity in the workforce, where women hold a slight majority in the field. According to the Bureau of Labor Statistics (2023), women make up 57% of accountants, leaving men as a minority at 43%. Additionally, the racial demographics in accounting show some variation compared to the broader U.S. population: 68.9% of accountants identify as white, while 11.2% are Black or African American, 8% are Hispanic or Latino, and 12% are Asian.

Figure 9

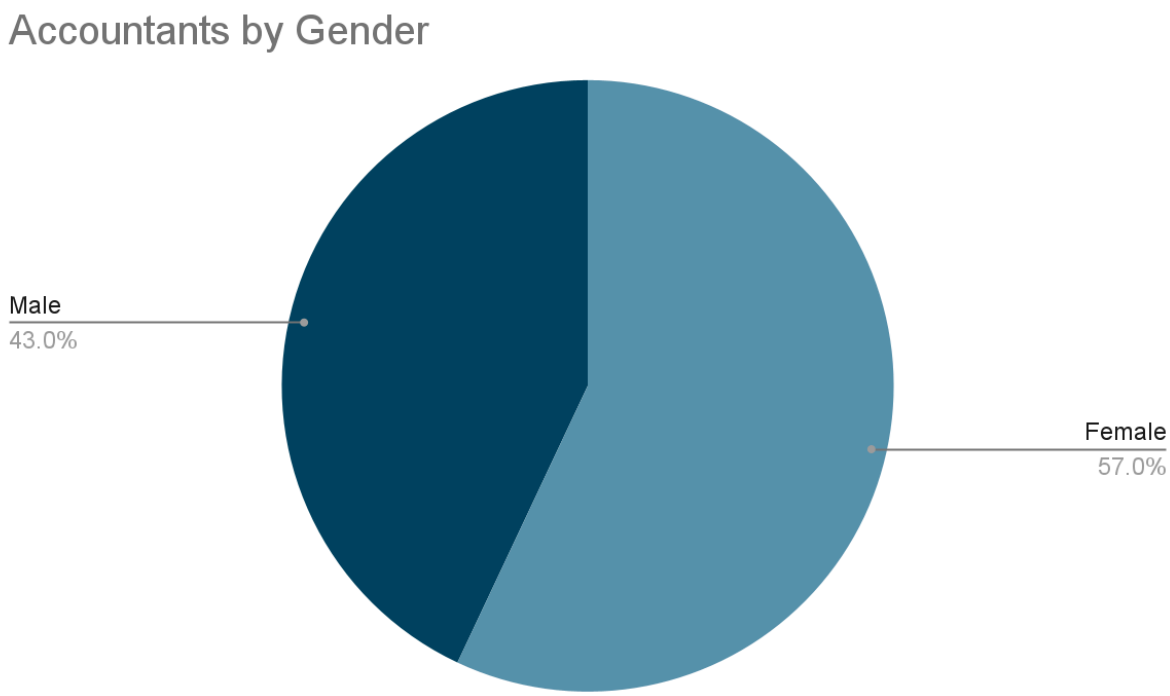
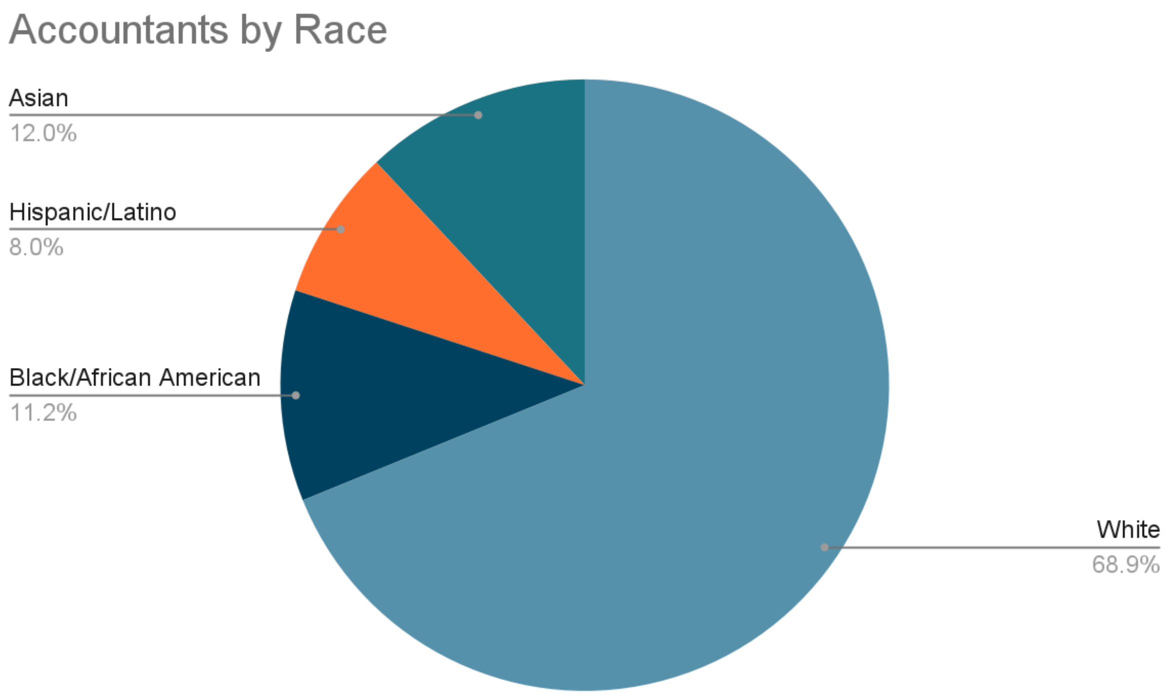


Figure 10.



U.S. Bureau of Labor Statistics. (n.d.). *Employed persons by detailed occupation, sex, race, and Hispanic or Latino ethnicity*. U.S. Bureau of Labor Statistics. Retrieved November 6, 2024, from https://www.bls.gov/cps/cpsaat11.htm

Upon examining Craiyon.ai-generated images for the "Accountant" keyword, we found some disparities when compared to real-world demographics as documented by the BLS:

- Gender Representation: Research perceived 70.6% of the generated images to depict men, with only 28.1% perceived as women, and a small proportion, 1.3%, perceived as non-binary. This distribution diverges significantly from the actual gender representation within the accounting profession, suggesting that the model may have a bias toward depicting accountants as male, despite women being the majority in this field.

- Racial Representation: The racial makeup in the AI-generated images showed closer alignment to real-world statistics but still with notable differences. Of the generated images:
  - 69% of subjects were perceived as white
  - 17.2% were perceived as Black or African American
  - 2.4% were perceived as Hispanic or Latino
  - 11.4% were perceived as Asian

Figure 11.

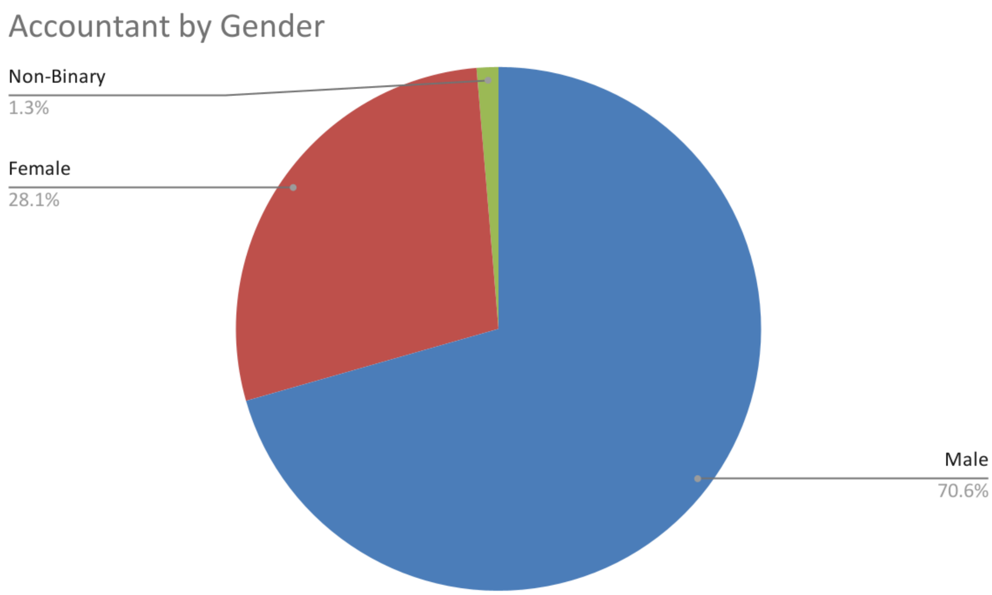
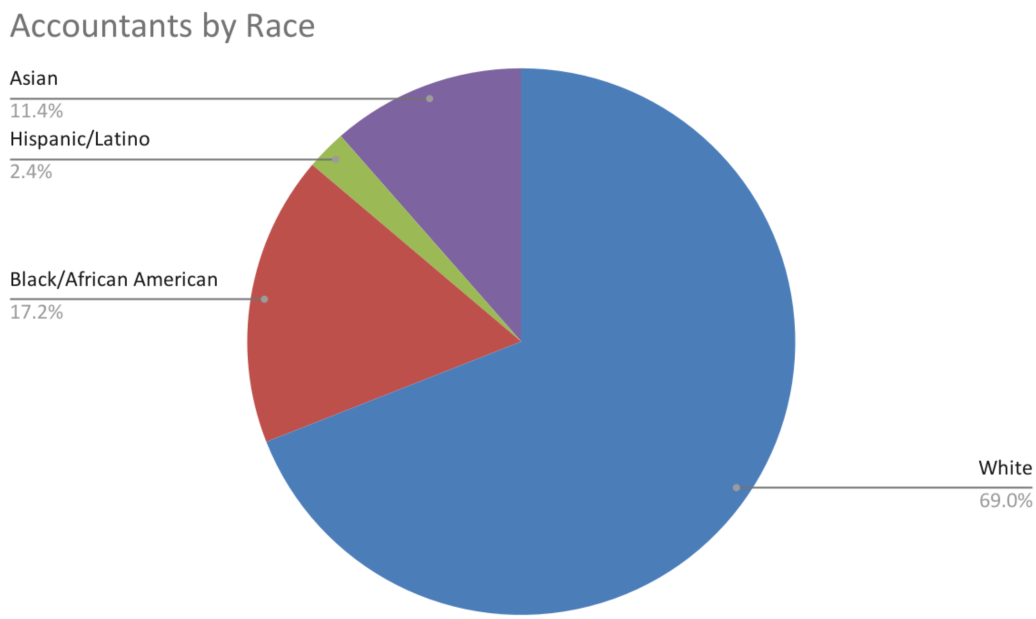


Figure 12.



These results reveal that while Craiyon.ai's model reasonably reflects racial diversity in the accounting field, it lacks accuracy in gender representation, favoring male depictions despite the actual female majority. This discrepancy may stem from the training data used to develop the AI model, where existing biases in professional portrayals could lead to an overrepresentation of men in traditionally gender-neutral professions. This highlights the importance of curating balanced training datasets to mitigate the influence of such biases in AI-generated content

5.4 Image Collection Comparison to Other Sources

In our study, researchers observed a significant lack of diversity across both race and gender in the generated images across all selected categories. Noting this pattern, we hypothesized that underlying biases in machine learning (ML) models could stem from the datasets and media sources used to train these models. Recognizing this correlation, we investigated the types of datasets that AI image generation models typically rely on and how these datasets may contribute to the bias observed in AI outputs.

A notable example that informed our research approach is the recent partnership between Lionsgate, a prominent movie studio, and Runway, a company specializing in generative AI tools. Through this collaboration, Lionsgate granted Runway access to its extensive licensed media content for training and testing purposes, potentially reinforcing the influence of media portrayals on AI models. We focused on media portrayals of lead and supporting actors to explore how the patterns within these datasets could affect model bias.

Industry data illustrates a striking gender and racial imbalance in casting: for example, in 2023, men represented 67% of lead roles in major productions (Carollo, 2024). In terms of racial representation, individuals identifying as white accounted for 69.2% of lead roles, while Black or African descent actors made up only 17.6%, and Hispanic or Latino actors comprised 6.6% (Common Senes Media, 2021). These statistics suggest that if training datasets reflect similar proportions, models trained on them might replicate and even reinforce these disparities, impacting the diversity and accuracy of AI-generated content.

6. Conclusion

Our study highlights the inherent biases that can permeate AI models, particularly in fields like image generation. Throughout our research, we observed significant imbalances in gender and racial representation across various occupational categories when comparing AI-generated images to real-world data. This suggests that, despite the potential of machine learning models to emulate human creativity and insights, they are often constrained by the biases embedded in their training data. These biases are a product of multiple factors: limitations in dataset diversity, historical representation gaps in media, and the oversights in data curation processes by developers and researchers.

For instance, in professions like "Mathematician," where real-world diversity is notable, Craiyon.ai outputs displayed a limited representation of gender and race, heavily skewing towards white, male subjects. Similarly, in "Construction Worker" and "Accountant" categories, while there was some alignment with actual gender representation, racial diversity was underrepresented. Such disparities suggest that the model's training data may reflect societal biases or rely on visual cues and media sources that have historically lacked diverse representation. As with media partnerships, like that between Lionsgate and Runway, when models are trained on narrow or stereotypical depictions, they reinforce these biases rather than provide an accurate reflection of the workforce diversity.

Our findings underscore the importance of conscientious data selection and the establishment of bias mitigation strategies during AI development. Training AI models requires more than just a large volume of data; it requires diverse, representative datasets that encapsulate the real-world variety in gender, race, and other demographic factors. Without this, AI models may propagate a limited, often inaccurate worldview, inadvertently marginalizing certain groups and reinforcing stereotypes. These biases call for an industry-wide commitment to improving data curation practices and implementing ongoing oversight to make AI outputs as inclusive and representative as possible.

Ultimately, our study serves as a reminder that while AI models are powerful tools, they are not neutral. They reflect the datasets and societal norms they are exposed to and therefore must be trained and monitored with careful attention to diversity and fairness if they are to serve as equitable resources in society.

Bolukbasi, T., Chang, K.-W., Zou, J. Y., Saligrama, V., & Kalai, A. T. (2016). Man is to computer programmer as woman is to homemaker? Debiasing word embeddings. *Proceedings of the 30th International Conference on Neural Information Processing Systems*, 4356–4364.

Carollo,Laura 2024 *Distribution of lead actors in movies in the U.S. 2011-2023, by gender*. Retrieved from https://www.statista.com/statistics/692465/distribution-lead-actors-gender/
Crawford, K. (2021). *Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence*. Yale

Common Sense Media. 2021 *Representation fact sheets*. Common Sense Media. Retrieved from https://www.commonsensemedia.org/sites/default/files/research/report/representation-fact-sheets_final-release_0.pdf

LionsGate (2024). Runway Partners with Lionsgate in First-of-its-Kind AI Collaboration https://investors.lionsgate.com/news-and-events/press-releases/2024/09-18-2024-140126979

Schmidhuber, J. (2015). Deep learning in neural networks: An overview. *Neural Networks*, *61*, 85-117. https://doi.org/10.1016/j.neunet.2014.09.003

What Did My AI Learn? How Data Scientists Make Sense of Model Behavior *Communications of the ACM*. https://doi.org/10.1145/3542921

University Press.OpenAI. (2022, July 18). Reducing bias and improving safety in DALL·E 2. OpenAI. https://openai.com/index/reducing-bias-and-improving-safety-in-dall-e-2/

U.S. Bureau of Labor Statistics. (n.d.). *Employed persons by detailed occupation, sex, race, and Hispanic or Latino ethnicity*. U.S. Bureau of Labor Statistics. Retrieved November 6, 2024, from https://www.bls.gov/cps/cpsaat11.htm

Zhang, Y., et al. (2020). A comprehensive review of AI and deep learning models and techniques.

Zippia. (n.d.). *Data analyst demographics and statistics in the U.S.* Zippia. Retrieved November 6, 2024, from https://www.zippia.com/data-scientist-jobs/demographics/