Facial Demography Analysis of the LAION Dataset

IRIS DOMINGUEZ-CATENA, Public University of Navarre, Spain
DANIEL PATERNAIN, Public University of Navarre, Spain
MIKEL GALAR, Public University of Navarre, Spain

Large-scale image-text datasets have become fundamental building blocks for modern AI systems, raising concerns about the demographic biases they may encode and propagate. We present a comprehensive analysis of LAION, one of the largest and most influential datasets in this domain, focusing on demographic representation and intersectional biases across age, gender and race. Our methodology combines state-of-the-art face detection (RetinaFace) with specialized demographic classifiers (FairFace and EMO-AffectNet) to analyze a random sample of 500,000 image URLs from ReLAION-2B-en, yielding over 37,000 faces. We analyze both general representational biases, revealing severe overrepresentation of certain groups—such as white people and individuals aged 20-29—and intersectional biases, notably the underrepresentation of women over 30 years old and non-White infants. These results highlight the importance of considering not just individual demographic attributes, but their intersections when evaluating and mitigating bias in large-scale datasets.

Keywords: LAION, Demographic Bias, Intersectional bias, Dataset Analysis, Fairness

Reference Format:
Iris Dominguez-Catena, Daniel Paternain, and Mikel Galar. 2025. Facial Demography Analysis of the LAION Dataset. In *Proceedings of Fourth European Workshop on Algorithmic Fairness (EWAF'25).* Proceedings of Machine Learning Research, 5 pages.

1 Introduction

Large-scale image-text datasets like LAION [1] have become fundamental in training modern vision-language models. Understanding the demographic composition and potential biases in these datasets is crucial for ensuring fair and representative AI systems [4], even if it's not the single driving cause for bias in the model predictions [11].

Unlike previous works that focus on bias in trained models [2, 8, 12], in this work, we examine the source of these biases, providing the first detailed analysis of demographic representational and intersectional bias in LAION across multiple dimensions: age, gender and race.

2 Methodology

We begin by randomly sampling 500,000 URLs from ReLAION-2B-en [1], the main variant of LAION, the largest public image-text pairs dataset. From the successfully downloaded images (227,748 total), we detect faces using RetinaFace [9, 10], a model well adapted to challenging conditions and scale variations, removing those smaller

Authors' Contact Information: Iris Dominguez-Catena, Public University of Navarre, Pamplona, Spain, iris.dominguez@unavarra.es; Daniel Paternain, Public University of Navarre, Pamplona, Spain, daniel.paternain@unavarra.es; Mikel Galar, Public University of Navarre, Pamplona, Spain, mikel.galar@unavarra.es.

EWAF'25, June 30–July 02, 2025, Eindhoven, NL

than 48×48 pixels. Finally, we analyze the demographic attributes of these faces using FairFace [6], a model specifically designed for robust multi-demographic classification, to obtain predictions for age, race, and gender.

Subsequently, we perform both a general analysis of the proportions of each demographic group within their corresponding variable, and an intersectional analysis using the Ducher's Z metric [5], following the recommendations in [3]. Intuitively, Z compares the observed co-occurrence of group $g \in G$ and class $y \in Y$ to the expected co-occurrence if they where independent. It is defined as:

$$\mathrm{Z}(X,g,y) = \begin{cases} \frac{p_{g \wedge y} - p_g p_y}{\min[p_g, p_y] - p_g p_y} & \text{if } p_{g \wedge y} - p_g p_y > 0 \\ \frac{p_{g \wedge y} - p_g p_y}{p_g p_y - \max[0, p_g + p_y - 1]} & \text{if } p_{g \wedge y} - p_g p_y < 0 \\ 0 & \text{otherwise,} \end{cases} \tag{1}$$

where p_g, p_y and $p_{g \wedge y}$ are the proportions of samples in the population X that belong to group g, class y or both, respectively. The values of Z are in the range $[-1, 1]$, with 1 being the maximum overrepresentation of the combination of group and class, 0 being no correlation and -1 being the maximum underrepresentation of the combination.

3 Results

3.1 Demographic Distribution

Figure 1 shows the proportion of the dataset corresponding to each demographic group. The analysis reveals substantial imbalances across all demographic attributes in ReLAION-2B-en. Regarding **age** distribution, there is a strong bias towards young adults, with individuals aged 20-29 representing 44% of all detected faces. This overrepresentation sharply contrasts with other age groups, particularly the extremes over 70 or under 2 years old, which account for less than 2% of the dataset each. **Gender** distribution shows a notable male bias, with men appearing in approximately 57% of images. In terms of **racial** representation, White individuals are heavily overrepresented, comprising 55% of all detected faces. Other racial groups show significantly lower representation, with Southeast Asian in particular being underrepresented at only 1% of the dataset.

3.2 Intersectional Bias Analysis

We analyze intersectional bias by examining correlations between demographic attributes using Ducher's Z [5]. This metric ranges from -1 to +1, where, for each combination of demographic groups, -1 indicates complete absence, 0 represents expected proportional representation, and +1 denotes maximum overrepresentation. Figure 2 presents the Ducher's Z scores for all pairwise combinations of age, gender, and race.

Regarding the **age-gender** correlation, we observe moderate to strong correlations across all age ranges between 30 and 69 years, with middle-aged women being underrepresented. In contrast, women are weakly overrepresented in the 10-29 age range.

For the **age-race** correlation, we identify several significant demographic biases. White infants (0-2 years) are moderately overrepresented, while all other racial groups except Southeast and East Asian are strongly underrepresented in this age range. East Asian individuals show overrepresentation in the 20-29 age range but moderate underrepresentation between ages 30-69. In the 70+ age group, Latino/Hispanic individuals exhibit strong underrepresentation, while Black individuals show moderate underrepresentation.

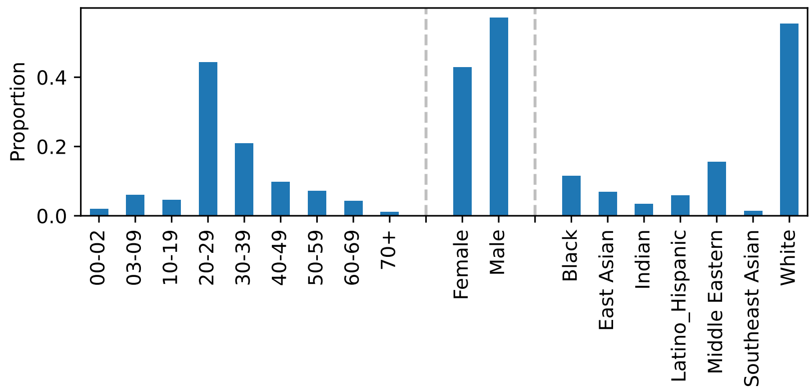
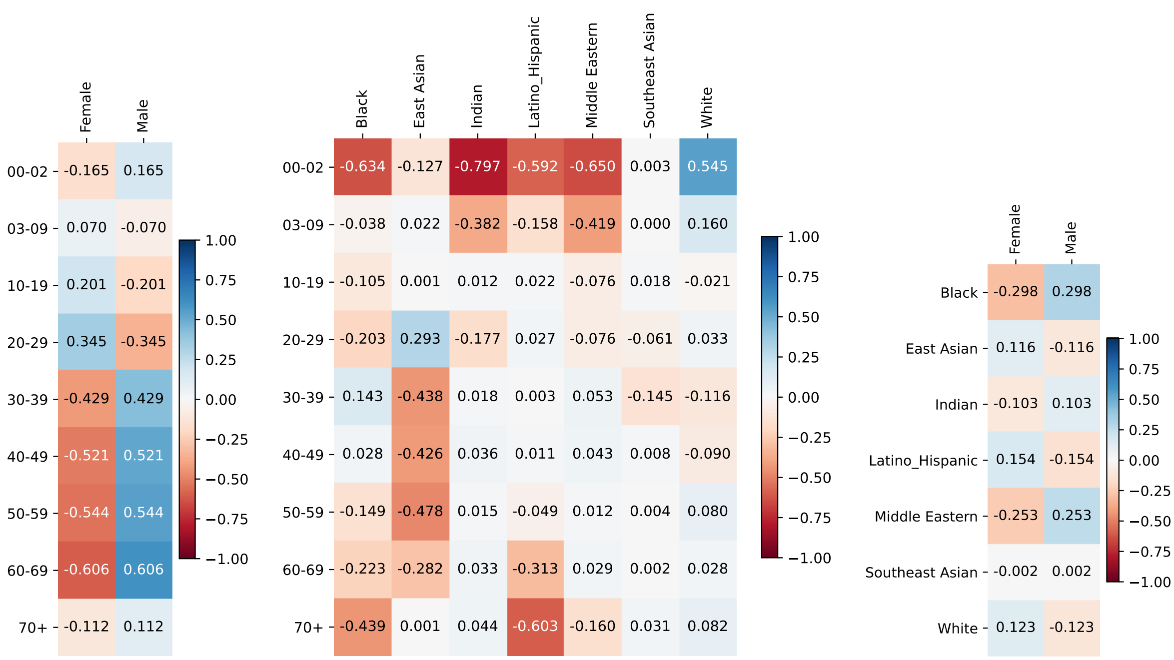


Fig. 1. Distribution of demographic attributes in LAION: age, gender and race.

Finally, **race-gender** correlations show weak correlations for Black and Middle Eastern individuals, in both cases overrepresenting men.



(a) Age-Gender (b) Age-Race (c) Race-Gender

Fig. 2. Intersectional bias analysis using Ducher's Z metric across different demographic attribute pairs.

4 Discussion

Our analysis reveals significant demographic biases in ReLAION-2B-en that mirror those found in other Internet sourced datasets, such as those for Facial Expression Recognition (FER) [3]. We observe a strong bias towards young adults (20-29 age range) and White individuals (55%), aligned with biases identified in image generation models trained on LAION [8]. The most notable difference to previous datasets is the substantial male overrepresentation (57%), which contrasts with the generally balanced gender distribution found in FER datasets. This gender imbalance is particularly concerning as it persists across multiple demographic intersections and could significantly impact downstream applications.

The intersectional analysis reveals particularly concerning patterns in gender-age correlations, notably the significant underrepresentation of middle-aged women. The age-race correlations show critical disparities, especially in the representation of infants and elderly individuals across racial groups. White infants are disproportionately overrepresented while other racial groups are underrepresented in early age ranges, and Latino/Hispanic, Black and East Asian individuals face substantial underrepresentation in older age groups. These biases could perpetuate and amplify existing societal inequities when such datasets are used to train large-scale AI models.

4.1 Limitations

Several limitations should be considered when interpreting our results. Our analysis relies on auxiliary models for face detection and demographic classification, which may introduce their own biases. Future work should validate these findings against self-reported or human-labeled data to distinguish between biases inherent in LAION and those introduced by our measurement tools. Additionally, the demographic categories are constrained by the pre-defined labels in the auxiliary models, which may not capture the full spectrum of human diversity, particularly regarding gender identity and racial categories.

Finally, the use of different operational definitions of fairness and bias [7] implies that the goals we pursue at the model level can vary substantially. As a result, there is not a single ideal dataset composition, and enforcing a perfectly balanced representation of all demographic groups is not always the most appropriate strategy for a given fairness objective. Furthermore, recent work indicates that dataset imbalances do not translate into model biases in a straightforward manner [4].

5 Conclusions

This study demonstrates significant demographic biases in LAION that could have substantial implications for downstream applications. While certain biases, such as age and racial representation imbalances, mirror patterns seen in other datasets [3], LAION's pronounced gender bias presents unique concerns.

The impact of these biases may vary depending on the application. While previous research suggests that demographic disproportions might have limited impact on some classification tasks like FER [4], the implications for generative AI systems could be more severe [8]. This is particularly relevant given LAION's widespread use in training large generative models.

Future work should focus on developing more accurate and comprehensive demographic classification tools, creating methods to mitigate these biases during dataset curation, and investigating the specific impact of these biases on different downstream applications, particularly in generative AI.

These findings underscore the importance of careful dataset curation and the need for diverse, representative training data in machine learning applications, especially for general-purpose datasets like LAION that are used across multiple domains.

6 Funding

This work was funded by the Spanish MICIN (PID2022-136627NB-I00/AEI/10.13039/501100011033 FEDER, UE), the Government of Navarre (PC24-GÉNESIS-041-002), and the support of the 2024 Leonardo Grant for Researchers and Cultural Creators from the BBVA Foundation. The BBVA Foundation is not responsible for the opinions, comments, and content included in the project and/or its derived results, which are the sole responsibility of the authors.

References

[1] [n. d.]. Releasing Re-LAION 5B: Transparent Iteration on LAION-5B with Additional Safety Fixes | LAION. https://laion.ai/blog/relaion-5b

[2] Abeba Birhane, Sepehr Dehdashtian, Vinay Prabhu, and Vishnu Boddeti. 2024. The Dark Side of Dataset Scaling: Evaluating Racial Classification in Multimodal Models. In *Proc. 2024 ACM Conf. Fairness Account. Transpar. (FAccT '24)*. Association for Computing Machinery, New York, NY, USA, 1229–1244. https://doi.org/10.1145/3630106.3658968

[3] Iris Dominguez-Catena, Daniel Paternain, and Mikel Galar. 2024. Metrics for Dataset Demographic Bias: A Case Study on Facial Expression Recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* (2024), 1–18. https://doi.org/10.1109/TPAMI.2024.3361979

[4] Iris Dominguez-Catena, Daniel Paternain, Mikel Galar, MaryBeth Defrance, Maarten Buyl, and Tijl De Bie. 2025. Biased Heritage: How Datasets Shape Models in Facial Expression Recognition. https://doi.org/10.48550/arXiv.2503.03446 arXiv:2503.03446 [cs]

[5] M. Ducher, C. Cerutti, M. P. Gustin, and C. Z. Paultre. 1994. Statistical Relationships between Systolic Blood Pressure and Heart Rate and Their Functional Significance in Conscious Rats. *Med. Biol. Eng. Comput.* 32, 6 (Nov. 1994), 649–655. https://doi.org/10.1007/BF02524241

[6] Kimmo Karkkainen and Jungseock Joo. 2021. FairFace: Face Attribute Dataset for Balanced Race, Gender, and Age for Bias Measurement and Mitigation. In *2021 IEEE Winter Conf. Appl. Comput. Vis. WACV*. IEEE, Waikoloa, HI, USA, 1547–1557. https://doi.org/10.1109/WACV48630.2021.00159

[7] Shira Mitchell, Eric Potash, Solon Barocas, Alexander D'Amour, and Kristian Lum. 2021. Algorithmic Fairness: Choices, Assumptions, and Definitions. *Annu. Rev. Stat. Appl.* 8, 1 (March 2021), 141–163. https://doi.org/10.1146/annurev-statistics-042720-125902

[8] Leonardo Nicoletti and Dina Bass Technology + Equality. 2023. Humans Are Biased. Generative AI Is Even Worse. *Bloomberg.com* (April 2023). https://www.bloomberg.com/graphics/2023-generative-ai-bias/

[9] Sefik Serengil and Alper Ozpinar. 2024. A Benchmark of Facial Recognition Pipelines and Co-Usability Performances of Modules. *J. Inf. Technol.* 17, 2 (2024), 95–107. https://doi.org/10.17671/gazibtd.1399077

[10] Sefik Ilkin Serengil and Alper Ozpinar. 2020. LightFace: A Hybrid Deep Face Recognition Framework. In *2020 Innov. Intell. Syst. Appl. Conf. ASYU*. IEEE, 23–27. https://doi.org/10.1109/ASYU50717.2020.9259802

[11] Harini Suresh and John Guttag. 2021. A Framework for Understanding Sources of Harm throughout the Machine Learning Life Cycle. In *Equity Access Algorithms Mech. Optim.* ACM, – NY USA, 1–9. https://doi.org/10.1145/3465416.3483305

[12] Yixin Wan, Arjun Subramonian, Anaelia Ovalle, Zongyu Lin, Ashima Suvarna, Christina Chance, Hritik Bansal, Rebecca Pattichis, and Kai-Wei Chang. 2024. Survey of Bias In Text-to-Image Generation: Definition, Evaluation, and Mitigation. https://doi.org/10.48550/arXiv.2404.01030 arXiv:2404.01030 [cs]