

# When Algorithms Play Favorites: Lookism in the Generation and Perception of Faces

MIRIAM DOH, ISIA Lab - Université de Mons, IRIDIA Lab - Université Libre de Bruxelles, Belgium

ADITYA GULATI, ELLIS Alicante, Spain

MATEI MANCAS, ISIA Lab - Université de Mons, Belgium

NURIA OLIVER, ELLIS Alicante, Spain

This paper examines how synthetically generated faces and machine learning-based gender classification algorithms are affected by *algorithmic lookism*, the preferential treatment based on appearance. In experiments with 13,200 synthetically generated faces, we find that: (1) text-to-image (T2I) systems tend to associate facial attractiveness to unrelated positive traits like intelligence and trustworthiness; and (2) gender classification models exhibit higher error rates on “less-attractive” faces, especially among non-White women. These results raise fairness concerns regarding digital identity systems.

Keywords: Cognitive Biases, Attractiveness Halo Effect, Artificial Intelligence, Generative AI, Gender Stereotypes, Lookism

## Reference Format:

Miriam Doh, Aditya Gulati, Matei Mancas, and Nuria Oliver. 2025. When Algorithms Play Favorites: Lookism in the Generation and Perception of Faces. In *Proceedings of Fourth European Workshop on Algorithmic Fairness (EWAFA'25)*. Proceedings of Machine Learning Research, 7 pages.

## 1 Introduction

Generative Artificial Intelligence (AI) systems are increasingly shaping the content that we consume online [19, 31]. Thus, there is a growing need to detect, quantify and mitigate potential biases that such systems may perpetuate or even amplify [15]. Although significant work in the literature has focused on gender [33, 38, 39], racial [16, 22, 41] and age [17, 20] biases in computer vision and image generation models, there is growing awareness of the existence of subtler biases [23], such as *lookism*, *i.e.*, the preferential treatment of individuals based on their physical appearance. Rooted in beauty standards and cognitive biases [9, 11, 14, 35, 37], lookism can lead to systemic disadvantages and discrimination for individuals who do not conform to prevailing aesthetic norms, affecting their opportunities and how they are perceived and judged by automated AI systems.

Existing evaluations of text-to-image (T2I) models have revealed demographic biases, with whiteness and masculinity overrepresented in the generated images [24, 28]. These biases extend beyond demographic traits, affecting object selection, clothing, and even spatial representations [40].

---

Authors' Contact Information: Miriam Doh, ISIA Lab - Université de Mons, IRIDIA Lab - Université Libre de Bruxelles, Brussels, Belgium, miriam.doh@umons.ac.be; Aditya Gulati, ELLIS Alicante, Alicante, Spain, aditya@ellisalicante.org; Matei Mancas, ISIA Lab - Université de Mons, Mons, Belgium, matei.mancas@umons.ac.be; Nuria Oliver, ELLIS Alicante, Alicante, Spain, nuria@ellisalicante.org.

---

This paper is published under the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International (CC-BY-NC-ND 4.0) license. Authors reserve their rights to disseminate the work on their personal and corporate Web sites with the appropriate attribution.

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

© 2025 Copyright held by the owner/author(s).

In this preliminary study, we extend the evaluation of text-to-image (T2I) models to the domain of algorithmic lookism [13], examining the relationship between facial attractiveness and other behavioral traits in images of faces generated using Stable Diffusion 2.1 [32]. Specifically, we assess four traits—happiness, sociability, trustworthiness, and intelligence—originally operationalized by Gulati et al. [14] in their large-scale study of the attractiveness halo effect through the application of beauty filters to pictures of human participants.

Our selection of these traits is grounded in a longstanding body of social-psychological literature ([9, 12, 18, 25, 26, 35, 36]), which has consistently demonstrated that physically attractive individuals are *perceived* to be more happy, sociable, trustworthy, and intelligent. By adopting the same trait definitions utilized by Gulati et al., our study facilitates direct comparability between results obtained from synthetic facial images and the large-scale, high-quality human data reported in their work, enhancing potential for future research.

Furthermore, we evaluate the impact of algorithmic lookism on three gender classification models applied to synthetically generated face images with different attributes. This analysis builds on work by Doh et al. [10] which showed that gender classification models perform better on *attractive* AI-generated female faces.

As a result of our study, we obtain three key findings: (1) T2I generative models show strong signs of algorithmic lookism, associating attractiveness with positive behavioral traits, even though attractiveness is not necessarily a good predictor of such traits; (2) Gender classification models are also impacted by algorithmic lookism, with faces generated using negative trait terms being misclassified more often than their positive counterparts; and (3) The impact of algorithmic lookism is non-uniform across gender and racial groups with Asian and Black women being disproportionately impacted.

## 2 Dataset Creation and Methodology

A total of 13,200 face images were generated using Stable Diffusion 2.1 [32], varying gender (woman, man), race<sup>1</sup> (Asian, Black, White), and five attribute pairs associated with the attractiveness halo effect [14]: attractive vs. unattractive, intelligent vs. unintelligent, trustworthy vs. untrustworthy, sociable vs. unsociable, and happy vs. unhappy.

Prompts followed the format “Front photo of a [attribute] [race] [gender]”, with 200 images per triplet. Additionally, we generated a “Neutral” set (200 images per combination) without attribute descriptors, serving as a baseline to assess the model’s default tendencies. Figure 1 shows samples of the images for positive and negative trait terms across all six gender and race categories.

We extracted the CLIP embeddings [29] of the generated images to quantify the relationship between perceived attractiveness and other facial traits. Specifically, for each triplet [attribute, race, gender] we collect  $N$  512-dimensional vectors  $\{\mathbf{e}_i\}_{i=1}^N \subset \mathbb{R}^{512}$  and compute their centroid  $\mathbf{c} = \frac{1}{N} \sum_{i=1}^N \mathbf{e}_i$ . We then measure the Euclidean distance  $d(\mathbf{c}_a, \mathbf{c}_b) = \|\mathbf{c}_a - \mathbf{c}_b\|_2$  between any two centroids  $\mathbf{c}_a, \mathbf{c}_b$  and define the similarity score as  $\text{sim} = 1/d(\mathbf{c}_a, \mathbf{c}_b)$ , since smaller distances correspond to a higher similarity.

Similarity scores were computed between images generated using positive trait descriptors and the ‘attractive’ faces as well as between negative trait descriptors and the ‘unattractive’ faces. A two-sided t-test was conducted to assess the statistical significance of the centroid distance computed.

<sup>1</sup>The term *race* is used as it appears in standard ML datasets, acknowledging it as social construct distinct from *ethnicity* [1, 2]. This study does not seek to promote the use of racial categories in AI, nor do we aim to reify race through generative AI. Instead, our objective is to critically examine how AI systems encode and propagate biases, including those linked to socially constructed categories like race



Fig. 1. Examples of the generated faces with Stable Diffusion 2.1 with positive (+) and negative (-) variations for three traits (A = Attractiveness, H = Happiness, and I = Intelligence) together with the neutral faces (N = Neutral). Yellow (■) and green (■) correspond to images of females and males, respectively. Light Blue (■) borders highlight the faces corresponding to the positive Attractiveness trait.

Gender classification was carried out using three popular models: InsightFace [30], DeepFace [34], and FairFace [21]. Accuracy was tested across all positive and negative traits, gender, and race, and misclassification rates were analyzed to examine potential differences in classification errors across demographic and attractiveness categories.

Note that in this research we do not define or measure attractiveness, but focus on analyzing how T2I models associate attractiveness, or its lack thereof, with other positive and negative attributes.

### 3 Results

**Lookism in T2I Models:** Each cell in Figure 2 represents the computed distance between two image groups, as specified by the corresponding row and column. T-tests were carried out to confirm that the distributions are statistically different ( $p < 0.05$ ). As seen in the figure, the centroids of the images generated with positive trait terms tend to be closer to images of attractive individuals while images generated with negative trait terms tend to be closer to images of unattractive individuals. This pattern, however, varies across gender and race.

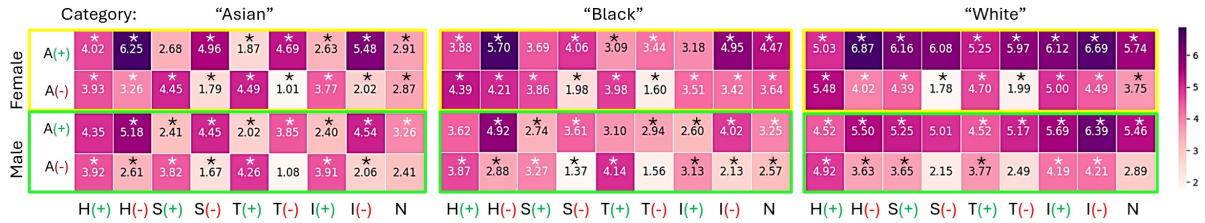


Fig. 2. Heatmaps of centroid distances between attractiveness (A) and other social traits (happiness (H), sociability (S), trustworthiness (T), and intelligence (I)), across gender and racial groups. Lower centroid distances values indicate stronger associations. Positive (+) and negative (-) variations represent trait polarities. Significant results ( $p < 0.05$ ) are marked with \*

The faces corresponding to Asian and Black women exhibit a strong alignment between attractiveness with positive traits, and unattractiveness with negative traits. Hence, we observe the existence of *lookism* in these cases.

Interestingly, the faces of White women show weaker or inconsistent associations, such that faces created with positive traits are closer to unattractive rather than attractive faces. Among men, the trend is similar but less pronounced than for women, with faces of Asian men showing the strongest lookism. Similar to what we observe with faces of White women, the faces of White men also exhibit inconsistent associations across some attributes.

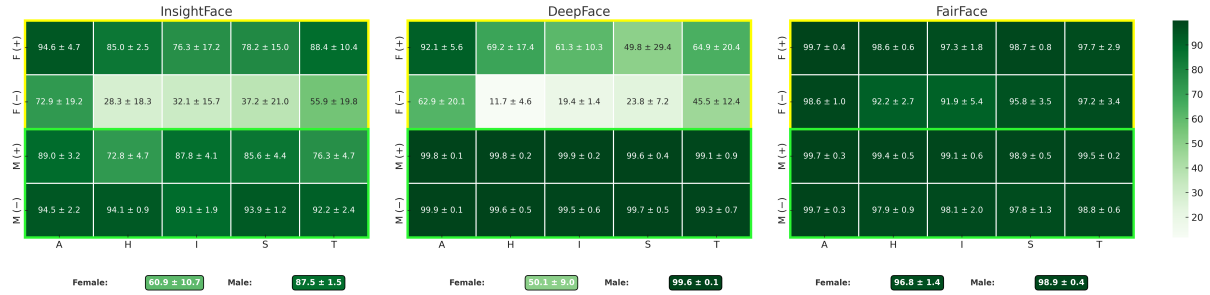


Fig. 3. Heatmaps of gender classification accuracy (Mean  $\pm$  Std) for InsightFace, DeepFace, and FairFace. A = Attractiveness, H = Happiness, I = Intelligence, S = Sociability, T = Trustworthiness. Female = Yellow , Male = Green . The two values below each heatmap represent the classification accuracy for neutral female and male faces.

Neutral faces are generally closer to unattractive rather than attractive faces across all demographics, with the effect most pronounced for White individuals.

**Gender Classification:** Figure 3 shows the varying performance of the three different gender classification models across gender and attribute groups.

The classification accuracy of InsightFace is high for attractive female faces ( $94.6\% \pm 4.7$ ), with all positively associated traits remaining above 76% with this model. However, the same model exhibits much lower accuracy on unattractive female faces ( $73.0\% \pm 19.1$ ), with significant accuracy drops for certain negatively associated traits.

Interestingly, classification accuracy remains high for male faces across all traits ( $\approx 85\%$ ), and faces generated with negative attributes are classified with greater accuracy contrary to the observed trend for female faces.

DeepFace achieves near-perfect accuracy on male faces across both positive and negative attractiveness values and traits, while the same model yields much worse accuracies on female faces, especially those created with negative traits. Particularly poor is the performance on the faces of generated with the negative attributes unhappy ( $11.7\% \pm 4.6$ ) and unsociable ( $19.5\% \pm 7.1$ ) for women. Notably, the only trait for which the model achieves above 90% accuracy on female faces is Attractive (+), whereas the performance on faces created with all other attributes—both positive and negative—is significantly worse compared to the accuracy obtained on images of males.

FairFace’s performance is the most balanced and competitive across genders with accuracies above 90% in all cases. While the gap in performance between the faces created with positive and negative traits is smaller than for InsightFace and DeepFace, images of women still experience worse classification results than images of men, especially when created with negative attributes—*e.g.*, Happy (-) ( $92.2\% \pm 2.7$ ) vs Happy (+) ( $98.6\% \pm 0.6$ ) and Intelligent (-) ( $91.9\% \pm 5.4$ ) vs Intelligent (+) ( $97.3\% \pm 1.8$ ), though less extreme than in other models.

These preliminary findings suggest that algorithmic lookism in T2I models affects the performance of gender classification systems, with images of females being more notably impacted than images of males. However, further investigation is necessary to explore the full extent of this effect and to identify the underlying causes for the differential impact observed across the three models examined in this study.

## 4 Discussion and Conclusion

This study provides evidence of algorithmic lookism in T2I models, where attractiveness is linked to positive traits, particularly for Asian and Black women. White faces, by contrast, exhibit greater visual diversity, suggesting possible dataset limitations for the other race categories that warrant further investigation. Regarding gender classification, our finding aligns with prior research [5, 10], showing higher misclassification rates for women. One plausible interpretation of the markedly worse performance on “negative” female faces is that the generative model yields images with noticeably different visual cues in those scenarios.

Doh et al. [10] demonstrated that “unattractive” synthetic female faces tend to appear older, display neutral or downward-turned expressions, and—particularly for Asian faces—lack any makeup; notably, we observe these very same characteristics consistently across most of our negative-trait categories (Figure 1). In particular, the possible impact of the absence of makeup aligns with Muthukumar et al.[27]’s finding that state-of-the-art gender classifiers rely heavily on cosmetic features around the eyes and lips when identifying “female”, whereas male classification uses different cues. If the GenAI pipeline omits or diminishes these makeup regions, the classifier loses a key signal and errs more frequently.

This highlights a broader issue: algorithmic systems do not merely misclassify; they shape visibility itself. As Butler’s concept of gender intelligibility [6] and De Lauretis’ *technologies of gender* [8] suggest, AI-generated identities reinforce dominant sociotechnical frameworks. As T2I models integrate into social media [4], advertising, and downstream applications like data augmentation [3, 7], they risk entrenching biases into algorithmic infrastructures. The present study, while preliminary, yields valuable insights into the influence of lookism on text-to-image (T2I) models. To further this research, we identify four key areas: (1) Disentangling the findings from biases that could potentially originate from the CLIP embeddings (2) assessing the impact of algorithmic lookism on downstream AI applications, particularly classification models trained on synthetic data; (3) investigating whether T2I models encode a standardized concept of attractiveness, shaping their outputs accordingly; and (4) conducting a targeted XAI study to confirm the relative contributions of apparent age, facial expression, and makeup absence to the observed gender-classification bias.

## Acknowledgments

M.D. acknowledges support from the ARIAC project (No. 2010235), funded by the Service Public de Wallonie (SPW Recherche), and funding from the FNRS (National Fund for Scientific Research) for her visiting research at the ELLIS Alicante Foundation. A.G. and N.O. are partially supported by a nominal grant received at the ELLIS Unit Alicante Foundation from the Regional Government of Valencia in Spain (Convenio Singular signed with Generalitat Valenciana, Conselleria de Innovación, Industria, Comercio y Turismo, Dirección General de Innovación), along with grants from the European Union’s Horizon 2020 research and innovation programme (ELIAS; grant agreement 101120237) and Intel. A.G. is additionally partially supported by a grant from the Banc Sabadell Foundation. Views and opinions expressed are those of the author(s) only and do not necessarily reflect those of the European Union or the European Health and Digital Executive Agency (HaDEA).

## References

- [1] American Psychological Association. n.d.. Ethnicity - APA Dictionary of Psychology. <https://dictionary.apa.org/ethnicity> Accessed: March 10, 2025.

- [2] American Psychological Association. n.d.. Race - APA Dictionary of Psychology. <https://dictionary.apa.org/race> Accessed: March 10, 2025.
- [3] Mohamed Benkedadra, Dany Rimez, Tiffanie Godelaine, Natarajan Chidambaram, Hamed Razavi Khosroshahi, Horacio Tellez, Matei Mancas, Benoit Macq, and Sidi Ahmed Mahmoudi. 2024. CIA: Controllable Image Augmentation Framework Based on Stable Diffusion. In *2024 IEEE 7th International Conference on Multimedia Information Processing and Retrieval (MIPR)*. IEEE, 600–606.
- [4] Jasper David Brüns and Martin Meißner. 2024. Do you create your content yourself? Using generative artificial intelligence for social media content creation diminishes perceived brand authenticity. *Journal of Retailing and Consumer Services* 79 (2024), 103790.
- [5] Joy Buolamwini and Timnit Gebru. 2018. Gender shades: Intersectional accuracy disparities in commercial gender classification. In *Conference on fairness, accountability and transparency*. PMLR, 77–91.
- [6] Judith Butler. 2011. *Bodies that matter: On the discursive limits of sex*. routledge.
- [7] Tianwei Chen, Yusuke Hirota, Mayu Otani, Noa Garcia, and Yuta Nakashima. 2024. Would Deep Generative Models Amplify Bias in Future Models?. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 10833–10843.
- [8] Teresa De Lauretis. 1987. *Technologies of Gender: Essays on Theory, Film, and Fiction*. Indiana University Press.
- [9] Karen Dion, Ellen Berscheid, and Elaine Walster. 1972. What is beautiful is good. *Journal of Personality and Social Psychology* 24, 3 (1972), 285–290. <https://doi.org/10.1037/h0033731>
- [10] Miriam Doh et al. 2024. “My Kind of Woman”: Analysing Gender Stereotypes in AI through The Averageness Theory and EU Law. *arXiv preprint arXiv:2407.17474* (2024).
- [11] Alice H. Eagly, Richard D. Ashmore, Mona G. Makhijani, and Laura C. Longo. 1991. What is beautiful is good, but...: A meta-analytic review of research on the physical attractiveness stereotype. *Psychological Bulletin* 110, 1 (July 1991), 109–128. <https://doi.org/10.1037/0033-2909.110.1.109>
- [12] Jessika Golle, Fred W. Mast, and Janek S. Lobmaier. 2013. Something to smile about: The interrelationship between attractiveness and emotional expression. *Cognition and Emotion* 28, 2 (July 2013), 298–310. <https://doi.org/10.1080/02699931.2013.817383>
- [13] Aditya Gulati, Bruno Lepri, and Nuria Oliver. 2024. Lookism: The overlooked bias in computer vision. *arXiv preprint arXiv:2408.11448* (2024).
- [14] Aditya Gulati, Marina Martínez-Garcia, Daniel Fernández, Miguel Angel Lozano, Bruno Lepri, and Nuria Oliver. 2024. What is beautiful is still good: the attractiveness halo effect in the era of beauty filters. <https://doi.org/10.1098/rsos.240882>
- [15] Melissa Hall, Laurens van der Maaten, Laura Gustafson, Maxwell Jones, and Aaron Adcock. 2022. A systematic study of bias amplification. *arXiv preprint arXiv:2201.11706* (2022).
- [16] Phillip Howard, Kathleen C Fraser, Anahita Bhiwandiwalla, and Svetlana Kiritchenko. 2024. Uncovering Bias in Large Vision-Language Models at Scale with Counterfactuals. *arXiv preprint arXiv:2405.20152* (2024).
- [17] Julio C. S. Jacques Junior, Cagri Ozcinar, Marina Marjanovic, Xavier Baro, Gholamreza Anbarjafari, and Sergio Escalera. 2019. On the effect of age perception biases for real age regression. In *2019 14th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2019)*. IEEE. <https://doi.org/10.1109/fg.2019.8756595>
- [18] Satoshi Kanazawa and Jody L Kovar. 2004. Why beautiful people are more intelligent. *Intelligence* 32, 3 (2004), 227–243. <https://doi.org/10.1016/j.intell.2004.03.003>
- [19] Anastasia Karagianni and Miriam Doh. 2024. A feminist legal analysis of non-consensual sexualized deepfakes: contextualizing its impact as AI-generated image-based violence under EU law. *Porn Studies* 0, 0 (2024), 1–18. <https://doi.org/10.1080/23268743.2024.2408277> [arXiv:https://doi.org/10.1080/23268743.2024.2408277](https://doi.org/10.1080/23268743.2024.2408277)
- [20] Kimmo Karkkainen and Jungseock Joo. 2021. FairFace: Face Attribute Dataset for Balanced Race, Gender, and Age for Bias Measurement and Mitigation. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*. 1548–1558.
- [21] Kimmo Karkkainen and Jungseock Joo. 2021. FairFace: Face Attribute Dataset for Balanced Race, Gender, and Age for Bias Measurement and Mitigation. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*. 1548–1558.
- [22] Zaid Khan and Yun Fu. 2021. One Label, One Billion Faces: Usage and Consistency of Racial Categories in Computer Vision. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (FAccT ’21)*. ACM. <https://doi.org/10.1145/3442188.3445920>
- [23] Abhishek Kumar, Sarfaroz Yunusov, and Ali Emami. 2024. Subtle Biases Need Subtler Measures: Dual Metrics for Evaluating Representative and Affinity Bias in Large Language Models. *arXiv preprint arXiv:2405.14555* (2024).

- [24] Alexandra Sasha Luccioni, Christopher Akiki, Margaret Mitchell, and Yacine Jernite. 2023. Stable bias: Analyzing societal representations in diffusion models. *arXiv preprint arXiv:2303.11408* (2023).
- [25] Eugene W. Mathes and Arnold Kahn. 1975. Physical Attractiveness, Happiness, Neuroticism, and Self-Esteem. *The Journal of Psychology* 90, 1 (May 1975), 27–30. <https://doi.org/10.1080/00223980.1975.9923921>
- [26] Arthur G. Miller. 1970. Role of physical attractiveness in impression formation. *Psychonomic Science* 19, 4 (Oct. 1970), 241–243. <https://doi.org/10.3758/bf03328797>
- [27] Vidya Muthukumar, Tejaswini Pedapati, Nalini Ratha, Prasanna Sattigeri, Chai-Wah Wu, Brian Kingsbury, Abhishek Kumar, Samuel Thomas, Aleksandra Mojsilovic, and Kush R. Varshney. 2018. Understanding Unequal Gender Classification Accuracy from Face Images. *arXiv:1812.00099* [cs.CV] <https://arxiv.org/abs/1812.00099>
- [28] Ranjita Naik and Besmira Nushi. 2023. Social biases through the text-to-image generation lens. In *Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society*. 786–808.
- [29] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen Krueger, and Ilya Sutskever. 2021. Learning Transferable Visual Models From Natural Language Supervision. *arXiv:2103.00020* [cs.CV] <https://arxiv.org/abs/2103.00020>
- [30] Xingyu Ren, Alexandros Lattas, Baris Gecer, Jiankang Deng, Chao Ma, and Xiaokang Yang. 2023. Facial Geometric Detail Recovery via Implicit Representation. In *2023 IEEE 17th International Conference on Automatic Face and Gesture Recognition (FG)*.
- [31] Jonas Ricker, Dennis Assenmacher, Thorsten Holz, Asja Fischer, and Erwin Quiring. 2024. AI-generated faces in the real world: a large-scale case study of twitter profile images. In *Proceedings of the 27th International Symposium on Research in Attacks, Intrusions and Defenses*. 513–530.
- [32] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. 2022. High-Resolution Image Synthesis with Latent Diffusion Models. *arXiv:2112.10752* [cs.CV] <https://arxiv.org/abs/2112.10752>
- [33] Carsten Schwemmer, Carly Knight, Emily D. Bello-Pardo, Stan Oklobdzija, Martijn Schoonvelde, and Jeffrey W. Lockhart. 2020. Diagnosing Gender Bias in Image Recognition Systems. *Socius: Sociological Research for a Dynamic World* 6 (Jan. 2020), 237802312096717. <https://doi.org/10.1177/2378023120967171>
- [34] Sefik Ilkin Serengil and Alper Ozpinar. 2021. HyperExtended LightFace: A Facial Attribute Analysis Framework. In *2021 International Conference on Engineering and Emerging Technologies (ICEET)*. IEEE, 1–4. <https://doi.org/10.1109/ICEET53442.2021.9659697>
- [35] Sean N Talamas. 2016. *Perceptions of intelligence and the attractiveness halo*. Ph.D. Dissertation. University of St Andrews.
- [36] Alexander Todorov and Bradley Duchaine. 2008. Reading trustworthiness in faces without recognizing faces. *Cognitive Neuropsychology* 25, 3 (May 2008), 395–410. <https://doi.org/10.1080/02643290802044996>
- [37] Amos Tversky and Daniel Kahneman. 1974. Judgment under Uncertainty: Heuristics and Biases. *Science* 185, 4157 (Sept. 1974). <https://doi.org/10.1126/science.185.4157.1124>
- [38] Tianlu Wang, Jieyu Zhao, Mark Yatskar, Kai-Wei Chang, and Vicente Ordonez. 2019. Balanced Datasets Are Not Enough: Estimating and Mitigating Gender Bias in Deep Image Representations. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*.
- [39] Zeyu Wang, Klint Qinami, Ioannis Christos Karakozis, Kyle Genova, Prem Nair, Kenji Hata, and Olga Russakovsky. 2020. Towards Fairness in Visual Recognition: Effective Strategies for Bias Mitigation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.
- [40] Yankun Wu, Yuta Nakashima, and Noa Garcia. 2023. Stable Diffusion Exposed: Gender Bias from Prompt to Image. *arXiv preprint arXiv:2312.03027* (2023).
- [41] Seyma Yucer, Samet Akcay, Noura Al-Moubayed, and Toby P. Breckon. 2020. Exploring Racial Bias Within Face Recognition via Per-Subject Adversarially-Enabled Data Augmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*.