

We sincerely thank the ethical meta reviewer for their thoughtful feedback. Owing to the reviewers suggestions, we have made the following changes in the camera-ready version of our paper.

- 1. Annotation Instructions:** We provide the details of the instructions given to the human annotators in Appendix C. Additionally, we include the screenshot of the interface in Figure 2. In the original version, we did include a *Cannot judge* option for the case where the annotators cannot recognize the gender. However, we had not considered the uncertainty for skin color. Following the suggestions, we updated the interface, re-ran the annotation process and have obtained new annotation results. Based on the new results, we count the number of images for which at least 2 out of 3 annotators choose *Cannot judge* (either due to blurry images or both features being visible) option in terms of gender and skin colors. We find that the proportion is small, around 6.67%.
- 2. Individual Annotations:** As promised during the rebuttal phase, we have open-sourced our annotation results for MinDALL.E, DALL.E-mini and Stable Diffusion at https://github.com/Hritikbansal/entigen_emnlp. We shall actively resolve all the issues researchers encounter on our repo. We hope that our transparency of individual annotations helps in further studying the variance amongst the annotators.
- 3. Binary Categories:** In the revised version of our experiments, we account for uncertainty in assigning a gender or skin color through CLIP model as well as Humans. Section 3.3 clarifies this further in the main text. In addition to the human interface that clearly prevents the annotators from assigning any gender or skin color whenever they are uncertain, Table 4 in Appendix C contains a new prompt “T_uncertain” for CLIP based evaluation. We understand that despite all these changes, we operationalize our evaluation metrics for the binary categories and hence, we have added detailed discussion on that in the last paragraph of the Limitations section and parts of the Ethical statement.
- 4. Race and Skin Color:** As mentioned in the rebuttal phase, we totally agreed with the differences in studying race and skin color. Hence, we have positioned the revised paper towards studying skin color and have refrained from any racial associations based on the generated images. We re-ran all our experiments with *skin color* as a social axis. All our tables and graphs in the main text mention those results. After considering uncertain situations and replacing *race* with *skin color* in ethical interventions, we find that our conclusions still hold. Our results

indicate that ethical interventions do lead to reduction in the score (Table 1). We provide detailed discussion of our observations in Section 3.

5. **Citations on Culture:** We have cited three papers ([1],[2],[3]) that support the assumptions we make in our work in Section 2.1.
6. **Ethics:** We discussed the harms of generating images in general as well as the kind of work we are doing, as well as potential solutions in the third paragraph of our Ethics Statement.

References:

[1] Bell CM. Ritual: Perspectives and dimensions. Oxford University Press on Demand; 1997.

[2] Xu L, Xu M. Comparison on Wedding Culture between China and Western Countries. In 8th International Conference on Education, Management, Computer and Society 2018 (pp. 423-426).

[3] Acharya A, Talamadupula K, Finlayson MA. An atlas of cultural commonsense for machine reasoning. In AAAI Conference on Artificial Intelligence 2021 Feb 2.
