

Dear Reviewers,

Thank you very much for the helpful feedback. We revised the paper in the following way:

- To make the Preliminaries section clearer, we added a short running example in Sections 2.1 and 2.2 as proposed by the reviewer.
- Concerning the question about the second post-hoc query regarding counterfactual fairness for specific subgroups, we added the formal definition of this query in Section 4.4.
- We added a paragraph in section 4.2 regarding the interplay between satisfaction level of our counterfactual fairness axioms and its original definition (Eq. 1 in our paper).
- We also reviewed the paper again in order to further improve style and flow.
- In particular, we have accepted and integrated the following comments and corrections:

(text below appearing [*in brackets*] is a citation of the original review)

1. [*“The paper by Wagner and d’Avila Garcez (2021) has recently inaugurated” --> this phrasing should be updated, as the work is no longer recent.*] We have updated this phrasing. Thank you for pointing that out.
2. [*“We integrate counterfactual fairness by estimating counterfactuals examples via current methods” --> perhaps clarify “current methods for counterfactuals.”*] We have added a citation here to clarify that we employ a contribution by Javaloy et al., which is the state of the art as far as we know.
3. [*“can further devide” --> should be “can further divide.”*] It should; thank you.
4. [*“this query could logically formulates as:” --> should be “this query could be logically formulated as.”*] It should; thank you.