

## Explanation of revisions

We thank the reviewers and the meta-reviewer for their feedback. We report below on our revisions, following all the suggestions of the meta review, to address the concerns raised in the reviews while maintaining its strengths and contributions.

**Experimental setup / framing.** The previous reviewing cycle considered the absence of non-Western languages as a major shortcoming of the paper as framed. One possible remedy would be to add a wider range of languages to the experimental setup.

At the same time, our extensive experiments and analyses were recognized as a core strength with our overarching goal to provide in-depth insights into cross-lingual opinion transfer in multilingual large language models. Unfortunately, adding a wider range of languages conflicts with the goal of analytical depth, given the strict ARR paper length limit.

The meta-review explicitly mentions two ways to improve the paper: either by adding more languages, or by re-framing the paper as a Western-focused investigation. In order to maintain the analytical depth in the revised version, we did not add new languages, but instead choose the reframing option. This limitation is now stated in the title, emphasized throughout the main body of the paper, and discussed in even more detail in the limitations section.

More specifically, we see the following challenges in extending our original experimental setup:

1. The evaluation data is rooted in Western European political discourse, and all evaluated languages are spoken primarily in Western contexts.
2. Introducing non-Western languages would introduce significant confounding variables, especially due to differing political and cultural environments, which would require deeper contextual analysis beyond the paper’s scope.
3. Some of the models used in our experiments are not officially available in other languages. We must be sure that we get valid responses from the models.
4. The data is currently unavailable in non-Western languages, and high-quality translations would require native speakers for validation.

We include these reasons both in the method section and in the limitation section of the paper to justify our decision. We strongly believe that our original experimental setup make a worthwhile scientific contribution to the study of multilingual LLMs.

**Choice of LLMs.** We did not align or fine-tune additional models, as the results for RQ2 are highly consistent across the two models we examine. For RQ1, the findings are likewise robust and show similar trends across models. Given this consistency, we do not anticipate substantive changes in further alignment experiments and prioritize analytical depth over breadth.

**Presentation changes.** We have moved the example of the evaluation task into the main body of the paper for better accessibility and clarity. We have also implemented all suggested wording changes from the reviews both the text and the figures. In particular, we have reworded the section on the beta regression analysis to improve clarity. We replaced the term “more conservative” with “less left-leaning” to avoid misinterpretation, especially considering the overall left-leaning tendencies observed in the results.