

## SUMMARY OF CHANGES

# Demystifying Representation Spaces of Multilingual and Multimodal Aspects in Large Audio Language Models

1. We revised the title from “Demystifying Representation Alignment...” to “Demystifying Representation Spaces...” in response to earlier feedback regarding the misleading implication that good alignment necessarily entails good clustering. The new title better reflects the paper’s core claim, which is focused specifically on how representation spaces form clusters and align, not on multimodal or multilingual alignment more broadly.
2. In the abstract, we modified the final sentence to better emphasize our claim regarding the clustering behavior of representations.
3. In the Introduction, we revised the wording in the list of contributions to ensure clarity and alignment with the paper’s focus.
4. In Section 3, we added a brief justification for our choice of latent mapping technique, to clarify our methodological reasoning.
5. We revised the Discussion section based on the feedback we received, particularly concerning alignment failures, and also added commentary on broader implications and potential impact.
6. We included several new appendices that I believe will be useful to readers. These additions aim to provide supporting materials for our analysis and clarify implementation details.