

## Explanation of Revisions

We thank the reviewers and area chair for their valuable feedback. We have addressed the major concerns and improved our paper as follows:

1. Translation and Data Quality:
  - Added detailed description of translation using GPT-4o-mini and Gemini-2.5-flash combined with manual native speaker review.
  - Provided examples and described quality control measures including filtering, back-translation validation, and manual checks.
2. Experimental Methodology:
  - Included exact hyperparameters for each model-task pair, early stopping criteria, and number of runs for reproducibility.
  - Clarified evaluation metrics and rationale for their selection.
3. Model Evaluation Scope:
  - Justified focus on BERT-like models as state-of-the-art for Nepali NLU and publicly available models.
  - Highlighted future work to explore other model families.
4. Analysis and Presentation:
  - Expanded analysis on model performance challenges with small datasets and complex tasks.
  - Added a formal conclusion summarizing key findings and future directions.
5. Reproducibility Commitment:
  - Committed to releasing code and data under permissive licenses with clear documentation.
6. Minor Improvements:
  - Corrected all typographical issues and improved tables and figures.