

Responsible NLP Checklist

Paper title: *One Pair Suffices: Unlocking Universal Zero-Shot Translation via Cross-Architecture Alignment*

Authors: *Hao Zong, CongHuYuan, Chao Bei, Wentao Chen, Huan Liu, Kaiyu Huang, Degen Huang*

How to read the checklist symbols:

- the authors responded 'yes'
- the authors responded 'no'
- the authors indicated that the question does not apply to their work
- the authors did not respond to the checkbox question

For background on the checklist and guidance provided to the authors, see the [Responsible NLP Checklist](#) page at ACL Rolling Review.

A. Questions mandatory for all submissions.

A1. Did you describe the limitations of your work?
This paper has a Limitations section.

A2. Did you discuss any potential risks of your work?
(left blank)

B. Did you use or create scientific artifacts? (e.g. code, datasets, models)

B4. Did you discuss the steps taken to check whether the data that was collected/used contains any information that names or uniquely identifies individual people or offensive content, and the steps taken to protect/anonymize it?
We did not collect new raw data from the web. All training and evaluation datasets used in this work (NLLB corpus, NTREX, and FLORES-200) are standard, widely-adopted, and publicly available open-source benchmarks that have been previously vetted and anonymized by the research community (Costa-juss et al., 2022).

B6. Did you report relevant statistics like the number of examples, details of train/test/dev splits, etc. for the data that you used/created?
Relevant dataset statistics, including the specific languages, data sources, and the exact number of sentence pairs used for training and testing splits, are thoroughly reported in Section 4.1 (Datasets and Evaluation Protocols) and Table 1.

C. Did you run computational experiments?

C2. Did you discuss the experimental setup, including hyperparameter search and best-found hyperparameter values?
We thoroughly discussed our experimental setup, including the hardware used (NVIDIA H200 GPUs), learning rates, batch sizes, optimizer choices, and early stopping strategies. Please refer to Section 4.4 (Implementation Details) and Section 5 (Training Protocol).

C3. Did you report descriptive statistics about your results (e.g., error bars around results, summary statistics from sets of experiments), and is it transparent whether you are reporting the max, mean, etc. or just a single run?
Due to the massive scale of the models involved (up to 13B parameters for baselines and 5.25B

The Responsible NLP Checklist used at ACL Rolling Review is adopted from NAACL 2022, with the addition of ACL 2023 question on AI writing assistance and further refinements based on ARR practice. ACL 2026 used a subset of ARR checklist form.

for our HCA framework), running multiple trials with different random seeds to report error bars was computationally prohibitive. However, we ensure robustness by reporting summary statistics (averages across high-resource and low-resource language groups) across dozens of languages to clearly demonstrate generalized trends.

D. Did you use human annotators (e.g., crowdworkers) or research with human subjects?

D1. Did you report the full text of instructions given to participants, including e.g., screenshots, disclaimers of any risks to participants or annotators, etc.?

Not applicable. We did not conduct research involving human subjects or human annotators; all evaluations rely on standardized automatic metrics (COMET, chrF++).

D2. Did you report information about how you recruited (e.g., crowdsourcing platform, students) and paid participants, and discuss if such payment is adequate given the participants' demographic (e.g., country of residence)?

(left blank)

D3. Did you discuss whether and how consent was obtained from people whose data you're using/curating (e.g., did your instructions explain how the data would be used)?

(left blank)

D4. Was the data collection protocol approved (or determined exempt) by an ethics review board?

(left blank)

E. Did you use AI assistants (e.g., ChatGPT, Copilot) in your research, coding, or writing?

E1. If you used AI assistants, did you include information about their use?

We used AI assistants (e.g., ChatGPT/Claude) exclusively for proofreading, grammatical correction, and polishing the English phrasing of the manuscript. No AI assistants were used to generate experimental data, formulate the core scientific hypotheses, or write the original technical content.