Responsible NLP Checklist

Paper title: *Where the Cat Sat: A Multilingual Framework for Spatial Language Understanding*
Authors: *Demian Inostroza, Ekaterina Vylomova, Charles Kemp, Mae Carroll, Wanchun Li, Meladel Mistica*

How to read the checklist symbols:

☑ the authors responded 'yes'

☒ the authors responded 'no'

N/A 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.

N/A 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)**

N/A 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?
(left blank)

☑ 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?
Section 3.4 (Resulting Gold Standard): 189 sentences across three languages, 63 unique stasis dynamicity combinations per language.

☑ **C. Did you run computational experiments?**

☑ C2. Did you discuss the experimental setup, including hyperparameter search and best-found hyperparameter values?
Section 4.3 (Inference Configuration): temperature 0.2, maximum token limit 8,192, three independent runs per model-language pair via OpenRouter API.

☑ 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?
Appendix (Tables section): all results are reported with standard deviation across three independent runs (e.g., Tables 5, 6, 7, 8, 9, 10).

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.

☑ **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.?
Appendix B (Annotation Guidelines): full operational definitions and instructions given to annotators are provided for all six components (Figure, Ground, Spatial Predicate, Spatial Markers, Dynamicity, Stasis).

☑ 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)?
native speakers were recruited through personal networks and volunteered without financial compensation for this brief linguistic elicitation task.

☑ 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)?
Section 7 (Ethical Considerations): all participants provided informed consent and were informed that their anonymized sentence productions would be used for NLP research and made publicly available as part of a benchmark dataset.

☑ D4. Was the data collection protocol approved (or determined exempt) by an ethics review board?
Section 7 (Ethical Considerations): the research was approved by the University of Melbourne Human Research Ethics Committee.

☑ **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?
AI assistants were used to improve code efficiency and to rephrase structures from Spanish to English during the writing process.