Responsible NLP Checklist

Paper title: *Leveraging Label Semantics and Entity Description Generation for LLM-based Fine-grained Entity Typing*
Authors: *Yingying Ma, Hongliang Dai, Xia Zhang, Piji Li*

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?
Our work focuses on a standard sequence labeling task using public benchmarks and does not involve sensitive data or high-risk applications.

☑ **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?
Our experiments are conducted on OntoNotes and BBN, which are standard, widely-used academic benchmarks in the NLP community. These datasets have been pre-processed and vetted by their original creators to ensure they do not contain personally identifying information or offensive content.

☑ 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?
Yes, we report the detailed statistics of the datasets used, including the number of entity types and the sizes of train/dev/test splits in Section 4.1 (Experimental Setup) 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 provide a comprehensive description of our experimental setup, including the hardware environment and the best-found hyperparameter values for LoRA fine-tuning, in Section 4 and and Appendix.

☑ 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?
Yes, we report the mean and standard deviation across five independent runs for all main results. These detailed statistics are reported in Appendix D(Table 8) to ensure transparency and demonstrate the stability of our method.

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?**

☒ N/A 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.?
(left blank)

☒ N/A 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)

☒ N/A 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)

☒ N/A 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?
AI assistants (e.g., ChatGPT/Gemini) were used to assist in polishing the linguistic quality and refining the LaTeX formatting of the manuscript. The core research ideas, experimental designs, and data analysis were conducted solely by the authors.