

Responsible NLP Checklist

Paper title: *UCS: Estimating Unseen Coverage for Improved In-Context Learning*

Authors: *Jiayi Xin, Xiang Li, Evan Qiang, Weiqing He, Tianqi Shang, Weijie J Su, Qi Long*

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?

We do not see a direct potential risk from this work because it is fundamental machine learning research on improving in-context learning for large language models. It does not introduce a high-stakes application or deploy a system that directly affects users, so any potential harms are indirect and minimal.

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 include a dedicated discussion of personally identifying or offensive content because our study does not involve human-subject data or newly collected personal data. The experiments use established public NLP benchmarks for intent classification and reasoning, which are standard research datasets and were not used to identify real individuals.

- 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 4, Section5, Section6

C. Did you run computational experiments?

- C2. Did you discuss the experimental setup, including hyperparameter search and best-found hyperparameter values?

Appendix D

- 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?

Section 4, Section5, Section6

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.?

Not applicable since we did not use human annotators or research with human subjects in this work.

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)?

Not applicable since we did not do any recruitment/payment.

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)?

Not applicable since we used public dataset.

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

Not applicable since no human subjects are involved.

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 utilized large language models (specifically ChatGPT and Gemini) for polishing paper and literature search. Paper polishing by LLMs was exclusively for sentence-level editing and grammatical error correction. All cited references were manually verified by the authors to ensure accuracy.