

## Responsible NLP Checklist

Paper title: *Bridging the Memorization-Utilization Gap: Near-Lossless Context Compression via Reinforcement Learning*

Authors: *Yujan Ting, Xu Tang, Terrence Chen, Weijing Huang*

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.*

- A2. Did you discuss any potential risks of your work?

*Our work proposes an efficiency-focused context compression technique that reduces computational cost and improves long context utilization. As our method compresses existing context, it limits the risk surface to information fidelity rather than introducing new risk-relevant capabilities. Accordingly, the primary potential risk shared by compression-based methods is information loss or reduced faithfulness in downstream use. We mitigate this through outcome-based reinforcement learning, which encourages faithful utilization and demonstrates near-lossless recovery in our evaluations. While residual failure modes may remain, we believe the approach is more likely to mitigate than amplify risks associated with the faithful utilization of compressed inputs. Because these issues are treated as core technical objectives and evaluated empirically throughout the paper, we do not include a separate risk discussion.*

### 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 use publicly available research corpora widely adopted in the NLP community (HotpotQA, TriviaQA, Natural Questions, DCLM, etc.). These datasets are derived from public web sources or standardized benchmarks and have been curated or filtered for general quality by their original authors. As our work focuses on compression methodology and downstream utilization rather than data collection or curation, we do not include an explicit discussion of this concern in the paper.*

- 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, Appendix M - Evaluation Setup*

### C. Did you run computational experiments?

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

*Appendix O - Training Hyperparameters*

---

*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.*

- 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?

*We report single-run results without error bars due to computational constraints in training 7B and 32B models through multiple stages (continual learning, instruction tuning, RL) and evaluating across contexts of up to 120K tokens. We note that our findings show consistent trends across model scales, context lengths, and evaluation benchmarks, providing indirect evidence of robustness. For latency measurements, where variance is a primary concern and runs are inexpensive, we report statistics over 20 runs (Appendix J).*

**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.?

*This question is not applicable to our work. We did not involve any human participants in this study.*

- 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)?

*This question is not applicable to our work. All training and evaluation data are sourced from existing public datasets without human recruitment.*

- 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)?

*This question is not applicable to our work. All datasets used are publicly released for research purposes, and no new data were collected from human subjects.*

- D4. Was the data collection protocol approved (or determined exempt) by an ethics review board?
- This question is not applicable to our work. Our research does not involve human subjects and therefore does not require ethics board approval.*

**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 Claude Code and ChatGPT to help debug code, improve phrasing, and check grammar. We did not include this information in the paper because these tools were used only to assist with implementation and refine wording. All research ideas, experimental designs, and results analysis were composed by the authors.*