

Responsible NLP Checklist

Paper title: *CondenseFlow: Scalable Latent Space Collaboration via Semantic Compression for Multi-Agent Systems*

Authors: *xiaoyu chen, Fengge Wu, Zhao Junsuo, Yun Fan*

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?

This work uses only publicly available mathematical reasoning benchmarks (AIME 2024/2025, HMMT 2025), scientific QA datasets (GPQA-Diamond, MedQA), and code generation benchmarks (MBPP-Plus, LiveCodeBench). None of these datasets 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?

Section 4.1 reports the number of benchmarks (7), dataset categories, and model configurations. Training data statistics (1,000 problems from GSM8K and MBPP training splits) are reported in Appendix C.3.

C. Did you run computational experiments?

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

Section 4.1 describes the experimental setup including evaluation protocols (Standard Protocol and Stress Test Protocol). Complete hyperparameter settings are provided in Appendix C.2 (Table 6, training hyperparameters) and Appendix D.1 (Table 7, inference hyperparameters).

- 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?

Tables 1 and 2 report mean standard deviation over 5 runs for all model-task combinations. Section 4.2 and 4.3 report average accuracy and percentage gaps () between methods.

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.?

This work does not involve human annotators or human subjects. All evaluations are conducted on existing benchmarks using automated metrics.

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)?

No human participants were recruited or compensated in this study.

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 work only uses publicly available benchmark datasets. No new data collection from human subjects was conducted.

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

No human subjects research was conducted. Ethics review board approval was not required.

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?

The use of AI assistants is disclosed in the Limitations section of the paper: "We used AI assistants for code debugging, grammar checking, and manuscript proofreading. All scientific contributions including methodology design, theoretical analysis, experimental setup, and result interpretation were conducted entirely by the authors."