

Responsible NLP Checklist

Paper title: *NeuralFSM: Adaptive Multi-Agent Coordination via Learning Finite-State Execution Policy*

Authors: *Jiye Wang, Yu Wang, Jianbin Li, Shiduo Yang, Kenan Guo, Yuanhe Zhao*

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 are confident that there are no ethical issues with this paper in terms of motivation, design, experimentation, and data use. The attack experiments are designed solely to reduce the performance of the constructed multi-agent systems on benchmarks such as mathematical reasoning and code generation, and do not construct any datasets to train agents to produce anomalous interactive behaviors. Furthermore, as long as the underlying large language model used is aligned with human values, our system will not generate harmful content.

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?

The benchmark datasets used in this paper are all from Hugging Face's public datasets, covering mathematical reasoning, code synthesis, and knowledge-intensive QA, etc., and do not contain personally identifying information or offensive content. Furthermore, all benchmark datasets used are cited 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?

In Appendix C

C. Did you run computational experiments?

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

In section 5.1, 5.4, and Appendix E.5

- 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?

In Appendix E

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.?

We did not use human annotators or research with human subjects

- 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)?

We did not hire anyone to participate in this work.

- 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)?

In Appendix C

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

The data we use does not require approval from an ethics review 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?

We use large language models to assist in grammatical and vocabulary correction, as well as bug fixing during the code debugging process. The design of the technical methods, experimental implementation, result analysis, manuscript writing, and conclusion formulation in this paper were all independently completed by the authors, and have been reviewed and verified by the authors.