

Responsible NLP Checklist

Paper title: *Evolving Agentic Workflow Driven by Human-Agent Collaboration*

Authors: *Yuxin Liu, Jinxuan Zhang, Yuezhang Peng, Hefeng Zhou, Xiangfeng Wang, Jiong Lou, Chentao Wu, Jie LI, Jingjing Qu, Chaochao Lu*

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?

(left blank)

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?

(left blank)

- ^{N/A} 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?

(left blank)

C. Did you run computational experiments?

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

Experimental setup and hyperparameters are provided in Section 5 and Appendix of the paper.

- 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?

Descriptive statistics are reported in Section 5 of the paper.

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.?

Evaluator information is reported in Appendix of the paper.

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](#).

- 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)?

Evaluator information is reported in Appendix of the paper.

- 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)?

Information is provided in Section 5.

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

This study involved human evaluation, but as the feedback was provided voluntarily by experts with full informed consent, and no personally identifiable information was collected or stored, the study was determined to be exempt from ethics review. The evaluation process was in line with standard academic practices, and all participants' rights were protected throughout the study.

- 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 did not use AI assistants directly for research, coding, or the primary writing of the paper. However, AI tools were referenced for some language refinement and editing tasks. These tools were not used for generating original content or conducting research but were employed for improving the clarity and presentation of the text. Author Submission Checklist: yes