

Responsible NLP Checklist

Paper title: *Visual Inception: Compromising Long-term Planning in Agentic Recommenders via Multi-modal Memory Poisoning*

Authors: *Jiachen Qian*

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?

Section 6.2 and Section 6.5 discuss platform-scale risks, potential misuse, and ethical mitigation strategies

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?

Section 5.1 describes synthetic benchmark sessions, and Section 6.5 states ethical safeguards; no personally identifying or offensive user data is used.

- 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 5.1 reports dataset/session scale (500/300/200 across domains), and Appendix A.1 reports reproducibility statistics (meanstd over 5 runs).

C. Did you run computational experiments?

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

Section 5.1 and Appendix A.2A.4 provide environment, implementation settings, and key 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?

Section 5.3 reports primary results with meanstd, and Appendix A.1 clarifies statistics are aggregated over 5 runs.

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.?

Human verification was performed internally by authors/research staff for sanity checking, not as an external participant study with formal participant instructions.

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)?

N/A (no external recruitment or paid participants).

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)?

N/A (evaluation used synthetic data and model outputs; no personal data from participants was collected).

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

N/A (no human-subject data collection protocol requiring IRB review).

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?

AI assistants were used for language polishing and drafting support. All technical content, experimental design, result validation, and final manuscript decisions were performed and verified by the authors