

## Responsible NLP Checklist

Paper title: *MAViS: A Multi-Agent Framework for Long-Sequence Video Storytelling*

Authors: *Qian Wang, Ziqi Huang, Ruoxi Jia, Paul Debevec, Ning Yu*

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 Ethical Considerations*

### B. Did you use or create scientific artifacts? (e.g. code, datasets, models)

- B1. Did you cite the creators of artifacts you used?

*Section 4 in main paper and Section C in Appendix*

- B2. Did you discuss the license or terms for use and/or distribution of any artifacts?

*We did not explicitly discuss artifact license consistency, as all datasets and models employed are widely used in open academic research. Our code will be open-sourced.*

- B3. Did you discuss if your use of existing artifact(s) was consistent with their intended use, provided that it was specified? For the artifacts you create, do you specify intended use and whether that is compatible with the original access conditions (in particular, derivatives of data accessed for research purposes should not be used outside of research contexts)?

*We did not explicitly discuss artifact usage consistency, as all datasets and models employed are widely used in open academic research.*

- 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 synthetic or automatically generated data and does not involve any personally identifiable or offensive content.*

- B5. Did you provide documentation of the artifacts, e.g., coverage of domains, languages, and linguistic phenomena, demographic groups represented, etc.?

*(left blank)*

- 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, Section C and Section F*

*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.*

**C. Did you run computational experiments?**

- C1. Did you report the number of parameters in the models used, the total computational budget (e.g., GPU hours), and computing infrastructure used?

*Section H*

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

*Section G*

- 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 due to high computational cost of large-scale video generation experiments, consistent with common practice in prior works.*

- C4. If you used existing packages (e.g., for preprocessing, for normalization, or for evaluation, such as NLTK, SpaCy, ROUGE, etc.), did you report the implementation, model, and parameter settings used?

*We used standard evaluation tool (AutoGen) with its default settings and did not explicitly detail parameter configurations.*

**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.?

*Full instructions provided to human evaluators are included in Appendix E.*

- 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)?

*Human annotators were recruited internally within our research group and did not receive monetary compensation. Therefore, recruitment and payment details were not included in 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)?

*This study does not use any human-related or personally identifiable data.*

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

*We did not seek IRB approval because our work exclusively uses publicly available datasets that do not contain personally identifiable information, consistent with standard academic practice.*

- D5. Did you report the basic demographic and geographic characteristics of the annotator population that is the source of the data?

*We did not collect or report demographic information of annotators due to platform anonymity and to protect privacy. All the human annotators were recruited internally within our research group.*

**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 ChatGPT to assist with English language polishing and minor text restructuring under the supervision of the authors. No scientific content, data analysis, or experimental design was generated by AI.*