

Responsible NLP Checklist

Paper title: *LitVISTA: A Benchmark for Narrative Orchestration in Literary Text*

Authors: *Mingzhe Lu, Yiwen Wang, Yanbing Liu, Qi You, Chong Liu, Ruize Qin, Haoyu Dong, Wenyu Zhang, JiaRui Zhang, Yue Hu, Yunpeng Li*

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 10 (Ethical Considerations), especially the paragraph Potential Risks and Subjectivity.

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 10 (Ethical Considerations), especially "Data Source, Licensing, and Privacy".

- 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 3.2.3 (Annotation Summary), Table 1.

C. Did you run computational experiments?

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

Section 4.1 states that detailed experimental configurations, including hyperparameter settings and prompt designs, are provided in Appendix D.

- 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 3.3.2, Table 2, and Figure 4. We report Precision, Recall, F1, and Harmonic Mean; the results are presented as single-run evaluations.

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.?

Appendix B (B.1 VISTA Annotator Manual; B.2 VISTA Theoretical Codebook).

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)?
Section 3.2.1 and Section 10. We report that the annotators were Masters students and PhD candidates, and that all participants took part on a voluntary basis and received no monetary compensation.
- 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)?
Section 10 (Ethical Considerations), especially Annotator Compensation and Process
- D4. Was the data collection protocol approved (or determined exempt) by an ethics review board?
No formal ethics review board approval or exemption determination was obtained. As described in Section 10, the study uses public-domain literary texts from Project Gutenberg, contains no PII of living individuals, and involves informed volunteer annotators without intervention beyond standard annotation work.
- 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?
Section 10 (Use of AI Tools). We permitted annotators to use AI tools solely for summarizing broader literary contexts and clarifying plot backgrounds. The core annotation tasks were performed manually by human annotators, and no AI-generated labels were used in constructing the gold-standard dataset.