

## Responsible NLP Checklist

Paper title: *Deep-Reporter: Deep Research for Grounded Multimodal Long-Form Generation*

Authors: *Fangda Ye, Kuicai Dong, Xie Zhifei, Yuxin Hu, Yihang Yin, Shurui Huang, Shikai Dong, Chen Zhang, Jianzhu Bao, Shuicheng YAN*

How to read the checklist symbols:

- the authors responded 'yes'
- the authors responded 'no'
- <sup>N/A</sup> 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 include an Ethical Considerations section in the paper to discuss potential risks.*

### 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?

*In Appendix Section A.1, We describe our procedures for handling personally identifiable information and potentially offensive content.*

- 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 A.2, we report statistics of the data associated with the proposed benchmark. In Sections 2.2 and Appendix C, we describe how the data used for training are collected, along with their corresponding statistics.*

### C. Did you run computational experiments?

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

*In Appendix C, we present detailed experimental settings, where Sections C.1 and C.2 describe the optimal hyperparameters used in our experiments.*

- 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?

*Due to the high computational cost of the experiments, we adopt a lower decoding temperature and report single-run results in Table 2 and Table 3.*

---

*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.?

*In Sections 2.2, 3, and D.3, we describe how participants were involved in the work.*

- 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)?

*In Section A.3, we report the information about recruitment and payment.*

- 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. Our benchmark uses publicly available analytical reports and webpages from authoritative repositories intended for open information dissemination. No private user-generated data was collected, and annotators provided relevance judgments on public technical content rather than contributing personal data.*

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

*N/A. Our work uses publicly available, non-personal, research-oriented materials and does not involve human subjects research requiring IRB approval. Annotation tasks involve only relevance judgments on public technical content.*

**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?

*In the Ethical Considerations section, we report information about the use of AI assistants.*