

Responsible NLP Checklist

Paper title: *Telling Speculative Stories to Help Humans Imagine the Harms of Healthcare AI*
Authors: *Xingmeng Zhao, Tongnian Wang, Dan Schumacher, Veronica Rammouz, Anthony Rios*

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?
Discussed in Section 8 (Ethics Statement), including risks of bias in generated stories, misleading narratives, and potential societal impacts

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?
Discussed in Section 8 (Ethics Statement). The paper states that no identifying information was collected and that outputs were reviewed to remove harmful or misleading 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?
Discussed in Section 4 (Experiments) and Section 5 (User Study), including number of AI applications (38), scenario variations, and number of participants (45)

C. Did you run computational experiments?

C2. Did you discuss the experimental setup, including hyperparameter search and best-found hyperparameter values?
Discussed in Section 4 (Experiments, Configuration), including models used, hardware setup, temperature (0.1), and token limits

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?
While results are reported with win rates and averages, the paper does not include full descriptive statistics such as variance, confidence intervals, or error bars. The study relies on pairwise comparisons and aggregated metrics rather than distributional reporting.

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.?

Instructions and task structure (speculative model card task and conditions) are described in Section 5 (User Study), with additional materials referenced in the appendix

- 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)?

IRB is mentioned and payment is mentioned in the user study section.

- 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)?

Discussed in Section 8 (Ethics Statement), which states that participants provided informed consent

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

Explicitly stated in Section 5 (User Study) and Section 8 (Ethics Statement) that the study was approved by an IR

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?

See A.11