

Responsible NLP Checklist

Paper title: *KnowMe-Bench: Benchmarking Person Understanding for Lifelong Digital Companions*
Authors: *Tingyu Wu, Zhisheng Chen, Ziyang Weng, Shuhe Wang, Shuo Zhang, Sen Hu, Silin Wu, Qizhen Lan, Huacan Wang, Ronghao Chen*

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?

Ethical considerations section (Page 9).

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 ("De-identification & Ethics")

- 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 ("Datasets and Narrative Modalities") and Table 1.

C. Did you run computational experiments?

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

Section 5.1

- 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 4.4, Section 5 and Table 2

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 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. 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)?

Appendix F reports that the human study used three expert annotators with M.A.-level training in literature or linguistics and that they were compensated at a reasonable rate consistent with standard expert-annotation practice. However, the paper does not fully report recruitment source/platform or demographic-specific payment adequacy.

- 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 5.1 and the Ethical considerations section

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

The Ethical considerations section states that the benchmark uses publicly releasable narratives and does not introduce additional risks regarding data privacy or human rights violations, but no specific IRB approval or exemption number is reported.

- 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 3.1 describes the multi-agent data generation pipeline utilizing LLMs, and Section 4.4 describes the use of GPT-4o in the LLM-as-a-Judge evaluation protocol.