

## Responsible NLP Checklist

Paper title: *ImplicitMemBench: Measuring Unconscious Behavioral Adaptation in Large Language Models*

Authors: *Chonghan Qin, Xiachong Feng, Weitao Ma, Xiaocheng Feng, Lingpeng Kong*

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.*

- <sup>N/A</sup> A2. Did you discuss any potential risks of your work?

*(left blank)*

### 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?

*No. The benchmark data consists of constructed/synthetic task instances, prompts, and evaluation templates rather than personal user data, and we do not use data that names or uniquely identifies individuals in the benchmark itself.*

- 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?

*Yes. See Section 3.4, Table 6, and Figure 3.*

### C. Did you run computational experiments?

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

*Yes. See Section 4.1.*

- 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?

*Yes. See Section 4.3 and Table 7.*

### 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.?

*No. We report the human-baseline protocol and scoring procedure, but do not include the full*

---

*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.*

*instruction text given to participants in the paper or appendix. The human study is used only to provide a high-level reference baseline rather than to analyze participant behavior in depth.*

- 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)?

*No. We report that the human baseline was collected from five computer science Ph.D. students, but do not provide a full description of recruitment, demographics, or compensation. Our goal was to establish a small internal reference baseline rather than conduct a broad human-subject study.*

- 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)?

*No. We do not explicitly discuss consent procedures in the paper. The human baseline involves a small internal study with non-sensitive benchmark tasks, and participant responses are used only in aggregate as a reference baseline.*

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

*No. We do not report ethics board approval in the paper. The human baseline is limited to a small, minimal-risk internal evaluation on non-sensitive benchmark tasks, used only to establish a reference baseline.*

- 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?

*No. We used AI assistants to polish the writing*