

## Responsible NLP Checklist

Paper title: *Rethinking Data Mixing from the Perspective of Large Language Models*

Authors: *Yuanjian Xu, Tianze Sun, Changwei Xu, XinLong Zhao, Jianing Hao, Ran Chen, Yang Liu, Ruijie Xu, Stephen Chen, Guang Zhang*

How to read the checklist symbols:

- the authors responded 'yes'
- the authors responded 'no'
- N/A 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?

*Limitation*

### 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 A.2 mentions SlimPajamas cleaned and deduplicated nature (627B Token Cleaned and Deduplicated Version of RedPajama), ensuring no personally identifying or 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?

*Elaboration: Section A.2 Table 3 & A.2 Table 5 provide dataset proportions, model sizes, and training step counts full quantitative details for data use.*

### C. Did you run computational experiments?

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

*Section A.2 Table 4 provides model parameters (124M300M) and architectural specs. Section A.2 Table 5 specifies training steps, batch sizes, and learning rates.*

- 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 3.2 & 3.3 (Tables 12) report mean accuracies and validation perplexities for multiple baselines full comparison transparency.*

*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.?

*(left blank)*

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)?

*(left blank)*

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)?

*(left blank)*

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

*(left blank)*

**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?

*AI assistants were used solely for text polishing and formatting*