

## Responsible NLP Checklist

Paper title: *Improving the Throughput of Diffusion-based Large Language Models via a Training-Free Confidence-Aware Calibration*

Authors: *Jucheng Shen, Gaurav Sarkar, Yeonju Ro, Sharath Nittur Sridhar, Zhangyang Wang, Aditya Akella, Souvik Kundu*

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?

*This work focuses on improving inference efficiency of existing diffusion-based LLMs and does not introduce new datasets, deployment settings, or application domains. Therefore, it does not introduce additional societal risks beyond those already associated with the underlying base models.*

### B. Did you use or create scientific artifacts? (e.g. code, datasets, models)

#### <sup>N/A</sup> 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?

*We use standard public benchmarks (GSM8K, MATH, MBPP, HumanEval) that are widely used in the community and do not contain personally identifying information. We did not collect or curate new data.*

#### 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?

*We use the standard test splits of GSM8K (5-shot), MATH (4-shot), MBPP (3-shot, pass@1), and HumanEval (0-shot, pass@1) as provided by the original benchmarks. Shot settings and evaluation protocols are specified in section 5.1. We do not create new data or modify splits.*

### C. Did you run computational experiments?

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

*Experimental setup is described in section 5.1. Hyperparameters for CadLLM on LLaDA and DREAM, and for the Fast-dLLM baselines, are listed in Appendix Tables 4, 5, and 6. A 20% hyperparameter sensitivity sweep is provided in Appendix section 7.4.7 (Tables 1216).*

#### 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,

---

*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.*

etc. or just a single run?

*Statistical uncertainty (mean standard error over three seeds) for throughput and accuracy on LLaDA across all four benchmarks is reported in Appendix section 7.4.8, Table 17. Other numbers in Table 1, Table 10, and ablation tables are single-run results, as noted in the paper.*

**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?

*(left blank)*