

Responsible NLP Checklist

Paper title: *RACC: Regret-Aware Confidence Calibration for Consistent Masked Discrete Diffusion Decoding*

Authors: *Qinglin Zeng, Jusheng Zhang, Jing Yang, Ningyuan Liu, Keze Wang*

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.

^{N/A} 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)

^{N/A} 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?
(left blank)

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?
In Section 4.2 (Experimental Setup), we detail the datasets used (HumanEval, MBPP, GSM8K, MATH, Sudoku, Countdown) and cite the original papers. We utilize the standard test splits and evaluation protocols associated with these established benchmarks.

C. Did you run computational experiments?

C2. Did you discuss the experimental setup, including hyperparameter search and best-found hyperparameter values?
We specify the exact hyperparameter values used for our reported results in Section 4.2 (Implementation Details). Furthermore, we explicitly discuss the hyperparameter search space and sensitivity analysis in Section 4.4 (Ablation Study), justifying our selection of the best-found values.

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?
All primary performance metrics are reported as point estimates derived from standard single-run evaluations. Due to the significant computational overhead inherent to the iterative denoising process of large-scale masked diffusion models, executing multi-seed statistical variance analysis across the comprehensive suite of benchmarks and baselines is computationally intractable. This 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.

protocol remains strictly consistent with the established evaluation practices in the diffusion language model literature.

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?

We disclose our AI usage directly here: We utilized generative AI tools (e.g., Gemini) exclusively for writing refinement, grammatical error correction, and LaTeX formatting (Case A in ACL Policy). All scientific claims, experimental designs, and results are original contributions of the authors.