

Responsible NLP Checklist

Paper title: *EvoMD-LLM: Learning the Language of Species Evolution in Reactive Molecular Dynamics*

Authors: *Zhichen Tang, Zhengzheng Dang, Yulin Chen, Jixin Wu, Haiwen Li, Yanming Wang*

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?
Limitations

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?
The paper does not explicitly discuss checks for personal identifiers or offensive content. All data used in this work is generated from molecular dynamics simulations and derived symbolic representations, which do not contain natural language, personal information, or human-related content by construction. Therefore, no anonymization or additional privacy protection steps are required. We will clarify this explicitly in the final version.

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?
3.2, Appendix A

C. Did you run computational experiments?

C2. Did you discuss the experimental setup, including hyperparameter search and best-found hyperparameter values?
3.2, Appendix B.3

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?
3.2, 3.3, Appendix B.2

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.?

The data in this study was entirely generated by physical simulation (molecular dynamics simulation) and annotated by an automated script (based on bond-level analysis), without involving any human participants or manual annotators.

- 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)?

The data in this study was entirely generated by physical simulation (molecular dynamics simulation) and annotated by an automated script (based on bond-level analysis), without involving any human participants or manual annotators.

- 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)?

The data in this study was entirely generated by physical simulation (molecular dynamics simulation) and annotated by an automated script (based on bond-level analysis), without involving any human participants or manual annotators.

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

The data in this study was entirely generated by physical simulation (molecular dynamics simulation) and annotated by an automated script (based on bond-level analysis), without involving any human participants or manual annotators.

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 for code implementation and manuscript preparation. Specifically, GitHub Copilot assisted in drafting data preprocessing and evaluation scripts, and ChatGPT was used to polish language and enhance manuscript clarity. All AI-generated content/suggestions were manually reviewed, edited, and verified by the authors. The authors take full responsibility for the final work, and no AI assistant contributed to core research innovations (e.g., method design, experimental conception, key results interpretation) that would qualify for authorship.