Responsible NLP Checklist

Paper title: *SynthFix: Adaptive Neuro-Symbolic Code Vulnerability Repair*
Authors: *Yifan Zhang, Jieyu Li, Kexin Pei, Yu Huang, Kevin Leach*

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?
Yes. Discussed in the 'Ethical Considerations' section (after Limitations), which covers risks of automated patches inducing over-reliance and incomplete fixes.

☑ **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?
N/A. We use established public code benchmarks (FixJS from public GitHub bug-fixing commits, CodeFlaws from Codeforces). These are source-code datasets without user-identifying content, and the task does not involve human subjects or personal 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?
Yes. Dataset sizes (~300k FixJS, ~4k CodeFlaws) and the fixed 80/10/10 train/validation/test split are reported in Section 3.2. Appendix C provides distributions of CWE categories and functionality categories for the test sets used in RQ3.

☑ **C. Did you run computational experiments?**

☑ C2. Did you discuss the experimental setup, including hyperparameter search and best-found hyperparameter values?
Yes. Section 3 (Experimental Design) and Appendix B (Detailed Experimental Setup) describe the model architectures, PPO configuration, learning rates, schedulers, and batch sizes. Appendix D describes the router's training objective.

☒ 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?
No. For reproducibility we train with a single fixed random seed (Section 3.2). Multi-seed variance is not reported because the evaluation spans four base models up to 7B parameters with PPO loops;

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.

multi-seed runs were computationally infeasible. Reported numbers are single-run values. We note this as a limitation.

☒ **D. Did you use human annotators (e.g., crowdworkers) or research with human subjects?**

☐ N/A 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)

☐ N/A 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)

☐ N/A 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)

☐ N/A 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?
Yes. See the 'Use of AI Assistants' paragraph in the Ethical Considerations section. AI assistants were used only for minor editorial assistance (phrasing, LaTeX formatting) and not for research ideation, experiment design, implementation, or result analysis; no substantive technical content was AI-generated.