

## Responsible NLP Checklist

Paper title: *ElasticFlow: One-Step Physics-Consistent Policy with Elastic Time Horizons for Language-Guided Manipulation*

Authors: *Kewei Chen, Yayu Long, Shuai Li, Mingsheng Shang*

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.*

- A2. Did you discuss any potential risks of your work?

*See Section 7 Ethical Considerations. The paper discusses physical safety hazards related to robot control (specifically high-frequency jitter) in the "Physical Safety and Green Computing" subsection. It also addresses potential misuse and data biases in the "Social Responsibility and Misuse" subsection.*

### 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?

*See Section 7 Ethical Considerations. The paper explicitly state that the study was conducted in standard simulation and real-world environments involving no sensitive 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?

*See Section 4 Experiments and Appendix E Real-World Experiment Details. Specifically, Section 4.2 mentions the ABC-D split for the CALVIN benchmark. Section E.1 and Section E.3 (Table 8) detail the data scale for real-world experiments, including the number of trajectories (20-50 per task) and total trials (280 trials).*

### C. Did you run computational experiments?

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

*See Section 4.5 (subsection "Sensitivity Analysis of CFG Guidance Weight") and Appendix C.3. The paper discuss the sensitivity analysis for the Classifier-Free Guidance (CFG) weight  $w$ , identifying the optimal interval and the best-found value. Hardware configurations are detailed in Appendix D.*

- 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?

See Table 1 (Section 4.1), which reports the average success rates for the LIBERO benchmark. Table 2 (Section 4.2) reports the average chain length for CALVIN. Section E.3 and Table 8 report success rates derived from 20 trials per task for real-world experiments.

**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?

*See Appendix A Use of AI Assistants. The paper explicitly states that AI assistants were used solely for grammar checking and polishing.*