

Responsible NLP Checklist

Paper title: *ToolDNA: Autonomous Evolution of Tool Metadata for Robust Dialogue Agents*

Authors: *Qiuyuan Ai, Cong Wang, jiaqi zhang, zengxin han, Jie Song*

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?
Section Number: Ethics Statement

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?
Section Number: Ethics Statement

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?
Section Number: Section 4 & appendix C

C. Did you run computational experiments?

C2. Did you discuss the experimental setup, including hyperparameter search and best-found hyperparameter values?
Section Number: Section 4 & appendix C

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?
Section Number: Section 4

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 annotation for this dataset was performed by internal domain experts (sales professionals) within the enterprise. The specific annotation guidelines and operational workflows involve proprietary business logic and commercial trade secrets, which cannot be disclosed in full. However, the core logic for tool invocation and error analysis used by the system is reflected in the Feedback LLM prompt templates provided in Appendix B.

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.

- 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 annotators were existing internal employees (professional sales staff) of the company and were not specifically recruited for this academic study. As they performed these tasks as part of their professional roles within the organization, specific recruitment details and salary information are confidential and governed by internal corporate policies.

- 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)?

Section Number: Ethics Statement.

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

The dataset used in this study consists of historical, de-identified customer service logs provided by the enterprise. The data collection and research protocols were conducted in compliance with internal corporate data privacy policies and ethical guidelines for academic research. Since the data was retrospective and fully anonymized to remove all PII before research began, it was determined to be exempt from formal Institutional Review Board (IRB) approval.

- 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 didn't use AI assistants.