

## Responsible NLP Checklist

Paper title: *WebUncertainty: Dual-Level Uncertainty Driven Planning and Reasoning For Autonomous Web Agent*

Authors: *Lingfeng Zhang, yongan sun, Jinpeng Hu, Hui Ma, Ying Yang, Kuiwen Liu, Zenglin Shi, Meng 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?

*Ethics Statement Section*

### 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?

*Our study uses public benchmarks (WebArena, WebVoyager) with functional web tasks (e.g., information retrieval, workflow execution) from open platforms. The data does not involve personally identifying information (PII) or offensive content, so steps for checking/protecting such content are not applicable.*

- 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?

*Experiment Section*

### C. Did you run computational experiments?

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

*Experiment Section*

- 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?

*Experiment Section*

---

*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.?

*Our study conducts automated evaluations on public benchmarks (WebArena and WebVoyager) without involving human participants, annotators, or any individuals in data collection or experimental procedures. Since there are no participants/annotators in the research, there are no instructions, screenshots, or risk disclaimers to report.*

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)?

*Our research conducts fully automated evaluations on public benchmarks (WebArena and WebVoyager) without involving any human participants, annotators, or individuals in data collection or experimental procedures. Since there are no participants in the study, there is no recruitment process, payment arrangement, or related demographic-specific payment adequacy to report.*

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)?

*Our study uses publicly available benchmarks (WebArena, WebVoyager) consisting of functional web tasks from open platforms. The data does not involve personal data or require curation of individual-specific information, so obtaining consent from data subjects is not applicable.*

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

*Our study only uses publicly available benchmarks (WebArena, WebVoyager) for model evaluation and does not involve any data collection protocols, human subjects, or original data curation. Therefore, ethics review board approval or exemption is not applicable.*

**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 only used AI assistants for grammatical and syntactic polishing of the manuscript, which did not involve any changes to the core content of the paper (including framework design, experimental setup, results analysis, formula derivation, etc.). The AI assistant was merely a tool for improving language expression fluency, and all research ideas, technical innovations, and experimental results are original work completed independently by the authors.*