

## Responsible NLP Checklist

Paper title: *Reviving Iterative Refinement in Diffusion-based NER with an Initializer-Restorer Approach*

Authors: *Long Hai Trieu, Ph Minh Hiu, Makoto Miwa*

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?

*Potential risks are discussed in the Ethics Statement section. The proposed method aims to improve NER accuracy for beneficial downstream applications such as knowledge graph construction and document analysis. We acknowledge the potential environmental impact of training large diffusion models and have employed techniques like EMA and efficient initialization to mitigate unnecessary computation.*

### 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?

*All datasets used (ACE2004, GENIA, CleanCoNLL) are standard, publicly available NLP benchmarks. They contain news and biomedical text and do not include any privately collected or personally identifiable information beyond what is already present in these well-established benchmarks.*

- 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?

*Dataset statistics including the number of train/dev/test splits are reported in Appendix B, Table 5.*

### C. Did you run computational experiments?

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

*The experimental setup including backbone models, EMA decay rate, noise schedule, GPU details, and the number of candidate boxes  $k=180$  are described in Section 3.3. The default inference step  $=5$  is justified in Section 4.1.*

- 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?

---

*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.*

*Due to computational constraints, all reported results are based on a single random seed. We do not report error bars or summary statistics across multiple runs. This limitation is explicitly acknowledged in Section 4.1, where we note the potential for run-to-run variance while arguing that consistent trends across three datasets support the validity of our approach.*

**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.?

*No human annotators or participants were used in this work.*

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)?

*No human annotators or participants were recruited or paid in this work.*

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)?

*All datasets used (ACE2004, GENIA, CleanCoNLL) are standard publicly available benchmarks with established data usage agreements. No new data was collected and no consent process was required.*

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

*No new data was collected in this work. All datasets used are existing publicly available benchmarks (ACE2004, GENIA, CleanCoNLL). No ethics review board approval was required.*

**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?

*The use of Google Gemini for language editing during manuscript preparation is disclosed in the Acknowledgments section. All research design, experimentation, and analysis were conducted by the authors.*