

Responsible NLP Checklist

Paper title: *Reframing Responsibility: Framing-Aware Event Causality Identification*

Authors: *Jin Zhao, Jiayi Yao, Xinrui Hu, Nianwen Xue*

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?
Ethical Considerations 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?
Ethical Considerations section. The dataset is built from publicly available Wikipedia articles and contains no private individual-level information; however, it covers politically contentious topics that may be perceived as sensitive or offensive. We discuss these considerations and mitigations in the Ethical Considerations section.

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 3 (Table 1 reports per-language counts of topics, documents, tokens, events, coreference chains, causal relations, and responsibility targets; Table 2 reports framing/source/certainty distributions). Section 4.1 reports the train/dev/test split (36/5/10 topics), with the topic-level breakdown in Appendix A.3.

C. Did you run computational experiments?

C2. Did you discuss the experimental setup, including hyperparameter search and best-found hyperparameter values?
Section 4 (main experimental setup, candidate generation, models, evaluation metrics). Hyperparameters and training details are in Appendix A.6 (Joint FrECI model: optimizer, learning rate, epochs, batch size, max sequence length) and Appendix A.7 (SFT LLaMA-3.1-8B baseline: QLoRA configuration, learning rate, LoRA rank/alpha/dropout, batch size, decoding settings).

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.

Section 4.3 (Table 4 reports precision, recall, F1, attribution-component scores, and Full-Claim EM for all models on the held-out test set). Appendix A.5 reports a cross-lingual breakdown of Joint FrECI performance grouped by source language. All numbers are from a single run on the topic-held-out test split (10 topics).

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.?

Appendix A.1 (FrECI Annotation Guidelines, full guidelines reproduced as Figure 4). Appendix A.2 also documents the annotation interface (Figure 5).

- 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)?

Annotators were research assistants who performed annotation as part of their research training and academic responsibilities; no separate monetary compensation was provided. Annotator backgrounds and the calibration/training process are described in Section 3.6.1 (Annotator Recruitment and Training).

- 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)?

The dataset is built entirely from publicly available Wikipedia articles released under Wikipedia's CC BY-SA license. No personal, private, or individually identifying data is collected from human subjects, so individual consent is not applicable.

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

The work uses publicly available Wikipedia text and does not collect new data from human subjects. Annotators are members of the research team performing structured annotation of public text, which does not constitute human-subjects research requiring IRB approval at our institution.

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?

AI is mainly used for polishing language purposes.