Responsible NLP Checklist

Paper title: *DongbaMIE: A Multimodal Information Extraction Dataset for Evaluating Semantic Understanding of Dongba Pictograms*
Authors: *Xiaojun Bi, Shuo Li, Junyao Xing, Ziyue Wang, Fuwen Luo, Weizheng Qiao, Lu Han, Ziwei Sun, Peng Li, Yang Liu*

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?
We answered No because the dataset is derived from public Dongba manuscripts, contains no sensitive content, and poses no foreseeable risks.

☑ **B. Did you use or create scientific artifacts? (e.g. code, datasets, models)**

☑ B1. Did you cite the creators of artifacts you used?
Yes, we cited the DeepSeek model used in constructing the DongbaMIE dataset (see Section 3.3: Hybrid Annotation Pipeline).

☑ B2. Did you discuss the license or terms for use and/or distribution of any artifacts?
Yes, we specified that the DongbaMIE dataset is released under the CC BY-NC 4.0 license for non-commercial research use (see Section 3: Constructing DongbaMIE Dataset).

N/A B3. Did you discuss if your use of existing artifact(s) was consistent with their intended use, provided that it was specified? For the artifacts you create, do you specify intended use and whether that is compatible with the original access conditions (in particular, derivatives of data accessed for research purposes should not be used outside of research contexts)?
(left blank)

☒ 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?
The dataset is derived from publicly available Dongba manuscripts, which do not contain personally identifying information or offensive content.

☑ B5. Did you provide documentation of the artifacts, e.g., coverage of domains, languages, and linguistic phenomena, demographic groups represented, etc.?
Yes, we provide basic documentation of the DongbaMIE dataset, including its source, content type, and annotation pipeline (see Section 3: Constructing DongbaMIE Dataset).

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.

☑ 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?
Yes, we report the number of examples and the train/dev/test splits for the DongbaMIE dataset (see Section 3: Constructing DongbaMIE Dataset).

☑ **C. Did you run computational experiments?**

N/A C1. Did you report the number of parameters in the models used, the total computational budget (e.g., GPU hours), and computing infrastructure used?
(left blank)

☑ C2. Did you discuss the experimental setup, including hyperparameter search and best-found hyperparameter values?
Yes, we describe the experimental setup in Section 4: Methodology.

☒ 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?
No, we report results from a single run for each experiment, as detailed in Section 5 Experimental Analysis

☒ C4. If you used existing packages (e.g., for preprocessing, for normalization, or for evaluation, such as NLTK, SpaCy, ROUGE, etc.), did you report the implementation, model, and parameter settings used?
No, we used standard default settings for all external packages and tools, so we did not report specific parameters.

☑ **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.?
see Section 3.3: Hybrid Annotation Pipeline

☑ 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)?
Yes, we report the recruitment method for annotators and relevant statistics in Section 3.3: Hybrid Annotation Pipeline. The annotation was performed by volunteer researchers within our team, so no monetary payment was provided.

N/A 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)

N/A D4. Was the data collection protocol approved (or determined exempt) by an ethics review board?
(left blank)

N/A D5. Did you report the basic demographic and geographic characteristics of the annotator population that is the source of the data?
(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?
No, we did not use AI assistants in the preparation of this work.