Responsible NLP Checklist

Paper title: *Financial Risk Relation Identification through Dual-view Adaptation*
Authors: *Wei-Ning Chiu, Yu-Hsiang Wang, Andy Hsiao, Yu-Shiang Huang, Chuan-Ju Wang*

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?
Discussed in Section 7 (Limitations), where we note risks such as over-reliance on 10-K filings, lack of real-time adaptability, and the potential absence of qualitative expert input

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

☑ B1. Did you cite the creators of artifacts you used?
Citations to pretrained encoders (e.g., BERT, Contriever, DPR, FinBERT, SEC-BERT, Llama-3.2) are provided in Section 2 (Related Works) and Section 4.3 (Encoders for Comparison)

☒ B2. Did you discuss the license or terms for use and/or distribution of any artifacts?
While we used publicly available pretrained models (e.g., Hugging Face releases), we did not explicitly discuss licenses. A justification is that all models used are broadly accessible for research under standard open-source or research-use licenses.

☑ 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)?
In Section 4.3 and 4.5 (Encoders for Comparison & Evaluation) we use the cited models in line with their intended purpose (retrieval/encoding for text similarity). For the artifact we create (our dual-view encoder), intended use is specified for financial document retrieval and relation discovery

☒ 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 dataset is entirely based on public corporate filings (Form 10-K) and stock market data neither contains 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.?

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.

Dataset scope and artifact coverage are described in Section 4.4.1 (Data Sources): 2,136 filings from 337 S&P 500 companies (20182024), restricted to Item 1A and Item 7A risk-related sections

☑ 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?
Reported in Section 4.1 and 4.4.1 8,500 training pairs, 1,000 validation pairs, 2,136 filings, 337 companies. Also, train/validation/test splits are detailed for downstream ADGAT experiments.

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

☑ C1. Did you report the number of parameters in the models used, the total computational budget (e.g., GPU hours), and computing infrastructure used?
In Section 4.1, we state that training used BERT-base-uncased (110M parameters) on a single NVIDIA V100 GPU for ~8 hours.

☑ C2. Did you discuss the experimental setup, including hyperparameter search and best-found hyperparameter values?
Full details are given in Section 4.1 (Encoder Training Details) and Section 4.7.1 (ADGAT Experimental Setup). Includes batch size, learning rate, optimizer, dropout, epochs, and hyperparameter search ranges

☑ 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?
We report multiple runs with averages and standard deviations (e.g., ADGAT stock prediction, Table 3). We also provide sensitivity analysis (Appendix A, Figure 3) to show stability under threshold variations

☑ 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?
Section 4.1 specifies Hugging Face BERT-base-uncased, Adam optimizer, L2 regularization, and learning scheduler. Section 4.8 mentions BAAI/bge-reranker-v2-m3 for re-ranking

☒ **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.?
We did not involve human annotators or human subjects all data comes from publicly available financial filings and market data.

☒ 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)?
We did not involve human annotators or human subjects all data comes from publicly available financial filings and market data.

☒ 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)?
We did not involve human annotators or human subjects all data comes from publicly available financial filings and market data.

☒ D4. Was the data collection protocol approved (or determined exempt) by an ethics review board?
We did not involve human annotators or human subjects all data comes from publicly available financial filings and market data.

☒ D5. Did you report the basic demographic and geographic characteristics of the annotator population that is the source of the data?
We did not involve human annotators or human subjects all data comes from publicly available financial filings and market data.

☑ **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 assistants (e.g., ChatGPT) were used to support writing and organization. This is acknowledged here in the checklist response; the models were not used for generating results or experiments, only as an auxiliary tool in drafting and refinement.