Responsible NLP Checklist

Paper title: *ConText-LE: Cross-Distribution Generalization for Longitudinal Experiential Data via Narrative-Based LLM Representations*
Authors: *Ahatsham Hayat, Bilal Khan, Mohammad Rashedul Hasan*

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?
Section 6

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

☑ B1. Did you cite the creators of artifacts you used?
Section 3.4 and 4.2

☒ B2. Did you discuss the license or terms for use and/or distribution of any artifacts?
We did not discuss licenses or terms of use because our paper focuses on the methodological contribution of the ConText-LE framework rather than on artifact distribution. While we utilize existing datasets and pre-trained models, we reference them primarily to demonstrate our approach rather than for redistribution.

☒ 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)?
We did not explicitly discuss whether our use of existing artifacts was consistent with their intended use, nor did we specify intended use conditions for our created artifacts. While we used the datasets (GLOBEM, LifeSnaps, MFAFY) for behavioral prediction research, which aligns with their original purposes, we did not explicitly verify or state this alignment in the paper. Similarly, for the pre-trained models (Llama 3.1, Mistral-7B, Falcon-7B), we did not discuss whether our fine-tuning for behavioral forecasting was within their intended use scope.

☒ 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?
We did not explicitly discuss the steps taken to check for personally identifying information or offensive content in the datasets, nor did we detail anonymization procedures. While we worked

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.

with longitudinal experiential data that inherently contains sensitive personal information (activity patterns, sleep data, mood reports, academic behaviors), we did not describe specific protocols for identifying and protecting personally identifiable information. We acknowledge in our limitations section that "LE data often contains sensitive personal information, and narrative representations may inadvertently make this information more interpretable or identifiable," but we did not discuss concrete steps taken to verify anonymization or check for offensive content in the original datasets. Since we relied on existing datasets (GLOBEM, LifeSnaps, MFAFY) that presumably underwent ethical review and anonymization by their original creators, we did not detail additional privacy protection measures.

☑ B5. Did you provide documentation of the artifacts, e.g., coverage of domains, languages, and linguistic phenomena, demographic groups represented, etc.?
Section 4

☑ 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 4.1 and 4.2

☑ **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?
While we reported model parameter counts (Llama 3.1 8B: 8 billion parameters, Mistral-7B: 7.3 billion parameters, Falcon-7B: 7 billion parameters) in Section 4.6, we did not provide comprehensive computational budget information such as total GPU hours or detailed computing costs. We mentioned using "NVIDIA A40 48GB GPUs", but we did not report the total computational budget across all experiments.

☑ C2. Did you discuss the experimental setup, including hyperparameter search and best-found hyperparameter values?
Section 4.2

☑ 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?
Section 4.5

☑ 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 4.2

☒ **D. Did you use human annotators (e.g., crowdworkers) or research with human subjects?**

N/A 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.?
(left blank)

N/A 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)?
(left blank)

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?
We did not include information about AI assistant use in the paper itself. While we used AI assistants during the writing and revision process to help refine language, structure arguments, and ensure clarity of presentation, we did not explicitly acknowledge this assistance in any section of the paper.