

## Responsible NLP Checklist

Paper title: *Beyond Single Representations: Multi-Model Embedding Fusion for Stable Text Classification*

Authors: *Jiho Gwak, Yuchul Jung*

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.*

- N/A A2. Did you discuss any potential risks of your work?

*This paper focuses on a fundamental analysis of embedding fusion and layer-wise representational properties for text classification using publicly available benchmark datasets. Since the research does not involve the generation of harmful content, human-subject studies, or the deployment of models in sensitive real-world applications, it does not pose significant ethical risks or potential for misuse.*

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

- N/A 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?

*(left blank)*

- 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?

*Detailed statistics for all datasets used in our experiments, including the number of classes, training samples, and test samples, are reported in Table 1 (Section 3.2, page 4). Additionally, the train/test split ratios and specific data partitioning methods are described in Section 4 (page 4, 5).*

### C. Did you run computational experiments?

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

*We provide the details of our experimental environment, including hardware specifications (NVIDIA RTX 4090 GPUs) and specific hyperparameters such as batch size (1024), learning rate (1e-4), optimizer (Adam), and epochs (120) in Section 4*

- 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 both Accuracy and F1-scores for our results. It is explicitly stated in Table 2 that last-layer*

---

*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.*

*results are averaged over five runs, while best-layer results correspond to the peak performance observed across those runs. Furthermore, for multi-model fusion experiments in Table 4, we report the minimum and maximum performance ranges across all possible combinations.*

**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.?

*(left blank)*

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)*

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)*

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

*(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?

*AI assistants were used solely for language translation and grammatical refinement of the manuscript. The core research, including the formulation of the study, experimental design, data analysis, and coding, was conducted entirely by the authors without the use of AI for content generation.*