

Responsible NLP Checklist

Paper title: *Do We Need Distinct Representations for Every Speech Token? Unveiling and Exploiting Redundancy in Large Speech Language Models*

Authors: *Bajian Xiang, Tingwei Guo, Xuan Chen, Yang Han*

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?

This work focuses on improving the inference efficiency of large speech language models through training-free token compression, and does not introduce new datasets, human-subject interaction, deployment in safety-critical settings, or capabilities that materially increase misuse risk. The main limitations are discussed in the Limitations section, which covers evaluation scope, alignment accuracy, and baseline comparisons.

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?

This work does not involve new data collection. We use previously released benchmark datasets and pretrained models, and therefore did not perform our own data collection, anonymization, or screening pipeline for personally identifying information or offensive content. We rely on the documentation and release practices of the original data sources.

- 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 5.1 and Appendix C.1 (Benchmark Details). We report the datasets used, their evaluation splits (e.g., test-clean/test-other), and relevant statistics such as the number of examples for ASR, speech QA, and speech translation benchmarks.

C. Did you run computational experiments?

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

Sections 3.1, 5.1, and 5.4. We discuss the experimental setup and specify the key hyperparameters used in our method, including the injection layers, cosine similarity thresholds, and lookback window sizes. We also include parameter sensitivity analyses to justify the selected values.

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.

- 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?

Sections 3.2, 3.3, 4.2, and 5.1-5.4. As specified in the paper, all reported results are dataset-averaged (mean) evaluation scores. Since we use greedy decoding in our experiments, these numbers are deterministic and indicate mean performance on the evaluation set.

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?

The use of AI assistants was limited to minor editing and was not explicitly documented in the paper.