

Responsible NLP Checklist

Paper title: *Beyond Variance: Knowledge-Aware LLM Compression via Fisher-Aligned Subspace Diagnostics*

Authors: *Ibne Farabi Shihab, Sanjeda Akter, Anuj Sharma*

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?
(left blank)

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?
We do not collect new data; we use established public datasets for calibration/evaluation. We did not include a dedicated discussion of PII/offensive-content screening steps in this work.

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 3 (Experimental Setup) reports calibration dataset choice and size (e.g., 4096 samples from C4) and describes alternative calibration distributions (WikiText-2, Alpaca-format)

C. Did you run computational experiments?

C2. Did you discuss the experimental setup, including hyperparameter search and best-found hyperparameter values?
Section 3 details model families, calibration setup, and compression-rate settings; Section 4.2 reports inference measurement settings (GPU, batch size, sequence length). Appendix D provides threshold and calibration-size sensitivity analysis

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?
Appendix D reports bootstrap sampling (1000 resamples) and 95% confidence intervals for estimates across calibration sizes; the paper also reports stability across multiple random seeds.

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.

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?

Acknowledged in Acknowledgement and Reproducibility section