

Responsible NLP Checklist

Paper title: *Fine-Tuned Thoughts: Leveraging Chain-of-Thought Reasoning for Industrial Asset Health Monitoring*

Authors: *Shuxin Lin, Dhaval C Patel, Christodoulos Constantinides*

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?

Our work is methodological and does not introduce new datasets, sensitive applications, or deployment settings; therefore, no direct risks were identified.

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

- B1. Did you cite the creators of artifacts you used?

Section 4.3 discusses the benchmark dataset we use in this paper. The dataset is properly cited.

- B2. Did you discuss the license or terms for use and/or distribution of any artifacts?

The dataset we have used is licensed with Apache 2.0. We do not release any artifacts with this work, and therefore a license discussion is not applicable.

- 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 only used publicly available artifacts (datasets/models) in ways consistent with their intended use as specified by their licenses and documentation.

- 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?

The datasets used are standard publicly available without PII or offensive content. Therefore, no additional anonymization steps were required.

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

See Section 4.3, which discusses the benchmark dataset used in this paper, including the domains and assets covered. Additional details (e.g., linguistic and demographic documentation) are available in the datasets official public documentation linked therein.

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.

- 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?
See Sections 4.1 and 4.3, which document the number of examples.
- 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?
See Section 4.6, which reports the number of parameters, as well as details of the computing infrastructure used.
- C2. Did you discuss the experimental setup, including hyperparameter search and best-found hyperparameter values?
We did not conduct a hyperparameter search. The experiments were run with fixed hyperparameter values based on standard practice.
- 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?
See Sections 4.54.8, which present the experimental results with descriptive statistics and clearly indicate whether reported values correspond to single runs or aggregated measures.
- 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?
See Section 4.4, which details the evaluation metrics and packages used, including proper citation of PertEval. See Section 4.6, which documents the implementation and parameters used for finetuning.
- 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 do not conduct any human annotation or other crowdsourcing activities.
- 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 do not conduct any human annotation or other crowdsourcing activities.
- 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 do not conduct any human annotation or other crowdsourcing activities.
- D4. Was the data collection protocol approved (or determined exempt) by an ethics review board?
We do not conduct any human annotation or other crowdsourcing activities.
- D5. Did you report the basic demographic and geographic characteristics of the annotator population that is the source of the data?
We do not conduct any human annotation or other crowdsourcing activities.
- 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 do not use AI assistants in research or writing.