

## Responsible NLP Checklist

Paper title: *HCFD: A Benchmark for Audio Deepfake Detection in Healthcare*

Authors: *Mohd Mujtaba Akhtar, Girish, Muskaan Singh*

How to read the checklist symbols:

- the authors responded 'yes'
- the authors responded 'no'
- 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?

*The paper discusses the risk context of healthcare audio manipulation and states that the models and benchmark are not intended for medical diagnosis, treatment decisions, or standalone clinical security deployment.*

### 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?

*Section 3. We curated a pathology-aware deepfake dataset by generating codec-based fake speech from existing healthcare speech corpora using multiple publicly available neural audio codecs. The paper states that no new data were collected, no personally identifiable information was included, and all source corpora were used under their respective data-use agreements.*

- 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?

*Sections 3.1 and 3.3. We report the source datasets used to curate the pathology-aware deepfake benchmark, including corpus-level details and the controlled generation protocol. The paper also describes the train/dev/test setup, one-to-one mapping between bona fide and codec-generated samples, and speaker-disjoint splitting.*

### C. Did you run computational experiments?

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

*he paper reports optimizer, epochs, batch size, weight decay, gradient clipping, evaluation setup, and detailed model hyperparameters in the appendix.*

- 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,

---

*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.*

etc. or just a single run?

*The paper reports Accuracy, macro-F1, and EER, but it does not report variance estimates, confidence intervals, error bars, or indicate that results are averaged across multiple runs.*

**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.?

*The study does not involve newly recruited participants or annotators.*

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)?

*The study does not involve recruitment or payment of participants/annotators by the authors.*

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)?

*Section 3. The pathology-aware deepfake dataset was curated from existing healthcare speech corpora, and the paper states that access to the underlying corpora remains subject to their original licenses, consent provisions, and access agreements.*

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

*The study uses existing datasets available for research use under their respective licenses and access conditions, and does not involve a new data-collection protocol by the authors.*

**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?

*(left blank)*