

Reproducibility Checklist

Instructions for Authors:

This document outlines key aspects for assessing reproducibility. Please provide your input by editing this `.tex` file directly.

For each question (that applies), replace the “Type your response here” text with your answer.

Example: If a question appears as

```
\question{Proofs of all novel claims  
are included} { (yes/partial/no) }  
Type your response here
```

you would change it to:

```
\question{Proofs of all novel claims  
are included} { (yes/partial/no) }  
yes
```

Please make sure to:

- Replace ONLY the “Type your response here” text and nothing else.
- Use one of the options listed for that question (e.g., **yes**, **no**, **partial**, or **NA**).
- **Not** modify any other part of the `\question` command or any other lines in this document.

You can `\input` this `.tex` file right before `\end{document}` of your main file or compile it as a stand-alone document. Check the instructions on your conference’s website to see if you will be asked to provide this checklist with your paper or separately.

1. General Paper Structure

- 1.1. Includes a conceptual outline and/or pseudocode description of AI methods introduced (yes/partial/no/NA) **yes** - We give conceptual outline for **Cobra** backend and prompt-CoT tuning in Sec. 3.
- 1.2. Clearly delineates statements that are opinions, hypothesis, and speculation from objective facts and results (yes/no) **yes** - The methodology, assumptions, and error cases are explicitly labeled and discussed.
- 1.3. Provides well-marked pedagogical references for less-familiar readers to gain background necessary to replicate the paper (yes/no) **partial** - References to **GPT-4o**, **Gemini**, **DADA-2000**, **Whisper** etc are included; no tutorial-style explanations are provided.

2. Theoretical Contributions

- 2.1. Does this paper make theoretical contributions? (yes/no) **no**

If yes, please address the following points:

- 2.2. All assumptions and restrictions are stated clearly and formally (yes/partial/no) **NA**
- 2.3. All novel claims are stated formally (e.g., in theorem statements) (yes/partial/no) **NA**
- 2.4. Proofs of all novel claims are included (yes/partial/no) **NA**
- 2.5. Proof sketches or intuitions are given for complex and/or novel results (yes/partial/no) **NA**
- 2.6. Appropriate citations to theoretical tools used are given (yes/partial/no) **NA**
- 2.7. All theoretical claims are demonstrated empirically to hold (yes/partial/no/NA) **NA**
- 2.8. All experimental code used to eliminate or disprove claims is included (yes/no/NA) **NA**

3. Dataset Usage

- 3.1. Does this paper rely on one or more datasets? (yes/no) **yes**
If yes, please address the following points:
 - 3.2. A motivation is given for why the experiments are conducted on the selected datasets (yes/partial/no/NA) **yes** - We explain the relevance of **DADA-2000** and justify secondary testing on a Bilibili subset (Sec. 4.1).
 - 3.3. All novel datasets introduced in this paper are included in a data appendix (yes/partial/no/NA) **yes** - Our curated 20-video Bilibili subset is described in Sec. 4.1 with frame/display metadata.
 - 3.4. All novel datasets introduced in this paper will be made publicly available upon publication of the paper with a license that allows free usage for research purposes (yes/partial/no/NA) **yes** - The Bilibili-subset JSON labels and sampling pipeline will be released under CC BY-NC.
 - 3.5. All datasets drawn from the existing literature (potentially including authors’ own previously published work) are accompanied by appropriate citations (yes/no/NA) **yes** - **DADA-2000** and other external datasets are cited.
 - 3.6. All datasets drawn from the existing literature (potentially including authors’ own previously published work) are publicly available (yes/partial/no/NA) **yes** - **DADA-2000** dataset is open to the research community.
 - 3.7. All datasets that are not publicly available are described in detail, with explanation why publicly available alternatives are not scientifically satisfying (yes/partial/no/NA) **partial** - Bilibili videos may be removed by the uploader; archiving scripts will be

provided in README.

4. Computational Experiments

- 4.1. Does this paper include computational experiments? (yes/no) **yes**

If yes, please address the following points:

- 4.2. This paper states the number and range of values tried per (hyper-) parameter during development of the paper, along with the criterion used for selecting the final parameter setting (yes/partial/no/NA) **yes - Prompt-tuning used three variants with different instruction templates. Temperature parameter fixed at 0.0 for all models to ensure deterministic outputs.**

- 4.3. Any code required for pre-processing data is included in the appendix (yes/partial/no) **yes - The Cobra chunking and frame-sampling scripts are included in supplementary materials.**

- 4.4. All source code required for conducting and analyzing the experiments is included in a code appendix (yes/partial/no) **yes - Evaluation scripts for all baseline models and metric computation tools are included in supplementary materials.**

- 4.5. All source code required for conducting and analyzing the experiments will be made publicly available upon publication of the paper with a license that allows free usage for research purposes (yes/partial/no) **yes - GitHub repository will use MIT License or equivalent upon acceptance.**

- 4.6. All source code implementing new methods have comments detailing the implementation, with references to the paper where each step comes from (yes/partial/no) **yes - Key routines documented inline with references to Section 3 methodology are included in supplementary materials.**

- 4.7. If an algorithm depends on randomness, then the method used for setting seeds is described in a way sufficient to allow replication of results (yes/partial/no/NA) **NA**

- 4.8. This paper specifies the computing infrastructure used for running experiments (hardware and software), including GPU/CPU models; amount of memory; operating system; names and versions of relevant software libraries and frameworks (yes/partial/no) **yes - All experiments conducted on MacBook Pro with Apple M3 Pro chip (10-core CPU, 16-core GPU), 48GB unified memory, macOS Sonoma 14.5.0. Complete software dependency list provided in supplementary materials.**

- 4.9. This paper formally describes evaluation metrics used and explains the motivation for choosing these metrics (yes/partial/no) **yes - F1-score as harmonic mean of precision and recall for balanced evaluation.**

Precision for detection accuracy, recall for sensitivity measure.

- 4.10. This paper states the number of algorithm runs used to compute each reported result (yes/no) **yes - All performance metrics averaged over three independent runs.**

- 4.11. Analysis of experiments goes beyond single-dimensional summaries of performance (e.g., average; median) to include measures of variation, confidence, or other distributional information (yes/no) **yes - Reported standard deviation (SD = 0.499) across aligned videos, 95% CI of difference ([−0.126, 0.126]), Cohen's d = 0.000, examined per-video F1-score distribution and confusion matrices.**

- 4.12. The significance of any improvement or decrease in performance is judged using appropriate statistical tests (e.g., Wilcoxon signed-rank) (yes/partial/no) **yes - Paired t-test conducted over 95 aligned videos - t(94) = 0.000, p = 1.000, with full statistical report including t-value, p-value, degrees of freedom, and Cohen's d.**

- 4.13. This paper lists all final (hyper-)parameters used for each model/algorithm in the paper's experiments (yes/partial/no/NA) **NA**