

SYSTASYS Resource-Lite V2

MIREX 2026 - Music Evaluation via CMI-RewardBench

Author: Hichem Benali

System summary

Resource-Lite V2 is a zero-training, inference-time pairwise music preference system built for the MIREX 2026 CMI-RewardBench task. It compares Audio A and Audio B against the text prompt using a pinned LAION-CLAP checkpoint and a lightweight deterministic decision layer.

Evaluation protocol

The submitted predictions were produced in one blind pass on the official private test package. The full 618-item prediction file was frozen before any feedback or oracle was read. No hidden-set tuning, threshold tuning, reranking, or post-freeze repair was performed.

Reproducibility

The first-attempt prediction file contains 618 unique sample identifiers. Frozen SHA-256: 4e2125a60b397b258a5d4c22a473788481801e3a3ff6e47d2c84e467f1aa831a. The local freeze receipt records `oracle_exposed_before_freeze=false` and `feedback_read_before_freeze=false`.

Compute and dependencies

The method was executed locally on an already-owned Windows workstation using the pinned model checkpoint and local audio package. No paid API, hosted inference service, or external live evaluation platform was used to generate the submitted predictions.

Limitations

This is a deliberately resource-bounded system. It does not fine-tune on the private set and does not claim calibration to hidden human labels. The official MIREX score is therefore treated as the first external measurement of this frozen system.