

Experiment Status: Not production-ready (current run)

60-second takeaway

In this benchmark, CosyVoice2 baseline audio sounded acceptable while our CosyVoice3 finetune run did not meet production quality.

This is a run-specific outcome under one configuration, not a broad claim that CosyVoice3 is inherently worse.

We are treating CosyVoice2 as a control baseline and CosyVoice3 as an active rerun candidate.

Where this fits

- **For founders:** do not deploy this CosyVoice3 run as-is.
- **For engineers:** use this page as a diagnostic handoff for the next rerun.

Series overview:

- https://instavar.com/blog/IMDA_NSC_Voice_Cloning_Finetuning_Benchmark_2026

Result summary

CosyVoice2 is included as a baseline/control and produced acceptable qualitative output on our selected sample. CosyVoice3 finetuning in this run did not reach production-ready quality, with unstable long-form behavior and weaker linguistic consistency in listening checks.

Audio evidence

CosyVoice2 baseline/control

CosyVoice3 representative sample from this run

What this does and does not mean

This conclusion is specific to our exact setup: dataset shape, checkpoint path, prompt handling, and inference configuration. It should not be interpreted as a

universal model-family ranking.

Likely contributors in this run

- Checkpoint quality drift after early epochs.
- Sensitivity to prompt formatting and long-text decoding behavior.
- Operational fragility from large checkpoint churn and unstable inference zones.

Next rerun plan

- Rebuild a controlled CosyVoice3 run with stricter checkpoint gating and explicit long-text eval checkpoints.
- Keep a fixed prompt formatting harness for all comparisons.
- Re-evaluate using the same rubric and side-by-side audio panel used in this series.

Engineer appendix

Key paths from this run

- CosyVoice LoRA tools added: /mnt/work/chee-wei-jie/voice-models/CosyVoice/tools/train_cosyvoice3_lora.py, /mnt/work/chee-wei-jie/voice-models/CosyVoice/tools/infer_cosyvoice3_lora.py
- CV3 run artifacts referenced in sessions: /mnt/work/chee-wei-jie/voice-models/CosyVoice_runs/female01_cv3_lr1e5_run1/

Related deep dives

- [VoxCPM](#)
- [Qwen3-TTS LoRA](#)
- [IndexTTS2](#)