60-second takeaway
IndexTTS2 gave us a usable full-SFT baseline with strong operational predictability once we stabilized restart behavior.
In this run, `model_step14000.pth` was the practical checkpoint to keep.
The major challenge was process reliability and checkpoint retention policy, not core output quality.

Where this fits

- **For founders:** IndexTTS2 is a steady full-finetune option in this benchmark.
- **For engineers:** this page focuses on run recovery and checkpoint management as much as quality.

Series overview:

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

Experiment setup

- **Model:** IndexTTS2
- **Dataset:** IMDA NSC `FEMALE_01_44k` processed manifests
- **Hardware:** RTX 3090 Ti 24 GB
- **Training mode:** full SFT with resume

Best checkpoint logic

- Best validation region was around step ~13800.
- Saved checkpoints available around that region were `model_step14000.pth` and later `model_step15949.pth`.
- We treated `model_step14000.pth` as the practical best anchor for this run.

Audio evidence

Representative sample

`Settings:` long-text prompt comparison path, step 14000 checkpoint.

Failure modes we saw

- Training runs were interrupted multiple times and required explicit resume management.
- Some crashes were low-level (`pt_autograd_0` segfault signs), which made clean logs critical.
- Retention policy kept only recent checkpoint windows, so older steps disappeared automatically.

Recommended inference settings

- Keep checkpoint selection tied to both listening and nearest validation region.
- Prefer explicit run logs and resume metadata over implicit state.
- For this run, compare step 14000 and step 15949 only when you need late-run style differences.

Engineer appendix

Key paths from this run

- Checkpoints: `/mnt/work/chee-wei-jie/voice-models/FEMALE_01_44k/trained_ckpts_female01`
- Representative files: `model_step14000.pth`, `model_step15949.pth`

Operational note

- If you need longer history for post-hoc analysis, increase checkpoint retention before rerun.

Related deep dives

- VoxCPM
- Qwen3-TTS LoRA
- CosyVoice2 vs CosyVoice3