

# The Voice-Cloning Scorecard

8 voice engines evaluated, 5 tested hands-on — which one to clone your voice with, and what it costs.

ENGINE	TYPE	MIN REFERENCE	VERDICT
★ Cartesia Pro	Commercial · fine-tune	54 min studio	Winner — most natural in blind A/B
ElevenLabs	Commercial · zero/pro	~76 s to ceiling	Strong baseline; neutral prosody, tunable
★ LMNT	Freemium	5 s clip	Best free — 15k chars/mo, unlimited clones
Fish Audio	Commercial	15 s clip	Solid zero-shot clone
★ Coqui XTTS	Open source	5 s clip	Best OSS — 512-dim embedding, ~29 langs

## THE VERDICT

Cartesia for quality, LMNT for a free start, Coqui XTTS for open-source control. StyleTTS2, OpenAI and Deepgram were evaluated but not bench-tested. The ecosystem around the voice — lip-sync, avatars, distribution — matters more than the raw engine.

- Training-data quality follows a logarithmic curve: gains from 5s→60s of reference are steep, 60s→54min flatten.
- Only 2 engines handled German well — verify every non-English clone before you ship it.