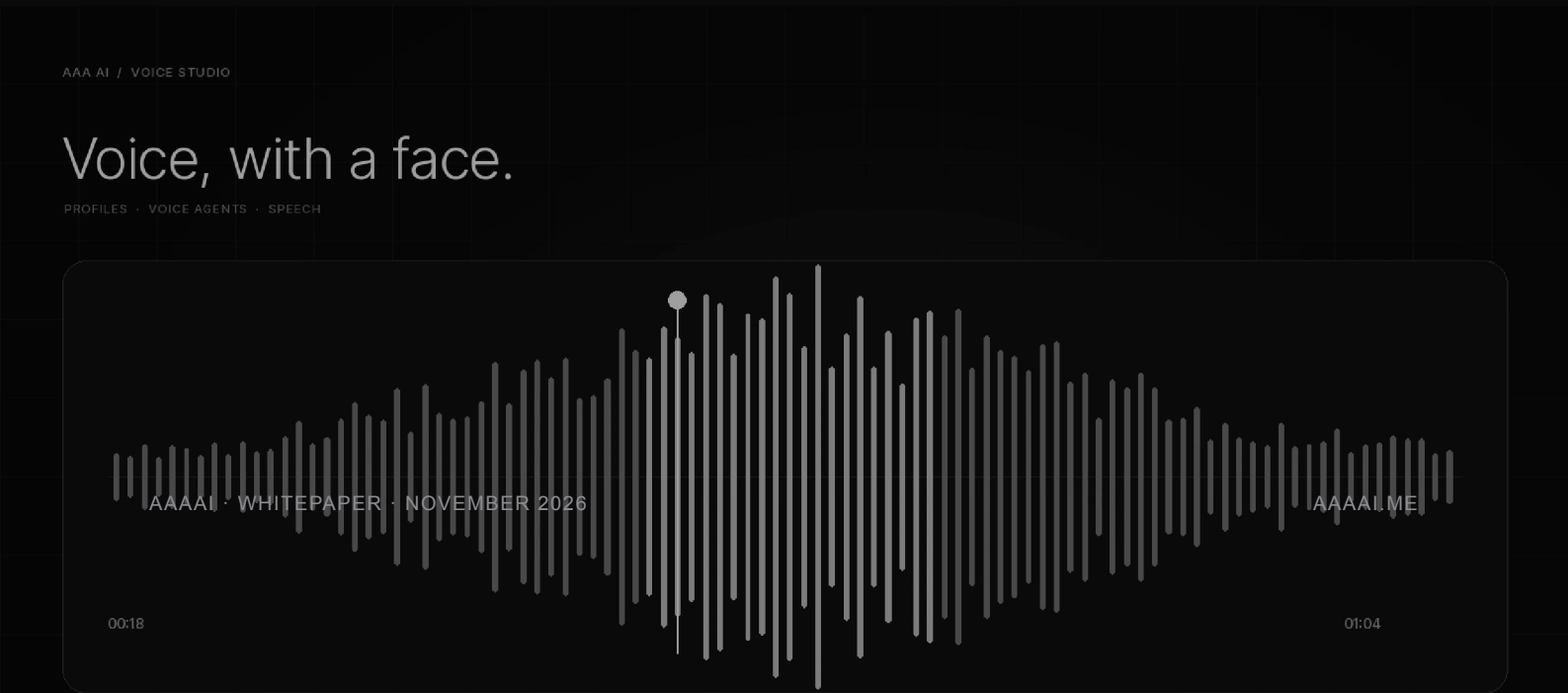
A|A|A AI

VOICE STUDIO

Voice

Reusable voice profiles, multi-provider TTS, captures and voice agents that answer — with read-only tools and explicit knowledge.



WHITEPAPER · 02

A voice layer, not a voice demo.

Voice Studio is the full voice I/O workspace: reusable profiles (cloned, preset or designed), personality and effects defaults, multi-provider speech generation, long-script chunking with crossfade, eight DSP effects, versioned takes and an async queue.

Captures use Whisper transcription and refinement. Voice agents add a persona, greeting, language, voice profile and optional LLM on top of the same profile.

Voice agents

- **Knowledge entries** retrieved into the prompt.
- **Pattern rules** that answer before the LLM or call an allowlisted webhook.
- **Talk/type sessions** plus POST /api/voice/agents/<id>/reply and /api/voice/speak.
- **Providers**: OpenAI, ElevenLabs, Coqui, OpenAI-compatible endpoints, offline system voice and Voicebox.

WHITEPAPER · 03

Latency, data and rollout.

Production notes

- Cloud or GPU-backed providers are recommended for low-latency agent calls; local CPU synthesis can take 10–20 seconds per short phrase.
- Metadata lives in PostgreSQL when configured; audio records write through to S3 per user.
- Reference samples need transcripts for cloning; store only consented voices.
- Voice agents should be bounded by explicit tools and knowledge, not open-ended actions.

Product URL: web.aaaai.me/voice-studio. Whitepaper covers profiles, cloning, agents, storage and latency.