MCP SERVER | NO CODE | CLOUD HOSTED

Cartesia (Voice AI) MCP for AI Agents

Generate high-fidelity speech synthesis and transcribe spoken word data

Cartesia (Voice AI) brings state-of-the-art voice synthesis and speech recognition to your AI client. Clone voices using just five seconds of audio, generate high-fidelity text-to-speech streams, or transcribe any audio file with industry-leading latency. It's built for building truly human conversational experiences.

A+ Quality Score 100/100

text-to-speech | speech-to-text | voice-synthesis | low-latency | ai-voice | audio-streaming

MAINTAINED BY VINKIUS CLOUD FOR AI AGENTS · JUNE 2026
vinkius.com/mcp/cartesia-voice-ai

SCAN TO VIEW

The connectivity layer between AI and the world's software.

Vinkius sits between AI and every application. All communication passes through Vinkius Cloud via the Model Context Protocol (MCP) — with governance, observability, and security at every layer.

Your AI Connections Run Through Vinkius Cloud

The world's largest managed MCP catalog

Vinkius is the connectivity layer where AI connects to the software your business already runs. We handle the hosting, the security, the credentials, the uptime — you get agents that actually do things.

We operate the world's largest managed MCP catalog. Major SaaS platforms, CRMs, databases, and cloud providers — running, monitored, production-ready. This MCP server is hosted and maintained by the Vinkius Cloud for AI Agents.

> *The agent doesn't manage credentials, doesn't manage uptime, doesn't manage security. Vinkius does.*
>
> — Architecture principle

Four Pillars of the Vinkius Runtime

01 — Security by design

Credentials stay encrypted at rest via AES-256. The AI agent never touches raw keys — they're injected into a sandboxed V8 isolate at runtime. Actions are logged, and connections have an emergency kill switch.

02 — Built on MCP Fusion

This MCP server was built with **MCP Fusion**, the open-source framework (Apache 2.0) that powers the entire Vinkius catalog. Schema-as-firewall strips undeclared fields, compiled PII redaction runs at zero overhead, and cryptographic lockfiles produce git-diffable audit trails.

03 — Deterministic observability

Eight immutable metrics per endpoint: request volume, p95 latency, error rate, active connections, cost attribution. A live payload feed logs every tool call with mutation detection.

04 — Autonomous operations

Servers are deployed, monitored, and patched autonomously. New capabilities and security patches ship weekly. Zero-downtime deployments ensure continuous availability across all managed MCP servers.

AES-256	Ed25519	24h TTL	V8 Isolate
Encryption at rest	PKI vault signatures	Ephemeral session keys	Sandboxed execution

One Token. Instant Access.

Every MCP server on Vinkius is accessed through a **Connection Token**. Tokens are generated in the cloud dashboard and produce a unique MCP endpoint URL. Paste this URL into any MCP-compatible client — no SDK required.

A single token can serve **multiple AI clients simultaneously**, or you can issue separate tokens per client for granular access control. Each token tracks its own request count, last activity timestamp, and can be individually enabled or revoked.

MCP ENDPOINT

```
https://edge.vinkius.com/{token}/mcp
```

Claude · Cursor · VS Code · Windsurf · Grok · Gemini

Security Is the Architecture

Security in Vinkius is not a feature — it's the foundation of the runtime. The gateway enforces multiple independent protection layers between AI agents and third-party APIs.

01 — Ed25519 PKI Vault

Every workspace has an Ed25519 Master Key. Session keys are generated ephemerally (24h TTL) and signed by the Master Key. Credentials never leave the vault boundary.

02 — V8 Isolate Sandboxing

Tool code runs inside isolated-vm V8 isolates with 64 MB memory caps and per-request timeouts. No filesystem access, no network access except through the SSRF-guarded fetch bridge.

03 — SSRF Guard

All outbound HTTP requests are DNS-resolved and validated before execution. Private IP ranges (10.x, 172.16-31.x, 192.168.x, AWS metadata 169.254.x) are blocked at the network layer.

04 — DLP & PII Redaction

A ResponseGuard pipeline intercepts every tool response. Configurable redaction patterns strip sensitive fields (emails, SSNs, card numbers) before data reaches the AI agent.

05 — Cryptographic Audit Trail

Every request is signed into a SHA-256 hash chain with Ed25519 signatures. Events form a tamper-proof, SIEM-exportable forensic record.

06 — Honeytoken Trap System

Phantom credentials are injected into isolated environments. If a honeytoken is used outside Vinkius infrastructure, the server is quarantined instantly.

Emergency Kill Switch

EU AI Act Art. 14(1) Compliant

The kill switch is an **emergency halt** mechanism — not a simple toggle. When triggered, it executes three actions atomically:

01 — Server deactivated

The MCP server is immediately taken offline across the entire cluster.

02 — All tokens revoked

Every connection token is invalidated. Total lockout — reconnection blocked until new tokens are issued.

03 — WebSocket connections killed

Active connections terminated via Redis pubsub broadcast. Propagates to every runtime node in the cluster.

Full Visibility. Zero Guesswork.

The Vinkius cloud dashboard includes a full MCP Governance suite — real-time analytics and security controls for production AI operations.

Control Plane

KPI dashboard with request volume, latency, success rate, token consumption, and AI-generated operational briefings.

Agent Activity

Which AI clients are connecting, how often, and what they're doing — real-time session tracking.

FinOps

Cost tracking per tool, payload compression savings, budget optimization signals, and consumption trends.

Tool Health

Slowest and most error-prone tools, with actionable root-cause insights and performance baselines.

Firewall & DLP

PII redaction activity, sensitive data protection counters, and security event timeline.

Incident Log

Error trends, failure rates, status-code breakdowns, and forensic audit trail access.

> Get started at **cloud.vinkius.com** — connect your AI agent in under 60 seconds.

Cartesia (Voice AI) MCP

20 tools available
Cloud-hosted on Vinkius

This MCP connects powerful voice processing into anything your agent runs on. You can build applications where the AI speaks and understands like a person—not a robot reading text.

Need to generate natural audio? Use high-fidelity models to synthesize speech, or stream it out in real time via SSE for low latency. Want to make sure your brand voice is consistent? Clone voices from minimal samples of audio input, then adapt that voice to different languages and dialects. Need the AI to understand something complicated? Transcribe any spoken audio file into text using advanced models that support multiple languages.

It’s also great for maintaining context. You can manage custom pronunciation dictionaries so the AI says specialized or technical terms correctly every time, even across complex agent orchestration flows. If you're building a sophisticated application, Vinkius makes connecting this voice intelligence to your existing workflows simple and reliable.

Core Capabilities

01 — Generate realistic speech audio

Convert text into high-quality audio bytes or stream the output instantly using advanced TTS models.

02 — Transcribe spoken word to text

Process and convert any audio file, regardless of language, into accurate written text.

03 — Create custom voice profiles

Build entirely new, personalized voices using short samples of existing human speech.

04 — Modify and manage voices

Get details about available voices, update their metadata, or even delete them when they're no longer needed.

05 — Control specific pronunciations

Create and maintain custom dictionaries to ensure the AI pronounces technical names or foreign words exactly right.

One Click on Vinkius — From Prompt to Execution

Available at **vinkius.com/mcp/cartesia-voice-ai** — connect your AI agent in three steps.

01 Subscribe to this MCP and provide your Cartesia API Key.

02 Your agent calls a function, specifying the action (e.g., generating audio) and providing the necessary input data like text or an audio file.

03 The MCP processes the request using its voice models and returns the resulting audio stream or transcribed text to your client.

The bottom line is that you just tell your AI agent what you need—a voice, a transcription, or a spoken message—and it handles the complex generation process.

Built For

This MCP serves anyone building applications where speech and audio are core features. It's for product teams needing conversational agents that sound human, content creators automating voiceovers globally, or developers integrating real-time audio into existing systems.

Conversational AI Developer

Build complex agent pipelines where the AI must not only process text but also speak and react with natural, low-latency voices.

Media Localization Specialist

Automate the voiceover process for global content. Use cloned voices to adapt a single script into dozens of languages while maintaining brand identity.

UX/Product Designer

Integrate speech synthesis directly into product workflows, ensuring that user feedback or system alerts are delivered with professional quality audio and timing.

What Changes When You Connect

01 Achieve true conversational depth. Use `tts_sse` to stream audio in real time, making your agent feel responsive instead of delayed.

02 Maintain brand consistency globally. Clone a voice using just five seconds of audio via `clone_voice`, then adapt it across regions using `localize_voice`.

03 Eliminate mispronunciation errors. Use `create_pronunciation_dict` to lock down how your AI agent speaks specialized terminology, ensuring technical accuracy every time.

04 Process large amounts of data easily. Run bulk transcriptions on hours of audio files using `stt_batch`, saving manual effort across content teams.

05 Build sophisticated call tracking. Use `list_agent_calls` to track exactly what your agents talked about and how many credits were used.

Real-World Applications

Building a multilingual customer service bot

A support company needs their agent to handle calls in Spanish, German, and French. They use `localize_voice` on one core voice model, ensuring the tone remains consistent while adapting the audio output for each language.

Automating video podcast production

A content creator has many interviews to turn into episodes. Instead of hiring a voice actor, they use `clone_voice` on their own voice and then run `tts_bytes` to generate the entire script's audio track instantly.

Analyzing recorded user feedback

A product team records hundreds of video calls with users. Instead of listening manually, they feed all the audio into `stt_batch`, getting clean text transcripts that can be analyzed for key pain points.

Creating dynamic narrative audiobooks

An audiobook developer needs a narrator who sounds consistent but also needs to speak specialized scientific terms correctly. They use `create_pronunciation_dict` and then generate the entire book's narration using high-quality TTS.

Patterns to Avoid

Treating audio like a file upload

✗ AVOID

Manually uploading large batches of audio files one by one into a portal and waiting hours for results. This is slow and doesn't scale past small projects.

✓ INSTEAD

Use the `stt_batch` tool to process entire folders of audio files in one go, making bulk transcription quick and efficient.

Assuming voice consistency across languages

✗ AVOID

Taking a single recorded English voice model and simply hoping it sounds natural when translated into Japanese or Arabic. The result is usually robotic and unnatural.

✓ INSTEAD

Always use `localize_voice` to adapt your core voice profile, ensuring the resulting audio sounds native and appropriate for the new dialect.

Ignoring technical jargon

✗ AVOID

Having an agent explain a complex medical term like 'myocardial infarction' and having it pronounced incorrectly because the system doesn't know how to say it.

✓ INSTEAD

Define custom word rules using `create_pronunciation_dict` so your AI agent speaks every specialized term with perfect, intended accuracy.

The Right Fit

Use this MCP if generating or understanding human speech is central to your product's core value. For instance, if you need an agent to read content aloud or summarize a voice call, Cartesia handles it. However, don't use this just because you want basic text-to-speech; you need the low latency and control offered by `tts_sse`. Also, if your primary need is merely storing recordings for later analysis, other simple storage solutions might suffice. But when you need to *process* that audio—cloning a voice, adapting it, or transcribing it in bulk—this MCP provides the necessary depth and controls.

Cartesia (Voice AI) MCP: Solving complex audio localization challenges

Right now, localizing content is a nightmare. You record an actor for English, then you have to hire a completely different person in Mandarin who might sound slightly different, and even if they nail the accent, matching the original emotional tone is nearly impossible.

With Cartesia (Voice AI), you clone your core voice once. Then, using `localize_voice`, you adapt that single profile for multiple languages. You get consistent quality, perfect vocal fidelity, and a massive time savings without compromising brand identity.

Cartesia (Voice AI) MCP: Ensuring accurate speech recognition in agents

Manual transcription is slow. You record a meeting and then have to copy the audio into a separate service, hoping it captures every technical term correctly. It's tedious, time-consuming, and prone to error.

The MCP lets your agent run `stt_batch` directly on large volumes of recorded speech. This gives you accurate, machine-processed text outputs right where you need them—integrated into your workflow.

Cartesia (Voice AI): 20 Tools for Speech Synthesis and Audio Processing

Use these tools to manage voices, generate speeches, transcribe files, and control pronunciation within your agent's workflows.

#	TOOL	DESCRIPTION
01	`get_voice`	Retrieves specific metadata for a known voice model.
02	`list_agent_calls`	Shows a record of past calls and transcripts handled by a particular agent.
03	`update_voice`	Changes general information or metadata associated with an existing voice model.
04	`clone_voice`	Creates a custom, unique voice profile from a small audio clip of five seconds or longer.
05	`create_pronunciation_dict`	Establishes a new list of specific word pronunciations for the AI to follow.
06	`delete_pronunciation_dict`	Removes an existing custom pronunciation dictionary entirely.
07	`delete_voice`	Permanently removes a voice model from the system.
08	`generate_access_token`	Creates a temporary token needed for running client-side requests securely.
09	`get_agent`	Fetches detailed information about a specific configured voice agent.
10	`get_usage_credits`	Retrieves current statistics on the account's remaining usage credits and billing history.
11	`infill_bytes`	Generates audio content to smoothly bridge a gap between two existing audio segments.
12	`list_agents`	Provides an overview of all configured voice agents within the account.
13	`list_pronunciation_dicts`	Lists all custom pronunciation dictionaries that have been created.
14	`list_voices`	Returns a comprehensive list of every available voice model in the system.

#	TOOL	DESCRIPTION
15	`localize_voice`	Adapts an existing voice profile to sound natural in a new language or regional dialect.
16	`stt_batch`	Transcribes multiple audio files into text format efficiently, suitable for bulk processing.
17	`tts_bytes`	Generates and returns the full audio data bytes from a given text input.
18	`tts_sse`	Streams generated speech audio in real time using Server-Sent Events for immediate playback.
19	`update_pronunciation_dict`	Modifies or corrects specific word pronunciations within an existing dictionary.
20	`voice_changer_bytes`	Alters the voice of a provided audio clip while carefully preserving its original emotion and intonation.

See It in Action

Real prompts you can use once this MCP is connected to your AI agent through Vinkius Cloud.

U I just finished a call with a customer. Can I get a summary transcript and also check how many calls we've handled this month?

Call Summary & Usage Report

- **Date:** 2026-06-19
- **Duration:** 8:45
- **Key Topics:** Billing dispute, feature request (dark mode), next steps.
- **Action Items:** Follow up with user on billing by end of day.

📊 **Account Usage**
You have used 12,300 credits this month. Last cycle refreshed: June 1st.

U I need to generate a sample audio file saying 'The next quarter's focus is on cloud migration.' What voice should I use?

Audio Generation Report

✅ **Voice Selected:** 'Baritone' (ID: 79a045e3...)

- **Model Used:** Sonic-3.5
- **Output Format:** WAV audio stream (ready for download)
- **Latency:** Sub-second generation confirmed.

The file is attached, ready to play.

U List all the custom voices we have and check if 'Spanish' is an option.

Available Voice Models

Here are your active models:

- ✅ Baritone (ID: 79a045e3...)
- ✅ British Lady (ID: 123x-456y...)
- 👤 Custom Clone - User A (ID: [C L O N E D]...)

I found a model labeled 'Latin American Female' which supports Spanish. You can use `get_voice` for its specific ID.

Frequently Asked Questions

01 **How do I make my AI agent sound like me, even if I only record myself briefly?**

You clone your voice using a short audio clip. This creates a unique digital model of your speaking patterns and tone that the AI can use across all its outputs, maintaining brand consistency.

02 **Does Cartesia (Voice AI) support transcribing different languages?**

Yes. The system handles multi-language transcription, meaning you don't have to worry about language switching when processing audio files into text for your agents.

03 **Is the generated speech low latency enough for a real-time chat agent?**

Absolutely. By streaming audio via Server-Sent Events, the system delivers synthesized sound almost instantly, making the conversation flow naturally and feel highly responsive to the user.

04 **What if my company has specialized terminology that sounds wrong when spoken by the AI?**

You solve this with pronunciation dictionaries. You define exactly how a specific word or acronym should sound, and the MCP forces the agent to say it correctly every time.

05 **Can I update my voice models if they need new metadata or changes?**

Yes, you can manage existing voices by calling `update_voice`. This lets you modify details like model descriptions or usage parameters without changing the actual sound profile.

Go Live in 60 Seconds

Get your connection token from **cloud.vinkius.com**, then paste the endpoint URL into any MCP-compatible client.

YOUR MCP ENDPOINT

```
https://edge.vinkius.com/[TOKEN]/mcp
```

CLIENT	WHERE TO CONFIGURE
Claude AI	Profile → Customize → Connectors → "+" → Add custom connector → Paste endpoint
Cursor	Settings → Features → MCP Servers → "+ Add New MCP Server" → Type: SSE → Paste endpoint
VS Code	Ctrl/Cmd+Shift+P → "MCP: Add Server" → add `"cartesia-voice-ai": { "url": "…" }`
Windsurf	MCP Settings → `mcp_settings.json` → Add endpoint URL
ChatGPT	Settings → Tools & plugins → Add MCP server → Paste endpoint
Gemini	Extensions → Add MCP Server → Paste endpoint URL

ASK AN AI ABOUT THIS

Let your preferred AI explain this MCP server

- Ask ChatGPT
- Ask Claude
- Ask Perplexity
- Ask Gemini
- Ask Grok

READY TO CONNECT

Cartesia (Voice AI) is live on Vinkius Cloud.

Get your connection token, paste it into your AI agent, and start building. No SDK. No deployment. Just results.

Start at cloud.vinkius.com →

vinkius.com · support@vinkius.com

INDEPENDENT PLATFORM DISCLAIMER

Vinkius is an independent platform and is not affiliated with, endorsed by, sponsored by, verified by, or otherwise authorized by Cartesia (Voice AI). All third-party trademarks, logos, and brand names are the property of their respective owners. Their use in this document is strictly for informational purposes to identify service compatibility and interoperability.

DOCUMENT INFORMATION

Generated	June 2026
MCP Server	Cartesia (Voice AI) MCP
Server ID	`019e3874-a740-7258-9692-87f651d07053`
Platform	Vinkius Cloud for AI Agents
Endpoint	`https://edge.vinkius.com/{token}/mcp`

LICENSE & USAGE

This document is generated automatically by the Vinkius PDF Engine. Content reflects the MCP server configuration at the time of generation and may change as updates are deployed. For the most current information, visit vinkius.com/mcp/cartesia-voice-ai.