

MCP SERVER

NO CODE

CLOUD HOSTED

# Retell AI MCP

Orchestrate full-scale, real-world voice interactions.

Retell AI MCP connects your agent to a full telecommunications stack. It lets your AI client handle complex voice interactions, like initiating outbound calls or registering web-based sessions. You can build and manage entire AI personas, configure their underlying language models, and analyze call transcripts for sentiment analysis—all through simple chat commands.

**A+** Quality Score 100/100

voice-ai

conversational-agents

telephony

real-time-audio

voice-automation

customer-interaction



# 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.

### 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.

### 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.

### 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**

Encryption at rest

**Ed25519**

PKI vault signatures

**24h TTL**

Ephemeral session keys

**V8 Isolate**

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.

### 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.

### 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.

### 06 — Honeypot Trap System

Phantom credentials are injected into isolated environments. If a honeypot 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.

**FinOps**

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

**Firewall & DLP**

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

**Agent Activity**

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

**Tool Health**

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

**Incident Log**

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

Get started at [cloud.vinkius.com](https://cloud.vinkius.com) — connect your AI agent in under 60 seconds.

# Retell AI MCP

10 tools available

Cloud-hosted on Vinkius

Your agent needs to talk on the phone? This MCP connects it directly to Retell AI's voice platform. Instead of just chatting with text, your AI client can now orchestrate actual real-world calls. You can command your assistant to make an outbound call or set up a web browser session for testing. It's like giving your chatbot a functioning phone line and a full operational team behind it.

Need to manage the agents themselves? You can dynamically create new AI personalities, adjusting their core instructions and models in one go. When the calls are done, you don't have to manually sift through records; you simply ask your agent to list all historical activity or pull detailed transcripts for deep analysis. Because this MCP lives on Vinkius, you connect once from any compatible client, giving your team access to managing voice agents and call telemetry without needing to log into a dozen separate dashboards.

---

## Core Capabilities

### 01 — Make Outbound Calls

Directly initiate phone calls using specified source and destination numbers.

### 03 — Record Web Calls for Testing

Register a new web-based call session, useful for testing browser integrations.

### 05 — List Communication Assets

Check available phone numbers, text-to-speech voices, and configured agent profiles.

### 02 — Manage AI Agent Personas

Create or update the specific instructions and models that define how an AI voice agent behaves.

### 04 — Review Call History and Details

Pull logs of all past calls or retrieve the full transcript and sentiment analysis for one specific conversation.

# One Click on Vinkius — From Prompt to Execution

Available at [vinkius.com/mcp/retell-ai](https://vinkius.com/mcp/retell-ai) — connect your AI agent in three steps.

- 01 Install the Retell extension module into your MCP environment.
- 02 Get your API Key from your provider account and input it into the required authentication parameter.
- 03 Tell your assistant what you want to do, like: 'List all voices, create a support agent, and test a call to 555-1234.'

The bottom line is that you chat naturally with your AI client, and it runs the complex telephony actions for you.

---

## Built For

This MCP targets anyone running customer-facing automation. Think operations managers who are tired of manually checking call logs or support leads who need to spin up a new, temporary agent personality immediately.

### Customer Support Manager

Sets up autonomous agents that handle initial triage and qualify leads during inbound/outbound calls.

### DevOps Engineer

Prototypes complex voice workflows, using the MCP to manage agent configurations without building custom admin panels.

### Sales Operations Lead

Runs batch outbound calls for follow-up or qualification, then uses call details to generate post-call reports.

---

## What Changes When You Connect

- 01 You don't need to write code to manage agents. Use the `create_agent` and `update_agent` tools to define new AI personalities and adjust their instructions via plain chat commands.

- 
- 02** Analyze call outcomes instantly. After a conversation, use `get_call_details` to pull transcripts and sentiment analysis without opening any separate dashboard tabs.
- 
- 03** Test web integrations easily. The `register_web_call` tool lets you set up virtual call sessions for debugging and quality assurance before going live.
- 
- 04** Control the entire communication stack. You can check all available assets—from calling numbers via `list_phone_numbers` to voice options using `list_voices`.
- 
- 05** Execute complex actions in one chat session. Combine tools like `create_agent`, `create_llm`, and `create_phone_call` into a single, natural language request.
- 

---

## Real-World Applications

### Quality Assurance Scripting

A QA engineer needs to test how the new support bot handles objections. They ask their agent to 'Register a web call and then initiate a mock objection handling flow.' The system uses `register_web_call` and simulates the full interaction, allowing immediate bug reporting.

### Scaling Support Coverage

The support team suddenly needs a specialized agent. Instead of spending an hour in a dashboard, they ask their agent to 'Create a new fraud detection agent.' The system uses `create_agent` and `create_llm`, deploying the persona instantly.

### Post-Mortem Call Review

A sales manager needs to know why a deal stalled. They ask their agent, 'Show me the details for call ID 987.' The system uses `get_call_details`, providing the full transcript and confirming the sentiment score was negative.

### Compliance Auditing

A compliance officer needs to audit all activity from last week. They ask their agent to 'List calls between Monday and Friday.' The system uses `list_calls` to pull the required log data for review.

---

# Patterns to Avoid

---

## Treating it like a simple messaging tool

### ✗ AVOID

Thinking you just need to send text messages. Trying to use a generic chat bot for telephony tasks, which can't handle the complexity of real-time audio or number routing.

### ✓ INSTEAD

For actual voice interactions, always use ``create_phone_call``. This tool handles the full telecommunications stack and provides necessary call logging via ``list_calls``.

---

## Ignoring agent management

### ✗ AVOID

Building a bot that works perfectly in testing but can't adapt when product requirements change. The bot needs constant manual reconfiguration.

### ✓ INSTEAD

Use the MCP to manage your agents. You can use ``update_agent`` or ``create_llm`` to instantly adjust the core instructions of an existing persona without redeploying code.

---

## Not checking assets first

### ✗ AVOID

Attempting to dial a number, only to find out the calling line is inactive or the voice sounds robotic and generic.

### ✓ INSTEAD

Always check your resources first. Use ``list_phone_numbers`` before making calls, and use ``list_voices`` to ensure you select the right text-to-speech voice for quality.

---

## The Right Fit

Use this MCP if your core business process involves any form of two-way conversation—whether it's a phone call, a web chat simulation, or an internal routing system. Specifically, you need to programmatically *act* on the voice interaction (e.g., making calls, recording sessions) and then analyze the results (transcripts, sentiment). Don't use this if you just need basic data lookup; for example, listing all agents is fine with `list_agents`, but if you need to *change* those agents or *make* a call, you must use these tools. If your primary need is purely text-based messaging and never involves voice calls, look at dedicated chat integration MCPs instead.

---

## The current process for testing voice interactions is slow and painful.

Right now, setting up a test call means logging into the telephony dashboard, manually selecting numbers, configuring agent scripts in a separate GUI, hitting 'test,' waiting 20 minutes to see the transcript, then copying data into a spreadsheet for manual review. It's a cycle of clicks and context switching.

With this MCP, you simply tell your agent what you want done—for instance, 'Run a test call using Agent X.' The system handles the number selection, the connection, the recording, and makes all the results available instantly in the chat window. You get immediate action and detailed data.

---

## Retell AI MCP Gives You Full Voice Control.

You no longer have to leave your main workflow just to manage communication assets. Listing numbers, checking voices, or adjusting an agent's core identity (using `update_agent` ) happens right where you are working.

This MCP moves voice automation from a series of siloed tools into one cohesive conversational process. Your AI client becomes the central operating system for all your telecommunication needs.

---

# Retell AI With 10 Tools

These tools let you build, configure, and audit the entire lifecycle of a voice agent's interaction, from initial setup to final transcript analysis.

#	TOOL	DESCRIPTION
01	<code>create_agent</code>	Creates and deploys a new AI voice agent personality.
02	<code>create_llm</code>	Configures the specific language model that powers an agent's conversation brain.
03	<code>create_phone_call</code>	Initiates a new outbound call by specifying the calling and receiving numbers.
04	<code>get_call_details</code>	Retrieves comprehensive data for a specific phone call, including transcripts and sentiment scores.
05	<code>list_agents</code>	Lists all AI agent profiles currently configured in the account.
06	<code>list_calls</code>	Provides a log of both historical and active calls.
07	<code>list_phone_numbers</code>	Shows all phone numbers associated with the account for calling.
08	<code>list_voices</code>	Displays every available text-to-speech voice option.
09	<code>register_web_call</code>	Sets up a new web browser session for call testing or integration.
10	<code>update_agent</code>	Modifies the instructions or parameters of an existing AI voice agent.

---

## See It in Action

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

**U** Can you show me the transcripts for call ID `c\_f3a123`?



I accessed the tool `get_call_details` probing for the unique identifier `c_f3a123`. The call lasted 4 minutes and the sentiment analysis registered as 'positive'. Below is the full transcript of the conversation generated by Retell!

**U** List all available agents I can use.



I sent a request evaluating all active profiles using `list_agents`. I discovered 3 agents in your account, including your 'Support Bot' (ID: agent\_1), 'Sales Executive', and 'Reception Voice'. Let me know if you would like me to modify any of their instructions using `update_agent`.

**U** We are testing out new numbers. Please use 'from\_number' `+18005551234` and dial `+14085551234` assigning my 'agent\_555'.



Done. I engaged `create_phone_call` injecting your JSON settings to ring out! Retell processing is initiated. Monitor your active calls using `list_calls` to see when it connects and check its metrics after completion.

---

## Frequently Asked Questions

### 01 How do I get an agent to make a call using Retell AI MCP?

You use the `create_phone_call` tool by providing the required JSON payload, specifying both the calling number and the receiving number.

---

---

**02 Can I see what voices are available with Retell AI MCP?**

Yes, you call ``list_voices``. This tool pulls a complete list of all text-to-speech options your account supports so you can pick the best one for your agent.

---

**03 What if I need to change an existing AI bot's behavior?**

You use ``update_agent``. This tool lets you modify the instructions and parameters of a deployed agent without having to recreate it from scratch.

---

**04 How do I find out what calls happened yesterday with Retell AI MCP?**

Call ``list_calls`` first for a list of all call IDs. Then, use the specific ID you found with ``get_call_details`` to retrieve transcripts and sentiment.

---

**05 Does Retell AI MCP work for web browser testing?**

Yes, it does. You use ``register_web_call`` to establish a simulated web session, allowing you to test your agents in a live-like environment before deployment.







---

# Go Live in 60 Seconds

Get your connection token from [cloud.vinkius.com](https://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
 <b>Claude AI</b>	Profile → Customize → Connectors → "+" → Add custom connector → Paste endpoint
 <b>Cursor</b>	Settings → Features → MCP Servers → "+ Add New MCP Server" → Type: SSE → Paste endpoint
 <b>VS Code</b>	Ctrl/Cmd+Shift+P → "MCP: Add Server" → add <code>"retell-ai": { "url": "..." }</code>
 <b>Windsurf</b>	MCP Settings → <code>mcp_settings.json</code> → Add endpoint URL
 <b>ChatGPT</b>	Settings → Tools & plugins → Add MCP server → Paste endpoint
 <b>Gemini</b>	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

# Retell 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](https://cloud.vinkius.com) →

[vinkius.com](https://vinkius.com) · [support@vinkius.com](mailto: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 Retell 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	Retell AI MCP
Server ID	019d75fe-fd3b-7223-b7a0-275b113c4ab7
Platform	Vinkius Cloud for AI Agents
Endpoint	<a href="https://edge.vinkius.com/{token}/mcp">https://edge.vinkius.com/{token}/mcp</a>

### 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/retell-ai](https://vinkius.com/mcp/retell-ai).