MCP SERVER | NO CODE | CLOUD HOSTED

Tavily MCP

Turn Web Research Into a Conversation

Tavily brings professional web research into your AI agent. It transforms complex data gathering—like auditing search context, pulling clean articles from URLs, or getting instant answers to obscure questions—into a simple conversation with your AI client.

A+ Quality Score 100/100

search-engine | ai-agents | data-extraction | context-retrieval | web-research | real-time-data

MAINTAINED BY VINKIUS CLOUD FOR AI AGENTS · JUNE 2026
vinkius.com/mcp/tavily

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.

Tavily MCP

6 tools available
Cloud-hosted on Vinkius

This MCP lets you give your agent the ability to act like a dedicated research assistant. Instead of opening ten tabs and copy-pasting information, your agent goes out onto the web and does the heavy lifting for you. It’s designed specifically so that when your agent talks to the internet, it gets optimized, structured data back—not just random search results. You can ask it about a topic, and it will instantly get the necessary context, summarize complex findings from multiple sites, or even pull clean text from articles you link it to. Whether you're deep into market analysis or tracking breaking news, your agent provides an intelligence layer that keeps everything grounded in real-time sources. Since Vinkius hosts this MCP, your agent gets immediate access to professional web research capabilities across any compatible client.

Core Capabilities

01 — Conducting optimized web searches

Your agent executes targeted queries that pull search results specifically formatted for AI consumption.

02 — Extracting direct answers from web sources

The agent retrieves concise, synthesized answers to complex questions without needing manual data synthesis.

03 — Pulling clean text from specific URLs

It strips away ads and clutter, giving you only the main, readable content from any webpage link.

04 — Monitoring current events

You can ask your agent to track specialized news results to stay updated on industry changes in real time.

05 — Gathering visual research context

The agent searches and provides high-quality image sets, maintaining a necessary visual context for reports.

One Click on Vinkius — From Prompt to Execution

Available at **vinkius.com/mcp/tavily** — connect your AI agent in three steps.

01 Subscribe to this MCP and provide your API key.

02 Connect the service to your preferred AI client (Claude, Cursor, etc.).

03 Ask your agent a research question; it executes the search and returns structured data.

The bottom line is you get reliable web data fed directly into your conversation flow without ever leaving your chat interface.

Built For

This is for researchers, analysts, developers building complex agents, and content strategists who spend too much time copy-pasting. If you get frustrated with manual web browsing or having to manually piece together data from multiple sources, this MCP is your fix.

Market Analyst

They use the agent to run systematic research on industry trends, collecting context and extracting key findings from competitor websites.

Software Developer

They integrate this MCP into their agents to pull live documentation or external data sources when writing code, keeping it relevant.

Content Strategist

They ask the agent to monitor trending topics and retrieve AI-generated answers for article ideas, speeding up content planning cycles.

What Changes When You Connect

01 Get structured context immediately. Instead of wading through raw search result pages, using `get_search_context` gives your agent optimized snippets that are ready for the AI to analyze.

02 Skip manual synthesis entirely. When you need a quick answer on a complex topic, asking the agent to use `get_answer` provides a summarized conclusion right away.

03 Eliminate web clutter. If you find an important article but hate reading boilerplate text and ads, running `extract_content` pulls only the clean, actionable body copy.

04 Stay current with zero effort. Monitoring niche updates or industry news is easy when your agent runs `search_news`, keeping you instantly updated on breaking events.

05 Enhance visual reports. When context requires images—like for a product pitch—using `search_images` ensures the agent finds relevant, high-quality visuals.

Real-World Applications

Comparing competitor pricing and features

A marketing manager asks their agent to gather data on three competitors. The agent runs multiple searches and uses `extract_content` on the top result pages for each company, giving a side-by-side comparison of product tiers.

Writing an article on emerging tech

A content strategist asks their agent about 'quantum computing breakthroughs.' The agent performs a general `search_web` and then uses the retrieved context to give a direct summary, preventing the need for manual source checking.

Monitoring regulatory changes

An operations lead needs to know if there were any changes in compliance laws this week. The agent runs `search_news` and filters results specifically for legal updates, providing a focused report on the newest regulations.

Developing an AI chatbot knowledge base

A developer needs external data to fine-tune their agent's responses. They ask the MCP to retrieve optimized context using `get_search_context` for a specific topic, building a robust and reliable internal dataset.

Patterns to Avoid

Using generic search queries

✗ AVOID

Asking your agent to 'tell me about AI' without telling it where to look. It gets overwhelming results that are too broad or too thin.

✓ INSTEAD

Instead, focus the query. Ask for specific context using `get_search_context` on a narrow topic, like 'LLM deployment strategies in finance.' This guarantees structured, actionable data.

Trying to read full articles manually

✗ AVOID

Copy-pasting the whole text from a long whitepaper into your prompt and asking the AI to summarize it. You risk passing too much noise.

✓ INSTEAD

Use `extract_content` first on the PDF's link, or target specific sections. This gives the agent clean text that is ready for synthesis, not just raw data.

Ignoring real-time updates

✗ AVOID

Relying on old blog posts when a major industry announcement happened hours ago. Your intelligence is out of date.

✓ INSTEAD

Always start with `search_news` for the absolute latest information. This ensures your entire workflow operates on current, time-sensitive data.

The Right Fit

Use this MCP if you need reliable web context feeding into an AI agent's conversation flow. You need to transform open research questions into structured, usable data points. If your goal is synthesis—taking scattered pieces of information and making them coherent—this is essential. Don't use it if you just want a simple Google search result list; for that, traditional search engines work fine. Also, don't use it if your task involves complex calculations or internal database queries; those require a dedicated data source MCP instead. When in doubt, remember: if the information needs to come from the open web and must be structured for an agent, this is the tool.

The Pain of Researching Anything Online

You know the process. You start with a question, then you open five different browser tabs: one for news, one for a competitor's site, and three others just to find supporting data points. Then, you spend the next hour clicking through paywalls, scrolling past ads, and manually copying specific paragraphs into a single document.

With this MCP, that whole sequence disappears. You simply talk to your agent: 'Find me the latest news on quantum computing and pull out three key stats from Company X.' The agent handles the tab opening, the clicking, the data extraction, and gives you one clean, actionable summary.

Tavily MCP: Structured Web Data for Your Agent

No more manual copy-pasting. You don't have to worry about the noise of the web; the MCP handles filtering out ads, boilerplate text, and irrelevant fluff automatically.

What changes now is speed and reliability. Instead of wasting hours compiling data, you get instant access to structured context from multiple sources, making your agent smarter faster.

Tavily: Web Research Tools (6)

These tools let you search the web for AI-optimized results, extract content from specific URLs, and gather structured context for complex agent queries.

#	TOOL	DESCRIPTION
01	`extract_content`	Pulls only the clean, readable text from a specific list of URLs.
02	`get_answer`	Generates a direct, synthesized answer for a complex search query.
03	`get_search_context`	Retrieves highly structured and optimized context snippets perfect for feeding into an AI model.
04	`search_images`	Searches the web specifically for high-quality images relevant to a topic.
05	`search_news`	Monitors and retrieves results focused on real-time news and current events.
06	`search_web`	Conducts general web searches using results optimized for AI processing.

See It in Action

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

U Search for the latest breakthroughs in 'Quantum Computing' using Tavily.

I've retrieved the latest research context. Recent breakthroughs include new error-correction methods and increased qubit coherence times. Would you like the links to the full papers?

U Get an AI answer for 'How does photosynthesis work?'.

According to the latest sources, photosynthesis is the process used by plants to convert light energy into chemical energy. I can provide a more detailed breakdown if needed.

U Extract content from https://vinkius.com.

Content extracted! I've retrieved the main text from Vinkius, which discusses their latest AI research and MCP server ecosystem. Would you like a summary of the findings?

Frequently Asked Questions

01 **How does the Tavily MCP improve search results for my AI agent?**

It uses specialized web searches that format results specifically for LLMs. This means instead of a list of links, your agent receives optimized snippets and context ready to be analyzed.

02 **Can I use Tavily MCP to summarize articles?**

Yes. You can use the `extract_content` tool to pull clean text from a link, which then allows your agent to summarize it accurately without needing the whole messy webpage.

03 **Is Tavily MCP better than just using Google Search?**

It's different. While Google gives you links, this MCP gives your agent structured data and context optimized for understanding, not just clicking.

04 **How do I get real-time information with Tavily MCP?**

Use the `search_news` tool. This function monitors current events and specializes in providing timely updates on industry shifts so your research is always up to date.

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 `"tavily": { "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

Tavily 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 Tavily. 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	Tavily MCP
Server ID	`019d8488-0208-7339-b04e-0ec7cdef04cb`
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/tavily.