

MCP SERVER

NO CODE

CLOUD HOSTED

Perplexity AI MCP

Get research-grade answers with verifiable sources.

Perplexity AI MCP connects your agent to real-time web search capabilities. Get answers that are grounded in current information, complete with source citations and links. Stop relying on generalized knowledge; use this for deep research, fact-checking, and complex analysis directly from any compatible client.

A+ Quality Score 95.83/100

web-search

research-assistant

citations

natural-language

information-retrieval



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 — connect your AI agent in under 60 seconds.

Perplexity AI MCP

14 tools available

Cloud-hosted on Vinkius

This MCP turns your agent into a dedicated, academic research assistant. You stop manually switching between search engines, reading through dozens of results, and then trying to figure out which sources are trustworthy. Instead, you ask a question—whether it's about the latest quantum computing breakthroughs or the historical market share of electric vehicles—and get an answer synthesized from multiple real-time web sources. The AI doesn't just guess; it builds comprehensive answers while linking every claim back to its original source citation. This makes it indispensable for anyone who needs accuracy, whether you're writing a technical report or building a knowledge base in your application. Connecting this MCP through Vinkius gives your agent access to best-in-class research synthesis, making guesswork obsolete.

Core Capabilities

01 — Retrieve verifiable facts with sources

The system provides answers and links every claim back to its original source URL.

03 — Focus searches on specific industries or academic domains

Restrict the search results to a list of trusted websites, like government sites or journals, eliminating irrelevant noise.

05 — Structure data into predictable formats

Force the AI to return results as JSON, making it simple for your code to ingest structured information immediately.

02 — Perform deep, exhaustive literature reviews

Run extended research queries that generate detailed reports covering complex topics and multiple angles.

04 — Maintain context across multi-step research sessions

The agent remembers previous questions and answers, allowing you to refine complex queries naturally over time.

One Click on Vinkius — From Prompt to Execution

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

- 01 Subscribe to this MCP and enter your Perplexity API key into your agent's configuration.
- 02 Instruct your AI client to perform a research query, specifying any necessary filters like domain restrictions or recency requirements.
- 03 Receive an answer that is not just text, but a synthesized report complete with source links, related questions, and structured data.

The bottom line is you get reliable, verifiable web knowledge delivered directly into your workflow without leaving the chat interface.

Built For

This MCP solves the pain point of 'source attribution.' It's for anyone who has to write reports or build applications based on external data. If you spend time verifying facts across multiple tabs, this is for you.

Market Analyst

Runs deep research queries comparing competitor strategies and market share using domain filters and structured outputs.

Academic Researcher

Conducts literature reviews for papers, ensuring every single fact is linked back to a reputable academic source.

Software Developer

Generates code analysis or extracts specific data points from complex web results using JSON schema definitions and reasoning tools.

What Changes When You Connect

- 01 **Citation Reliability:** Instead of accepting general statements, the `chat_with_citations` tool provides a link for every fact, eliminating guesswork and ensuring academic integrity.

-
- 02** Depth Control: Use `deep_research` when you need an exhaustive report on a complex topic. It takes longer but gives vastly more breadth than quick queries.
-
- 03** Domain Specificity: The `chat_with_domain_filter` tool lets you narrow searches to trusted domains, which is critical for industry reports or legal research.
-
- 04** Data Structuring: If your agent needs data for an application, use `structured_query`. It forces the output into reliable JSON format, ready for code consumption.
-
- 05** Time Sensitivity: When reporting on breaking news, `chat_with_recency_filter` ensures the information is based only on hours or days of old, not outdated knowledge.
-

Real-World Applications

Comparing global EV market incentives

A market analyst needs to compare tax subsidies across three countries. They use `chat_with_domain_filter` to restrict searches only to national government sites, ensuring the pricing data is current and official.

Extracting contact data from a webpage

A developer finds a complex article and needs only names and phone numbers. They use `structured_query` with a defined schema to programmatically pull out clean, usable JSON records, rather than messy text.

Verifying a historical claim for a paper

A student writes about quantum computing and needs proof. They run `chat_with_citations`, which guarantees that every technical detail they include has an immediate, verifiable source link attached to the answer.

Tracking breaking news on AI policy

A journalist wants the latest updates on EU AI regulations from the last month. They use `chat_with_recency_filter` and restrict results to specific regulatory bodies' websites for maximum relevance.

Patterns to Avoid

Using general chat for fact-checking

✗ AVOID

Asking a simple question using `chat_completion` when the answer involves disputed facts or hard numbers. You get confidence, but no proof.

✓ INSTEAD

Always use `chat_with_citations`. This forces the agent to link every single claim it makes back to an original source, making your information auditable.

Searching for niche technical data

✗ AVOID

Running a general query about 'topological qubits' which might pull in general physics articles that don't help with specific engineering details.

✓ INSTEAD

Use `chat_with_domain_filter` and restrict your search to known academic or industry sites, like 'arxiv.org'. This keeps the focus tight.

Needing structured data for a database

✗ AVOID

Getting a long text block from an article that you then have to manually copy and parse into columns in Excel.

✓ INSTEAD

Use `structured_query` and provide the exact JSON schema you need. The agent returns clean, machine-readable data right away.

The Right Fit

Use this MCP if your workflow requires verifiable information from the live web—anything that needs a source citation or structured output falls into this camp. If you are building an application that must act on facts (like pulling market prices, quoting regulations, or summarizing research), this is essential. Don't use it if you just need creative text generation, brainstorming ideas, or general philosophical discussion; for those tasks, your agent doesn't need web access. Similarly, don't run a basic query if you know the topic is complex; in that case, use `deep_research` instead of a standard chat prompt to guarantee maximum depth.

The headache of verifying every fact online

Today, writing a comprehensive report means bouncing between Google results, clicking through dozens of articles, and frantically copying text snippets. You lose track of which source said what, spending more time managing citations than actually writing.

The whole process is fragile: you manually copy-paste data into your agent, then spend another half hour cross-referencing those sources to build a bibliography. It's tedious, error-prone work that slows down the entire project.

Getting grounded answers with Perplexity AI MCP

The manual steps disappear entirely. You ask your agent to 'Compare X and Y,' and it handles the web searches, synthesizes the common ground, and automatically cites every single piece of data in one clean output.

What's different now is that you get a final, ready-to-use report where accuracy isn't an assumption—it's verifiable by linking directly back to the source.

Perplexity AI MCP: 14 Tools for Research

These tools give your agent granular control over search parameters, allowing you to filter results by domain, time, and structure the output like a professional database.

#	TOOL	DESCRIPTION
01	<code>chat_completion</code>	Ask a general question to get an answer with real-time web search and source citations for quick lookups.
02	<code>chat_with_citations</code>	Get answers where every fact or claim is linked directly to its original published source URL, perfect for academic work.
03	<code>chat_with_domain_filter</code>	Restrict the search results to specific domains you provide (like government sites) to ensure high-trust sources only.
04	<code>chat_with_history</code>	Ask follow-up questions by providing a JSON array of previous messages so the model understands context.
05	<code>chat_with_images</code>	Request visual results alongside text answers, which include URLs to relevant images found during the search.
06	<code>chat_with_recency_filter</code>	Filter search results by time period (hour, day, week) so you only get information on recent events or news.
07	<code>chat_with_related_questions</code>	Receive a list of suggested follow-up questions to help guide your research and discover related topics.
08	<code>deep_research</code>	Run extensive, in-depth investigations that generate detailed reports with thorough citations on complex subjects.
09	<code>follow_up</code>	Continue a research conversation by supplying the history of messages along with your next question.
10	<code>list_models</code>	Check which specific search models are available before running your query to ensure you pick the right tool for the job.
11	<code>reasoning</code>	Use this model for complex tasks like mathematical proofs, code reviews, or logical deductions requiring step-by-step analysis.
12	<code>search_query</code>	Perform the most comprehensive search possible, combining citations, images, and related follow-up questions in one result.
13	<code>structured_query</code>	Force the AI to return its answer as a JSON object that matches a specific data schema you define for programmatic use.

#	TOOL	DESCRIPTION
14	system_prompt_query	Define a custom role or set of rules (e.g., 'Act like a financial auditor...') so the model adopts a specific persona and tone.

See It in Action

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

What are the latest developments in quantum computing as of this week?



Based on recent developments this week, there are three major breakthroughs in quantum computing:

1. **IBM's 1000+ qubit processor** achieved new error correction milestones, reducing error rates by 50% compared to previous generations.
2. **Google Quantum AI** demonstrated quantum advantage in optimization problems for logistics, showing 100x speedup over classical algorithms.
3. **A team at MIT** developed a new topological qubit design that maintains coherence at higher temperatures.

 **Sources:** Nature.com, MIT News, IBM Research Blog

Would you like me to research any of these breakthroughs in more detail?

- U** Do deep research on the competitive landscape of electric vehicle manufacturers in Southeast Asia, including market share, pricing strategies, and government incentives.



I've conducted comprehensive research on the EV competitive landscape in Southeast Asia. Here's the analysis:

Market Leaders


1. **BYD** — 35% market share, aggressive pricing (\$15K-30K), strong government partnerships in Thailand and Indonesia
2. **Tesla** — 18% market share, premium positioning (\$40K+), expanding Supercharger network
3. **Neta (Hozon)** — 12% market share, budget-friendly (\$10K-20K), popular in Vietnam

Government Incentives

- **Thailand:** Up to ฿150,000 (\$4,200) subsidy per EV, 0% import duty until 2025
- **Indonesia:** 0% luxury tax for EVs, \$3,500 purchase incentive
- **Vietnam:** 0% registration fee until 2027, reduced road tax

Key Strategies

- Local manufacturing to avoid import tariffs
- Battery swapping infrastructure partnerships
- Fleet sales to ride-hailing platforms (Grab, Gojek)

 **15 sources cited** including Reuters, BloombergNEF, and regional government reports.

Would you like me to dive deeper into any specific country or manufacturer?

- U** Search for news about AI regulation in the European Union from the last month, restricted to europa.eu and reuters.com domains.



I found 5 recent articles about AI regulation in the EU from the last month, restricted to europa.eu and reuters.com:

1. **EU Commission publishes implementation guidelines for the AI Act** — europa.eu, 2 weeks ago
2. **European countries scramble to meet AI Act compliance deadlines** — reuters.com, 1 week ago
3. **EU establishes new AI Office to enforce regulations** — europa.eu, 3 weeks ago
4. **Industry leaders respond to EU AI Act risk classification system** — reuters.com, 4 days ago
5. **EU member states finalize national AI regulatory frameworks** — europa.eu, 5 days ago

 **Citations:** All 5 sources from europa.eu and reuters.com with direct links.

Would you like me to summarize any of these articles in detail?

Frequently Asked Questions

01 How does `chat_with_citations` work with Perplexity AI MCP?

The `chat_with_citations` tool ensures that every single fact or claim in the response is linked back to its original source URL, which is essential for academic integrity.

02 Can I restrict my search results using `chat_with_domain_filter`?

Yes. You provide a comma-separated list of domains (e.g., 'nih.gov, arxiv.org'), and the agent will only use sources from those trusted websites.

03 What is the difference between `chat_completion` and `search_query`?

Use `chat_completion` for a quick answer, but use `search_query` when you want the fullest possible result, which bundles citations, images, and related questions all in one go.

04 How do I make sure my research is up-to-date?

You use `chat_with_recency_filter` and specify the time window (like 'day' or 'week'). This guarantees that your answers are based on recent events, not outdated knowledge.

05 Can I ask a follow-up question using `chat_with_history`?







Yes. By supplying the conversation history in JSON format, you allow the model to maintain context, so your agent understands exactly what you mean when you ask a follow-up question.

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 <code>"perplexity-ai": { "url": "..."</code>
 Windsurf	MCP Settings → <code>mcp_settings.json</code> → 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

Perplexity 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 · 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 Perplexity 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	Perplexity AI MCP
Server ID	019d75f1-7bfa-734e-891c-8eabcb1f904e
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/perplexity-ai.