MCP SERVER | NO CODE | CLOUD HOSTED

NVIDIA Vision MCP

Go from text prompt to analyzed, structured data.

NVIDIA Vision connects powerful visual APIs to your AI client, letting you generate images from text prompts or analyze existing visuals. Use it to ask questions about photos, detect objects in complex scenes, or extract data from scanned documents and forms. It handles everything from artistic style transfers to detailed business understanding.

A+ Quality Score 100/100

computer-vision | image-generation | object-detection | visual-qa | image-captioning | generative-ai

MAINTAINED BY VINKIUS CLOUD FOR AI AGENTS · JUNE 2026
vinkius.com/mcp/nvidia-vision

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.

NVIDIA Vision MCP

9 tools available
Cloud-hosted on Vinkius

This MCP lets you treat images like structured data. Instead of manually running through different services—one for object counting, another for captioning, and a third for document reading—you just ask your agent a question about the image. You can generate brand-new concepts using Stable Diffusion models based only on text prompts, or feed it a scanned receipt and have it pull out the total amount due and the vendor name. When you subscribe through Vinkius, your AI client gets access to this entire suite of visual tools in one place. It's built for professionals who need deep understanding from visuals, whether they are creating marketing assets or analyzing financial records.

Core Capabilities

01 — Create new images from text

Generate high-quality, unique images instantly using Stable Diffusion models based on detailed written descriptions.

02 — Answer questions about visuals

Upload a photo and ask specific questions; the agent reads the image content and provides a detailed answer.

03 — Extract data from documents

Process scanned forms, receipts, or business papers to accurately identify and pull out key pieces of information.

04 — Identify objects in images

List every object visible in a picture, or locate specific items within the frame using visual grounding.

05 — Describe image contents

Get rich, detailed captions that summarize everything happening in an image without needing to ask follow-up questions.

One Click on Vinkius — From Prompt to Execution

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

01 Subscribe to this MCP and provide your API Key from NVIDIA's developer site.

02 Direct your AI client (like Cursor or Claude) to the visual task, providing either a text prompt or an image URL.

03 Your agent uses the appropriate tool—whether it's generating an asset or analyzing data—and returns the structured result directly to you.

The bottom line is that your AI client can seamlessly switch between creating visual content and deeply understanding existing images, all through one connection.

Built For

This MCP is for anyone whose job involves bridging the gap between raw media and actionable data. If you work with marketing assets, legal documents, or complex product visuals daily, this tool saves massive amounts of time by automating analysis steps that used to require multiple specialized tools.

Digital Designer

Needs to generate mockups for social media campaigns. They use the image generation and style transfer tools to rapidly iterate through dozens of visual concepts without hiring a dedicated illustrator.

Business Analyst

Receives stacks of scanned invoices or forms from different departments. The analyst uses document_qa to automatically extract revenue figures, dates, and vendor IDs into a clean spreadsheet format for immediate reporting.

Content Creator

Needs descriptive copy for a product catalog. They feed the image captioning tool photos of their goods, which instantly generates detailed descriptions they can use on e-commerce sites.

What Changes When You Connect

01 Stop guessing what an image means. Use visual_question_answering to ask your agent specific questions about any photo—like 'What brand is this watch?' or 'When did this meeting happen?' and get a definitive answer.

02 Never start from scratch again. The generate_image tool lets you build marketing concepts instantly, simply by typing out what you need in a text prompt, skipping the initial brainstorming phase entirely.

03 Process paperwork faster than ever. Instead of manually reading tables on scanned receipts, document_qa extracts figures like tax IDs and subtotals into clean data points you can use immediately.

04 Gain visual control over your assets. The style_transfer tool lets a designer take an existing photo and make it look like a Renaissance painting or a cyberpunk graphic with one command.

05 Improve searchability of visuals. Use image_captioning to get detailed, searchable descriptions for every photo you upload, making large archives instantly discoverable.

Real-World Applications

Analyzing competitor product shots

A market researcher uploads multiple photos of competing products. They use detect_objects to count the number of visible features (like ports or buttons) and then use visual_question_answering to confirm if a specific brand logo is present on each device.

Processing old legal contracts

A paralegal receives dozens of poorly scanned, handwritten agreements. They feed the batch into document_qa, which accurately reads and extracts key clauses like 'Effective Date' and 'Termination Clause', saving hours of manual transcription.

Designing a mood board for a client

A designer is stuck on a concept. They use generate_image to create several visual options—like 'a brutalist building covered in moss' or 'futuristic beach at twilight'—and then uses image_segmentation to isolate key elements from the best result.

Cataloging scientific research photos

A biologist uploads images of local flora. They use detect_objects to list all visible species and run visual_grounding to pinpoint exactly where specific plant parts (like seeds or root systems) are located in the photo.

Patterns to Avoid

Treating images like raw files

✗ AVOID

Trying to copy and paste an image of a document into a general-purpose LLM prompt, hoping it 'just knows' what the numbers mean.

✓ INSTEAD

You have to use document_qa. This tool specifically processes scanned documents and forms, ensuring the agent understands that those blurry lines are actually revenue figures or dates.

Trying to generate art without context

✗ AVOID

Asking a general AI client to 'make something pretty' using only vague instructions. The result is generic and uninspired.

✓ INSTEAD

Use generate_image with detailed prompts, specifying the model (like stabilityai/stable-diffusion-3-medium) and desired dimensions. Be specific about style, mood, and subject matter.

Confusing description with data

✗ AVOID

Using image_captioning to read a bank statement, resulting in flowery language ('a collection of financial figures') instead of the actual numbers needed.

✓ INSTEAD

For structured data extraction from forms or receipts, always use document_qa. It's built for OCR and understanding transactional fields.

The Right Fit

Use this MCP if your primary bottleneck is visual intelligence—when you need to either create complex imagery or read deep meaning from existing photos and documents. This tool excels at transforming unstructured pixels into structured data points (like dates, names, product counts) and high-fidelity assets. Don't use it if your problem is pure text generation; for that, a standard LLM connection will suffice. You should also avoid using this MCP if you

just need simple classification (e.g., 'Is this picture of a cat or a dog?'); while some tools can do that, dedicated image classifiers are often faster and more reliable. However, if your goal is complex reasoning over visual data—like asking the AI to summarize all the financial implications shown in a tax document—this MCP's combination of detect_objects, document_qa, and visual_question_answering makes it essential.

Manually processing visuals slows down every department.

Right now, if you get a stack of marketing photos or scanned contracts, the workflow is brutal. You open one tool to count objects, another service to write captions, and then maybe a third app just to extract dates from forms. It's a cycle of copy-pasting data between five different tabs, wasting hours before you even start your actual work.

With this MCP connected through Vinkius, the process collapses into one prompt. You give your agent the image or document, and it handles the analysis—whether it's listing objects using detect_objects or pulling a revenue total via document_qa—and hands you clean, usable data back to work with.

Get instant visual understanding with NVIDIA Vision.

The days of multiple specialized APIs are over. Instead of switching between object detection services and general captioning models, you're running it all through one unified connection. You get the power to segment images into specific regions while simultaneously asking natural language questions about what those segments represent.

It means your team can focus on strategy, not plumbing. The visual intelligence is simply available when you need it.

NVIDIA Vision: 9 Tools for Visual AI

These tools let you perform every visual task imaginable, from generating new artwork with text prompts to extracting structured data from scanned business forms.

#	TOOL	DESCRIPTION
01	`image_captioning`	Generates a descriptive text summary detailing the contents and context of an image.
02	`detect_objects`	Identifies and provides a list of every physical object present in an uploaded picture.
03	`document_qa`	Reads scanned documents, forms, or receipts and answers specific questions about the contained text and data.
04	`generate_image`	Creates a brand-new image file from scratch based on a written text prompt using Stable Diffusion models.
05	`visual_grounding`	Pinpoints and isolates specific objects or phrases within an image, telling you exactly where they are located.
06	`image_segmentation`	Separates an image into distinct regions, allowing you to identify and isolate every major object present.
07	`style_transfer`	Applies the artistic look or style of one picture onto another existing visual asset.
08	`list_vision_models`	Retrieves a list of all available vision models that can be used with the NVIDIA API Catalog.
09	`visual_question_answering`	Allows you to ask natural language questions about an image and receive a direct answer based on its visual content.

See It in Action

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

U Generate an image of a futuristic city at sunset.

Image generated successfully! Base64 data available for display.

U What objects do you see in this image: https://example.com/photo.jpg

I detect: 1. A red car (center). 2. A tree (left). 3. A building (background). 4. Two people walking (right).

U Describe this image in detail: https://example.com/document.png

The image shows a business document dated March 2026. It contains a table with revenue figures totaling $2.4M.

Frequently Asked Questions

01 **Can I use NVIDIA Vision to generate images for a website?**

Yes, absolutely. You use the generate_image tool by providing a text prompt (e.g., 'minimalist corporate office') and selecting your desired model parameters.

02 **Does NVIDIA Vision help with legal documents?**

It does. The document_qa tool is specifically designed to work with scanned forms, receipts, and contracts, allowing you to ask questions about the text it finds inside.

03 **What is the difference between image_captioning and visual_question_answering?**

Image captioning provides a general description of everything in an image. Visual question answering requires you to ask a specific query, like 'Who is this person?' or 'What year was this built?' for a targeted answer.

04 **Do I need a developer background to use NVIDIA Vision?**

No. You connect the MCP using your API key, but after that, you interact with it through natural conversation via your AI client, which handles all the complex coding for you.

05 **Can I isolate specific parts of an image using NVIDIA Vision?**

Yes. You can use visual_grounding to pinpoint a specific object or phrase and image_segmentation to cleanly separate that object from the rest of the picture.

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 `"nvidia-vision": { "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

NVIDIA Vision 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 NVIDIA Vision. 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	NVIDIA Vision MCP
Server ID	`019d75e1-6da6-72c6-9a76-f7027431578c`
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/nvidia-vision.