MCP SERVER | NO CODE | CLOUD HOSTED

Twelve Labs MCP

Search videos with natural language queries.

Twelve Labs (Video Understanding) provides your agent with multimodal AI capabilities to index video content and extract deep semantic insights. You can query vast video libraries using natural language, locating specific objects, actions, or speakers across hours of footage without manual tagging. This MCP handles the entire process: asset upload, indexing, embedding generation, and complex search retrieval.

A+ Quality Score 98.33/100

video-search | multimodal-ai | semantic-search | video-indexing | computer-vision

MAINTAINED BY VINKIUS CLOUD FOR AI AGENTS · JUNE 2026
vinkius.com/mcp/twelve-labs-video-understanding

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.

Twelve Labs (Video Understanding) MCP

18 tools available
Cloud-hosted on Vinkius

This MCP gives your agent the ability to 'watch' videos and understand what it sees and hears. Instead of manually reviewing massive video archives, you can now query them using plain language—asking things like, "Show me every time someone mentions Q3 revenue" or "Find all shots featuring a red car." When you connect this MCP via Vinkius, your agent gets direct access to the tools needed to index videos and run deep analyses. It handles everything from uploading assets to creating searchable indexes. Your agent can process visual and audio data simultaneously, generating structured insights about entities and moments in time. You simply tell your AI client what it needs, and this MCP does the heavy lifting of turning raw video files into actionable, machine-readable data points.

Core Capabilities

01 — Find specific moments in videos

Search through indexed video content using natural language queries to pinpoint exact time stamps for objects or actions.

02 — Upload and manage entire video libraries

Handle large-scale asset uploads, confirming multipart sessions and monitoring the indexing status of videos from URLs or local files.

03 — Generate deep data embeddings

Create multimodal vector embeddings asynchronously or synchronously to power advanced machine learning workflows on video assets.

04 — Structure entity data for analysis

Define and populate collections of entities, such as people or objects, allowing your agent to track specific subjects across multiple videos.

05 — Run targeted video analyses

Execute deep analysis tasks on video assets to extract structured data points from both visual frames and audio tracks.

One Click on Vinkius — From Prompt to Execution

Available at **vinkius.com/mcp/twelve-labs-video-understanding** — connect your AI agent in three steps.

01 First, you create a new index using the `create_index` tool. This establishes the container where your searchable video metadata will live.

02 Next, you upload your videos, either by confirming a multipart session or calling `index_asset`. The system then processes and embeds the content, making it ready for search.

03 Finally, you query the index using natural language prompts via the `search` tool. Your agent returns specific time stamps, objects, or actions found within the indexed footage.

The bottom line is that your videos transform from static files into a fully searchable, structured knowledge base accessible through simple conversation.

Built For

This MCP is for technical teams drowning in video data. It's the media manager who spends hours manually tagging footage, or the security analyst who needs to find a specific event that happened months ago—you need this to cut through the noise and get direct answers.

Media Asset Manager

Uses this MCP to organize massive video libraries by creating indexes and running deep analysis tasks, ensuring all b-roll or source footage is cataloged and searchable.

Security Analyst

Queries hours of surveillance or body camera footage using natural language to find specific objects (like a red vehicle) or actions without watching every second.

AI Developer

Integrates video understanding into complex agent workflows, utilizing tools like `embed_async` and `create_entity_collection` to build multimodal pipelines.

What Changes When You Connect

01 You stop wasting time watching hours of footage. By using the `search` tool, your agent finds specific moments—like 'a person laughing' or 'the car passing by'—and gives you exact time codes instantly.

02 Indexing huge libraries used to be a multi-step pipeline nightmare. Now, you just create an index and use `index_asset` to process videos in bulk, making them immediately available for deep searches.

03 You don't need specialized ML teams for every project. By calling tools like `embed_async`, your agent generates the necessary multimodal embeddings so you can build advanced search logic without writing complex vector code.

04 Tracking specific subjects across videos is simple. You create a collection with `create_entity_collection` and use `create_entity` to tag people or objects, letting your agent monitor them over time.

05 Handling massive files used to crash systems. This MCP manages uploads using multipart sessions (`create_multipart_upload`), ensuring stable transfer of huge video archives regardless of size.

Real-World Applications

Finding evidence in old security footage

A security analyst needs to know when a specific employee entered the restricted area. Instead of watching days of footage, they ask their agent: 'Search index idx_security for any instance of Employee X near Door Y.' The MCP uses `search` and returns precise time stamps.

Curating marketing b-roll quickly

A content creator needs quick shots of a product in use. They upload 50 videos, create an index using `create_index`, and then ask their agent to find 'close up shots of the packaging' across all assets.

Building knowledge retrieval for legal cases

A developer needs to build a system that answers questions based on video testimony. They use `embed_async` and populate an index, allowing their agent to retrieve visual context using semantic search via the `get_indexed_asset` tool.

Analyzing product failure points

An operations team needs to find where a machine fails. They run synchronous analysis (`analyze_sync`) on video feeds, which extracts structured data about component failure rates and automatically logs them into an entity collection.

Patterns to Avoid

Treating videos like PDFs

✗ AVOID

Trying to search for a keyword ('meeting') using only text transcription results fails when the conversation is interrupted by visual cues or background noise.

✓ INSTEAD

You must use this MCP's native multimodal capabilities. Use the `search` tool against an indexed asset; it finds 'meeting' even if it's implied visually, not just spoken.

Over-relying on basic indexing

✗ AVOID

Just uploading a video and expecting search results for specific people or objects that aren't labeled manually.

✓ INSTEAD

Always use the dedicated tools. First, establish your data with `create_index`, then populate it by running deep analysis tasks or using `embed_async` to generate proper embeddings.

Assuming single-shot processing

✗ AVOID

Trying to process a 10GB video file in one go, which often leads to timeouts and failed transfers.

✓ INSTEAD

Use the robust upload mechanisms. Start with `create_multipart_upload` to manage large files reliably, then call `report_multipart_progress` to confirm successful transfer.

The Right Fit

Use this MCP if your primary challenge is finding information *inside* video content—if the data you need exists visually or audibly within a stream of footage. You should use it when simple text search, like searching PDFs or databases, won't cut it because context matters. Don't use this if all you need to do is store metadata about videos; for that, basic cloud storage tools suffice. If you only need to

transcribe speech, other dedicated transcription services might work better. However, if you need the AI to *understand* what the people in the video are doing, or who they are interacting with, this MCP gives you the full multimodal tooling set via `create_entity` and `search` that no simple file storage system can match.

Finding a single moment in hours of footage is painful work.

Today, if you're reviewing an incident or curating content for marketing, the process is brutal. You open your video management system, and instead of getting a direct answer, you are faced with endless thumbnails and time sliders. You have to click play, scrub through minutes of irrelevant footage, pause it, take notes, and copy down timestamps manually. It's an exercise in exhaustion.

With this MCP, your agent handles the slog. Instead of manual clicking, you just ask a natural language question like, 'Find every instance of the blue widget being handled.' The system processes that request against your indexed videos and spits out exactly what you need—the time codes, the relevant clips, and sometimes even structured data about the object itself.

Twelve Labs (Video Understanding) MCP provides search access to visual context.

You eliminate the manual steps of tagging. You don't have to wait for a human editor or ML engineer to go through footage and label every object or action point; you simply establish an index using `create_index`, and your agent does the deep work automatically.

The difference is radical: instead of viewing video archives as passive recordings, they become active, queryable knowledge bases. Your AI client can now interact with them like a highly specialized research assistant.

Twelve Labs (Video Understanding): 18 Tools

These tools give your agent granular control over every stage of multimodal AI workflow, from uploading assets to generating detailed search indexes.

#	TOOL	DESCRIPTION
01	analyze_async	Starts a background job to break down and analyze video content into segments.
02	create_multipart_upload	Starts a segmented upload session, allowing you to reliably transfer very large video files in parts.
03	get_index	Retrieves the full details of a specific index using its unique ID number.
04	analyze_sync	Analyzes and breaks down video content instantly, returning results immediately.
05	confirm_multipart_upload	Verifies the details of a large, segmented file upload session before starting the transfer.
06	create_asset	Uploads raw video content to begin the process of creating an indexed digital asset.
07	update_index	Changes the descriptive name of an existing video index without affecting its underlying data.
08	create_entity_collection	Sets up a group or collection designed to hold and categorize specific types of entities, like people or places.
09	create_entity	Adds a single entity, such as a person's name, into an existing defined collection.
10	create_index	Initializes and names a new index that will store all the video metadata for later searching.
11	delete_index	Removes an entire index from the system when the associated project or data is no longer needed.
12	embed_async	Generates vector embeddings for video content in the background, preparing it for semantic search.
13	embed_sync	Creates vector embeddings for video content instantly, useful for small-scale testing and immediate use cases.

#	TOOL	DESCRIPTION
14	get_indexed_asset	Fetches all the structured data and metadata associated with an already uploaded and indexed video file.
15	index_asset	Sends a specific asset to be processed and added to a pre-existing index.
16	list_indexes	Retrieves a list of all available indexes, showing their names and IDs.
17	report_multipart_progress	Checks and reports the current progress status of a large, ongoing multipart upload session.
18	search	Runs a natural language query against an index to find specific moments or time segments within videos.

See It in Action

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

U List all my Twelve Labs video indexes.

I've retrieved your indexes. You have 3 active indexes: 'Marketing-2024' (ID: idx_abc123), 'Product-Demos' (ID: idx_def456), and 'Security-Feed' (ID: idx_ghi789).

U Create a new index named 'Webinar-Archive' using the Marengo 3.0 model with visual and audio options.

Index 'Webinar-Archive' has been created successfully with ID: idx_web987. It is configured with the Marengo 3.0 model for both visual and audio analysis.

U Search for 'a person presenting a slideshow' in index idx_abc123.

Searching... I found 2 relevant matches in index idx_abc123. The first is at 02:15 in 'Q3-Review.mp4' and the second is at 10:45 in 'Strategy-Session.mp4'.

Frequently Asked Questions

01 **How do I start using the Twelve Labs (Video Understanding) MCP?**

You subscribe to this MCP on Vinkius and provide your API key. Then, you use the `create_index` tool first to establish a searchable container for your data.

02 **Can I search my videos using natural language with Twelve Labs (Video Understanding) MCP?**

Yes, that's its main purpose. The `search` tool lets you type what you are looking for—like 'a person arguing about contracts'—and it returns time codes across your indexed content.

03 **Is Twelve Labs (Video Understanding) MCP better than just uploading videos to Google Drive?**

Absolutely. Standard storage services only hold the file; this MCP actually analyzes the contents, creating structured indexes and allowing semantic search based on objects or actions.

04 **What is the difference between `analyze_sync` and `analyze_async` with Twelve Labs (Video Understanding) MCP?**

`analyze_sync` gives you immediate results for small tasks, but `analyze_async` handles large videos or complex jobs in the background without timing out your agent session.

05 **How do I upload a massive video file to Twelve Labs (Video Understanding) MCP?**

You use the multipart tools. First, call `create_multipart_upload` to start the session, then send chunks of data and monitor progress with `report_multipart_progress`.

06 **How do I list all my existing video indexes?**

You can use the `list_indexes` tool. It will return a list of all indexes available in your Twelve Labs account, including their IDs and configuration.

07 **Can I search for a specific moment inside my videos using text?**

Yes! Use the `search` tool by providing an `index_id` and a search query. The AI will find the most relevant timestamps and video segments based on your description.

08 **How do I add a new video to an index for analysis?**

First, use `create_asset` with a public URL to upload the video. Then, use the `index_asset` tool with the resulting `asset_id` and your target `index_id` to start the processing.

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 `"twelve-labs-video-understanding": { "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

Twelve Labs (Video Understanding) 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 Twelve Labs (Video Understanding). 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	Twelve Labs (Video Understanding) MCP
Server ID	`019e38ff-bc3e-73d5-bb9b-0eb183ea0793`
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/twelve-labs-video-understanding.