



# Financial Chat Agent

Hosted chat API for natural-language access to viaNexus data

INTEGRATION GUIDE · MAY 08, 2026

## CONTENTS

---

[01 Overview](#)

[02 Authentication & Endpoints](#)

[03 Quickstart](#)

[04 Session Management](#)

[05 Error Handling](#)

[06 Sample Queries](#)

[07 Agent Capabilities](#)

[08 What's running today](#)

[09 Support](#)

**Just want to make your first call?** Skip to [§03 Quickstart](#). The two sections before it ([Overview](#), [Authentication & Endpoints](#)) give you the context you'll want eventually but aren't blockers.

**Note on AskNexus:** AskNexus (at `vianexus.com/asknexus`) is our reference chat UI. It's powered by the Financial Chat Agent (FCA) underneath. This guide shows you how to call the FCA directly so you can build the same kind of experience in your own app.

## 01 | Overview

---

A hosted chat API. You POST a question with your viaNexus token and receive a natural-language answer grounded in viaNexus data. The service runs on Google Cloud Run and is publicly reachable; no GCP IAM is required on your side.

```
your app → Financial Chat Agent (FastAPI, Cloud Run) → viaNexus MCP server → via
```

**How it works.** Each request runs an LLM (Anthropic Claude `sonnet-4.5` by default) inside a tool-calling loop. The model's tools are wired into the viaNexus MCP server, which exposes the viaNexus dataset catalog: quotes, fundamentals, news, corporate actions, sector performance, and more. The LLM typically calls `search` to discover relevant datasets, then `fetch` to pull rows, can chain additional calls if a question requires multiple data sources, and finally synthesizes a reply. The system prompt restricts answers to financial topics and prohibits fabricated data — if a dataset isn't available the model says so plainly.

### What you don't manage.

The service handles the OAuth handshake against MCP, the per-session lifecycle (1-hour TTL with automatic cleanup), and conversation memory so that multi-turn context (the agent remembering "AAPL" across follow-up questions) is preserved server-side. You make HTTPS requests; everything else is opaque to you.

**What runs on the LLM side.** Anthropic Claude is the default. The agent SDK also supports OpenAI and Google models — see [§08 What's running today](#) for the current production model and [§07 Agent Capabilities](#) for the boundaries the system prompt enforces.

## 02 | Authentication & Endpoints

### Base URL

```
https://financial-chat-agent-857389207619.us-central1.run.app
```

### Authentication

You authenticate with a single **viaNexus API token** passed as the `token` query parameter. It is the same token you use against `https://api.blueskyapi.com`. Personal-key tokens are prefixed `pk_...`.

On the first call for a new session, the service exchanges your token for a **software statement** (a signed credential issued by `https://api.blueskyapi.com/v1/agents/register`) and uses it to open an OAuth-backed session against the viaNexus MCP server. You never see or manage the software statement directly. Subsequent calls on the same `session_id` reuse the open MCP connection.

### HTTP surface

Method	Path	Purpose
POST	<code>/chat</code>	Send a message; create or continue a session
GET	<code>/health</code>	Service health and active session count
DELETE	<code>/sessions/{session_id}</code>	Close a session and free its MCP connection

## 03 | Quickstart

Make your first call in under a minute. You'll need a viaNexus token (it looks like `pk_...`).

```
TOKEN="your-vianexus-token"
BASE="https://financial-chat-agent-857389207619.us-central1.run.app"

curl -X POST "$BASE/chat?token=$TOKEN" \
  -H "Content-Type: application/json" \
  -d '{"message": "What is AAPL trading at right now?"}'
```

Response:

```
{
  "response": "Apple (AAPL) is currently trading at $292.59 (+1.79%)...",
  "session_id": "5f942fb6fadf4791bc7dece76d4a7332",
  "artifacts": []
}
```

The service returns a `session_id` automatically. Pass it back on the next call to continue the conversation (see [Session Management](#) below). If you don't, every call starts fresh.

### Request schema

Field	Type	Required	Description
<code>message</code>	string	yes	The user prompt.
<code>client_context</code>	object	no	Hint for tool filtering. e.g. <code>{"type": "openbb"}</code> exposes OpenBB-specific tools.
<code>widget_context</code>	string	no	Pre-formatted dashboard or widget state to be prepended to the prompt for context.

### Response schema

Field	Type	Description
<code>response</code>	string	The assistant's natural-language reply (may contain markdown tables and formatting).
<code>session_id</code>	string	32-character hex MCP session identifier. Persist this client-side to enable multi-turn conversations.
<code>artifacts</code>	array	Structured artifacts emitted during the turn. Empty array when the agent answered with text only.

## Performance

The first call on a new session typically takes **20–35 seconds** (token exchange, MCP handshake, dataset priming, first LLM round-trip). Subsequent calls on the same session are around **5 seconds**.

---

## 04 | Session Management

---

Pass the `session_id` back as a query param to continue an existing conversation:

```
curl -X POST "$BASE/chat?token=$TOKEN&session_id=5f942fb6fadf4791bc7dece76d4a7332" \  
-H "Content-Type: application/json" \  
-d '{"message": "How does that compare to its 52-week high?"}'
```

The agent remembers AAPL from the previous turn and answers in context (e.g. "AAPL is at \$292.59, just above its 52-week high of \$288.62...").

Sessions expire after 1 hour of inactivity. To close one early, call `DELETE /sessions/{session_id}`.

---

## 05 | Error Handling

---

Code	Meaning	Handling
200	Success	Use the response.
422	Missing required parameter (e.g. no <code>token</code> )	Validate request before sending.
404	<code>session_id</code> not found	Drop it; omit <code>session_id</code> to start a new session.
410	Session expired (1-hour TTL)	Start a new session.
500	Upstream error (token exchange, MCP, or LLM)	Inspect the <code>detail</code> field; retry after backoff.

All non-200 responses include a JSON body with a `detail` field describing the error.

---

## 06 | Sample Queries

---

All of these were tested live against the endpoint. Drop them into your demo as-is or use them as templates.

### Quote / price

What is AAPL trading at right now?

Returns the current price, change %, day range, volume, market cap, P/E.

### News headlines

What are the latest news headlines for NVDA?

Returns a numbered list of headlines with sources (Benzinga, CNBC, etc.) and one-line summaries.

### Sector performance

How is the technology sector performing today?

Returns sector % change plus top gainers/losers with their moves.

### Multi-ticker comparison

Compare the P/E ratios of AAPL, MSFT, and GOOGL.

Returns a **markdown table** rendered inside the `response` field — comparison data laid out cleanly with key insights.

**Follow-up in the same session** (pass `session_id` back)

How does that compare to its 52-week high?

The agent remembers the previous turn's ticker and answers in context.

### **Symbol resolution from common names**

What is Coca Cola trading at?

The agent resolves "Coca Cola" → KO from its general knowledge before fetching data.

---

## 07 | Agent Capabilities

---

- **Financial questions only.** Off-topic prompts get a polite redirect.
  - **No fabricated data.** If a dataset isn't available, the reply will say so plainly rather than guess.
  - **Tickers from common names work.** "Apple" → AAPL, "Coca Cola" → KO.
  - **Tables are rendered in markdown inside `response`.** Comparison queries come back nicely formatted.
-

## 08 | What's running today

---

<b>Model</b>	Anthropic Claude ( <code>claude-sonnet-4-5-20250929</code> )
<b>Response length cap</b>	1,000 tokens per reply — long analyses may be truncated
<b>Session TTL</b>	1 hour of inactivity
<b>MCP backend</b>	viaNexus MCP server ( <code>vast-mcp.blueskyapi.com</code> )

The underlying agent supports OpenAI and Google models in addition to Anthropic. If you need a different LLM, longer responses, a custom system prompt, or a dedicated MCP scope, contact us for a custom deployment.

---

## 09 | Support

---

<b>General inquiries</b>	<a href="mailto:contact@vianexus.com">contact@vianexus.com</a>
<b>Technical support</b>	<a href="mailto:support@vianexus.com">support@vianexus.com</a>
<b>Web</b>	<a href="http://vianexus.com">vianexus.com</a>
<b>Reference UI (AskNexus)</b>	<a href="http://vianexus.com/asknexus">vianexus.com/asknexus</a>