

Agents of Change

Elevating Enterprise AI with Agentic Flows

Andrew Dunsmoor
Gaurav Khanna

March 18, 2026



Agenda

1. The Augmented LLM
2. Agents vs. Agentic Flows
3. Agentic Design Patterns
4. Agentic Frameworks
5. Scaling Agentic AI
6. Agentic AI Ops, AI Canvas

**Agents will collaborate to
solve problems like a team
of subject matter experts**

AI Agents: The Origin Story

(Brief) History of the Post-GPT World

ChatGPT Launch
November 30th, 2022

“Oh, cool!”



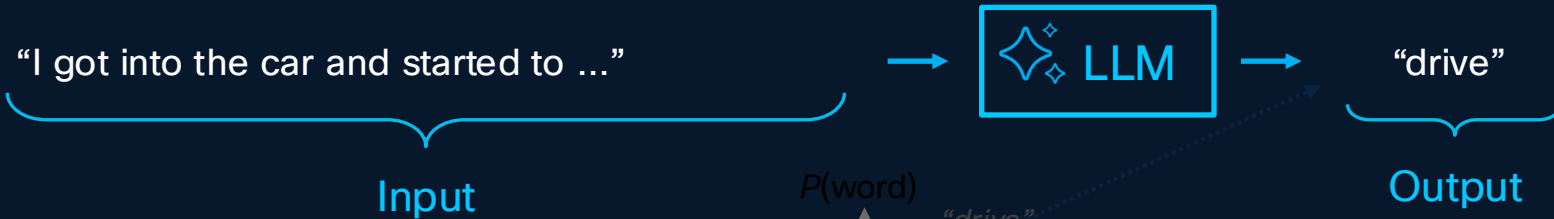
Hallucinations

“Oh, no!”

LLMs: Word-by-Word Prediction

Language models perform word-by-word prediction

Transformers / Attention Mechanism →
Use prompt and previously-generated text to predict the next word



Word-by-word predictions based on patterns they have **learned** from training on enormous corpora of text



Agentic Flows Origin Story: Prompting Techniques

Early 2023: Best practices for prompting and code generation

A **PROMPT** is a set of instructions provided to an LLM for it to execute.



PROMPT ENGINEERING is the art and science of designing prompts to give optimal results.

Prompt Engineering

A well-engineered prompt can contain one or more of the following elements:

- Persona
- Task / objective / instruction
- Tone / style
- Target audience
- Context
- Output format

Prompts: powerful “knobs” to get the most out of LLMs
Engineering prompts → Effective results

Agentic Origin Story: Prompting Techniques

Prompt Chaining

Scenario: Generate a professional email proposal based on a product description.

Prompt 1:

Given the following product descriptions, create a summary of the key features.

Prompt 2:

Draft an email structure with the following elements...

Prompt 3:

Generate the email...

Iterative prompting is effective, especially in lengthy interactions

Agentic Origin Story: Prompting Techniques

Chain of Thought

Simple Prompt

Prompt:

A juggler has 8 items. Half of the items are cubes. Half of the cubes are blue. How many blue cubes does the juggler have?

Response:

2

Mimics step-by-step reasoning that model would have seen in training data

Prompt:

Think about a problem in a step-by-step manner.

I start with x items.

I am told what percent (pA) of x items have attribute A.

Multiply x by pA to get a new number, y

Then I am told what percent (pB) of y items have attribute B.

Multiply y by pB to get a new number, z

Keep going until you arrive at final answer after you are done evaluating the attributes in succession.

Show your reasoning.

Problem:

A juggler has 8 items. Half of the items are cubes. Half of the cubes are blue. How many blue cubes does the juggler have?

Show your reasoning.

Response:

If half of the items are cubes, then the juggler has 4 cubes in total. Since half of the cubes are blue, the juggler has 2 blue cubes.

Agentic Origin Story: Coding

Natural Language



Code

Python

Prompt:

Write a Python function that returns the first n numbers in the Fibonacci sequence.

```
def generate_fibonacci(n):  
    sequence = []  
    a, b = 0, 1  
    for _ in range(n):  
        sequence.append(a)  
        a, b = b, a + b  
    return sequence  
  
# Example  
print(generate_fibonacci(10))
```

SQL

Prompt:

Write a SQL query to find all rows in a table called customers where the last_name field is missing.

```
SELECT *  
FROM customers  
WHERE last_name IS NULL;
```

Custom

Prompt:

Organize all the data in this file using a format that is a modification of JSON:

```
{  
  "key_1", "key_2":  
<column_{i,1}>,  
  "key_3": <column_{i,3}>,  
  "key_4": {  
    <column_{i,4}>  
  }  
}
```

Where the "key_x" refers to the column header in the data file and the {i,x} format is the cell value.

Agentic Origin Story: Tool Usage

Natural Language



APIs, Function Calls

API Call

Prompt:
Get the weather forecast for San Francisco using OpenWeatherMap.

```
import requests

api_key = "YOUR_API_KEY"
city = "San Francisco"
url = f"http://api.openweathermap.org/data/2.5/weather?q={city}&appid={api_key}"

response = requests.get(url)
data = response.json()

print(data["weather"][0]["description"])
```

Function Call

Prompt:
Schedule a meeting with John for Tuesday at 3pm.

```
{
  "function_call": {
    "name": "scheduleMeeting",
    "arguments": {
      "time": "2023-05-09T15:00:00",
      "invitees": ["John"]
    }
  }
}
```

Web Search

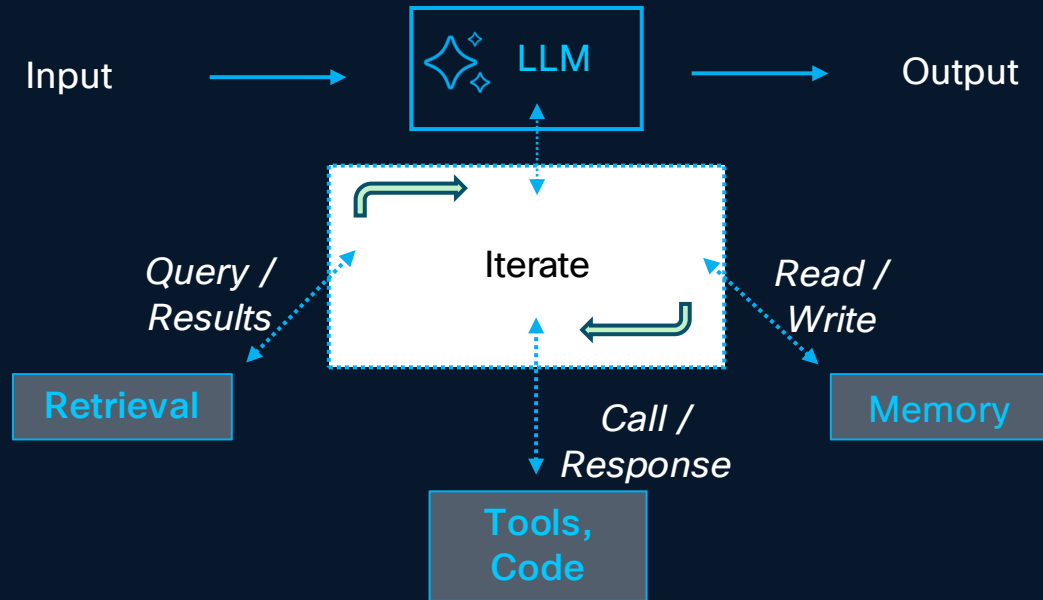
Prompt:
Search the web for recent news about CSCO earnings report.

```
from some_search_tool import search_web

results = search_web("Cisco CSCO Q2 2023 earnings report")
print(results[0]['title'], results[0]['snippet'])
```

Agentic Building Block: The Augmented LLM

Augmented LLM:
Chained Prompts + Coding + Tool Use



Agentic Flows Analogy: Essay Writing

Non-Agentive Workflow (Zero-Shot)

Prompt (Human Analogy):

Please type out an essay on topic X from start to finish in one go without using backspace



Usually not our best work!

Agentic Workflow (Iterative, Tool Use, etc.)

Prompt (Human Analogy):

- Create an **outline** on the topic
- What **information** needed? Where to get it?
- Do the **research**
- Write a **first draft**
- Consider **revisions** or further research.
- ...

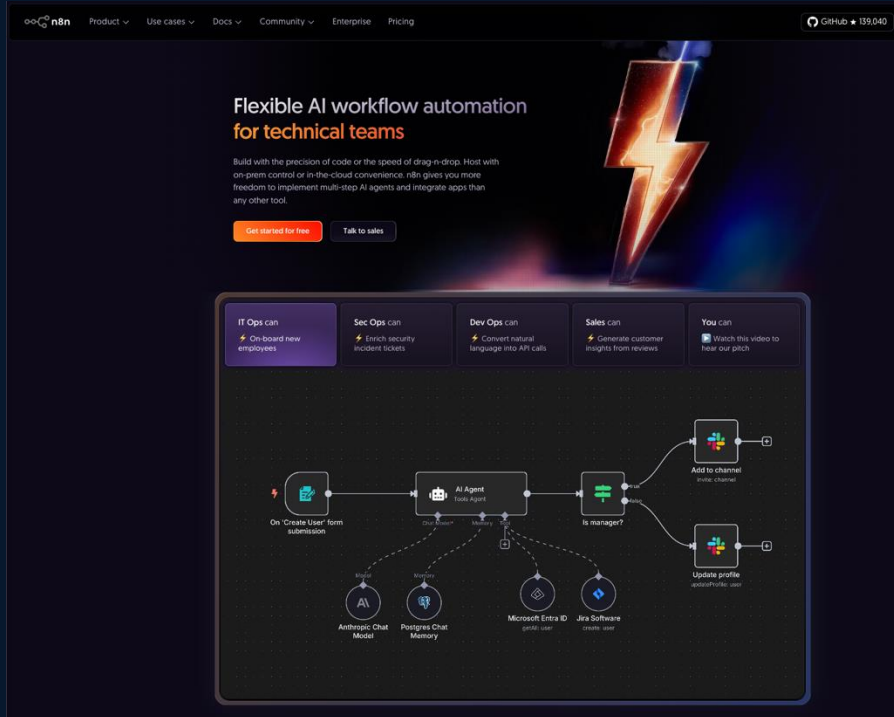


Time consuming, (sometimes messy), iterative ...
but usually better results!

Inspired by:
Andrew Ng: <https://www.youtube.com/watch?v=KrRD7i7y7NY>

Demo 01

Fetching Information



Flexible AI workflow automation for technical teams

Build with the precision of code or the speed of drag-n-drop. Host with on-prem control or in-the-cloud convenience. n8n gives you more freedom to implement multi-step AI agents and integrate apps than any other tool.

[Get started for free](#) [Talk to sales](#)

Use Cases:

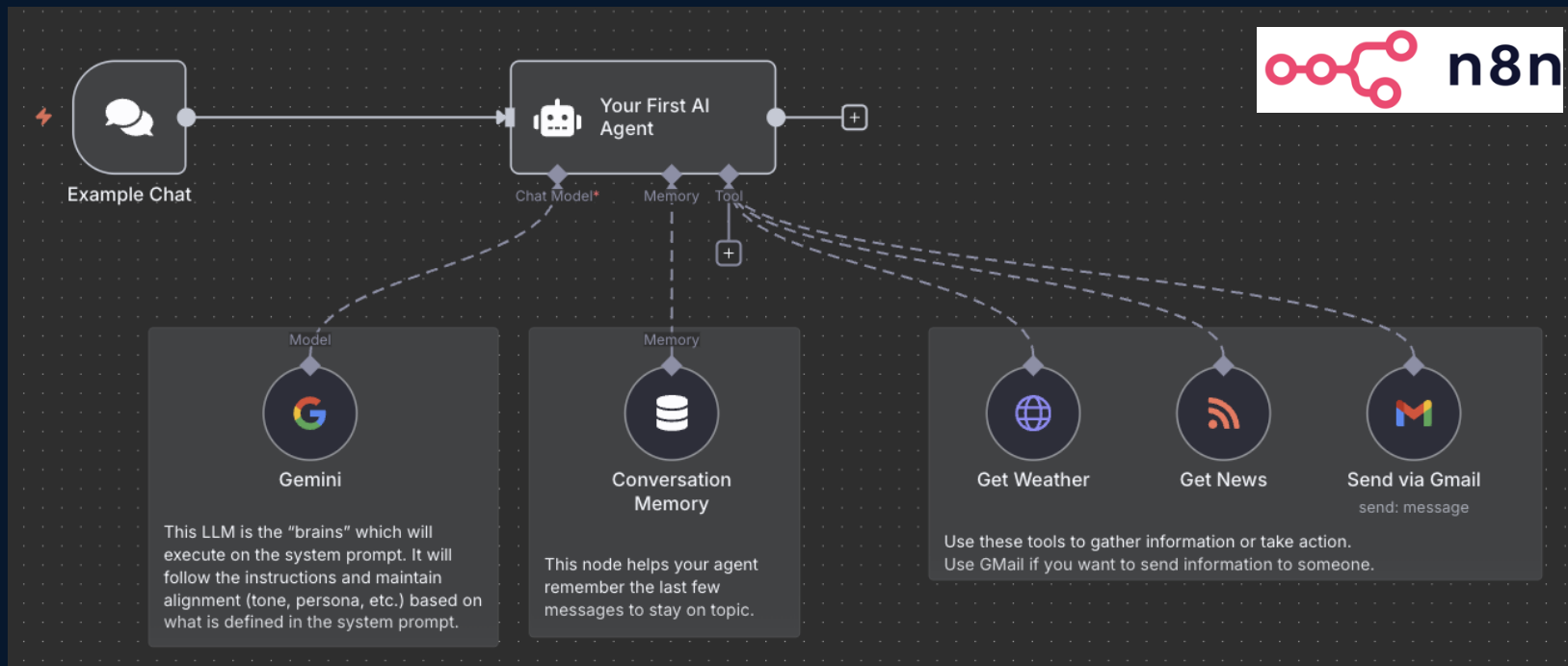
- IT Ops can:** On-board new employees
- Sec Ops can:** Enrich security incident tickets
- Dev Ops can:** Convert natural language into API calls
- Sales can:** Generate customer insights from reviews
- You can:** Watch this video to hear our pitch

Workflow Diagram:

The diagram shows a central **AI Agent (code agent)** node. It is connected to several input nodes: **On 'Create User' form submission**, **Anthropic Chat Model**, **Postgres Chat Memory**, **Microsoft Entra ID (generate user)**, and **Jira Software (create user)**. The AI Agent node is connected to an **is manager?** node, which then branches into **Add to channel (create channel)** and **Update profile (updateProfileUser)** nodes.

- **Visual, Node-Based Workflow Builder:**
 - Intuitive drag-and-drop interface, debugging tools, execution logs.
- **Hybrid Low-Code / Pro-Code Flexibility:**
 - Supports “no-code” operation for speed
 - Custom logic (JavaScript, Python), enabling granular control that low-code cannot address.
- **Extensibility and Integration:**
 - 400+ native integrations; “HTTP Request” nodes allow connecting to virtually any API.
 - Reusable workflow templates.
- **Strong platform growth and adoption:**
 - 250K active users; 3000 enterprise customers.

Demo 01: Fetching the News and Weather



Demo 01: Fetching the News and Weather

[Video]

The screenshot displays the n8n workflow editor for a project named "Simple_Weather_News_Age...". The main workflow consists of the following components:

- Example Chat**: A starting node that triggers the workflow.
- Your First AI Agent**: The central agent node, which is currently inactive. It has three sub-nodes: **Chat Model***, **Memory**, and **Tool**.
- Gemini**: A model node connected to the Chat Model sub-node. A tooltip explains: "This LLM is the 'brains' which will execute on the system prompt. It will follow the instructions and maintain alignment (tone, persona, etc.) based on what is defined in the system prompt."
- Conversation Memory**: A memory node connected to the Memory sub-node. A tooltip explains: "This node helps your agent remember the last few messages to stay on topic."
- Tools**: Three tool nodes connected to the Tool sub-node:
 - Get Weather**: A tool with a globe icon.
 - Get News**: A tool with a RSS icon.
 - Send via Gmail**: A tool with a Gmail icon and the label "send: message". A tooltip explains: "Use these tools to gather information or take action. Use Gmail if you want to send information to someone."

At the bottom of the interface, there is a **Chat** panel with a session ID "cc5fc..." and a **Logs** panel. The chat panel contains a message input field with the placeholder "Type a message" and a "Send a message below to trigger the chat workflow" instruction. The logs panel is currently empty with the text "Nothing to display yet. Execute the workflow to see execution logs."

Demo 01: Fetching the News and Weather

[Vi

The screenshot displays the n8n workflow editor interface. At the top, the workflow is titled "Simple_Weather_News_Age...". The main workspace shows a workflow starting with an "Example Chat" node connected to a "Your First AI Agent" node. The AI Agent node is configured with "Chat Model", "Memory", and "Tools". Below the AI Agent node, three tool nodes are visible: "Gemini", "Conversation Memory", and "Send via Gmail". The "Gemini" node is described as the "brains" of the agent. The "Conversation Memory" node helps the agent remember recent messages. The "Send via Gmail" node is used to send information to someone. To the right of the tool nodes, there are three more tool nodes: "Get Weather", "Get News", and "Send via Gmail". The "Get Weather" and "Get News" nodes are used to gather information, and the "Send via Gmail" node is used to send information to someone. Below the workflow editor, there is a "Chat" panel with a session ID "0843b..." and a "Logs" panel. The chat panel shows a message input field with the text "What is the latest news in the technology space?". The logs panel is currently empty.

Demo 01: Fetching the News and Weather

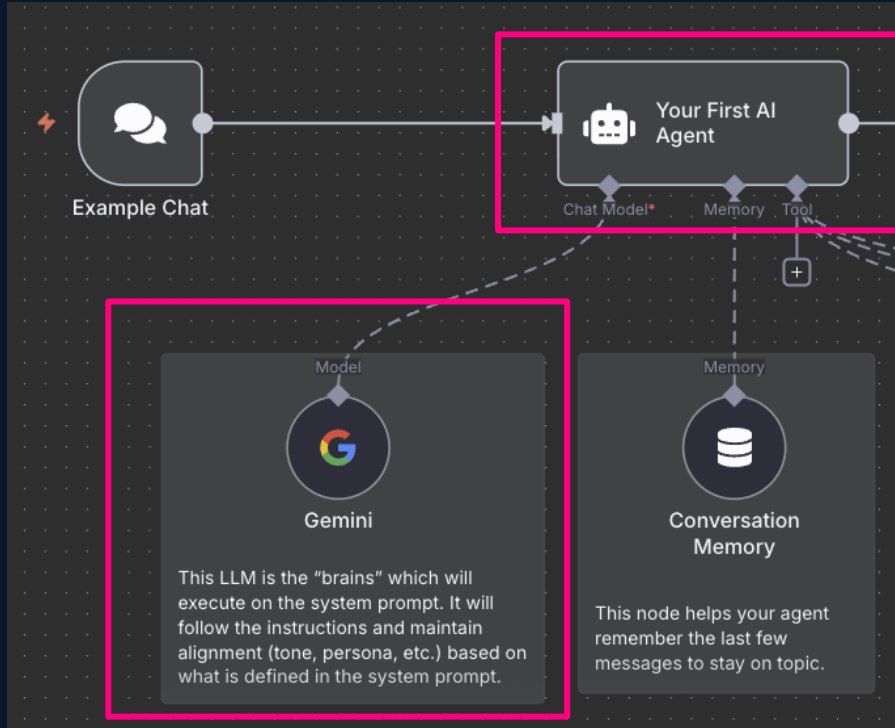
The screenshot displays the n8n workflow editor for a project named "Simple_Weather_News_Age...". The workflow is titled "Your First AI Agent" and consists of the following components:

- Example Chat**: The starting point of the workflow.
- Chat Model**: Connects the chat to the AI agent.
- Memory**: A node for storing and retrieving information.
- Tools**: A group of nodes that the AI agent can use:
 - Gemini**: A Large Language Model (LLM) node. Description: "This LLM is the 'brains' which will execute on the system prompt. It will follow the instructions and maintain alignment (tone, persona, etc.) based on what is defined in the system prompt."
 - Conversation Memory**: A node for remembering recent messages. Description: "This node helps your agent remember the last few messages to stay on topic."
 - Get Weather**: A tool to fetch weather information.
 - Get News**: A tool to fetch news information.
 - Send via Gmail**: A tool to send messages via email. Description: "Use these tools to gather information or take action. Use Gmail if you want to send information to someone."

The interface includes a sidebar with navigation options (Overview, Personal, Projects, Admin Panel, Templates, Variables, Insights, Help, What's New) and a user profile for "Gaurav Khanna". The bottom section shows a chat window with the message: "Can you please get the weather forecast fro San Jose, CA and send it to gkhanna1@gmail.com ?" and a logs window that is currently empty.

N8N: Breaking Down The Components

System Prompt and LLM



Role

You are the n8n Demo AI Agent, a friendly and helpful assistant...

How I Work

1. ****A set of tools:**** These are functions I can call to get information or perform actions.
2. ****Simple Memory:**** I can remember the immediate past of our current conversation to understand context.

My Tools Instructions

You must choose one of your available tools if the user's request matches its capability...

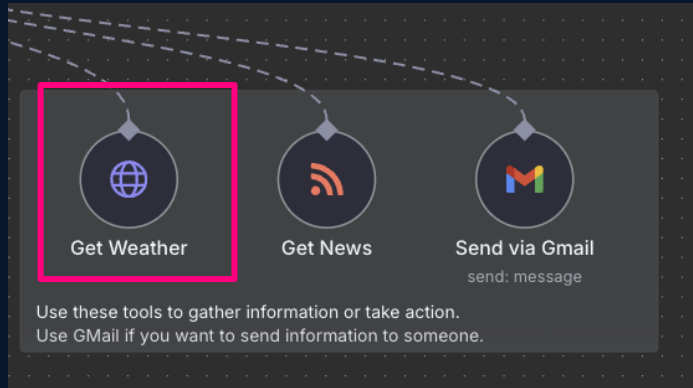
Output Tone and Style

- Respond in a friendly, conversational, and helpful tone...

Design practice: system prompt is logically separate from the LLM

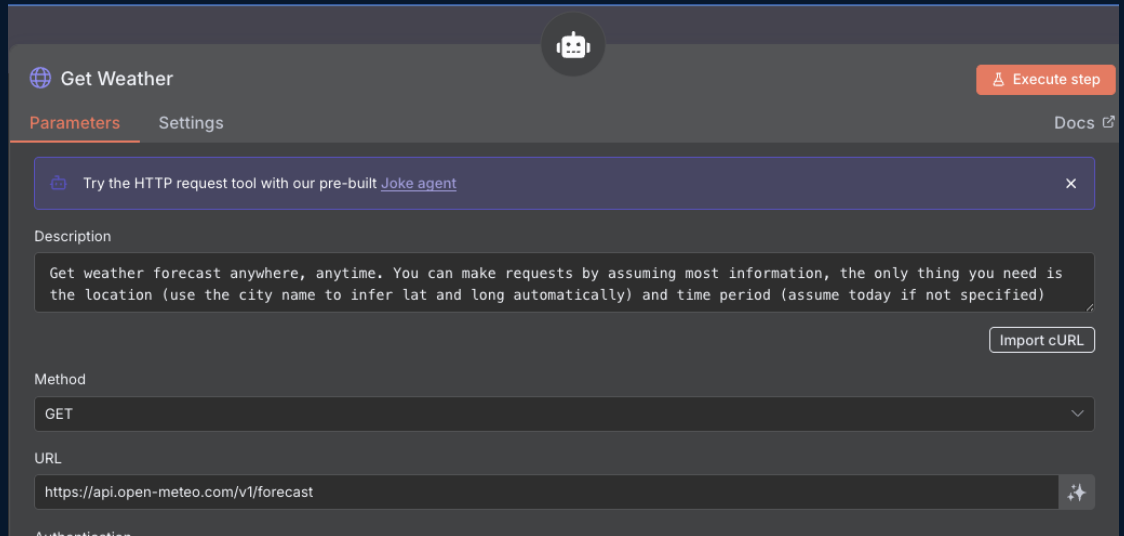
N8N: Breaking Down The Components

Tools



Use these tools to gather information or take action.
Use GMail if you want to send information to someone.

- Get Weather
- Get News
- Send via Gmail
send: message



Get Weather

Execute step

Parameters Settings Docs

Try the HTTP request tool with our pre-built [Joke agent](#)

Description

Get weather forecast anywhere, anytime. You can make requests by assuming most information, the only thing you need is the location (use the city name to infer lat and long automatically) and time period (assume today if not specified)

Import cURL

Method

GET

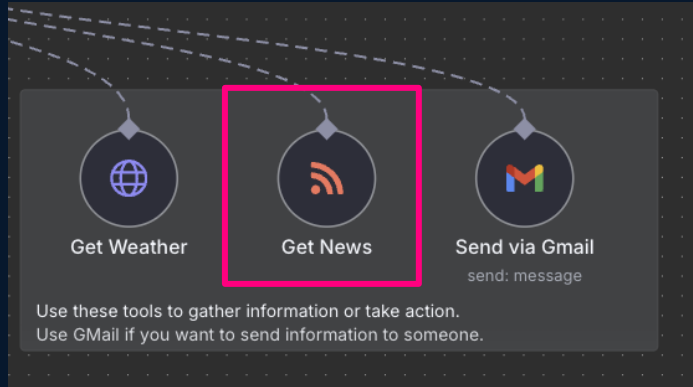
URL

https://api.open-meteo.com/v1/forecast

Authentication

N8N: Breaking Down The Components

Tools

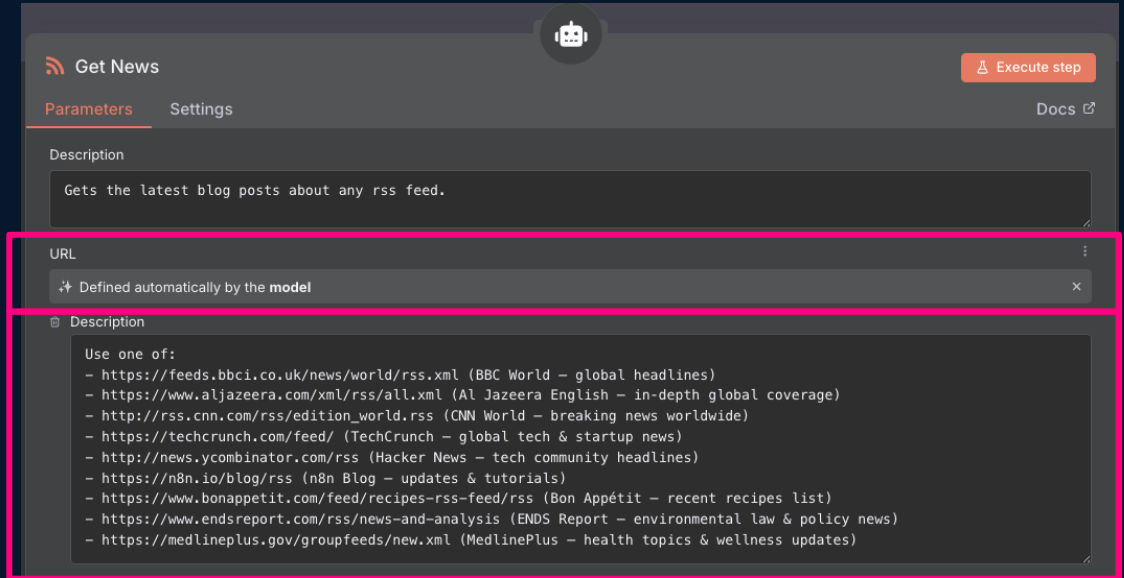


Use these tools to gather information or take action.
Use GMail if you want to send information to someone.

Get Weather

Get News

Send via Gmail
send: message



Get News

Execute step

Parameters Settings Docs

Description

Gets the latest blog posts about any rss feed.

URL

Defined automatically by the model

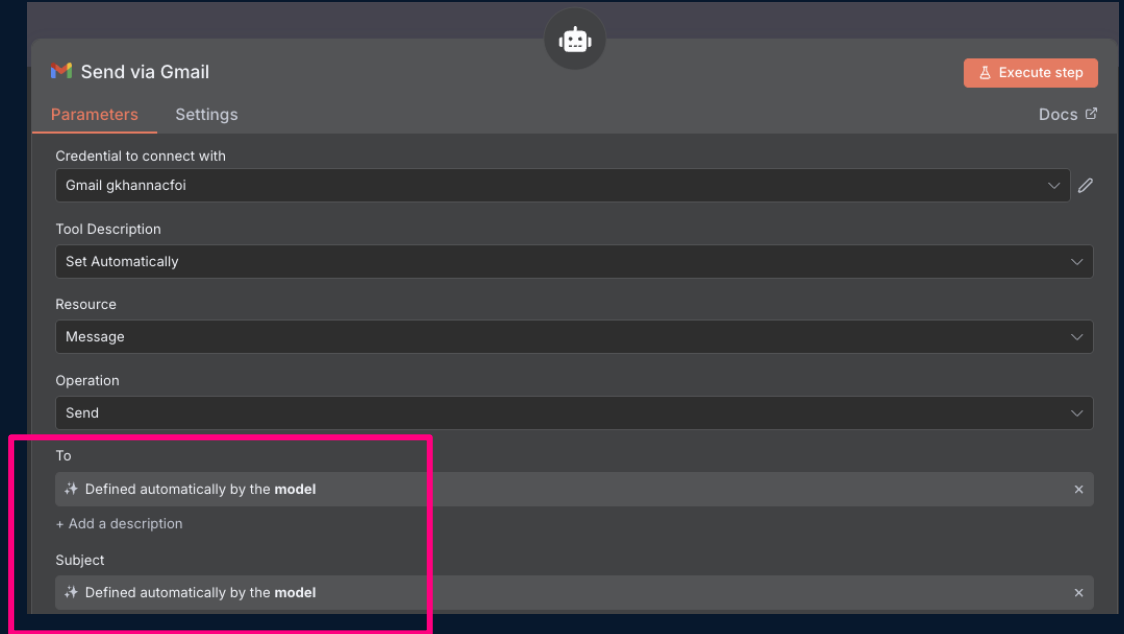
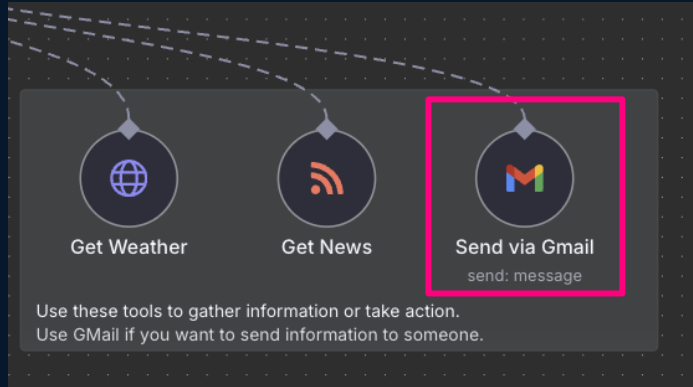
Description

Use one of:

- <https://feeds.bbc.co.uk/news/world/rss.xml> (BBC World – global headlines)
- <https://www.aljazeera.com/xml/rss/all.xml> (Al Jazeera English – in-depth global coverage)
- http://rss.cnn.com/rss/edition_world.rss (CNN World – breaking news worldwide)
- <https://techcrunch.com/feed/> (TechCrunch – global tech & startup news)
- <http://news.ycombinator.com/rss> (Hacker News – tech community headlines)
- <https://n8n.io/blog/rss> (n8n Blog – updates & tutorials)
- <https://www.bonappetit.com/feed/recipes-rss-feed/rss> (Bon Appétit – recent recipes list)
- <https://www.endsreport.com/rss/news-and-analysis> (ENDS Report – environmental law & policy news)
- <https://medlineplus.gov/groupfeeds/new.xml> (MedlinePlus – health topics & wellness updates)

N8N: Breaking Down The Components

Tools



N8N: Breaking Down The Components

Web-based Chatbot Interface

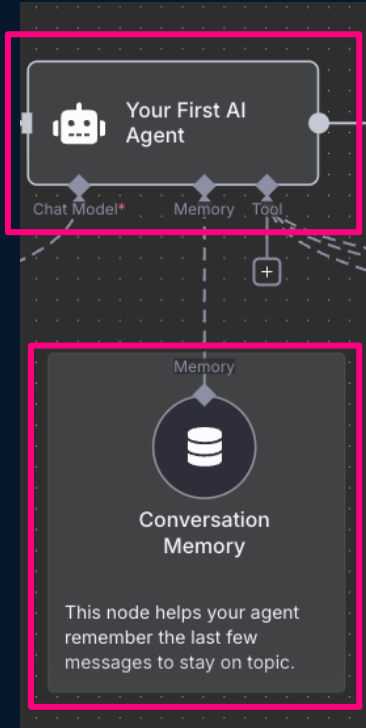
- **Stateful:** The platform maintains conversation context
- **Persistent context window:** Previous messages remain available automatically
- **Automatic context management:** The interface handles history seamlessly
- **Memory responsibility:** The platform (OpenAI, Google, etc.) manages the conversation state

N8N Agent (API-based)

- **Stateless:** Each API call is independent
- **No persistent context:** LLM doesn't retain conversation history between calls
- **Manual context management:** N8N must explicitly pass the full conversation history in each request
- **Memory responsibility:** The workflow system (N8N) handles all context preservation

N8N's conversation memory structure essentially recreating what a web-based chat interface does automatically

N8N: Breaking Down The Components



Generating the full context for the LLM

Most recent user query

System prompt, Formatting instructions

Chat history

Most recent LLM output

```
{
  "action": "saveContext",
  "input": {
    "input": "Can you suggest what to pack for my trip based on this forecast.",
    "system_message": "# Role\nYou are the n8n Demo AI Agent.. ",
    "formatting_instructions": "IMPORTANT: For your response to user..."
  },
  "output": {
    "input": "Can you suggest what to pack for my trip based on this forecast.",
    "system_message": "# Role\nYou are the n8n Demo AI Agent..",
    "formatting_instructions": "IMPORTANT: For your response to...",
    "chat_history": [
      {
        "type": "HumanMessage",
        "content": "What is the weather forecast for San Diego, CA for the next 5 days?"
      },
      {
        "type": "AIMessage",
        "content": "Here's the weather forecast for San Diego...[WEATHER DATA]"
      }
    ],
    "output": "That's a great question! Based on the forecast...[PACKING ADVICE]"
  }
}
```

Every API call includes: current user input, complete system prompt, full conversation history

Demo 01

Fetching Information

Agentic Design

What is An “Agent”?



Open AI



Agents vs. Agentic Flow

	Agent - Definition A system or software entity that can independently perform tasks, typically guided by goals or objectives, interacting dynamically with an environment.	Agentic Flow - Definition A structured or semi-structured workflow or series of tasks leveraging multiple agents (or agent-like behaviors), LLMs, and external tools or APIs.
Feature	(Single) Agent	Agentic Flow
Autonomy	High Makes decisions independently	Medium → High Structured but with some autonomy per step
Coordination	Low/None Standalone tasks	High Multiple tasks coordinated
Complexity	(Typically) singular goal-oriented tasks	Multiple steps, tools, and integrations
Scalability	Limited	Highly scalable, flexible (modular approach)

Definitions: Go With the Flow!

"It's an agent!"



"No it's not!"

"It's agentic!"



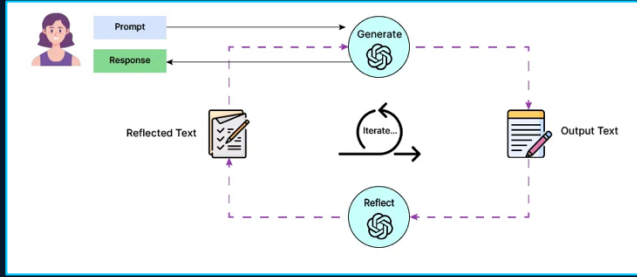
"Rather than having to choose whether or not something is an agent in a binary way, it would be more useful to think of systems as being agent-like to different degrees. Unlike the noun "agent," the adjective "agentic" allows us to contemplate such systems and include all of them in this growing movement."

-- Andrew Ng, DeepLearning.ai (2024)

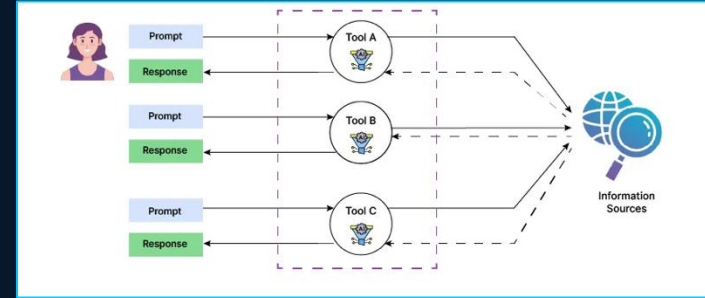
TL; DR: Prioritize practical capabilities and the "agentic" nature of workflows rather than debating definitions.

Agentic Reasoning: Design Patterns

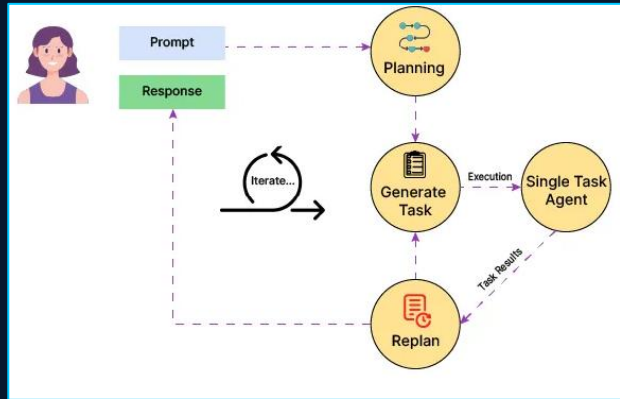
Iteration



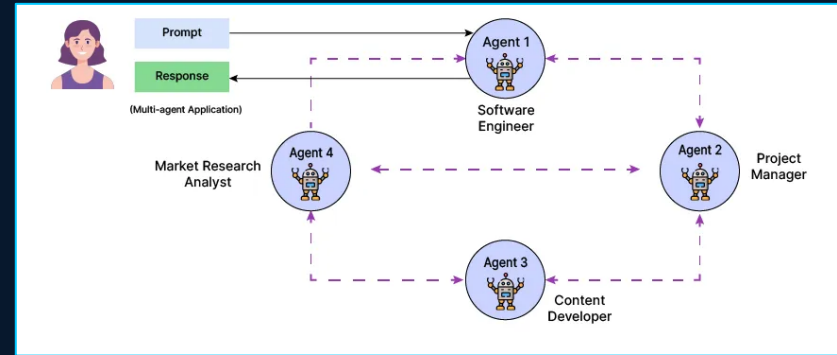
Tool Use



Planning



Multi-Agent Collaboration



Source(s):
 DeepLearning.ai: <https://www.deeplearning.ai/the-batch/lms-evolve-with-agentic-workflows-enabling-autonomous-reasoning-and-collaboration/>
 Analytics Vidhya: <https://www.analyticsvidhya.com/blog/2024/10/agentic-design-patterns/>

Significant Improvement on Benchmarks

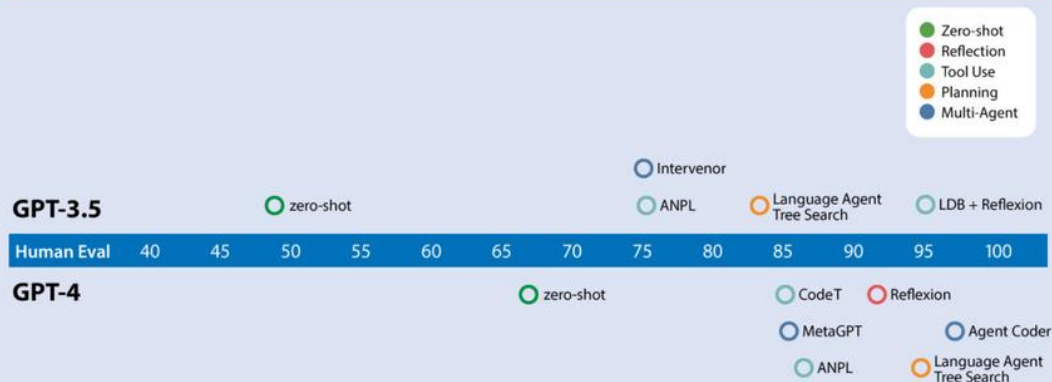
- HumanEval Benchmark (Coding)

- Zero shot:
- GPT-3.5: 48%
- GPT-4: 67%

- Using an agentic workflow with GPT-3.5: 95%

- Improvement just within GPT3.5 using agentic workflow dwarfs the improvement from GPT3.5 → GPT4 (which is a much better model overall).

GPT-3.5 and GPT-4 performance using zero-shot and agent workflows

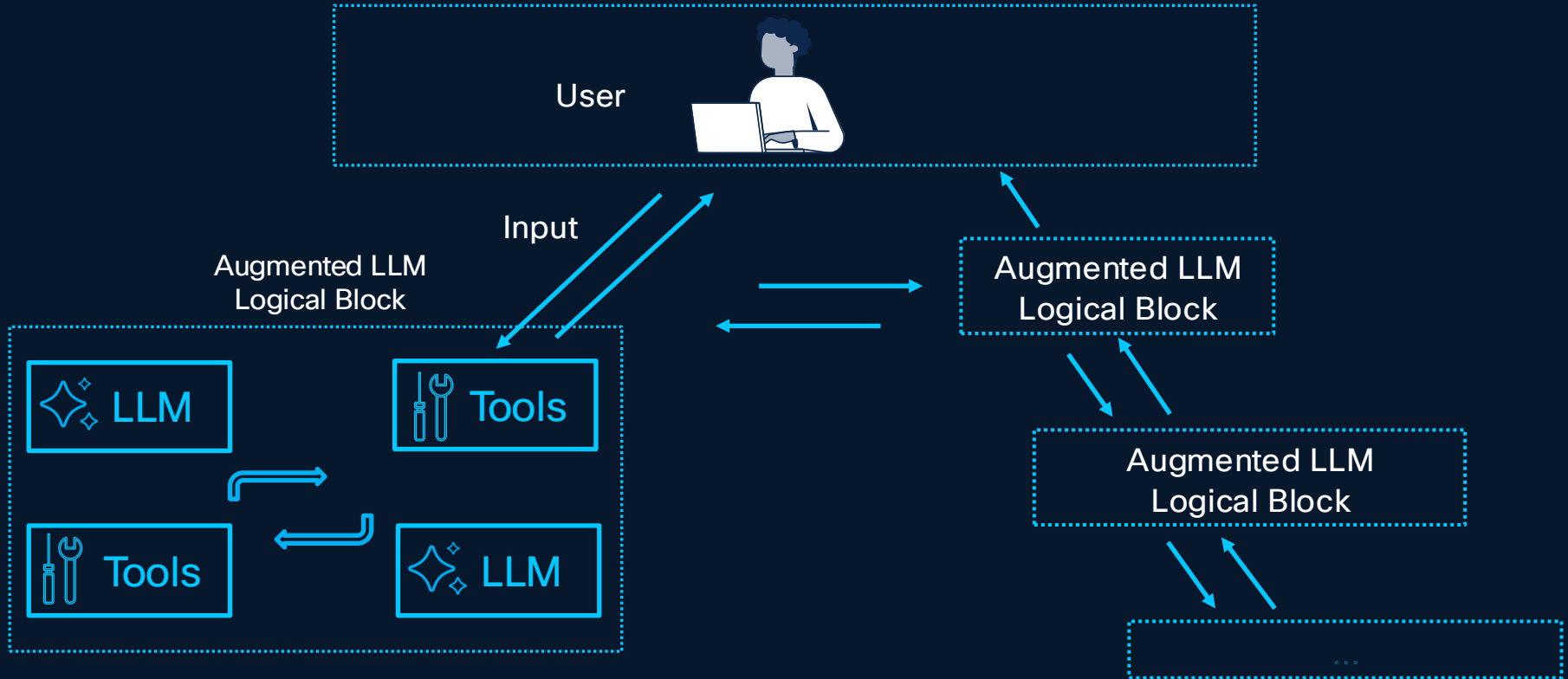


Performance of GPT-3.5 and GPT-4 (zero-shot) on HumanEval, along with algorithms that use agent workflows on top of GPT-3.5 or GPT-4. Thanks to Joaquin Dominguez and John Santerre for help with this analysis.

Sources:

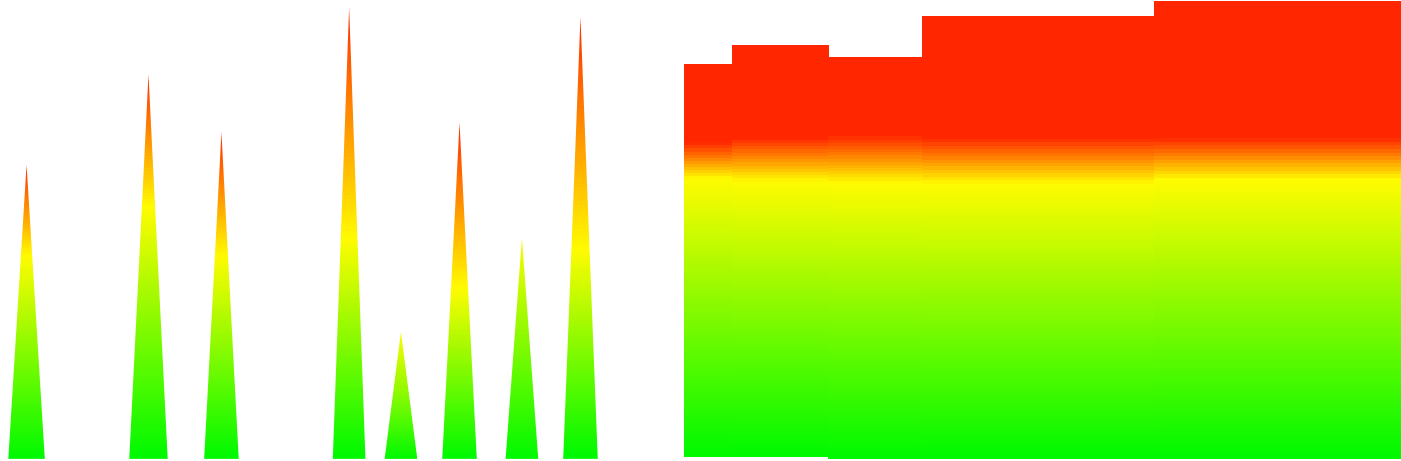
DeepLearning.ai: <https://www.deeplearning.ai/the-batch/how-agents-can-improve-llm-performance/>

Agentic Workflow Building Blocks



Systems can have agent-to-agent communication
(not necessary for user to observe every step)

Agentic AI and Inferencing Demand (Traffic)



Agentic Frameworks

Agentic Frameworks

Applications



Agentic
Orchestration Layer



Foundation Models,
Small LLMs



OpenAI
(ChatGPT)

ANTHROPIC

Anthropic
(Claude)



Meta
(LLaMA)

Gemini

Google
(Gemini)

Cloud



On-Premises

Infrastructure



GPU



CPU



Networking

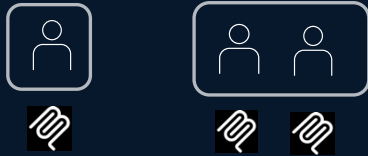


Storage

Scaling Agentic AI: Protocols

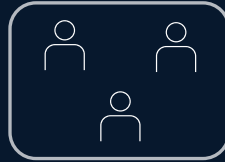
Agentic Market Evolution

Single Platform



Where Market is Today
90%

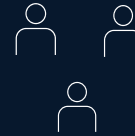
Agents created and hosted in a single platform using one or more MCP servers



Where the Market is Today
10%

Multiple agents created and hosted in a single platform interacting

Cross Platform



Where The Market is Headed

Multiple agents in different platforms interacting via protocols

Internet of Agents



A globally scaled network of interconnected agents, discovering and interacting with one another.

Source: Gartner Innovation Insight: Multiagent Systems, Key Reports

Agentic Computing: Needs a Collaboration Framework

come from different vendors

use different development frameworks

running on different clouds

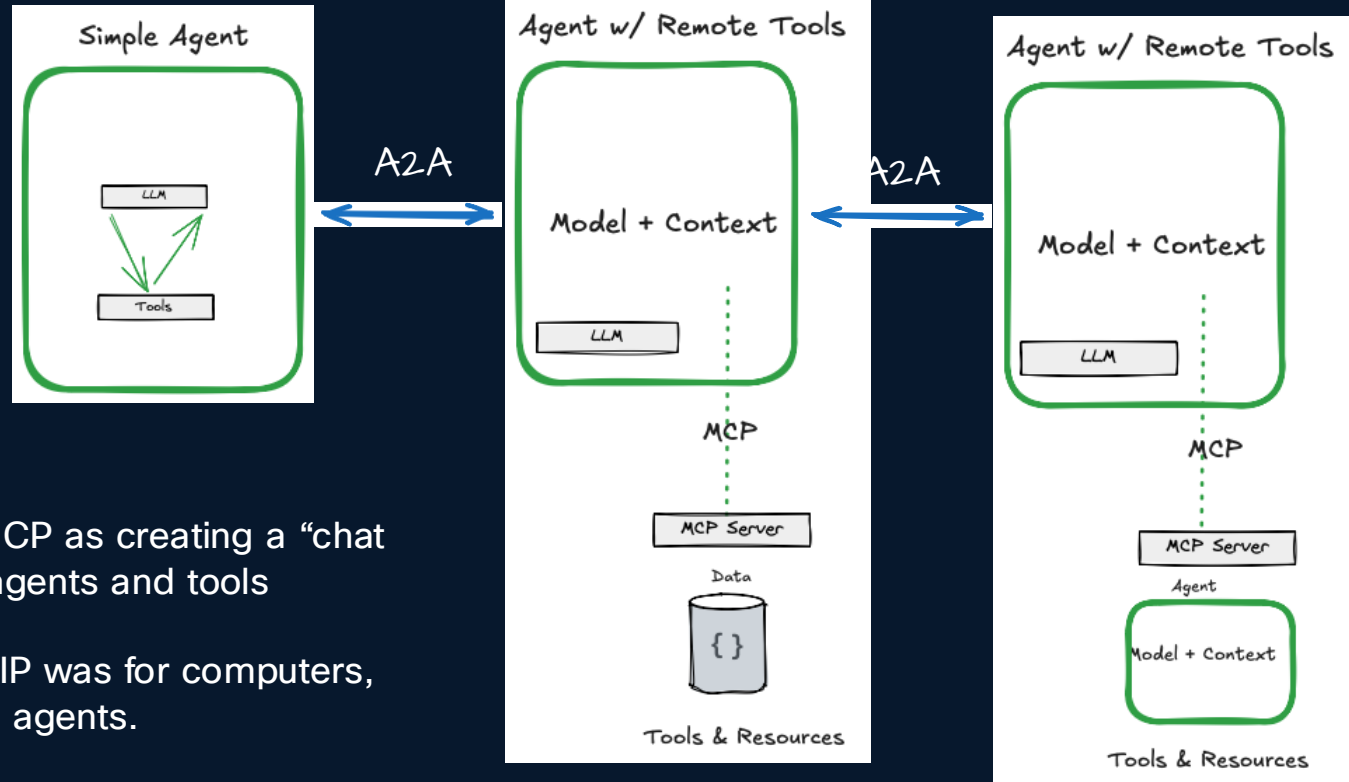
have different access controls

belong to different organizations

take on different profiles and personas

Inter-Agent and Tool Communication

What if agents and tools could have a “chat channel” to communicate?

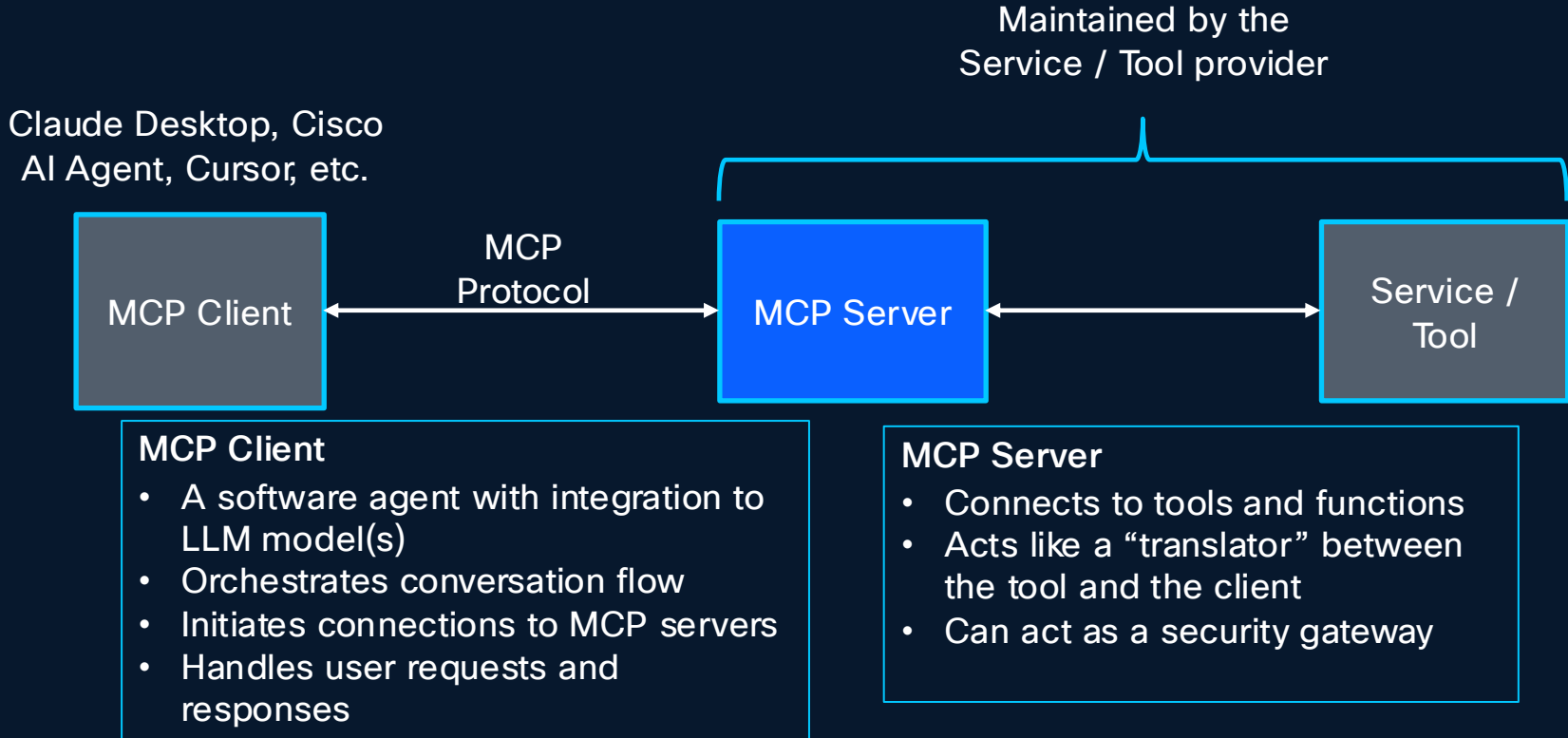


- Think of A2A and MCP as creating a “chat channel” between agents and tools
- Analogy: what TCP/IP was for computers, A2A/MCP are for AI agents.

Agentic Discovery Protocols: A2A, MCP

Feature	Agent-to-Agent (A2A)	Model Context Protocol (MCP)
Purpose	An open standard allowing AI heterogenous agents to communicate, collaborate effectively	An open standard designed to bridge AI assistants with the world of data and tools, to plug in many different sources of context
Original Creator	Google (April 2025)	Anthropic (November 2024)
Scope	Agent-to-agent collaboration	Agent-tool and agent-data integration
Discovery	Agent cards	Tools and services via MCP servers
Security Model	OAuth2.0, API keys	Application layer security

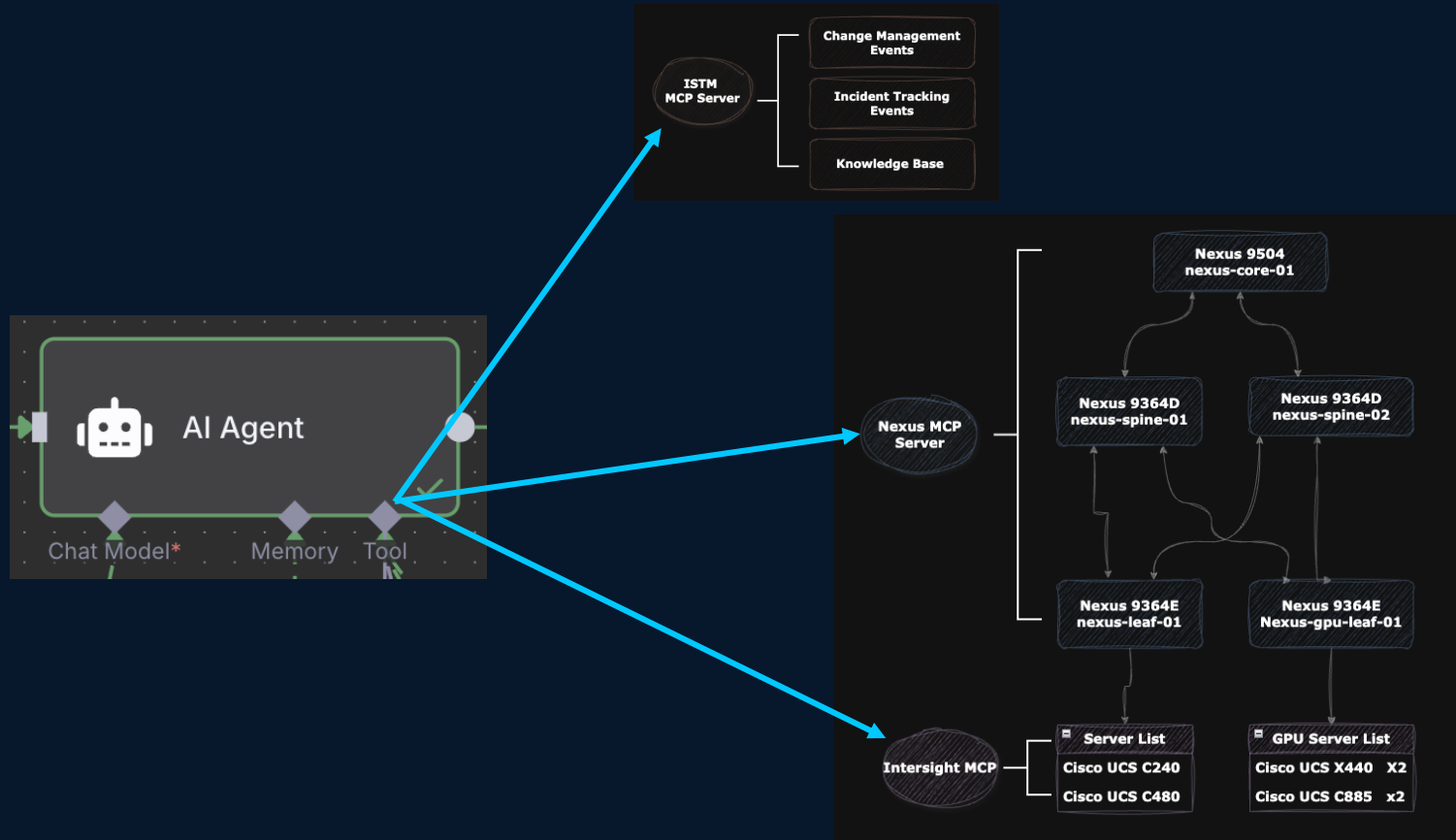
MCP: High-Level Flow



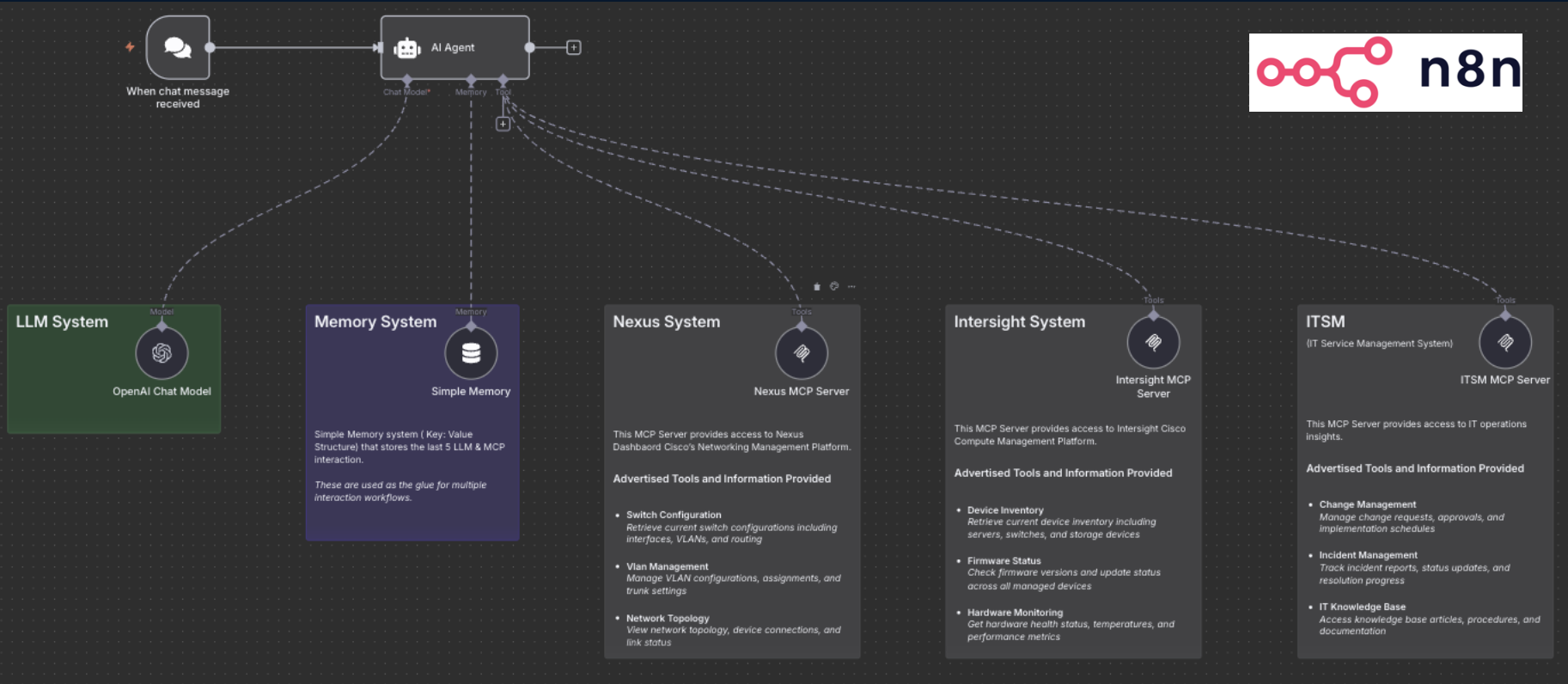
Demo 02

Basic Agentic AIOps

Demo 02: Agentic AI for Network Operations



Demo 02: Agentic AI for Network Operations



LLM System

Model

OpenAI Chat Model

Memory System

Memory

Simple Memory

Simple Memory system (Key: Value Structure) that stores the last 5 LLM & MCP interaction.

These are used as the glue for multiple interaction workflows.

Nexus System

Tools

Nexus MCP Server

This MCP Server provides access to Nexus Dashboard Cisco's Networking Management Platform.

Advised Tools and Information Provided

- **Switch Configuration**
Retrieve current switch configurations including interfaces, VLANs, and routing
- **Vlan Management**
Manage VLAN configurations, assignments, and trunk settings
- **Network Topology**
View network topology, device connections, and link status

Intersight System

Tools

Intersight MCP Server

This MCP Server provides access to Intersight Cisco Compute Management Platform.

Advised Tools and Information Provided

- **Device Inventory**
Retrieve current device inventory including servers, switches, and storage devices
- **Firmware Status**
Check firmware versions and update status across all managed devices
- **Hardware Monitoring**
Get hardware health status, temperatures, and performance metrics

ITSM
(IT Service Management System)

Tools

ITSM MCP Server

This MCP Server provides access to IT operations insights.

Advised Tools and Information Provided

- **Change Management**
Manage change requests, approvals, and implementation schedules
- **Incident Management**
Track incident reports, status updates, and resolution progress
- **IT Knowledge Base**
Access knowledge base articles, procedures, and documentation



Demo 02: Agentic AI for Network Operations

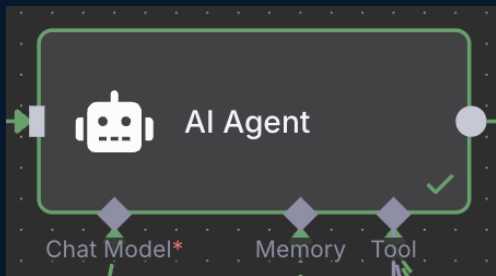
[VID

The screenshot displays the n8n AgenticOps interface. At the top, the workflow is titled "Personal AgenticOps-Wx2 WebexOne". The main workspace shows a workflow starting with a trigger "When chat message received" leading to an "AI Agent" node. The AI Agent is connected to several MCP (Model Context Protocol) servers:

- LLM System:** OpenAI Chat Model
- Memory System:** Simple Memory (Simple Memory system [Key: Value Structure] that stores the last 5 LLM & MCP interaction. These are used as the glue for multiple interaction workflows.)
- Nexus System:** Nexus MCP Server (This MCP Server provides access to Nexus Dashboard Cisco's Networking Management Platform. Advertised Tools and Information Provided: Switch Configuration, Vlan Management, Network Topology.)
- Intersight System:** Intersight MCP Server (This MCP Server provides access to Intersight Cisco Compute Management Platform. Advertised Tools and Information Provided: Device Inventory, Firmware Status, Hardware Monitoring.)
- ITSM:** ITSM MCP Server (This MCP Server provides access to IT operations insights. Advertised Tools and Information Provided: Change Management, Incident Management, IT Knowledge Base.)

At the bottom, there is a chat interface with a session ID "cd43e...", a "Send a message below to trigger the chat workflow" prompt, and a "Nothing to display yet. Execute the workflow to see execution logs." message.

Demo 02: Agentic AI for Network Operations



You are a network automation assistant specializing in Cisco network infrastructure, server operations, and IT Service Management.

AVAILABLE MCP TOOLS:

You have access to the following MCP tools and you MUST use them when users ask related questions:

****Intersight MCP Server Tools:****

- device_inventory: Call this tool for any inventory/device list requests

...

****Nexus MCP Server Tools:****

- switch_configuration: Call this tool for switch/interface configuration questions

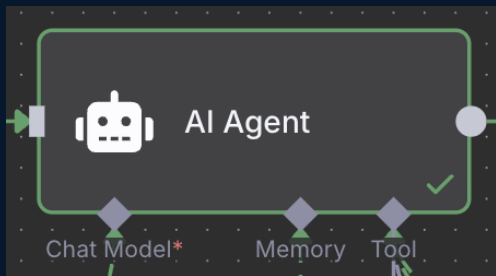
...

****ITSM MCP Server Tools:****

- change_management: Call this tool for change request/maintenance questions

...

Demo 02: Agentic AI for Network Operations



****INVENTORY WORKFLOW:****

When users request inventory information, follow this workflow:

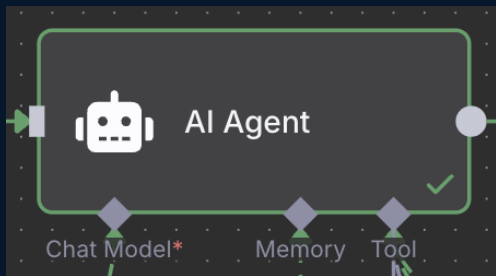
1. ****SERVER INVENTORY****: Call 'device_inventory' from Intersight MCP Server to get server information
2. ****NETWORK INVENTORY****: Call 'switch_configuration' from Nexus MCP Server to get switch information
3. ****COMBINED PRESENTATION****: Present both server and network device inventories in a unified format

Format the combined inventory response as:

- ****Servers Summary**** (count, types, status overview)

...

Demo 02: Agentic AI for Network Operations



```
**ITSM AUDIT**:
```

After receiving the firmware audit results, call the following tools from the ITSM MCP Server and pass the EXACT JSON output from the 'firmware_status' tool as context:

- 'change_management' tool to get planned changes that might conflict
- 'incident_tracking' tool to get current outages that might impact upgrades
- 'knowledge_base_access' tool to get relevant KB articles

This will provide:

- devices: The original device list from firmware_status
- itsm_changemanagement_items: Planned changes that might conflict
- itsm_outage_items: Current incidents that might impact upgrades
- itsm_knowledgebase_items: Relevant KB articles

Demo 02: Agentic AI for Network Operations

Prompt: Please build the plan to upgrade the servers to support the newest firmware

Success in 19.00s Overview Details

AI Agent

- Simple Memory
- OpenAI Chat Model
- Intersight MCP Server
- OpenAI Chat Model
- ITSM MCP Server
- ITSM MCP Server
- ITSM MCP Server
- OpenAI Chat Model
- Simple Memory

When chat message received

I have gathered the firmware audit data and ITSM context to build server firmware upgrade plan

High-Level Upgrade Steps:

Change Request
CHG0001234: Net
UTC (high prio
CHG0001235: Ser
UTC (medium pr

Priority Device
1 UCS-C480-001
2 UCS-C240-001

Issues:

Device Name
UCS-C885A-006
UCS-X440P-003
UCS-X440P-004
UCS-C885A-005
High-Level Upg

1. Schedule firmware upgrades during the approved Server Patching maintenance window (2024-01-22 01:00-04:00 UTC).
2. Upgrade UCS-C480-002 first due to critical update status.
3. Upgrade UCS-C240-001 next.
4. Follow KB001235 Server Maintenance Procedures for patching.
5. Monitor for any incidents or conflicts before and during the upgrade.
6. Avoid upgrades during ongoing critical incidents (e.g., Email Service Outage).
7. Notify stakeholders and obtain change approval per risk level.
8. Perform backups and snapshots before upgrades.
9. Validate system functionality post-upgrade.
10. Document changes and update firmware inventory.

Emerging Protocol Stack



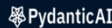
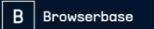
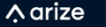
A Linux Foundation Project
An open-source collective for inter-agent collaboration

The AGNTCY is where we are building the Internet of Agents to be: A diverse, collaborative space to innovate, develop, and maintain software components and services for agentic workflows and multi-agent software.

Formative Partners

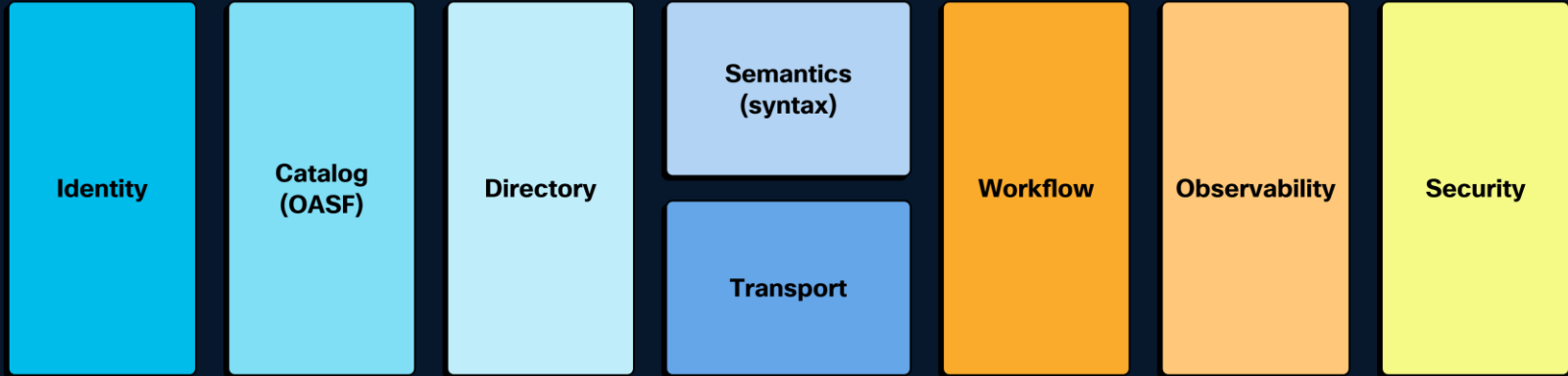


75+ collaborating organizations

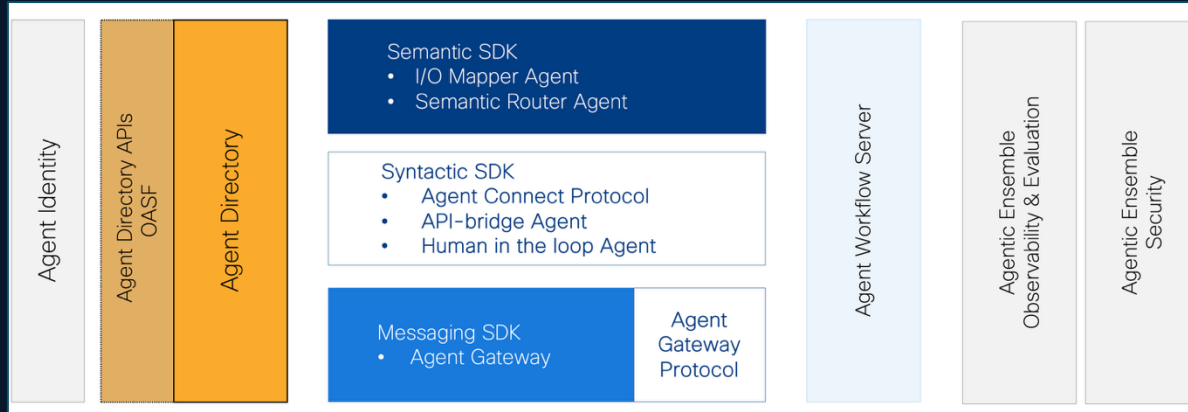


Agency: Protocol Stacks

AGNTCY



Agentic Protocol Stack: AGNTCY



Interoperability: Establish a common protocol that enables AI agents **from different vendors** and platforms to **communicate and work together**

Security: Ensure **secure interactions** between agents through robust authentication, authorization, and encryption mechanisms.

Scalability: Design a **scalable architecture** that leverages the cloud-native stack optimally, supporting a growing number of agents and interactions without compromising performance.

Standardization: Develop **standardized data models, schemas** to ensure consistent data representation and validation across ecosystem

Agency: Protocol Stacks

AGNTCY



Protocols

MCP

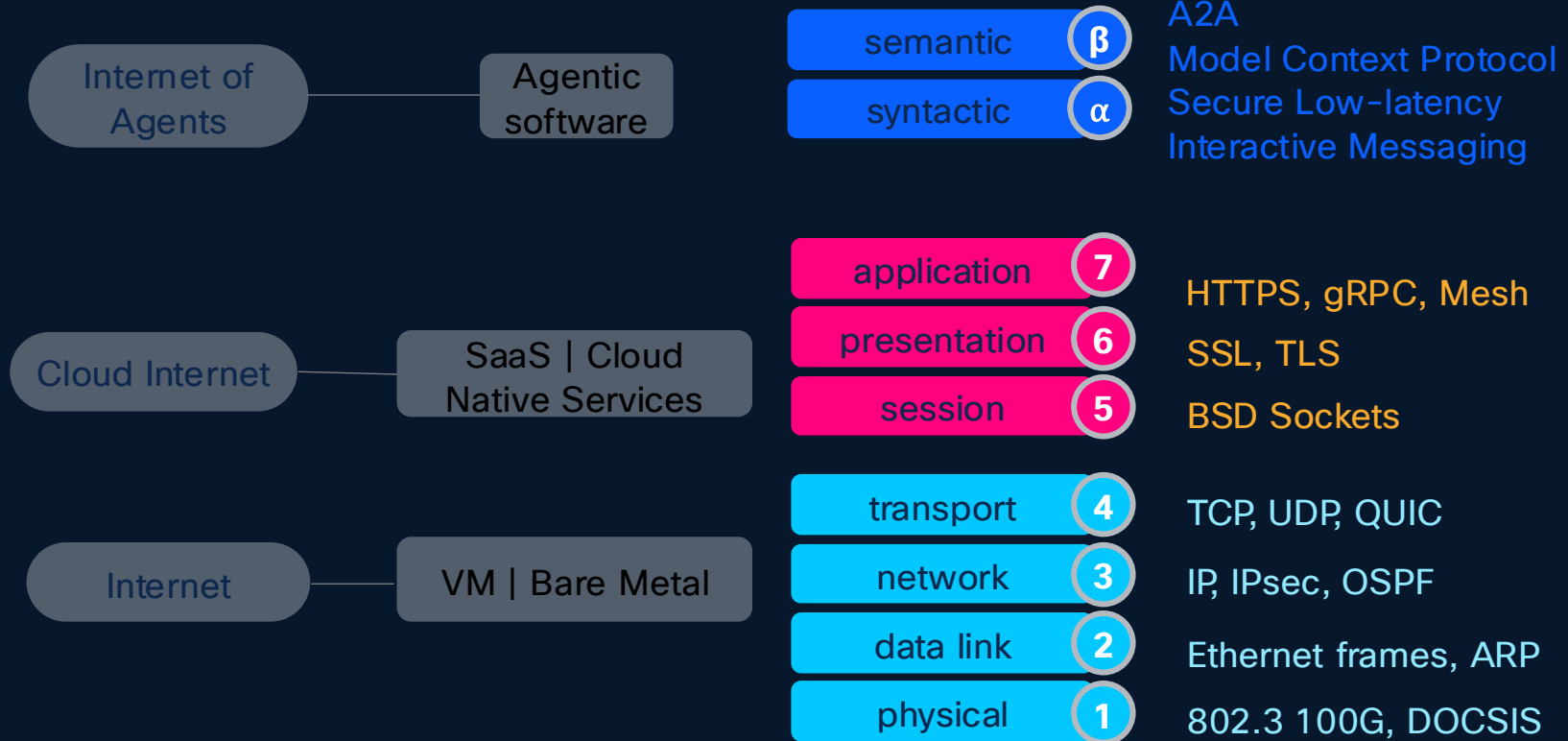
A2A

...

ACP

SLIM

Evolving the OSI model



AI Ops in Action: AI Canvas and Deep Network Model

AgenticOps

Once you have “chat channels” everywhere, you need Agentic AI Ops to monitor, secure, and orchestrate them



AIOps

Key Features of AI Canvas

AgenticOps

Cross-domain

Multiplayer

Purpose-built models

AI Canvas: Cisco's AgenticOps Platform

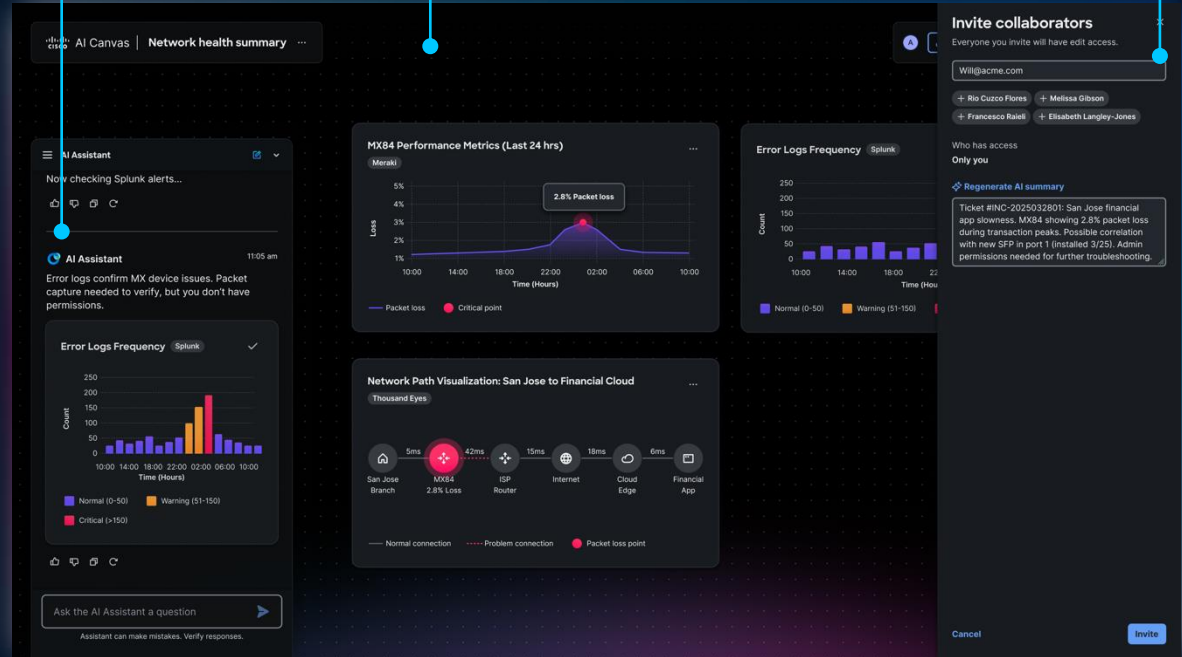
IT teams *interface* with AI Assistant and *solve* with AI Canvas

- Single canvas for cross domain troubleshooting
- Generative UI with reasoning built-in
- Keeps NetOps, SecOps, IT and execs on the same page

AI Assistant

Shared Workspace

Users



Disclaimer: Many of the products and features mentioned are still in development and will be made available as they are finalized, subject to ongoing evolution in development and innovation. The timeline for their release is subject to change.

AI Canvas: Cisco's AgenticOps Platform

AI Canvas | Application performance degradation

Generate report View activity Share

AI Assistant

Jackie shared this canvas with you. Here is the background:

Ticket #SRTK0023941: San Jose financial app failures. Cisco Meraki dashboard shows MX105 with significant packet loss on WAN1 interface. ThousandEyes Test ID 895623 segment visualization shows packet loss (16.4% total). Financial transactions experiencing frequent failures and timeouts, Maria Chen's email included and she is asking for a status update.

Ask the AI Assistant a question

Assistant can make mistakes. Verify responses.

ServiceNow SRTK0023941

Reported by System Administrator on 03/28/2025 at 09:45 AM PST

Description Received a ThousandEyes alert and reached out to Maria Chen to confirm. She said users at San Jose branch experiencing 3-5 second delays when processing financial transactions in EFR. Started approximately 30 minutes ago. Affects all 24 users at the branch. No recent changes reported.

ServiceNow Ticket SRTK0023941

SJ-MX105-01 WAN interface performance

Loss rate **16.4%** (2.3% change)

WAN interface latency 87ms (baseline <30ms)
Jitter 15ms (baseline <5ms)

SJ-MX105-01 performance (last 24 hrs)

Time (Hours)	Packet Loss (%)
20:00	0
24:00	0
04:00	0
08:00	0
12:00	16.4
16:00	0
20:00	0

Network segment performance analysis

Enterprise network 16.4% loss Gateway Internet Application financeapp.com

Maria Chen <mchen@company.com>
To: IT Support <tsupport@company.com>
Cc: Will Jenkins <wjenkins@company.com>, Jackie Torres <jtorres@company.com>

URGENT: Status Update Needed on Financial Application Performance Issues

Hello Team,

I'm following up regarding the ongoing performance issues with the Enterprise Financial Platform (EFP) affecting our San Jose branch. This performance degradation is severely impacting our ability to serve customers. We have end-of-month financial processing scheduled for tomorrow, and we need reassurance that this issue will be resolved by then.

Could you please provide an immediate status update on troubleshooting efforts and expected resolution timeframe? If there's a workaround we can implement in the meantime, that would be extremely helpful.

Thank you for your urgent attention to this matter.

Best regards, Maria Chen
Senior Director, Financial Operations

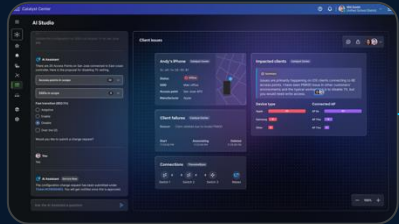
© 2025 Cisco and/or its affiliates. All rights reserved. Cisco Public

AI Canvas: Cisco's AgenticOps Platform

AI Assistant



AI Canvas



Campus and Branch



Topology, configuration, location



Voice and video experience



Cisco Meraki

Topology, configuration, location



WAN, Internet, App Insights



WAN Details



Identity Intelligence

Identity verification, user trust levels



Nexus Dashboard

Data center network management



Hyperfabric

Data center network management



Intersight

Unified management, automation, security

Data Center

Security and Observability



Cisco and third-party insights



Firewall

Security & connection events



ISE

Authentication Insights



Duo

Authentication & compliance



Secure Access

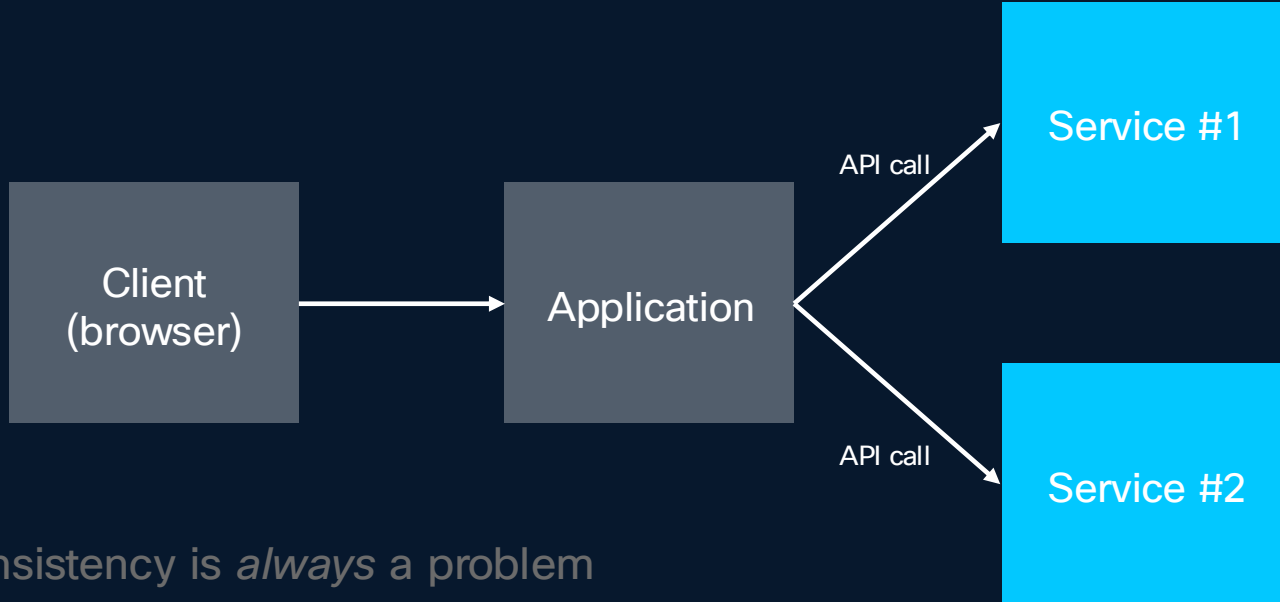
Private & SAAS resource access



XDR

Related threat incidents

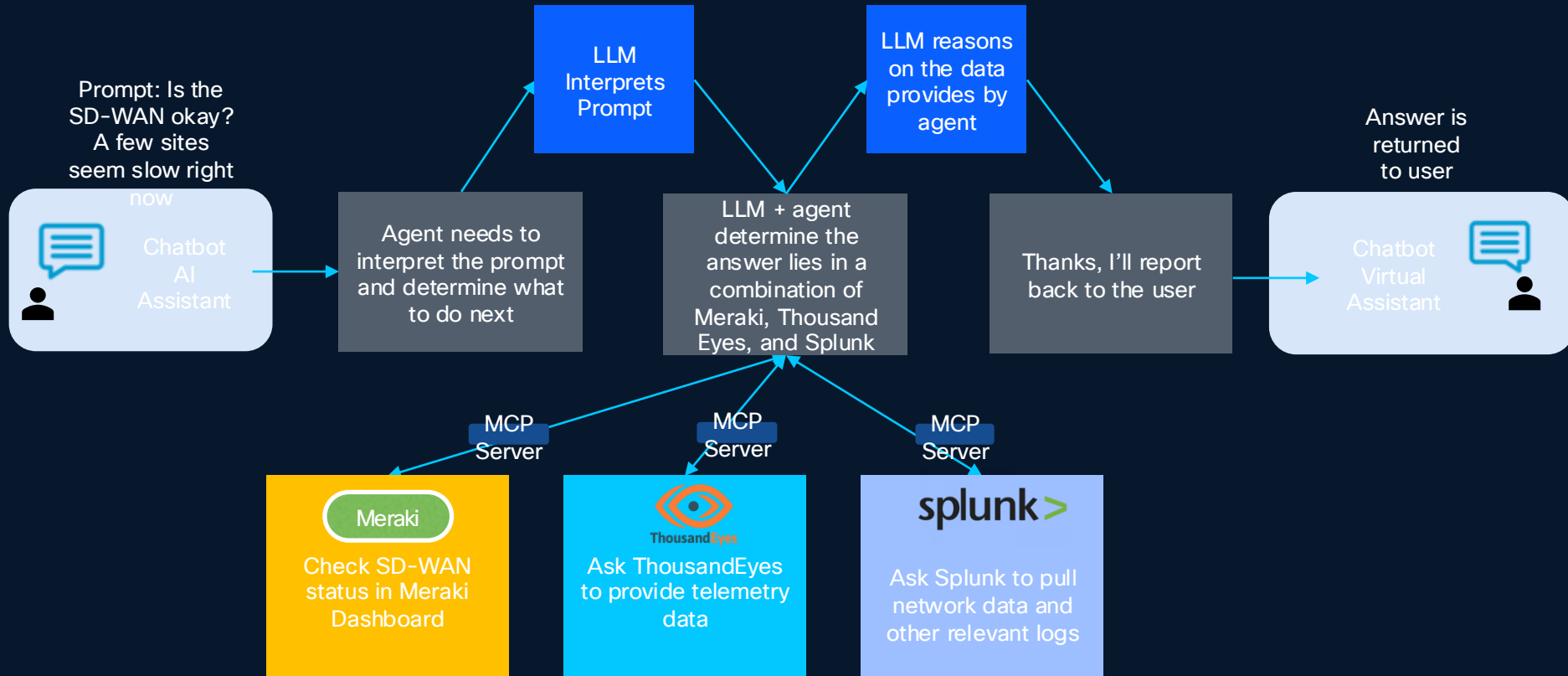
APIs and Consistency



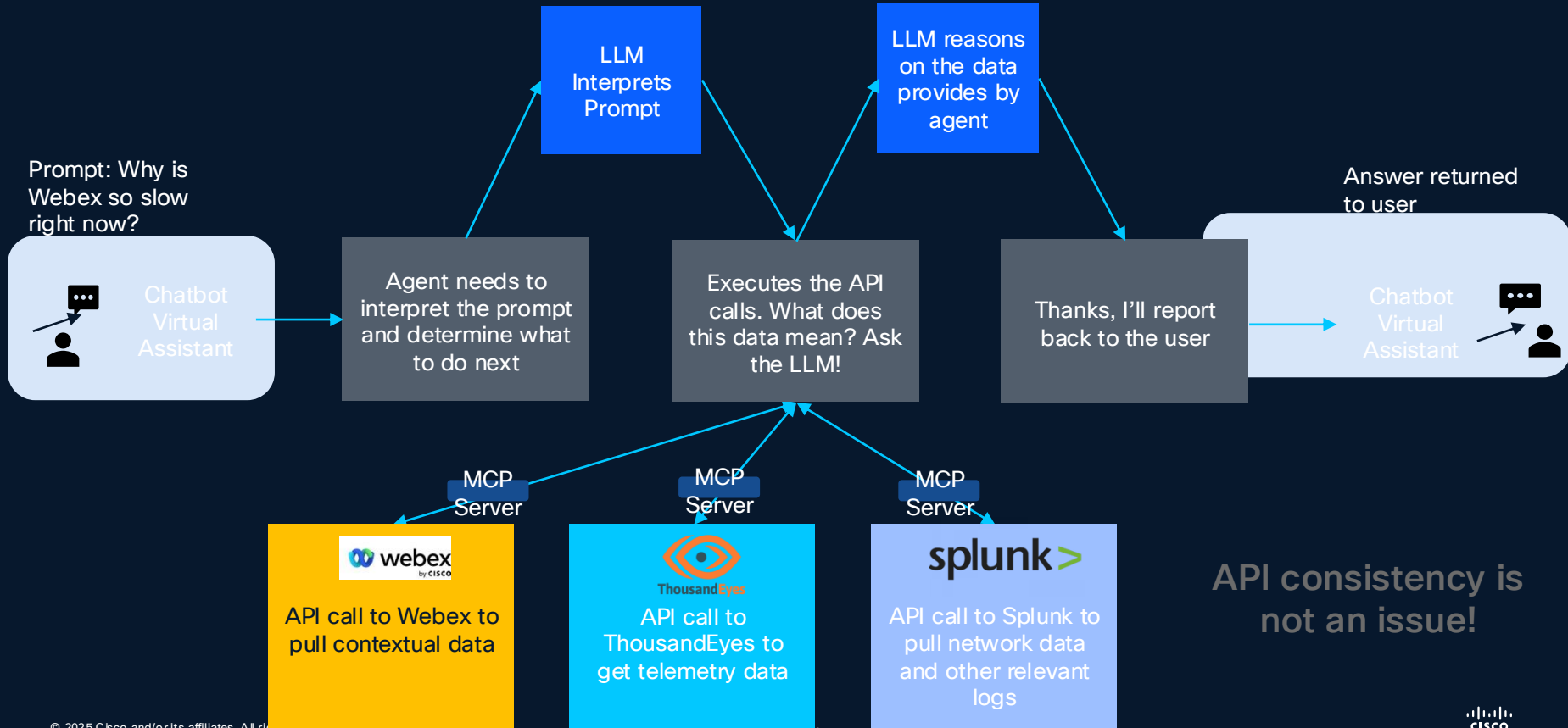
API consistency is *always* a problem

If service 1 changes their API, dependent applications will not work!

AI Ops → Agentic Ops



Webex Troubleshooting With MCP



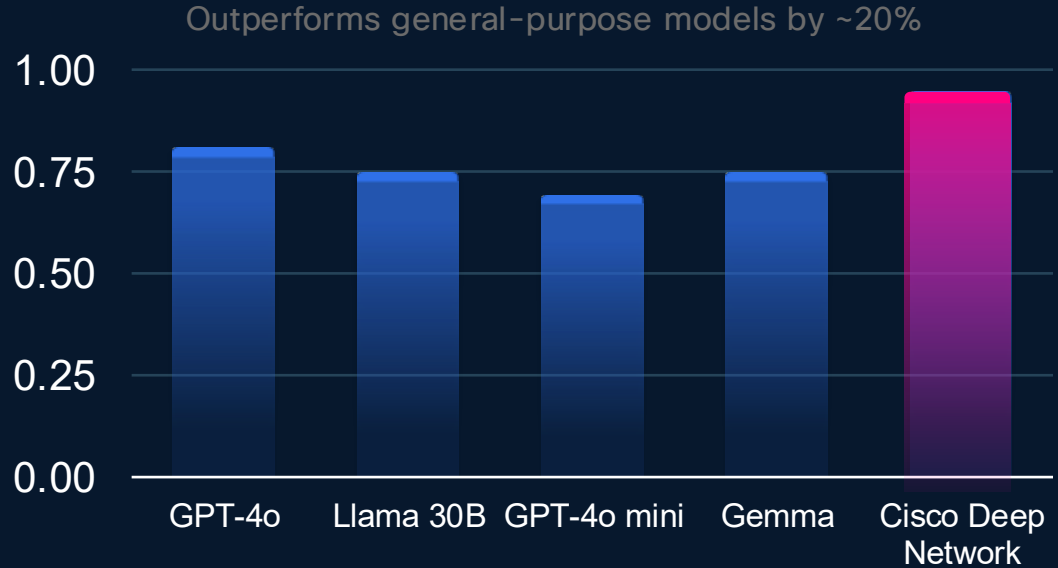
What are Agentic (IT) Tools?

- Any Cisco product
- External code execution
- API Call to external system
- A Web search
- Calculators (for math functions)
- Identity services (e.g. authentication, etc.)
- A database
- ...

Deep Network Model

Purpose-built for networking, expert accuracy

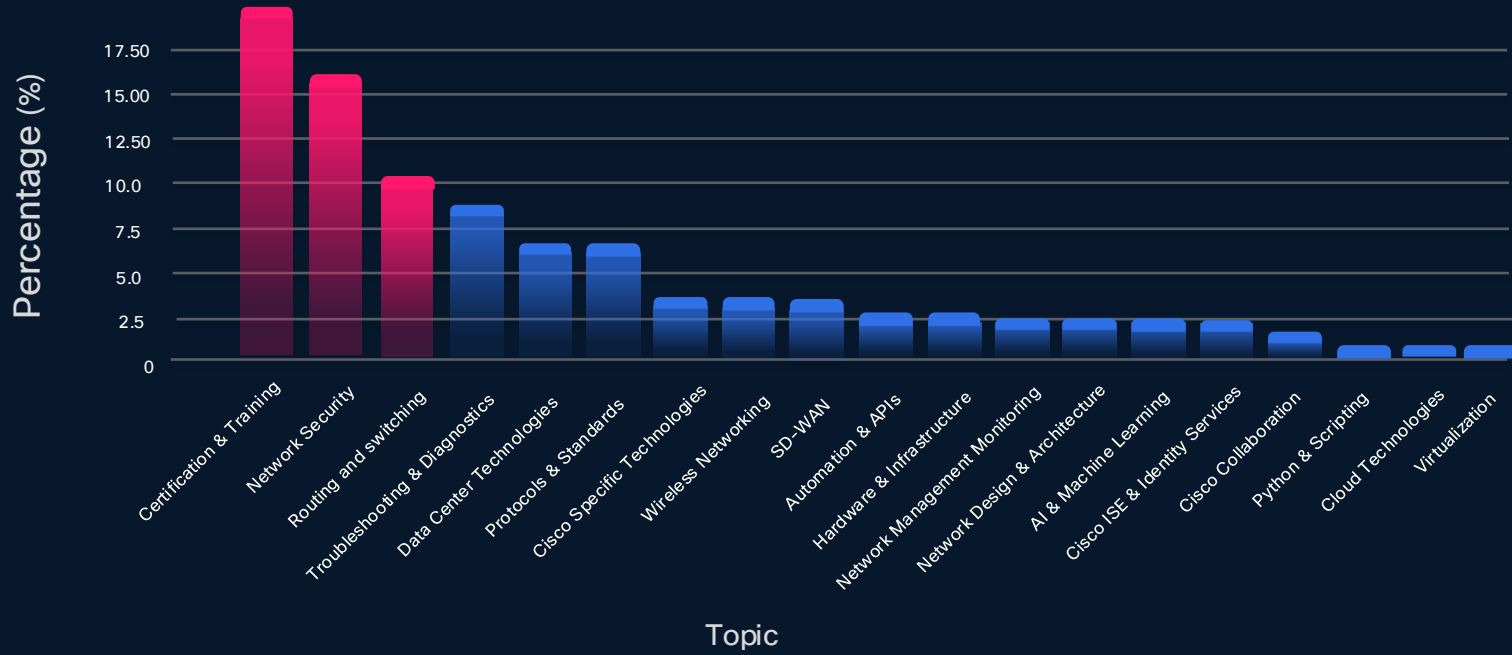
- More precise reasoning for troubleshooting, configuration, and automation
- Fine-tuned on 40+ years of expertise and expert-vetted for accuracy
- Evolves with live telemetry and real-world Cisco TAC and CX insights



Accuracy on CCIE-style multiple choice questions
(590-question benchmark), May 2025

Deep Network Model: Prompting

Percentage Distribution of Topics



Deep Network Model / Foundation

- Built upon an existing foundation model
- Various models are being used, and this will change over time
- Decision criteria is:
 - Foundation benchmark performance
 - Tunability
 - Model size (bigger isn't always better)

Datasets for Deep Network Model

- Cisco TAC Methodologies
- TAC Cases (without the PII data)
- Engineering traces
- Other useful internal documentation
- We need your help!!!

Agents don't take away our work – they take away the work that took us away from value.

What remains is where human creativity, judgment, and leadership shine.

Thank you



