Securely Integrating Large Language Model (LLM) and Proprietary Data for Deep Analytics

October 2025

We will be discussing leveraging Large Language Models (LLM) vast knowledge and analytics capabilities to securely unlock insights from proprietary real-time data and documents.

Agenda

1. Foundational Concepts

- Large Language Models (LLMs)
- Intelligent Agents
- Retrieval Augmented Generation (RAG)

2. Model Context Protocol (MCP)

- Model Context Protocol (MCP) – Definition & Features
- MCP Adoption & Momentum
- User, AI Agent, LLM, MCP Server Interaction

3. Real World Application/Demo

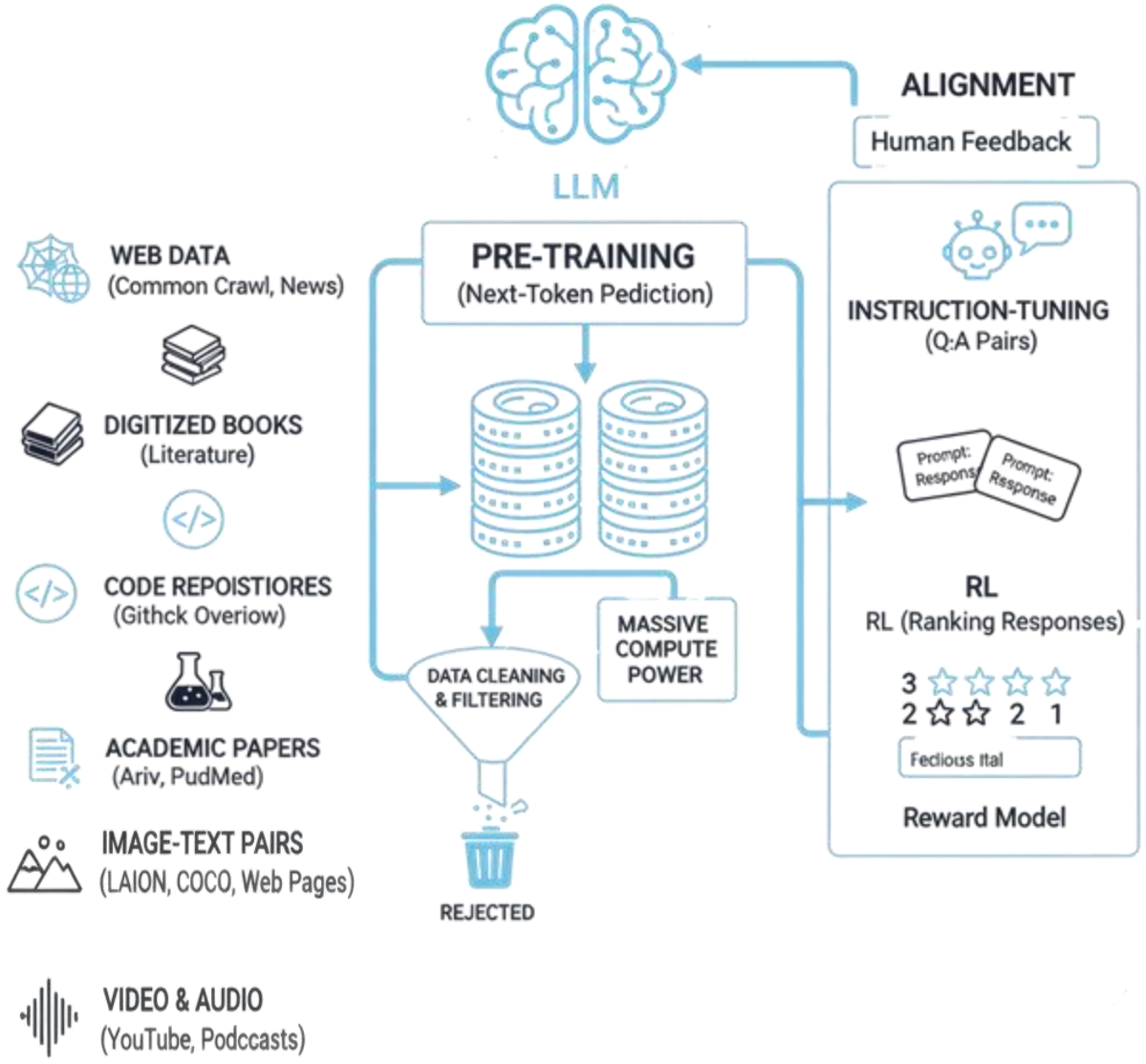
- Demo Setup & Architecture
- Live Demo Prompts & Scenarios
- Strategic Use Case Ideas

FOUNDATIONAL CONCEPTS

Large Language Models (LLMs)

- An LLM is an AI system trained on vast amounts of data
- They can understand and generate human language, answer questions, summarize information, translate language, and analyze complex and large amounts of data to find patterns and insights
  - **Examples:** GPT-4, Claude 3, Llama 3, Gemini 1.5

Think of it like a skilled research analyst who quickly reviews large volumes of documents to provide clear answers and meaningful summaries

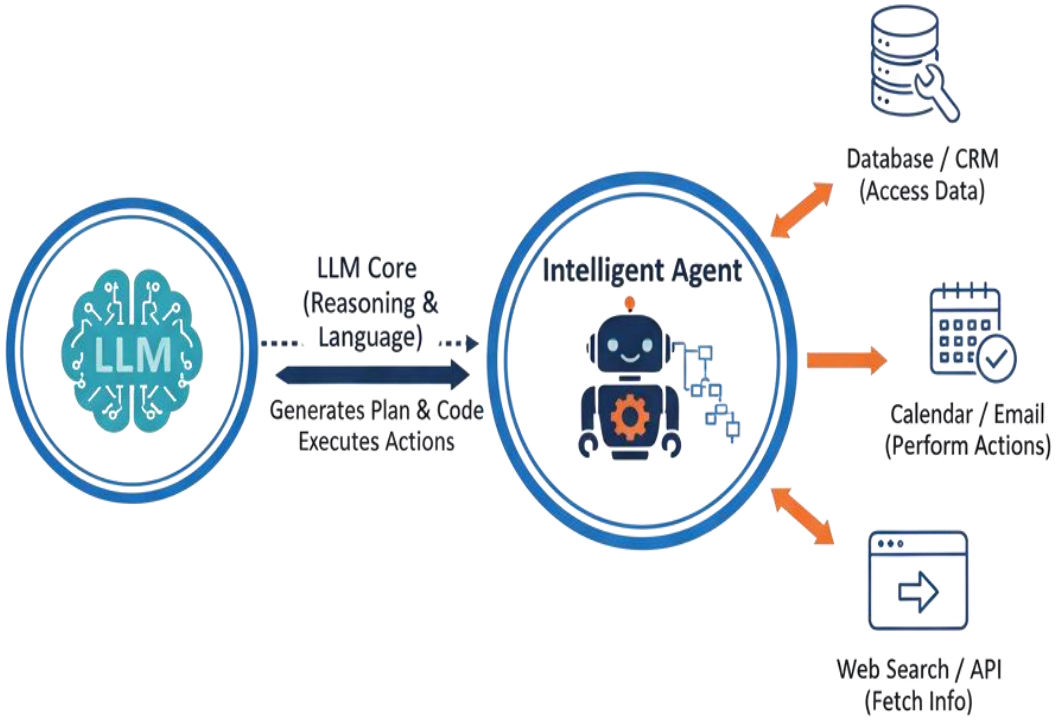


Intelligent Agents

- Intelligent Agents are autonomous software programs that uses an LLM as its “brain” to perform complex, multi-step tasks.
  - **Examples:** Gemini, ChatGPT, and Claude
  - **Example use case:** An enterprise agent could be a financial agent that uses live stock market data along with the **firm's proprietary risk parameters** to generate high-potential trade ideas."

Think of it as sales strategist who connects proprietary sales data with public market trends to automatically adjust strategy and maximize revenue

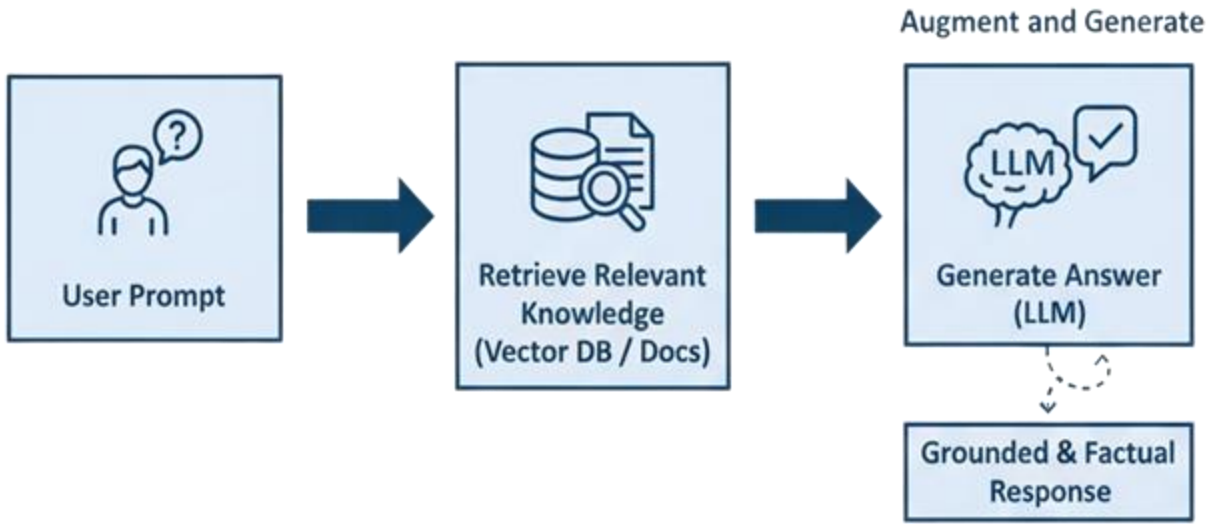
Intelligent Agents: Orchestrating LLM and Tools



LLM acts as the “Brain” to select and use the right tool

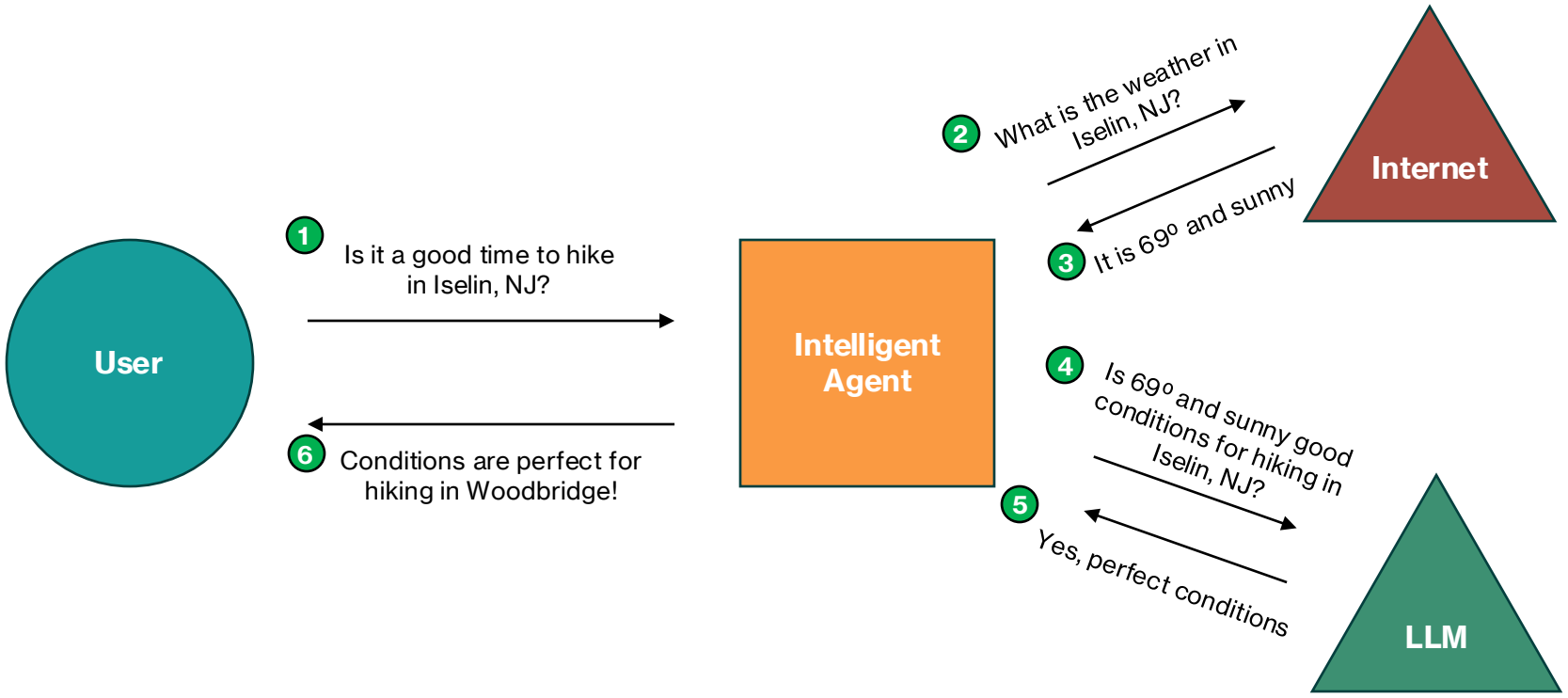
Retrieval Augmented Generation (RAG)

- RAG enhances LLMs by giving them access to up-to-date, **external/proprietary information** it was not originally trained on.
- RAG solves two critical problems for LLMs:
  1. **Knowledge Cutoff:** It allows the AI to use information that is new (like today's sport scores or news headlines).
  2. **Proprietary Data Access:** It allows the AI to use internal or private data (like a Company's internal records/databases) to provide valuable insights tailored to a Company's specific needs.



Retrieval Augmented Generation (RAG)

- **Examples**:
  - Modern LLM apps like Gemini and ChatGPT use RAG to provide current answers by fetching data from the web
    - Prior to RAG capabilities, LLM would often state that their knowledge was limited to a specific date as they could not fetch updated data in real time



Thanks to RAG **(steps 2 & 3)**, the intelligent agent is able to retrieve real-time data for LLM reasoning **(steps 4 & 5)** to provide an accurate, timely answer **(step 6)**

MODEL CONTEXT PROTOCOL (MCP)

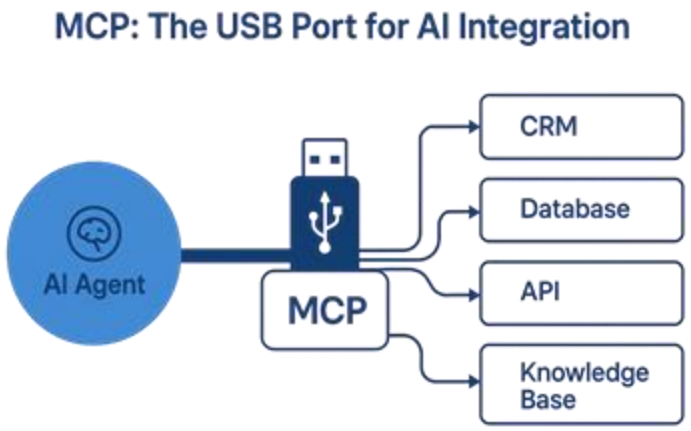
Model Context Protocol (MCP)

- While RAG is the specific technique for augmenting knowledge, MCP is a general-purpose, governed protocol for Agent-Environment interaction that enables RAG, Tool Use, and live data access.

- MCP serves as on open standard that allows AI models and agents to access proprietary external data and systems **securely** and efficiently.

- The MCP Server tells the AI Agent exactly what capabilities are available. If we were to update or change a backend system, the MCP immediately advertises these changes, and the **AI Agent automatically adapts**. This means the Agent can use new or updated tools instantly.

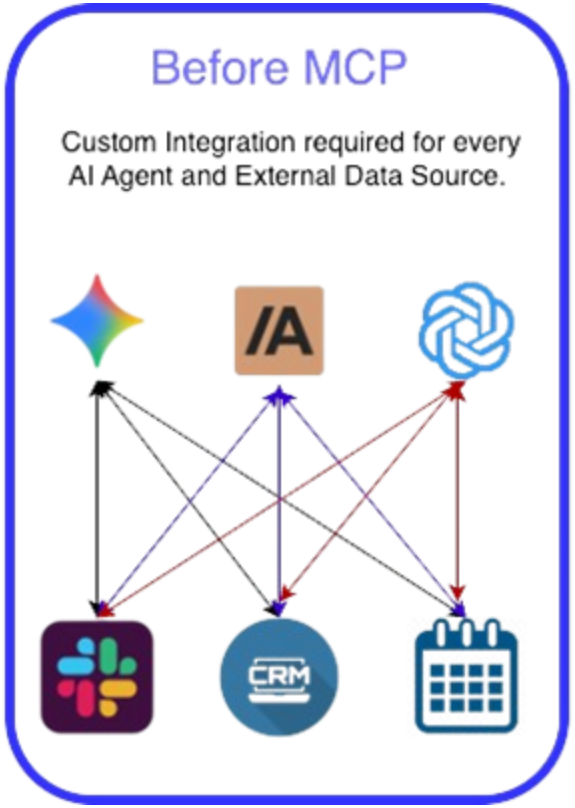
Think of MCP like a USB-C port for AI, enabling model to "plug in" to any compatible tool or database

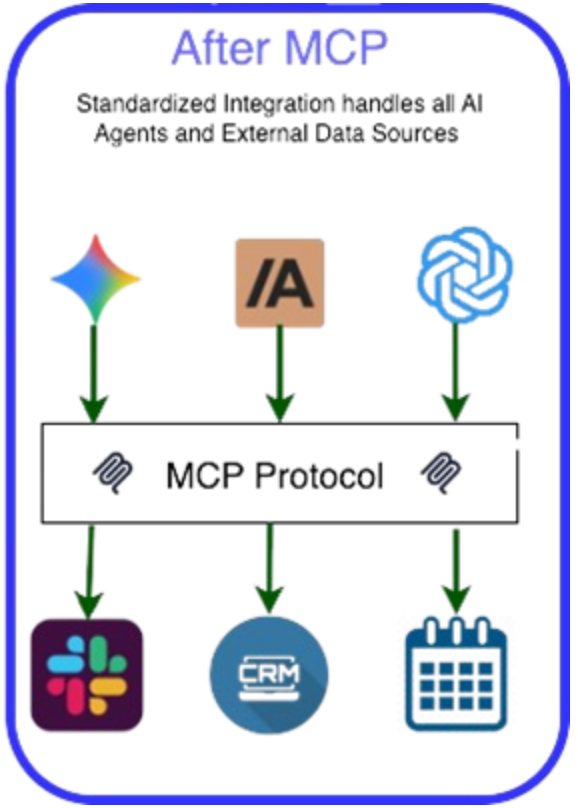


MCP can leverage a user's context to enforce **data entitlements**. This ensures the AI can only retrieve and access data the user is authorized to see

MCP - Adoption & Momentum

- **Value Proposition:** Before MCP, custom integration was required for every single AI model and external system. This created a tangled 'spaghetti architecture' that was costly, insecure, and less scalable

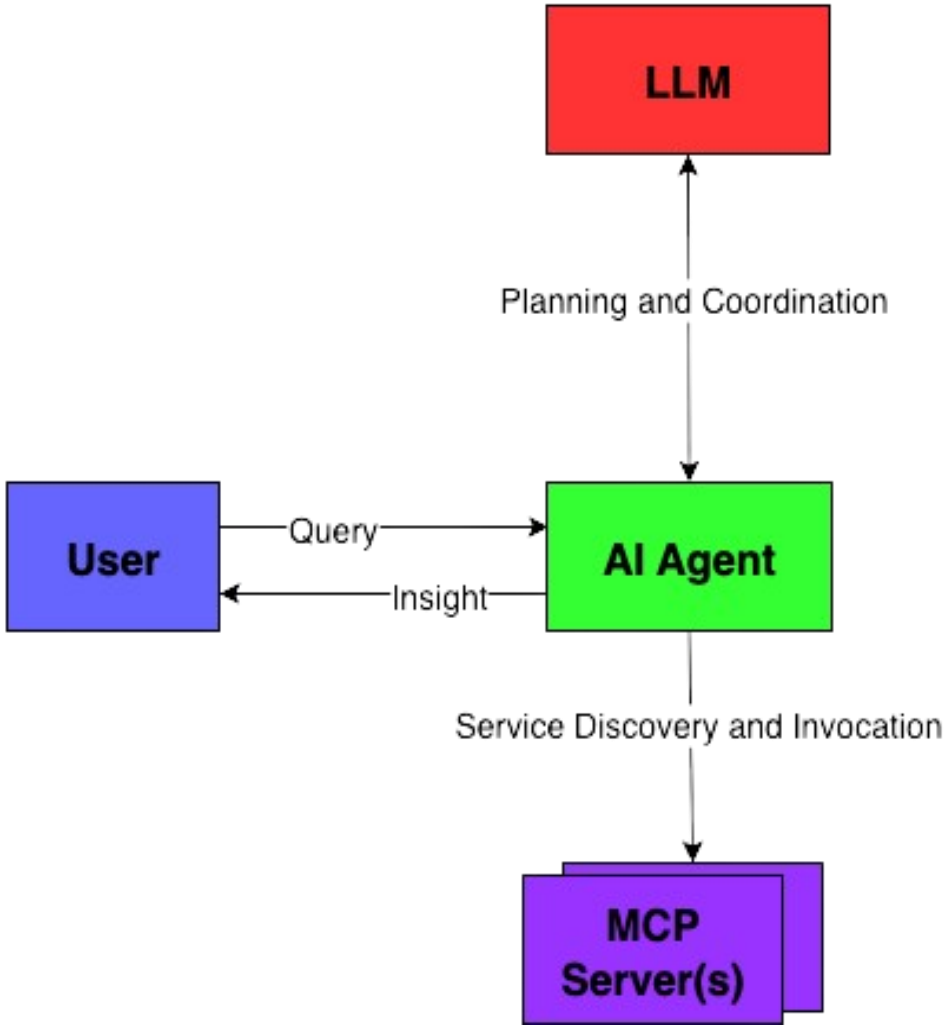




MCP - Adoption & Momentum

- ✓ **Protocol Creator:** Anthropic (Open-sourced the protocol in November 2024)
- ✓ **Governance:** Managed by an Open-Source Steering Group following a transparent governance model.
- ✓ **Major Adopters:** OpenAI, Google DeepMind, Microsoft (Azure), AWS (Amazon Bedrock, Amazon Q), Cloudflare, GitHub, Slack
- ✓ **SDK Availability:** Official SDKs for 8+ major languages, including Python, TypeScript, Java, C#, Go, and Rust.
- ✓ **Server Growth:** Nearly 16,000 unique MCP servers developed and available on marketplaces.
- ✓ **Enterprise Impact:** Early adopters like Block (Square) report saving 50–75% of time on common engineering tasks (e.g., integration, context management).
- ✓ **Future Outlook (2026):** 75% of API Gateway vendors and 50% of iPaaS vendors expected to include native MCP features.

User, AI Agent, LLM, MCP Server Interaction



REAL WORLD APPLICATION/DEMO

Demo Objective and Technical Setup

- **This demo shows how to combine:**

Real-time sales data and **proprietary product documents** (specifications, promotions, and warranties).

The vast knowledge and analytical abilities of a Large **Language Model (LLM)**.

By doing this, deep insights can be gained to inform critical business decisions, such as identifying new product bundles or creating targeted promotional strategies.

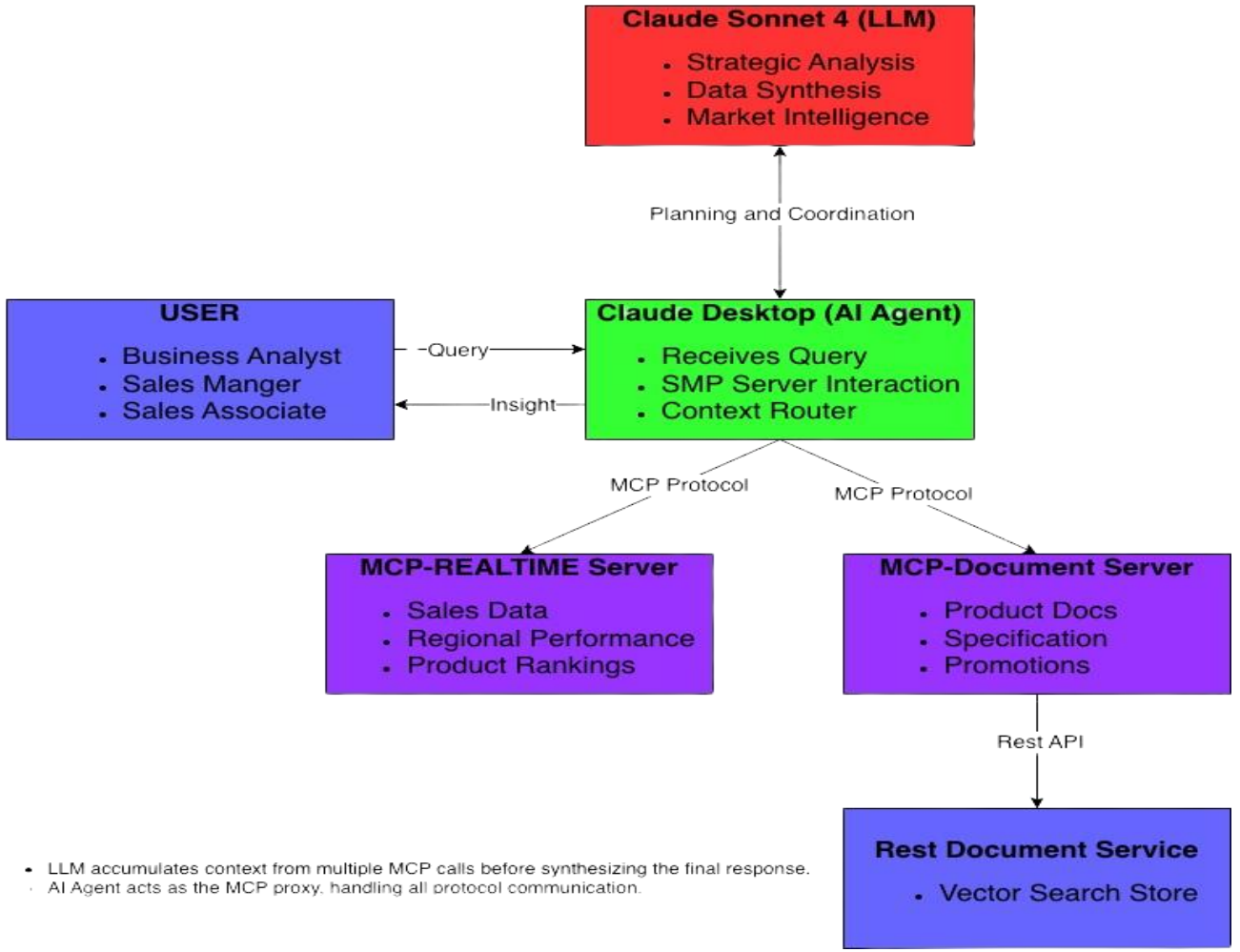
Demo Objective and Technical Setup

- This is made possible through a collaboration between an **Intelligent Agent** and **Model Context Protocol (MCP) Servers.**

Intelligent Agent (Claude Desktop)	MCP Servers
Powered by an advanced AI model, it acts as the central brain. It listens to questions, devises a plan to find the answers, and orchestrates the entire process.	Provide secure, efficient access to data and documents, enabling the agent to get the information it needs to answer questions.

- The intelligent agent (Claude Desktop) listens to a user's question, figures out which data sources it needs to consult, and then uses the MCP servers to collect the necessary information. Then, it combines the real-time data, relevant document details, and its own analytical abilities to provide a clear and actionable answer.

Demo Objective and Technical Setup



- LLM accumulates context from multiple MCP calls before synthesizing the final response.
- AI Agent acts as the MCP proxy, handling all protocol communication.

Demo Prompts

Prompt 1: Product Sales Alignment Analysis

Prompt	*"Our top-selling product is generating great revenue, but I want to make sure we're maximizing its potential. Analyze our best performer against our available accessories and tell me what bundles we should create."*		
What This Shows	**MCP Real Time**	**MCP Document**	**LLM Value-Add**
	Pulls top products by revenue	Searches accessories documentation	Correlates sales performance with product ecosystem, identifies bundle opportunities, calculates revenue potential

Demo Prompts

Prompt 2: Regional Strategy Optimization

Prompt

"Which region should I prioritize for our Q4 push, and what specific promotions from our catalog would work best there?"

What This Shows

MCP Real Time	MCP Document	LLM Value-Add
Regional sales distribution data	Current promotional strategies and campaigns	Regional performance analysis, promotion-to-performance matching, strategic recommendations

Demo Prompts

Prompt 3: Produce Performance Gap Analysis

Prompt	*"I notice our wireless mouse sales are low despite being featured in promotions. Help me understand why and what we should do about it."*		
What This Shows	**MCP Real Time**	**MCP Document**	**LLM Value-Add**
	Specific product performance metrics	Product specifications and promotional details	Gap analysis between marketing and performance, product positioning insights, competitive assessment

Demo Prompts

Prompt 4: Seasonal Strategy Development

Prompt

"Based on our current sales and available promotions, create a holiday shopping strategy that maximizes our revenue potential."

What This Shows

MCP Real Time	MCP Document	LLM Value-Add
Current product performance baseline	Holiday promotions and bundle documentation	Market timing insights, cross-selling strategies, revenue projection modeling

Demo Prompts

Prompt 5: Competitive Positioning Analysis

Prompt

"How does our UltraBook Pro 15 compare to market competitors, and should we adjust our pricing strategy based on our current sales performance?"

What This Shows

MCP Real Time	MCP Document	LLM Value-Add
Actual sales performance data	Product specifications and features	Market research integration, competitive analysis, pricing strategy recommendations

APPENDIX

Demo Installation

- All the demo code is available on Github. Please follow the instruction in ReadMe file to install and run
- MCP Realtime Server: https://github.com/AsifRajwani/MCP-Server
- MCP RAG Bridge Server: https://github.com/AsifRajwani/MCP-RAG-Bridge
- Rest RAG Service: https://github.com/AsifRajwani/RAG-service

Demo Installation

Once code is working, follow this steps to setup the Claude Desktop.

1. Download Claude Desktop for you operating system. (https://claude.ai/download)

2. Go to Claude Desktop menu item “Developer” and select “Edit Config”. This will take you to claude_desktop_config.json

3. Add following to claude_desktop_config.json and restart Claude desktop

```
{
        "mcpServers": {
                "MCP-Realtime": {
                        "command": "node",
                        "args": [    "Absolute Path to mcp-server.js"
                        ]
                },
                "MCP-RAG": {
                        "command": "node",
                        "args": [
                                "Absolute Path to rag-mcp-bridge.js"
                        ]
                }
        }
}
```

Demo Installation

4. Go to Claude Desktop menu item “Developer” and make sure both MCP Servers are there.

5. Go to Claude Desktop menu item “Connectors” and configure both server for all methods to “Allow unsupervised” access.

6. Restart the Claude Desktop.

7. Run the following prompt for Claude Desktop chat interface and see servers are called.

How does our UltraBook Pro 15 compare to market competitors, and should we adjust our pricing strategy based on our current sales performance?