MLOPS- The new branch in DevOps

Nikhil Ghodki
AIOperations Engineering Technical Leader

DEVNET-1371

What all OPs words have you heard of?

Cisco Webex App

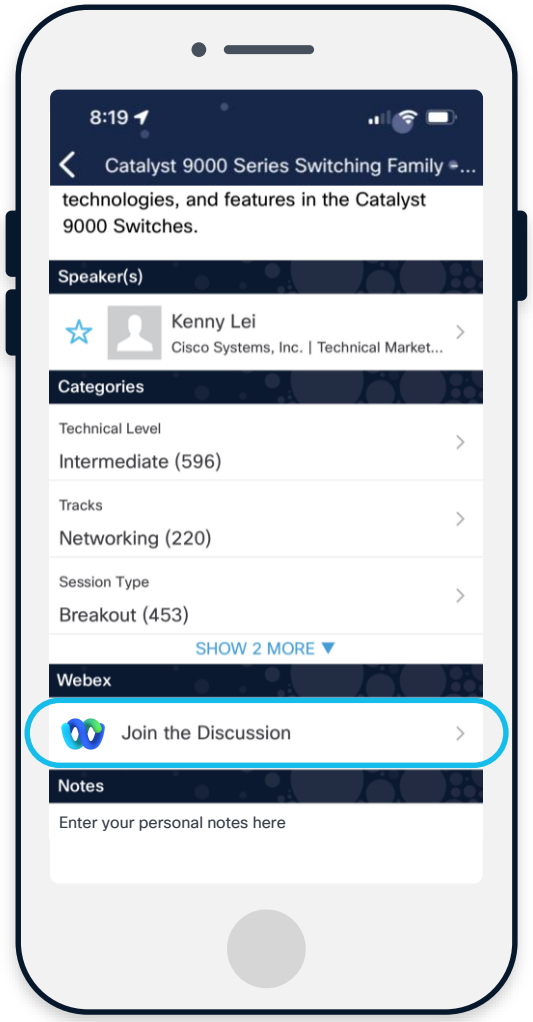
Questions?

Use Cisco Webex App to chat
with the speaker after the session

How

1. Find this session in the Cisco Live Mobile App
2. Click “Join the Discussion”
3. Install the Webex App or go directly to the Webex space
4. Enter messages/questions in the Webex space

Webex spaces will be moderated by the speaker until June 13, 2025.



Agenda

01 State of AI

02 Types of OPs

03 MLOPs

04 GenAIOPs

05 AgenticOPs

06 Sample Application

07 What it Takes for Application

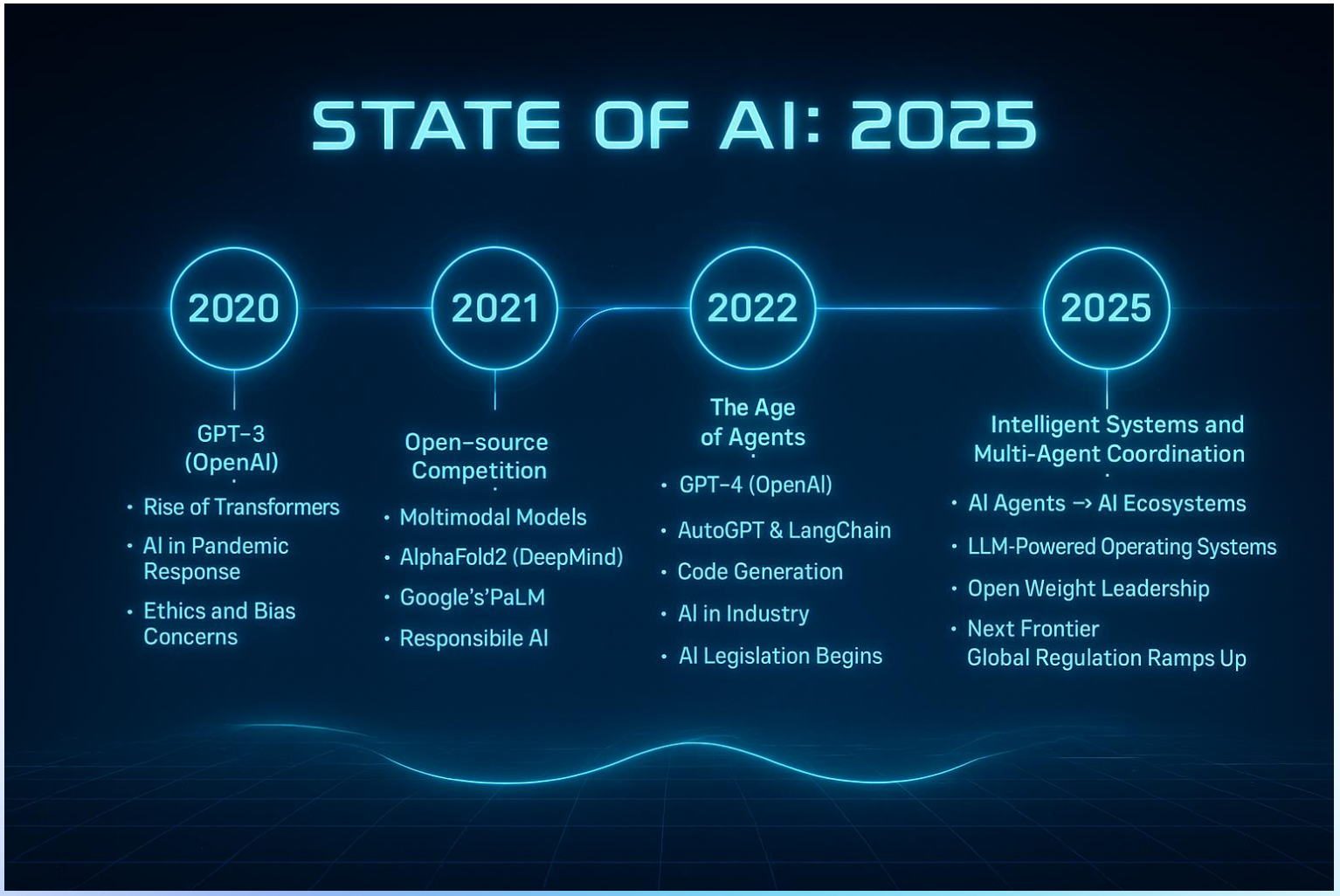
Who am I?

Nikhil Ghodki (He/Him)
AIOps Engineering Technical Leader at Cisco, AI/MLOPs

Happy to connect with you on LinkedIn
https://www.linkedin.com/in/nikghodki/

- Skillsets : MLOPs, Cybersecurity, AI Research, Infrastructure
- Personal : 10+ years of Security industry experience, wore hats for various roles, Double Masters (M.S. in Instrumentation, M.S. in Information Technology and Management)
- What I do : Architect Infrastructure for AI platform at Cisco, Design and implement MLOPs Pipeline, build small AI based tools for increasing team productivity, AI Research LoRA finetuning, etc

State of Artificial Intelligence



Is it really MLOps?

MLOPs

AIOPs

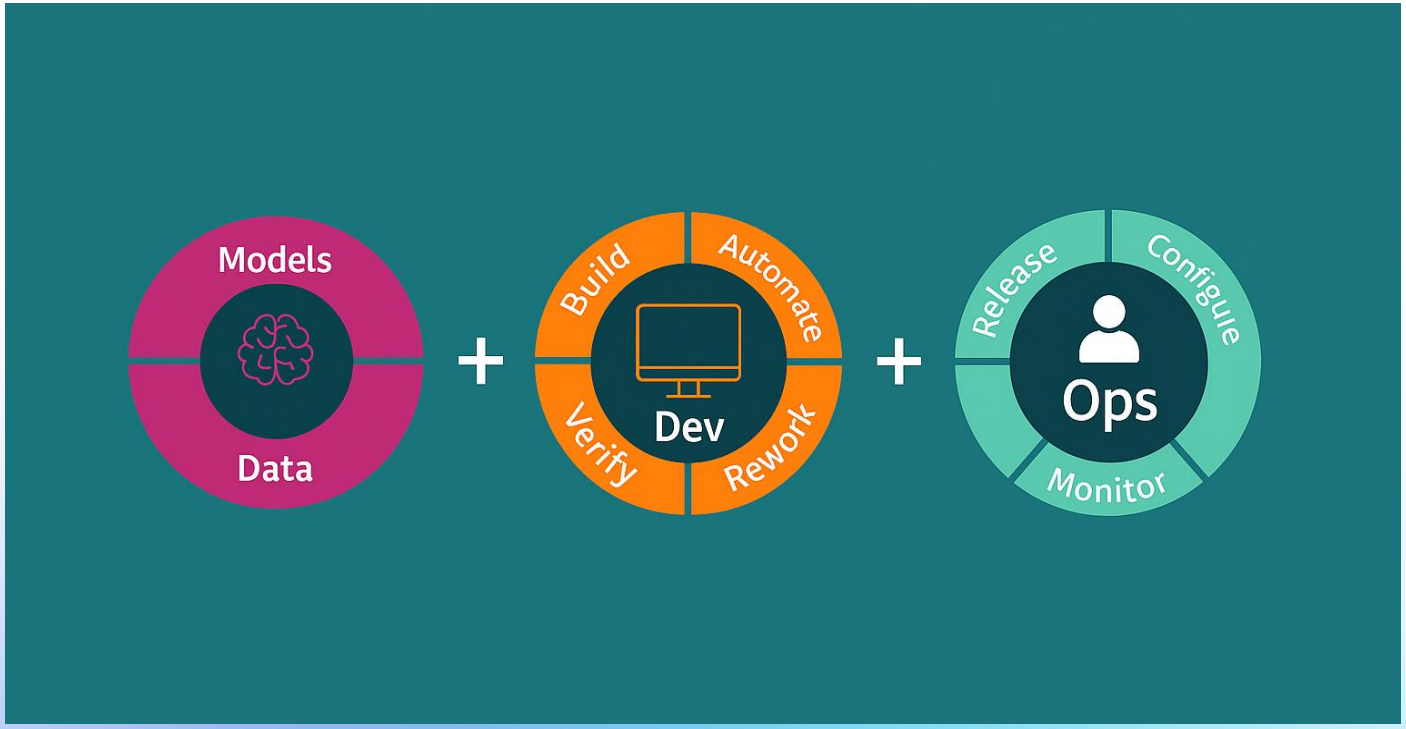
AgenticOps

GenAIOPs

MLOps (Ops for GenAI ?)

MLOPs – What is this?

ML + DEV + Ops = MLOps



Main challenges

“IT Leaders responsible for AI are discovering ‘AI pilot paradox’, where launching pilots is deceptively easy but deploying them into production is notoriously challenging.”

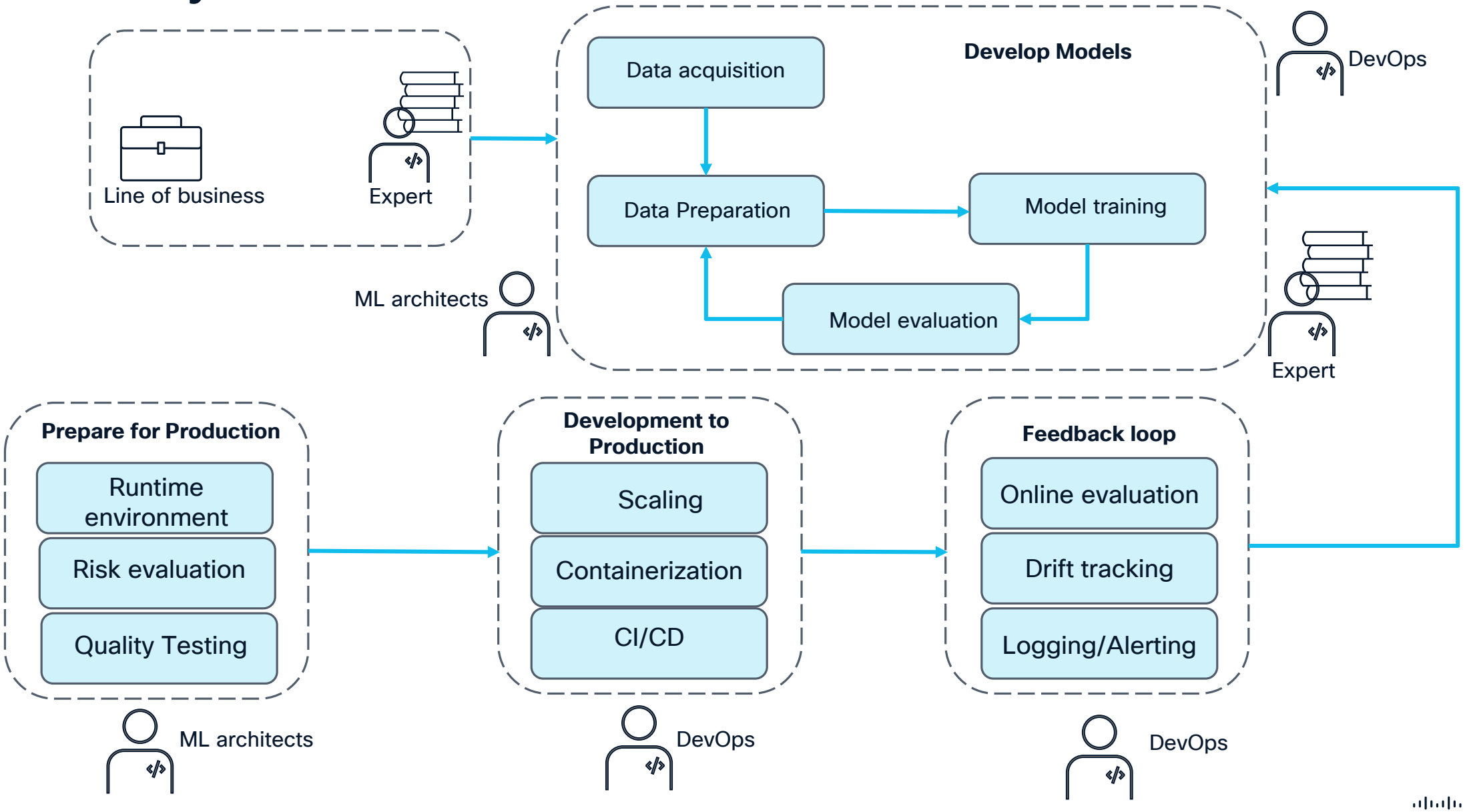
Publishing

Publishing ML Model and hosting the model is not enough

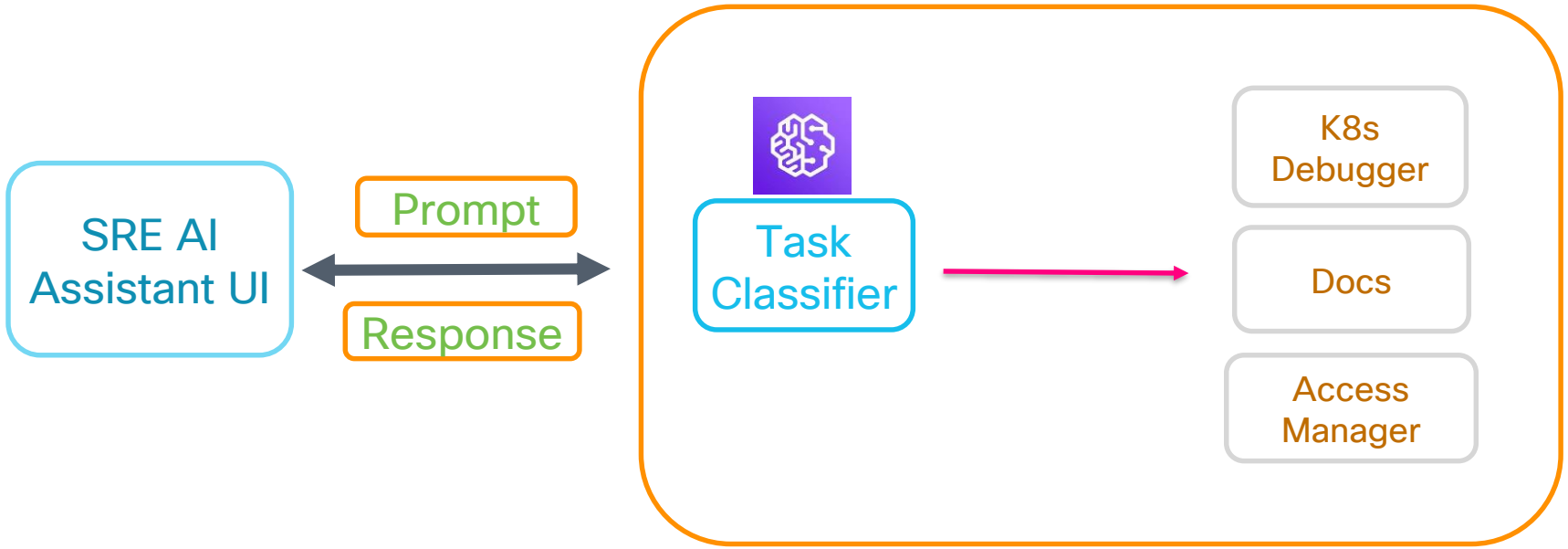
Managing

Managing the deployment, training, data curation is as equally important

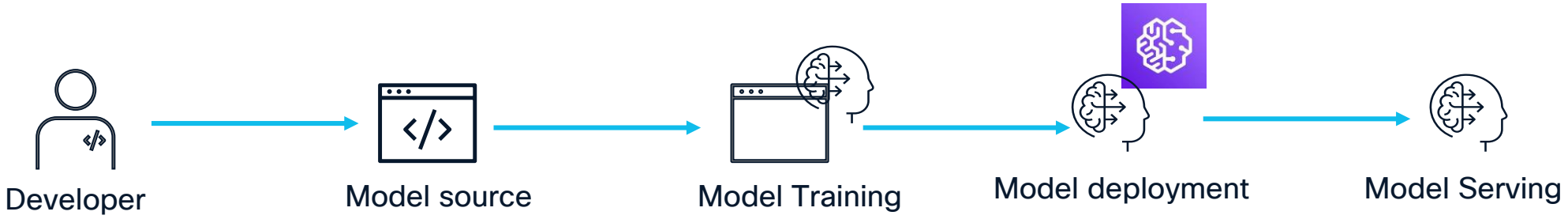
MLOps Lifecycle



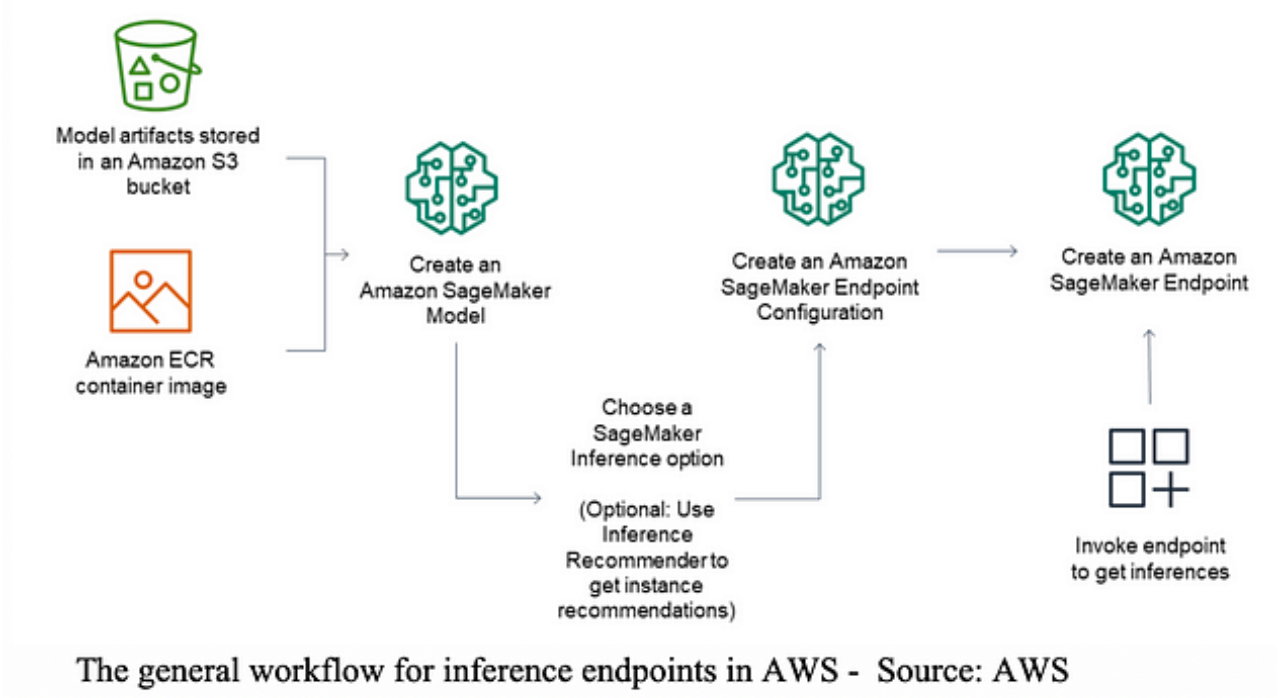
Demo Application



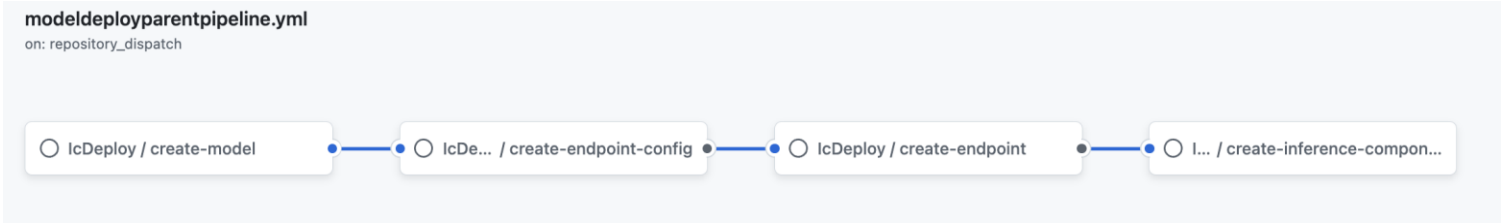
MLOps Pipeline



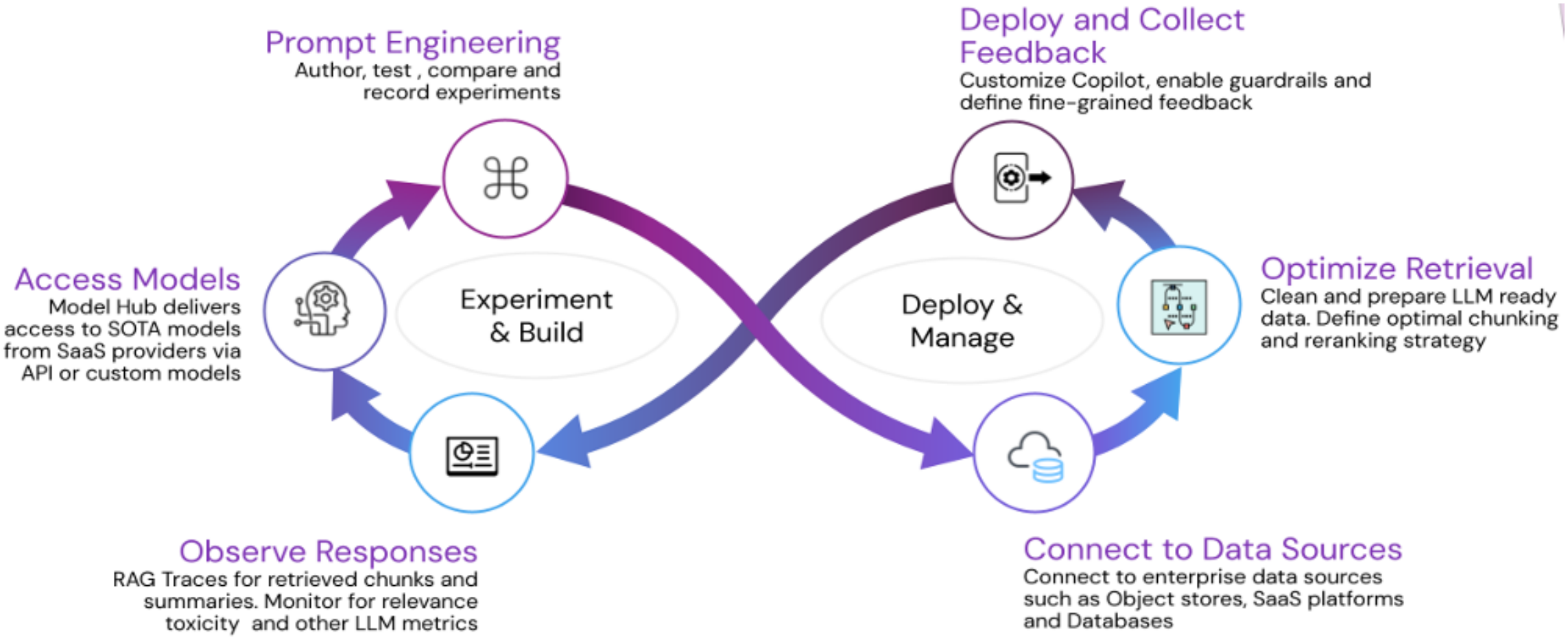
AWS SageMaker Deployment



The general workflow for inference endpoints in AWS - Source: AWS

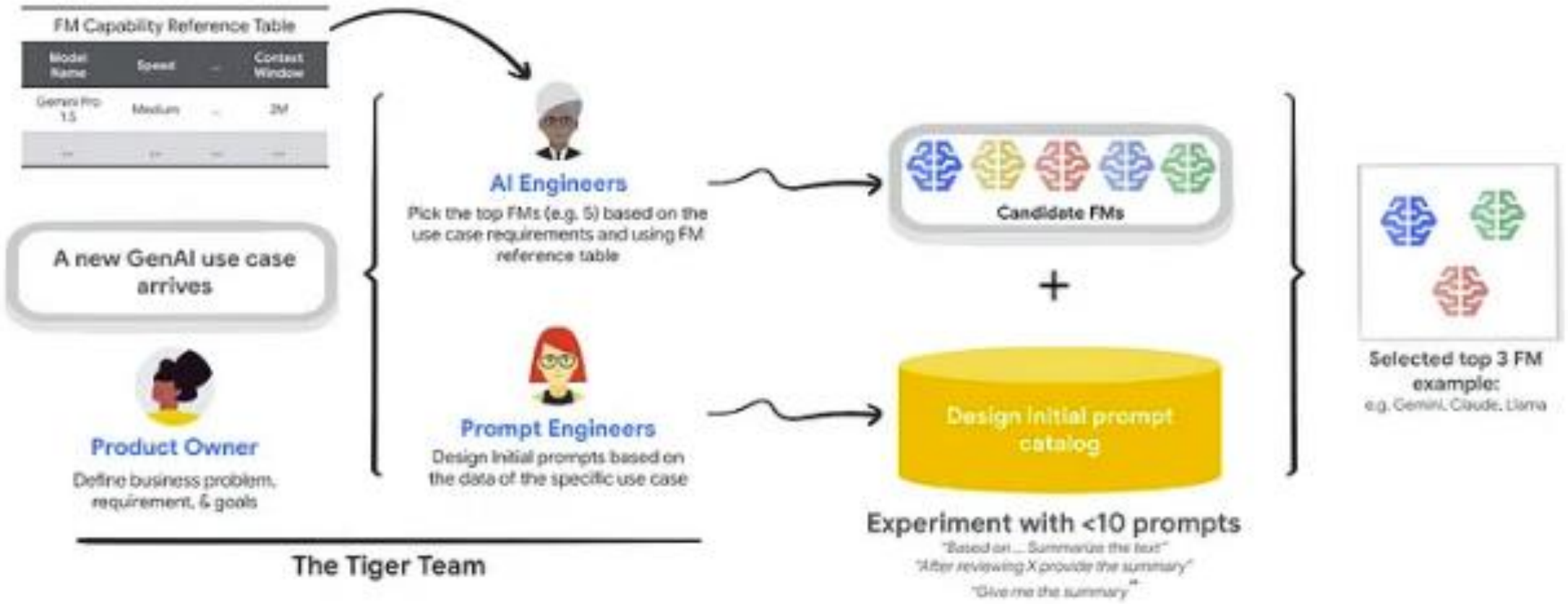


GenAIOps



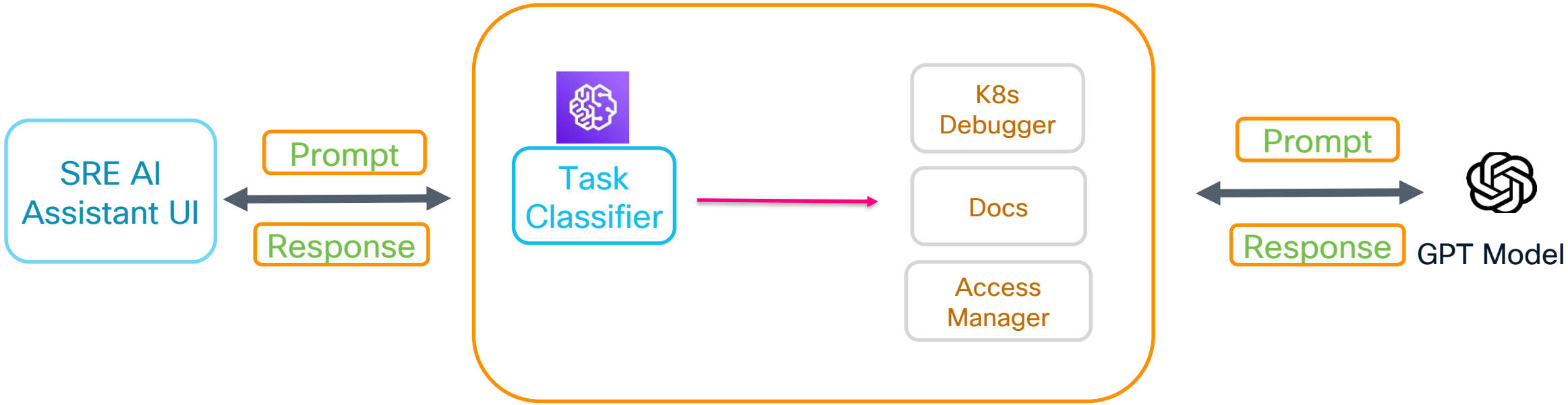
Ref: https://aws.amazon.com/blogs/machine-learning/migrating-to-amazon-sagemaker-karini-ai-cut-costs-by-23/

Prompt Engineering



Ref: https://medium.com/google-cloud/genaiops-operationalize-generative-ai-a-practical-guide-d5bedaa59d78

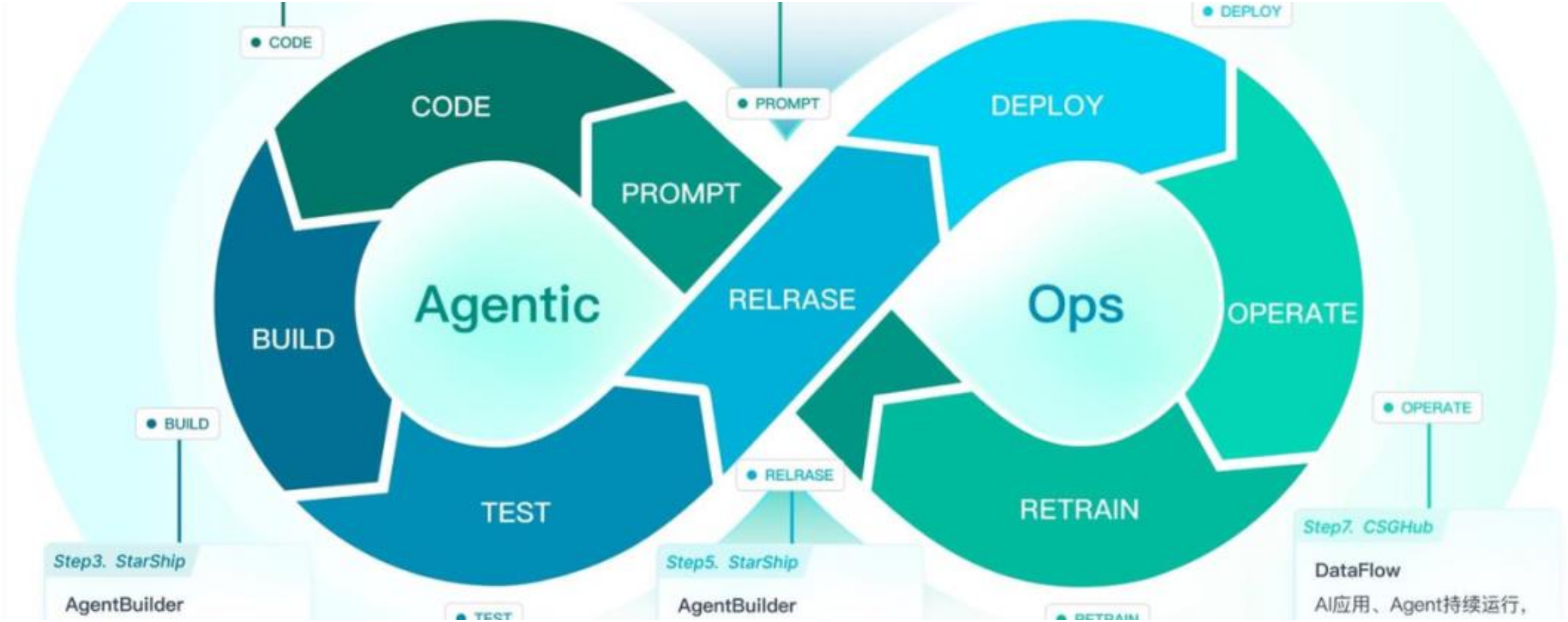
Demo Application



GenAI Reshaping OPs

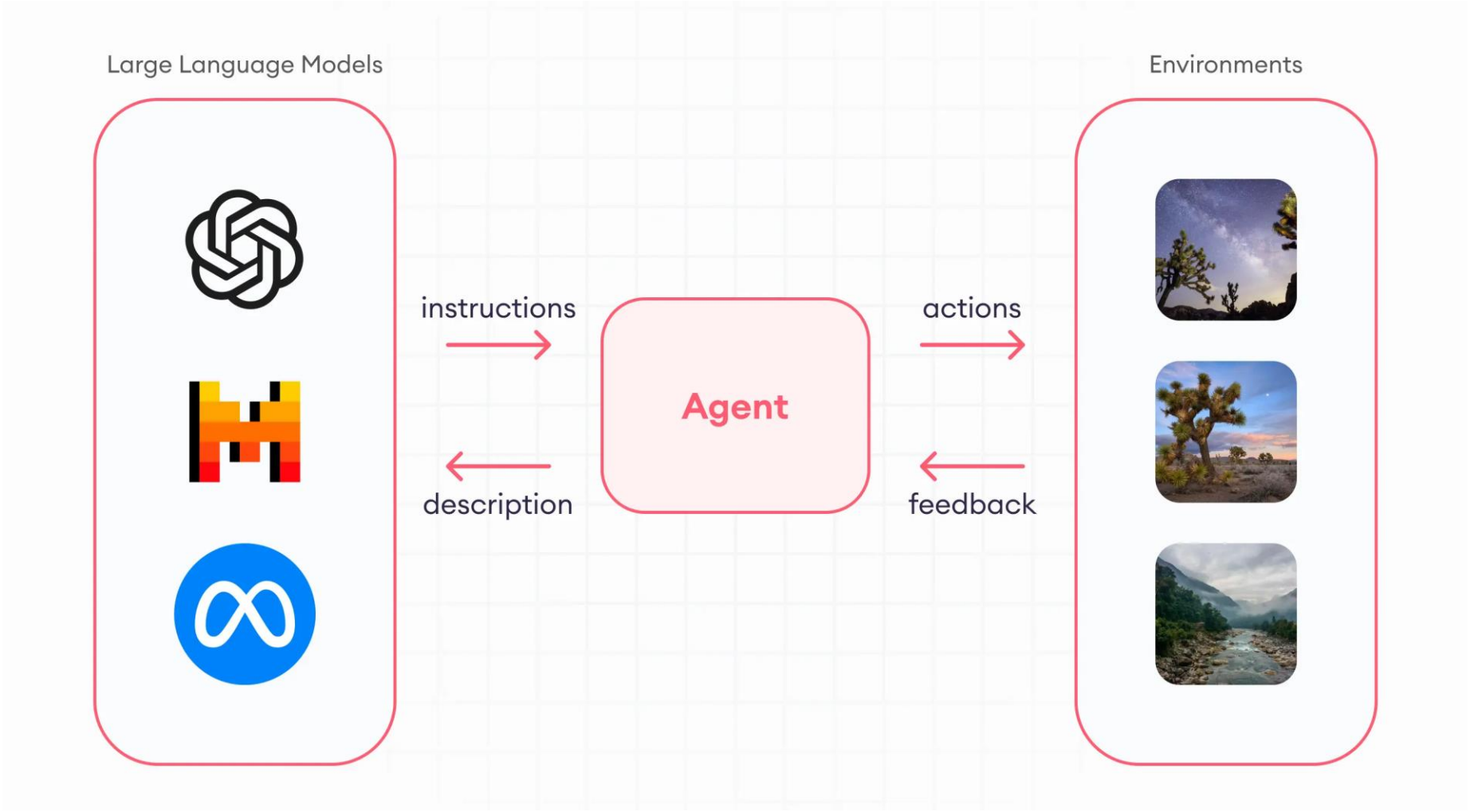
Ref: https://medium.com/google-cloud/genaiops-operationalize-generative-ai-a-practical-guide-d5bedaa59d78

AgenticOps

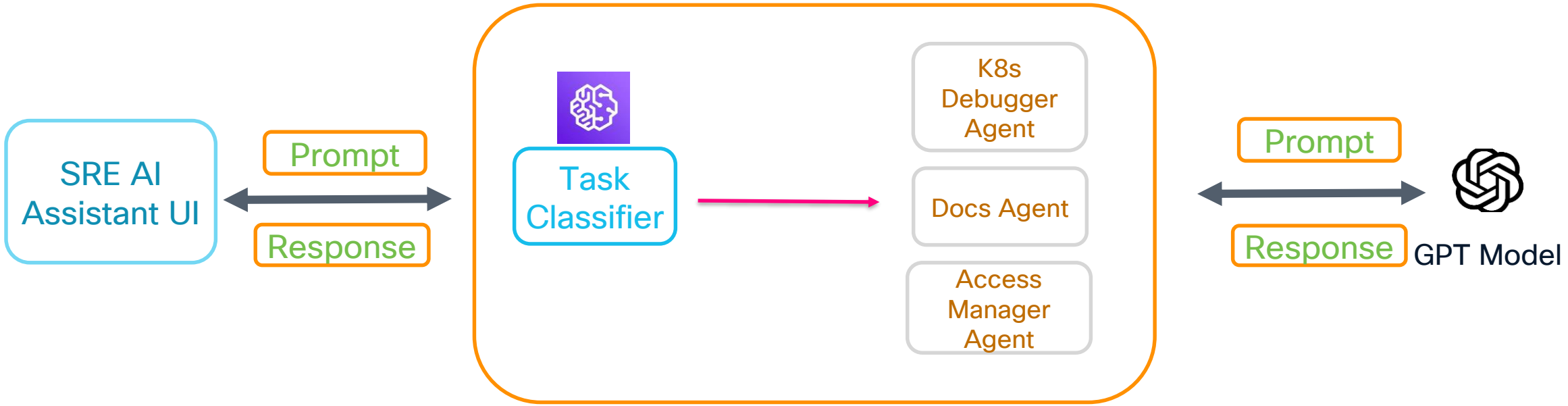


Ref: https://zhuanlan.zhihu.com/p/31620128944

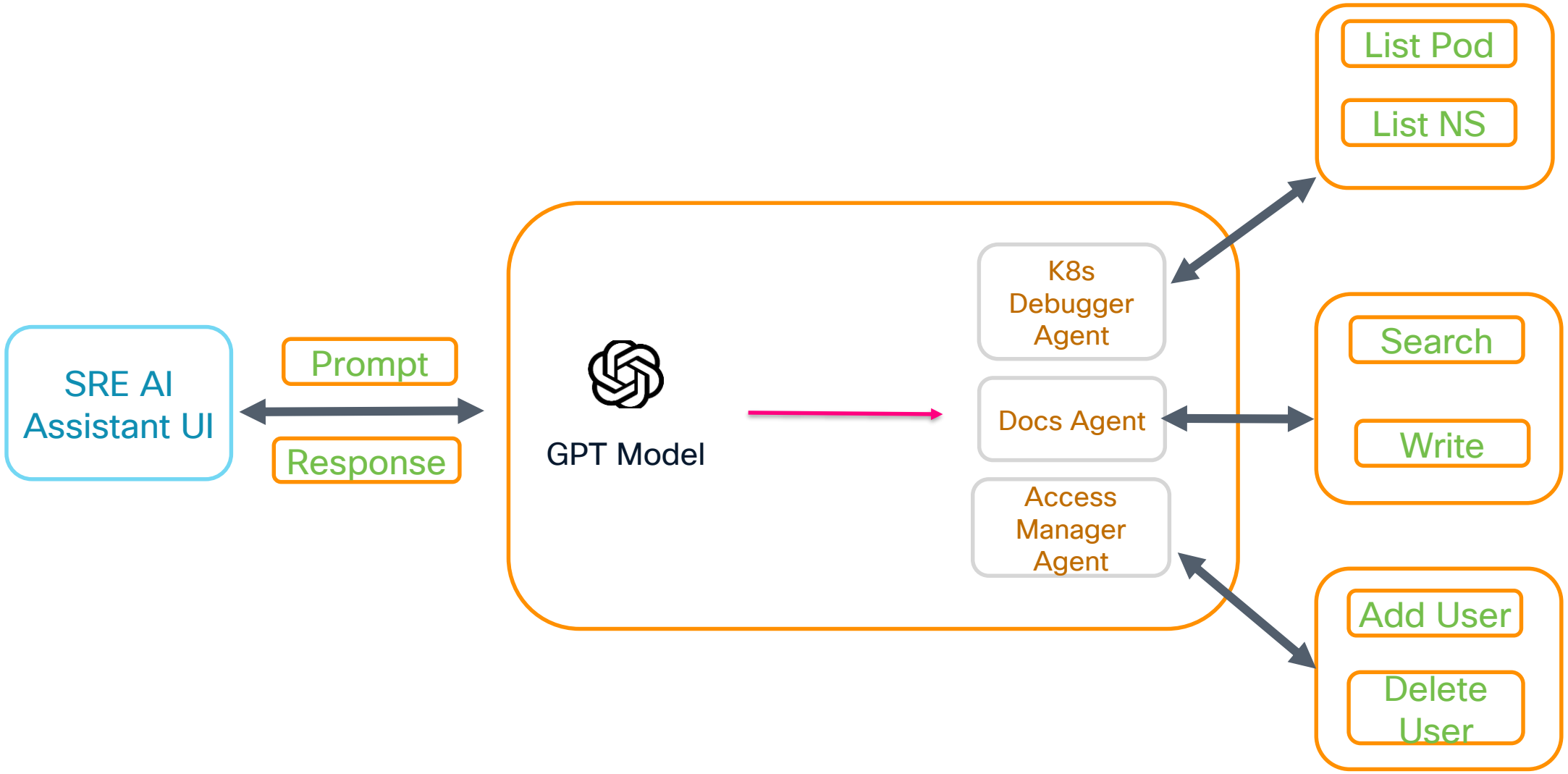
LLM Agent



Demo Application

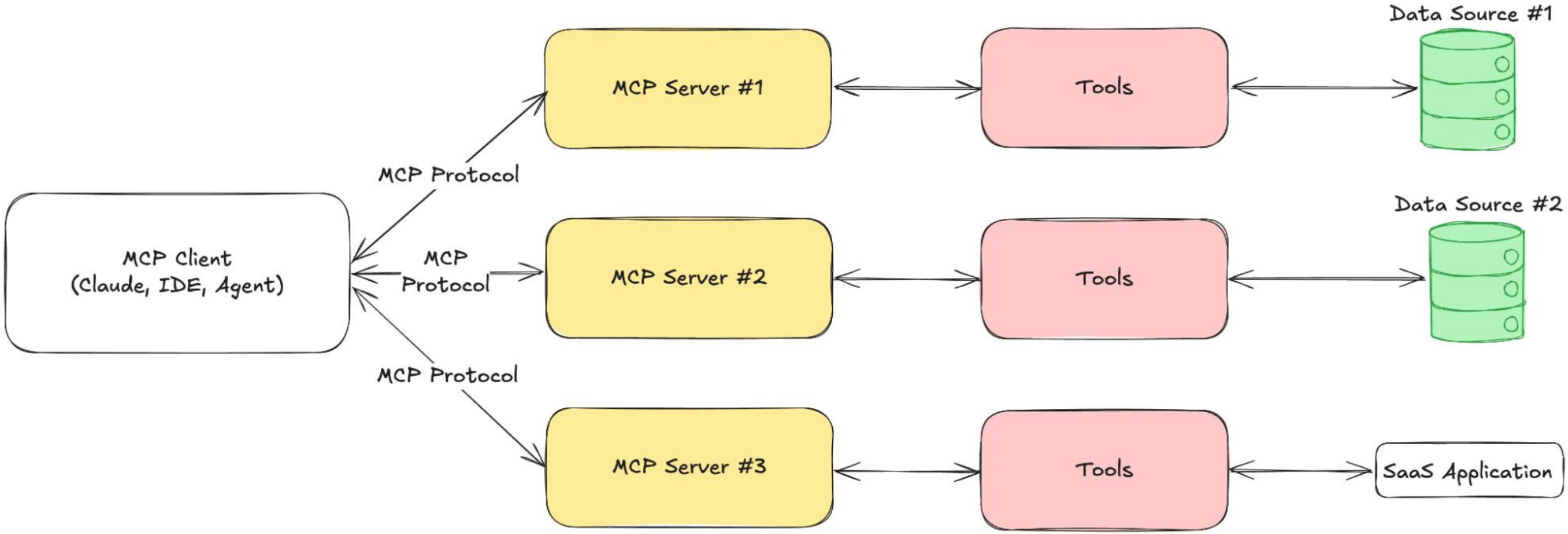


Demo Application



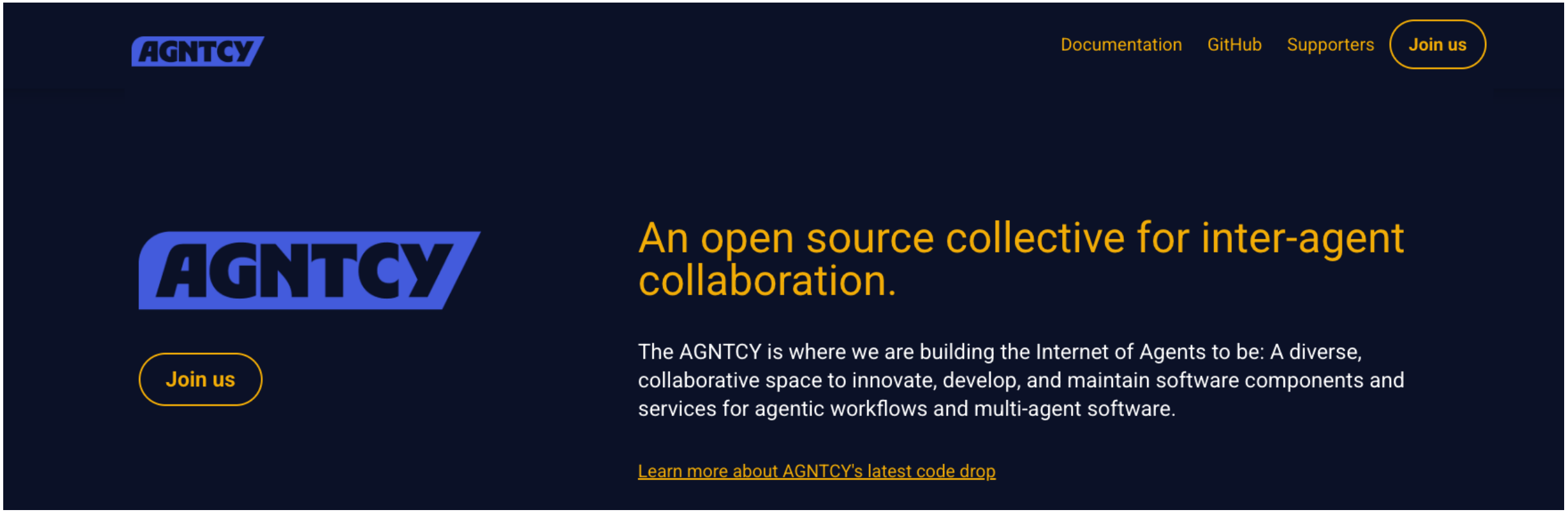
What all AI protocols have you heard about?

MCP

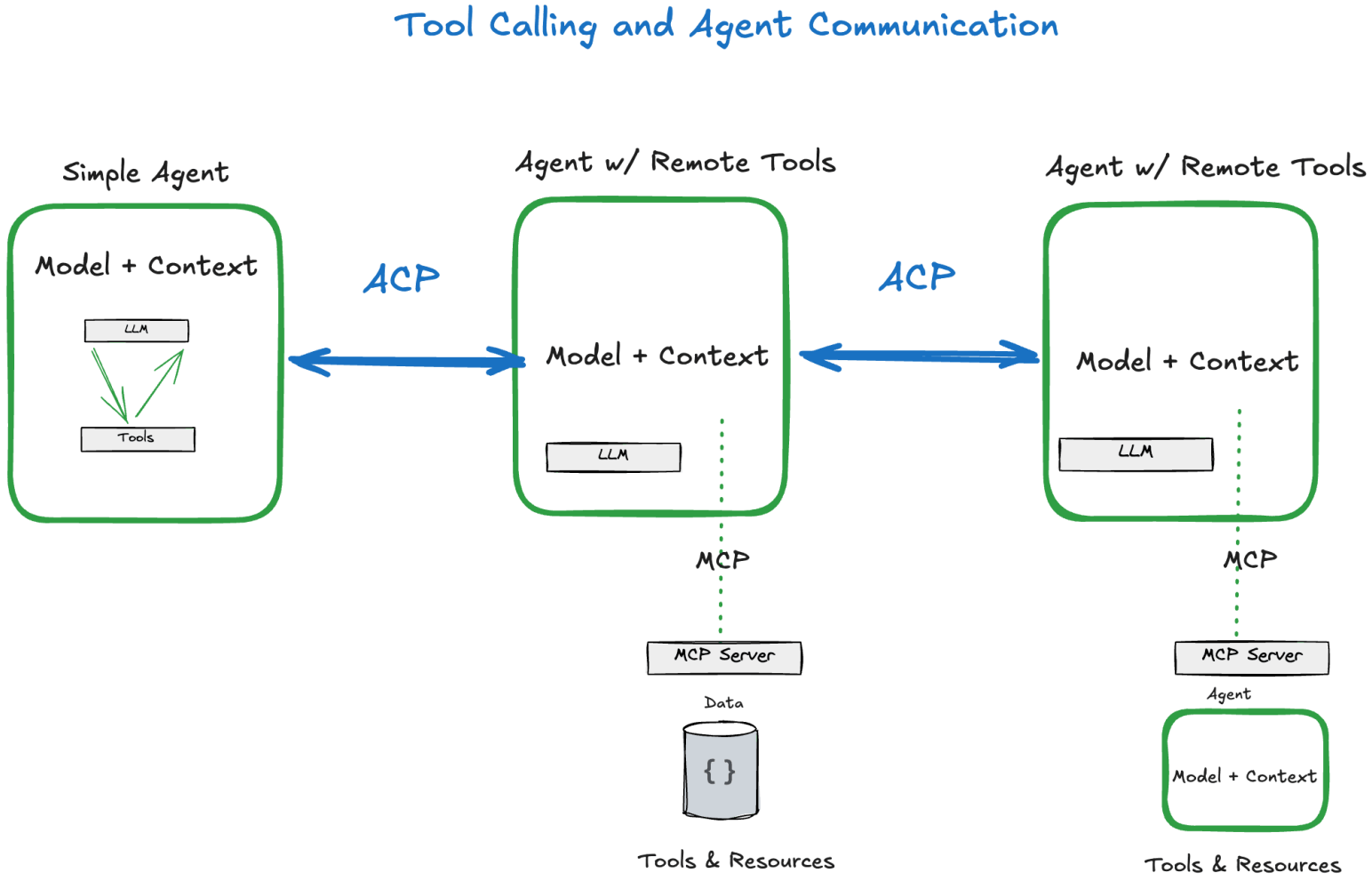


Fun Fact: What is the protocol developed by Cisco in LLM Agents space?

AGNTCY



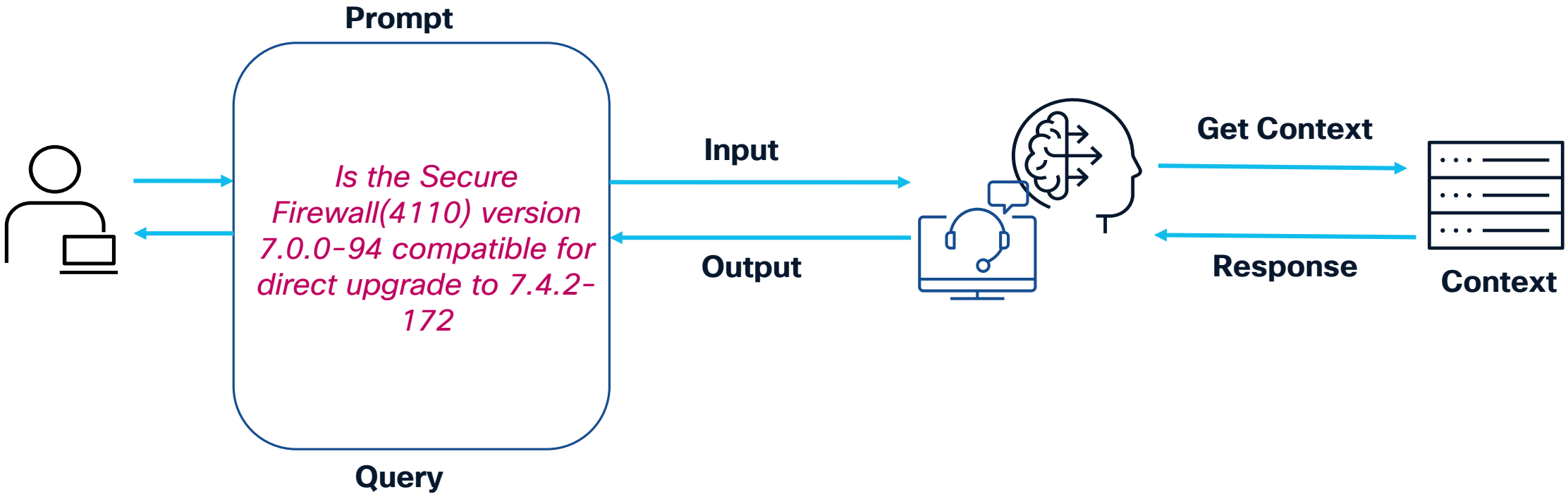
Agent Connect Protocol



Category	Focus	Sample Technologies	Level of Autonomy
MLOps	Operationalizing and managing ML models in production	MLflow, Kubeflow, Airflow, SageMaker	Low to Moderate (manual oversight)
AIOps	Automating IT operations using AI for monitoring and incident response	Dynatrace, Moogsoft, Splunk, IBM Watson AIOps	Moderate (decision support systems)
GenAIOps	Managing lifecycle and operations of generative AI models and systems	LangChain, LLMOps, Weights & Biases, Hugging Face Hub	High (dynamic tuning & adaptation)
AgenticOps	Orchestrating autonomous, goal-driven agents with memory and planning	AutoGPT, CrewAI, LangGraph, OpenDevin, AgentOS	Very High (autonomous behavior loops)

Sample LLM application

Enterprise based LLM application



Upgrade assistant which answers user queries for different Cisco products

Prompt Engineering

Model

```
compatibility_prompt = SystemMessage(f"""You are a TAC expert and your primary role is to solve compatibility queries for different Cisco Security products. For certain queries you might have to
                                         reference different tables which are present in a variable 'table_name' in tools. Do not generate table names.
                                         You need to retrieve tables from the tools and the tables would give the information in terms of matrices of compatibility for different product/versions.
                                         Read the table information and answer if the product/versions are compatible or not. If not provide a version that is compatible.
                                         Select the URL defined in list COMP_URL depending on the tool prefix.

    1) If the tool name starts with "fmc" kindly use the URL ending in "management-center-compatibility.html"
    2) If the tool name starts with "ftd" kindly use the URL ending in "threat-defense-compatibility.html"
    3) If the tool name starts with "asa" kindly use the URL ending in "asamatrix.html"
    4) If the tool name starts with "fxos" kindly use the URL ending in "fxos-compatibility.html"
    Below are few acronyms which you should be aware of
    FMC : Secure Firewall Management Center
    FTD : Secure Firewall Thread Defense
    FXOS: Firepower eXtensible Operating System
    ASA : Adaptive Security Appliance
    The query input would be given by the user
    The tools would give the context. Each tool has a description and try to map the query to the description provided. These would help in choosing the correct tool which would help in
    choosing the correct table for answering the query.
    The list of URL is defined in COMP_URL={COMP_URL}
    Return the response if the products are compatible to direct or step upgrade required.

""")
```

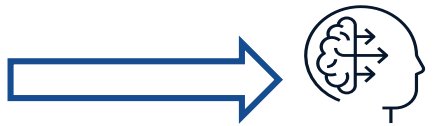
Knowledge Sources

Management Center Compatibility Guide → https://www.cisco.com/c/en/us/td/docs/security/secure-firewall/compatibility/management-center-compatibility.html

Secure Firewall Threat Defense Compatibility Guide → https://www.cisco.com/c/en/us/td/docs/security/secure-firewall/compatibility/threat-defense-compatibility.html

Firepower Compatibility Guide → https://www.cisco.com/c/en/us/td/docs/security/firepower/fxos/compatibility/fxos-compatibility.html

ASA Compatibility Guide → https://www.cisco.com/c/en/us/td/docs/security/asa/compatibility/asamatrx.html



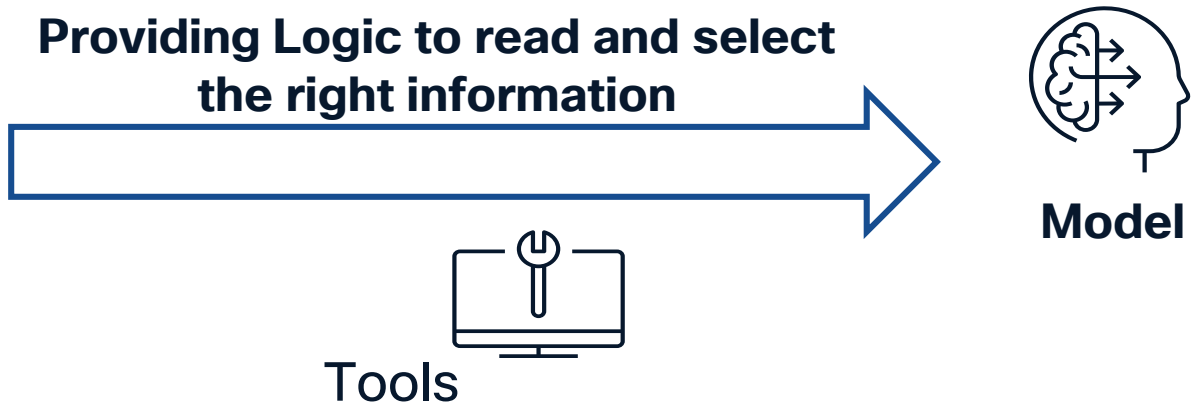
Model

Context

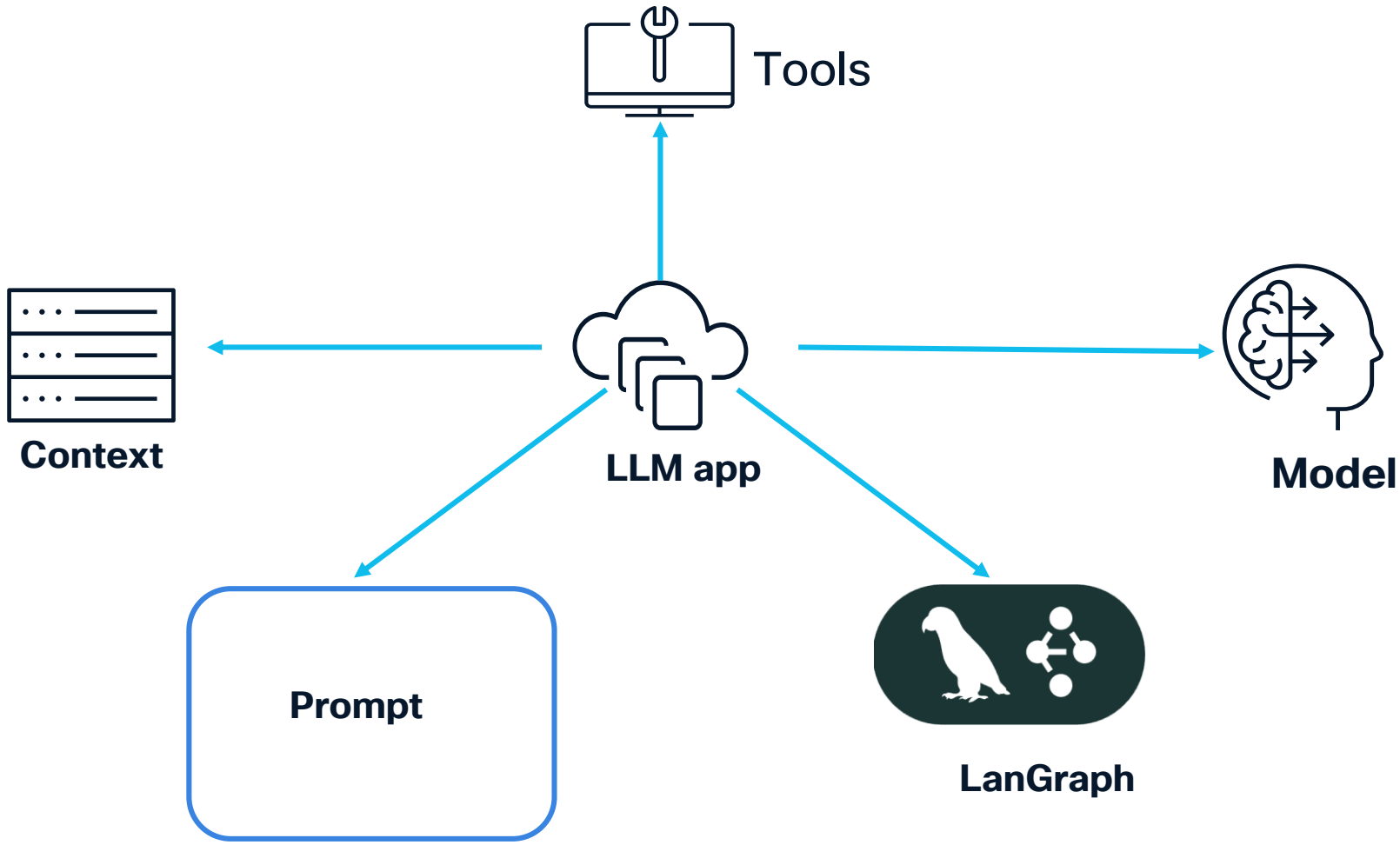
Functions

Table 9. On-Prem Management Center- Device Compatibility

Management Center Version	Oldest Device Version You Can Manage
7.7	7.2
7.6	7.1
7.4 Last support for NGIPS device management.	7.0
7.3	6.7
7.2	6.6
7.1	6.5
7.0	6.4
6.7	6.3
6.6	6.2.3
6.5	6.2.3
6.4	6.1
6.3	6.1
6.2.3	6.1
6.2.2	6.1
6.2.1	6.1
6.2	6.1
6.1	5.4.0.2/5.4.1.1
6.0.1	5.4.0.2/5.4.1.1
6.0	5.4.0.2/5.4.1.1
5.4.1	5.4.1 for ASA FirePOWER on the ASA-5506-X series, ASA5508-X, and ASA5516-X. 5.3.1 for ASA FirePOWER on the ASA5512-X, ASA5515-X, ASA5525-X, ASA5545-X, ASA5555-X, and ASA-5585-X series. 5.3.0 for Firepower 7000/8000 series and legacy devices.



Essential components of LLM Application



Demo

Live Demo

Complete your session evaluations

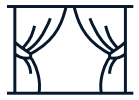
Complete a minimum of 4 session surveys and the Overall Event Survey to be entered in a drawing to win 1 of 5 full conference passes to Cisco Live 2026.

Earn 100 points per survey completed and compete on the Cisco Live Challenge leaderboard.

Level up and earn exclusive prizes!

Complete your surveys in the Cisco Live mobile app.

Continue your education



Visit the Cisco Showcase for related demos

Book your one-on-one Meet the Engineer meeting

Attend the interactive education with DevNet, Capture the Flag, and Walk-in Labs

Visit the On-Demand Library for more sessions at www.CiscoLive.com/on-demand

Contact me at nghodki@cisco.com

Thank you

cisco Live !


CISCO