29th Colloqium: CISSE
Cybersecurity Education in the Age of AI and Automation & Ambiguity

THREAT MODEL ON GOOGLE ADK AGENTS: AN OWASP AGENTIC SECURITY INITIATIVE PERSPECTIVE

Clark Ngo & Sam Chung
Smart and Secure Computing Research Group,
School of Technology & Computing
City University of Seattle

November 13, 2025 @ Seattle University

Current Job Market

Gemini

Company	WA Layoffs (2025)	Key Locations
Microsoft	Over 3,200 (May – July 2025, over 1,000 in SDE)	Redmond, etc.
Amazon	2,303 (October 2025, over 600 in SDE)	Seattle (1,887), Bellevue (416)
Total	Over 5,500 (over 1,600 in SDE)	Seattle Area

The Job Market in Seattle, Washington

- The Seattle metropolitan area added about 11,800 total jobs year-over-year through August 2025 (a modest 0.5% increase), with the primary growth coming from Healthcare, Clean Energy, and Trades, not the traditional tech sector.

- Tech employment in Washington fell 6% from mid-2022 to early 2025.

Current Situation

Gemini

Role Category	Growth Rate in Seattle (2025 Projections)	Context
Machine Learning Engineers	63% growth in job postings	One of the highest growth rates across all tech occupations in the region.
Data Scientists	48% growth in job postings	A significant increase, reflecting the need for professionals to manage and analyze data for AI models.
AI-Focused Positions (General)	**166% increase in job postings**	**Indicates the total volume of roles where AI skills (Applied Scientist, ML Engineer, AI SDE) are the primary requirement.**
Computer & Information Systems Managers	High Growth	This role, critical for managing enterprise data and security foundational to AI, is a fastest-growing tech talent role in the U.S.

“AI in X” vs. “X in AI”

Gemini

Feature	"AI in X" (e.g., AI in Full-Stack Development)	"X in AI" (e.g., Cybersecurity in AI)
Focus	Application of AI.	Influence on or Study of AI.
Direction	AI is the tool used to transform/optimize X.	X is the domain used to inform/shape AI.
Goal	Improve efficiency, solve problems, or automate tasks within X.	Enhance AI design, address ethical/social impact, or understand AI's mechanisms.
AI's Role	Enabler, technology, solution provider.	Object of study, system to be improved, or a technology with consequences to be managed.
Typical Questions	"How can AI optimize crop yields in Agriculture?"	"How does Cognitive Science inform deep learning architectures?"
Example Fields	AI in Medicine, AI in Finance, AI in Art.	Ethics in AI, Law in AI, Fairness in AI, HCI (Human Computer Interaction) in AI.
Computer & Information Systems Managers	High Growth	This role, critical for managing enterprise data and security foundational to AI, is a fastest-growing tech talent role in the U.S.

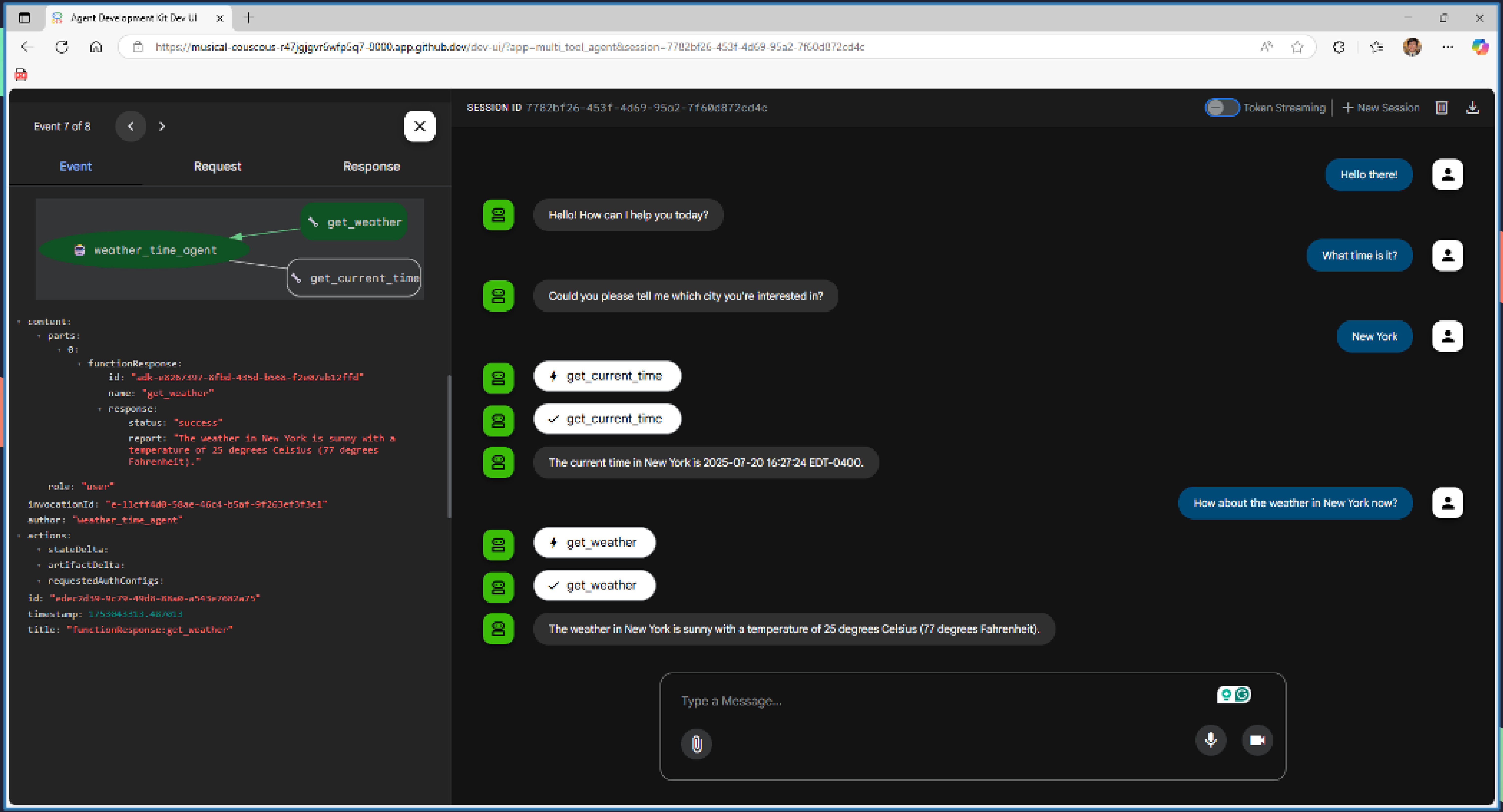
AI-enabled vs. AI-infused vs. AI-powered

Aspect	AI-Enabled	AI-Infused	AI-Powered
Core Meaning	Has AI as an add-on feature	Has AI deeply integrated into multiple components	Has AI as its core engine or driver
Role of AI	Supplementary feature; enhances a specific task	Pervasive enhancer; improves core functionality and UX	Fundamental driver; enables the primary function
Level of Integration	Shallow. A single feature or module	Deep. Woven into the system's architecture	Foundational. The system is built around the AI
Dependency	Low. The main product works without it	Medium-High. The product is significantly less useful without it	Absolute. The product cannot function without AI
Technologies	AI assistants (Gemini, ChatGPT, Copilot, etc.)	Vibe Coding, RAG (Retrieval-Augmented Generation), MCP (Model Context Protocol)	ADK (Agent Development Kit), A2A (Agent-to-Agent Protocol)

Examples of Google ADK Agents

- Brandon Hancock's ADK Crash Course
- Brandon Hancock's ADK Masterclass: Build AI Agents & Automate Workflows (Beginner to Pro) (3:12:47)
- Kruchten’s 4+1 view architectural models in UML for 12 use cases:
  - A Basic AI Agent, AI Agents with Tools, Lite LLM and Open Router,
  - Structured Outputs, Sessions and State Management, Persistent Storage
  - Multi-agent, Stateful Multi-agent, Callbacks
  - Sequential agent, parallel Agent, Loop Agent

An AI Agent with Tools


Agent Development Kit Dev UI
https://musical-couscous-r47jgjgvr5wfp5q7-8000.app.github.dev/dev-ui/?app=multi_tool_agent&session=7782bf26-453f-4d69-95a2-7f60d872cd4c
Event 7 of 8
Event
Request
Response
get_weather
weather_time_agent
get_current_time
SESSION ID 7782bf26-453f-4d69-95a2-7f60d872cd4c
Token Streaming
+ New Session
Type a Message...

```
content:
  parts:
    0:
      functionResponse:
        id: "adk-e8267397-8fbd-435d-b568-f2e07eb12ffd"
        name: "get_weather"
        response:
          status: "success"
          report: "The weather in New York is sunny with a
          temperature of 25 degrees Celsius (77 degrees
          Fahrenheit)."

  role: "user"
invocationId: "e-11cff4d0-58ae-46c4-b5af-9f263ef3f3e1"
author: "weather_time_agent"
actions:
  stateDelta:
  artifactDelta:
  requestedAuthConfigs:
id: "eder7d39-9c79-49d8-88a0-a543e7682a75"
timestamp: 1753043313.487013
title: "functionResponse:get_weather"
```

Hello there!

Hello! How can I help you today?

What time is it?

Could you please tell me which city you're interested in?

New York

⚡ get_current_time

✓ get_current_time

The current time in New York is 2025-07-20 16:27:24 EDT-0400.

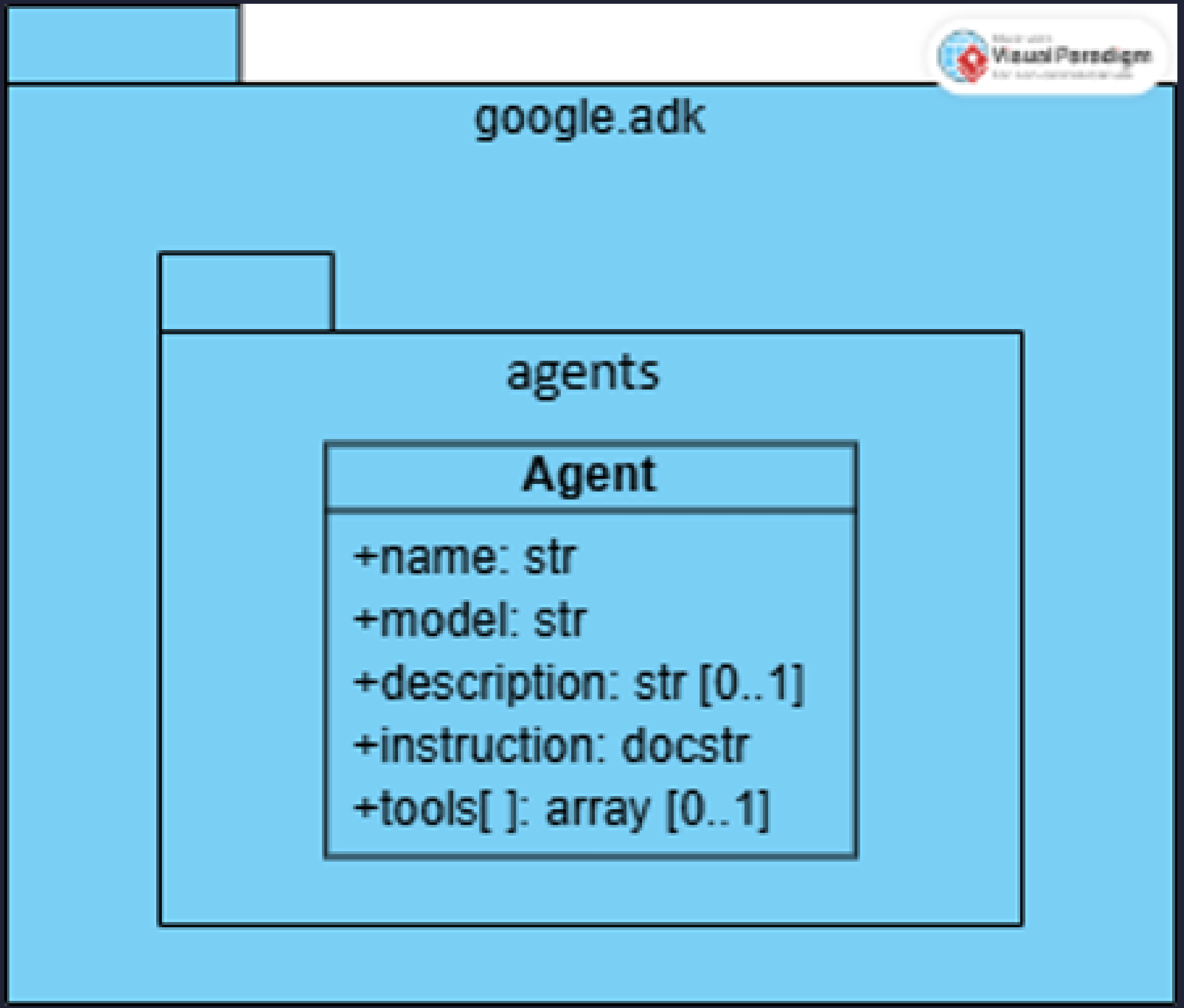
How about the weather in New York now?

⚡ get_weather

✓ get_weather

The weather in New York is sunny with a temperature of 25 degrees Celsius (77 degrees Fahrenheit).

Architecture – Logical Design View

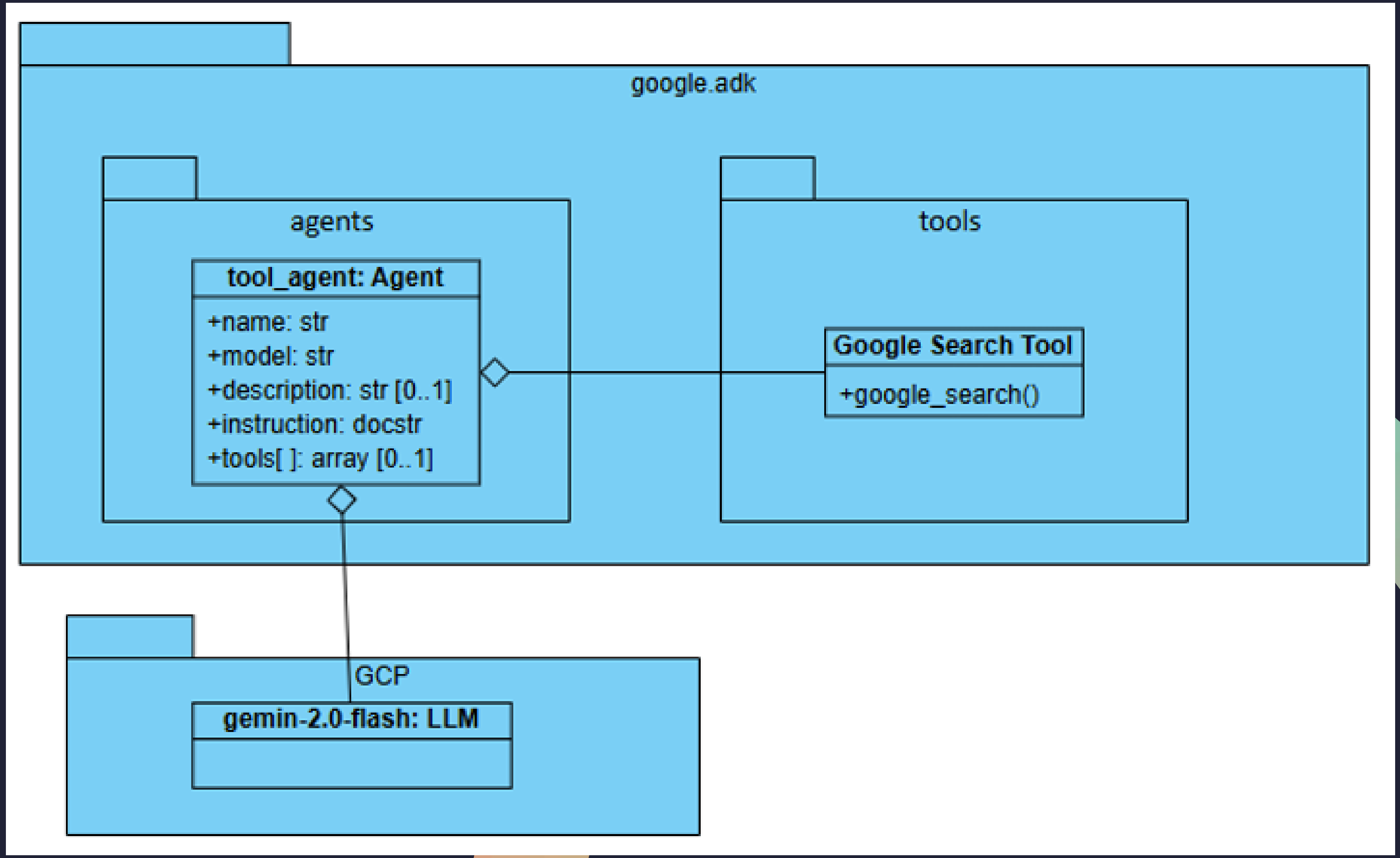


google.adk

agents

Agent

+name: str
+model: str
+description: str [0..1]
+instruction: docstr
+tools[ ]: array [0..1]



google.adk

agents

tool_agent: Agent

+name: str
+model: str
+description: str [0..1]
+instruction: docstr
+tools[ ]: array [0..1]

tools

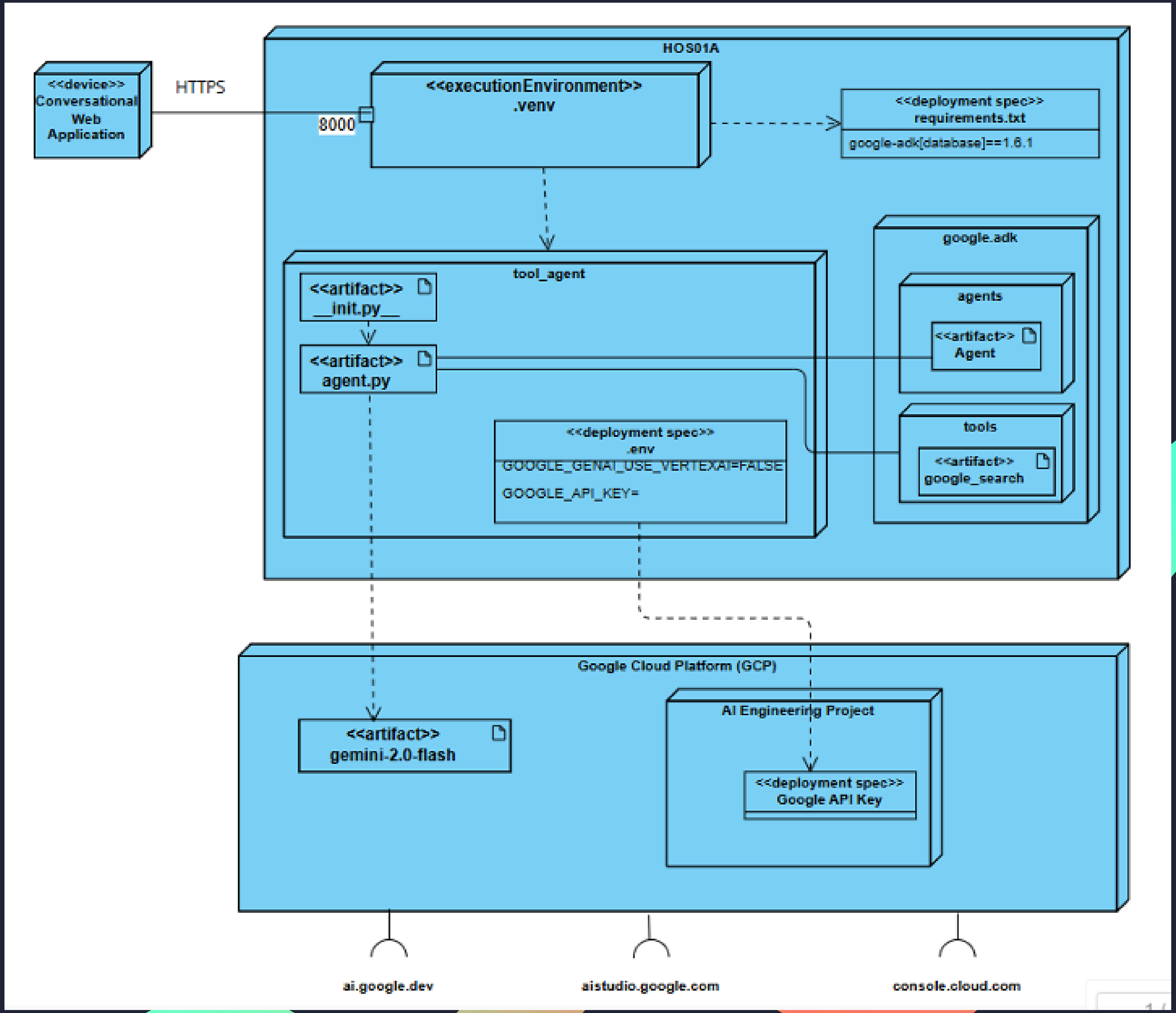
Google Search Tool

+google_search()

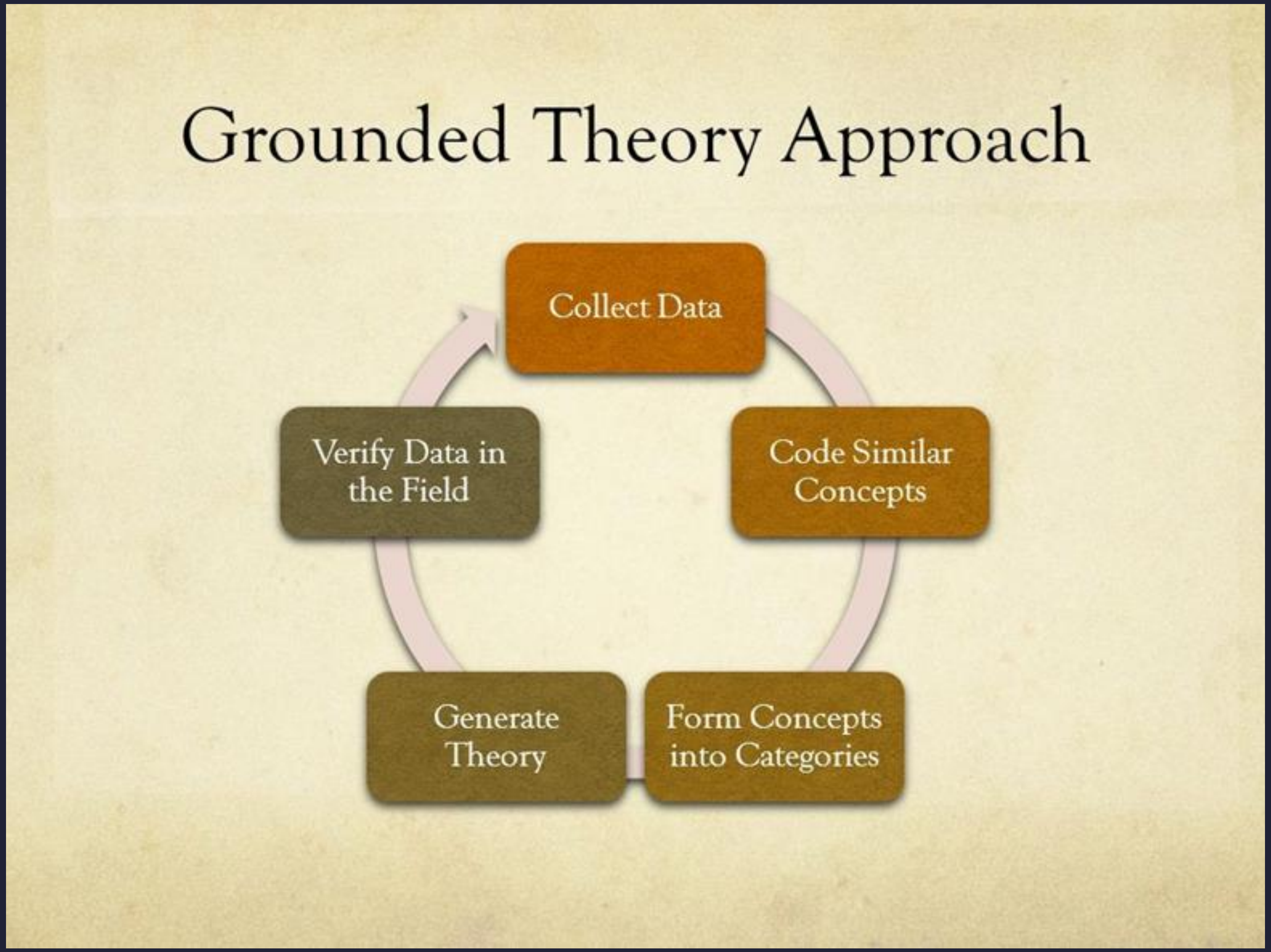
GCP

gemin-2.0-flash: LLM

Architecture – Deployment View



Grounded Theory



Grounded Theory Approach

Collect Data

Code Similar Concepts

Form Concepts into Categories

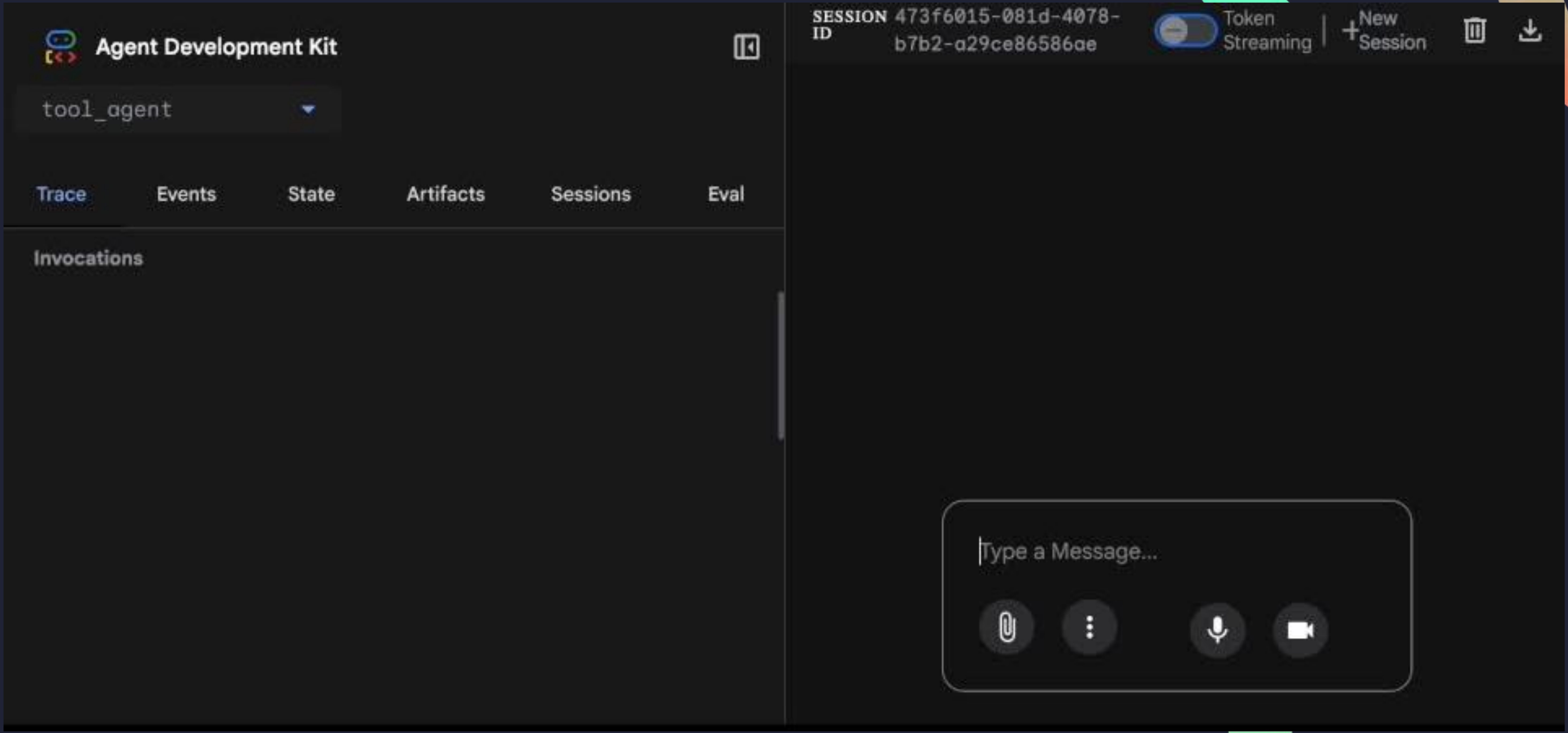
Generate Theory

Verify Data in the Field

- From coding to system architectures
- The n-tier architectures
- Microservices

Demo: Tool Agent

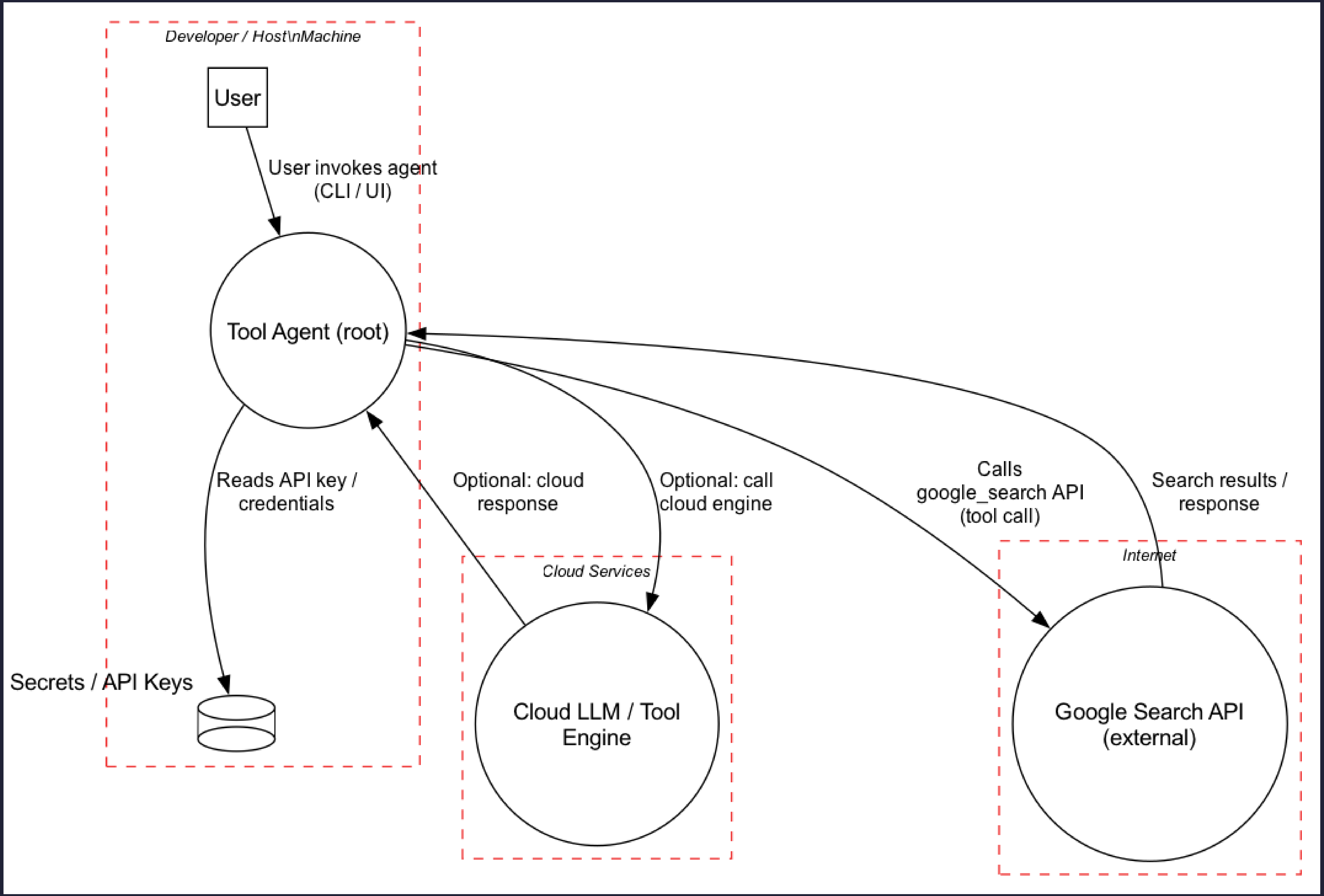
AGENT WITH TOOL FUNCTIONALITY OF GOOGLE SEARCH



SESSION ID 473f6015-081d-4078-b7b2-a29ce86586ae

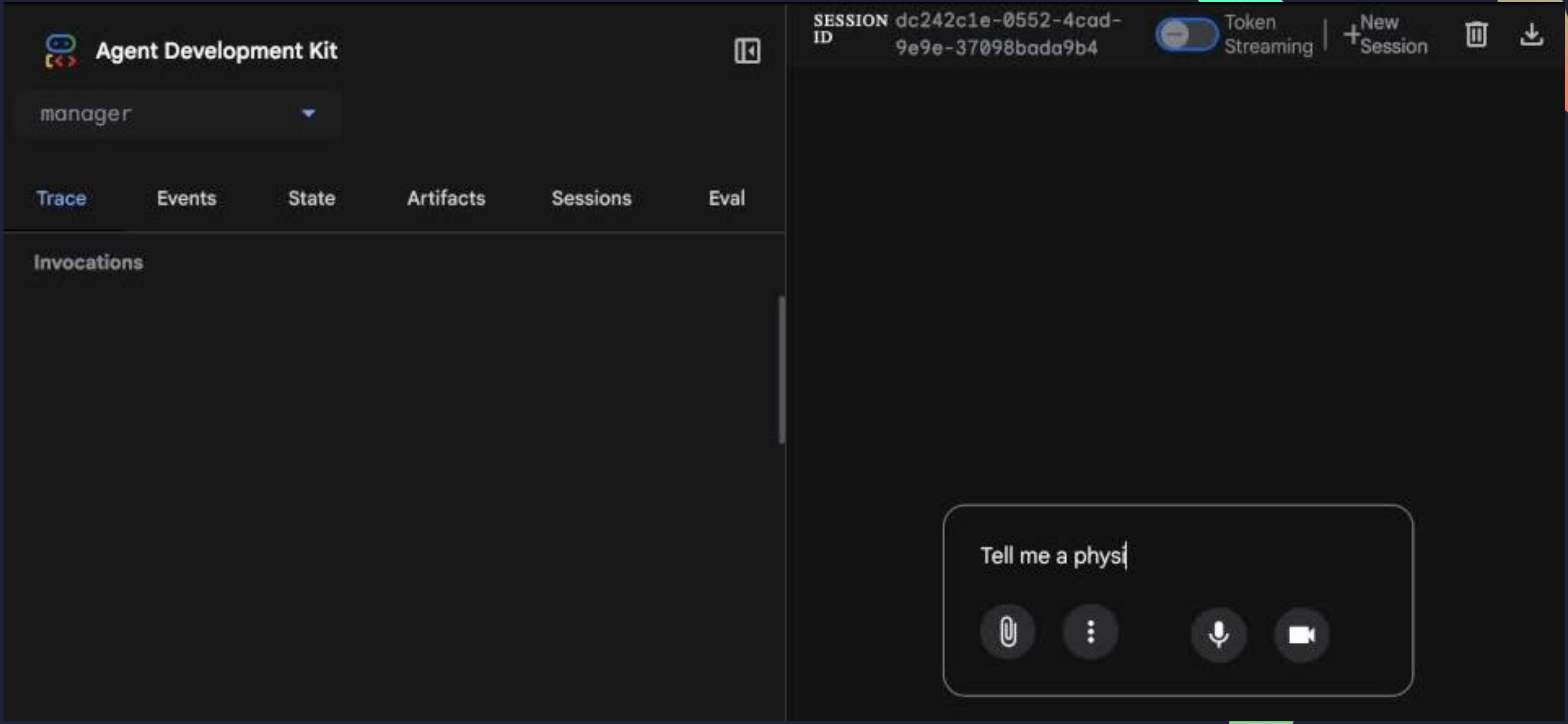
TOOL AGENT

With Google Search API
Data Flow Diagram


Developer / Host\nMachine
User
User invokes agent (CLI / UI)
Tool Agent (root)
Reads API key / credentials
Secrets / API Keys
Optional: cloud response
Optional: call cloud engine
Cloud Services
Cloud LLM / Tool Engine
Calls google_search API (tool call)
Search results / response
Internet
Google Search API (external)

Demo: Multi Agent

SUPERVISOR AGENT WITH SUB-AGENTS: FUNNY NERD AGENT, STOCK PRICE



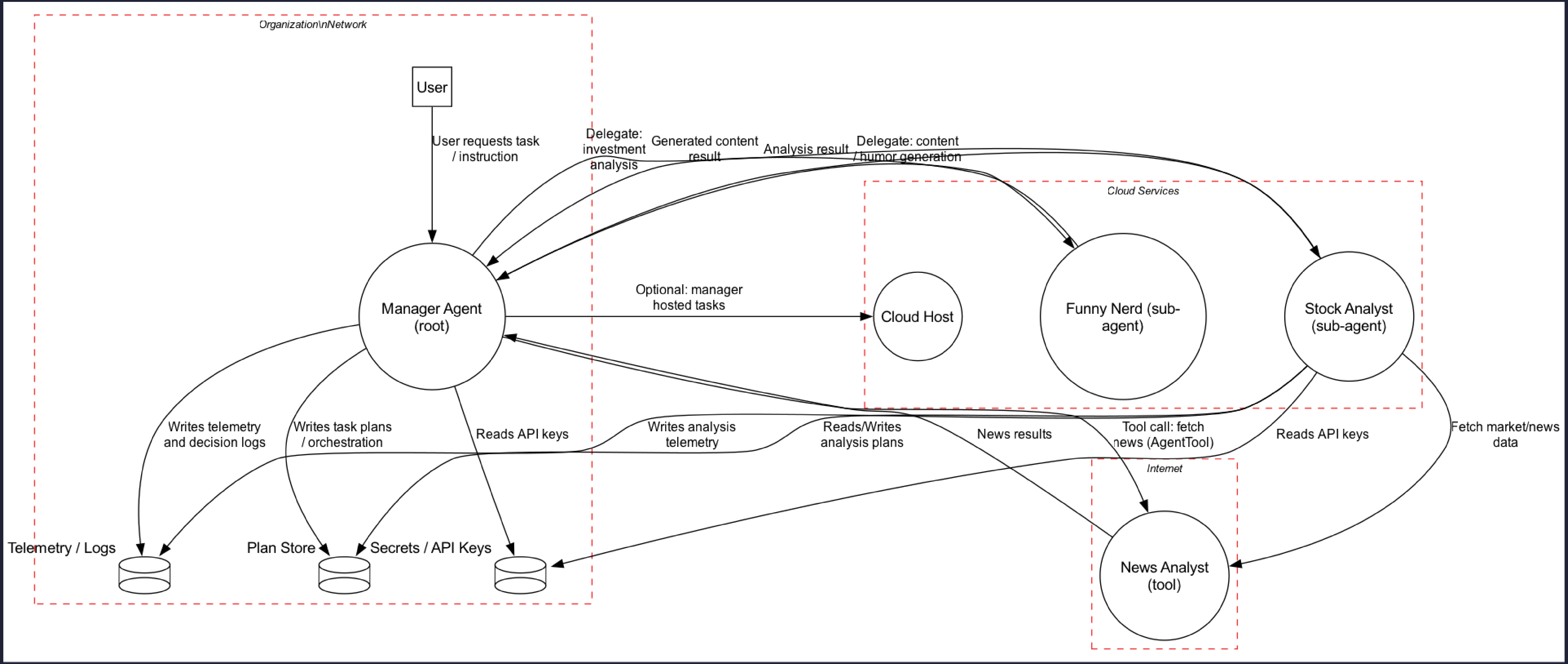
manager

SESSION ID dc242c1e-0552-4cad-9e9e-37098bada9b4

Tell me a physi

MULTI AGENT

Supervisor Agent: Manager | Sub-agents: Funny Nerd, Stock Analyst, News Analyst Data Flow Diagram


Organization\nNetwork
User
User requests task / instruction
Manager Agent (root)
Delegate: investment analysis
Generated content result
Analysis result
Delegate: content / humor generation
Cloud Services
Optional: manager hosted tasks
Cloud Host
Funny Nerd (sub-agent)
Stock Analyst (sub-agent)
Writes telemetry and decision logs
Writes task plans / orchestration
Reads API keys
Writes analysis telemetry
Reads/Writes analysis plans
News results
Tool call: fetch news (AgentTool)
Reads API keys
Fetch market/news data
Internet
Telemetry / Logs
Plan Store
Secrets / API Keys
News Analyst (tool)

Core Concept & Themes

Securing Autonomous AI

Applying the OWASP threat model to Google's ADK is essential for building responsible AI. It's the critical intersection of AI-driven agency, new security risks, and developer education.

The AI (Google ADK): A New Level of Power

Google's ADK isn't for chatbots; it's for autonomous agents. These agents have "agency"—the power to independently reason, plan, and use tools to take real-world actions.

The Security (OWASP): A New Class of Threats

This autonomy creates a massive new attack surface. The OWASP initiative provides the essential security framework for agent-specific threats like prompt injection, tool misuse, and excessive agency.

The Education (The Process): Building a Security-First Culture

The act of threat modeling is a critical educational tool. It teaches developers to think like attackers and helps the entire organization understand, govern, and deploy autonomous AI safely.

Why it matters

This framework is how we build trustworthy and responsible AI. It shifts the focus from "what can AI do?" to "how should AI do it securely?"

Threat Modeling

Threat modeling is a structured process to identify, quantify, and address potential security threats and vulnerabilities in a system.

Why is it Essential

1. Proactive Security
2. Risk-Based Decisions
3. Cost-Effective
4. Fosters Security Culture

The 4 Step Process

1. Deconstruct: "What are we building?"
2. Identify Threats: "What can go wrong?"
3. Mitigate: "What will we do about it?"
4. Validate: "Did we do it right?"

STRIDE for AI Agents

We've identified 6 scenarios where agents can fail.

Then, explore the concrete mitigation strategies for each.

Our focus is on threats unique to tool-calls and multi-agent interactions.

S: Spoofing

Scenario: An attacker agent impersonates a trusted **AuthAgent** to trick the **Root Agent** into performing an unauthorized action.

Mitigation Strategies:

- Implement mTLS (Mutual TLS)
- Use Signed Tokens (JWTs)
- Assume Zero Trust

T: Tampering

Scenario: A MitM attacker modifies a bank's API response, changing a user's balance from $1,000 to $10 before the agent sees it.

Mitigation Strategies:

- Enforce Strict TLS (1.2+)
- Verify Response Signatures
- Validate Data Schema

R: Repudiation

Scenario: A user denies giving a command to sell stock after the market changes, and the agent's vague logs can't prove the original instruction.

Mitigation Strategies

- Create an Immutable Audit Trail
- Require Explicit Confirmation
- Use Secure Timestamps

I: Information Disclosure

Scenario: A **Root Agent** passes a sensitive query (with PII) to a generic **ResearchAgent**, which then logs it to a public-facing database.

Mitigation Strategies

- Scrub PII Before Delegation
- Implement Contextual Logging
- Apply Data Masking by Default

D: Denial of Service

Scenario: A malicious user tricks an agent into calling the **send_email** tool 1,000 times in a loop, exhausting the API quota for all users.

Mitigation Strategies

- Enforce Per-User Rate Limits
- Implement "Cost" Analysis
- Require Pre-Execution Review

E: Elevation of Privilege

Scenario: A user prompt-injects a call to an admin-only tool, **delete_user('bob')**, and the LLM, trying to be helpful, obeys the command.

Mitigation Strategies

- Enforce ACLs at the *Tool Level*
- Use Dynamic Tool Scoping
- Treat the LLM as Untrusted

Primary Threats to for apps built with Google ADK

💉 Prompt Injection

A user provides a specially crafted prompt that tricks the LLM into bypassing its safety instructions or performing an unintended action.

⚠️ Improper Output Handling

The agent generates a string of text that is actually malicious data, like a piece of code or a database query. The application then executes this data, believing it's a safe output.

🐙 Excessive Agency

The agent is given too much power or an excessive number of permissions through its tools, and it uses that power in an unintended or harmful way.

STRIDE analysis for Google ADK

🕵️ Spoofing

User Spoofing: Attacker impersonates a User to send malicious prompts.

Agent Spoofing: Attacker creates a fake Agent to trick a User into connecting and stealing their prompts.

LLM Spoofing: Attacker impersonates Vertex AI via a Man-in-the-Middle (MitM) attack to intercept or modify API requests.

✍️ Tampering

Prompt Tampering: Attacker modifies the "User prompt" dataflow in transit

Request Tampering (Prompt Injection): A malicious User sends a crafted prompt to tamper with the Agent's logic

Response Tampering: Attacker modifies the "LLM response" or Agent Response

🚫 Repudiation

User Repudiation: A User denies sending a malicious prompt, which is possible if the GreetingAgent lacks sufficient authentication and logging.

Agent Repudiation: The GreetingAgent's actions (like logging or API calls) cannot be verified if its internal logs are missing or compromised.

STRIDE analysis for Google ADK

ℹ️ Information Disclosure

Data in Transit: Any dataflow (User prompt, LLM request, etc.) could be intercepted by a network attacker if TLS is weak or misconfigured.

Agent Data Leakage: The Agent could leak internal data via verbose error messages or expose one user's data to another.

LLM Data Leakage: The Vertex AI model could be tricked via prompt injection into revealing its system prompt or other sensitive, non-public data.

🛑 Denial of Service (DoS)

Agent DoS: The Agent server could be targeted by a network flood.

API Quota Exhaustion: A malicious User or attacker could exploit the agent to rapidly send requests to Vertex AI, burning through the API quota and taking the service offline.

⬆️ Elevation of Privilege (EoP)

Prompt Injection for EoP: A malicious User sends a prompt to bypass security controls or trick the agent into executing functions it shouldn't (e.g., accessing internal tools or data).

Agent Compromise: A vulnerability in the Agent's code (e.g., unsafe parsing of the LLM response) is exploited, allowing an attacker to gain control of the server.

Analysis

- Google's ADK is designed to build autonomous agents, not just chatbots. These agents possess "agency"—the power to independently reason, plan, and use tools to take real-world actions.

- This autonomy introduces a massive new attack surface that traditional security models do not fully cover.

- The STRIDE threat model proves to be a highly effective framework for deconstructing ADK agents

Conclusion

- The act of threat modeling is a **critical educational tool**.
- Applying the OWASP framework to Google's ADK is essential for **building responsible and trustworthy AI**.
- This security-focused approach shifts the fundamental development question from "What can AI do?" to "**How should AI do it securely?**".

Conclusion (Continues)

AI in Full-Stack Dev	Cybersecurity in AI
Google ADK	OWASP Framework
Kruchen’s 4+1 Architectural Views – UML Diagrams	Threat Modeling – Data Flow Diagrams
Grounded Theory	STRIDE Analysis

Best Practices for Secure AI-Powered Full-Stack Web Application / Mobile App Development

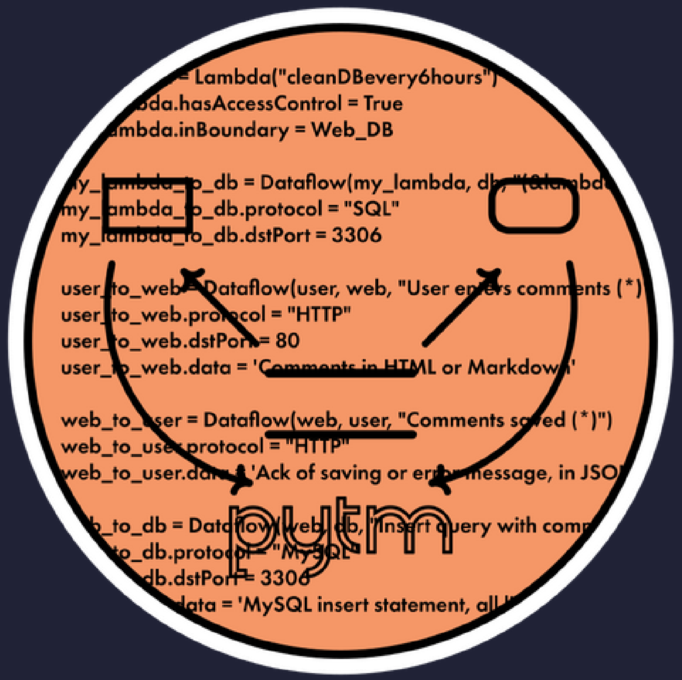
Resources

OWASP | TOP 10 LLM APPLICATIONS & GENERATIVE AI

Agentic AI – Threats and Mitigations

OWASP Top 10 for LLM Apps & Gen AI Agentic Security Initiative

Version 1.0
February 2025



```
= Lambda("cleanDBevery6hours")
bda.hasAccessControl = True
ambda.inBoundary = Web_DB

my_lambda_to_db = Dataflow(my_lambda, db, "(&lambd
my_lambda_to_db.protocol = "SQL"
my_lambda_to_db.dstPort = 3306

user_to_web = Dataflow(user, web, "User enters comments (*)
user_to_web.protocol = "HTTP"
user_to_web.dstPort = 80
user_to_web.data = 'Comments in HTML or Markdown'

web_to_user = Dataflow(web, user, "Comments saved (*)")
web_to_user.protocol = "HTTP"
web_to_user.data = 'Ack of saving or error message, in JSO

b_to_db = Dataflow(web, db, "Insert query with comm
to_db.protocol = "MySQL"
db.dstPort = 3306
data = 'MySQL insert statement, all
```

01 aiwithbrandon. (2025, May 3). Agent Development Kit (ADK) Masterclass: Build AI Agents & Automate Workflows (Beginner to Pro) [Video]. YouTube. Google. (2025). Agent Development Kit. Github.io. https://google.github.io/adk-docs/

02 Google. (2025). Agent Development Kit. Github.io. https://google.github.io/adk-docs/

03 OWASP GenAI Project. (2025, April 28). Agentic AI - Threats and Mitigations. OWASP Gen AI Security Project. https://genai.owasp.org/resource/agentic-ai-threats-and-mitigations/

04 OWASP pytm | OWASP Foundation. (2025). Owasp.org. https://owasp.org/www-project-pytm/

05 Threat Modeling | OWASP Foundation. (2020). Owasp.org. https://owasp.org/www-community/Threat_Modeling

06 Threat Modeling Process | OWASP Foundation. (2025). Owasp.org. https://owasp.org/www-community/Threat_Modeling_Process

THANK YOU!

DO YOU HAVE ANY QUESTIONS?