Google Cloud

Module 05
Gen AI agents: Transform your organization

Proprietary + Confidential

Generative AI Leader learning path

01
Gen AI: Beyond the chatbot

02
Gen AI: Unlock foundational concepts

03
Gen AI: Navigate the landscape

04
Gen AI apps: Transform your work

05
Gen AI agents: Transform your organization

Proprietary + Confidential

Module objectives

01 Define gen AI agent components and how they work together.

02 Explain how to combine components of gen AI agents to build powerful solutions.

03 Describe how to create gen AI agents and solutions using Google Cloud's gen AI products.

04 Explain how to effectively lead organizational transformation with gen AI.

Proprietary + Confidential

Agenda

01 Today’s agents

02 Building agents

03 Enhancing the customer experience with agents

Proprietary + Confidential

01

Today’s agents

Proprietary + Confidential

A brief history of agents

Proprietary + Confidential

A brief history of agents

01 Deterministic (traditional)

02 Generative

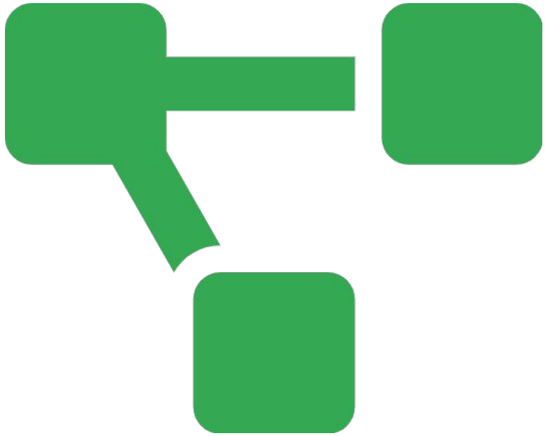
Proprietary + Confidential

A brief history of agents

01 Deterministic (traditional)

02 Generative

- Based on predefined paths and actions.
- Workflow-based and event-driven.
- Offers a high degree of control and predictability.
- The same input will always provide the same output.
- Includes: Reasoning loop + tools *(no foundation model!)*

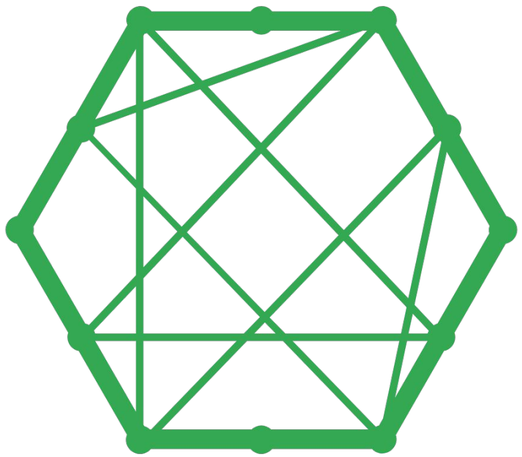


Proprietary + Confidential

A brief history of agents

01 Deterministic (traditional)

02 Generative



- Allows for much more natural and flexible conversations.
- Might give different answers to the same input.
- Includes: Reasoning loop + tools + foundation model.
- RAG enables integration from external sources.
- Hybrid agents.

Proprietary + Confidential

Components of a generative AI agent

Proprietary + Confidential

Components of a generative AI agent

- Foundation model
- Tools
- Reasoning loop

- The underlying large language model (LLM) that powers the agent.
- Select a model with training data relevant to the agent's intended use case.

Proprietary + Confidential

Components of a generative AI agent

- Foundation model
- Tools
- Reasoning loop

- Allow the agent to observe the world and to act upon it (enable interaction).
- Examples: Extensions that connect to APIs, functions that act as mock API calls, and data stores.

Proprietary + Confidential

Components of a generative AI agent

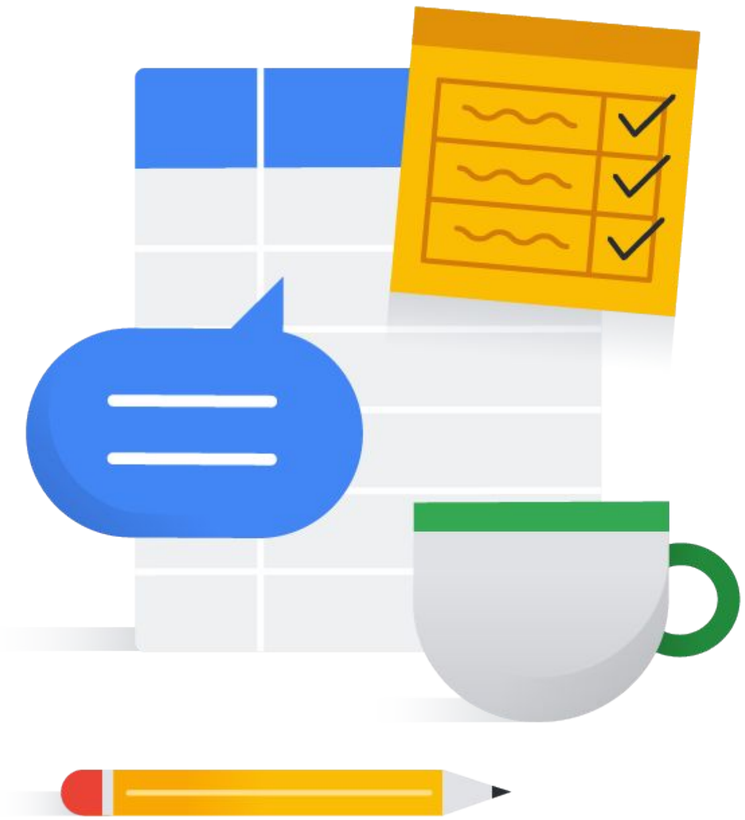
- Foundation model
- Tools
- Reasoning loop

- The core of the agent, responsible for making decisions and taking actions.
- An iterative process where the agent considers its goal, the available tools, and the information it has gathered.
- Frameworks guide the reasoning process.

Proprietary + Confidential

Key takeaways

- Generative AI agents evolved from inflexible deterministic types by adding natural language understanding. They use foundation models, tools for interaction, and a reasoning loop.



Proprietary + Confidential

Agenda

01 Today's agents

02 Building agents

03 Enhancing the customer experience with agents

Proprietary + Confidential

02

Building agents

Proprietary + Confidential

Using models

Proprietary + Confidential

Sampling parameters and settings

- Token count
- Temperature
- Top K
- Top P (nucleus sampling)
- Safety settings
- Output length

Proprietary + Confidential

Sampling parameters and settings

- Token count
- Temperature
- Top K
- Top P (nucleus sampling)
- Safety settings
- Output length

- Tokens are groups of characters (1 token ≅ 4 characters).
- Models have limits on how many tokens they can handle at once.

Proprietary + Confidential

Sampling parameters and settings

- Token count
- Temperature
- Top K
- Top P (nucleus sampling)
- Safety settings
- Output length

- Controls the "creativity" of the model by adjusting the randomness of word choices during text generation.
- Influences the diversity and unpredictability of the output.

Proprietary + Confidential

Sampling parameters and settings

- Token count
- Temperature
- Top K
- Top P (nucleus sampling)
- Safety settings
- Output length

- Lets the model randomly return a word from the top K number of words in terms of possibility.

Proprietary + Confidential

Sampling parameters and settings

- Token count
- Temperature
- Top K
- Top P (nucleus sampling)
- Safety settings
- Output length

- The cumulative probability of the most likely tokens considered during text generation.

Proprietary + Confidential

Sampling parameters and settings

- Token count
- Temperature
- Top K
- Top P (nucleus sampling)
- Safety settings
- Output length

- Allow you to filter out potentially harmful or inappropriate content from the model's output.

Proprietary + Confidential

Sampling parameters and settings

- Token count
- Temperature
- Top K
- Top P (nucleus sampling)
- Safety settings
- Output length

- Determines the maximum length of the generated text.

Proprietary + Confidential

Sampling parameters and settings

- Token count
- Temperature
- Top K
- Top P (nucleus sampling)
- Safety settings
- Output length

By experimenting with these parameters, you can significantly influence the AI model's behavior.

Proprietary + Confidential

Introduction to sampling parameters

Proprietary + Confidential

Accessing and changing model settings varies by **model source and access method**, often done through **application programming interfaces (APIs)**.

Proprietary + Confidential

Google gen AI tools

01 Google AI Studio

02 Vertex AI Studio

Proprietary + Confidential

Google gen AI tools

01 Google AI Studio

02 Vertex AI Studio

A web-based tool that allows developers, students, and researchers to try Gemini models and begin building with the Gemini Developer API.

Designed for ease of use and accessibility.

Proprietary + Confidential

Google gen AI tools

01 Google AI Studio

02 Vertex AI Studio

A Google Cloud console tool for rapidly prototyping and testing generative AI models.

Respects Google Cloud policies, ensuring governance and security throughout development.

Developers can:

- Test models using prompt samples.
- Design and save prompts.
- Tune foundation models.

Proprietary + Confidential

Comparison of Google AI Studio and Vertex AI Studio

01 Google AI Studio

- Offers a simpler interface for beginners to explore Gemini models, experiment with parameters, and generate creative text.
- Has usage limits and is better suited for prototyping with a standard Google account.

02 Vertex AI Studio

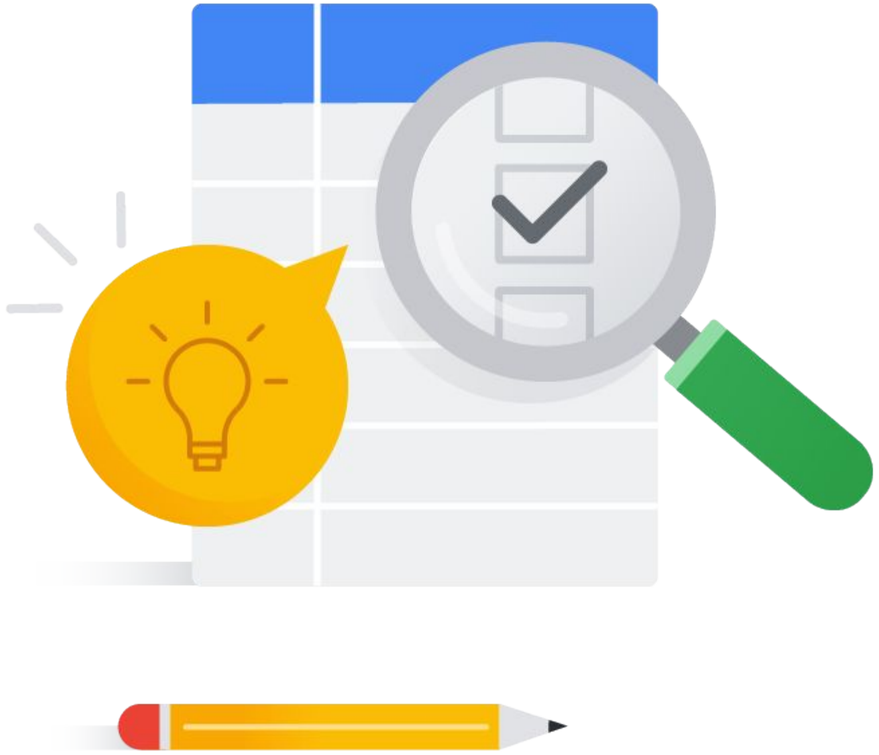
- A part of Vertex AI; a comprehensive platform within Google Cloud for professionals to build, train, and deploy ML models at scale.
- Provides enterprise-grade security and flexible quotas, incurring service charges.
- Easy integration with other tools, and grounding for better accuracy.

Proprietary + Confidential

Demo: Google AI Studio demo - Parameter playground

5 min

aistudio.google.com/



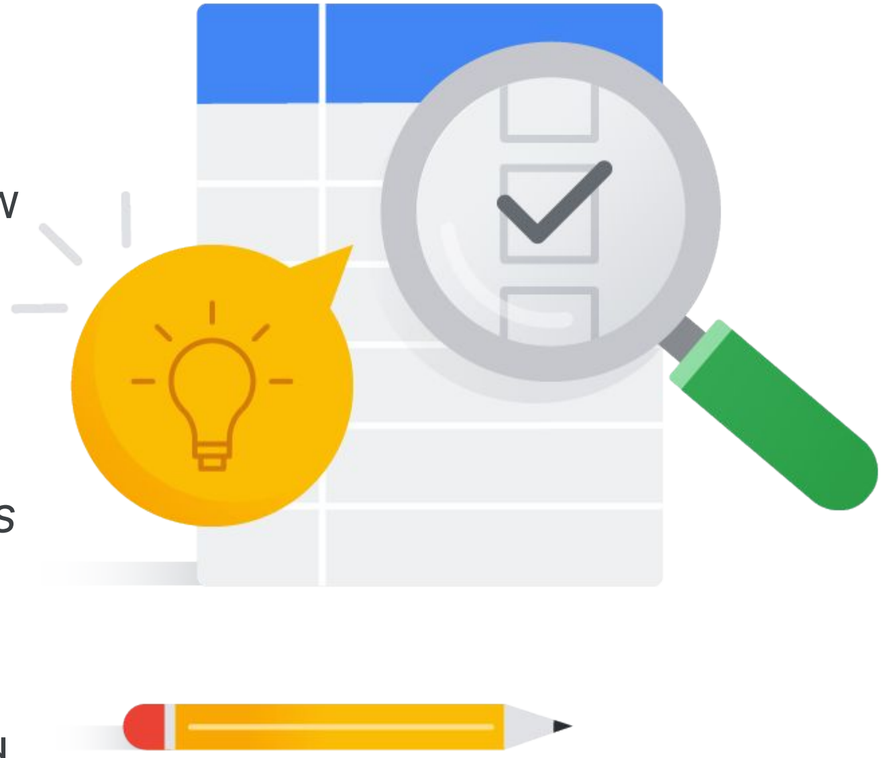
Proprietary + Confidential

Demo: Google AI Studio

Proprietary + Confidential

Takeaway try it: Parameter playground in Google AI Studio (optional)

The best way to understand these parameters is to try them yourself. This activity will help you develop an intuition for how to adjust your inputs to the AI model and achieve the desired results.



Note: This activity requires specific Google Cloud permissions that you might not have.

Additional resource: Takeaway try it - Parameter playground in Google AI Studio

Proprietary + Confidential

Adding prompt engineering to your reasoning loops

Proprietary + Confidential

The reasoning loop

Iterative process

Agent continuously evaluates its progress and determines the next best action to take.

Internal reasoning

Agent uses its underlying language model to think through the steps it needs to take to complete a task.

Decision making

Based on its internal reasoning, agent decides on the next course of action (choosing tools and determining inputs).

Reasoning frameworks

Reasoning loop utilizes prompt engineering frameworks and techniques to guide its reasoning and planning.

Proprietary + Confidential

Prompt engineering techniques

Chain-of-thought (CoT)

A technique where you guide a language model through a problem-solving process by providing examples with intermediate reasoning steps.

ReAct (Reason and Act)

A prompting framework that allows the language model to reason and take action on a user query, with or without in-context examples.

Proprietary + Confidential

Chain-of-thought (CoT)

- CoT enables reasoning capabilities through intermediate steps.
- Instead of just giving the LLM a prompt and expecting an answer, you guide it through the reasoning process.
- The intermediate steps are the “chain of thought” that leads to the correct answer.

Proprietary + Confidential

Ways to implement CoT

- Self-consistency
- Active-prompting
- Multimodal CoT

Proprietary + Confidential

Ways to implement CoT

- Self-consistency
- Active-prompting
- Multimodal CoT

Encouraging the LLM to generate multiple solutions and choose the most consistent one.

Proprietary + Confidential

Ways to implement CoT

- Self-consistency
- Active-prompting
- Multimodal CoT

Allowing the LLM to ask clarifying questions or request additional information.

Proprietary + Confidential

Ways to implement CoT

- Self-consistency
- Active-prompting
- Multimodal CoT

Combining text with other forms of data, like images or videos, to enhance reasoning.

Proprietary + Confidential

Why is CoT important?

Improved reasoning

Helps LLMs solve complex problems that require logical thinking.

Better accuracy

Breaking down problems into smaller steps can lead to more accurate results.

Enhanced explainability

Makes it easier to understand how the LLM arrived at an answer, building trust and transparency.

Proprietary + Confidential

CoT in action

Complex reasoning tasks

Explanation generation

Multi-step planning

Proprietary + Confidential

Key components of ReAct

ReAct allows the LLM to interact with the real world and gather information.

Proprietary + Confidential

Key components of ReAct

Think
The LLM generates a thought about the problem.

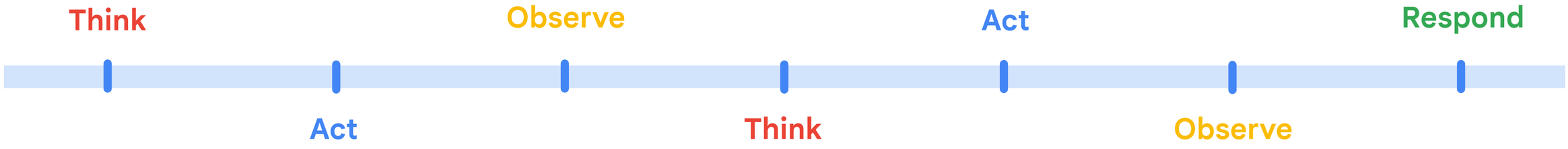
Act
The LLM decides what action to take, and specifies the input for the action.

Observe
The LLM receives feedback from the action.

Respond
The LLM generates a response.

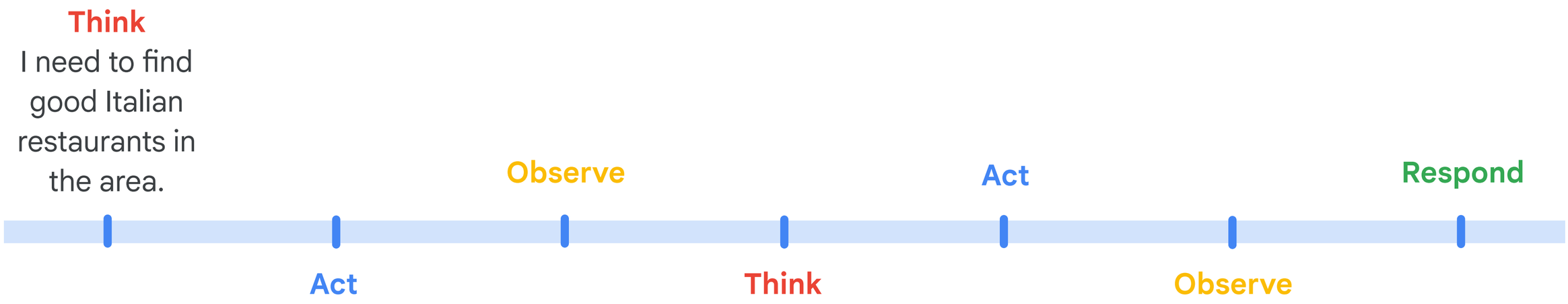
Proprietary + Confidential

ReAct: “thought-action-observation” loop



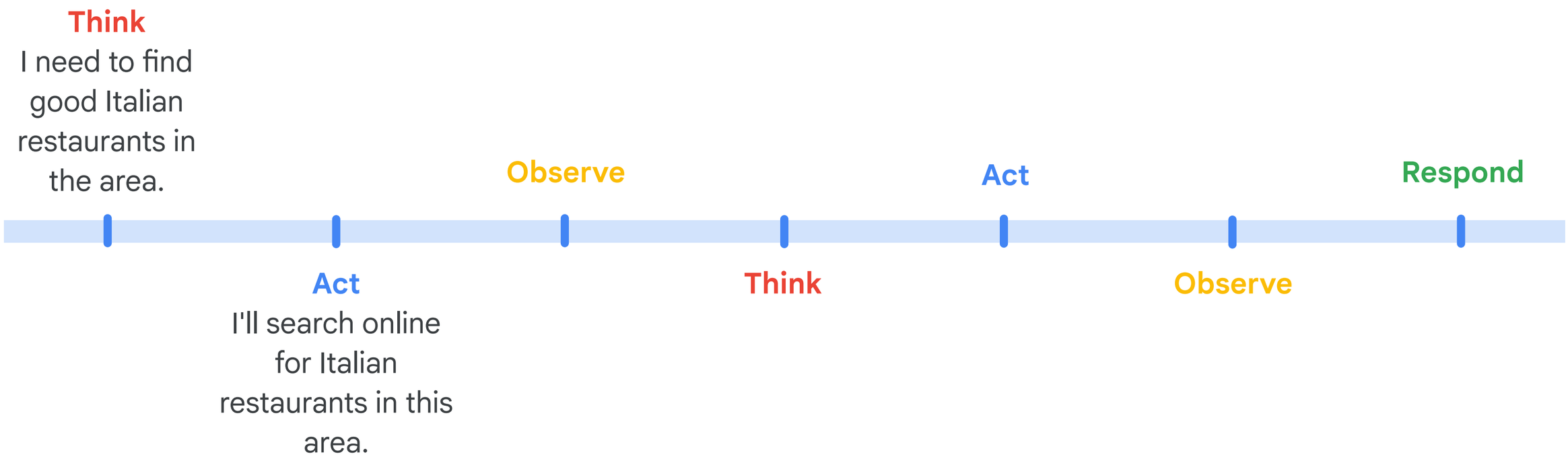
Proprietary + Confidential

ReAct: “thought-action-observation” loop example



Proprietary + Confidential

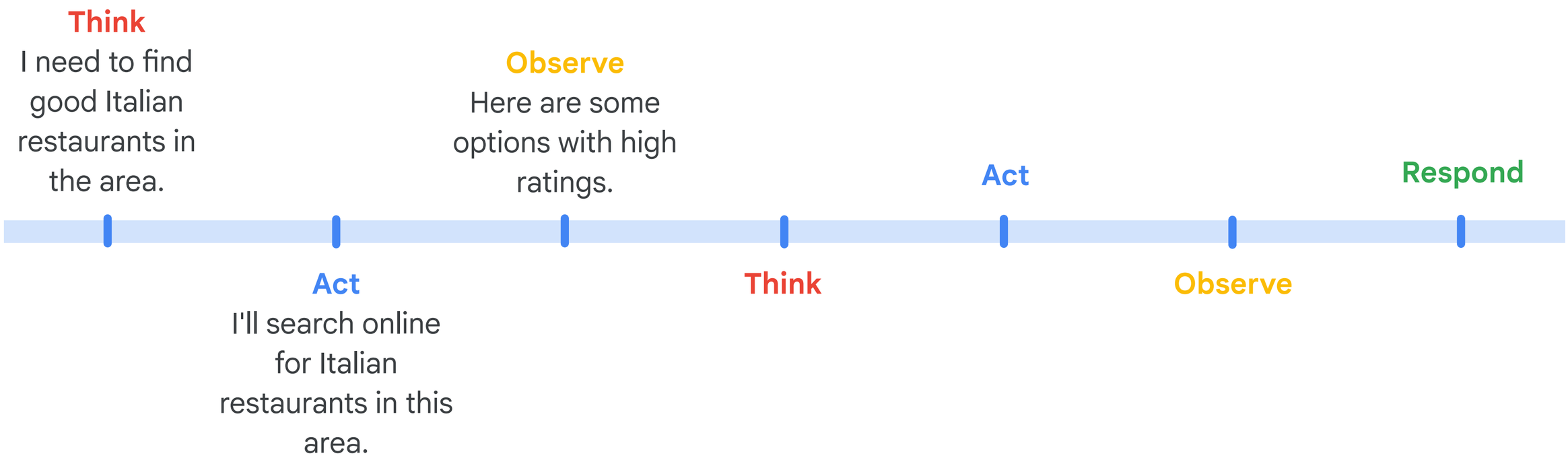
ReAct: “thought-action-observation” loop example



Google Cloud

Proprietary + Confidential

ReAct: “thought-action-observation” loop example



Proprietary + Confidential

ReAct: “thought-action-observation” loop example

Think
I need to find good Italian restaurants in the area.

Act
I'll search online for Italian restaurants in this area.

Observe
Here are some options with high ratings.

Think
Which one is the best fit based on the user's preferences?

Act

Observe

Respond

Proprietary + Confidential

ReAct: “thought-action-observation” loop example

Think
I need to find good Italian restaurants in the area.

Act
I'll search online for Italian restaurants in this area.

Observe
Here are some options with high ratings.

Think
Which one is the best fit based on the user's preferences?

Act
I'll check the menus to see which one has the most appealing dishes.

Observe

Respond

Proprietary + Confidential

ReAct: “thought-action-observation” loop example

Think
I need to find good Italian restaurants in the area.

Act
I'll search online for Italian restaurants in this area.

Observe
Here are some options with high ratings.

Think
Which one is the best fit based on the user's preferences?

Act
I'll check the menus to see which one has the most appealing dishes.

Observe
This one has delicious-looking pasta!

Respond

Proprietary + Confidential

ReAct: “thought-action-observation” loop example

Think
I need to find good Italian restaurants in the area.

Act
I'll search online for Italian restaurants in this area.

Observe
Here are some options with high ratings.

Think
Which one is the best fit based on the user's preferences?

Act
I'll check the menus to see which one has the most appealing dishes.

Observe
This one has delicious-looking pasta!

Respond
I recommend [name]. It has great reviews, is located nearby, and has a fantastic pasta selection.

Proprietary + Confidential

Why is ReAct important?

Dynamic problem solving	Reduced hallucination	Increased trustworthiness

Proprietary + Confidential

Why is ReAct important?

Dynamic problem solving

LLMs can tackle complex tasks that require interacting with external resources and adapting to new information.

Reduced hallucination

Increased trustworthiness

Proprietary + Confidential

Why is ReAct important?

Dynamic problem solving

LLMs can tackle complex tasks that require interacting with external resources and adapting to new information.

Reduced hallucination

Grounding reduces the risk of generating incorrect or nonsensical information.

Increased trustworthiness

Proprietary + Confidential

Why is ReAct important?

Dynamic problem solving

LLMs can tackle complex tasks that require interacting with external resources and adapting to new information.

Reduced hallucination

Grounding reduces the risk of generating incorrect or nonsensical information.

Increased trustworthiness

The LLM's reasoning process and how it interacts with external sources are visible, making responses transparent and trustworthy.

Proprietary + Confidential

ReAct in action

Question answering

Fact verification

Decision making

Proprietary + Confidential

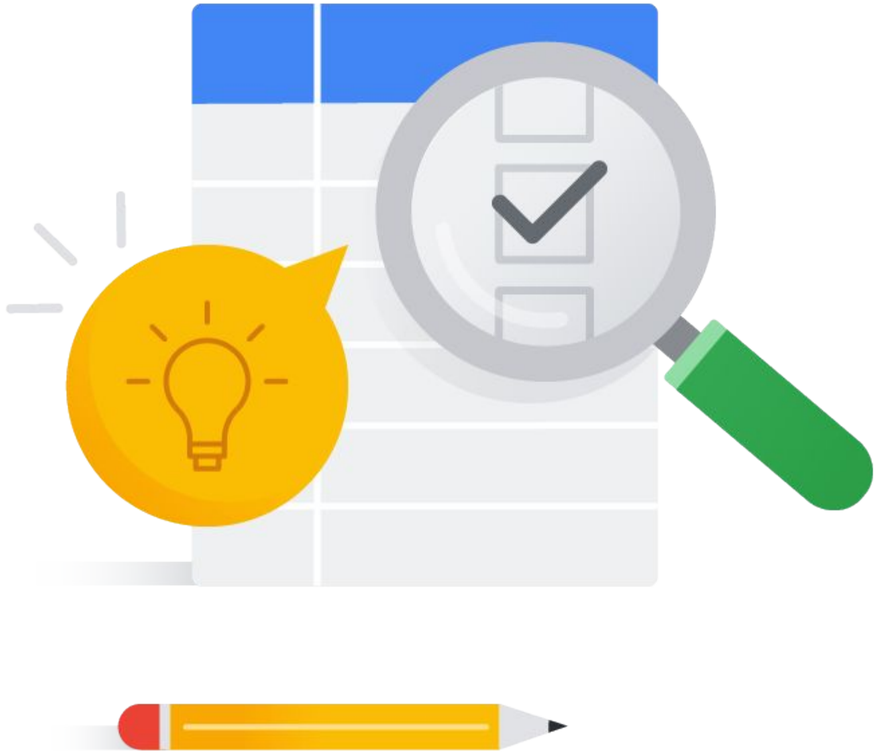
CoT focuses on internal reasoning, while ReAct enables external interaction. These prompt engineering techniques can be **combined for deeper reasoning and dynamic action**.

Proprietary + Confidential

Activity: The better prompt technique

5 min

1. Read the scenario.
2. Determine which would be a better primary prompt technique - ReAct or CoT? Provide a reason for your choice.
3. Put your answer in the chat.

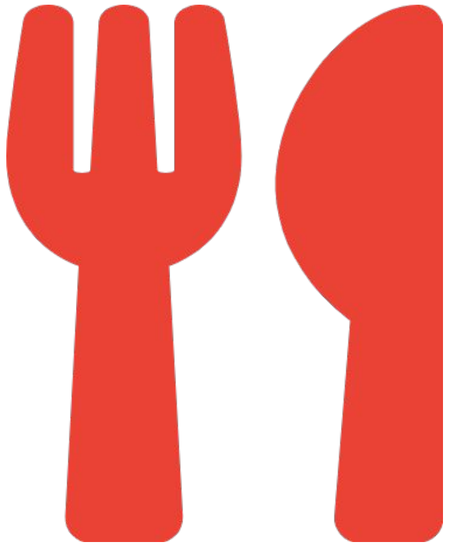


Proprietary + Confidential

Scenario: Find a restaurant

ReAct or CoT?

Find a highly-rated Italian restaurant near a current location.



Proprietary + Confidential

Scenario: Find a restaurant - Feedback

ReAct

Find a highly-rated Italian restaurant near a current location.

Reason

This would require reasoning and acting to get external information.

Proprietary + Confidential

Scenario: Summarize an article

ReAct or CoT?

Summarize a long article about the history of artificial intelligence.

Proprietary + Confidential

Scenario: Summarize an article - Feedback

CoT

Summarize a long article about the history of artificial intelligence.

Reason

This involves understanding and distilling information.

Proprietary + Confidential

Scenario: Brainstorm ideas

ReAct or CoT?

Brainstorm ideas for a new product.

Proprietary + Confidential

Scenario: Brainstorm ideas - Feedback

CoT

Brainstorm ideas for a new product.

Reason

This involves creative ideation, internal thought processes.

Proprietary + Confidential

Scenario: Search for research papers

ReAct or CoT?

Search for research papers on a particular scientific topic.

Proprietary + Confidential

Scenario: Search for research papers - Feedback

ReAct

Search for research papers on a particular scientific topic.

Reason

This would require reasoning and acting to get external information from academic databases.

Proprietary + Confidential

Adding tooling

Proprietary + Confidential

Types of agent tools

- Extensions (APIs)
- Functions
- Data stores
- Plugins

Proprietary + Confidential

Types of agent tools

- Extensions (APIs)
- Functions
- Data stores
- Plugins

Extensions (APIs) are sets of rules that govern how software interacts.

Extensions standardize agent interaction with external APIs, simplifying access to services and data despite varying API designs.

Proprietary + Confidential

Types of agent tools

- Extensions (APIs)
- Functions
- Data stores
- Plugins

Functions are specialized, reusable actions within an agent's toolbox. The reasoning system selects the appropriate function for the task at hand.

Proprietary + Confidential

Types of agent tools

- Extensions (APIs)
- Functions
- Data stores
- Plugins

Data stores give agents access to real-time, historical, and knowledge-based information, ensuring accurate and relevant responses.

Proprietary + Confidential

Types of agent tools

- Extensions (APIs)
- Functions
- Data stores
- Plugins

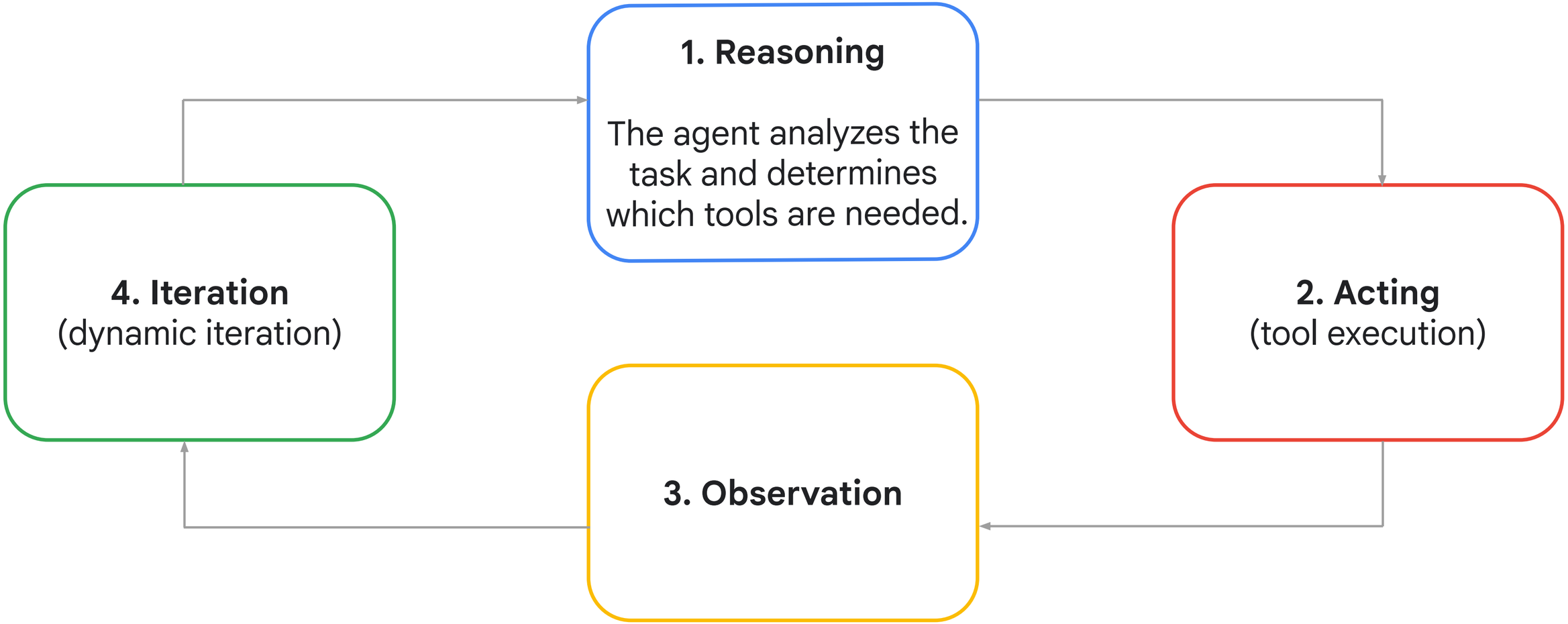
Plugins add new skills and integrations, extending an agent's capabilities by connecting it to services and specialized tools, or enabling interactions with other platforms.

Proprietary + Confidential

How does the reasoning loop work with tools?

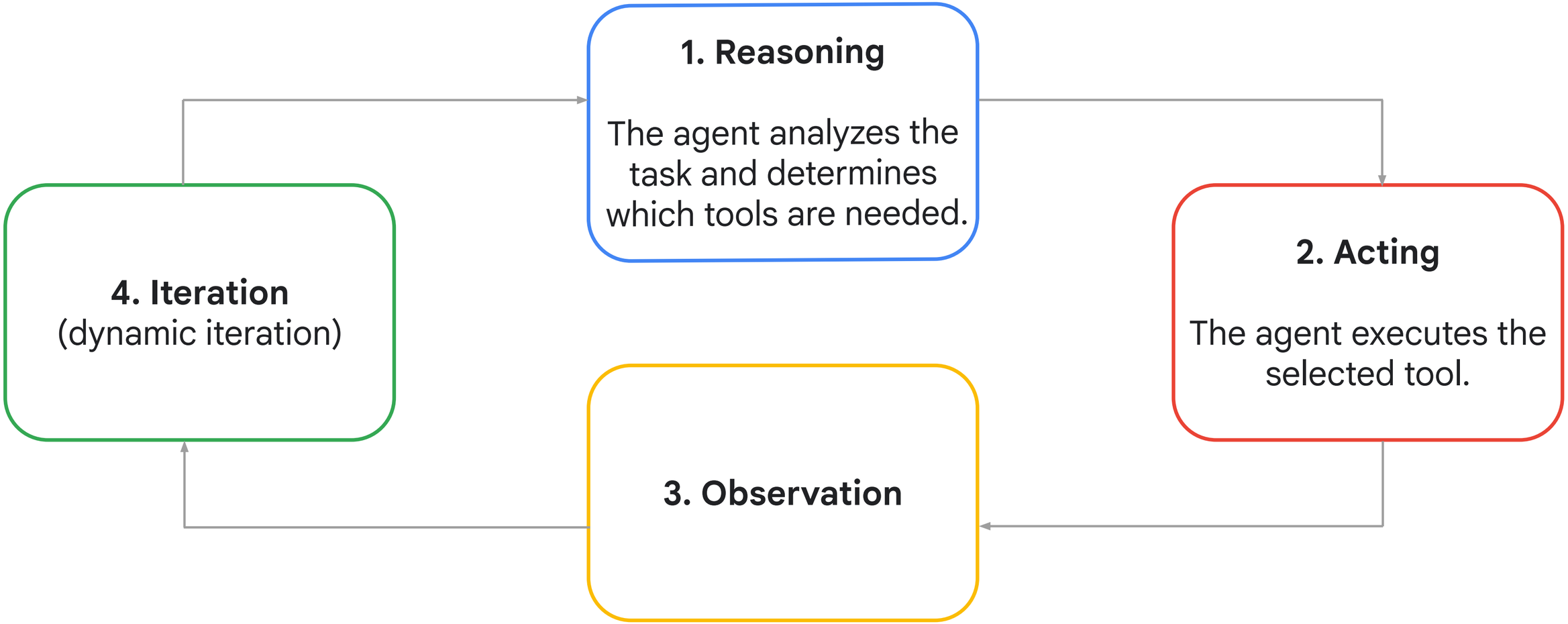
Proprietary + Confidential

How does the reasoning loop work with tools?



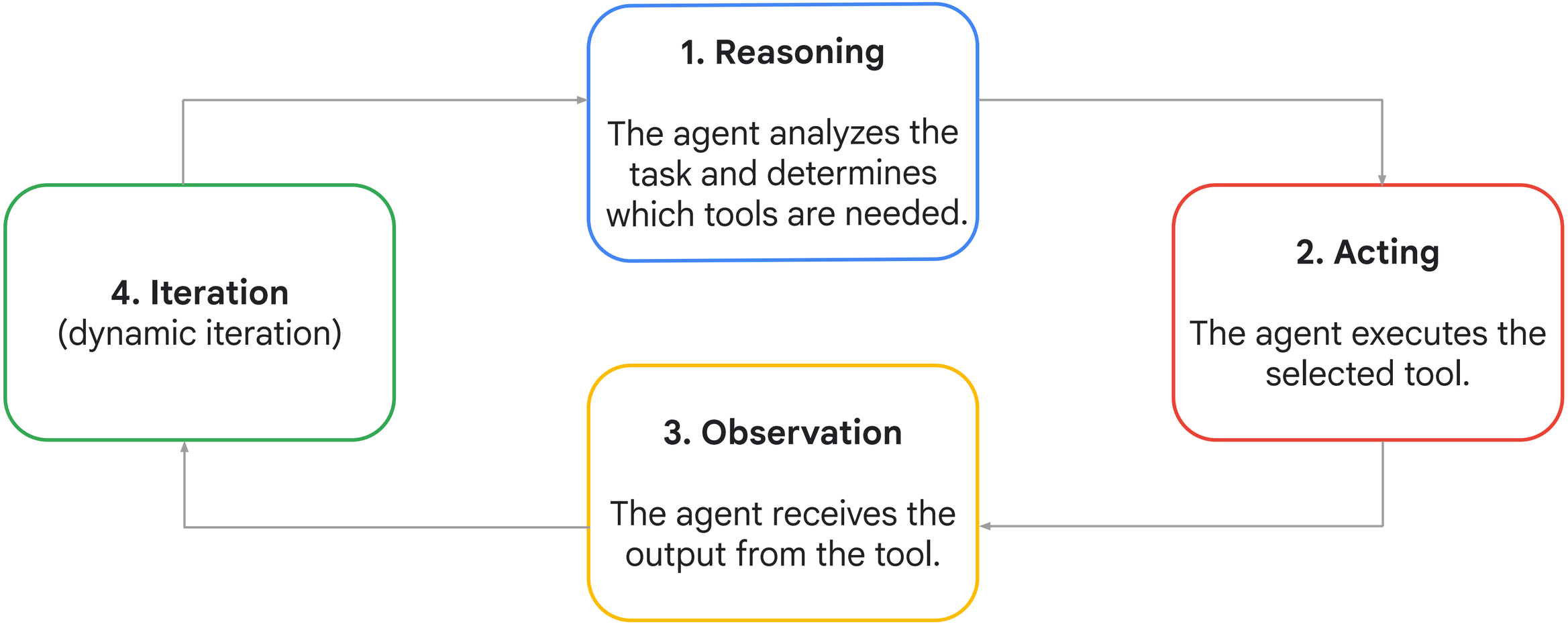
Proprietary + Confidential

How does the reasoning loop work with tools?



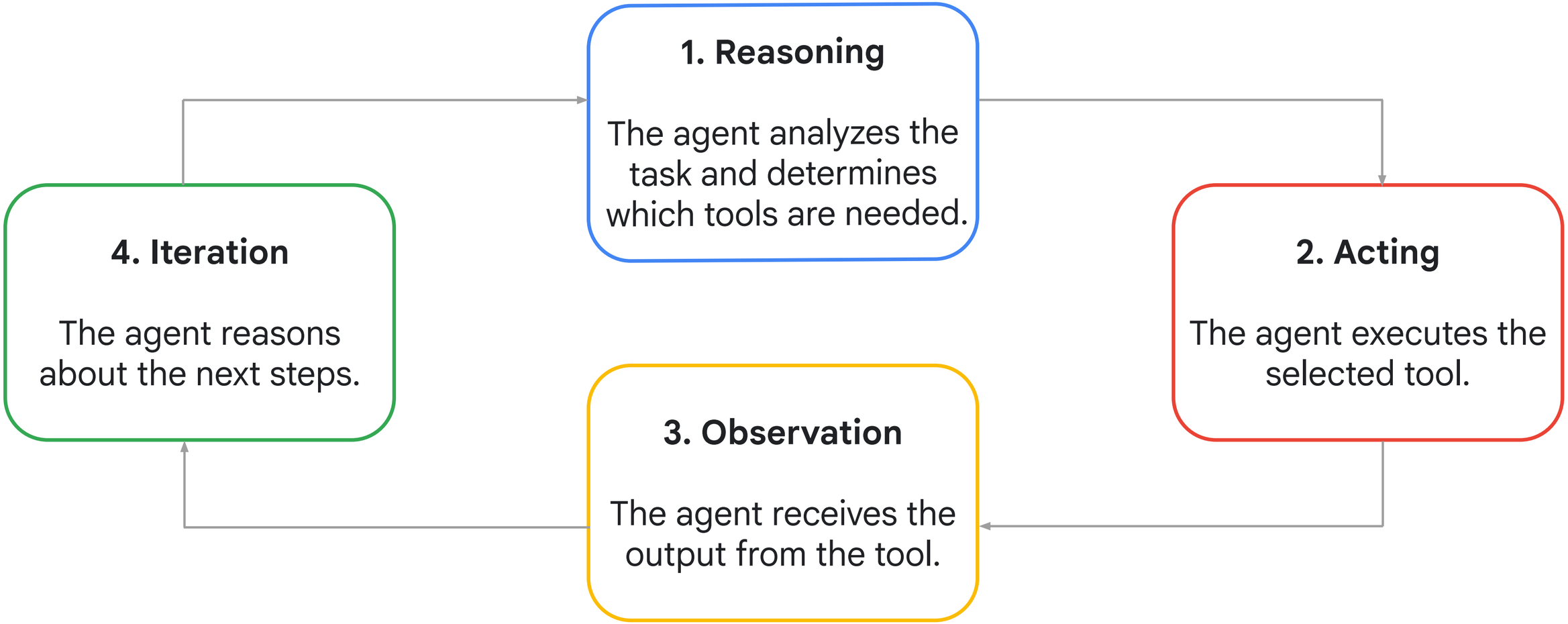
Proprietary + Confidential

How does the reasoning loop work with tools?



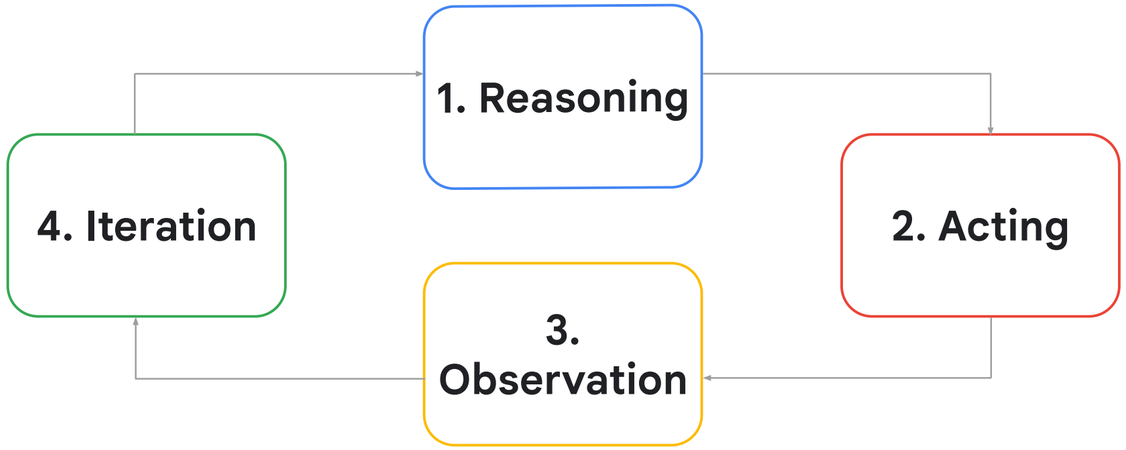
Proprietary + Confidential

How does the reasoning loop work with tools?



Proprietary + Confidential

Example: Scheduling a garden consultation



1. The agent needs to find an available time slot for a garden consultation.
2. The agent uses the scheduling plugin to check the gardener's availability for the next week, and accesses the customer database.
3. The scheduling plugin returns a list of available time slots.
4. The agent presents the available slots to the client.

Proprietary + Confidential

Google’s tooling: Key Google Cloud tools for agents

- Cloud Storage
- Databases
- Cloud Run functions
- Cloud Run
- Vertex AI

Proprietary + Confidential

Google’s tooling: Key Google Cloud tools for agents

- Cloud Storage
- Databases
- Cloud Run functions
- Cloud Run
- Vertex AI

- A highly scalable and durable object storage service.
- Used to store and retrieve data that your agent needs.

Proprietary + Confidential

Google’s tooling: Key Google Cloud tools for agents

- Cloud Storage
- Databases
- Cloud Run functions
- Cloud Run
- Vertex AI

- Cloud SQL, Spanner, Firestore.
- Agent uses databases to store and retrieve information, manage user data, or track its own progress.

Proprietary + Confidential

Google's tooling: Key Google Cloud tools for agents

- Cloud Storage
- Databases
- Cloud Run functions
- Cloud Run
- Vertex AI

- Create serverless functions that act as specialized tools for your agent.
- Easily triggered by your agent and scale automatically.

Proprietary + Confidential

Google's tooling: Key Google Cloud tools for agents

- Cloud Storage
- Databases
- Cloud Run functions
- Cloud Run
- Vertex AI

- Provides a serverless platform for deploying and running stateless containers.
- Ideal for custom tools that have specific dependencies or require more control.

Proprietary + Confidential

Google's tooling: Key Google Cloud tools for agents

- Cloud Storage
- Databases
- Cloud Run functions
- Cloud Run
- Vertex AI

- Use Vertex AI to create models or agents that are called as tooling by other agents.

Proprietary + Confidential

Google’s tooling: Pre-built AI APIs

Proprietary + Confidential

Google’s tooling: Pre-built AI APIs

Speech-to-Text API	Text-to-Speech API	Translation API	Document Translation API
Converts speech into text. Transcribes meetings, calls or video content.			

Proprietary + Confidential

Google's tooling: Pre-built AI APIs

Speech-to-Text API	Text-to-Speech API	Translation API	Document Translation API
Converts speech into text. Transcribes meetings, calls or video content.	Converts text into natural-sounding speech. Creates voice user interfaces and personalized communication.		

Proprietary + Confidential

Google's tooling: Pre-built AI APIs

Speech-to-Text API

Converts speech into text.

Transcribes meetings, calls or video content.

Text-to-Speech API

Converts text into natural-sounding speech.

Creates voice user interfaces and personalized communication.

Translation API

Translates text, documents, websites, and audio and video files.

Document Translation API

Proprietary + Confidential

Google's tooling: Pre-built AI APIs

Speech-to-Text API

Converts speech into text.

Transcribes meetings, calls or video content.

Text-to-Speech API

Converts text into natural-sounding speech.

Creates voice user interfaces and personalized communication.

Translation API

Translates text, documents, websites, and audio and video files.

Document Translation API

Translates formatted documents while keeping the original layout.

Proprietary + Confidential

Google's tooling: Additional pre-built AI APIs

Document AI API	Cloud Vision API	Cloud Video Intelligence API	Natural Language API
Extracts data from various document formats. Automates data capture, document processing. Summarizes documents.			

Proprietary + Confidential

Google's tooling: Additional pre-built AI APIs

Document AI API	Cloud Vision API	Cloud Video Intelligence API	Natural Language API
Extracts data from various document formats Automates data capture, document processing Summarizes documents	Allows developers to understand the content of an image. Image tagging, content moderation, and visual search.		

Proprietary + Confidential

Google's tooling: Additional pre-built AI APIs

Document AI API

Extracts data from various document formats.

Automates data capture, document processing.

Summarizes documents.

Cloud Vision API

Allows developers to understand the content of an image.

Image tagging, content moderation, and visual search.

Cloud Video Intelligence API

Allows developers to analyze video content and extract meaningful info.

Content recommendation, video search, and media analysis.

Natural Language API

Proprietary + Confidential

Google's tooling: Additional pre-built AI APIs

Document AI API

Extracts data from various document formats.

Automates data capture, document processing.

Summarizes documents.

Cloud Vision API

Allows developers to understand the content of an image.

Image tagging, content moderation, and visual search.

Cloud Video Intelligence API

Allows developers to analyze video content and extract meaningful info.

Content recommendation, video search, and media analysis.

Natural Language API

Helps derive insights from unstructured text.

Understands the sentiment of text, classify content, and extracts important entities.

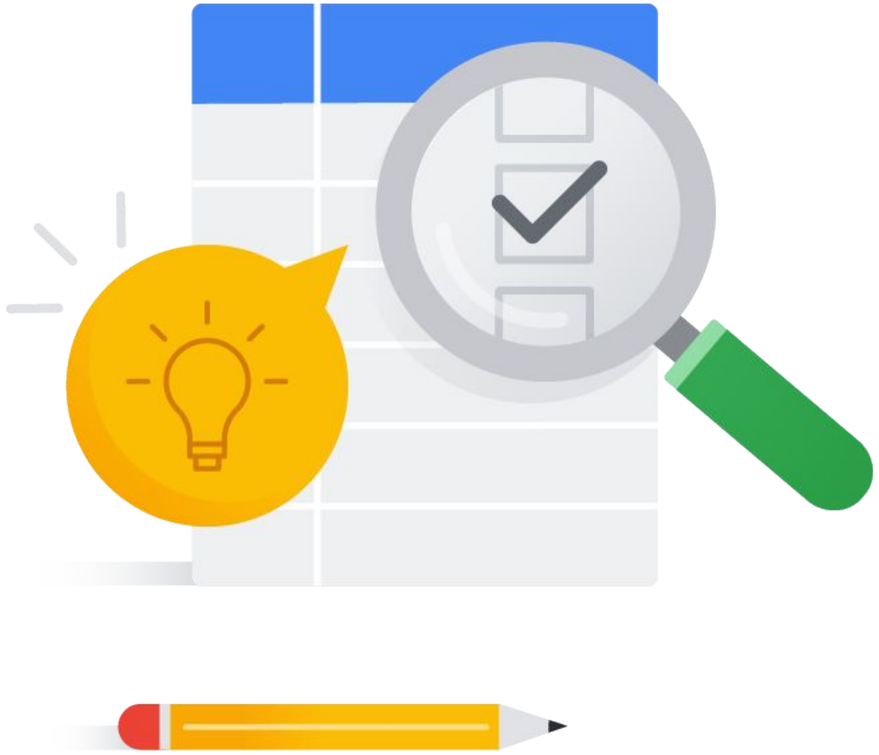
Proprietary + Confidential

Google Cloud also offers a vast API Library with APIs linked to many other Google products such as Google Maps, Google Workspace, YouTube, Google Photos, and more.

Proprietary + Confidential

Activity: Agent tooling example - Meeting location planner APIs

5 min



Scenario

A user uploads a document (e.g., a meeting invitation or a list of potential conference venues) and asks the agent to suggest the most convenient meeting location for attendees.

Proprietary + Confidential

Agent tooling example: Meeting location planner APIs

?

The agent uses this API to process the uploaded document and identify addresses mentioned within the document.

?

?

Proprietary + Confidential

Agent tooling example: Meeting location planner APIs

Document processing (Document AI API)

The agent uses this API to process the uploaded document and identify addresses mentioned within the document.

?

?

Proprietary + Confidential

Agent tooling example: Meeting location planner APIs

Document processing (Document AI API)

The agent uses this API to process the uploaded document and identify addresses mentioned within the document.

?

The agent takes the extracted addresses, and uses this API to convert them into latitude and longitude coordinates to calculate distances and travel times.

?

Proprietary + Confidential

Agent tooling example: Meeting location planner APIs

Document processing (Document AI API)

The agent uses this API to process the uploaded document and identify addresses mentioned within the document.

Geocoding (Google Maps Geocoding API)

The agent takes the extracted addresses, and uses this API to convert them into latitude and longitude coordinates to calculate distances and travel times.

?

Proprietary + Confidential

Agent tooling example: Meeting location planner APIs

Document processing (Document AI API)

The agent uses this API to process the uploaded document and identify addresses mentioned within the document.

Geocoding (Google Maps Geocoding API)

The agent takes the extracted addresses, and uses this API to convert them into latitude and longitude coordinates to calculate distances and travel times.

?

The agent uses this API, implemented within a Cloud function, to analyze the geocoded addresses and determine a suggested location.

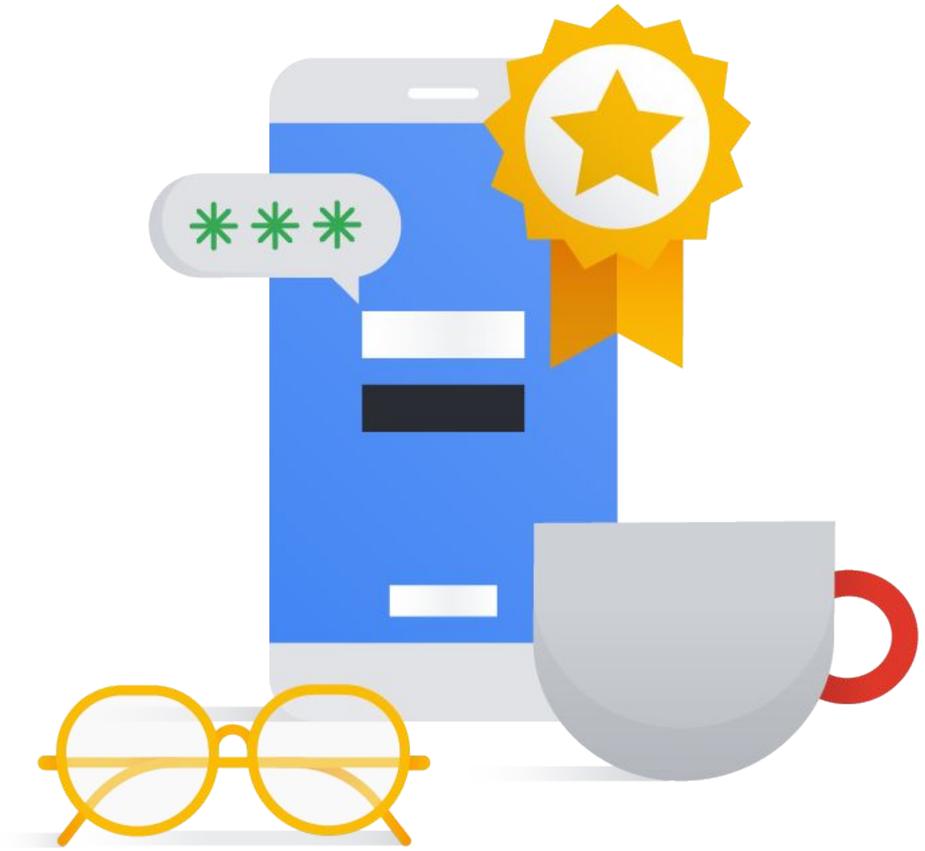
Proprietary + Confidential

Agent tooling example: Meeting location planner APIs

Document processing (Document AI API)	Geocoding (Google Maps Geocoding API)	Location analysis (custom logic)
The agent uses this API to process the uploaded document and identify addresses mentioned within the document.	The agent takes the extracted addresses, and uses this API to convert them into latitude and longitude coordinates to calculate distances and travel times.	The agent uses this API, implemented within a Cloud function, to analyze the geocoded addresses and determine a suggested location.

Proprietary + Confidential

Now let's do a short quiz to check your knowledge!



Proprietary + Confidential

Quiz | Question 01

Question

Which tool enables you to connect to external services (via APIs)?

A. Plugins

B. Functions

C. Data stores

D. Extensions

Proprietary + Confidential

Quiz | Question 01

Answer

Which tool enables you to connect to external services (via APIs)?

A. Plugins

B. Functions

C. Data stores

D. Extensions

Proprietary + Confidential

Quiz | Question 02

Question

Which tool enables you to define specific actions or tasks?

A. Plugins

B. Functions

C. Data stores

D. Extensions

Proprietary + Confidential

Quiz | Question 02

Answer

Which tool enables you to define specific actions or tasks?

A. Plugins

B. Functions

C. Data stores

D. Extensions

Proprietary + Confidential

Quiz | Question 03

Question

Which tool provides access to information?

A. Plugins

B. Functions

C. Data stores

D. Extensions

Proprietary + Confidential

Quiz | Question 03

Answer

Which tool provides access to information?

A. Plugins

B. Functions

C. Data stores

D. Extensions

Proprietary + Confidential

Quiz | Question 04

Question

Which tool enables you to add new skills and integrations?

A. Plugins

B. Functions

C. Data stores

D. Extensions

Proprietary + Confidential

Quiz | Question 04

Answer

Which tool enables you to add new skills and integrations?

A. Plugins

B. Functions

C. Data stores

D. Extensions

Proprietary + Confidential

Building applications from your agents

Proprietary + Confidential

How do you integrate APIs into your applications?

01

02

Proprietary + Confidential

How do you integrate APIs into your applications?

01

Generate an API key or set up authentication for secure access

Obtain credentials that allow your application to securely access and communicate with the Gemini API.

02

Proprietary + Confidential

How do you integrate APIs into your applications?

01

Generate an API key or set up authentication for secure access

Obtain credentials that allow your application to securely access and communicate with the Gemini API.

02

Start making API calls from your application code

API calls send your prompts and parameters to the Gemini model and receive the generated text in response, within your applications.

Proprietary + Confidential

Using the API: Enable API access for the tools

Vertex AI Studio

Google AI Studio

To use the API within your application, follow the directions for each tool.

Proprietary + Confidential

Using the API: Leverage your agent within applications

Cloud Run functions

Cloud Run

Proprietary + Confidential

Using the API: Leverage your agent within applications

Cloud Run functions

Cloud Run

- Small pieces of code that run in the cloud.
- Triggered by specific events.
- Use it to automatically process these events using your AI Studio API.
- No need to manage any servers.

Proprietary + Confidential

Using the API: Leverage your agent within applications

Cloud Run functions

Cloud Run

- A service that lets you run your applications in containers.
- Easy to deploy and scale these containerized applications.
- Useful for complex applications that might require specific software or configurations.

Proprietary + Confidential

Gemini API integrates into any application supporting HTTP requests, including web, mobile, desktop, and embedded systems.

Proprietary + Confidential

Using the API: Low-code or no-code tooling

Apps Script

AppSheet

- Is a low-code platform.
- Automates Google Workspace with JavaScript and built-in services.
- Uses Gemini APIs to create gen AI add-ons that enhance features in Google Workspace.

Proprietary + Confidential

Using the API: No code or low code tooling

Apps Script

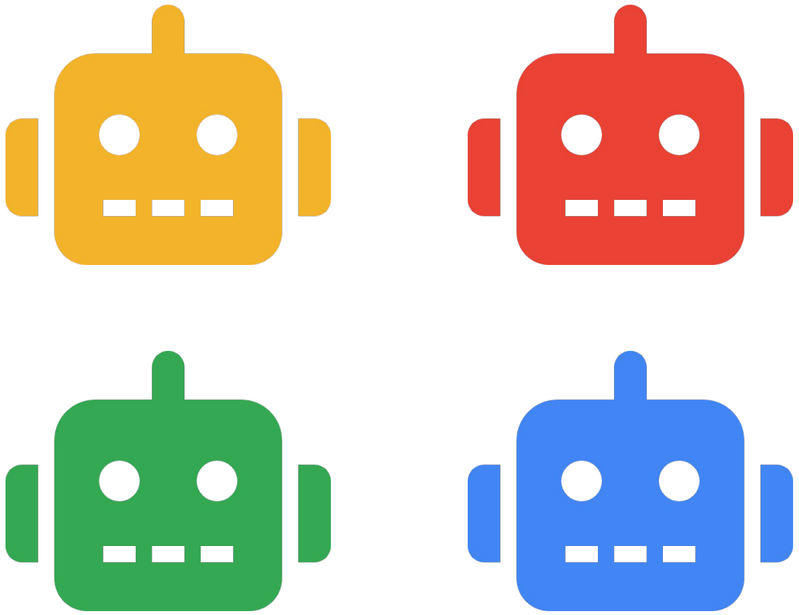
AppSheet

- Is a no-code platform.
- A platform for building custom business apps.
- Use AppSheet and the Gemini API for automation and intelligent features.

Proprietary + Confidential

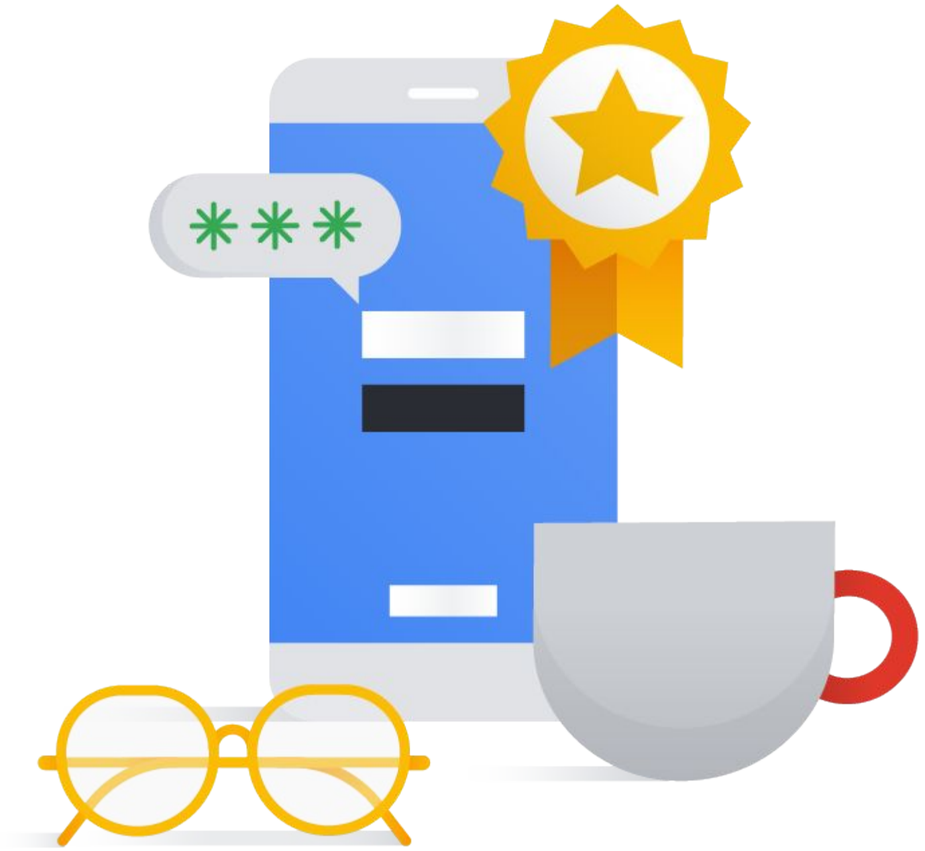
Multi-agent applications

- Systems that utilize multiple agents, each specialized for a specific task.
- Agents can work independently or interact with each other.
- Improves efficiency, allows for greater flexibility and scalability.
- An agent itself can be a tool within another agent.



Proprietary + Confidential

Now let's do a short quiz to check your knowledge!



Proprietary + Confidential

Quiz | Question 01

Question

Which parameter in Google AI Studio would you adjust to control the randomness and creativity of Gemini's output?

A. Token count

B. Temperature

C. Safety settings

D. Output length

Proprietary + Confidential

Quiz | Question 01

Answer

Which parameter in Google AI Studio would you adjust to control the randomness and creativity of Gemini's output?

A. Token count

B. Temperature

C. Safety settings

D. Output length

Proprietary + Confidential

Quiz | Question 02

Question

What is the primary distinction between Google AI Studio and Vertex AI Studio in terms of their target users?

A. Google AI Studio is for building custom models, while Vertex AI Studio is for using pre-trained models.

B. Google AI Studio is for accessing image generation models, while Vertex AI Studio is for text-based models.

C. Google AI Studio is designed for beginners and experimentation, while Vertex AI Studio caters to professionals and larger-scale projects.

D. Google AI Studio requires coding experience, while Vertex AI Studio offers a no-code interface for interacting with models.

Proprietary + Confidential

Quiz | Question 02

Answer

What is the primary distinction between Google AI Studio and Vertex AI Studio in terms of their target users?

A. Google AI Studio is for building custom models, while Vertex AI Studio is for using pre-trained models.

B. Google AI Studio is for accessing image generation models, while Vertex AI Studio is for text-based models.

C. Google AI Studio is designed for beginners and experimentation, while Vertex AI Studio caters to professionals and larger-scale projects.

D. Google AI Studio requires coding experience, while Vertex AI Studio offers a no-code interface for interacting with models.

Proprietary + Confidential

Quiz | Question 03

Question

Which of the following statements best describes the function of a "reasoning loop" within a generative AI agent?

A. It establishes a one-time set of instructions for the agent to follow without deviation.

B. It enables the agent to interact with users in a conversational manner, responding to prompts and questions.

C. It guides the agent through an iterative process of observation, internal reasoning, and decision-making to solve tasks.

D. It restricts the agent's actions to a predefined set of responses, ensuring predictable behavior.

Proprietary + Confidential

Quiz | Question 03

Answer

Which of the following statements best describes the function of a "reasoning loop" within a generative AI agent?

A. It establishes a one-time set of instructions for the agent to follow without deviation.

B. It enables the agent to interact with users in a conversational manner, responding to prompts and questions.

C. It guides the agent through an iterative process of observation, internal reasoning, and decision-making to solve tasks.

D. It restricts the agent's actions to a predefined set of responses, ensuring predictable behavior.

Proprietary + Confidential

Quiz | Question 04

Question

Imagine an AI agent tasked with scheduling a team meeting. The agent needs to access the team members' calendars, propose suitable time slots, and send out invitations. Which combination of agent tools would be most essential for this task?

A. Extensions and plugins

B. Data stores and functions

C. Plugins and data stores

D. Functions and databases

Proprietary + Confidential

Quiz | Question 04

Answer

Imagine an AI agent tasked with scheduling a team meeting. The agent needs to access the team members' calendars, propose suitable time slots, and send out invitations. Which combination of agent tools would be most essential for this task?

A. Extensions and plugins

B. Data stores and functions

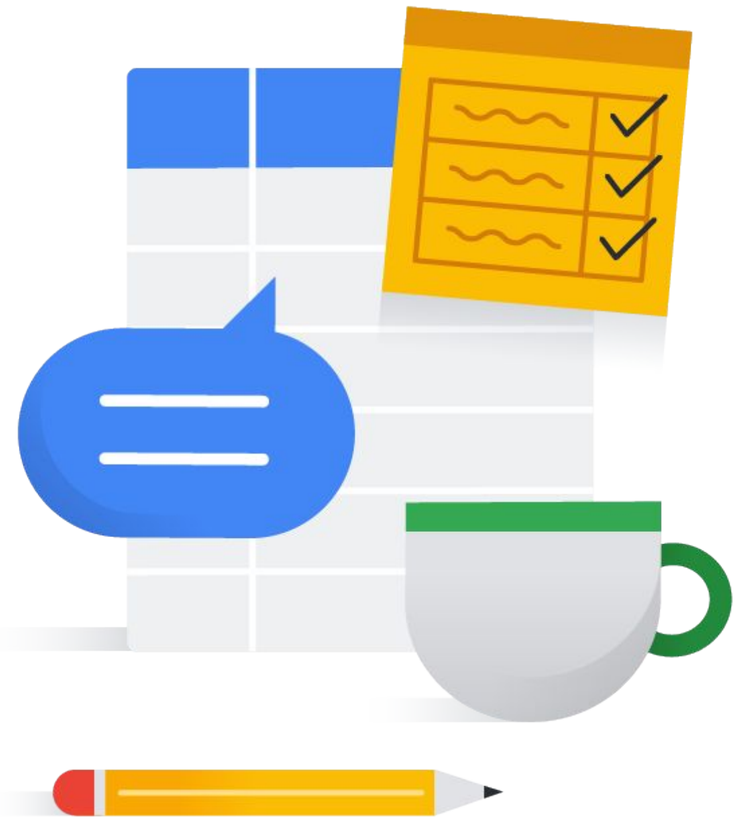
C. Plugins and data stores

D. Functions and databases

Proprietary + Confidential

Key takeaways

- Generative AI learns patterns to create text. Tools like Vertex and Google AI Studio allow parameter adjustments for customized application integration.
- The reasoning loop processes information for AI agent decisions. Prompt engineering enhances reasoning for improved interactions.
- Agent tooling enables AI to perform actions. Equip AI agents with tools for complex tasks and intelligent interaction.
- Integrate generative AI via various tools. Multi-agent systems enhance efficiency for diverse applications.



Proprietary + Confidential

Agenda

01 Today's agents

02 Building agents

03 Enhancing the customer experience with agents

Proprietary + Confidential

03

Enhancing the customer experience with agents

Proprietary + Confidential

Retrieval-augmented generation (RAG) and tooling

Proprietary + Confidential

How does the model work with RAG?

1. Retrieval
2. Augmentation
3. Generation

Proprietary + Confidential

How does the model work with RAG?

1. Retrieval
2. Augmentation
3. Generation

The LLM, equipped with retrieval tools, identifies relevant information from external sources.

Tools can include:

- Data stores
- Vector databases
- Search engines
- Knowledge graphs

Proprietary + Confidential

How does the model work with RAG?

1. Retrieval
2. Augmentation
3. Generation

The retrieved information is incorporated (or "augmented") into the prompt that is fed to the LLM.

This augmented prompt contains the user's original query and the relevant context retrieved from external sources.

Proprietary + Confidential

How does the model work with RAG?

1. Retrieval
2. Augmentation
3. Generation

The LLM processes the augmented prompt and generates a response.

Proprietary + Confidential

How does the model work with RAG?

1. Retrieval
2. Augmentation
3. Generation

✓ Iteration

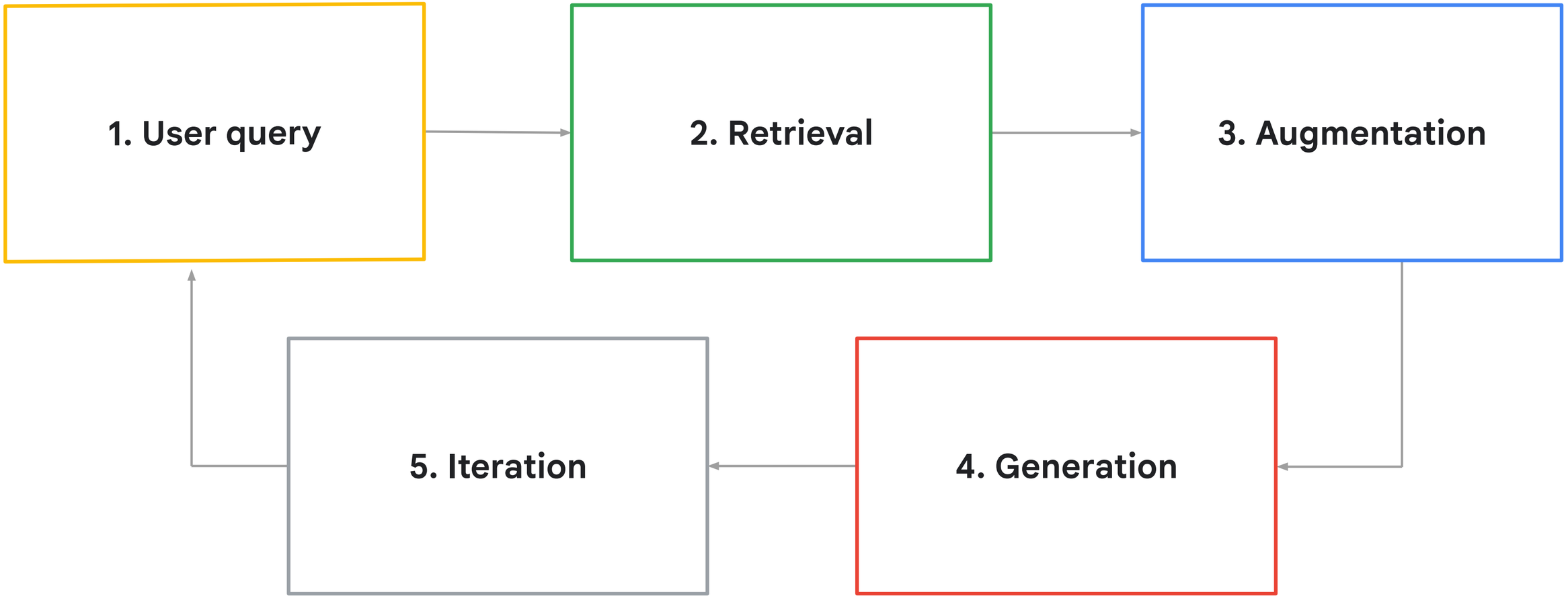
Note:

In some RAG systems, the LLM might iterate on the retrieval process.

This allows the LLM to continuously improve the quality and relevance of its responses.

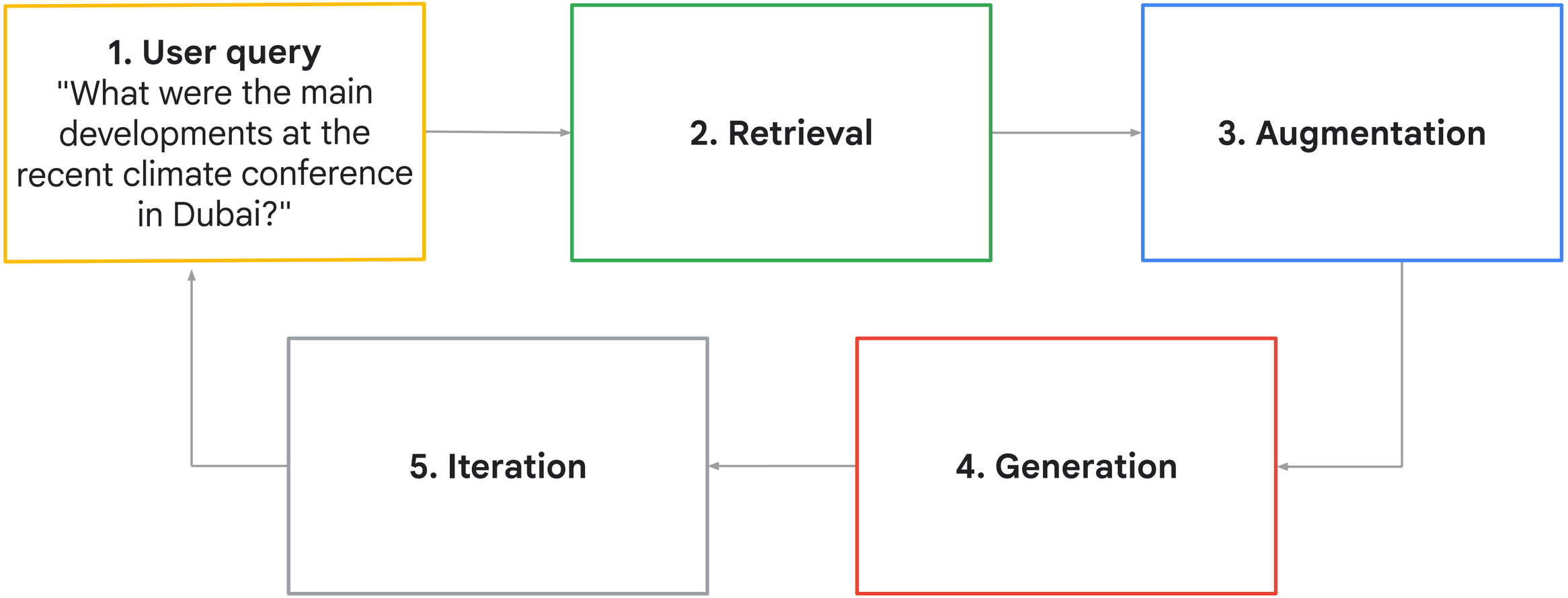
Proprietary + Confidential

RAG Example: Answering a question about a recent event



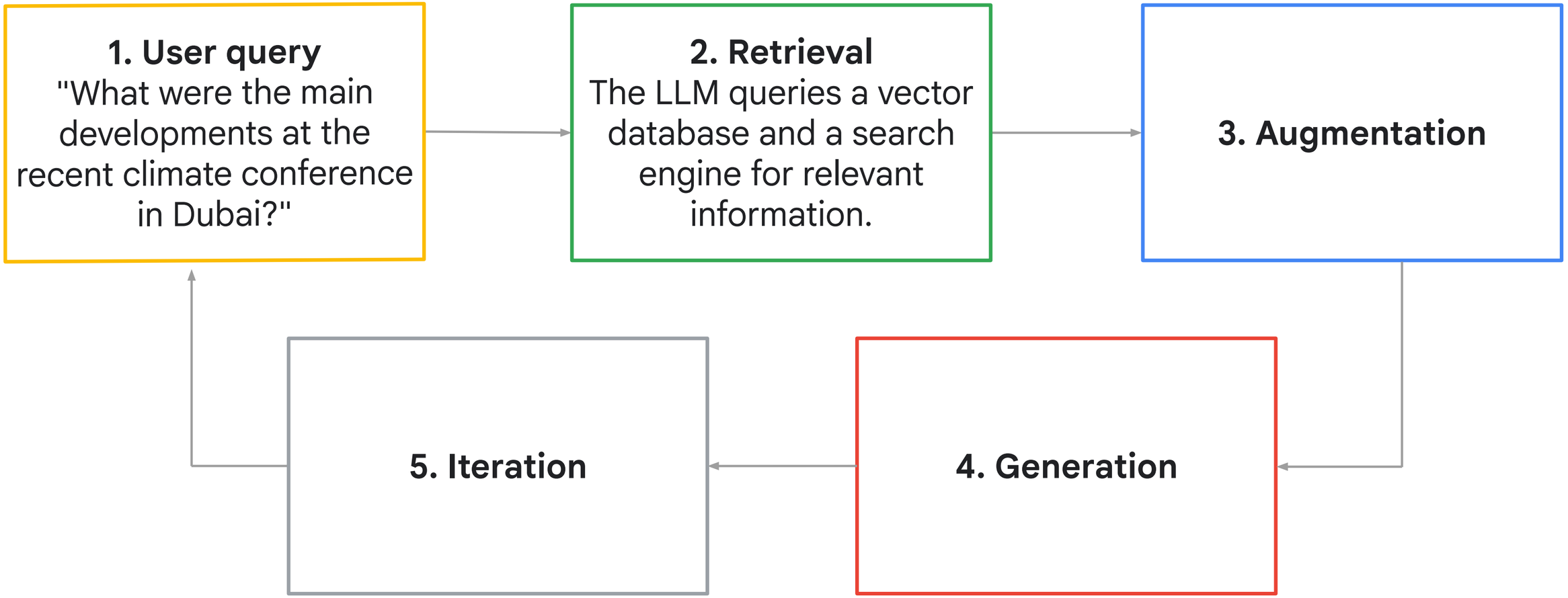
Proprietary + Confidential

RAG Example: Answering a question about a recent event



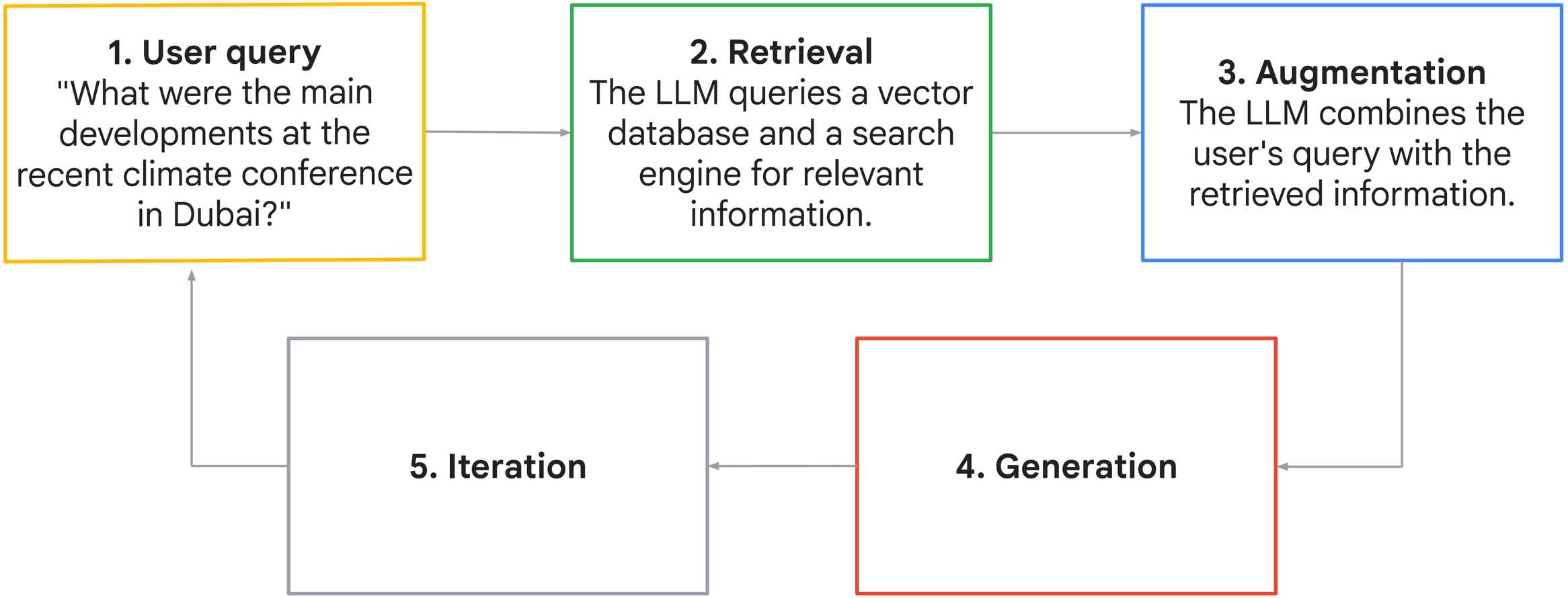
Proprietary + Confidential

RAG Example: Answering a question about a recent event



Proprietary + Confidential

RAG Example: Answering a question about a recent event



Proprietary + Confidential

RAG Example: Answering a question about a recent event

1. User query
"What were the main developments at the recent climate conference in Dubai?"

2. Retrieval
The LLM queries a vector database and a search engine for relevant information.

3. Augmentation
The LLM combines the user's query with the retrieved information.

4. Generation
The LLM generates a response summarizing the main developments at the climate conference, citing the sources.

5. Iteration

Proprietary + Confidential

RAG Example: Answering a question about a recent event

1. User query
"What were the main developments at the recent climate conference in Dubai?"

2. Retrieval
The LLM queries a vector database and a search engine for relevant information.

3. Augmentation
The LLM combines the user's query with the retrieved information.

4. Generation
The LLM generates a response summarizing the main developments at the climate conference, citing the sources.

5. Iteration
The LLM might try different search terms or query other data sources.

Proprietary + Confidential

Data stores

Data stores in AI applications act as structured and unstructured knowledge bases that your agent can draw upon.

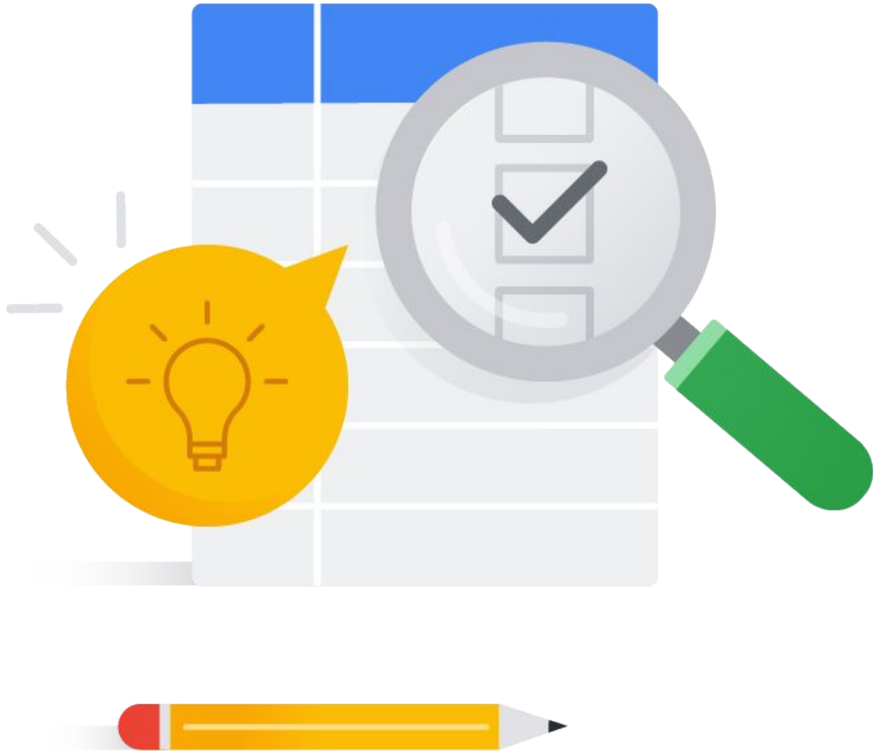
- Websites
- Structured databases
- Unstructured data

Proprietary + Confidential

Demo: RAG in action

5 min

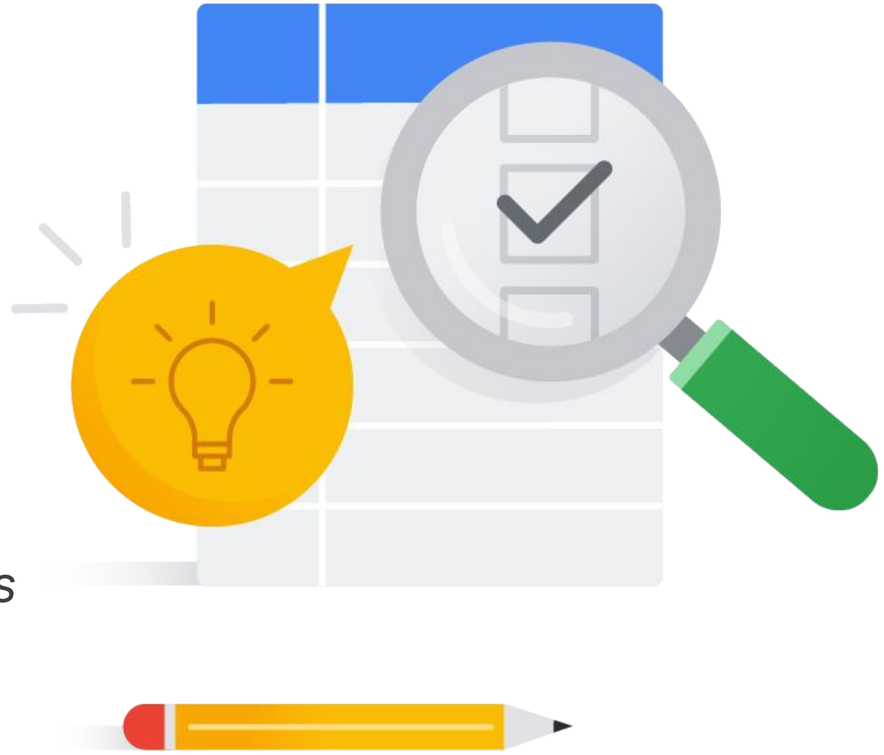
aistudio.google.com/



Proprietary + Confidential

Takeaway try it: RAG in action (optional)

The best way to understand how grounding can impact outputs is to try it yourself. In this activity you'll test using grounding with Google Search to see how RAG can impact outputs by providing more accurate and relevant information, drawing from real-time web data.

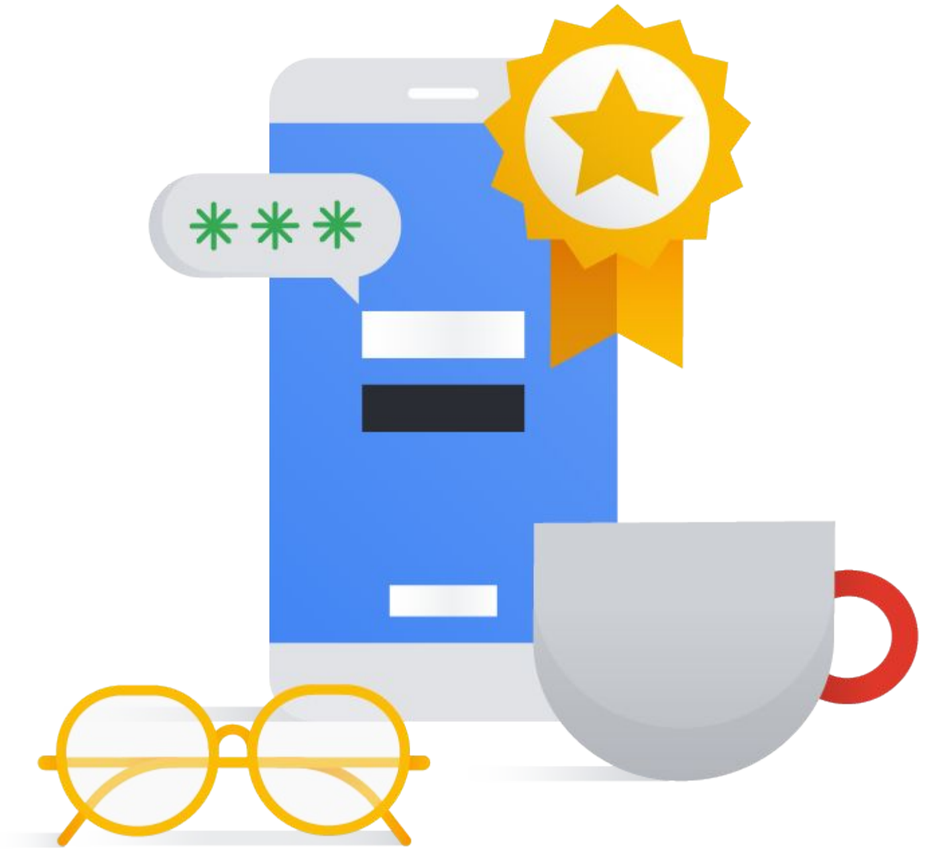


Note: This activity requires specific Google Cloud permissions that you might not have.

Additional resource: Takeaway try it - RAG in action

Proprietary + Confidential

Now let's do a short quiz to check your knowledge!



Proprietary + Confidential

Quiz | Question 01

Question

What stage in RAG involves fetching relevant documents from a knowledge base based on user query?

A. Retrieval

B. Augmentation

C. Generation

Proprietary + Confidential

Quiz | Question 01

Answer

What stage in RAG involves fetching relevant documents from a knowledge base based on user query?

A. Retrieval

B. Augmentation

C. Generation

Proprietary + Confidential

Quiz | Question 02

Question

What stage in RAG involves creating a response to the user's query in natural language?

A. Retrieval

B. Augmentation

C. Generation

Proprietary + Confidential

Quiz | Question 02

Answer

What stage in RAG involves creating a response to the user's query in natural language?

A. Retrieval

B. Augmentation

C. Generation

Proprietary + Confidential

Quiz | Question 03

Answer

What stage in RAG involves adding context to the query by including additional information from the knowledge base?

A. Retrieval

B. Augmentation

C. Generation

Proprietary + Confidential

Quiz | Question 03

Answer

What stage in RAG involves adding context to the query by including additional information from the knowledge base?

A. Retrieval

B. Augmentation

C. Generation

Proprietary + Confidential

The power of search agents

Proprietary + Confidential

Vertex AI: Search solutions

Document search	Media search	Healthcare search	Search for commerce

Proprietary + Confidential

Vertex AI: Search solutions

Document search	Media search	Healthcare search	Search for commerce
Use to search across a large repository of unstructured documents stored in Google Cloud Storage.			

Proprietary + Confidential

Vertex AI: Search solutions

Document search

Use to search across a large repository of unstructured documents stored in Google Cloud Storage.

Media search

Use with rich media libraries.

Understands and search in images, videos, and audio files.

Healthcare search

Search for commerce

Proprietary + Confidential

Vertex AI: Search solutions

Document search

Use to search across a large repository of unstructured documents stored in Google Cloud Storage.

Media search

Use with rich media libraries.

Understands and search in images, videos, and audio files.

Healthcare search

Enables searching across healthcare data while supporting regulatory compliance.

Search for commerce

Proprietary + Confidential

Vertex AI: Search solutions

Document search

Use to search across a large repository of unstructured documents stored in Google Cloud Storage.

Media search

Use with rich media libraries.

Understands and search in images, videos, and audio files.

Healthcare search

Enables searching across healthcare data while supporting regulatory compliance.

Search for commerce

Focuses on building a search app for your retail catalog.

Proprietary + Confidential

Vertex AI: Recommendations solutions

Media recommendations

Search for commerce

Proprietary + Confidential

Vertex AI: Recommendations solutions

Media recommendations

Tailored for consumer-focused media applications, like audio/video streaming and digital publishing.

Search for commerce

Optimized for e-commerce—drives sales by providing personalized product recommendations.

Proprietary + Confidential

How does Vertex AI Search work?

- Vertex AI Search uses intelligent data connection and generative AI. It acts as an agent, retrieving and recommending information from diverse data stores based on context.
- **Key strength:** Grounds gen AI LLM responses with your first-party data, curated third-party data, and Google's knowledge graph (using a RAG approach)

Proprietary + Confidential

Vertex AI Search: Extra generative AI features

- Search summaries
- Answers and follow-ups

Proprietary + Confidential

Vertex AI Search: Extra generative AI features

- Search summaries
- Answers and follow-ups

Generates concise, tailored summaries of search results, saving user time and providing quick overviews or comparisons.

Proprietary + Confidential

Vertex AI Search: Extra generative AI features

- Search summaries
- Answers and follow-ups

Adds AI-generated answers to search results.

Users can ask questions in natural language.

Proprietary + Confidential

Vertex AI Search offers **secure, scalable search with advanced analytics and easy integration** via APIs/SDKs for enterprises of all sizes, enhancing both internal and external applications.

Proprietary + Confidential

Use case: Enhancing the shopping experience with Vertex AI Search

Proprietary + Confidential

Customer engagement

Proprietary + Confidential

Google’s Customer Engagement Suite

Conversational Agents

Act as effective chatbots communicating with your customers.

Agent Assist

Conversational Insights

Proprietary + Confidential

Google’s Customer Engagement Suite

Conversational Agents

Act as effective chatbots communicating with your customers.

Agent Assist

Supports your live human contact center agents.

Conversational Insights

Proprietary + Confidential

Google's Customer Engagement Suite

Conversational Agents

Act as effective chatbots communicating with your customers.

Agent Assist

Supports your live human contact center agents.

Conversational Insights

Gain insights into all your communications with customers.

Proprietary + Confidential

Conversational Agents

01 Deterministic

02 Generative

03 Hybrid

Proprietary + Confidential

Conversational Agents

01 Deterministic

02 Generative

03 Hybrid

Deterministic agents follow explicitly defined rules and logic, requiring low to medium code for all actions.

Proprietary + Confidential

Conversational Agents

01 Deterministic

02 Generative

03 Hybrid

Generative agents use LLMs for natural conversation, determining actions based on prompts with minimal to no code.

Proprietary + Confidential

Conversational Agents

01 Deterministic

02 Generative

03 Hybrid

Google Cloud Conversational Agents enables hybrid agents, combining deterministic control with generative AI flexibility for improved customer service.

Proprietary + Confidential

Use case: Conversational agents

Proprietary + Confidential

Agent Assist

Supports live agents with real-time AI help:

- Generated responses
- Knowledge base suggestions
- Transcription
- Translation
- Summaries for faster, more accurate resolutions

Proprietary + Confidential

Conversational Insights

Analyzes customer interactions to boost efficiency and improve agent performance.

Generative FAQ identifies common questions and response effectiveness, highlighting areas for improvement.

Proprietary + Confidential

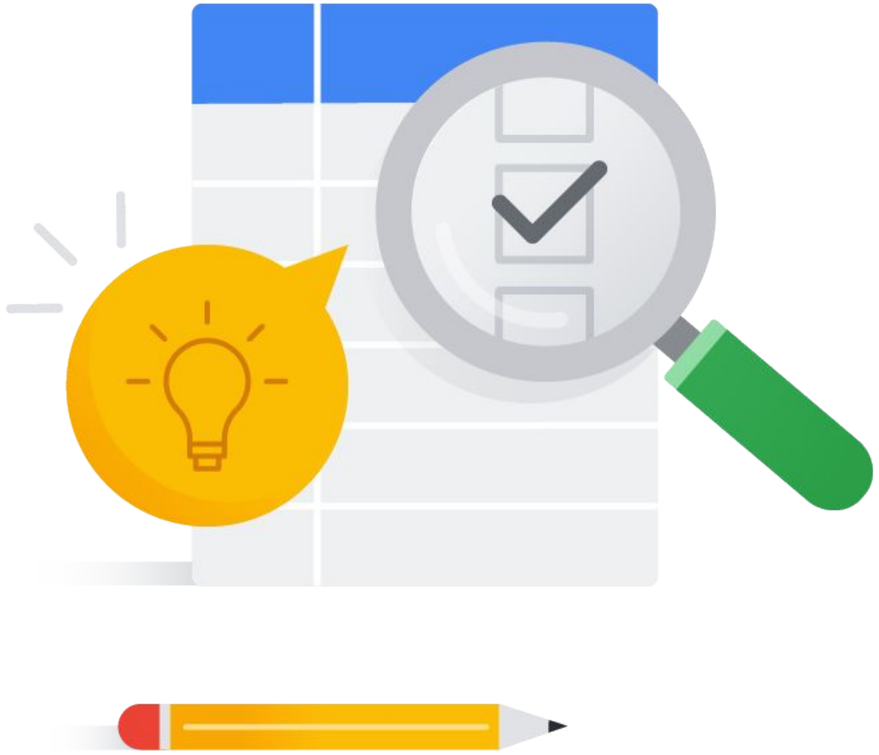
Contact Center as a Service (CCaaS) simplifies complex contact centers with 24/7 omnichannel support, CRM integration, and infrastructure management. It handles multichannel communication and integrates with agent and insight tools, focusing businesses on customer experience.

Proprietary + Confidential

Activity: Pick-a-product

5 min

Group



Identify the product or tool that matches the listed criteria.

Select from: Vertex AI Search, Vertex AI Recommendations, Conversational Agents, Agent Assist, or Conversational Insights

Proprietary + Confidential

Which product/tool...????

?	?	?	?
Delivers search experiences over your organization's data.			
Offers specialized tuning for industries like retail, media, and healthcare.			

Select from: Vertex AI Search, Vertex AI Recommendations, Conversational Agents, Agent Assist or Conversational Insights

Proprietary + Confidential

Which product/tool...????

Vertex AI Search	?	?	?
Delivers search experiences over your organization's data.			
Offers specialized tuning for industries like retail, media, and healthcare.			

Select from: Vertex AI Search, Vertex AI Recommendations, Conversational Agents, Agent Assist or Conversational Insights

Proprietary + Confidential

Which product/tool...????

Vertex AI Search	?	?	?
Delivers search experiences over your organization's data.	Provides guidance and suggestions to human contact center agents during live customer calls.		
Offers specialized tuning for industries like retail, media, and healthcare.	Aims to improve human agent efficiency, consistency, and first-call resolution.		

Select from: Vertex AI Search, Vertex AI Recommendations, Conversational Agents, Agent Assist, or Conversational Insights

Proprietary + Confidential

Which product/tool...????

Vertex AI Search	Agent Assist	?	?
Delivers search experiences over your organization's data.	Provides guidance and suggestions to human contact center agents during live customer calls.		
Offers specialized tuning for industries like retail, media, and healthcare.	Aims to improve human agent efficiency, consistency, and first-call resolution.		

Select from: Vertex AI Search, Vertex AI Recommendations, Conversational Agents, Agent Assist, or Conversational Insights

Proprietary + Confidential

Which product/tool...????

Vertex AI Search	Agent Assist	?	?
Delivers search experiences over your organization's data.	Provides guidance and suggestions to human contact center agents during live customer calls.	Provides personalized product or content suggestions to users.	
Offers specialized tuning for industries like retail, media, and healthcare.	Aims to improve human agent efficiency, consistency, and first-call resolution.	Aims to increase user engagement and user discovery.	

Select from: Vertex AI Search, Vertex AI Recommendations, Conversational Agents, Agent Assist, or Conversational Insights

Proprietary + Confidential

Which product/tool...????

Vertex AI Search	Agent Assist	Vertex AI Recommendations	?
Delivers search experiences over your organization's data.	Provides guidance and suggestions to human contact center agents during live customer calls.	Provides personalized product or content suggestions to users.	
Offers specialized tuning for industries like retail, media, and healthcare.	Aims to improve human agent efficiency, consistency, and first-call resolution.	Aims to increase user engagement and user discovery.	

Select from: Vertex AI Search, Vertex AI Recommendations, Conversational Agents, Agent Assist, or Conversational Insights

Proprietary + Confidential

Which product/tool...????

Vertex AI Search	Agent Assist	Vertex AI Recommendations	?
Delivers search experiences over your organization's data.	Provides guidance and suggestions to human contact center agents during live customer calls.	Provides personalized product or content suggestions to users.	Is used to design, build, and deploy AI-powered chatbots and voice bots.
Offers specialized tuning for industries like retail, media, and healthcare.	Aims to improve human agent efficiency, consistency, and first-call resolution.	Aims to increase user engagement and user discovery.	Understands user intent and extracts key information.

Select from: Vertex AI Search, Vertex AI Recommendations, Conversational Agents, Agent Assist, or Conversational Insights

Proprietary + Confidential

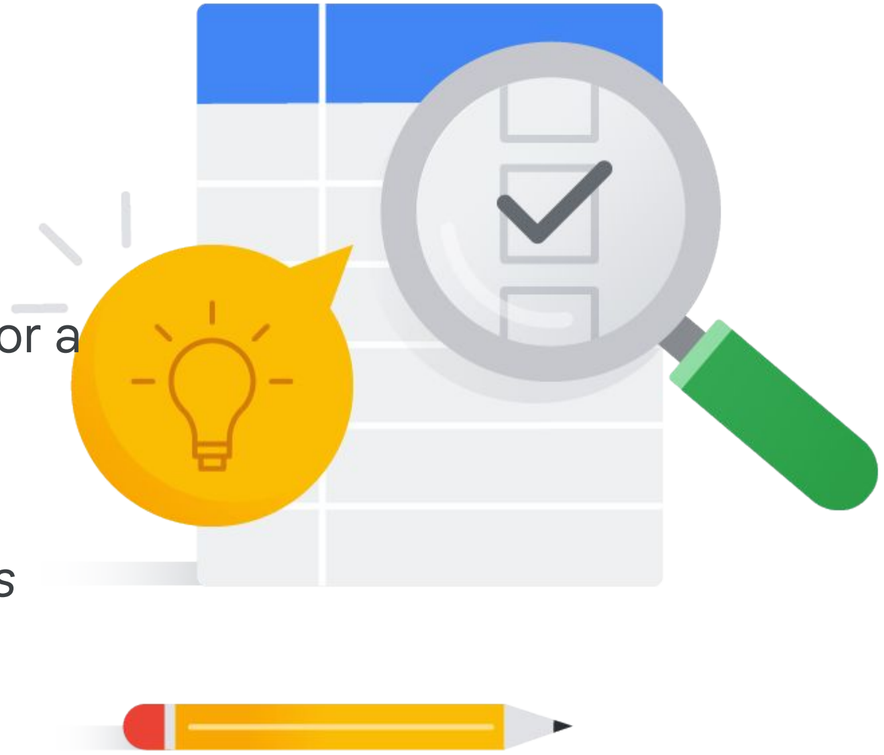
Which product/tool...????

Vertex AI Search	Agent Assist	Vertex AI Recommendations	Conversational Agents
Delivers search experiences over your organization's data.	Provides guidance and suggestions to human contact center agents during live customer calls.	Provides personalized product or content suggestions to users.	Is used to design, build, and deploy AI-powered chatbots and voice bots .
Offers specialized tuning for industries like retail, media, and healthcare.	Aims to improve human agent efficiency, consistency, and first-call resolution.	Aims to increase user engagement and user discovery.	Understands user intent and extracts key information.

Select from: Vertex AI Search, Vertex AI Recommendations, Conversational Agents, Agent Assist, or Conversational Insights

Proprietary + Confidential

Takeaway try it:
Build your own agent (optional)



Explore a very basic way of building a fully generative agent for a company, Cymbal Manufacturing, using Google AI Studio.

Note: This activity requires specific Google Cloud permissions that you might not have.

Additional resource: Takeaway try it - build your own agent

Proprietary + Confidential

Gemini Enterprise

Proprietary + Confidential

Gemini Enterprise creates AI-powered, customized agents that access and understand company data from any source, integrating into internal systems to act as personal work research assistants.

Proprietary + Confidential

Use case: Gemini Enterprise

Proprietary + Confidential

Gemini Enterprise: Launch point for enterprise-ready agents

- Connect to NotebookLM Enterprise
- Connect to multimodal search agents
- Add generative AI assistants
- Add custom agents

Proprietary + Confidential

Gemini Enterprise: Launch point for enterprise-ready agents

- Connect to NotebookLM Enterprise
- Connect to multimodal search agents
- Add generative AI assistants
- Add custom agents

Employees can upload information to analyze, get insights, and listen to audio summaries of data.

Proprietary + Confidential

Gemini Enterprise: Launch point for enterprise-ready agents

- Connect to NotebookLM Enterprise
- Connect to multimodal search agents
- Add generative AI assistants
- Add custom agents

Multimodal search agents are grounded in your data across multiple systems so your employees can find relevant information within your company.

Proprietary + Confidential

Gemini Enterprise: Launch point for enterprise-ready agents

- Connect to NotebookLM Enterprise
- Connect to multimodal search agents
- Add generative AI assistants
- Add custom agents

Generative AI assistants are grounded in your enterprise data and can be prompted to take actions through connectors (tools).

Proprietary + Confidential

Gemini Enterprise: Launch point for enterprise-ready agents

- Connect to NotebookLM Enterprise
- Connect to multimodal search agents
- Add generative AI assistants
- Add custom agents

Add custom agents through Conversational Agents.

Proprietary + Confidential

Strategy: Planning ahead

Proprietary + Confidential

Plan for generative AI integration

1. Establish a clear vision
2. Prioritize high-impact use cases
3. Invest in capabilities
4. Drive organizational change
5. Measure and demonstrate value
6. Champion responsible AI

Proprietary + Confidential

Plan for generative AI integration

1. Establish a clear vision
2. Prioritize high-impact use cases
3. Invest in capabilities
4. Drive organizational change
5. Measure and demonstrate value
6. Champion responsible AI

Align gen AI with strategy, secure leadership buy-in, and appoint an adoption champion.

Proprietary + Confidential

Plan for generative AI integration

1. Establish a clear vision
2. Prioritize high-impact use cases
3. Invest in capabilities
4. Drive organizational change
5. Measure and demonstrate value
6. Champion responsible AI

Align gen AI opportunities with business goals, considering feasibility and transformation potential.

Proprietary + Confidential

Plan for generative AI integration

1. Establish a clear vision
2. Prioritize high-impact use cases
3. Invest in capabilities
4. Drive organizational change
5. Measure and demonstrate value
6. Champion responsible AI

Develop gen AI skills via training and recruitment, fostering a learning and experimentation culture.

Proprietary + Confidential

Plan for generative AI integration

1. Establish a clear vision
2. Prioritize high-impact use cases
3. Invest in capabilities
4. Drive organizational change
5. Measure and demonstrate value
6. Champion responsible AI

Manage gen AI workflow shifts, and encourage agile collaboration and knowledge sharing for scaling.

Proprietary + Confidential

Plan for generative AI integration

1. Establish a clear vision
2. Prioritize high-impact use cases
3. Invest in capabilities
4. Drive organizational change
5. Measure and demonstrate value
6. Champion responsible AI

Track gen AI impact with data to continuously improve and demonstrate value.

Proprietary + Confidential

Plan for generative AI integration

1. Establish a clear vision
2. Prioritize high-impact use cases
3. Invest in capabilities
4. Drive organizational change
5. Measure and demonstrate value
6. Champion responsible AI

Implement a robust responsible AI framework ensuring fairness, transparency, privacy, and security.

Proprietary + Confidential

Plan for impact

- Define key metrics
- Collect and analyze data
- Iterate and improve

Proprietary + Confidential

Plan for impact

- Define key metrics
- Collect and analyze data
- Iterate and improve

Identify the metrics most relevant to your goals/business objectives.

Proprietary + Confidential

Plan for impact

- Define key metrics
- Collect and analyze data
- Iterate and improve

Set up tracking mechanisms, conduct surveys, or analyze existing data to understand the impact of your gen AI initiatives.

Proprietary + Confidential

Plan for impact

- Define key metrics
- Collect and analyze data
- Iterate and improve

Use the insights from your data analysis to refine your gen AI strategy and optimize performance.

Proprietary + Confidential

Plan for change

Regularly review and refine your strategy

Stay informed

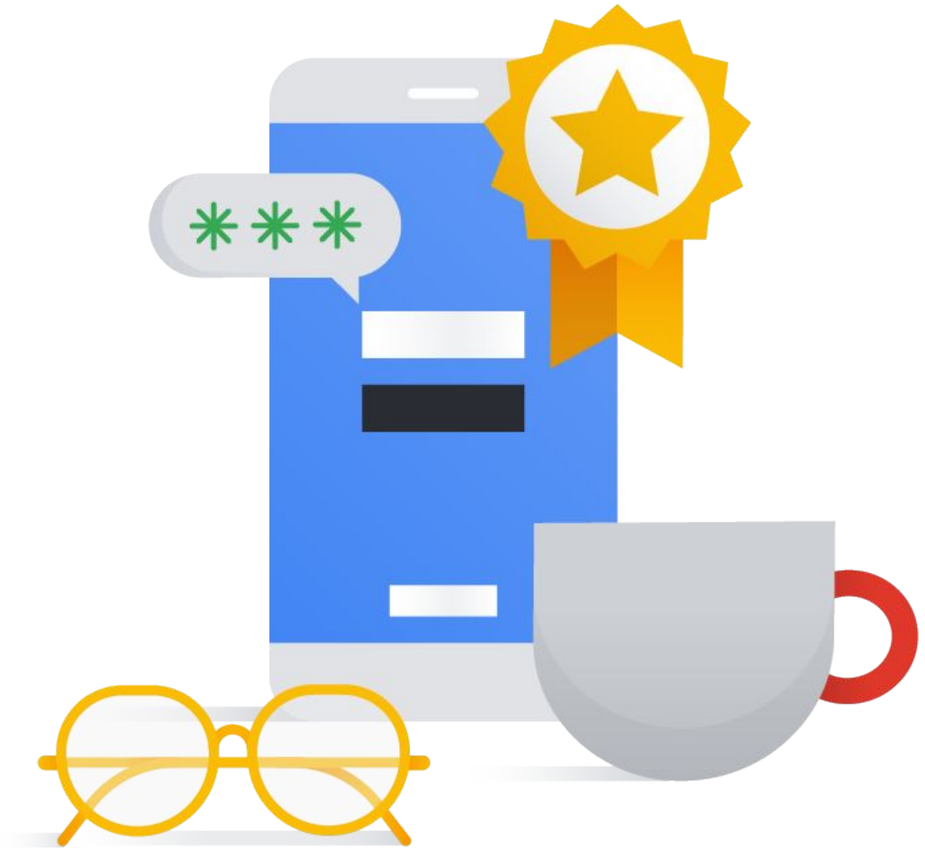
Engage with the generative AI community

Invest in training and development

Attract and retain top talent

Proprietary + Confidential

Now let's do a short quiz to check your knowledge!



Proprietary + Confidential

Quiz | Question 01

Question

How does a large language model (LLM) interact with data stores in a retrieval-augmented generation (RAG) workflow?

A. The LLM queries data stores to retrieve information relevant to a user's request, enriching its response.

B. The LLM uses data stores to permanently expand its base knowledge and parameters.

C. The LLM exclusively relies on data stores for generating responses, disregarding its internal knowledge.

D. The LLM passively receives information from data stores, waiting for updates to be pushed to it.

Proprietary + Confidential

Quiz | Question 01

Answer

How does a large language model (LLM) interact with data stores in a retrieval-augmented generation (RAG) workflow?

A. The LLM queries data stores to retrieve information relevant to a user's request, enriching its response.

B. The LLM uses data stores to permanently expand its base knowledge and parameters.

C. The LLM exclusively relies on data stores for generating responses, disregarding its internal knowledge.

D. The LLM passively receives information from data stores, waiting for updates to be pushed to it.

Proprietary + Confidential

Quiz | Question 02

Question

Which core capability of Vertex AI Search helps mitigate the issue of hallucinations in generative AI?

A. Multimodal search

B. Granular access controls

C. Advanced analytics

D. Grounding

Proprietary + Confidential

Quiz | Question 02

Answer

Which core capability of Vertex AI Search helps mitigate the issue of hallucinations in generative AI?

A. Multimodal search

B. Granular access controls

C. Advanced analytics

D. Grounding

Proprietary + Confidential

Quiz | Question 03

Question

What is the primary purpose of Gemini Enterprise within an organization?

A. To provide a centralized platform for organizing and accessing AI agents that can utilize company data to assist employees.

B. To enhance the organization's cybersecurity defenses by detecting and preventing data breaches.

C. To create a centralized repository for storing and managing all company data.

D. To provide a customer-facing interface for interacting with AI-powered chatbots.

Proprietary + Confidential

Quiz | Question 03

Answer

What is the primary purpose of Gemini Enterprise within an organization?

A. To provide a centralized platform for organizing and accessing AI agents that can utilize company data to assist employees.

B. To enhance the organization's cybersecurity defenses by detecting and preventing data breaches.

C. To create a centralized repository for storing and managing all company data.

D. To provide a customer-facing interface for interacting with AI-powered chatbots.

Proprietary + Confidential

Quiz | Question 04

Question

What are the key steps in planning for impact when implementing generative AI solutions?

A. Prioritize AI capabilities without considering ethical implications.

B. Focus on technology adoption and assume positive impact will follow.

C. Collect and analyze data with a focus on short-term goals in the beginning.

D. Define key metrics, collect and analyze data, and iterate and improve.

Proprietary + Confidential

Quiz | Question 04

Answer

What are the key steps in planning for impact when implementing generative AI solutions?

A. Prioritize AI capabilities without considering ethical implications.

B. Focus on technology adoption and assume positive impact will follow.

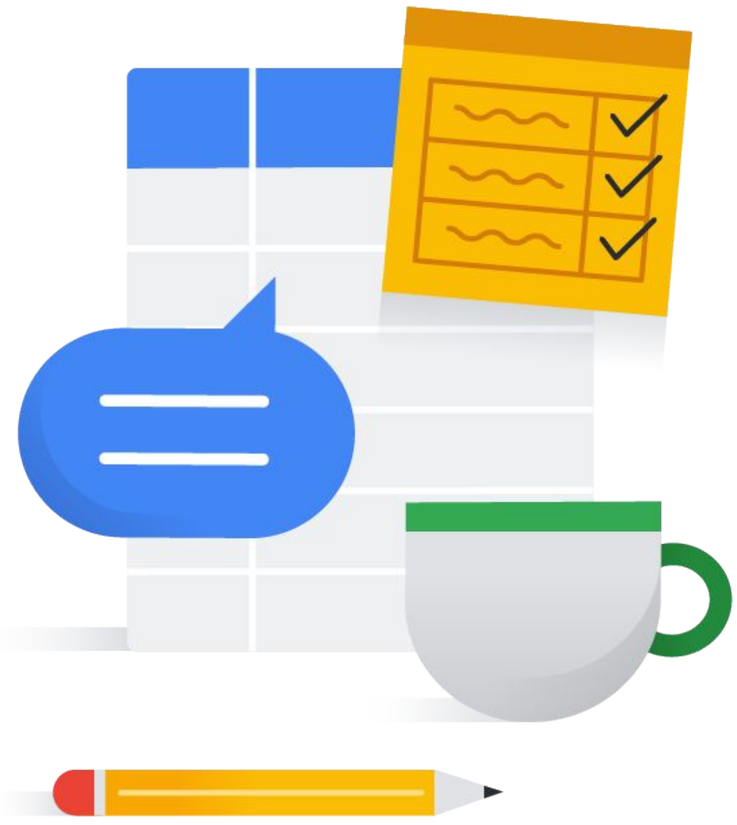
C. Collect and analyze data with a focus on short-term goals in the beginning.

D. Define key metrics, collect and analyze data, and iterate and improve.

Proprietary + Confidential

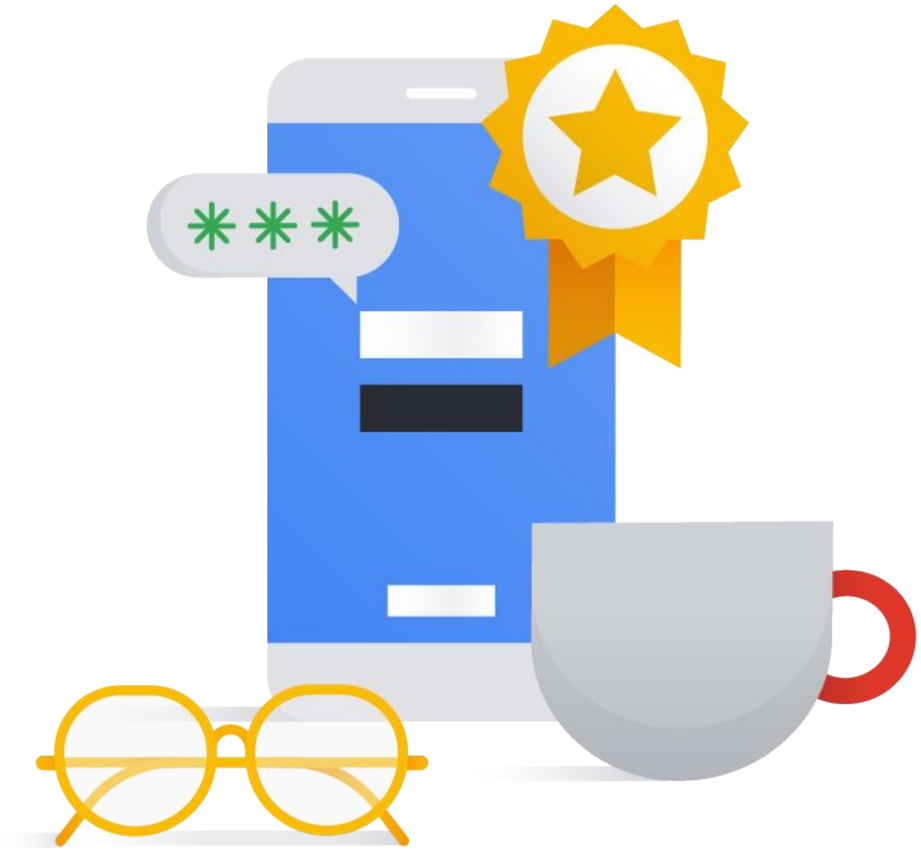
Key takeaways

- Data stores are key for reliable AI agents, managing knowledge for accurate and relevant responses using RAG.
- Integrating Vertex AI Search empowers customers to find and understand information faster for better decisions and efficiency.
- Customer Engagement Suite offers powerful agents using generative AI and other tech for seamless customer experiences.
- Gemini Enterprise is a central, secure platform for customized AI work assistants, enhancing productivity across various business teams.



Proprietary + Confidential

Now let's wrap up with a quiz to check your knowledge on Module 05.



Proprietary + Confidential

Quiz | Question 01

Question

Which of the following options best identifies the key components of an AI agent?

A. A complex algorithm, a user interface, and a data storage system

B. A reasoning loop, tools, and a foundational model

C. API connections, cloud functions, and a sophisticated database

D. A detailed user manual, customer service chatbots, and advanced marketing techniques

Proprietary + Confidential

Quiz | Question 01

Answer

Which of the following options best identifies the key components of an AI agent?

A. A complex algorithm, a user interface, and a data storage system

B. A reasoning loop, tools, and a foundational model

C. API connections, cloud functions, and a sophisticated database

D. A detailed user manual, customer service chatbots, and advanced marketing techniques

Proprietary + Confidential

Quiz | Question 02

Question

Which of the following scenarios best illustrates the concept of a multi-agent system in the context of generative AI?

A. A single AI agent capable of performing multiple tasks, such as writing emails and summarizing documents.

B. An AI agent that uses multiple tools, like a database and a web search API, to complete a task.

C. A system where multiple specialized AI agents collaborate to achieve a complex goal, such as booking a trip.

D. An AI agent that can be customized with different parameters to adapt to different tasks.

Proprietary + Confidential

Quiz | Question 02

Answer

Which of the following scenarios best illustrates the concept of a multi-agent system in the context of generative AI?

A. A single AI agent capable of performing multiple tasks, such as writing emails and summarizing documents.

B. An AI agent that uses multiple tools, like a database and a web search API, to complete a task.

C. A system where multiple specialized AI agents collaborate to achieve a complex goal, such as booking a trip.

D. An AI agent that can be customized with different parameters to adapt to different tasks.

Proprietary + Confidential

Quiz | Question 03

Question

What is a key advantage of using a hybrid approach (combining deterministic and generative AI) when building conversational agents with Google Cloud's Customer Engagement Suite?

A. It allows agents to handle only simple, predictable customer queries.

B. It eliminates the need for human agents in customer interactions.

C. It enables strict control over agent behavior while leveraging the flexibility of generative AI.

D. It restricts agent responses to predefined scripts, ensuring consistency.

Proprietary + Confidential

Quiz | Question 03

Answer

What is a key advantage of using a hybrid approach (combining deterministic and generative AI) when building conversational agents with Google Cloud's Customer Engagement Suite?

A. It allows agents to handle only simple, predictable customer queries.

B. It eliminates the need for human agents in customer interactions.

C. It enables strict control over agent behavior while leveraging the flexibility of generative AI.

D. It restricts agent responses to predefined scripts, ensuring consistency.

Proprietary + Confidential

Module objectives

01 Define gen AI agent components and how they work together.

02 Explain how to combine components of gen AI agents to build powerful solutions.

03 Describe how to create gen AI agents and solutions using Google Cloud's gen AI products.

04 Explain how to effectively lead organizational transformation with gen AI.

Proprietary + Confidential

Additional resources

Lesson 02

- Comparison of Vertex AI Studio and Google AI Studio
- Takeaway try it - Parameter playground in Google AI Studio
- API Library
- Set up authentication & authorization
- Generate an API key

Lesson 03

- Takeaway try it - RAG in action
- Takeaway try it - Build your own agent

Course study guide

Proprietary + Confidential

Module 05 complete!

Course summary

Generative AI Leader

Proprietary + Confidential

Course objectives

01 Discover the business value and impact of Gen AI on your organization.

02 Define core gen AI concepts.

03 Identify the core layers of the gen AI landscape.

04 Explain how to combine components of gen AI agents to build powerful solutions.

05 Identify Google Cloud's gen AI implementation and scaling solutions.

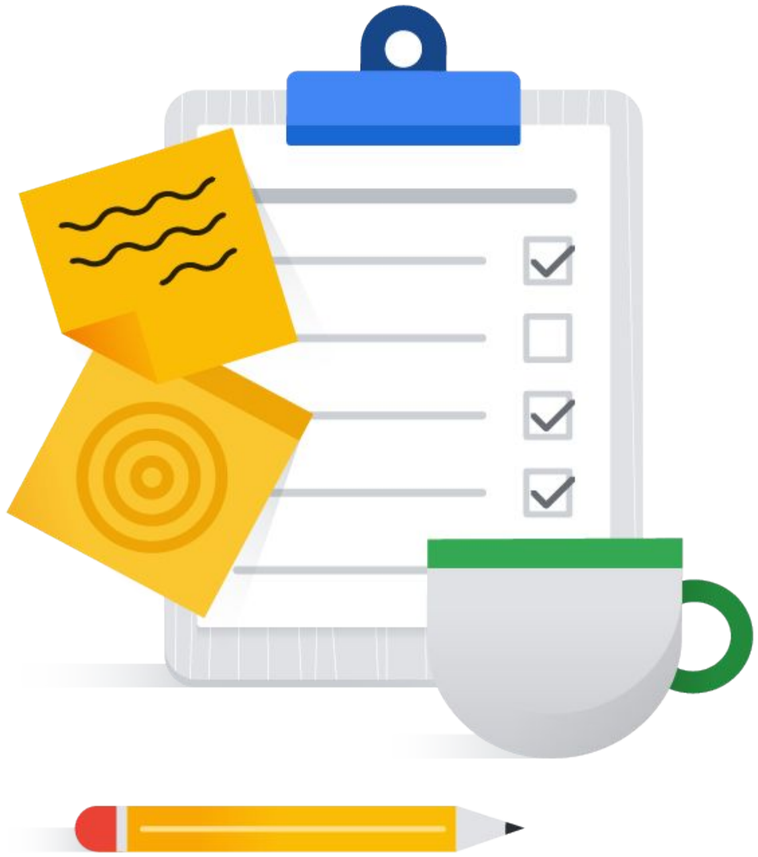
Proprietary + Confidential

Module 01 - Gen AI: Beyond the chatbot

01 Introduction to gen AI for businesses

02 Introduction to gen AI foundations

03 Gen AI strategy



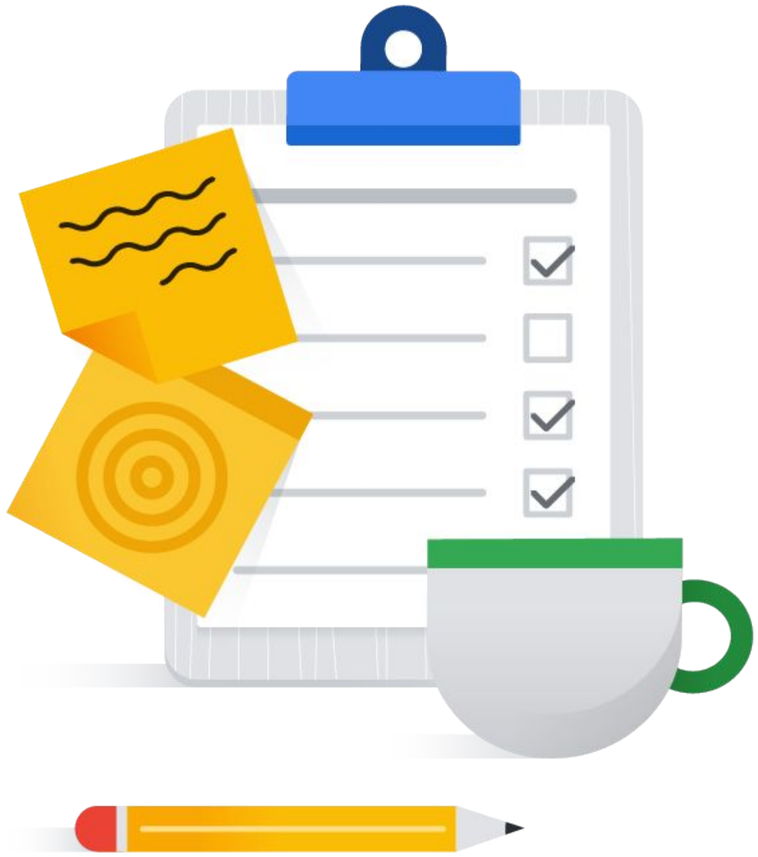
Proprietary + Confidential

Module 02 - Gen AI: Unlock foundational concepts

01 Core gen AI concepts

02 Foundation models

03 Building AI securely and responsibly



Proprietary + Confidential

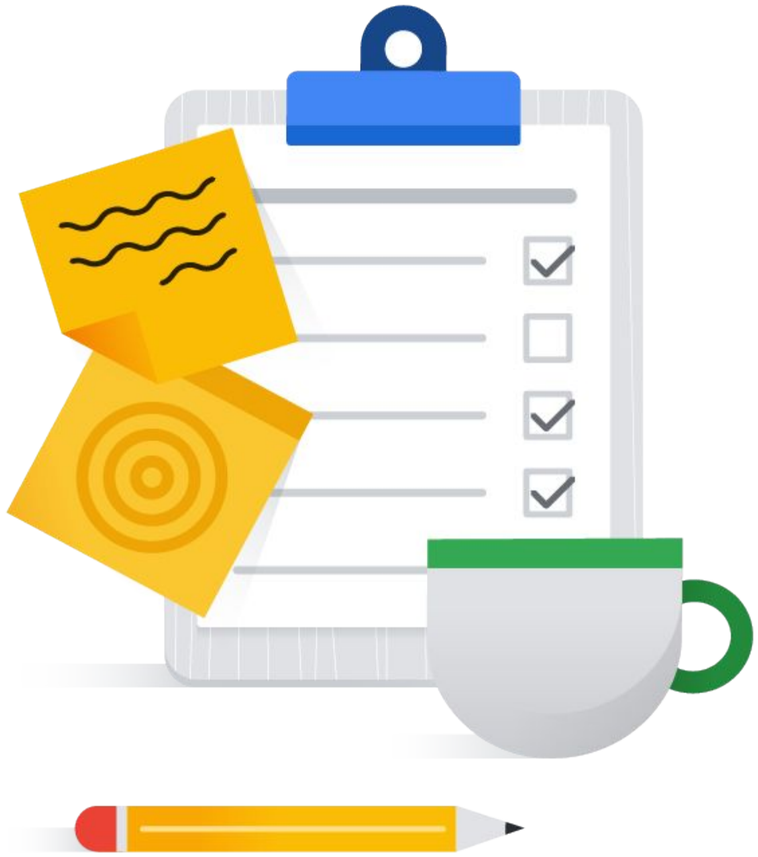
Module 03 - Gen AI: Navigate the landscape

01 The gen AI landscape

02 Gen AI agents and applications

03 Gen AI platform, model, and infrastructure

04 Gen AI project resources and management



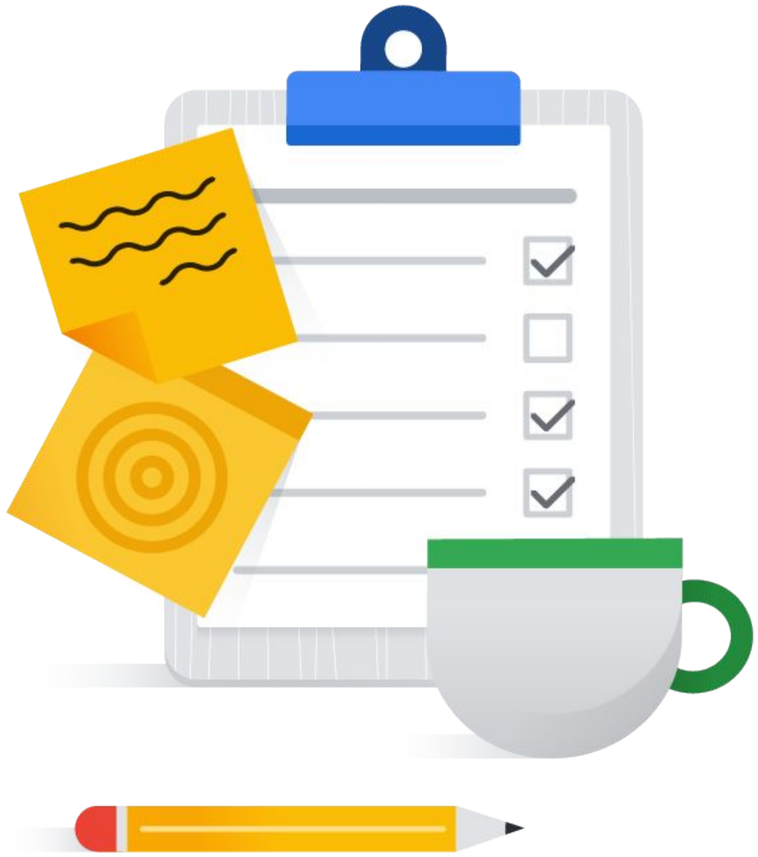
Proprietary + Confidential

Module 04 - Gen AI apps: Transform your work

01 Prompting techniques

02 Gen AI for productivity

03 Gemini for Google Cloud



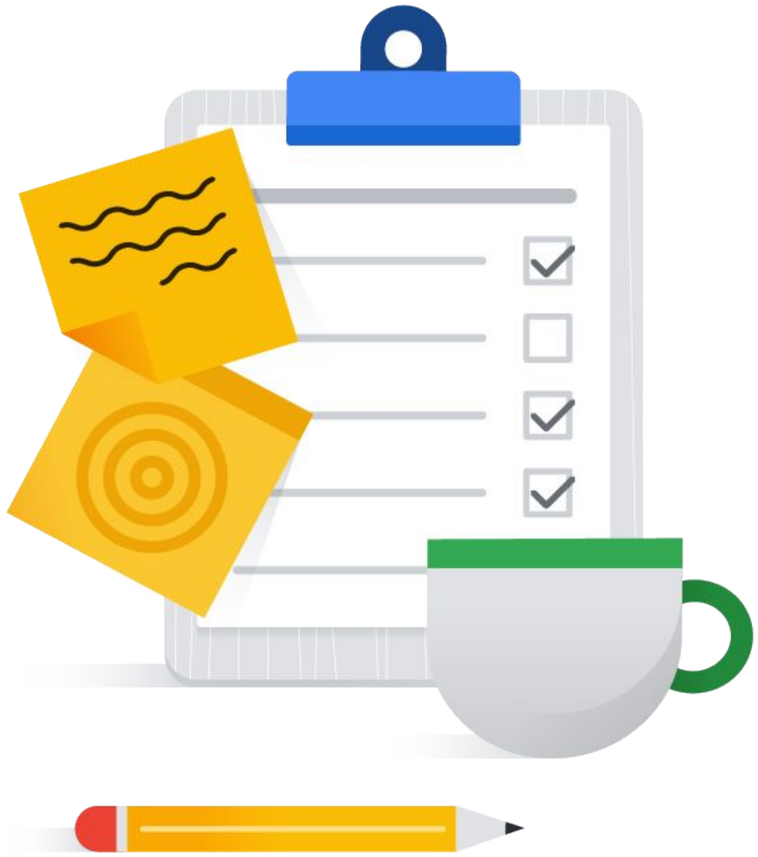
Proprietary + Confidential

Module 05 - Gen AI agents: Transform your organization

01 Today's agents

02 Building agents

03 Enhancing the customer experience with agents



Proprietary + Confidential

Generative AI Leader certification exam

A Google Cloud Certified **Generative AI Leader** is a visionary professional with comprehensive business knowledge of how generative AI can transform businesses using Google Cloud's gen AI products and services. They understand Google's AI-first approach and can drive innovative and responsible AI adoption by identifying use cases, fostering collaboration across technical and non-technical teams, and leveraging Google Cloud's enterprise offerings.

- Length: 90 minutes
- Exam format: 50-60 multiple choice questions
- For more info visit: cloud.google.com/learn/certification/generative-ai-leader

Proprietary + Confidential

What's next?

Prepare for the certification exam

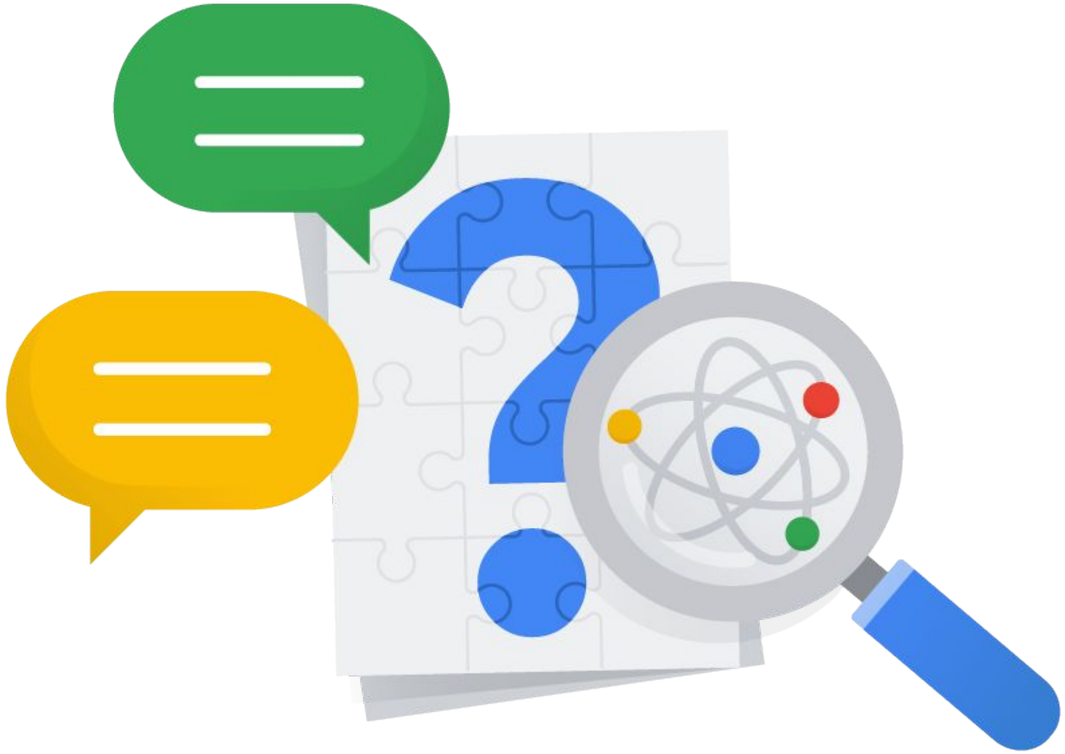
- Use resources and information from: http://cloud.google.com/learn/certification/generative-ai-leader
- Access the course study guide
- Revisit the **Generative AI Leader** material from this ILT course, or from the on-demand format: https://www.cloudskillsboost.google/paths/1951

Continue your learning journey

- Explore more and go deeper with generative AI with other learning paths and courses: https://www.cloudskillsboost.google/
- Gain hands-on experience with **Intermediate: Generative AI Labs with Gemini on Google Cloud**: https://www.cloudskillsboost.google/paths/1872
- Improve your productivity and efficiency with **Google Workspace with Gemini**: https://www.cloudskillsboost.google/paths/249

Proprietary + Confidential

Questions and answers



Proprietary + Confidential

Thank you for attending this training!

We love your feedback! Please take a minute to complete the survey and help us improve our courses.