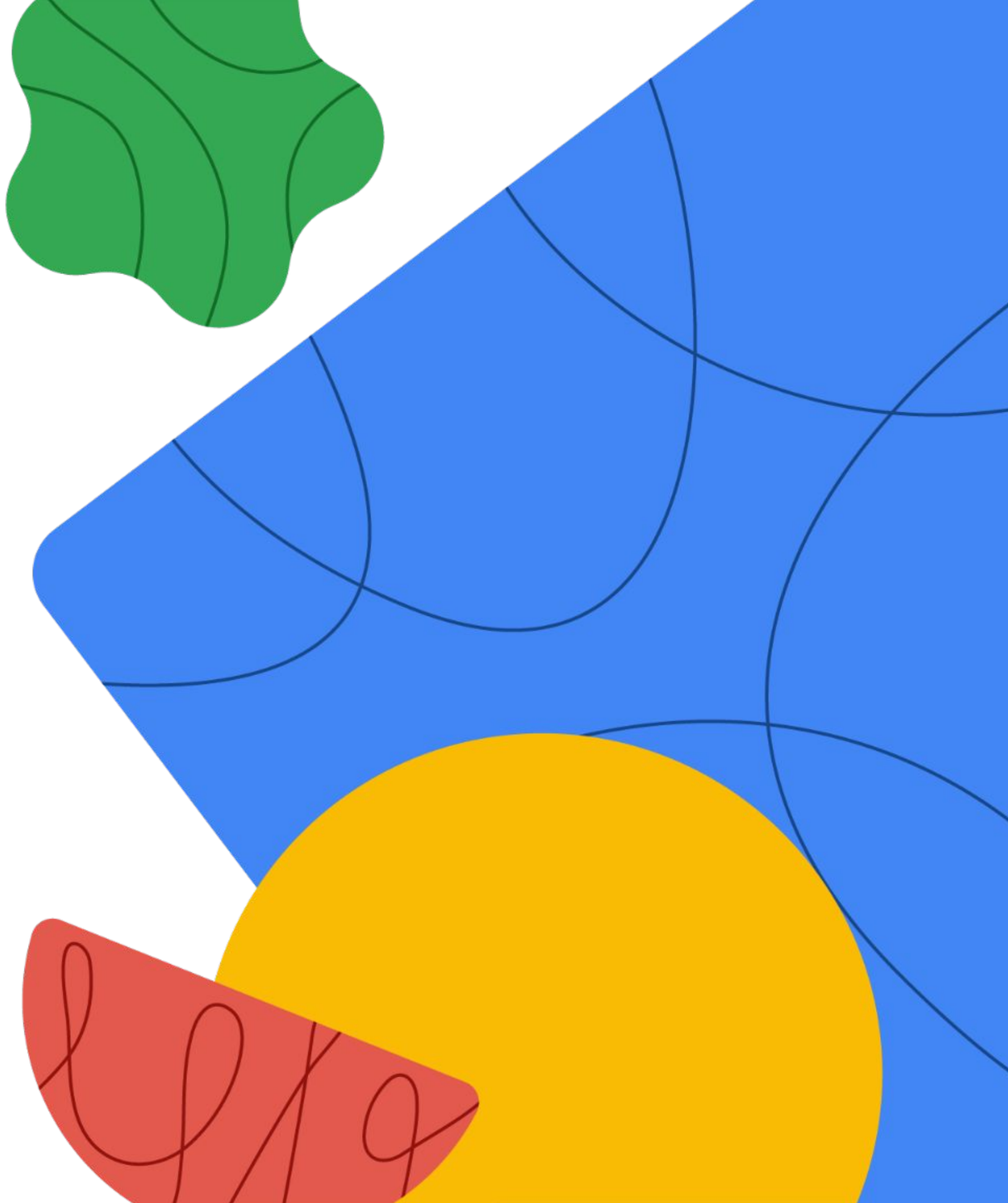
Google Cloud

Module 03
Gen AI: Navigate the landscape



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

Module objectives

01 Describe the layers of the gen AI landscape.

02 Identify entry points in the gen AI landscape to address business needs and innovation.

03 Describe components of the Google Cloud gen AI portfolio.

04 Explain how Google Cloud's AI-optimized resources support gen AI development.

05 Describe business factors to consider for specific applications.

Agenda

01 The gen AI landscape

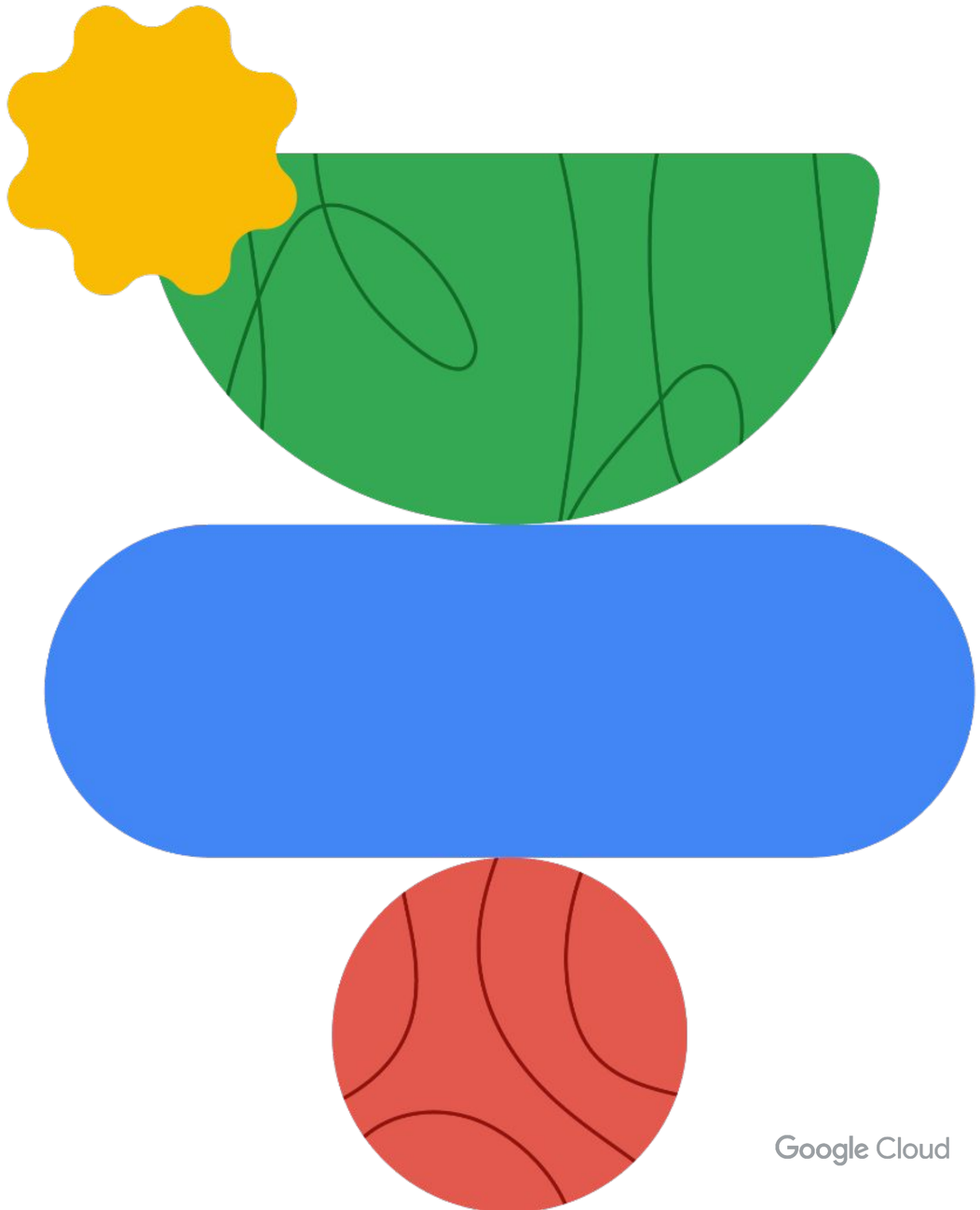
02 Gen AI applications and agents

03 Gen AI platform, models, and infrastructure

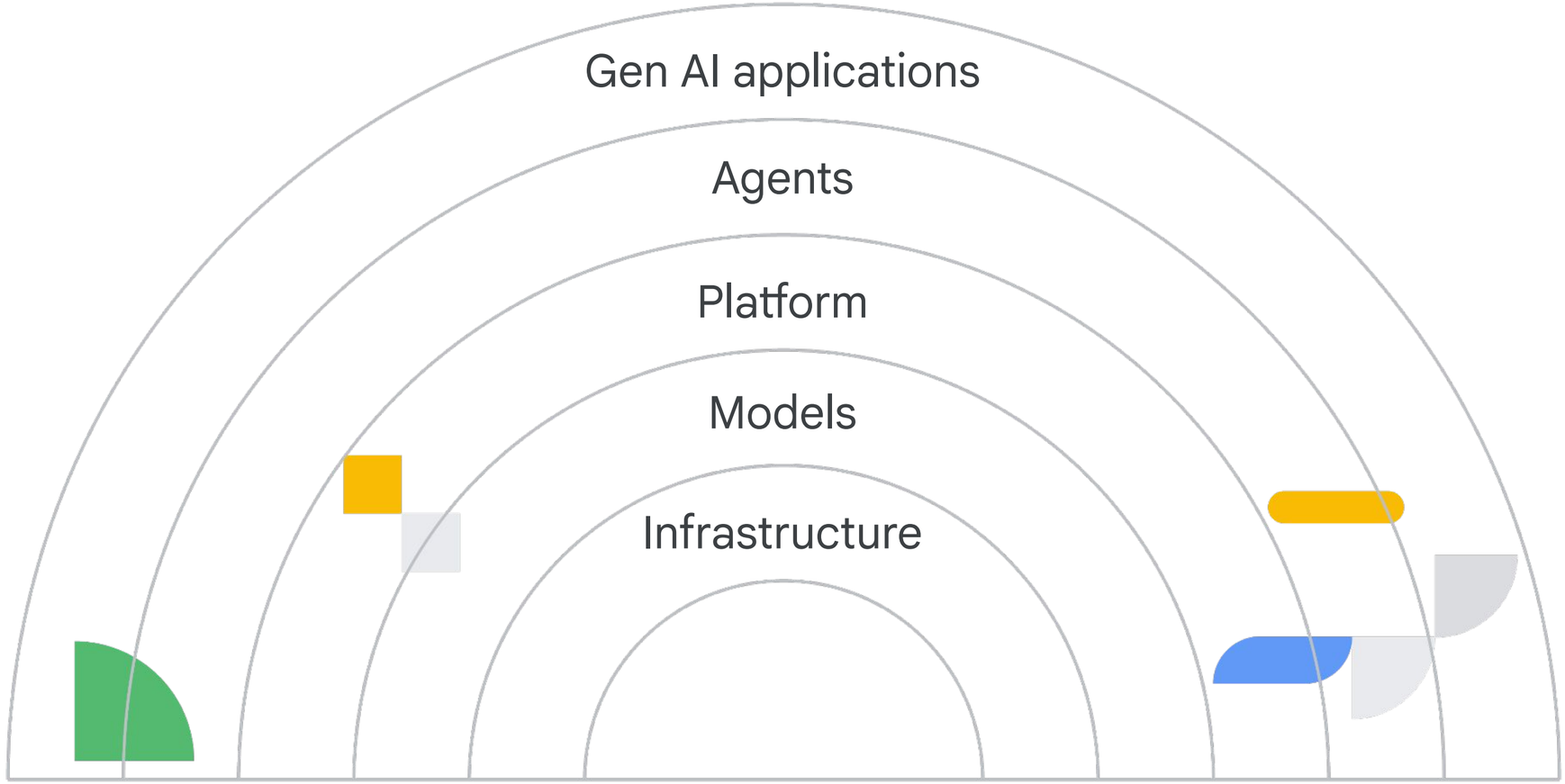
04 Gen AI project resources and management

The gen AI landscape

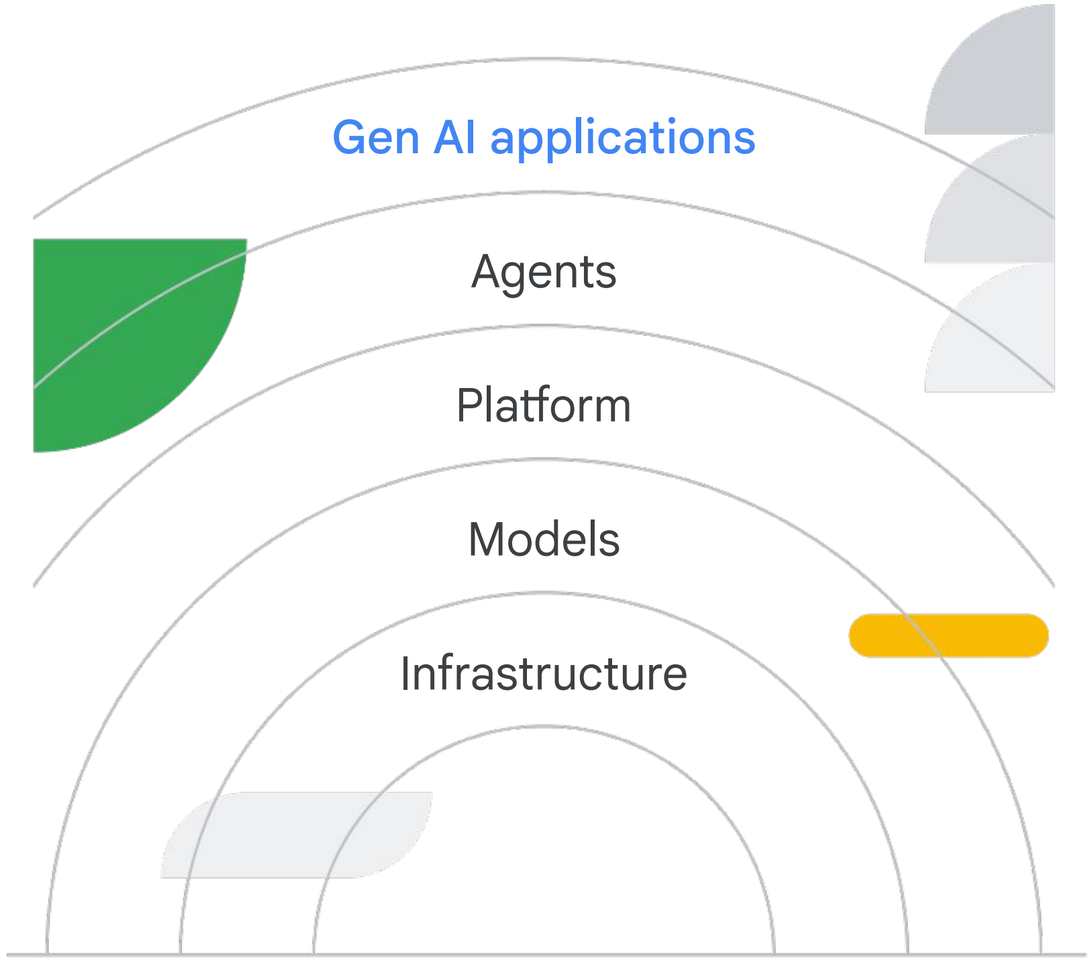
Understanding the layers of gen AI



Building blocks of generative AI

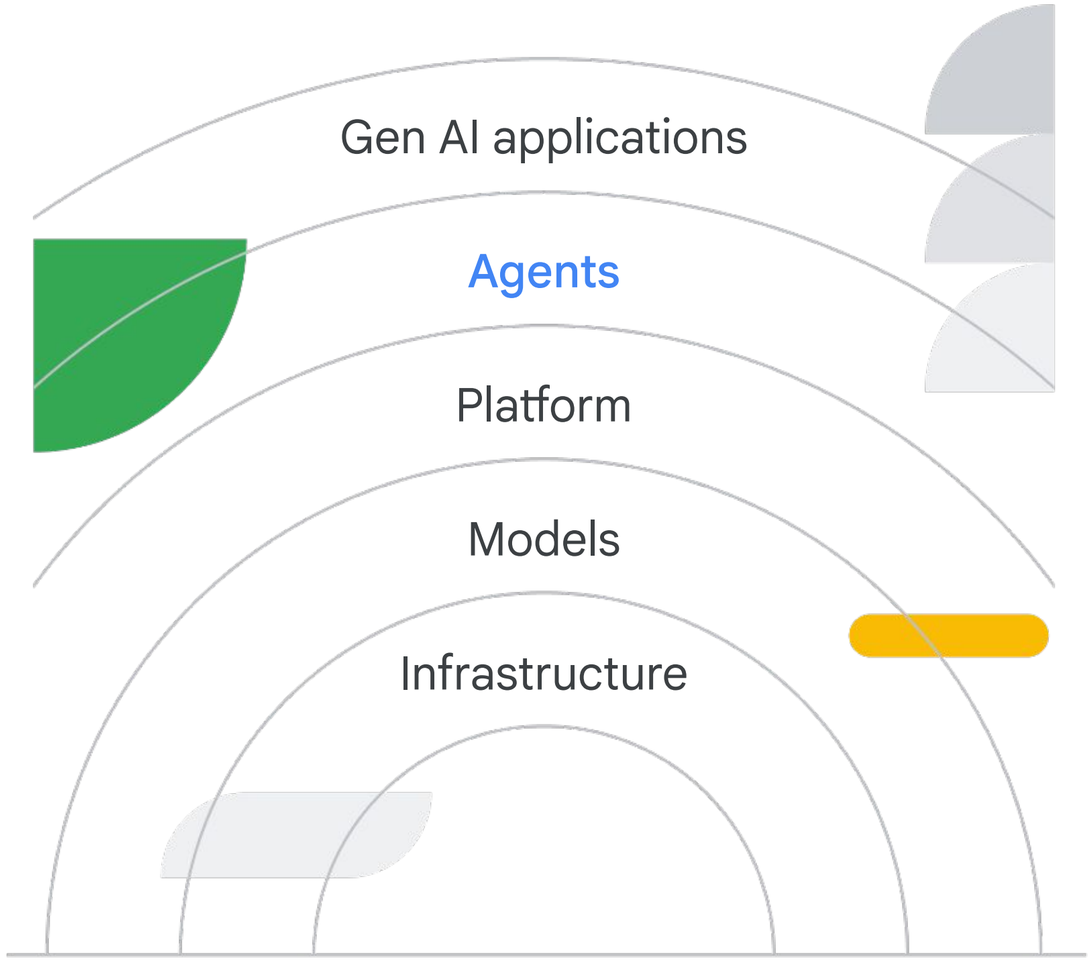


Building blocks of generative AI: Gen AI applications



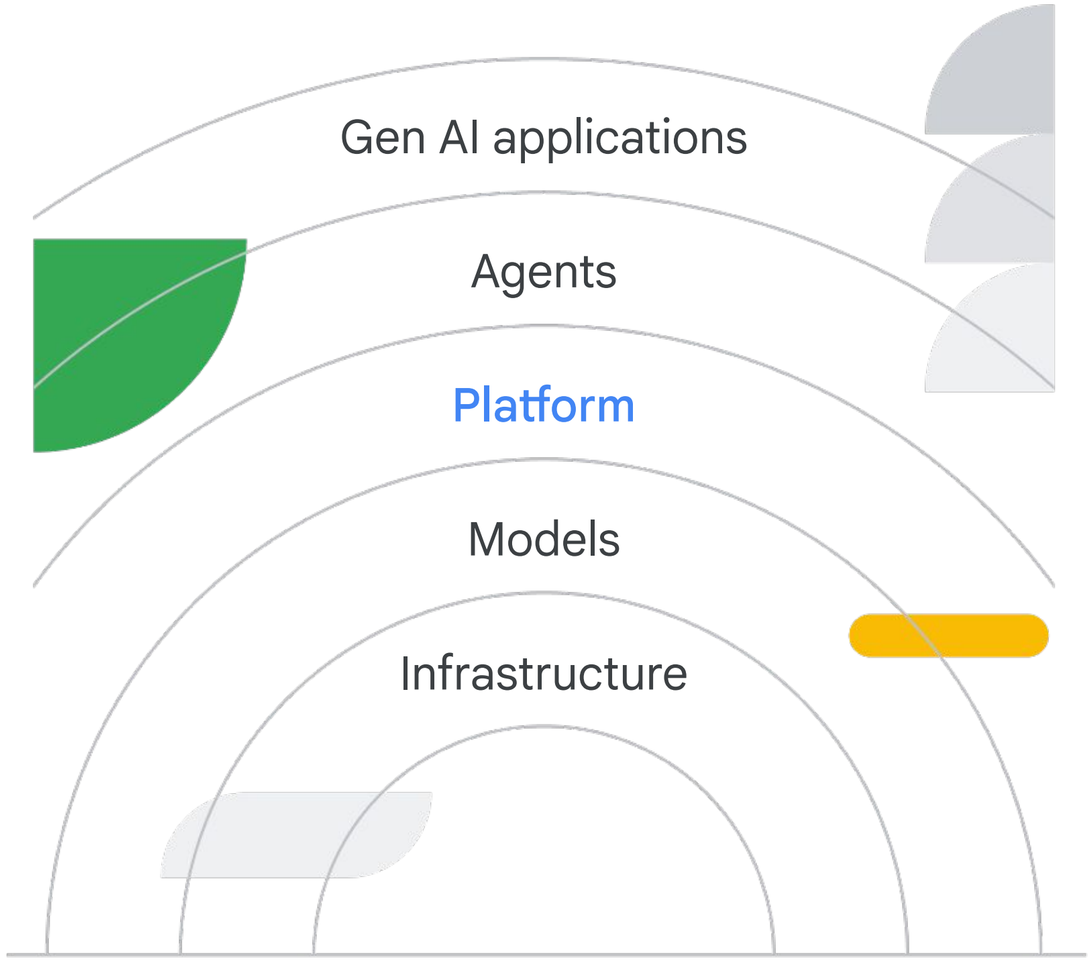
- A gen AI application is the user-facing part of generative AI (frontend).
- It allows users to interact with and leverage the capabilities of AI.
- Some examples are the Gemini app, Google Workspace with Gemini, and NotebookLM.

Building blocks of generative AI: Agents



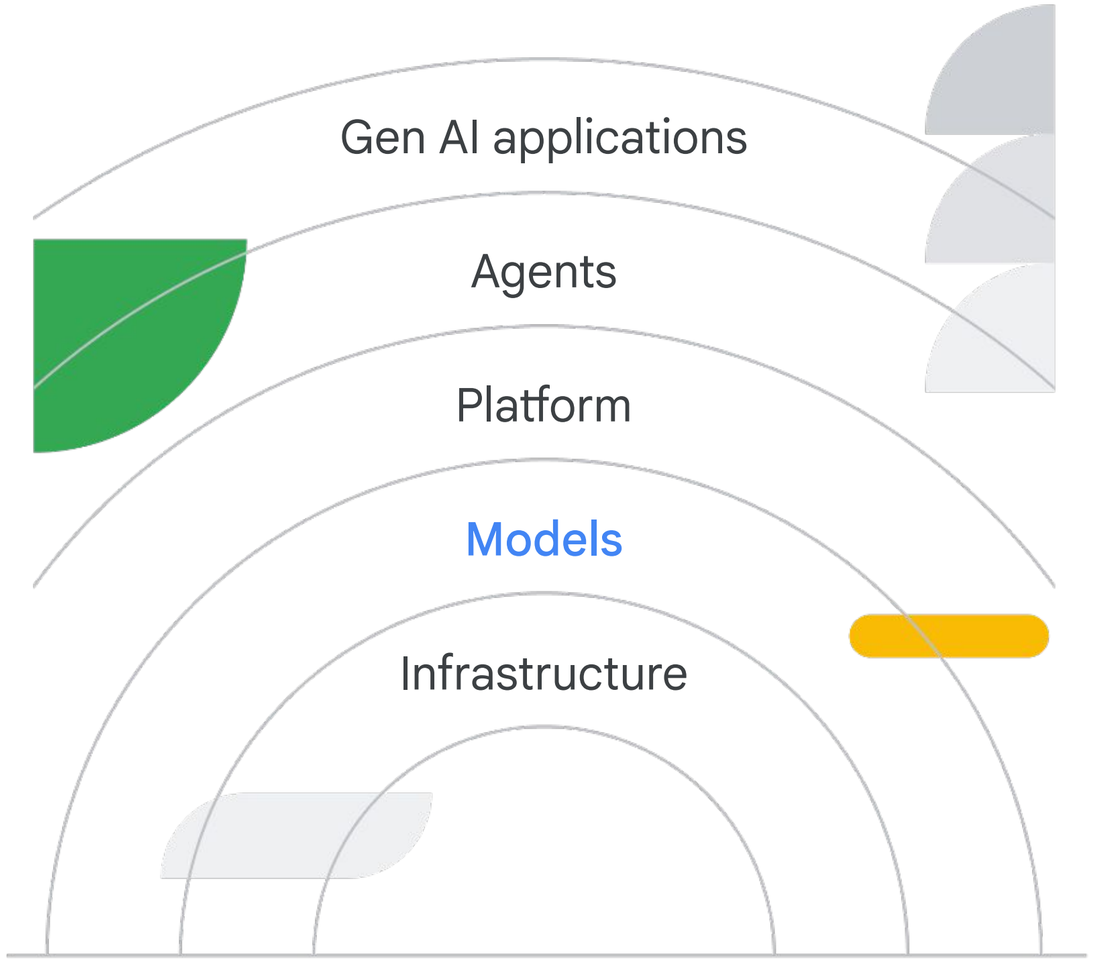
- An agent is a piece of software that learns how to best achieve a goal based on inputs and tools available to it.
- It focuses on autonomous action.
- Examples are customer agents, code agents, data agents, etc.

Building blocks of generative AI: Platform



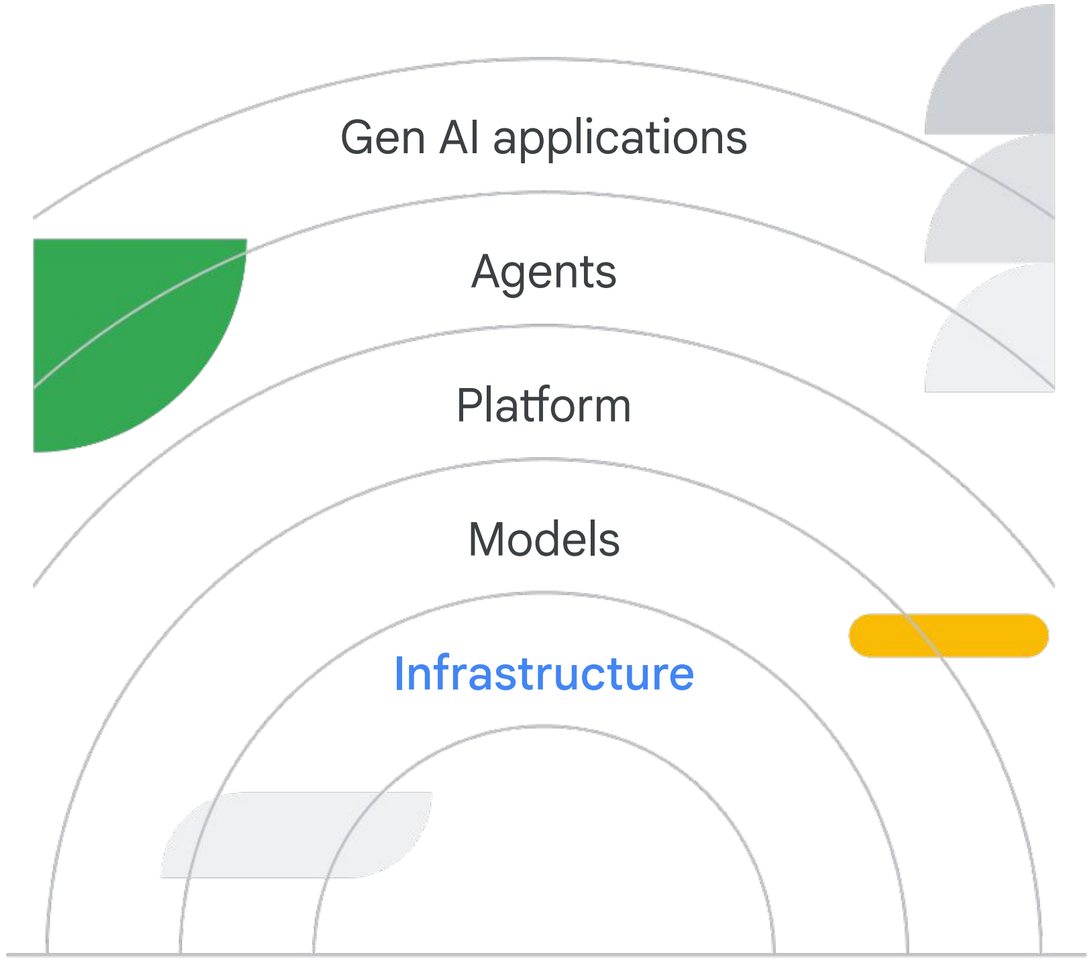
- A platform provides tools and services for agents and models to interact.
- It offers APIs, data management capabilities, and model deployment tools.

Building blocks of generative AI: Models



- The AI model is the brain of the agent.
- It is a complex algorithm trained on vast amounts of data.
- It generates new content, translates languages, answers questions, etc.
- Examples are large language models (LLMs), image recognition models, and recommendation systems.

Building blocks of generative AI: Infrastructure

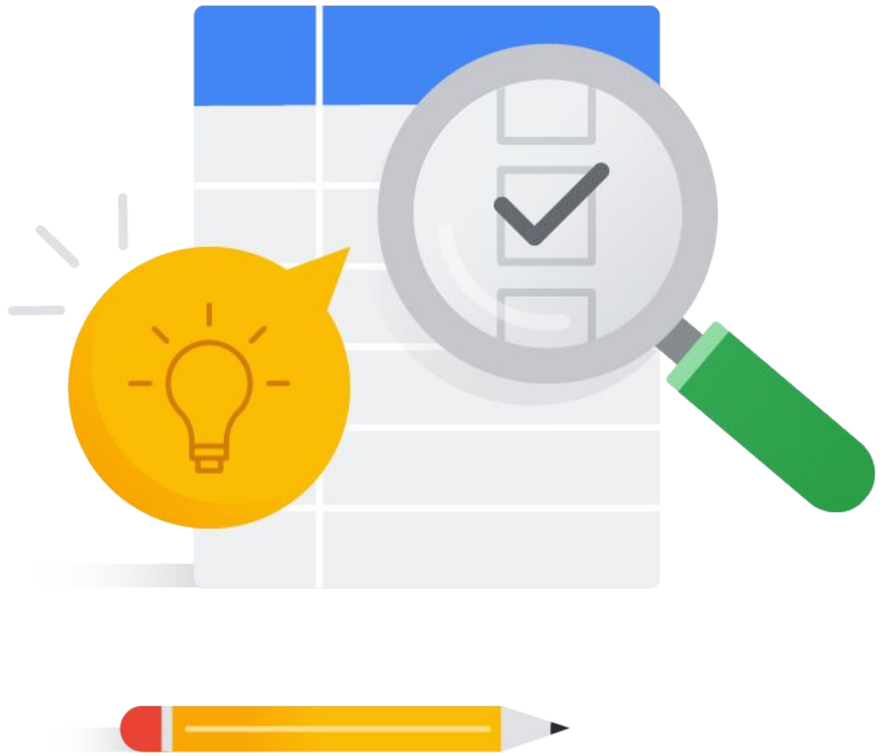


- The infrastructure provides the core computing resources needed for generative AI.
- It includes the hardware and software needed to store and run AI models and training data.

Activity: Layers of gen AI

5 min

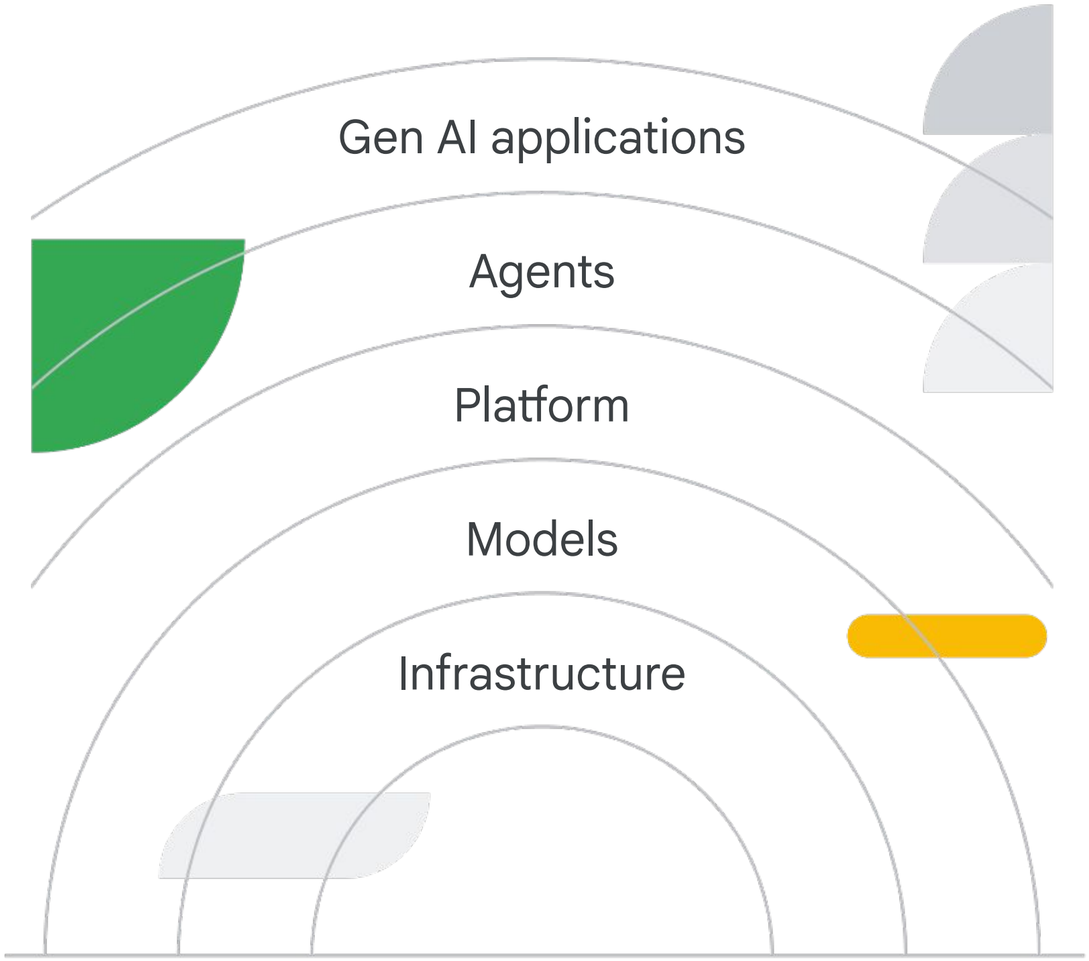
1. Read the statements.
2. Identify which layer each statement describes.



Building blocks of generative AI: Activity

Read the statement below, and identify which layer it describes.

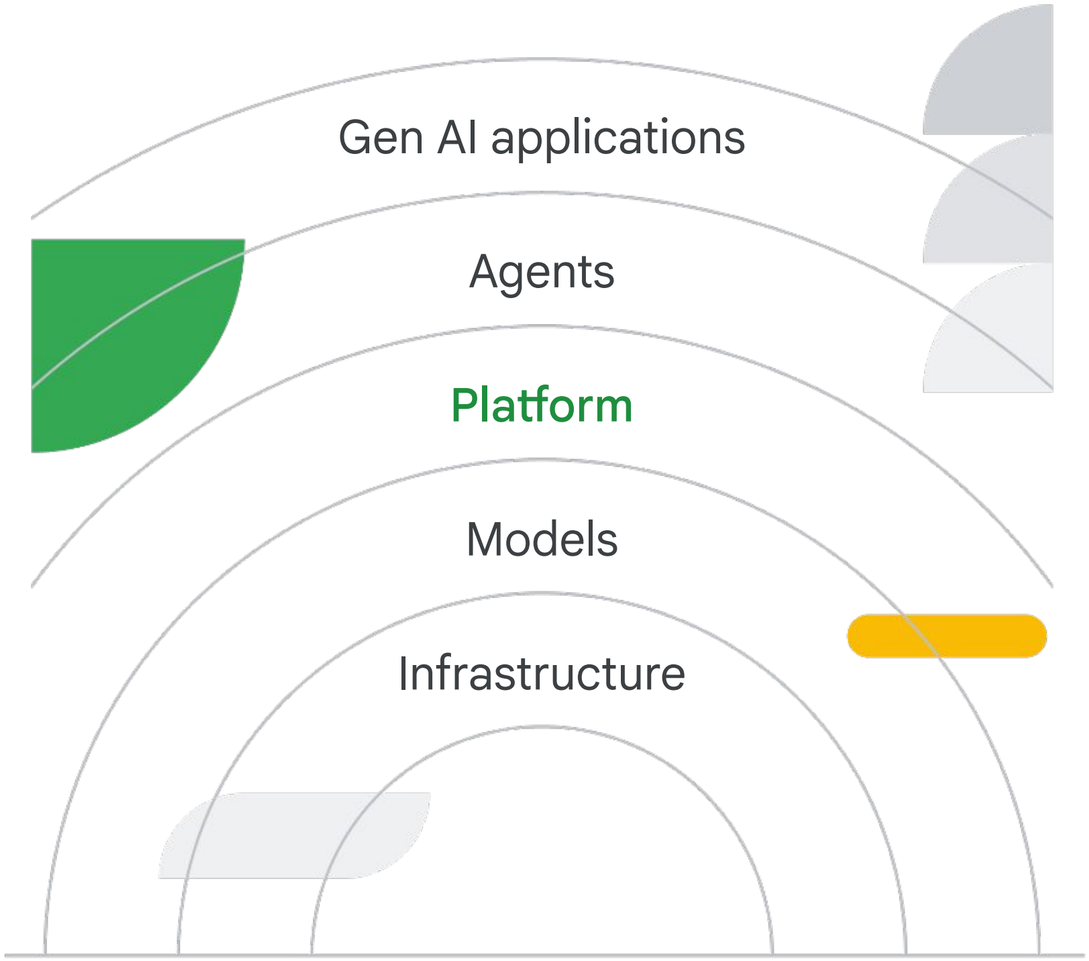
1. A team of data scientists needs a set of tools to label their training data, experiment with different neural network architectures, and deploy their finalized AI model into a production environment.



Building blocks of generative AI: Activity

Read the statement below, and identify which layer it describes.

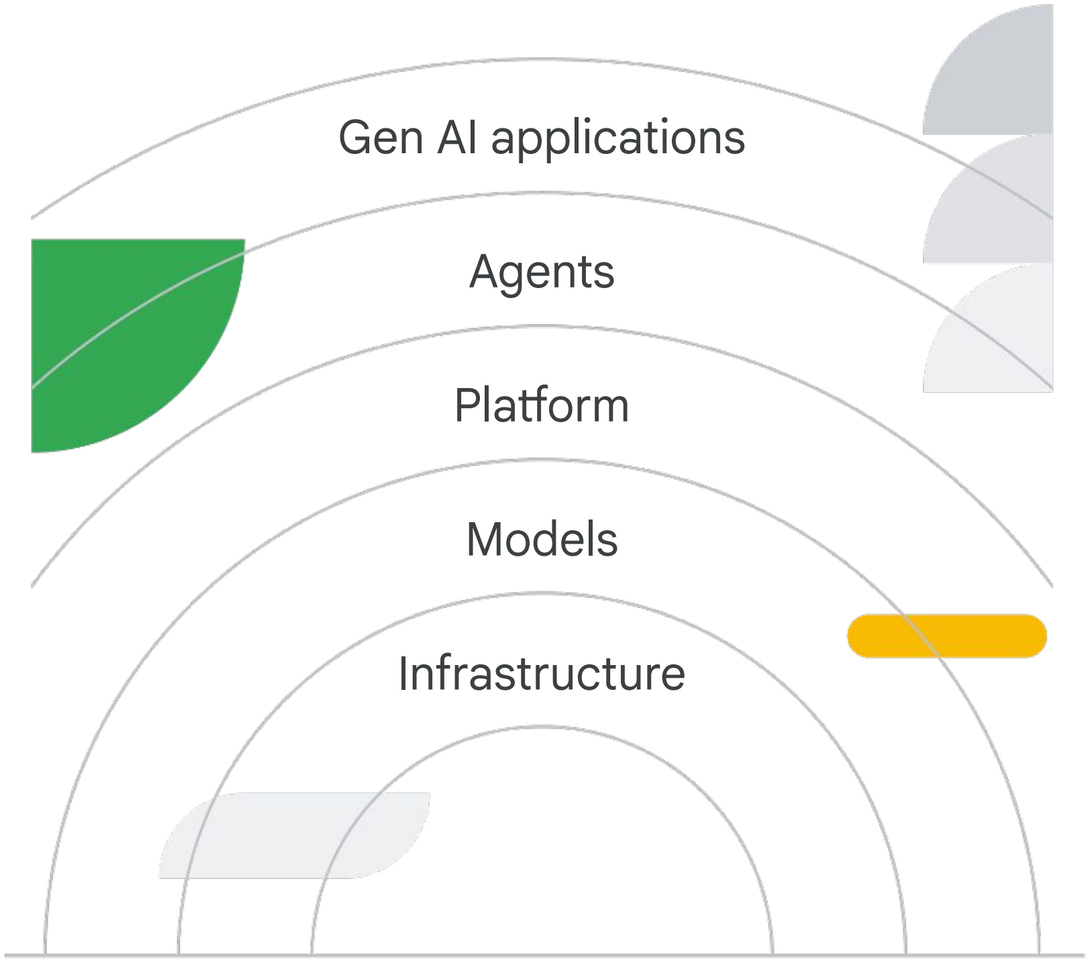
1. A team of data scientists needs a set of tools to label their training data, experiment with different neural network architectures, and deploy their finalized AI model into a production environment.



Building blocks of generative AI: Activity

Read the statement below, and identify which layer it describes.

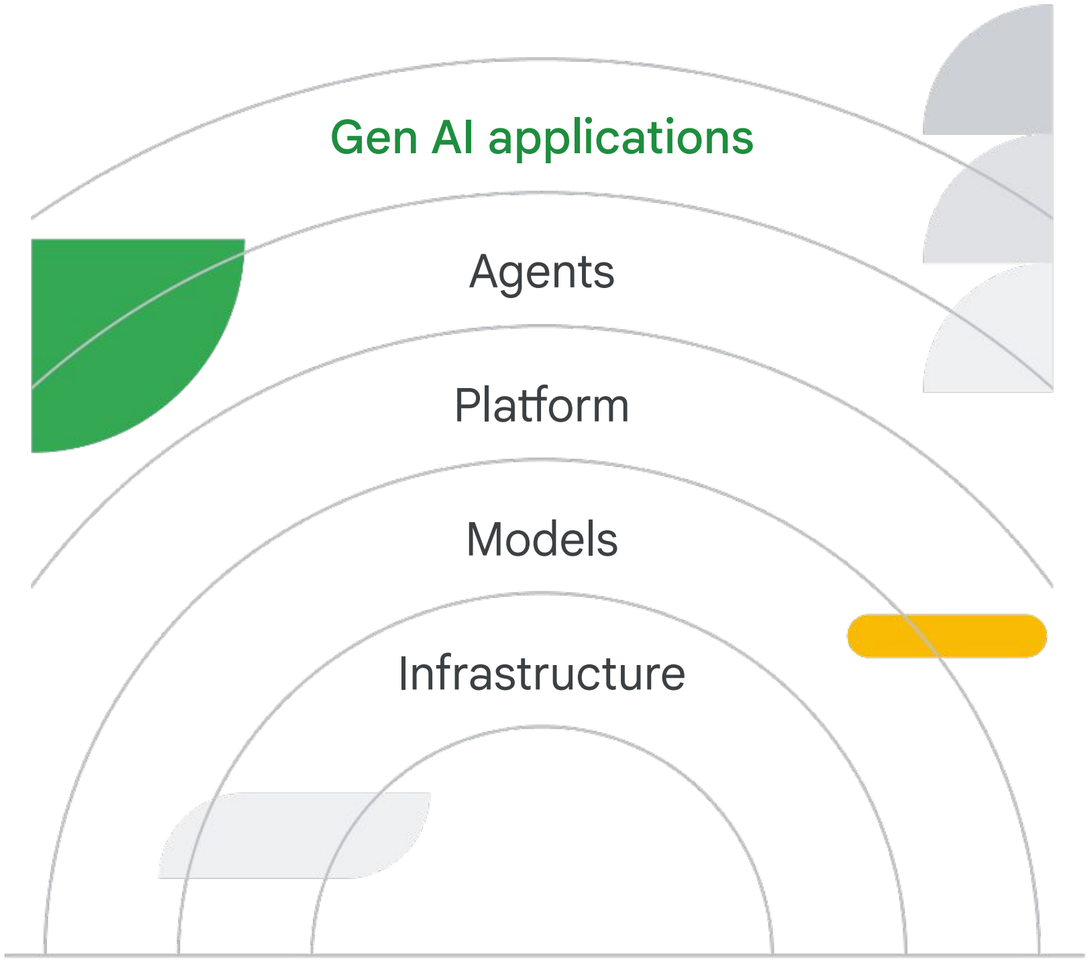
2. A customer is using a website where they can type in a description of an image they want. The website instantly generates that image for them.



Building blocks of generative AI: Activity

Read the statement below, then identify which layer it describes.

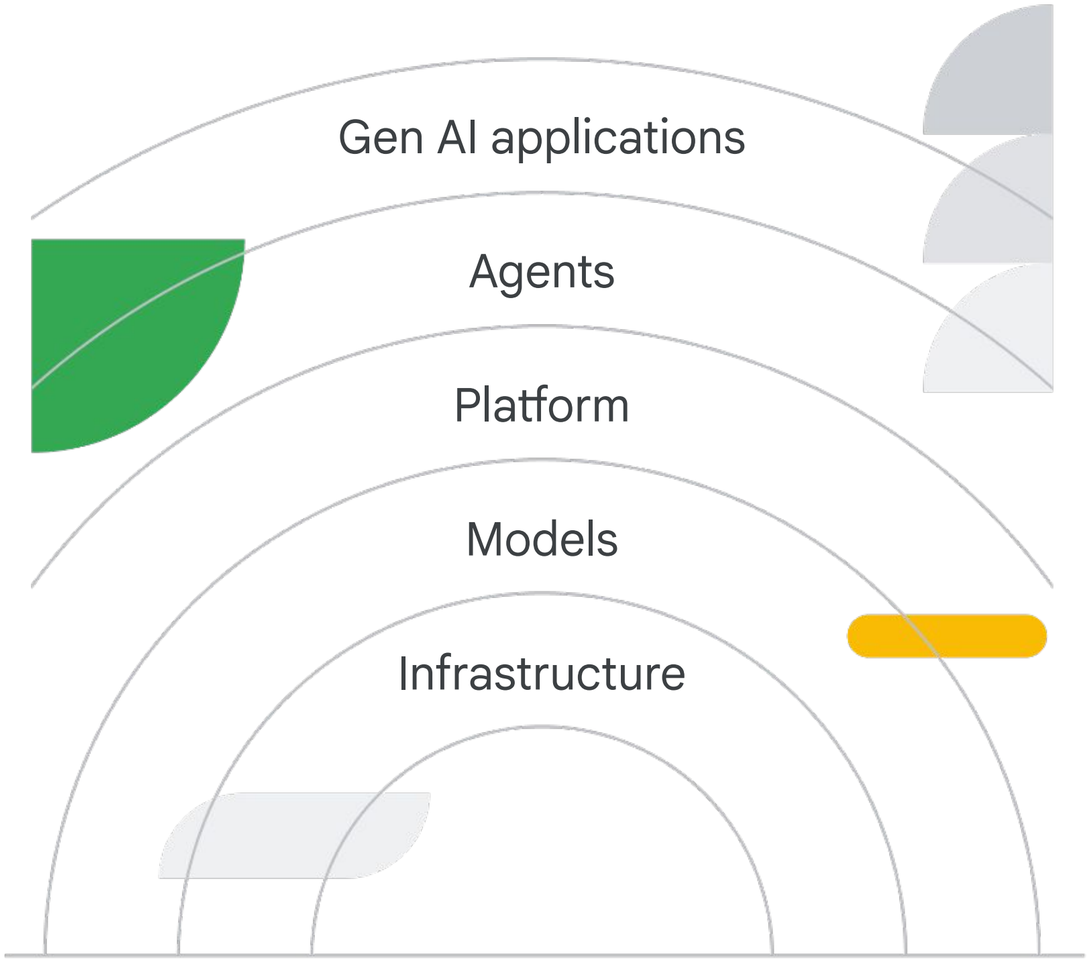
2. A customer is using a website where they can type in a description of an image they want. The website instantly generates that image for them.



Building blocks of generative AI: Activity

Read the statement below, and identify which layer it describes.

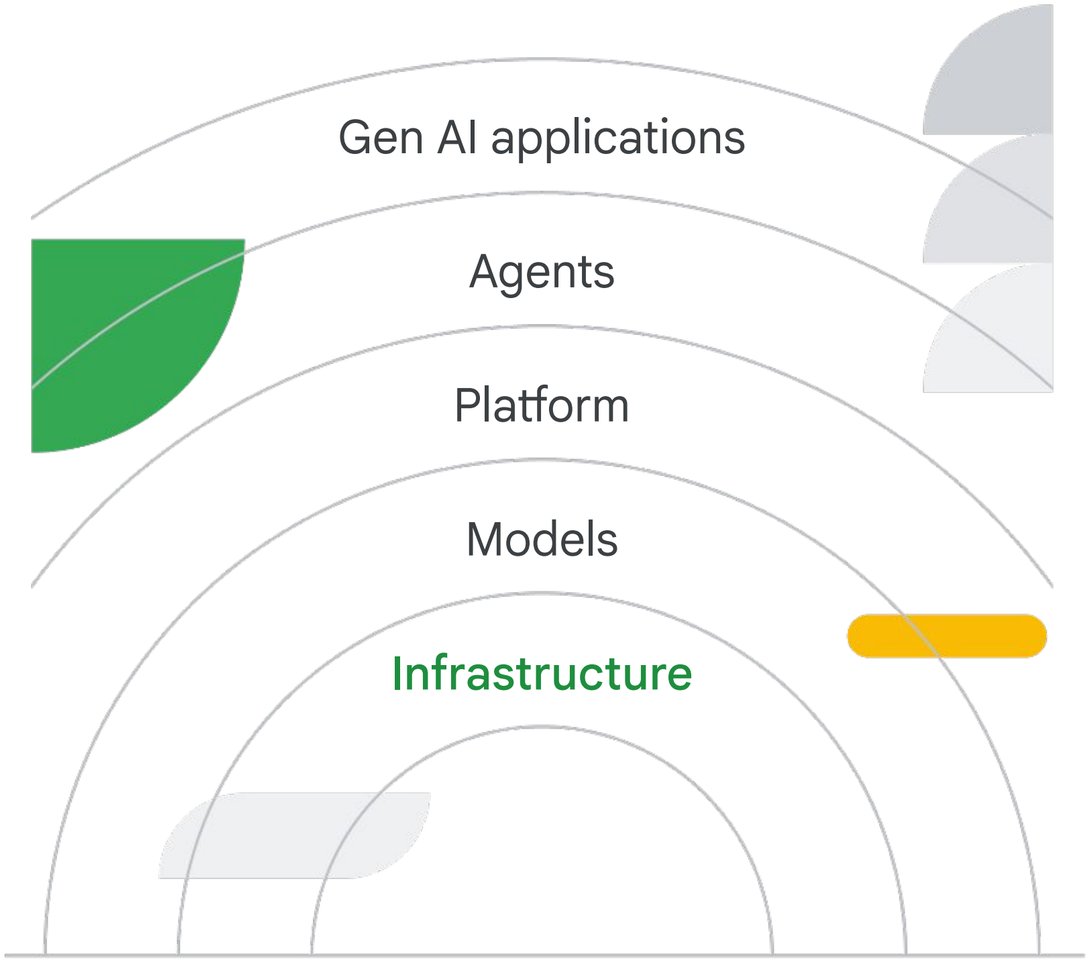
3. A company needs to significantly increase the processing power available to train a very large language model. This requires more specialized processors and high-speed data connections.



Building blocks of generative AI: Activity

Read the statement below, and identify which layer it describes.

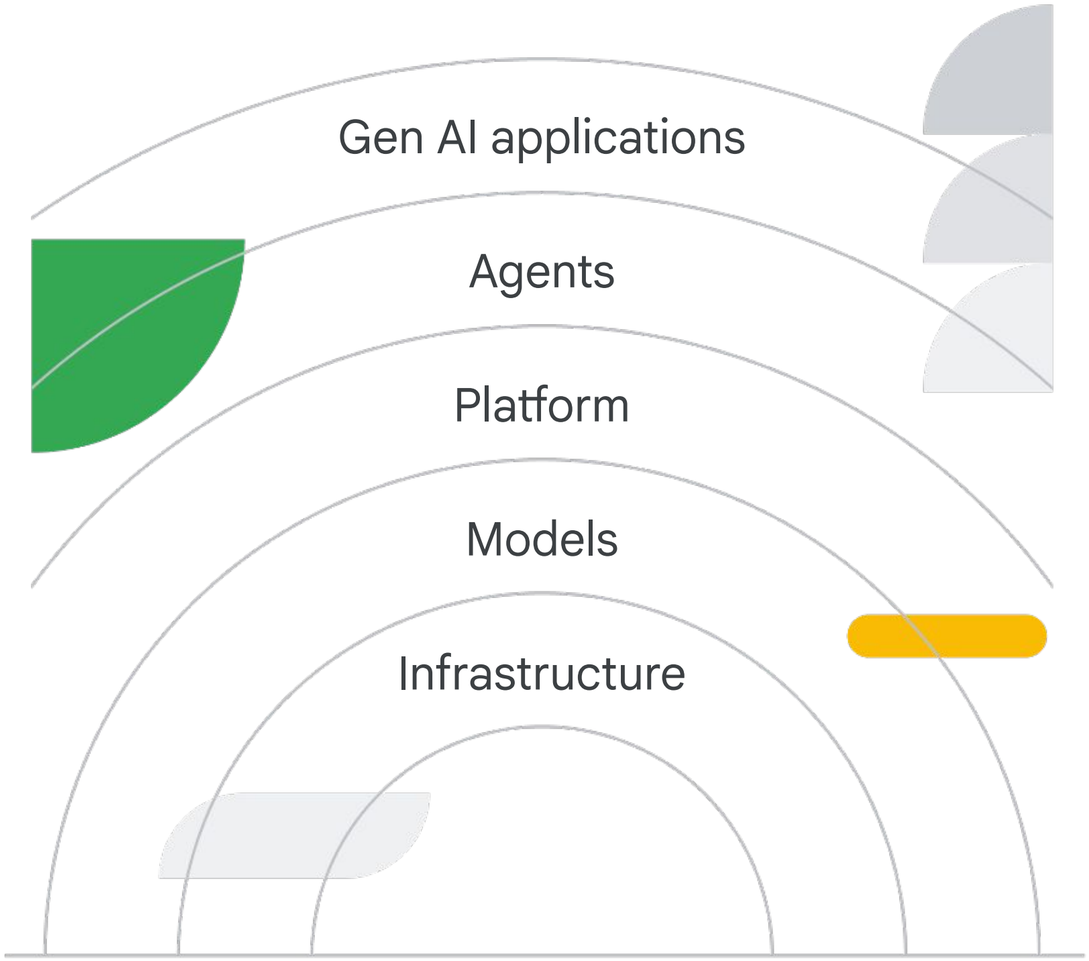
3. A company needs to significantly increase the processing power available to train a very large language model. This requires more specialized processors and high-speed data connections.



Building blocks of generative AI: Activity

Read the statement below, and identify which layer it describes.

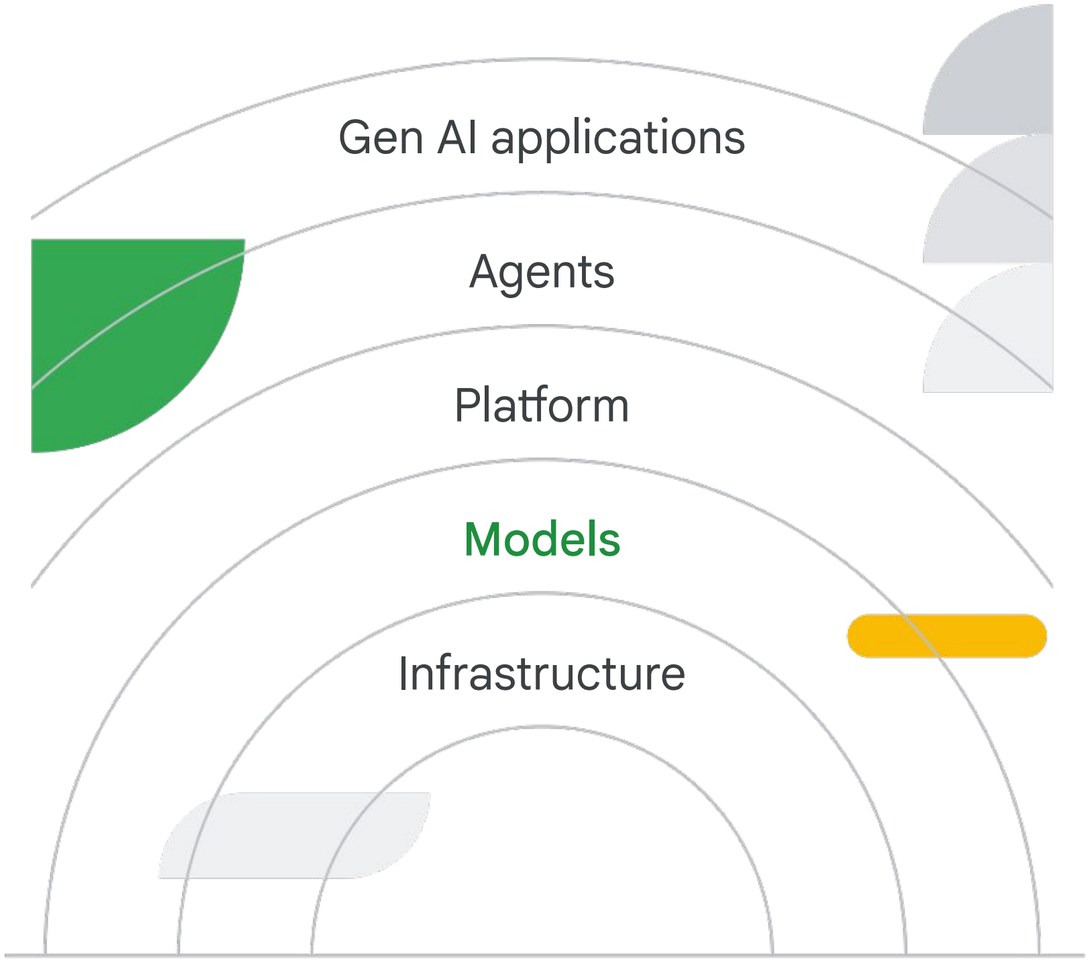
4. After being trained on millions of lines of code, a system can automatically suggest the next line of code a programmer is likely to write.



Building blocks of generative AI: Activity

Read the statement below, and identify which layer it describes.

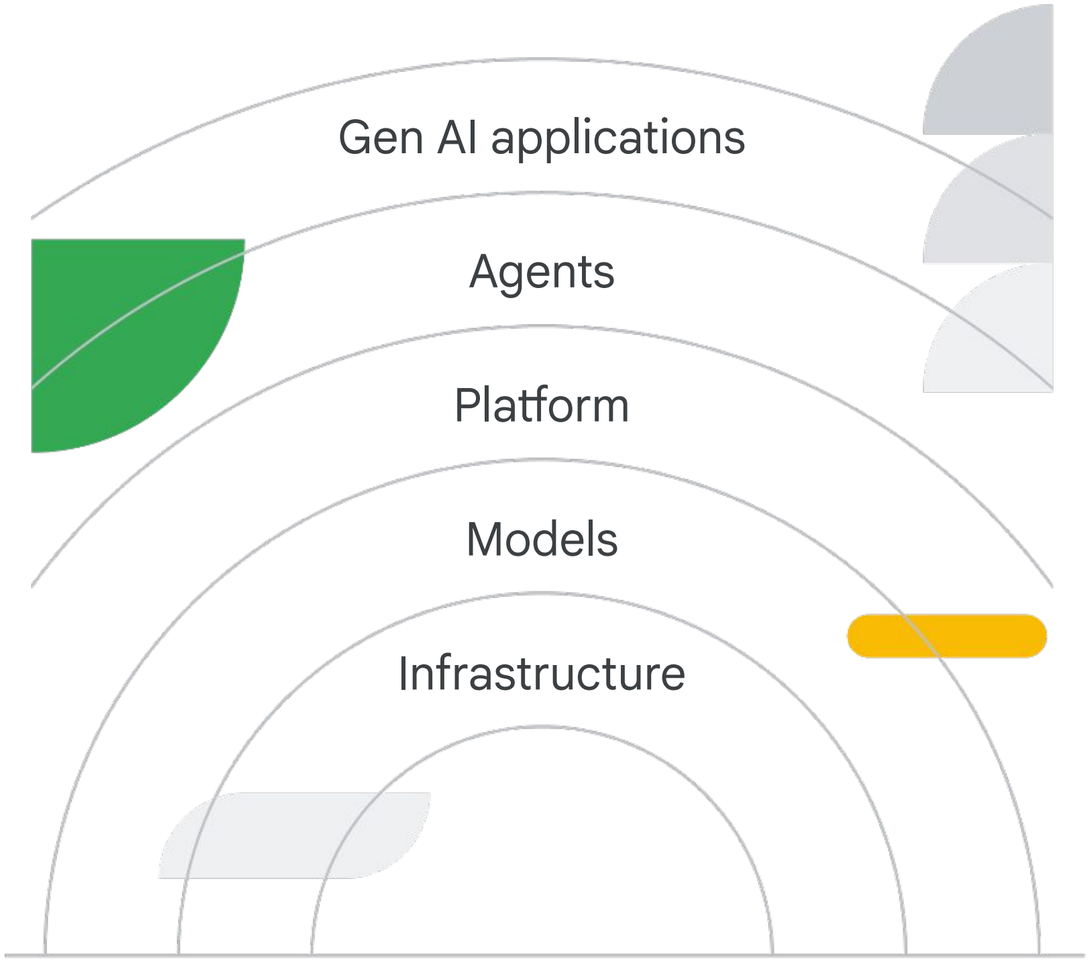
4. After being trained on millions of lines of code, a system can automatically suggest the next line of code a programmer is likely to write.



Building blocks of generative AI: Activity

Read the statement below, and identify which layer it describes.

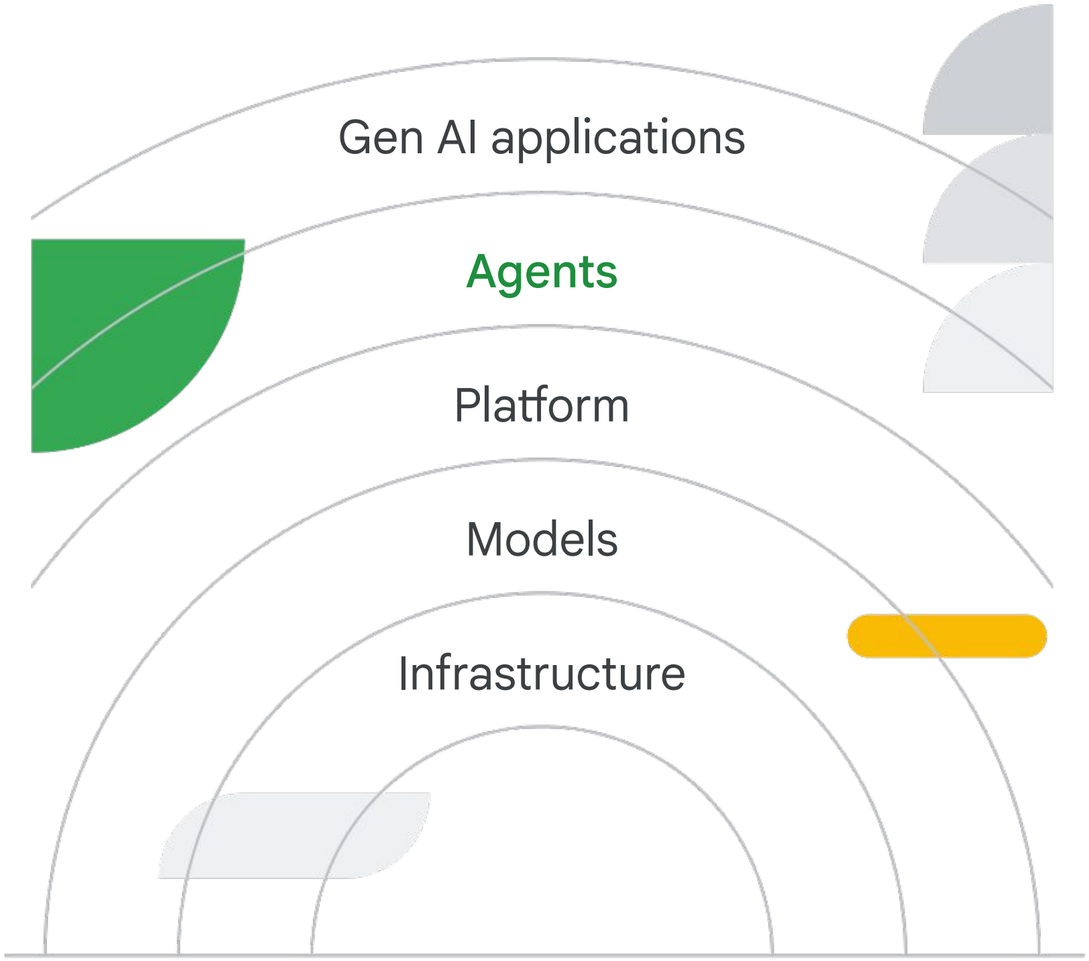
5. A software program can monitor a company's social media feeds, identify negative customer sentiment, and automatically draft a response to address the issue.



Building blocks of generative AI: Activity

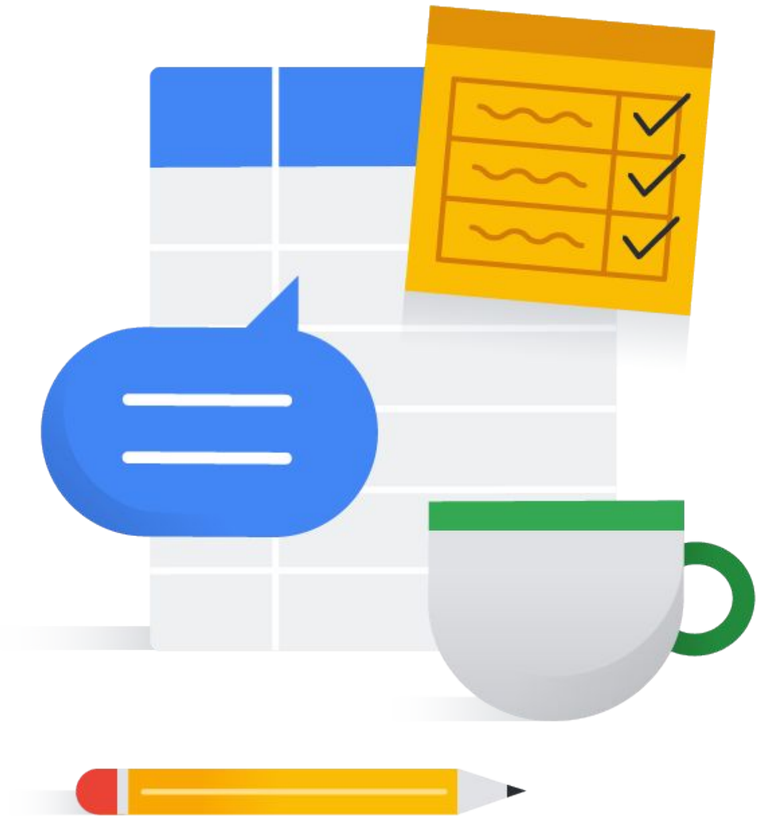
Read the statement below, and identify which layer it describes.

5. A software program can monitor a company's social media feeds, identify negative customer sentiment, and automatically draft a response to address the issue.



Key takeaways

- Generative AI relies on interconnected layers, from infrastructure to gen AI applications.
- Understanding these layers empowers informed decisions and business innovation.



Agenda

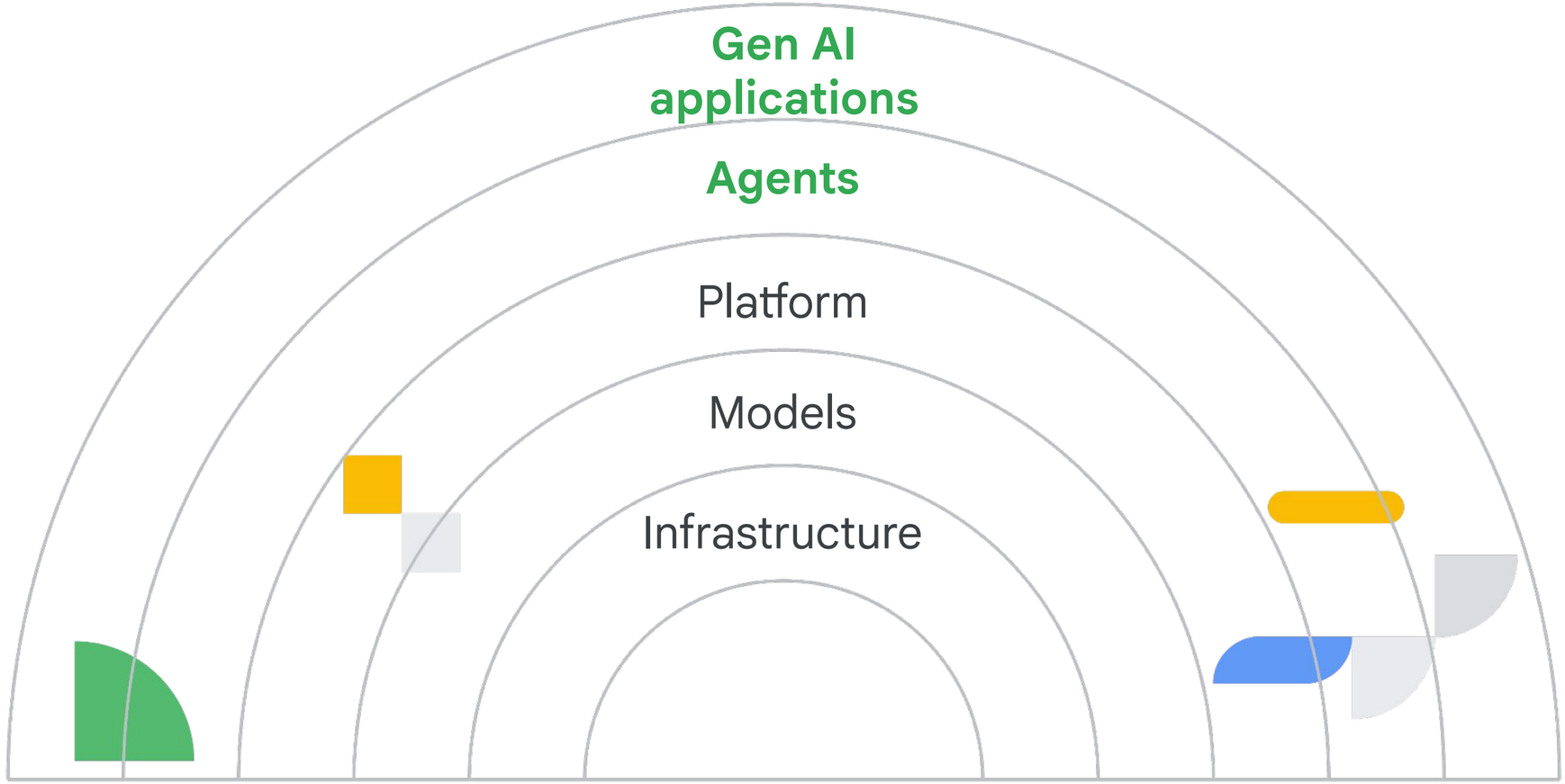
01 The gen AI landscape

02 Gen AI applications and agents

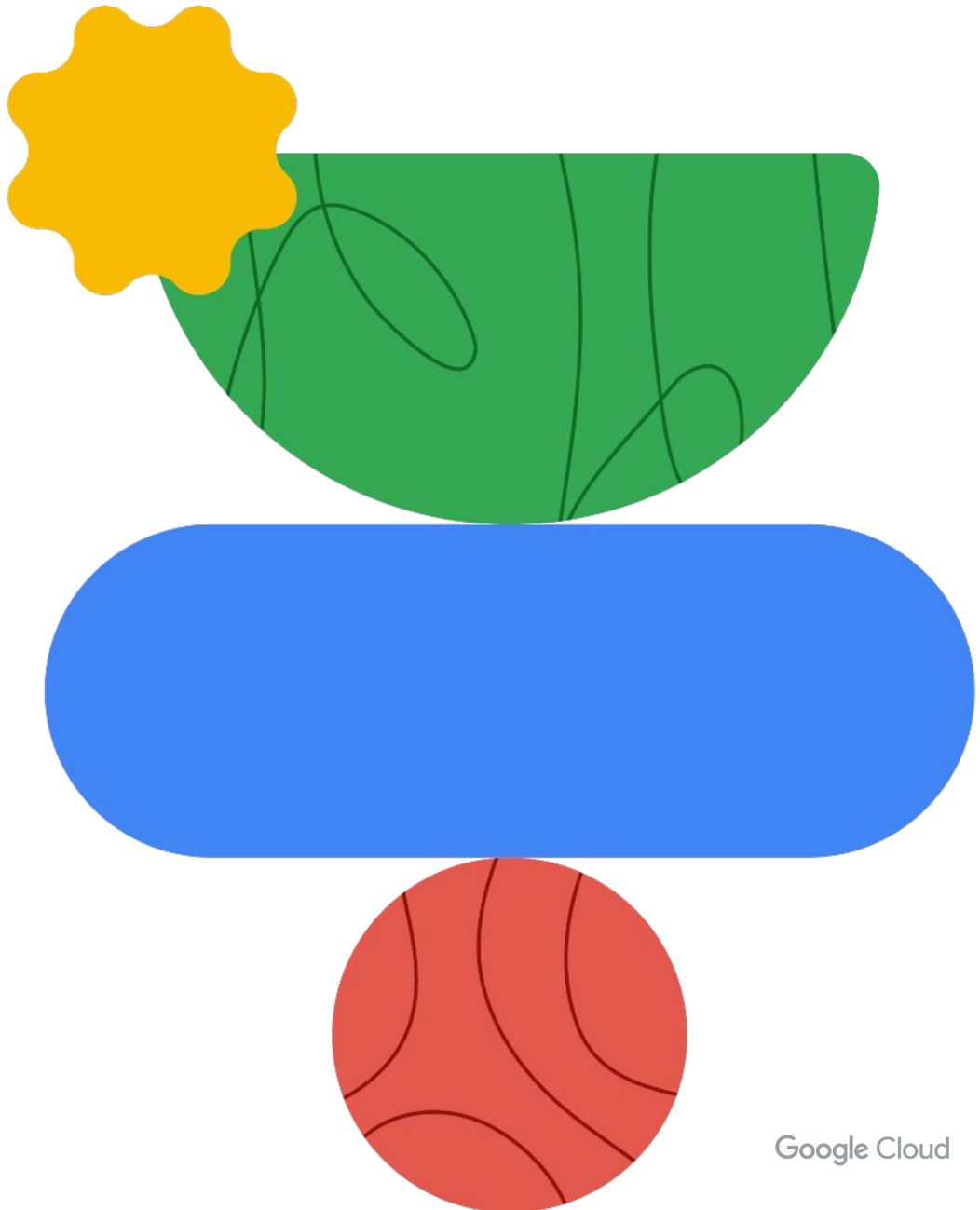
03 Gen AI platform, models, and infrastructure

04 Gen AI project resources and management

Gen AI applications and agents

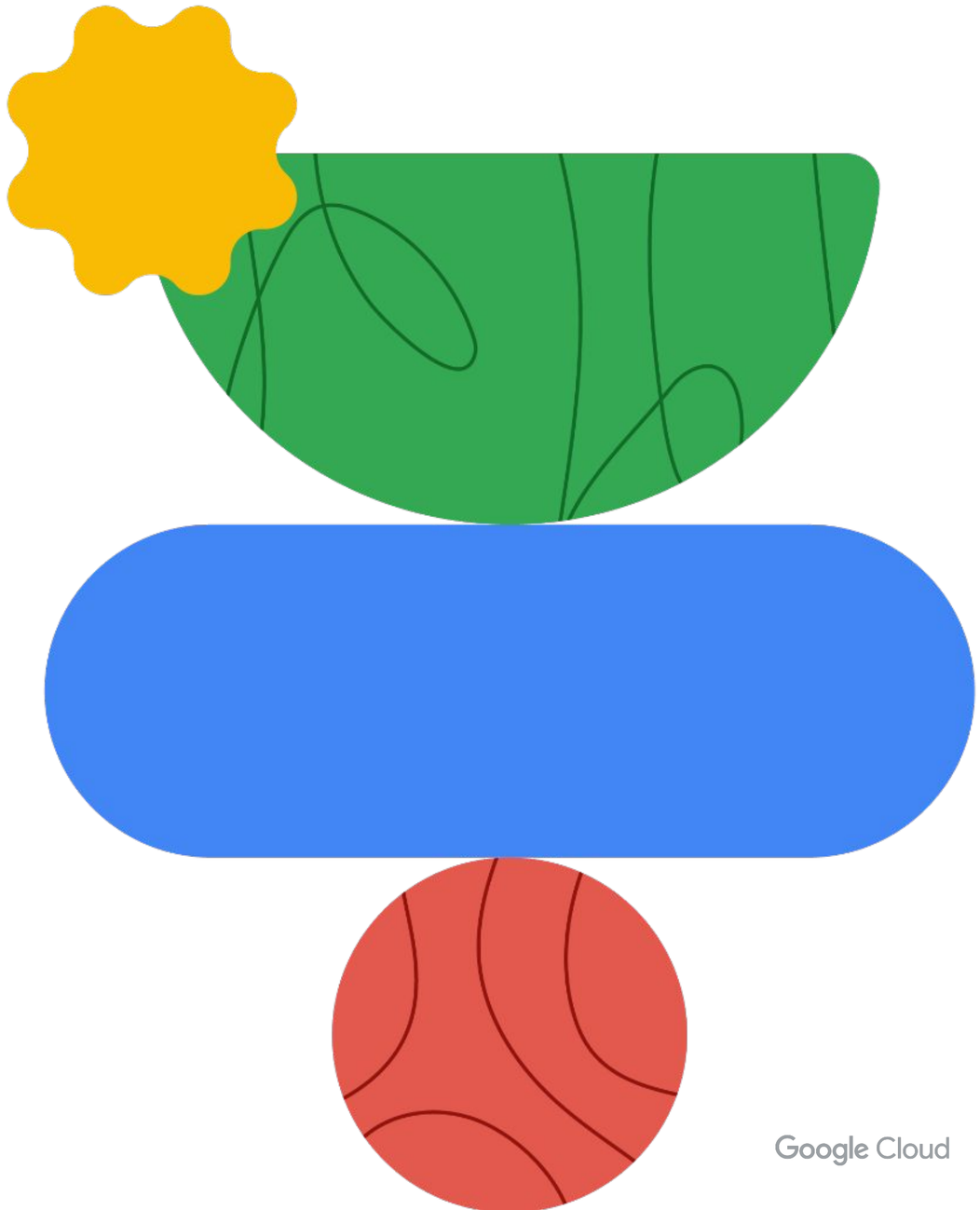


What can agents do?



A gen AI agent is an application that tries to achieve a goal by observing the world and acting upon it using the tools it has at its disposal.

How do agents relate to gen AI-powered applications?



Capabilities of agents

Understand and respond to natural language

Agents allow applications to have more intuitive interfaces that can understand complex requests.

Automate complex tasks

Personalization

Capabilities of agents

Understand and respond to natural language

Agents allow applications to have more intuitive interfaces that can understand complex requests.

Automate complex tasks

They can handle multi-step processes within an application.

Personalization

Capabilities of agents

Understand and respond to natural language

Agents allow applications to have more intuitive interfaces that can understand complex requests.

Automate complex tasks

They can handle multi-step processes within an application.

Personalization

They can learn user preferences and tailor the application experience accordingly.

Multi-agent systems in gen AI-powered applications

- A travel booking app
- A customer support app
- A personalized learning app

Multi-agent systems in gen AI-powered applications

- A travel booking app
- A customer support app
- A personalized learning app

- One **agent** handles the complex task of finding the best flights and hotels.
- Another **agent** specializes in suggesting relevant activities and attractions at the destination.
- The **application** provides the user with the interface for browsing options and making reservations.

Multi-agent systems in gen AI-powered applications

- A travel booking app
- A customer support app
- A personalized learning app

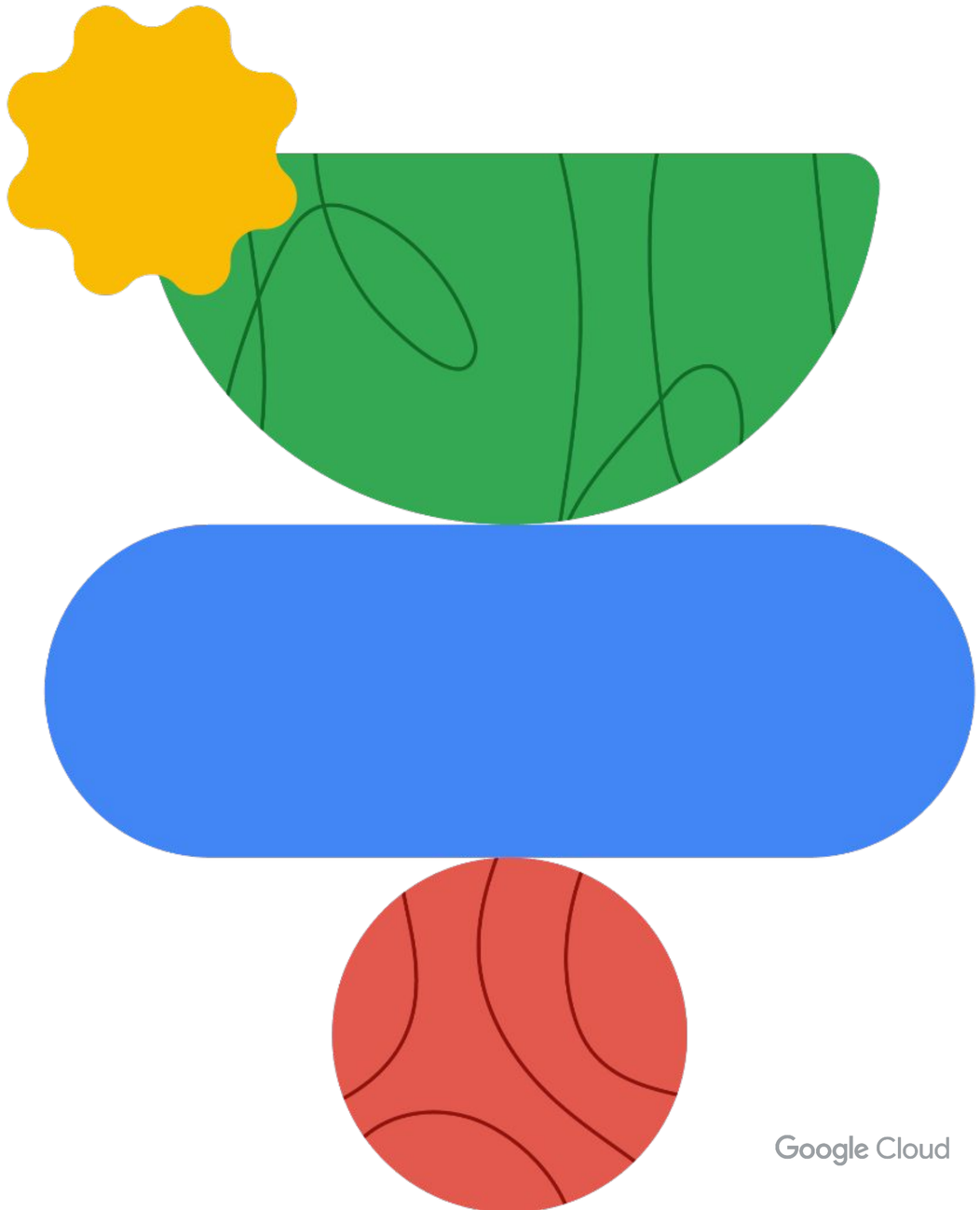
- An **agent** answers common questions, troubleshoots problems, and escalates complex issues to human representatives.
- The **application** provides the chat interface and integrates with other support systems.

Multi-agent systems in gen AI-powered applications

- A travel booking app
- A customer support app
- A personalized learning app

- An **agent** assesses a student's knowledge, recommends relevant learning materials, and generates personalized exercises.
- The **application** provides the structure for lessons and track progress.

How do agents work?



How agents work

01 **Conversational agents**

02 Workflow agents

How it works

1. You provide input: You type a message or speak to the agent.
2. The agent understands.
3. The agent calls a tool.
4. The agent generates a response.
5. The agent delivers the response.

How agents work

01 **Conversational agents**

02 Workflow agents

How it works

1. You provide input: You type a message or speak to the agent.
2. The agent understands.
3. The agent calls a tool.
4. The agent generates a response.
5. The agent delivers the response.

Examples

- Answering questions
- Chatting casually
- Accessing information

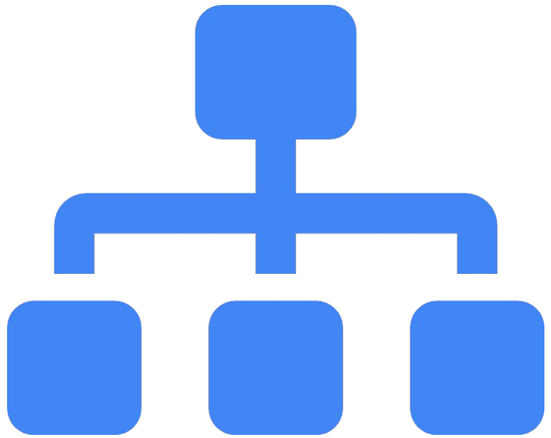
How agents work

01 Conversational agents

02 **Workflow agents**

How it works

1. You provide input: You define a task or trigger a process.
2. The agent understands.
3. The agent calls a tool.
4. The agent generates a result or output.
5. The agent delivers the result or output.



How agents work

01 Conversational agents

02 **Workflow agents**

How it works

1. You provide input: You define a task or trigger a process.
2. The agent understands.
3. The agent calls a tool.
4. The agent generates a result or output.
5. The agent delivers the result or output.

Examples

- Ecommerce order fulfillment
- Customer onboarding
- Automated research
- Security log parsing

Discussion: Conversational and workflow agents

5 min | Group

Consider these examples of agent solutions:

Conversational

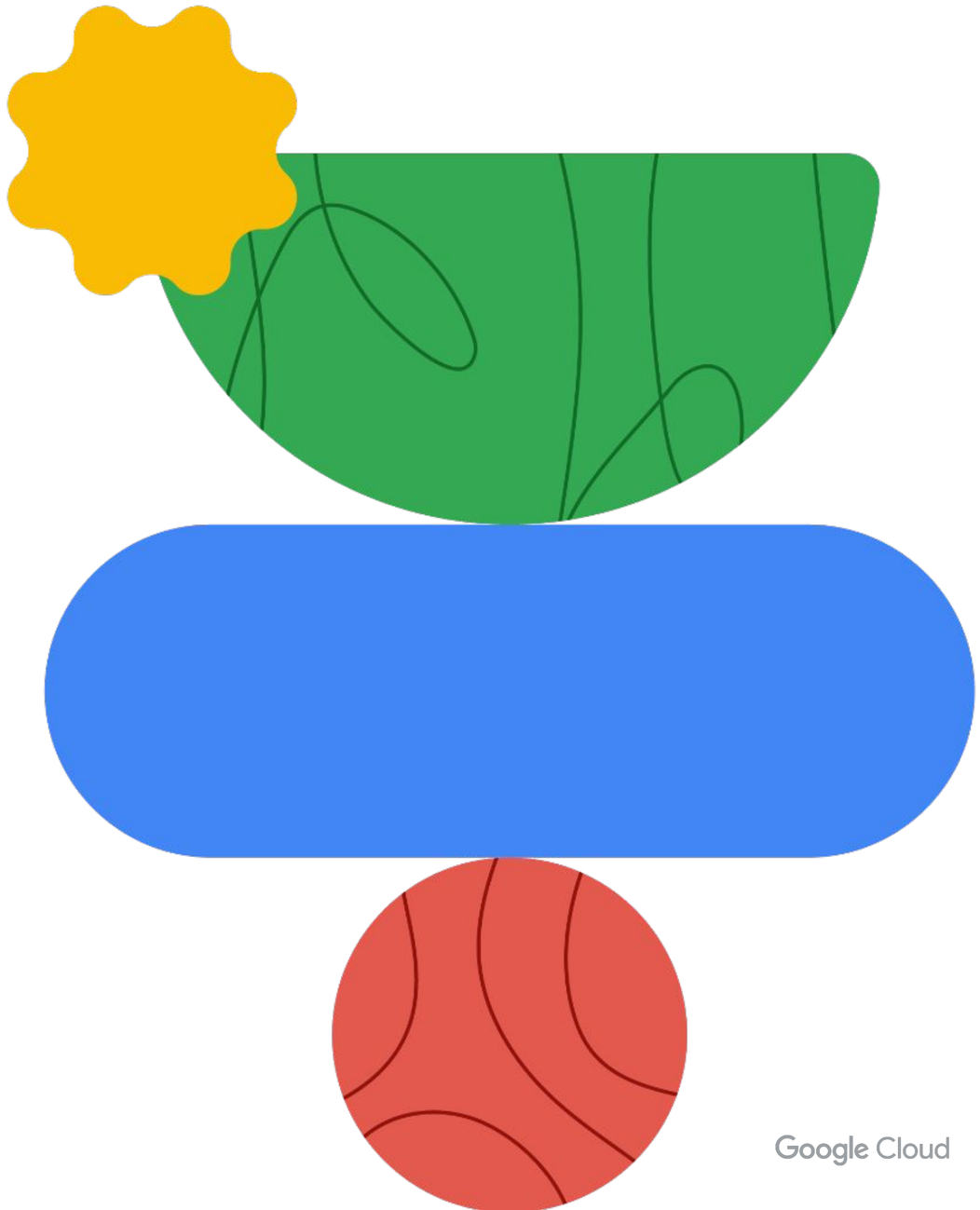
- Answering questions
- Chatting casually
- Accessing information

Workflow

- Ecommerce order fulfillment
- Customer onboarding
- Automated research
- Security log parsing

Select one of these solutions and describe a specific challenge in your organization that an agent could help address.

Agent use cases



Agent use cases

Customer service agents

Answer questions, resolve issues, and provide personalized recommendations.

Employee productivity agents

Creative agents

Agent use cases

Customer service agents

Answer questions, resolve issues, and provide personalized recommendations.

Employee productivity agents

Help workers find information, manage tasks, and automate workflows.

Creative agents

Agent use cases

Customer service agents

Answer questions, resolve issues, and provide personalized recommendations.

Employee productivity agents

Help workers find information, manage tasks, and automate workflows.

Creative agents

Generate new ideas, create content, and translate languages.

Additional agent use cases

Code agents

Assist developers in writing, reviewing, debugging code, and generating code.

Data agents

Security agents

Additional agent use cases

Code agents

Assist developers in writing, reviewing, debugging code, and generating code.

Data agents

Analyze large datasets, identify trends, and extract insights from data.

Security agents

Additional agent use cases

Code agents

Assist developers in writing, reviewing, debugging code, and generating code.

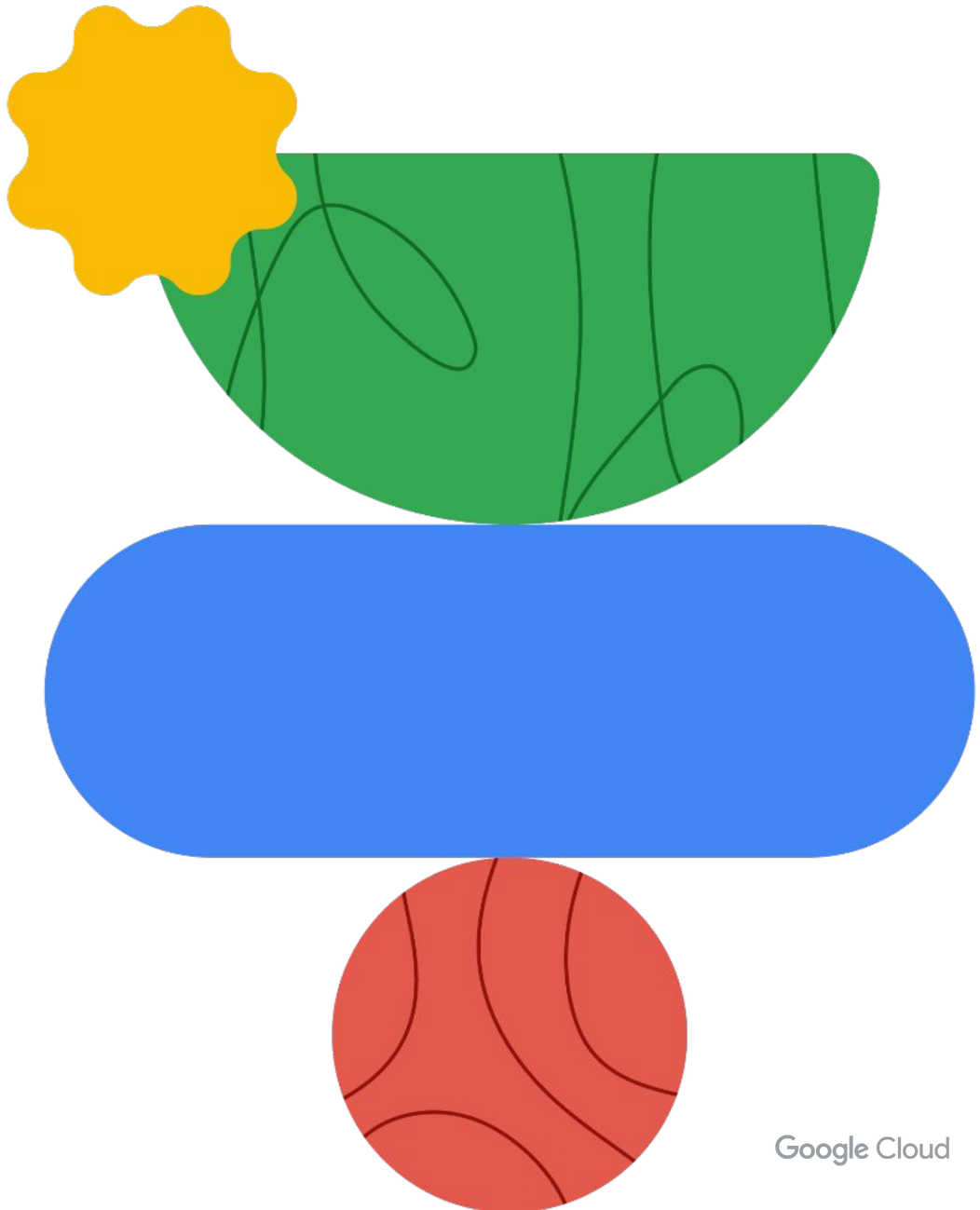
Data agents

Analyze large datasets, identify trends, and extract insights from data.

Security agents

Automate security tasks.

Gen AI agents:
Beyond just models



Gen AI agents: Beyond just models

The reasoning loop often **utilizes advanced prompt engineering frameworks** to guide its decision-making process.

Advanced prompt engineering frameworks

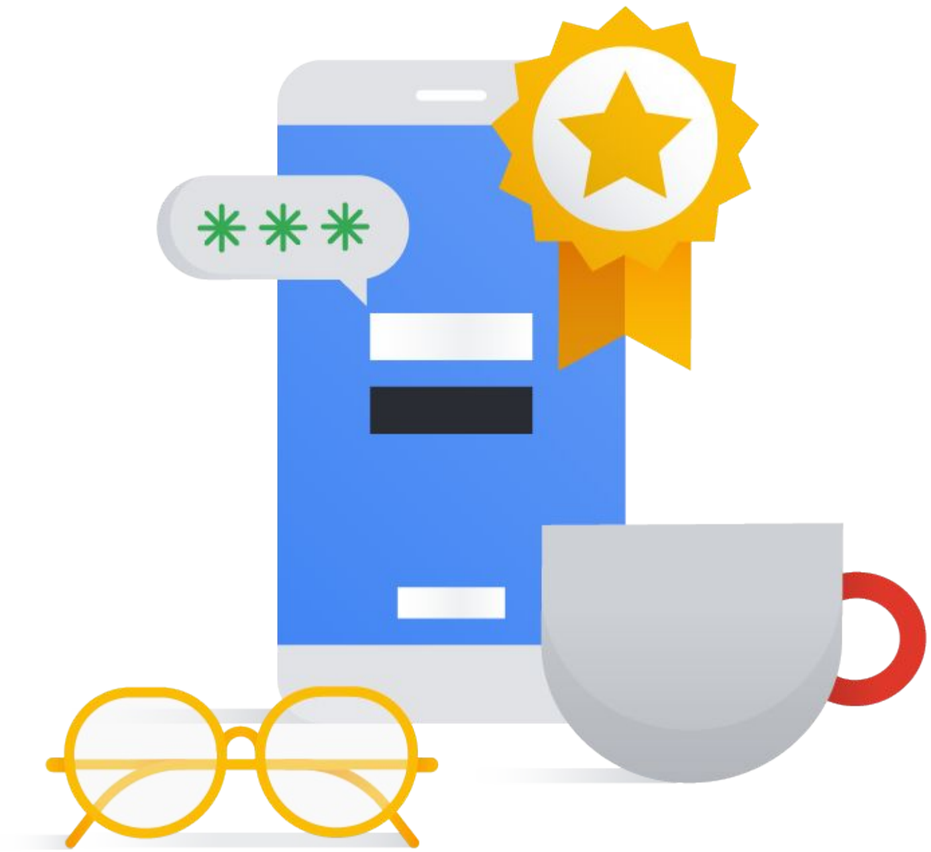
Frameworks include:

- Simple rule-based calculations.
- Complex thought chains.
- Machine learning algorithms.
- Probabilistic reasoning techniques.

Examples:

- ReAct (reasoning + acting) prompting
- Chain-of-thought (CoT) prompting

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



Quiz | Question 01

Question

What is the primary purpose of a gen AI agent?

A. To exclusively generate creative content like text and images.

B. To achieve a specific goal by observing its environment, reasoning, and acting upon it using available tools.

C. To function only as a platform for building other AI models.

D. To passively provide information without any ability to take action or use tools.

Quiz | Question 01

Answer

What is the primary purpose of a gen AI agent?

A. To exclusively generate creative content like text and images.

B. To achieve a specific goal by observing its environment, reasoning, and acting upon it using available tools.

C. To function only as a platform for building other AI models.

D. To passively provide information without any ability to take action or use tools.

Quiz | Question 02

Question

What is the relationship between gen AI agents and gen AI-powered applications?

A. Gen AI agents are intelligent components that function within a larger gen AI-powered application, which provides the user interface and overall goals.

B. Gen AI agents are standalone entities that operate independently of any application.

C. Gen AI-powered applications are simply another name for gen AI agents; the terms are interchangeable.

D. Gen AI agents define the user interface and overall goals, while applications handle the specific, complex tasks.

Quiz | Question 02

Answer

What is the relationship between gen AI agents and gen AI-powered applications?

A. Gen AI agents are intelligent components that function within a larger gen AI-powered application, which provides the user interface and overall goals.

B. Gen AI agents are standalone entities that operate independently of any application.

C. Gen AI-powered applications are simply another name for gen AI agents; the terms are interchangeable.

D. Gen AI agents define the user interface and overall goals, while applications handle the specific, complex tasks.

Quiz | Question 03

Question

A company is looking to implement a system where AI guides new customers through the account setup process, provides interactive tutorials, and answers their initial frequently asked questions. Which category of gen AI agent is best suited for this "customer onboarding" task?

A. A conversational agent strictly limited to providing factual answers like the capital of a country.

B. A creative agent specialized in generating novel content like marketing slogans.

C. A security agent focused on parsing logs and identifying abnormalities.

D. A workflow agent.

Quiz | Question 03

Answer

A company is looking to implement a system where AI guides new customers through the account setup process, provides interactive tutorials, and answers their initial frequently asked questions. Which category of gen AI agent is best suited for this "customer onboarding" task?

A. A conversational agent strictly limited to providing factual answers like the capital of a country.

B. A creative agent specialized in generating novel content like marketing slogans.

C. A security agent focused on parsing logs and identifying abnormalities.

D. A workflow agent. 

Quiz | Question 04

Answer

What two key elements distinguish AI agents from standalone large language models (LLMs) and empower them to tackle complex problems and manage multi-step tasks?

A. The ability to understand natural language and generate personalized responses.

B. A reasoning loop and the ability to use tools.

C. The option to be pre-built or custom-created and the capacity to operate within a user-facing layer.

D. The use of advanced prompt engineering frameworks and the ability to process information.

Quiz | Question 04

Answer

What two key elements distinguish AI agents from standalone large language models (LLMs) and empower them to tackle complex problems and manage multi-step tasks?

A. The ability to understand natural language and generate personalized responses.

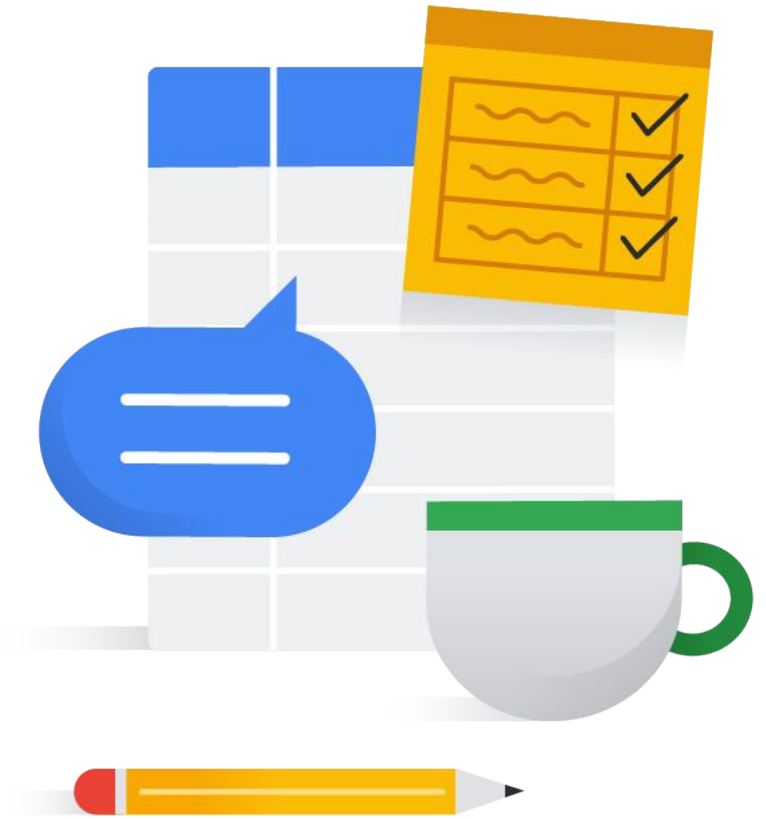
B. A reasoning loop and the ability to use tools.

C. The option to be pre-built or custom-created and the capacity to operate within a user-facing layer.

D. The use of advanced prompt engineering frameworks and the ability to process information.

Key takeaways

- Agents enhance applications by adding intelligence and automation.
- Applications provide the framework and purpose for agents.
- Gen AI agents, with reasoning and tools, solve complex problems beyond standalone models. Understanding their potential for innovation and efficiency is crucial.



Agenda

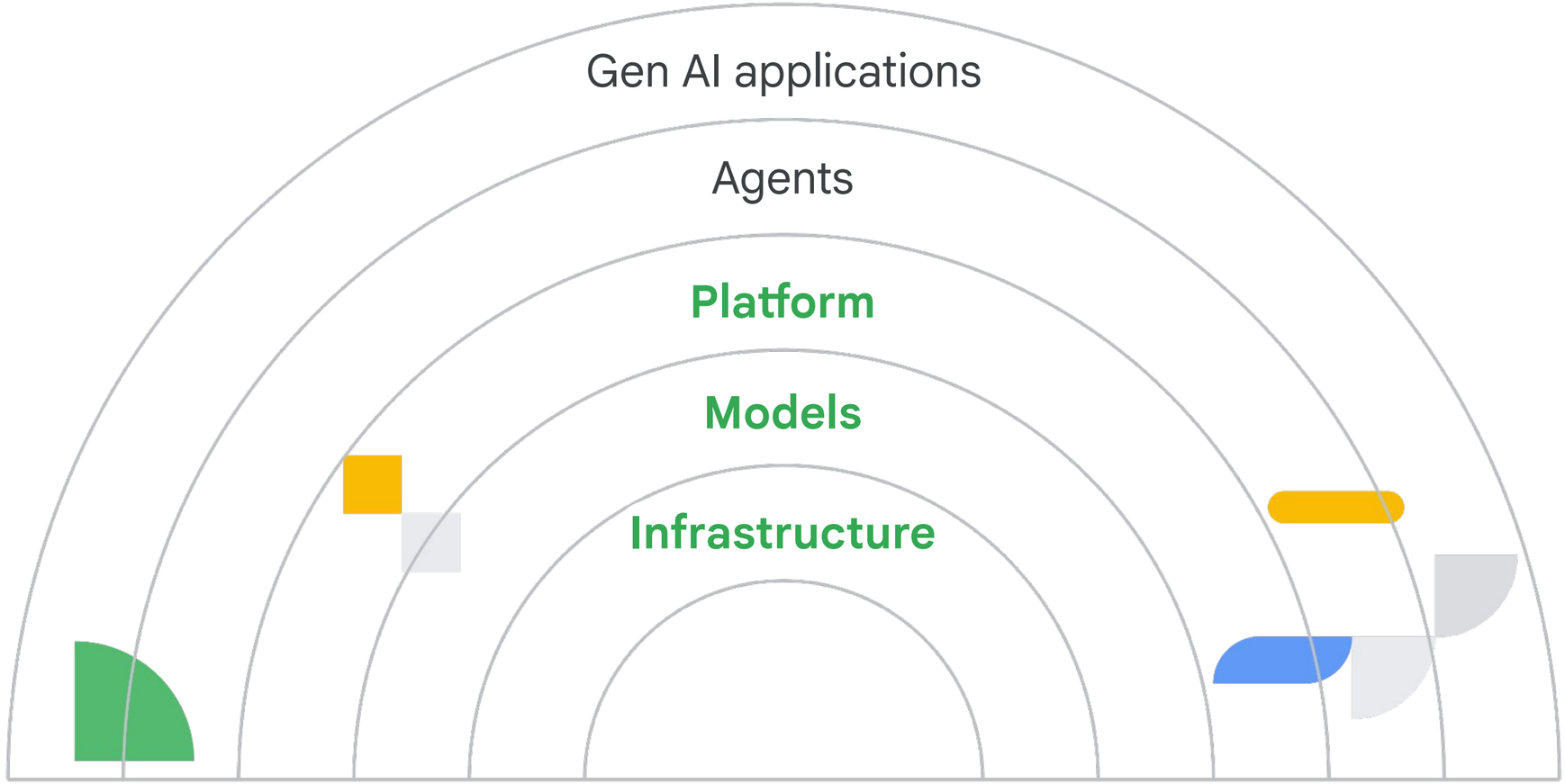
01 The gen AI landscape

02 Gen AI applications and agents

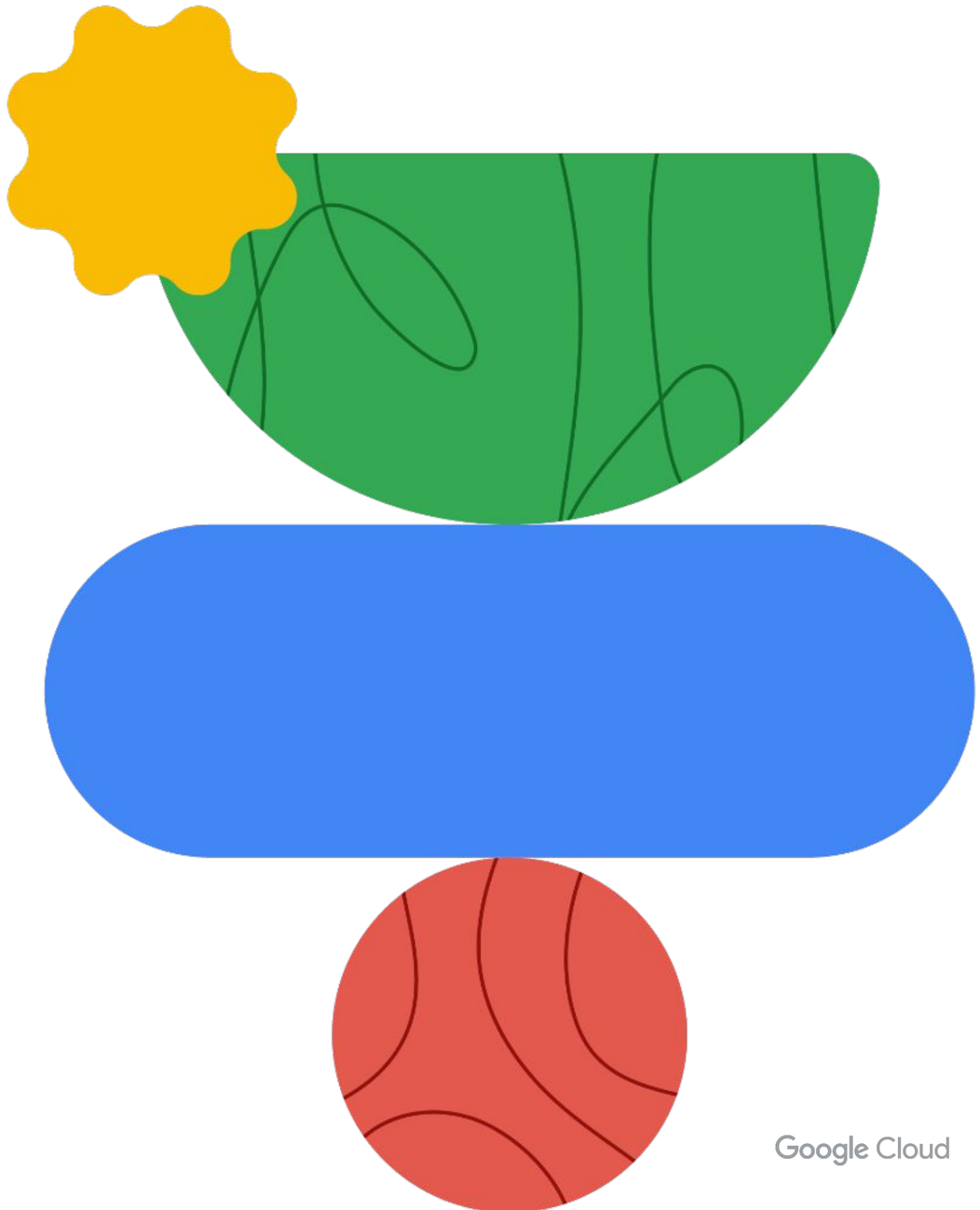
03 Gen AI platform, models, and infrastructure

04 Gen AI project resources and management

Gen AI platform, models, and infrastructure



The platform layer



The platform layer provides the **foundation for building and scaling** your AI initiatives.

Vertex AI

Vertex AI streamlines the entire ML workflow by providing the necessary:

- Infrastructure
- Pre-trained models
- Tools to build, deploy, and manage

Vertex AI: Key features and benefits

Open and flexible

Powerful infrastructure

Pre-trained models

Comprehensive tooling

Customization

Easy integration

Vertex AI's MLOps tools: Benefits

MLOps tools help you:

- Orchestrate end-to-end ML workflows.
- Perform feature engineering.
- Run experiments.
- Manage and iterate your models.
- Track ML metadata.
- Monitor and evaluate model quality.
- Automate, standardize, and manage the ML project lifecycle.

Vertex AI's MLOps tools

Feature Store

Share, serve, and reuse ML features to maintain consistency and efficiency.

Model Registry

Model evaluation

Vertex AI's MLOps tools

Feature Store

Share, serve, and reuse ML features to maintain consistency and efficiency.

Model Registry

Manage model versions, track changes, and organize your models throughout their lifecycle.

Model evaluation

Vertex AI's MLOps tools

Feature Store

Share, serve, and reuse ML features to maintain consistency and efficiency.

Model Registry

Manage model versions, track changes, and organize your models throughout their lifecycle.

Model evaluation

Evaluate and compare model performance.

Additional Vertex AI's MLOps tools

Workflow orchestration

Automates ML workflows using Vertex AI Pipelines.

Model Monitoring

Additional Vertex AI's MLOps tools

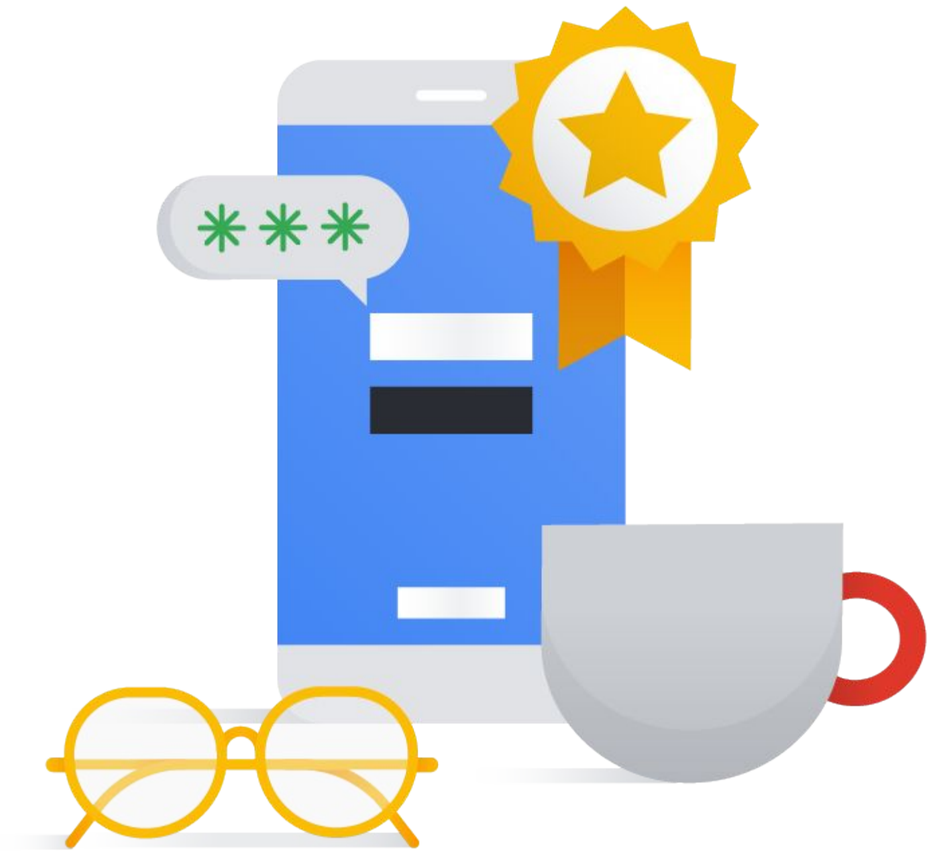
Workflow orchestration

Automates ML workflows using Vertex AI Pipelines.

Model Monitoring

Monitors models for performance degradation, detects input skew and drift, and triggers updates or retraining.

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



Quiz | Question 01

Question

You need to track experiments, manage different versions of your model, and keep a record of changes. Choose the Vertex AI tool that best suits your needs.

A. Model Monitoring

B. Workflow orchestrations

C. Model Registry

Quiz | Question 01

Answer

You need to track experiments, manage different versions of your model, and keep a record of changes. Choose the Vertex AI tool that best suits your needs.

A. Model Monitoring

B. Workflow orchestrations

C. Model Registry 

Quiz | Question 02

Question

Your deployed model’s performance is declining and you need to identify the cause and retrain it. Choose the Vertex AI tool that best suits your needs.

A. Model Monitoring

B. Workflow orchestrations

C. Feature Store

Quiz | Question 02

Answer

Your deployed model’s performance is declining and you need to identify the cause and retrain it. Choose the Vertex AI tool that best suits your needs.

A. Model Monitoring 

B. Workflow orchestrations

C. Feature Store

Quiz | Question 03

Question

You want to automate your entire ML pipeline, from data preprocessing to model deployment.

A. Model evaluation

B. Workflow orchestrations

C. Model Registry

Quiz | Question 03

Question

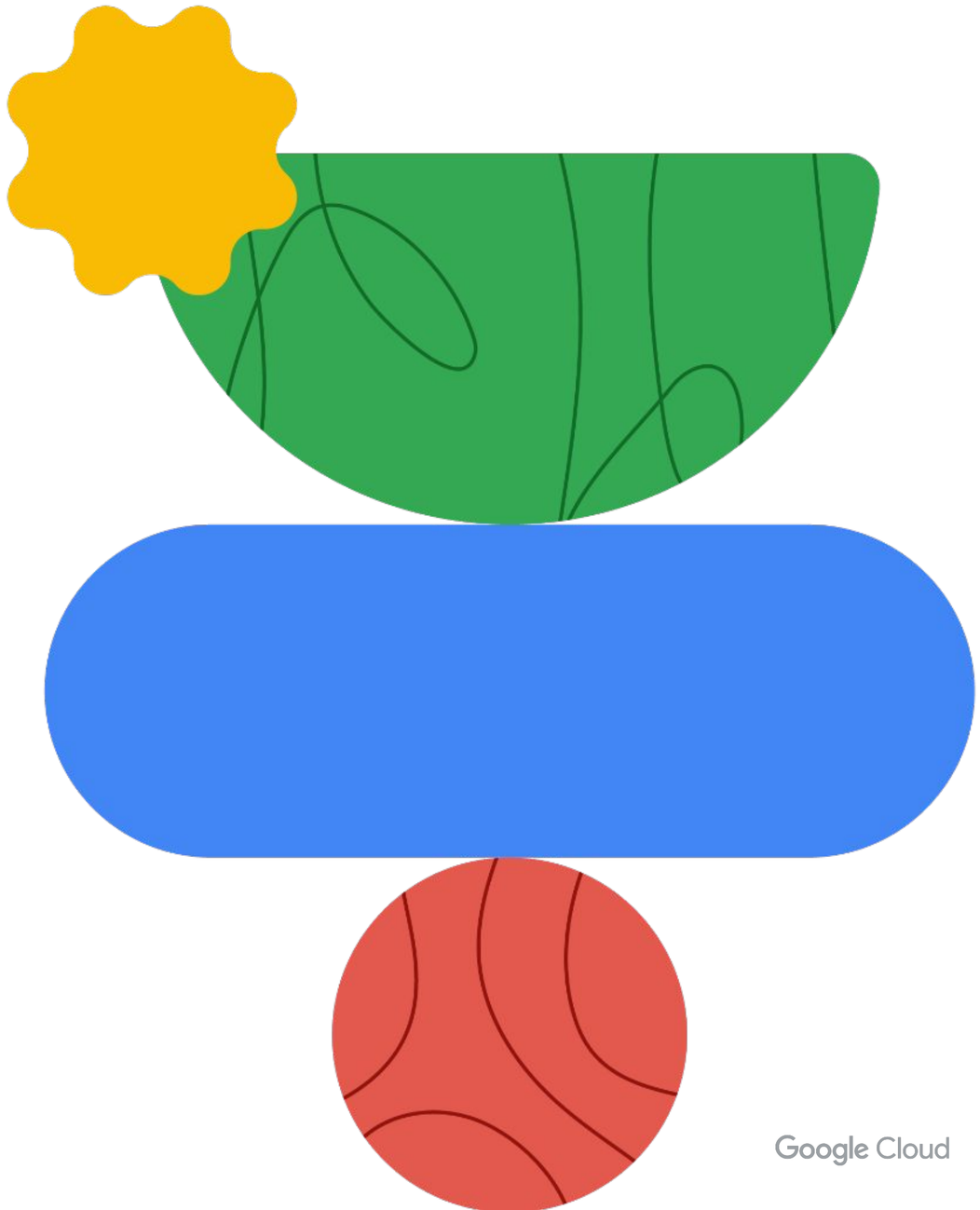
You want to automate your entire ML pipeline, from data preprocessing to model deployment.

A. Model evaluation

B. Workflow orchestrations 

C. Model Registry

The models layer



At the heart of every AI and machine learning system lies the model. They're sophisticated mathematical structures trained on massive amounts of data.

Vertex AI: A model hub

01 Use models with Model Garden

02 Build models with Vertex AI

Vertex AI: A model hub

01 Use models with Model Garden

02 Build models with Vertex AI

- Discover, customize, and deploy existing models.
- Select from over 160 models.
- Customize models within Vertex AI.
- Some pre-trained models can be used out-of-the-box.

Vertex AI: A model hub

01 Use models with Model Garden

02 Build models with Vertex AI

Examples

- **First-party foundation models:** Gemini, Imagen, Veo, Chirp
- **First-party pre-trained APIs:** Speech-to-Text, Natural Language Processing, Translation, Vision
- **Open models:** Gemma 2, CodeGemma, PaliGemma, Llama 3.1/3.2, Mistral AI/AI21, TII models
- **Third-party models:** Anthropic's Claude Model Family

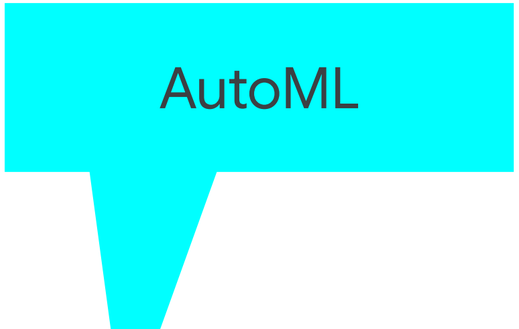
Vertex AI: A model hub

01 Use models with Model Garden

02 Build models with Vertex AI

Two options:

1. Go fully custom, and create and train models at scale.
2. Or use AutoML to create and train models .



Vertex AI: A model hub – AutoML

01 Use models with Model Garden

02 Build models with Vertex AI

Data type	Supported objectives
Image data	Classification, object detection
Video data	Action recognition, classification, object tracking
Tabular data	Classification or regression, forecasting

AutoML

Vertex AI: A model hub

01 Use models with Model Garden

02 Build models with Vertex AI

Standard workflow for model creation or tuning:

01 Gather your data

02 Prepare your data

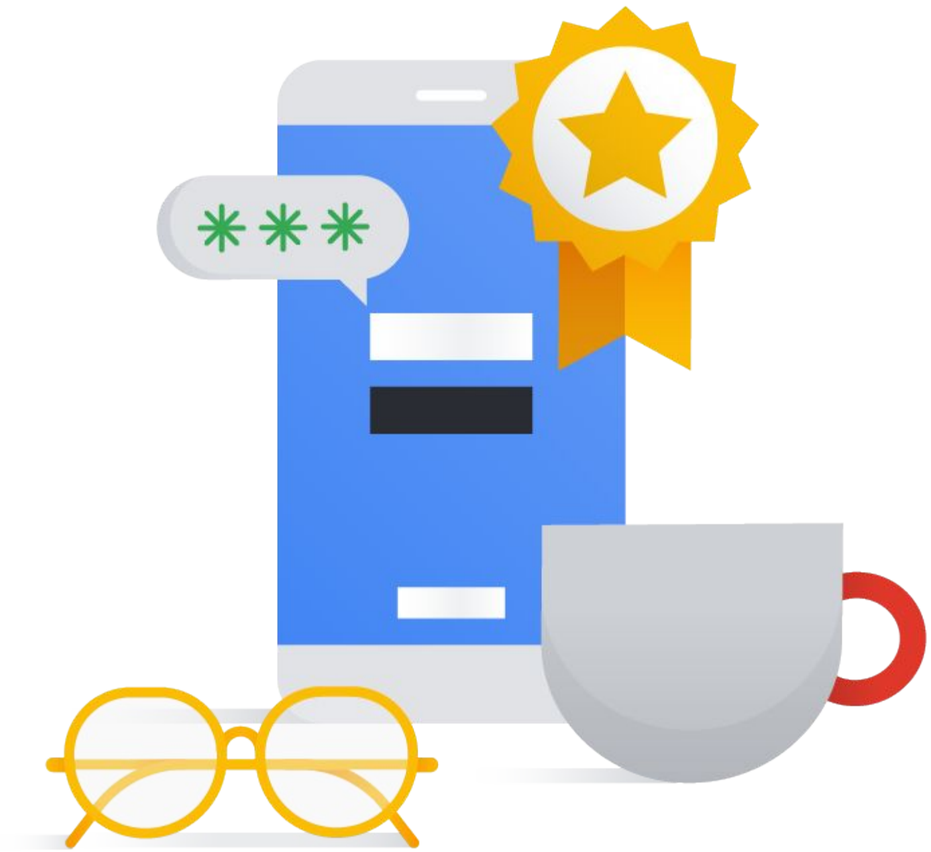
03 Train your model

04 Deploy and predict

05 Manage your model

Use case: Gen AI in manufacturing

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



Quiz | Question 01

Question

You need a model to translate languages for your global ecommerce platform. You want a readily available, high-performing solution. You have a moderate budget.

Which is the AI model that best suits your needs?

A. Vertex AI—build your own custom model

B. Vertex AI AutoML

C. Vertex AI Model Garden

Quiz | Question 01

Answer

You need a model to translate languages for your global ecommerce platform. You want a readily available, high-performing solution. You have a moderate budget.

Which is the AI model that best suits your needs?

A. Vertex AI - build your own custom model

B. Vertex AI AutoML

C. Vertex AI Model Garden 

Quiz | Question 02

Question

You're an experienced AI researcher developing a cutting-edge model for protein folding prediction. You need complete control over the model architecture and training process.

Which is the AI model that best suits the needs?

A. Vertex AI - build your own custom model

B. Vertex AI AutoML

C. Vertex AI Model Garden

Quiz | Question 02

Answer

You're an experienced AI researcher developing a cutting-edge model for protein folding prediction. You need complete control over the model architecture and training process.

Which is the AI model that best suits the needs?

A. Vertex AI - build your own custom model

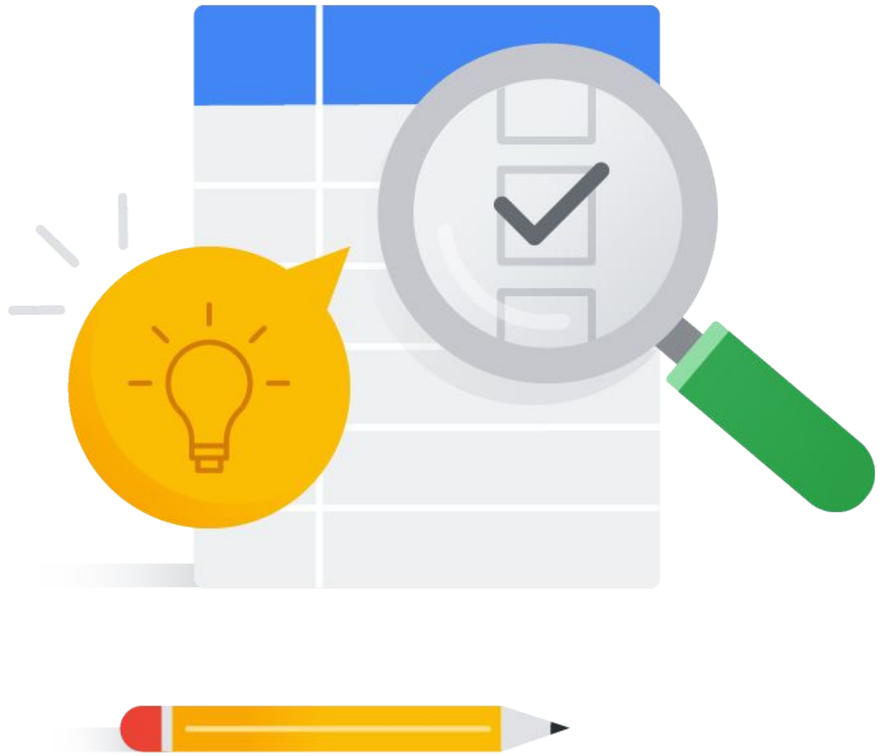
B. Vertex AI AutoML

C. Vertex AI Model Garden

Activity: Agent versus model

5 min

1. Read the scenario and related tasks.
2. Identify which layer handles the respective task - agent or model?
3. Put your answer in the chat.



Agent versus model: Creative content

Scenario:

An AI-powered writing assistant that helps users write different kinds of content, including social media posts, articles, and scripts.

Tasks:

1. Which layer generates grammatically correct text?
2. Which layer identifies and applies appropriate guidelines based on type of content?

Agent versus model: Creative content

Scenario:

An AI-powered writing assistant that helps users write different kinds of content, including social media posts, articles, and scripts.

Tasks:

1. Which layer generates grammatically correct text?
2. Which layer identifies and applies appropriate guidelines based on type of content?

Model layer or agent layer?

Agent versus model: Creative content

Scenario:

An AI-powered writing assistant that helps users write different kinds of content, including social media posts, articles, and scripts.

Tasks:

1. Which layer generates grammatically correct text?
2. Which layer identifies and applies appropriate guidelines based on type of content?

Model layer

Agent versus model: Creative content

Scenario:

An AI-powered writing assistant that helps users write different kinds of content, including social media posts, articles, and scripts.

Tasks:

1. Which layer generates grammatically correct text?
2. Which layer identifies and applies appropriate guidelines based on type of content?

Model layer

Model layer or agent layer?

Agent versus model: Creative content

Scenario:

An AI-powered writing assistant that helps users write different kinds of content, including social media posts, articles, and scripts.

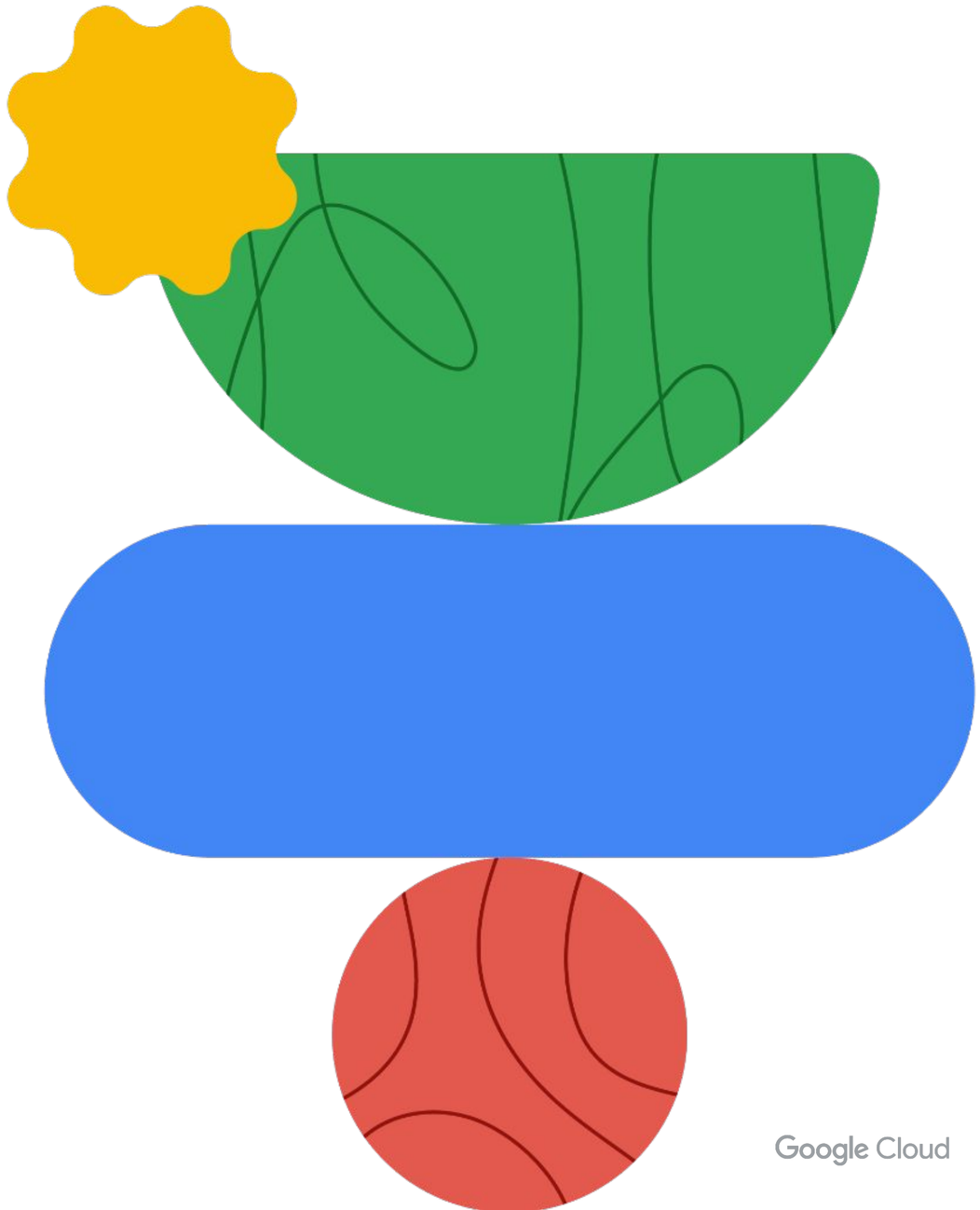
Tasks:

1. Which layer generates grammatically correct text?
2. Which layer identifies and applies appropriate guidelines based on type of content?

Model layer

Agent layer

The infrastructure layer



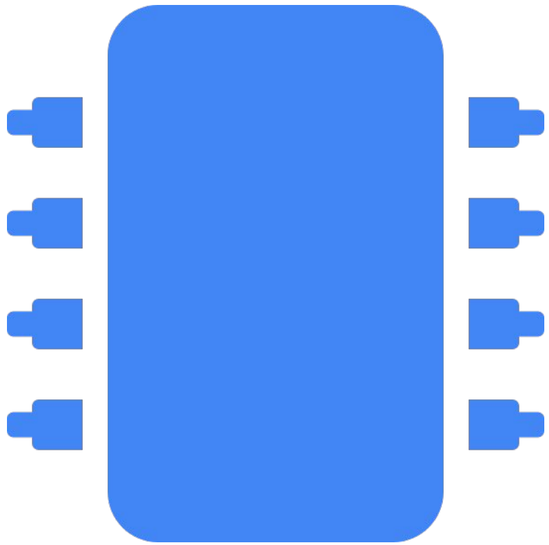
The infrastructure layer is **the foundation upon which any AI system is built**. It's the combination of **hardware and software** that provides the necessary resources to train, deploy, and scale AI models.

Key AI infrastructure components

- **High-performance computing**
  - GPUs and TPUs
  - Hypercomputers
- **High-performance storage**
  - Large-scale storage systems
  - Fast storage access
- **Networking**

Key AI infrastructure components

- **High-performance computing**
  - GPUs and TPUs
  - Hypercomputers
- **High-performance storage**
  - Large-scale storage systems
  - Fast storage access
- **Networking**



Key AI infrastructure components

- **High-performance computing**
  - GPUs and TPUs
  - Hypercomputers
- **High-performance storage**
  - Large-scale storage systems
  - Fast storage access
- **Networking**

- They are the workhorses of AI.
- They excel at parallel processing.
- GPUs offer general-purpose parallel processing power.
- TPUs are custom-designed, optimized for AI tasks.

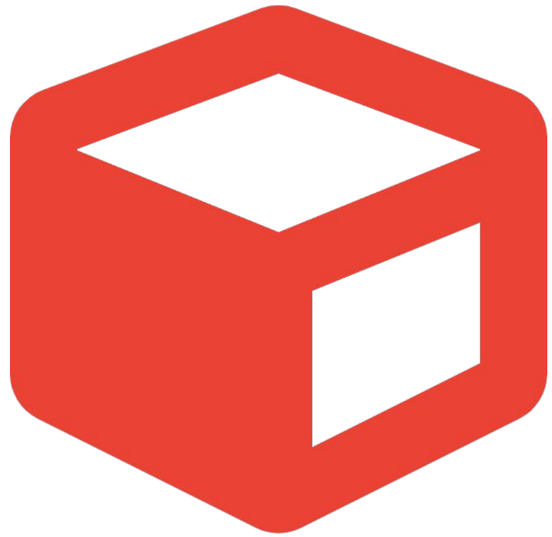
Key AI infrastructure components

- **High-performance computing**
  - GPUs and TPUs
  - Hypercomputers
- **High-performance storage**
  - Large-scale storage systems
  - Fast storage access
- **Networking**

- They are supercomputers built by connecting many individual computers (nodes) together.
- They provide the massive scale necessary for training and running generative AI models.

Key AI infrastructure components

- **High-performance computing**
  - GPUs and TPUs
  - Hypercomputers
- **High-performance storage**
  - Large-scale storage systems
  - Fast storage access
- **Networking**



Key AI infrastructure components

- High-performance computing
  - GPUs and TPUs
  - Hypercomputers
- **High-performance storage**
  - Large-scale storage systems
  - Fast storage access
- Networking

Google Cloud's storage infrastructure is optimized for AI workloads offering:

- High throughput
- Scalability
- Ability to create dense compute clusters for faster training and inference

Key AI infrastructure components

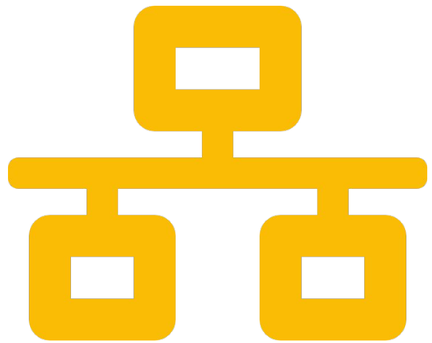
- High-performance computing
  - GPUs and TPUs
  - Hypercomputers
- **High-performance storage**
  - Large-scale storage systems
  - Fast storage access
- Networking

- It provides fast read and write speeds to keep up with the demands of the training process.

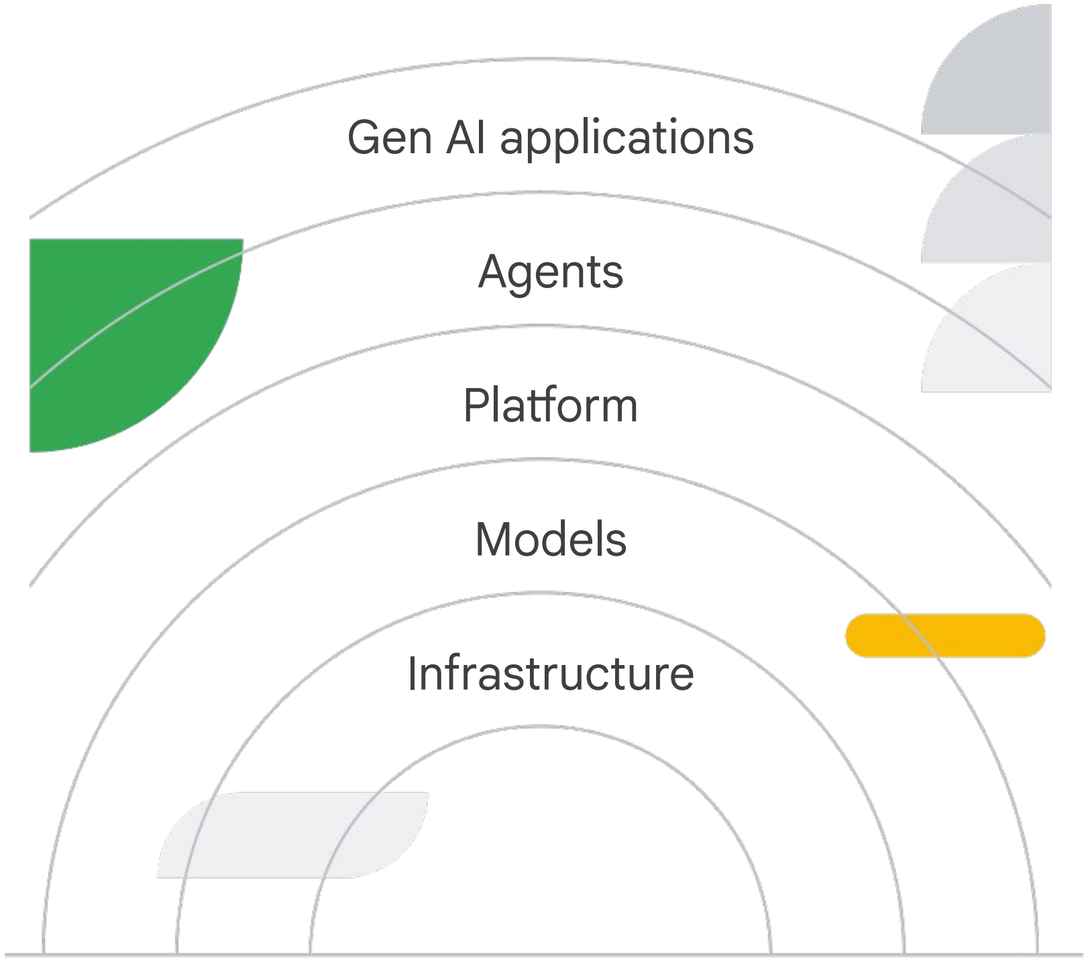
Key AI infrastructure components

- High-performance computing
  - GPUs and TPUs
  - Hypercomputers
- High-performance storage
  - Large-scale storage systems
  - Fast storage access
- **Networking**

- Fast and efficient communication is essential for coordinating the work all of the processors.
- Google's global fiber network provides high-bandwidth, low-latency connectivity.



Gen AI landscape: Travel chatbot example

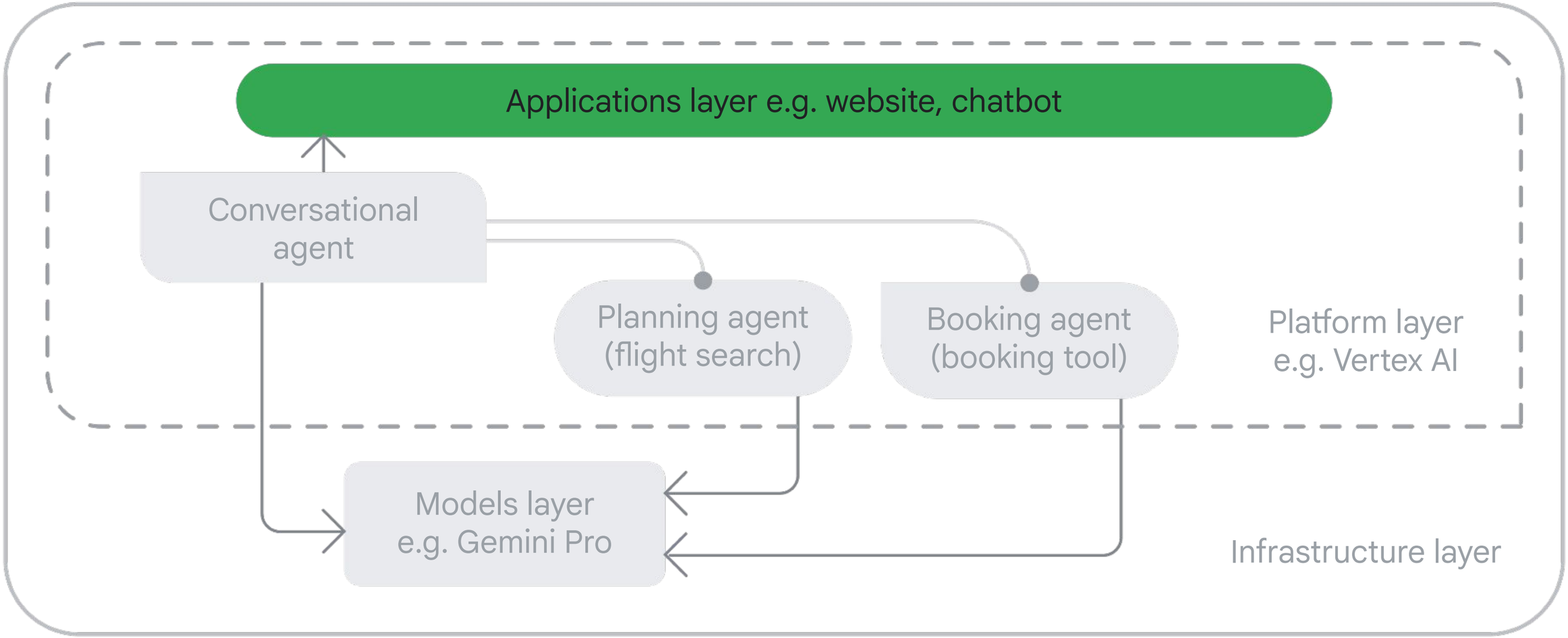


Scenario:

Your favorite air carrier has introduced a new chatbot feature for discovering and booking flights on their website.

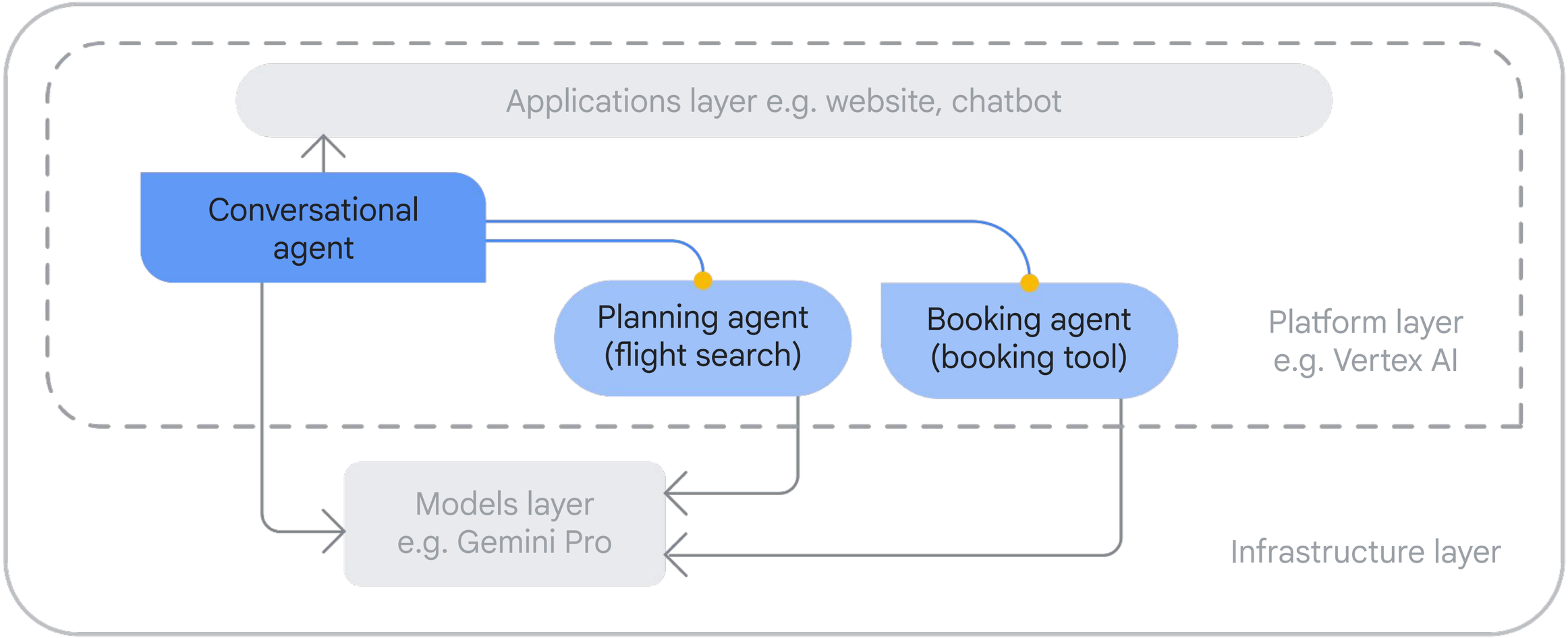
Gen AI landscape:

Travel chatbot example - Gen AI applications layer



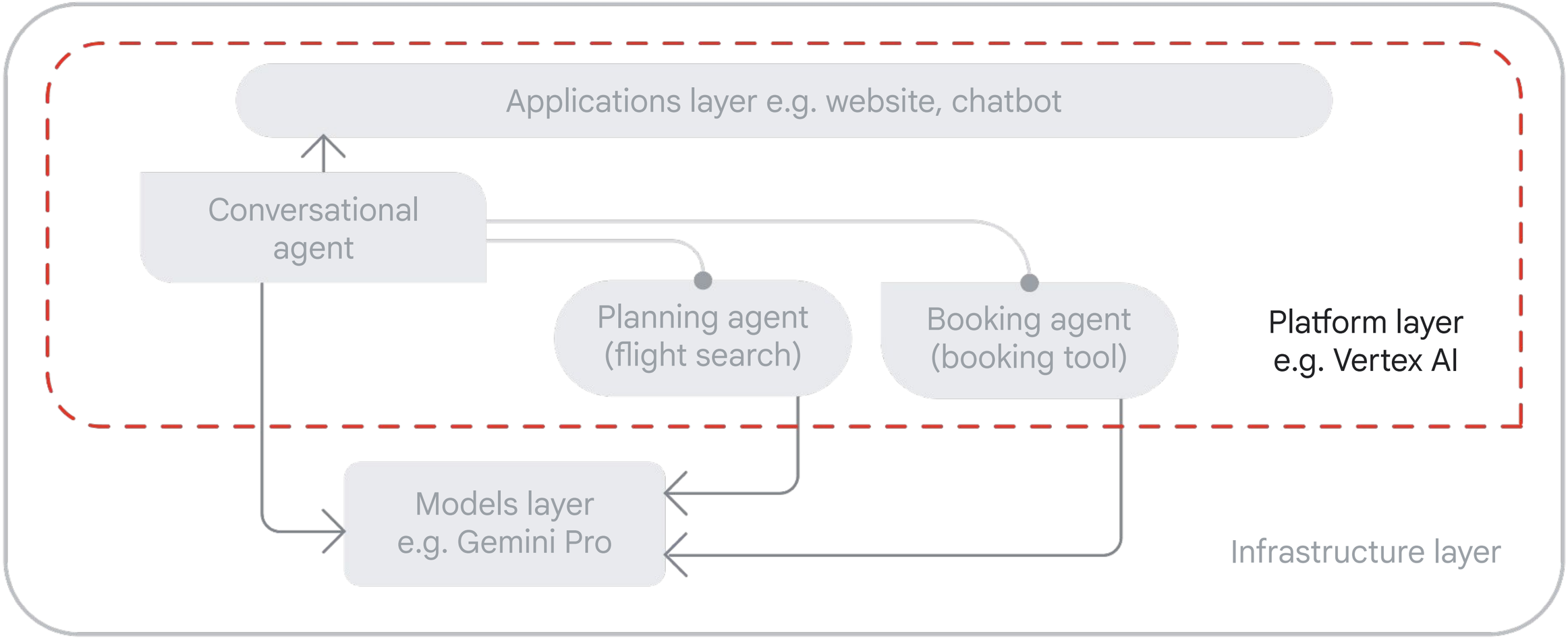
Gen AI landscape:

Travel chatbot example - agents layer



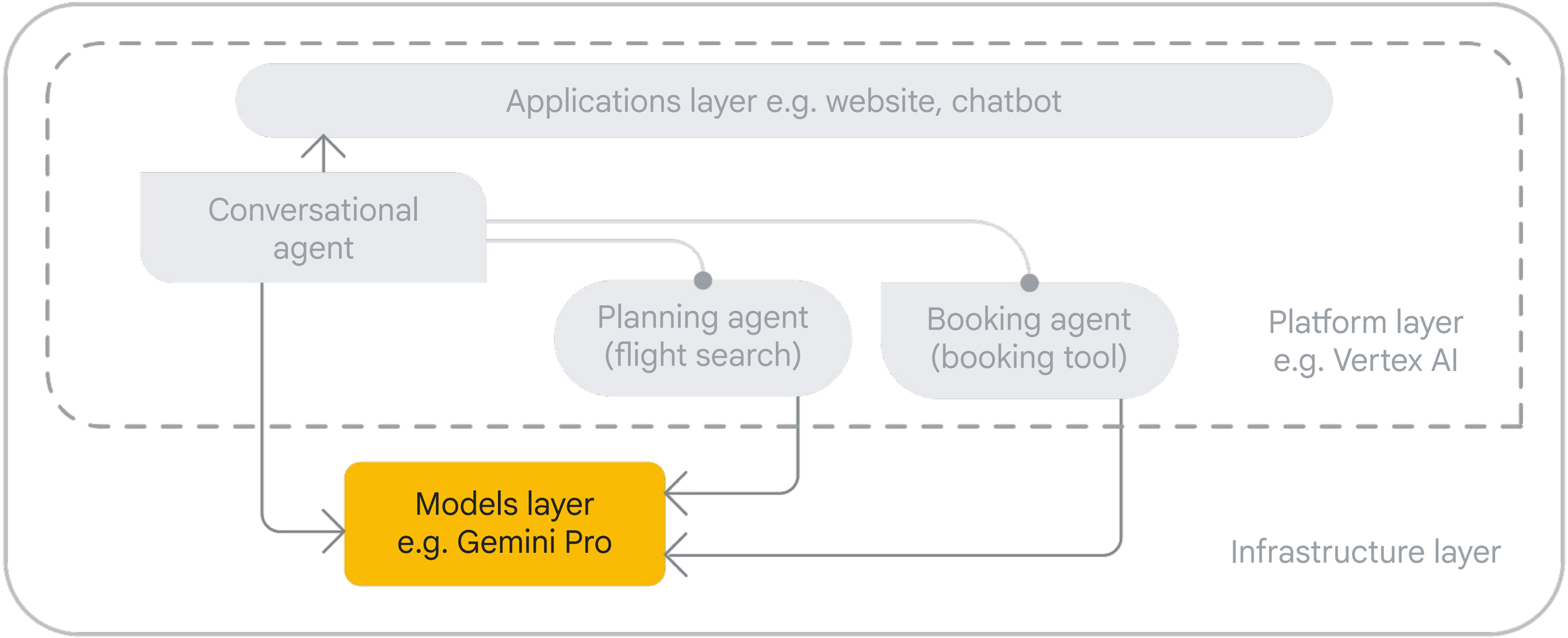
Gen AI landscape:

Travel chatbot example - platform layer



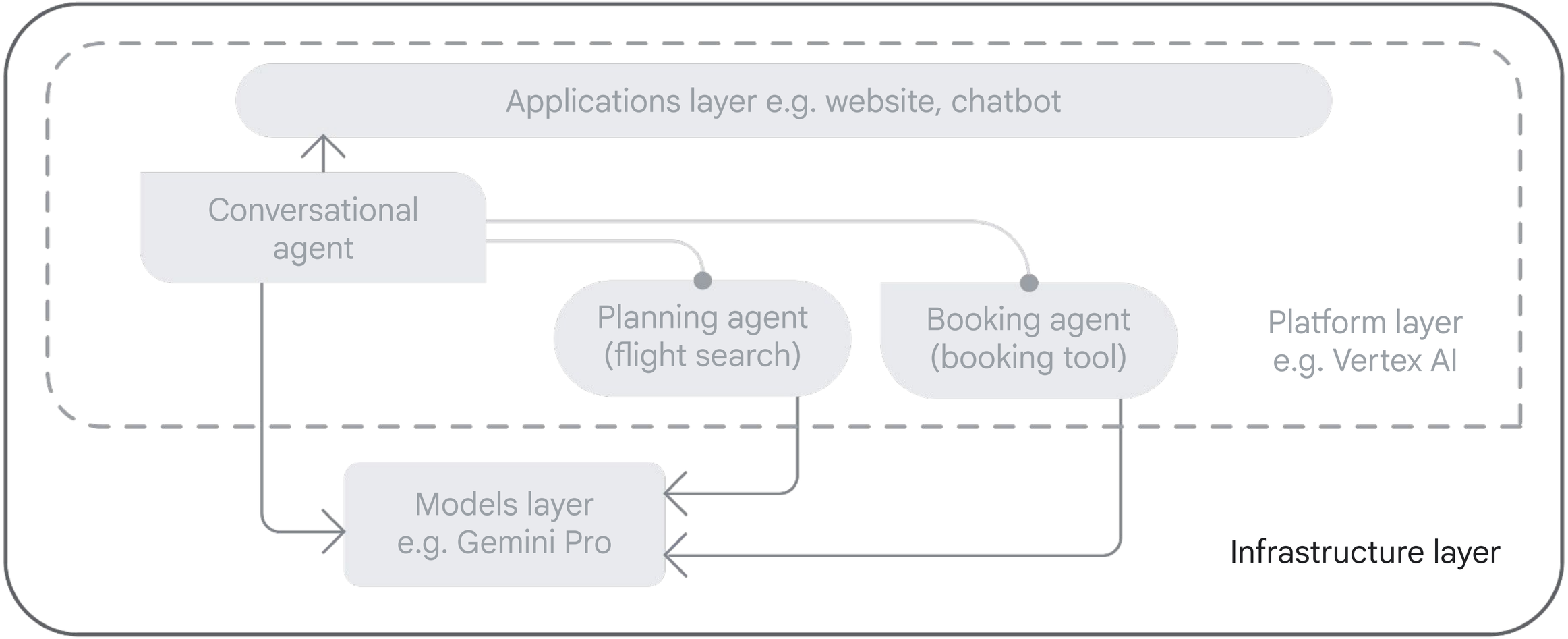
Gen AI landscape:

Travel chatbot example - models layer

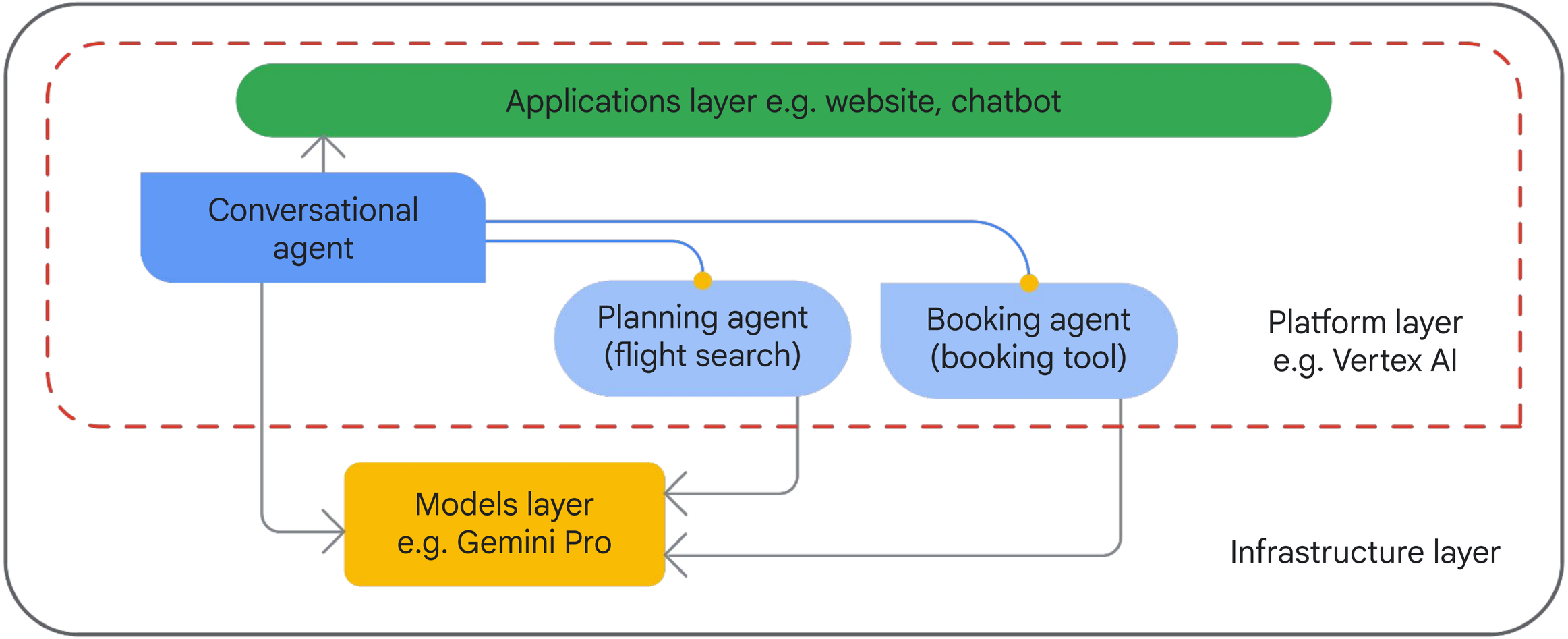


Gen AI landscape:

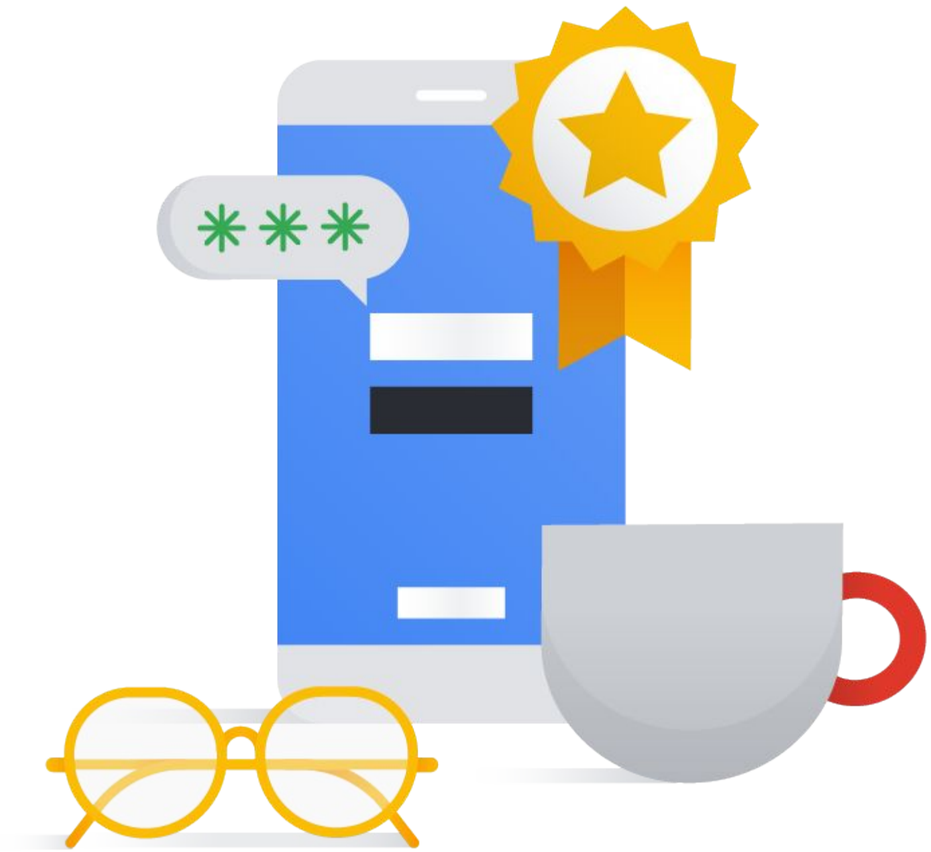
Travel chatbot example - infrastructure layer



Gen AI landscape:
Travel chatbot example



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



Quiz | Question 01

Question

At which of the following stages in the gen AI development process do you need to consider infrastructure requirements if you aren’t building on a platform that handles it for you? Select all that apply.

A. Model refinement: Improving your AI models based on feedback and new data.

B. Model monitoring: Tracking the performance of your deployed AI models and identifying potential issues.

C. Model deployment: Making your AI models available for use in applications or services.

D. Data collection: Gathering and preparing the data used to train your AI models.

E. Model training: Training your AI models on the prepared data.

Quiz | Question 01

Answer

At which of the following stages in the gen AI development process do you need to consider infrastructure requirements if you aren’t building on a platform that handles it for you? Select all that apply.

A. Model refinement: Improving your AI models based on feedback and new data.

B. Model monitoring: Tracking the performance of your deployed AI models and identifying potential issues.

C. Model deployment: Making your AI models available for use in applications or services.

D. Data collection: Gathering and preparing the data used to train your AI models.

E. Model training: Training your AI models on the prepared data.

Quiz | Question 02

Question

What are GPUs and TPUs in the context of AI infrastructure?

A. Software applications for managing AI models.

B. Storage devices for storing large AI datasets.

C. Networking protocols for connecting AI systems.

D. Specialized processors designed for parallel processing in AI tasks.

Quiz | Question 02

Answer

What are GPUs and TPUs in the context of AI infrastructure?

A. Software applications for managing AI models.

B. Storage devices for storing large AI datasets.

C. Networking protocols for connecting AI systems.

D. Specialized processors designed for parallel processing in AI tasks.

Quiz | Question 03

Question

Why is high-performance storage important for generative AI?

A. To store and efficiently access the massive datasets used in AI training.

B. To provide a user-friendly interface for interacting with AI models.

C. To automate the deployment and management of AI models.

D. To ensure the responsible use of AI systems.

Quiz | Question 03

Answer

Why is high-performance storage important for generative AI?

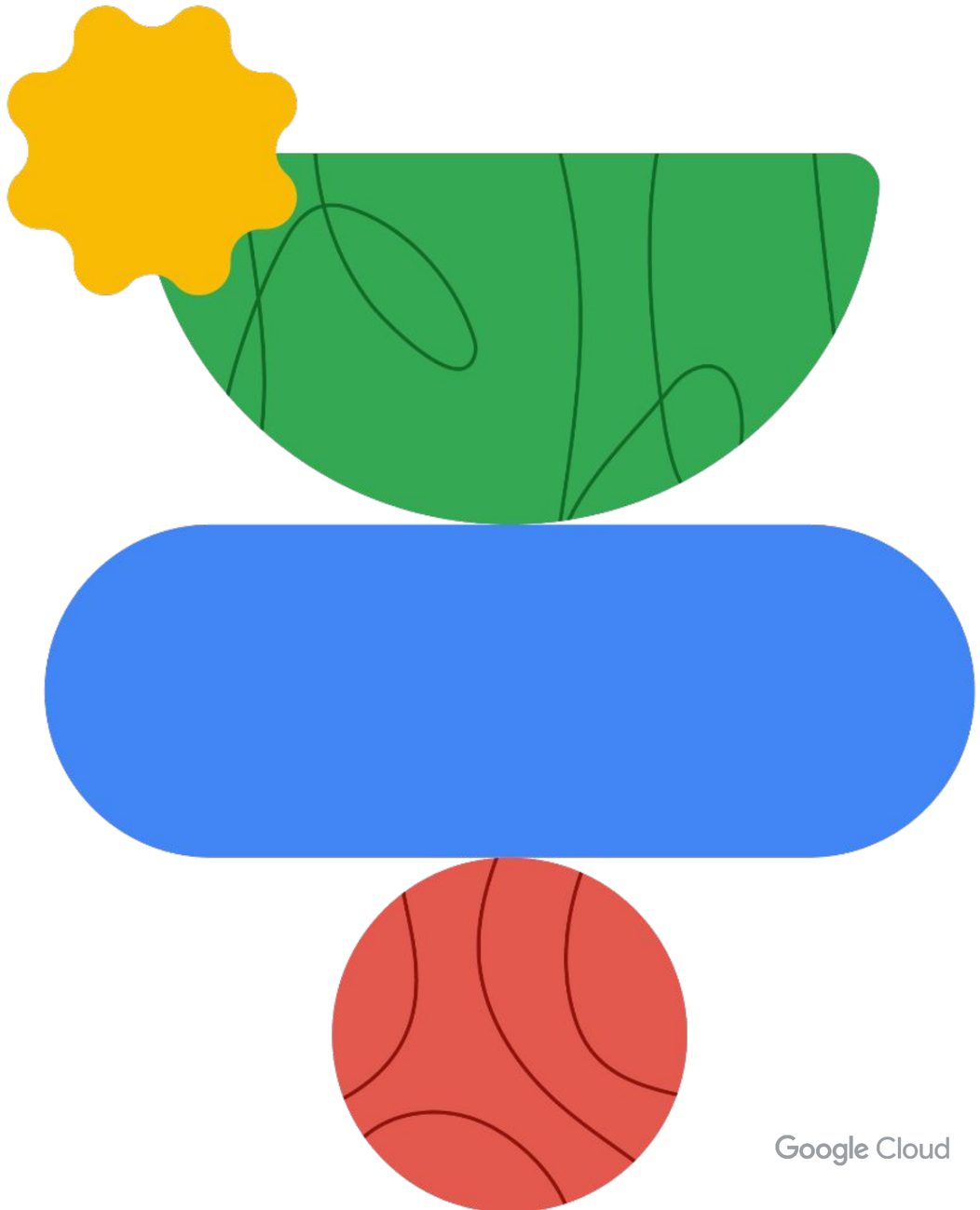
A. To store and efficiently access the massive datasets used in AI training.

B. To provide a user-friendly interface for interacting with AI models.

C. To automate the deployment and management of AI models.

D. To ensure the responsible use of AI systems.

Edge computing

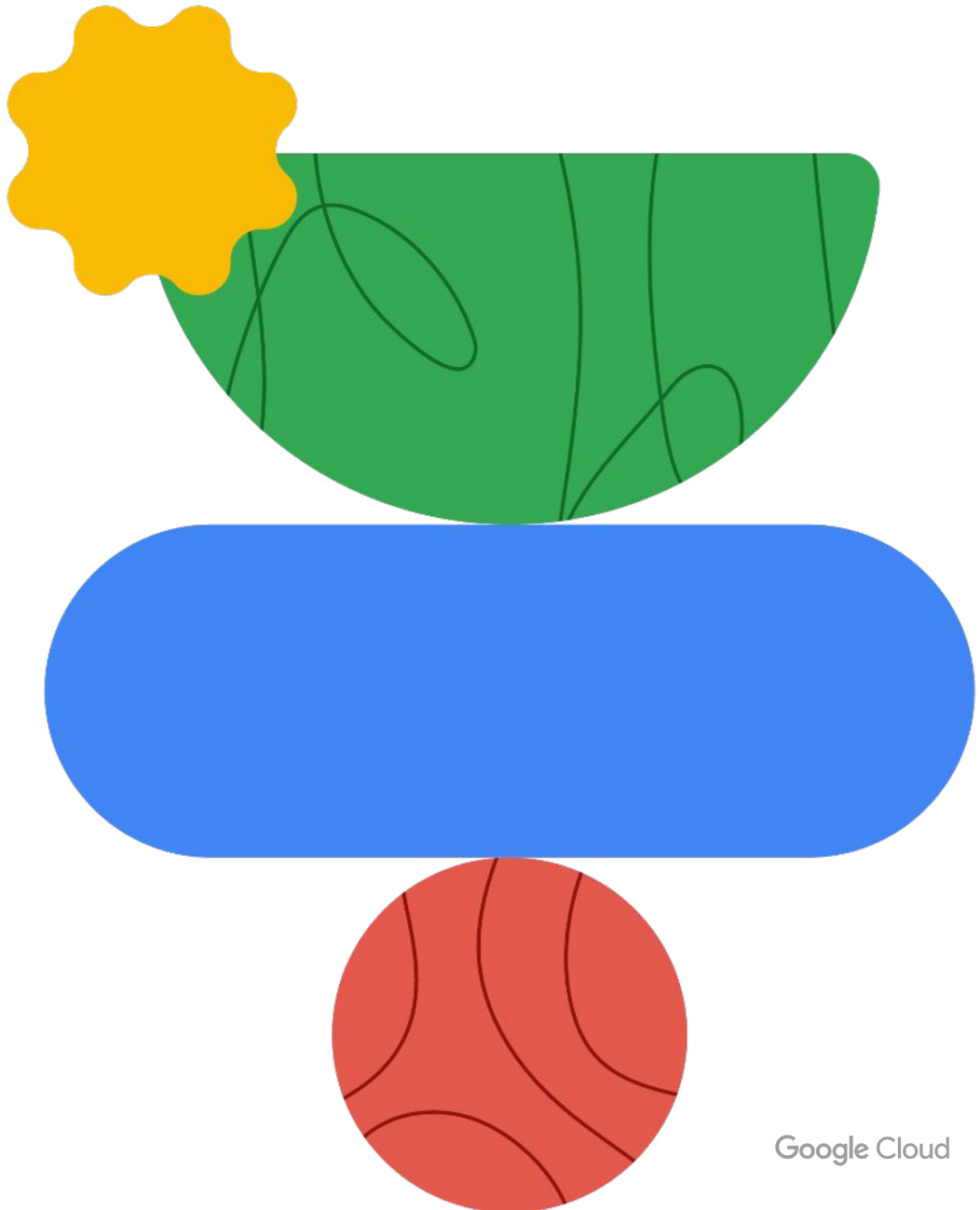


Edge computing runs AI on devices or servers closer to the data source or point of need.

Benefits of going local (or edge)

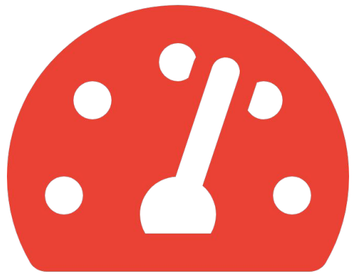
- It ensures real-time responsiveness.
- It increases data privacy.
- It reduces reliance on internet connectivity.
- Lite Runtime (LiteRT) helps ML models work efficiently on edge devices and mobile phones.
- Gemini Nano is an example of an AI model designed for edge.

Gemini Nano



Benefits of Gemini Nano

Privacy



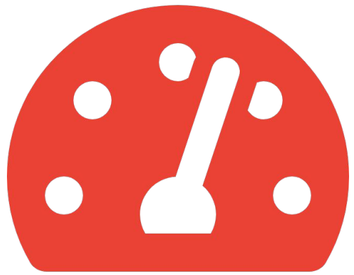
Speed

Offline access

Benefits of Gemini Nano

Privacy

Your data stays on your device, enhancing your privacy.



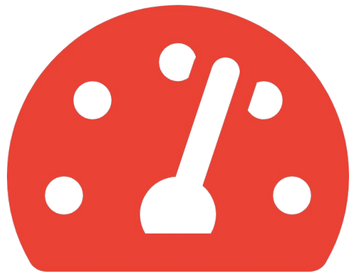
Speed

Offline access

Benefits of Gemini Nano

Privacy

Your data stays on your device, enhancing your privacy.



Speed

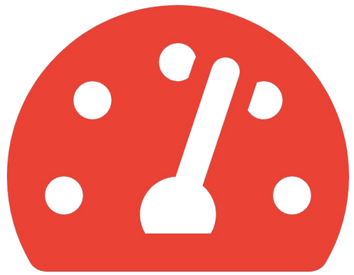
You get fast responses since there's no need to send data to the cloud.

Offline access

Benefits of Gemini Nano

Privacy

Your data stays on your device, enhancing your privacy.



Speed

You get fast responses since there's no need to send data to the cloud.

Offline access

Gemini Nano can work even without an internet connection.

Gemini Nano devices

Pixel phones

They power features like Call Notes, which summarizes phone conversations. Also Pixel Recorder, which summarizes voice recordings.

Android

Available to Android developers through the AI Edge SDK, enabling them to build innovative AI experiences into their apps.

Edge deployment: Vertex AI tools

Convert models

Convert models to Lite Runtime (LiteRT) for optimal performance on edge devices.

Package and deploy

Manage and monitor

Edge deployment: Vertex AI tools

Convert models

Convert models to Lite Runtime (LiteRT) for optimal performance on edge devices.

Package and deploy

Package models and dependencies into containers for deployment on edge hardware.

Manage and monitor

Edge deployment: Vertex AI tools

Convert models

Convert models to Lite Runtime (LiteRT) for optimal performance on edge devices.

Package and deploy

Package models and dependencies into containers for deployment on edge hardware.

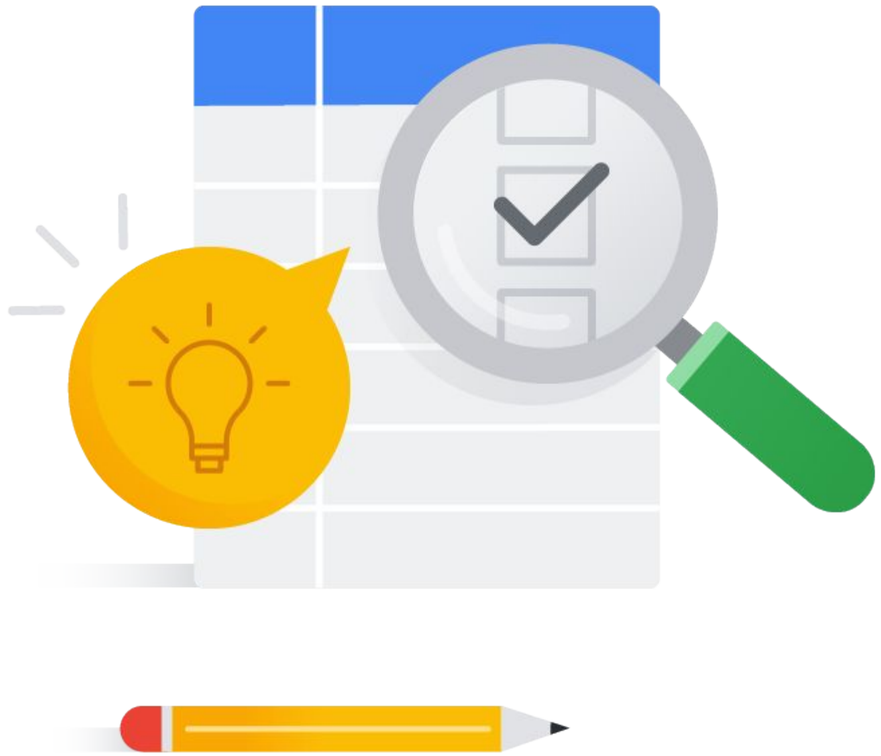
Manage and monitor

Manage edge deployments, track performance, and gather insights to improve models over time.

Activity: Edge or cloud?

5 min

1. Read the scenario.
2. Determine whether an edge deployment or a cloud deployment would be the most suitable solution for the AI. Provide a reason for your choice.
3. Put your answer in the chat.



Scenario: Medical device

Edge or cloud deployment?

A medical device that analyzes patient data in real-time to provide immediate feedback to doctors during surgery.

Scenario: Medical device - Feedback

Edge deployment

A medical device that analyzes patient data in real-time to provide immediate feedback to doctors during surgery.

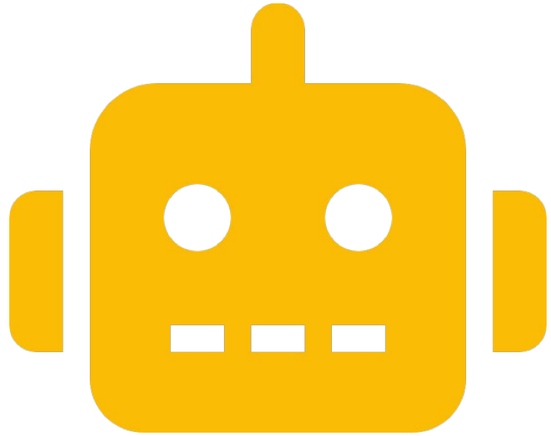
Reason

Real-time analysis is critical during surgery, and relying on the cloud could introduce unacceptable latency.

Scenario: Customer service chatbot

Edge or cloud deployment?

A customer service chatbot for a large ecommerce website that handles millions of inquiries daily.



Scenario: Customer service chatbot - Feedback

Cloud deployment

A customer service chatbot for a large ecommerce website that handles millions of inquiries daily.

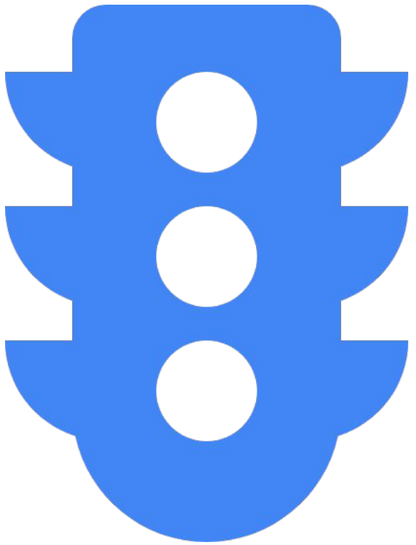
Reason

Handling millions of inquiries requires a scalable and robust cloud infrastructure.

Scenario: Traffic patterns

Edge or cloud deployment?

A system that uses AI to analyze traffic patterns and optimize traffic flow in a smart city.



Scenario: Traffic patterns - Feedback

Cloud deployment

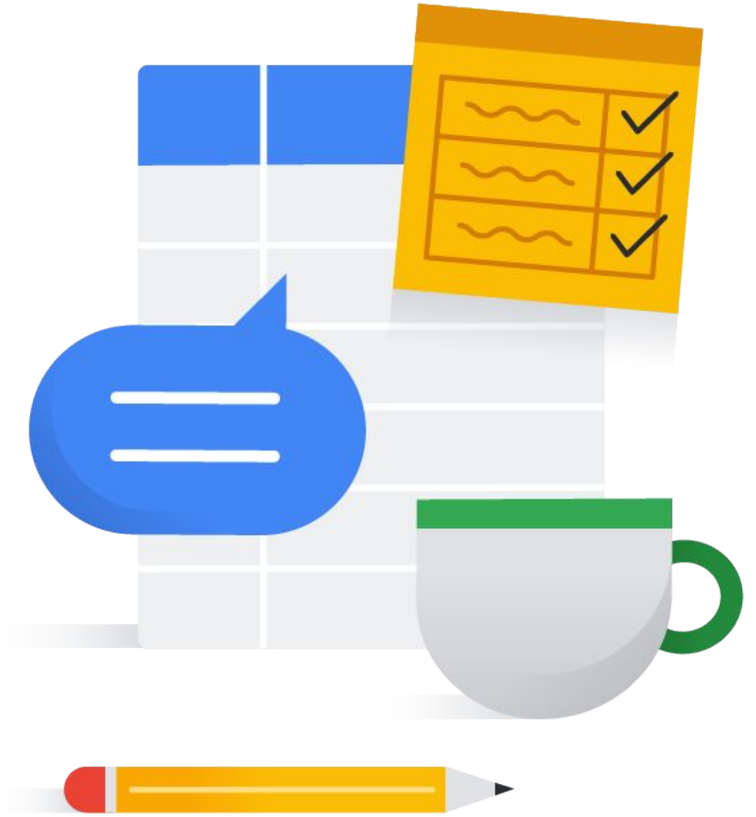
A system that uses AI to analyze traffic patterns and optimize traffic flow in a smart city.

Reason

Analyzing traffic data and optimizing flow requires a centralized system with the capacity of the cloud.

Key takeaways

- Vertex AI streamlines machine learning, offering tools and MLOps to simplify AI development and deployment.
- AI systems rely on models. Vertex AI provides model options, enabling AI innovation. Leaders must understand models for effective AI deployment.
- Google Cloud's infrastructure powers AI. Leaders must understand it for efficient, scalable AI deployment and resource management.
- Edge AI enables real-time responsiveness. Google's tools and Vertex AI support building and deploying models at the edge.



Agenda

01 The gen AI landscape

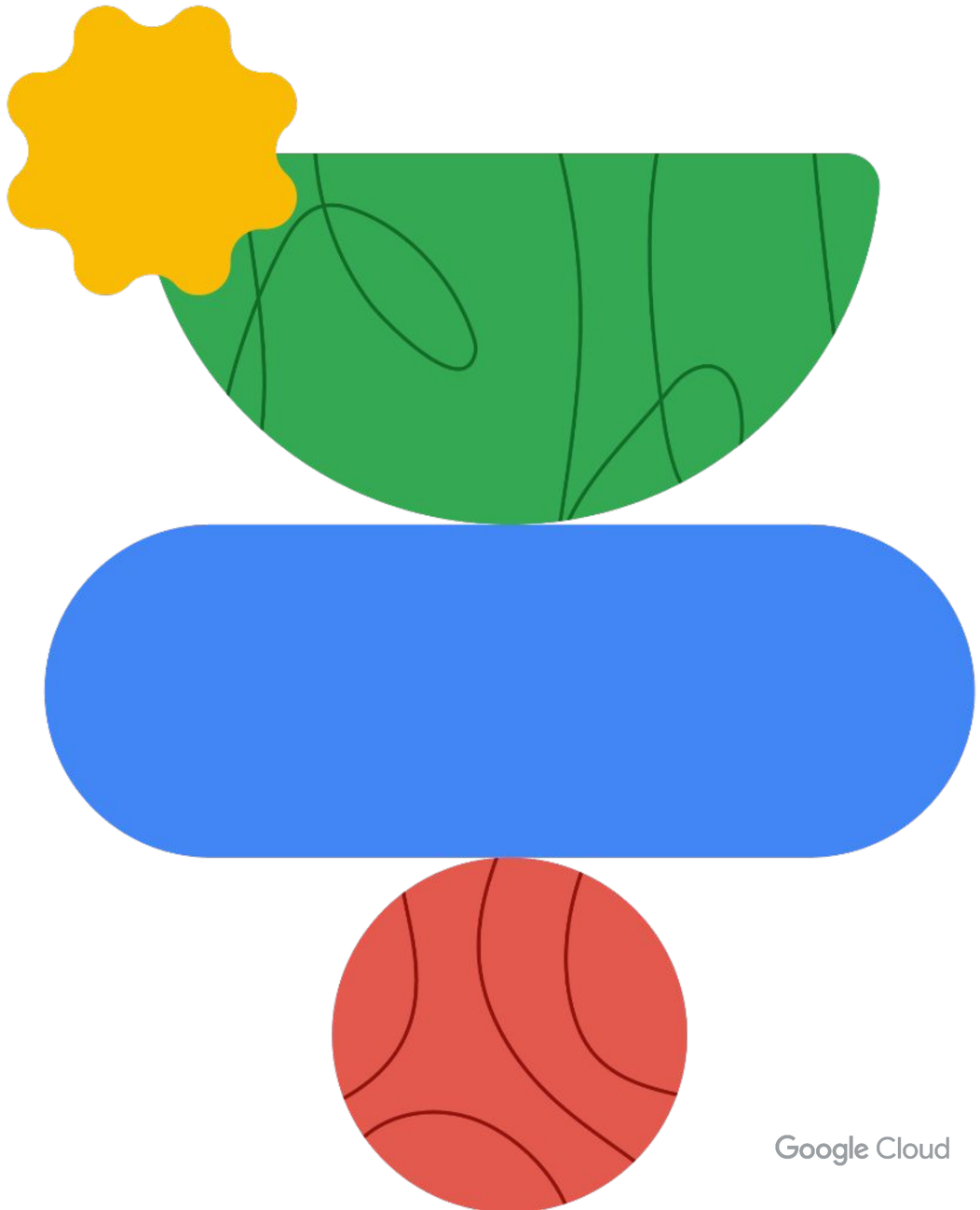
02 Gen AI applications and agents

03 Gen AI platform, models, and infrastructure

04 Gen AI project resources and management

Gen AI project resources and management

Gen AI project resources: People, cost, and time



Gen AI project resources

Scale: Individual, team, company, or millions of customers?

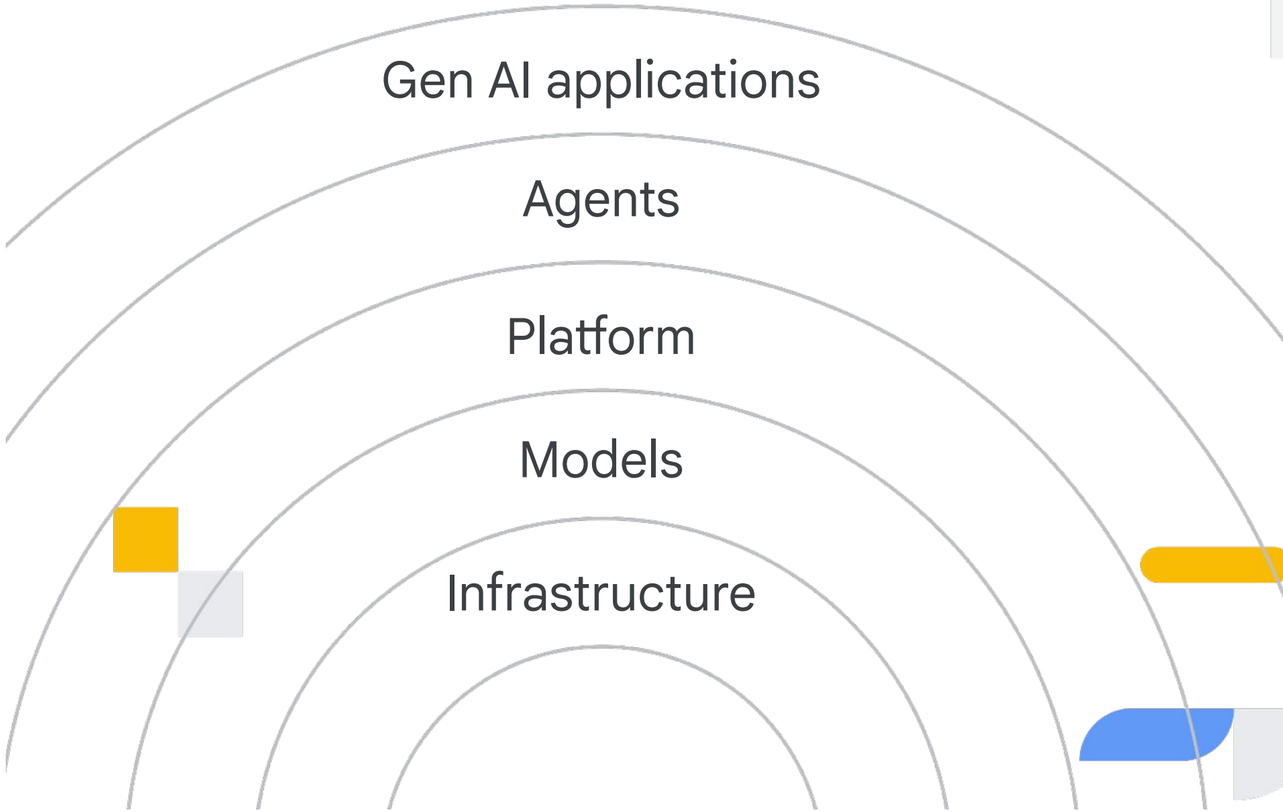
How, where, and when will users be interacting with the solution?

How custom, really? Fine-tuning or brand-new model?

How much time, how many people, and what's the budget?

Roles and responsibilities

- Business leaders
- Developers
- AI practitioners

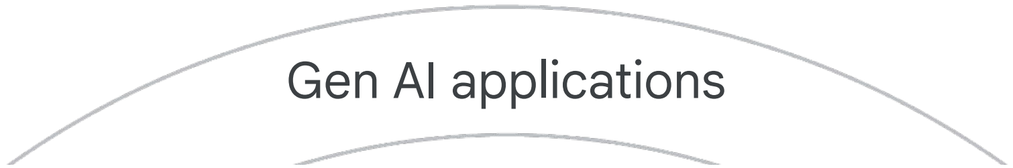


Roles and responsibilities

- Business leaders
- Developers
- AI practitioners

They interact with pre-built gen AI solutions to enhance daily operations and improve customer experiences.

Example: Google Workspace with Gemini.



Roles and responsibilities

- Business leaders
- Developers
- AI practitioners

They are responsible for building and deploying custom AI agents and integrating AI capabilities into existing applications.

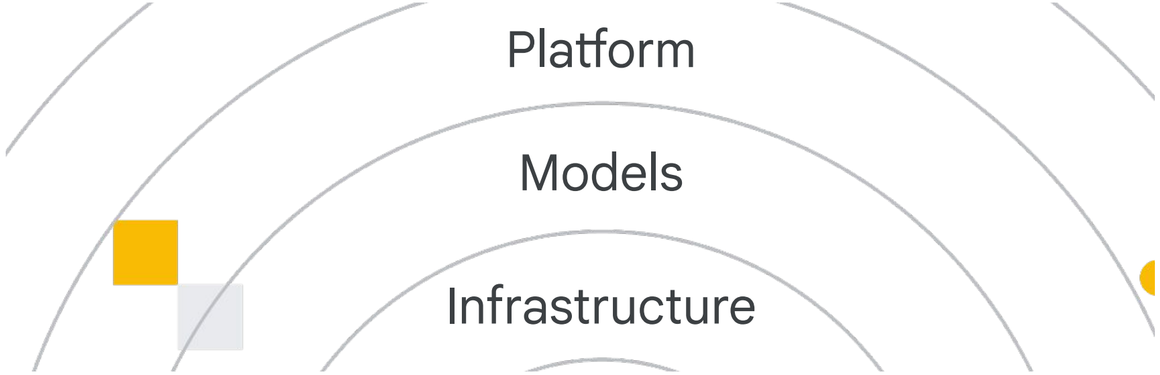
Examples: AI Applications, AI code generation, AI-driven data processing, pre-trained APIs, Vertex AI platform.

Roles and responsibilities

- Business leaders
- Developers
- AI practitioners

They are responsible for customizing, deploying, and optimizing generative AI models.

Examples: Vertex AI, scaling AI workloads, integrating models with BigQuery, implementing responsible AI measures.



Cost

When building gen AI solutions, you pay for three primary activities:

- **Training*** the model.
- **Deploying*** the model to an endpoint.
- **Using** the model to make predictions.

*Compute time, storage for the training data, and model outputs.

Pricing for using models

Usage-based

- Pay for the amount used.
- Pay for tokens or characters processed.

Subscription-based

Licensing fees

Free tiers

Pricing for using models

Usage-based

- Pay for the amount used.
- Pay for tokens or characters processed.

Subscription-based

- Pay a recurring fee for access to the model.
- Usage limits or features.

Licensing fees

Free tiers

Pricing for using models

Usage-based

- Pay for the amount used.
- Pay for tokens or characters processed.

Subscription-based

- Pay a recurring fee for access to the model.
- Usage limits or features.

Licensing fees

- One-time or recurring fees for using a model.
- Commercial purposes or embedding in products.

Free tiers

Pricing for using models

Usage-based

- Pay for the amount used.
- Pay for tokens or characters processed.

Subscription-based

- Pay a recurring fee for access to the model.
- Usage limits or features.

Licensing fees

- One-time or recurring fees for using a model.
- Commercial purposes or embedding in products.

Free tiers

- Free access with limited usage.
- Experimentation or non-commercial purposes.

Pricing metrics for using models

- Tokens
- Characters
- Requests
- Compute time

Pricing metrics for using models

- Tokens
- Characters
- Requests
- Compute time

They represent a piece of text, like a word or part of a word.

Pricing metrics for using models

- Tokens
- Characters
- Requests
- Compute time

Charge is based on the number of characters processed.

Pricing metrics for using models

- Tokens
- Characters
- Requests
- Compute time

Charge is made per request, regardless of the complexity or volume.

Pricing metrics for using models

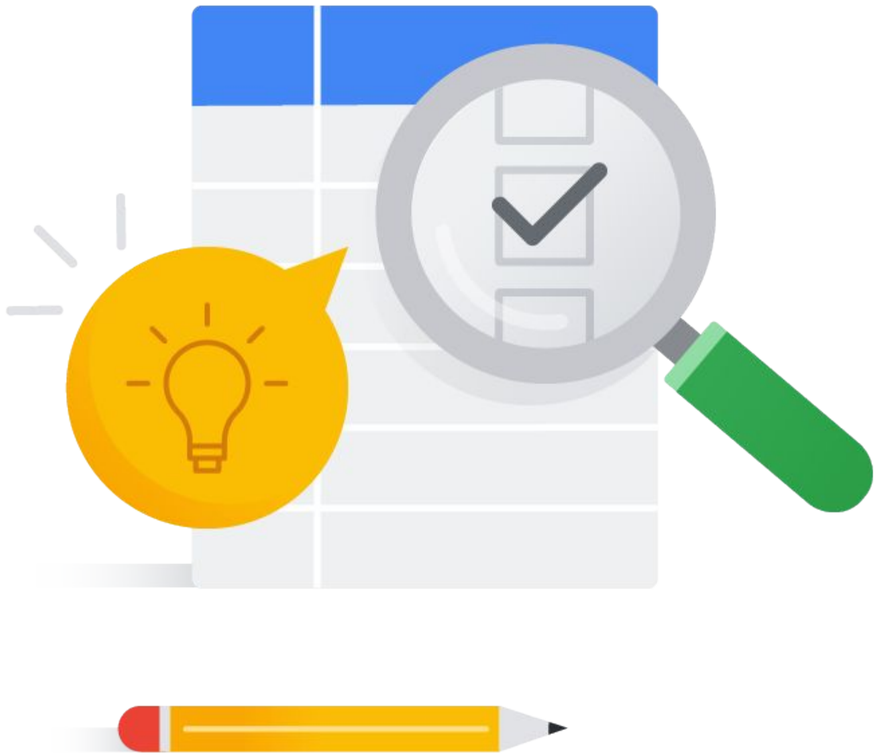
- Tokens
- Characters
- Requests
- Compute time

Time taken to process requests can also factor into the cost, especially for resource-intensive tasks.

Demo: Google Cloud's pricing calculator

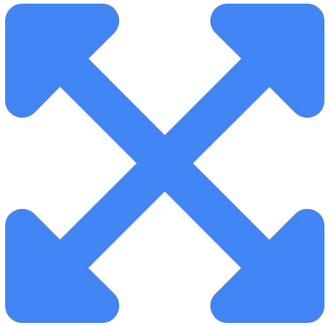
5 min

Google Cloud’s pricing calculator:
https://cloud.google.com/products/calculator?hl=en



Additional resource: Cost of building and deploying AI models in Vertex AI

Factors affecting cost



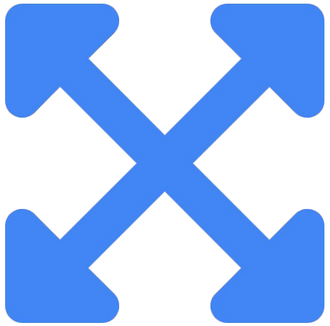
Model size and complexity

Context window

Features

Deployment

Factors affecting cost



Model size and complexity

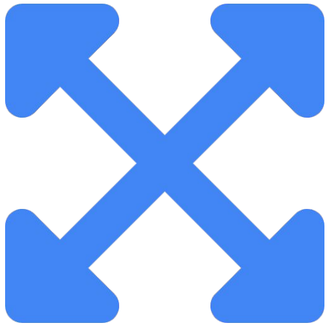
Larger, more capable models generally cost more.

Context window

Features

Deployment

Factors affecting cost



Model size and complexity

Larger, more capable models generally cost more.

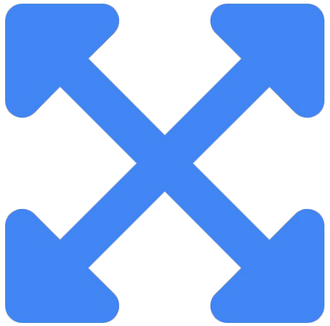
Context window

A larger context window can increase costs.

Features

Deployment

Factors affecting cost



Model size and complexity

Larger, more capable models generally cost more.

Context window

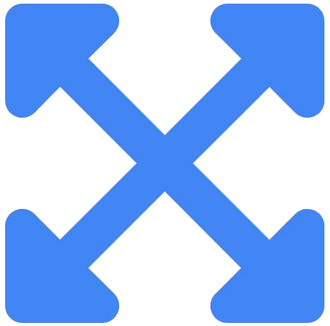
A larger context window can increase costs.

Features

Specialized features can have separate pricing.

Deployment

Factors affecting cost



Model size and complexity

Larger, more capable models generally cost more.

Context window

A larger context window can increase costs.

Features

Specialized features can have separate pricing.

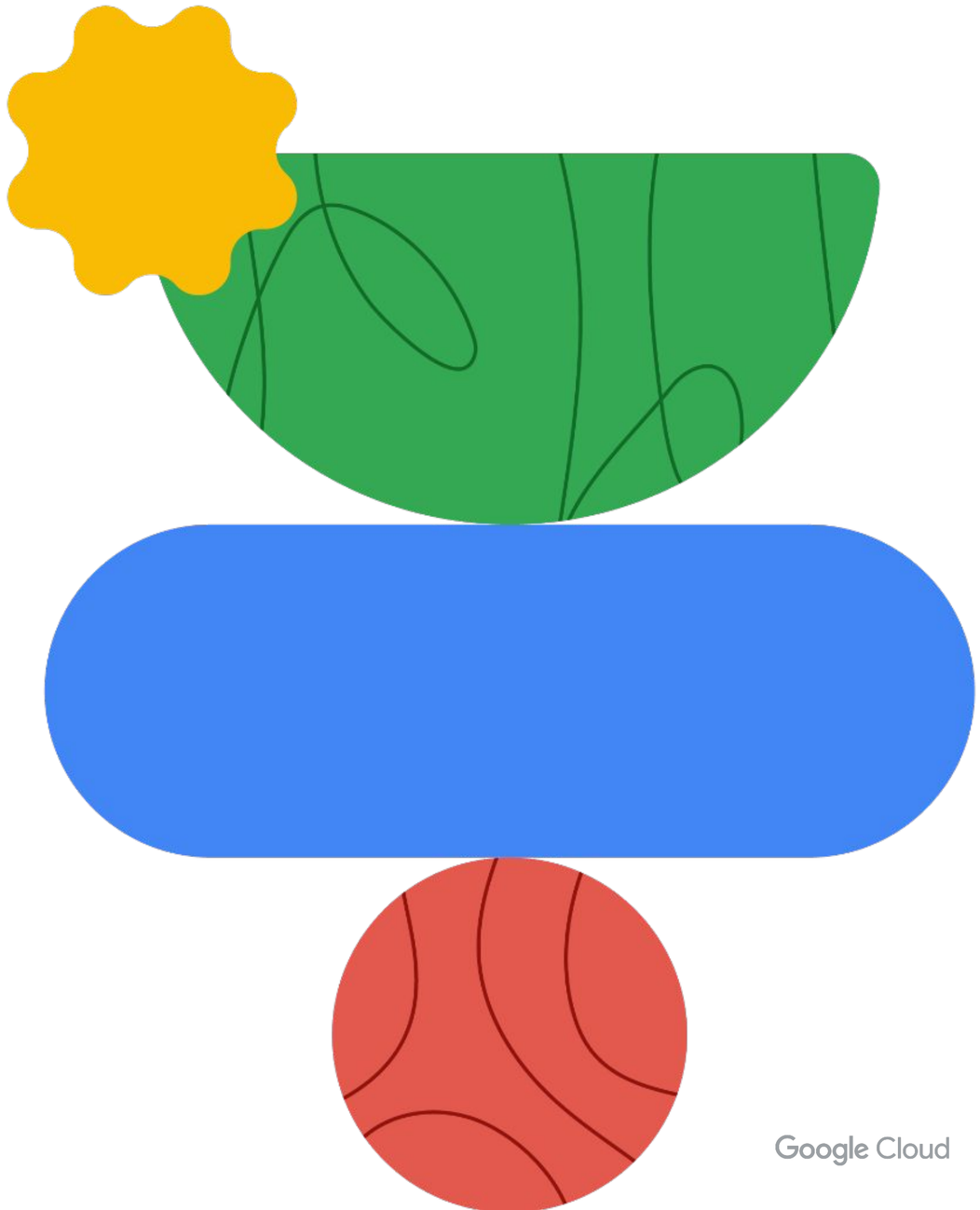
Deployment

It may have compute-based costs.

Time

- The more custom your solution is, the more time and resources it takes to build.
- Think about your project timelines, and evaluate them against your needs and requirements.

Gen AI solution needs



Scale

Small scale

- Select pre-built tools.
- Leverage existing gen AI-powered applications.

Large scale

- Select for scalability and security.
- Consider infrastructure costs, data storage, and latency challenges.

Customization

- Start with existing models
- Identify your unique needs
- Consider data specificity
- Consider task complexity

Customization

- Start with existing models
- Identify your unique needs
- Consider data specificity
- Consider task complexity

- Explore pre-trained models.
- Fine-tune these models on your specific data.

Customization

- Start with existing models
- Identify your unique needs
- Consider data specificity
- Consider task complexity

- What sets your project apart?
- Does it require specialized knowledge, handle complex tasks, or demand a unique user experience?

Customization

- Start with existing models
- Identify your unique needs
- Consider data specificity
- Consider task complexity

For specialized domains:

- Fine-tune with domain-specific datasets.
- Explore models specifically trained for those areas.

Customization

- Start with existing models
- Identify your unique needs
- Consider data specificity
- Consider task complexity

- Is it simple or intricate?
- Complexity impacts model choice and training approaches.

User interaction

User interface (UI)

User experience (UX)

User interaction

User interface (UI)

- Integrate AI seamlessly into your existing workflows.
- Use a dedicated interface, or embed AI capabilities within your current applications.

User experience (UX)

User interaction

User interface (UI)

- Integrate AI seamlessly into your existing workflows.
- Use a dedicated interface, or embed AI capabilities within your current applications.

User experience (UX)

- Aim for a user-friendly experience.
- Decide if it must be conversational, informative, or task-oriented.
- Consider the level of guidance and feedback users might need.

Privacy

Data security

- Measures to implement to protect data during processing and storage.
- Encryption, access controls, and secure data centers.

Compliance

- Specific regulations to adhere to.

Other considerations

- Latency
- Connectivity
- Accuracy
- Explainability

Other considerations

- Latency
- Connectivity
- Accuracy
- Explainability

- What are the acceptable response time limits?
- Consider real-time requirements, instantaneous versus some delay.

Other considerations

- Latency
- Connectivity
- Accuracy
- Explainability

- Will the solution always have internet access?
- Consider offline functionality.

Other considerations

- Latency
- Connectivity
- Accuracy
- Explainability

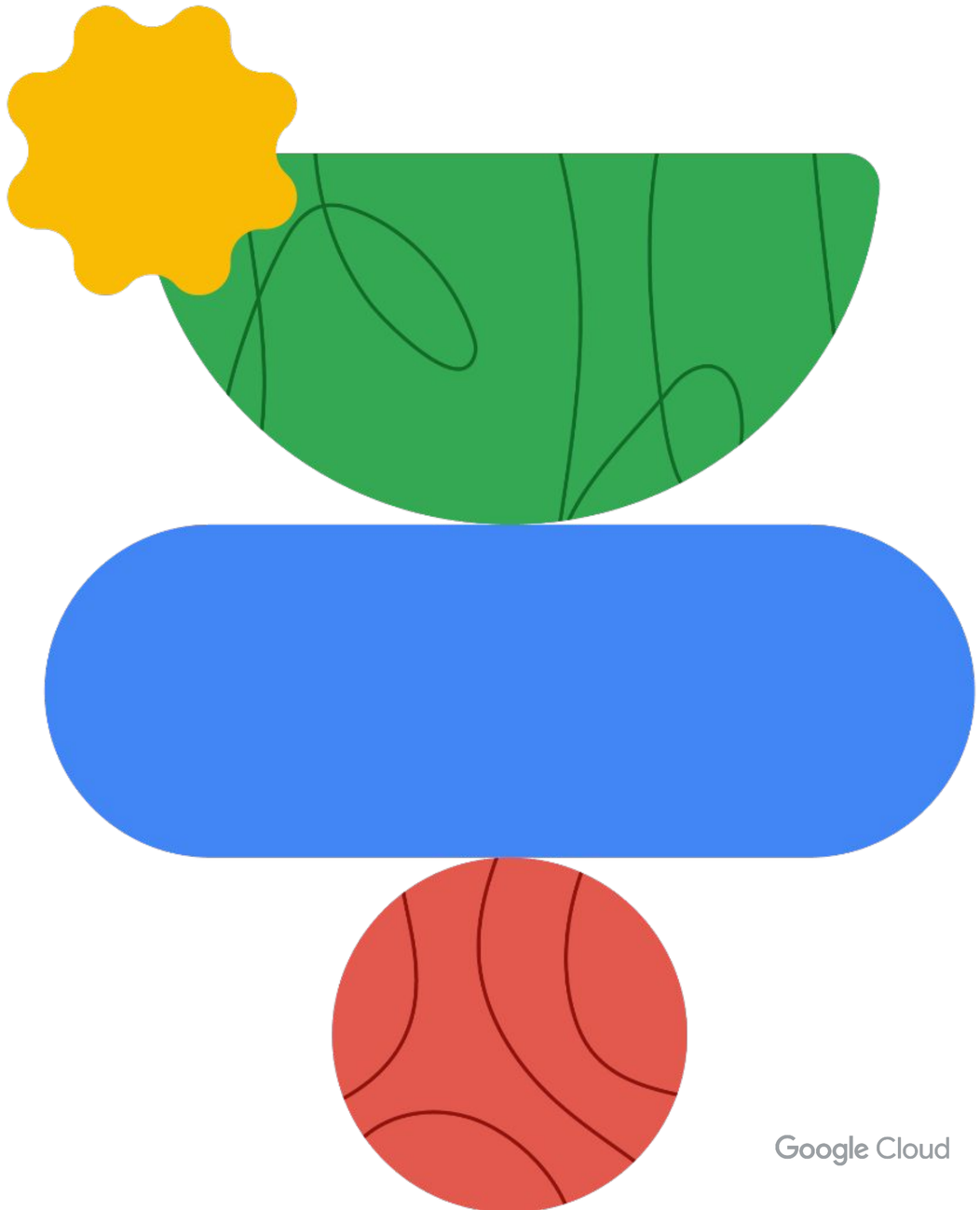
- What are the acceptable error tolerances?
- Define acceptable tolerances.

Other considerations

- Latency
- Connectivity
- Accuracy
- Explainability

- Do you need to understand the reasoning behind the AI's decisions?
- Transparency is key. In certain domains, understanding the AI's reasoning is crucial.

Decision-making and maintenance



Making decisions

01 Comparison of different companies and models

02 Key resources for comparison

Making decisions

01 Comparison of different companies and models

02 Key resources for comparison

- Evaluate model capabilities.
- Compare pricing structures.
- Factor in additional costs.
- Read the fine print.

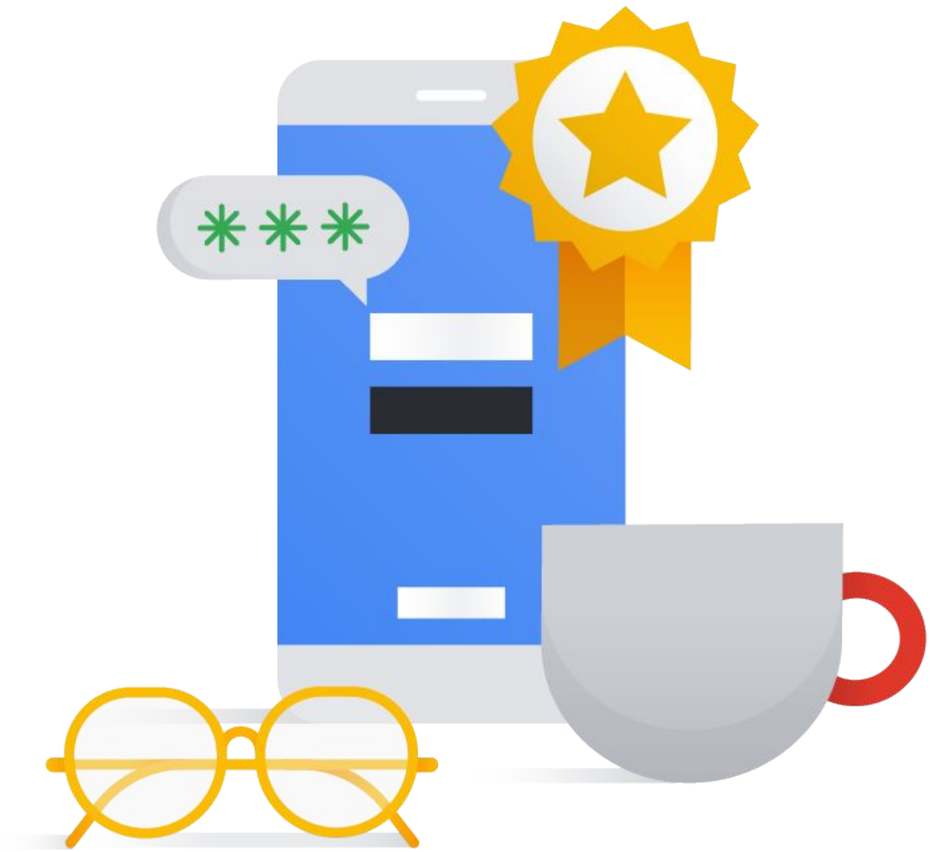
Making decisions

01 Comparison of different companies and models

02 Key resources for comparison

- Provider websites.
- Research papers and benchmarks.
- Community forums and discussions.

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



Quiz | Question 01

Question

What is a key factor to consider when determining how custom your gen AI solution needs to be?

A. The color scheme of the user interface.

B. Whether grounding or fine-tuning is sufficient or a brand-new model is required.

C. The number of lines of code used.

D. The font size of the text.

Quiz | Question 01

Answer

What is a key factor to consider when determining how custom your gen AI solution needs to be?

A. The color scheme of the user interface.

B. Whether grounding or fine-tuning is sufficient or a brand-new model is required.

C. The number of lines of code used.

D. The font size of the text.

Quiz | Question 02

Question

Which of the following best describes a primary responsibility of AI practitioners in the context of generative AI?

A. Utilizing pre-built gen AI solutions to improve daily business tasks.

B. Building and deploying custom AI agents and integrating AI into existing applications.

C. Customizing, deploying, and optimizing generative AI models while ensuring responsible AI practices.

D. Interacting with AI-powered applications to enhance user experiences.

Quiz | Question 02

Answer

Which of the following best describes a primary responsibility of AI practitioners in the context of generative AI?

A. Utilizing pre-built gen AI solutions to improve daily business tasks.

B. Building and deploying custom AI agents and integrating AI into existing applications.

C. Customizing, deploying, and optimizing generative AI models, while ensuring responsible AI practices.

D. Interacting with AI-powered applications to enhance user experiences.

Quiz | Question 03

Question

Which of the following is a primary cost associated with building gen AI solutions?

A. Initial server setup

B. Model training

C. User interface design

D. Regulatory compliance checks

Quiz | Question 03

Answer

Which of the following is a primary cost associated with building gen AI solutions?

A. Initial server setup

B. Model training 

C. User interface design

D. Regulatory compliance checks

Maintenance

- How will your project be maintained?
- Do you have the resources in place to maintain the project over time?
- What specific maintenance needs will your project have?

Key maintenance considerations

Model monitoring and retraining

- Continuously monitor model performance.
- Retrain it periodically.

Data updates

Software updates and bug fixes

Key maintenance considerations

Model monitoring and retraining

- Continuously monitor model performance.
- Retrain it periodically.

Data updates

- Plan regular updates to keep the model fresh and relevant.

Software updates and bug fixes

Key maintenance considerations

Model monitoring and retraining

- Continuously monitor model performance.
- Retrain it periodically.

Data updates

- Plan regular updates to keep the model fresh and relevant.

Software updates and bug fixes

- Stay informed about updates.
- Implement updates properly.

Additional key maintenance considerations

Hardware and infrastructure

- Server maintenance.
- Security updates.
- Capacity planning.

Security and compliance

Additional key maintenance considerations

Hardware and infrastructure

- Server maintenance.
- Security updates.
- Capacity planning.

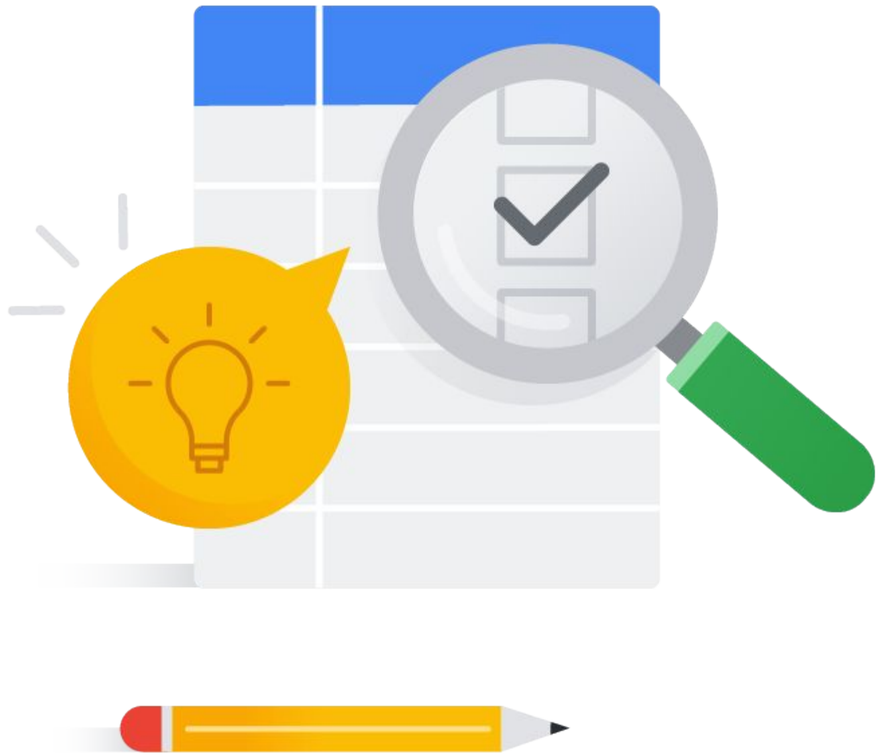
Security and compliance

- Regularly review and update security measures.
- Ensure ongoing compliance with data privacy regulations.

Activity: Best solution

5 min

1. Read the scenario and proposed solutions.
2. Determine which is the best proposed solution.
3. Put your answer in the chat.



Scenario 1

Requirement

You need a model that can generate realistic images from text descriptions for your new marketing campaign. You have some coding skills but are short on time.

What is the best proposal for this solution?

A. Build a custom image generation model from scratch.

B. Fine-tune a pre-trained text-to-image model from Model Garden.

C. Use a pre-trained API like Imagen to generate images quickly.

Scenario 1

Requirement

You need a model that can generate realistic images from text descriptions for your new marketing campaign. You have some coding skills but are short on time.

What is the best proposal for this solution?

A. Build a custom image generation model from scratch.

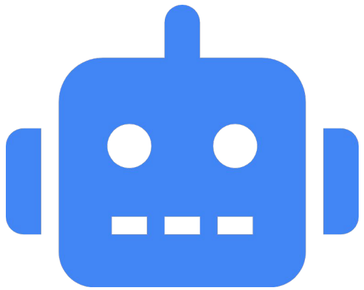
B. Fine-tune a pre-trained text-to-image model from Model Garden.

C. Use a pre-trained API like Imagen to generate images quickly.

Scenario 2

Requirement

You want to build a chatbot that can provide personalized movie recommendations to users. You have a basic understanding of AI concepts but no coding experience.



What is the best proposal for this solution?

A. Use a pre-trained language model from Model Garden and fine-tune it on movie data.

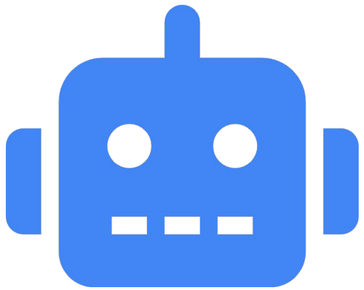
B. Use AI Applications to create a conversational agent.

C. Build a custom language model from scratch using Python.

Scenario 2

Requirement

You want to build a chatbot that can provide personalized movie recommendations to users. You have a basic understanding of AI concepts but no coding experience.



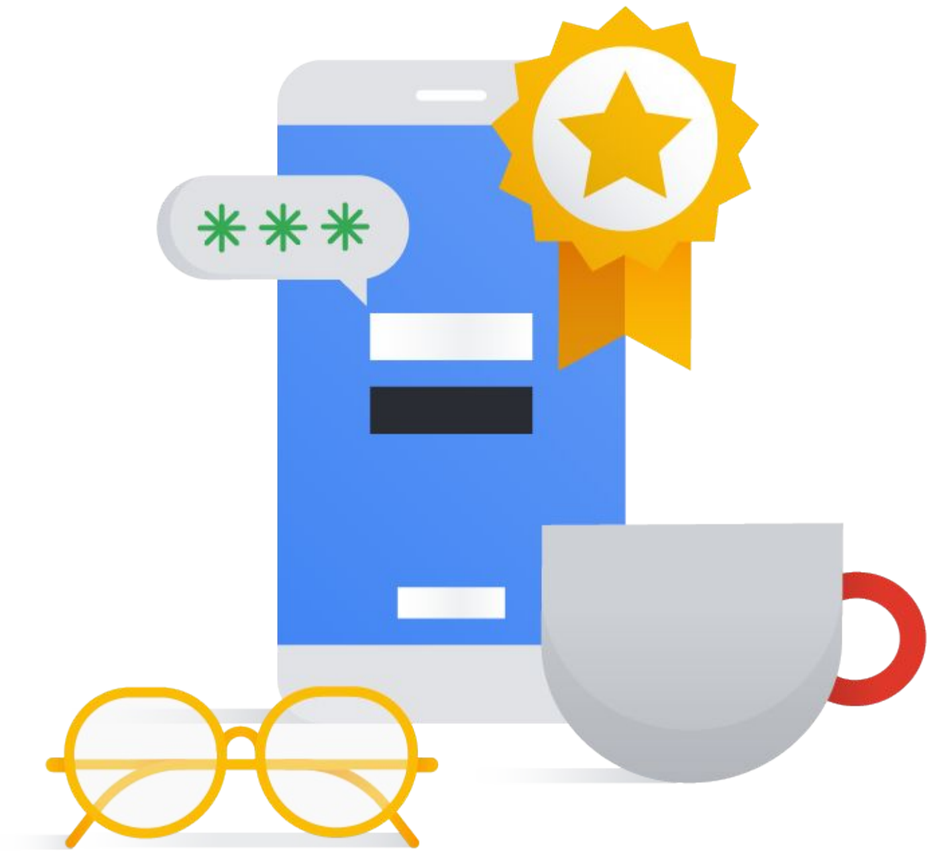
What is the best proposal for this solution?

A. Use a pre-trained language model from Model Garden and fine-tune it on movie data.

B. Use AI Applications to create a conversational agent.

C. Build a custom language model from scratch using Python.

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



Quiz | Question 01

Question

Who is responsible for building and deploying custom AI agents and integrating AI capabilities into applications?

A. Business leaders

B. Developers

C. AI practitioners

D. Network engineers

Quiz | Question 01

Answer

Who is responsible for building and deploying custom AI agents and integrating AI capabilities into applications?

A. Business leaders

B. Developers 

C. AI practitioners

D. Network engineers

Quiz | Question 02

Question

Your marketing team needs to quickly generate engaging product descriptions for an upcoming ecommerce sale. Which approach is the most efficient way to achieve this?

A. Build a custom model from scratch to generate product descriptions.

B. Fine-tune a pre-trained language model from Model Garden on existing product data.

C. Hire a team of copywriters to write each description manually.

D. Use a gen AI powered application like Google Workspace with Gemini to draft the descriptions.

Quiz | Question 02

Answer

Your marketing team needs to quickly generate engaging product descriptions for an upcoming ecommerce sale. Which approach is the most efficient way to achieve this?

A. Build a custom model from scratch to generate product descriptions.

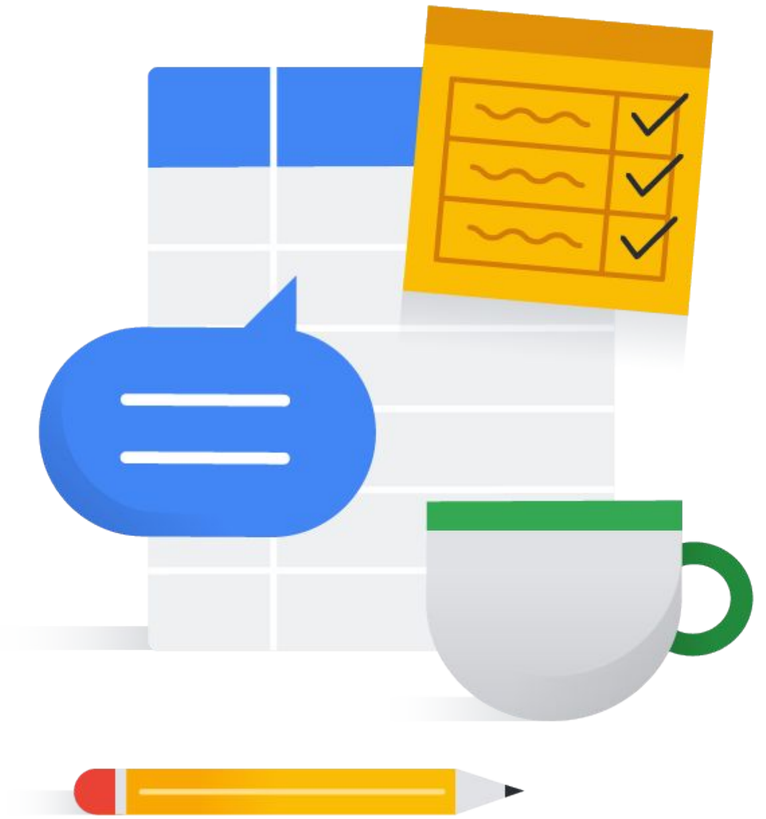
B. Fine-tune a pre-trained language model from Model Garden on existing product data.

C. Hire a team of copywriters to write each description manually.

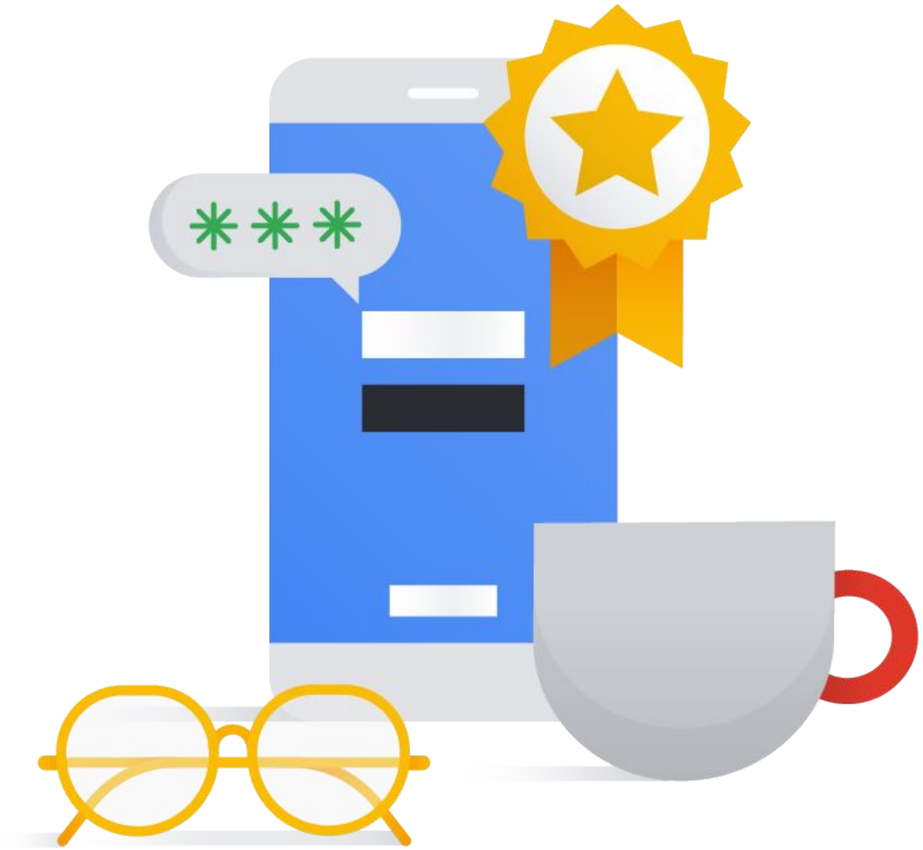
D. Use a gen AI powered application like Google Workspace with Gemini to draft the descriptions.

Key takeaways

- Generative AI success requires strategic resource allocation, cost awareness, and realistic timelines based on solution complexity.
- Plan for AI's future. Adapt, update, and optimize to ensure continued value and avoid costly errors.



Now let's wrap up with a quiz to check your knowledge on Module 03.



Quiz | Question 01

Question

What are the key elements that distinguish AI agents from standalone AI models?

A. Natural language processing and machine learning.

B. Reasoning loop and tools.

C. Data access and user interface.

D. Automation and personalization.

Quiz | Question 01

Answer

What are the key elements that distinguish AI agents from standalone AI models?

A. Natural language processing and machine learning

B. Reasoning loop and tools 

C. Data access and user interface

D. Automation and personalization

Quiz | Question 02

Question

What is the primary advantage of edge computing for AI applications?

A. Increased data storage capacity

B. Reduced need for data preprocessing

C. Real-time responsiveness and reduced latency

D. Lower development costs

Quiz | Question 02

Question

What is the primary advantage of edge computing for AI applications?

A. Increased data storage capacity

B. Reduced need for data preprocessing

C. Real-time responsiveness and reduced latency 

D. Lower development costs

Quiz | Question 03

Question

What is an important step when making decisions about Gen AI solutions?

A. Comparing different companies and models.

B. Prioritizing cutting-edge features.

C. Prioritizing ease of initial deployment over understanding data governance and compliance implications.

D. Focusing on the initial integration effort and later considering long-term maintenance and scalability.

Quiz | Question 03

Answer

What is an important step when making decisions about Gen AI solutions?

A. Comparing different companies and models. 

B. Prioritizing cutting-edge features.

C. Prioritizing ease of initial deployment over understanding data governance and compliance implications.

D. Focusing on the initial integration effort and later considering long-term maintenance and scalability.

Module objectives

01 Describe the layers of the gen AI landscape.

02 Identify entry points in the gen AI landscape to address business needs and innovation.

03 Describe components of the Google Cloud gen AI portfolio.

04 Explain how Google Cloud's AI-optimized resources support gen AI development.

05 Describe business factors to consider for specific applications.

Additional resources

Lesson 03

- Example models and specs

Lesson 03

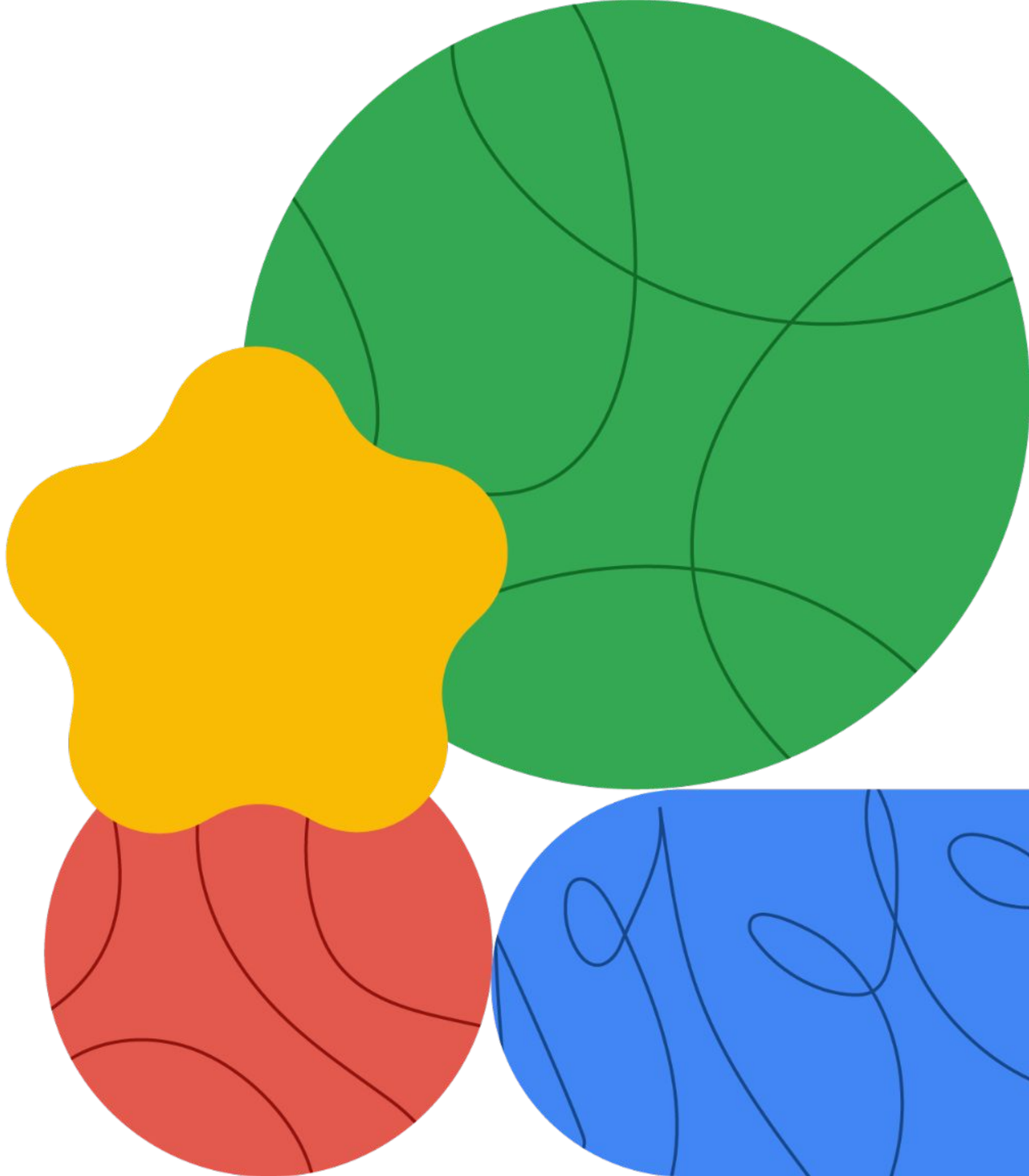
- Lite Runtime (LiteRT) Overview

- Cost of building and deploying AI models in Vertex AI

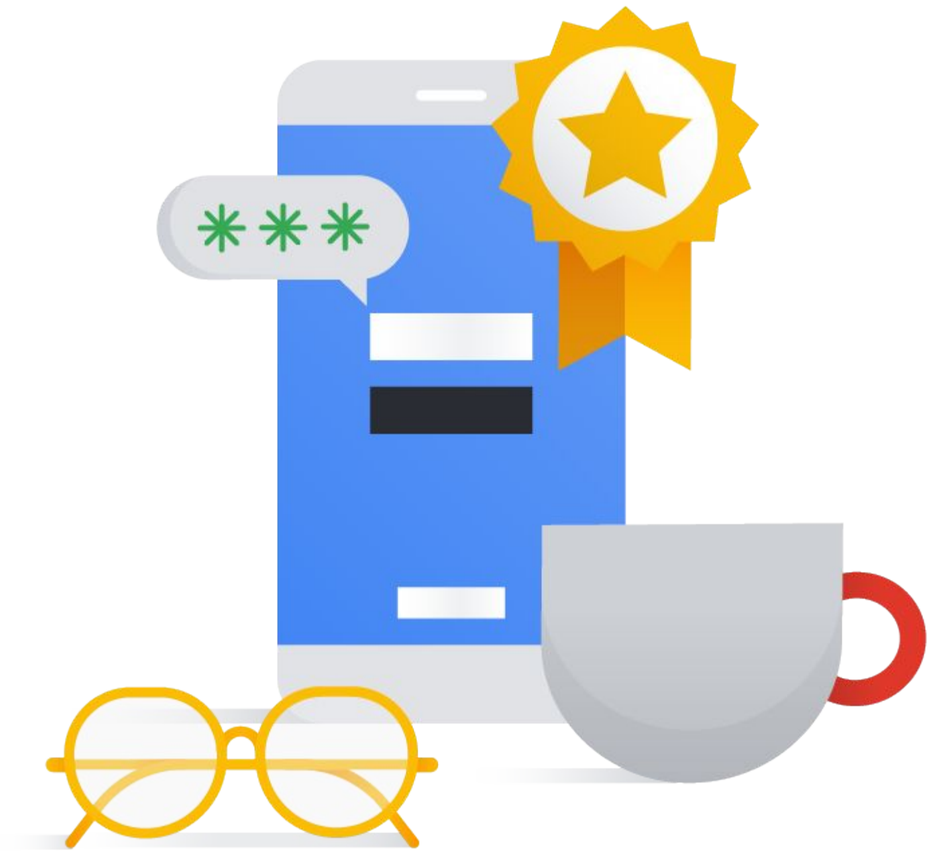
Module 03
Is complete

Google Cloud

Appendix 01: Lesson 02 quiz questions



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



Quiz | Question 01

Question

A data science team is training a new generative AI model on a massive dataset of images. They need access to powerful hardware and software resources to handle the computationally intensive training process.

Which layer of the GenAI landscape would provide the team with the necessary computational power and storage (Applications, Agents, Platform, Models or Infrastructure)?

Quiz | Question 01

Answer

A data science team is training a new generative AI model on a massive dataset of images. They need access to powerful hardware and software resources to handle the computationally intensive training process.

Which layer of the GenAI landscape would provide the team with the necessary computational power and storage (Applications, Agents, Platform, Models or Infrastructure)?

Infrastructure

Quiz | Question 02

Question

A game developer wants to create more realistic and engaging non-player characters (NPCs) in their game. They envision NPCs that can:

- Engage in dynamic conversations with the player
- React to the player's actions and choices
- Adapt their behavior based on the game's environment and storyline
- Exhibit unique personalities and backstories

Which layer of the gen AI landscape would be MOST crucial in defining the behaviors and capabilities of these AI-powered NPCs (Applications, Agents, Platform, Models or Infrastructure)?

Quiz | Question 02

Answer

A game developer wants to create more realistic and engaging non-player characters (NPCs) in their game. They envision NPCs that can:

- Engage in dynamic conversations with the player
- React to the player's actions and choices
- Adapt their behavior based on the game's environment and storyline
- Exhibit unique personalities and backstories

Which layer of the gen AI landscape would be MOST crucial in defining the behaviors and capabilities of these AI-powered NPCs (Applications, Agents, Platform, Models or Infrastructure)?

Agents

Quiz | Question 03

Question

A travel agency wants to use an AI agent to help customers plan their vacations. The agent should be able to:

- Gather customer preferences (budget, destination interests, travel dates)
- Search for flights, accommodations, and activities
- Create personalized itineraries with options and recommendations
- Book flights and hotels based on customer choices
- Provide ongoing support throughout the trip

How would the "agents" layer and the "applications" layer work together to create this AI-powered travel planning experience?

Quiz | Question 03

Question (continued)

How would the "agents" layer and the "applications" layer work together to create this AI-powered travel planning experience?

A. The agents layer would provide the user interface for the travel planning application.

B. The applications layer would determine the specific tasks the AI agent can perform, such as searching for flights or booking hotels.

C. The agents layer would define the AI's capabilities (searching, booking, recommending), while the applications layer would provide the user-facing tool (website or app) to interact with the agent.

D. The applications layer would provide the computational resources for the AI agent to operate.

Quiz | Question 03

Answer

How would the "agents" layer and the "applications" layer work together to create this AI-powered travel planning experience?

A. The agents layer would provide the user interface for the travel planning application.

B. The applications layer would determine the specific tasks the AI agent can perform, such as searching for flights or booking hotels.

C. The agents layer would define the AI's capabilities (searching, booking, recommending), while the applications layer would provide the user-facing tool (website or app) to interact with the agent.

D. The applications layer would provide the computational resources for the AI agent to operate.

Quiz | Question 04

Question

A news organization wants to develop an AI agent that delivers personalized news to each user. The agent should be able to:

- Learn the user's interests and reading habits
- Filter and prioritize news articles based on relevance
- Summarize key information from multiple sources
- Recommend related articles and diverse perspectives
- Adapt to the user's feedback and evolving preferences

How would the "agents" layer contribute to the functionality of this personalized news reader application?

Quiz | Question 04

Question (continued)

How would the "agents" layer contribute to the functionality of this personalized news reader application?

A. The agents layer would define the specific tasks the AI performs, such as filtering articles, summarizing information, and making recommendations.

B. The agents layer would provide the underlying AI models for natural language processing and understanding.

C. The agents layer would design the user interface of the news reader application.

D. The agents layer would provide the infrastructure for storing and processing the news data.

Quiz | Question 04

Answer

How would the "agents" layer contribute to the functionality of this personalized news reader application?

A. The agents layer would define the specific tasks the AI performs, such as filtering articles, summarizing information, and making recommendations.

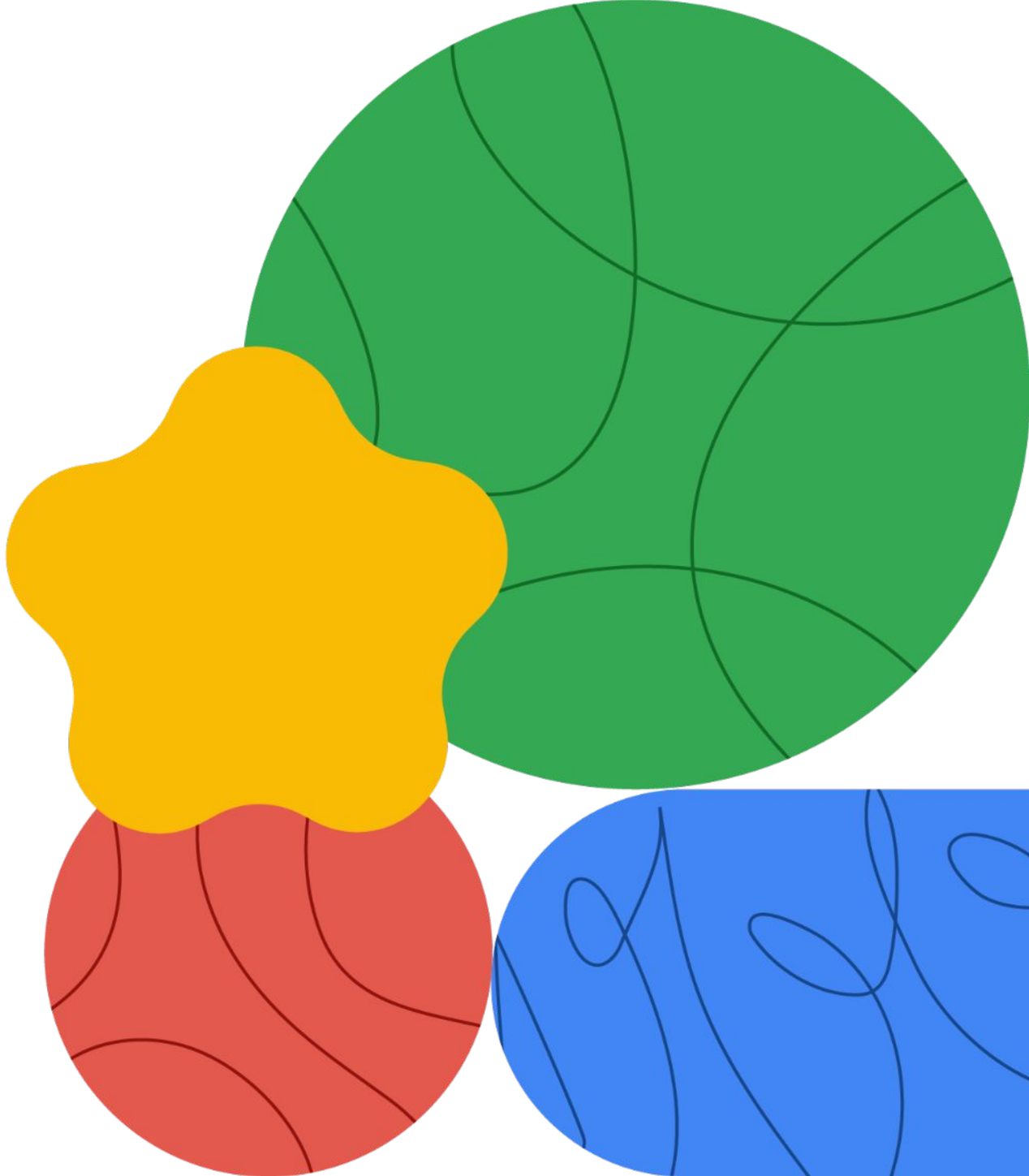
B. The agents layer would provide the underlying AI models for natural language processing and understanding.

C. The agents layer would design the user interface of the news reader application.

D. The agents layer would provide the infrastructure for storing and processing the news data.

Google Cloud

Appendix 02: Additional Lesson 04 Quiz questions



Quiz | Question 01

Question

Which factor can increase the cost of using a model?

A. Limiting the number of users

B. Using black and white displays

C. Smaller context window

D. Larger context window

Quiz | Question 01

Answer

Which factor can increase the cost of using a model?

A. Limiting the number of users

B. Using black and white displays

C. Smaller context window

D. Larger context window