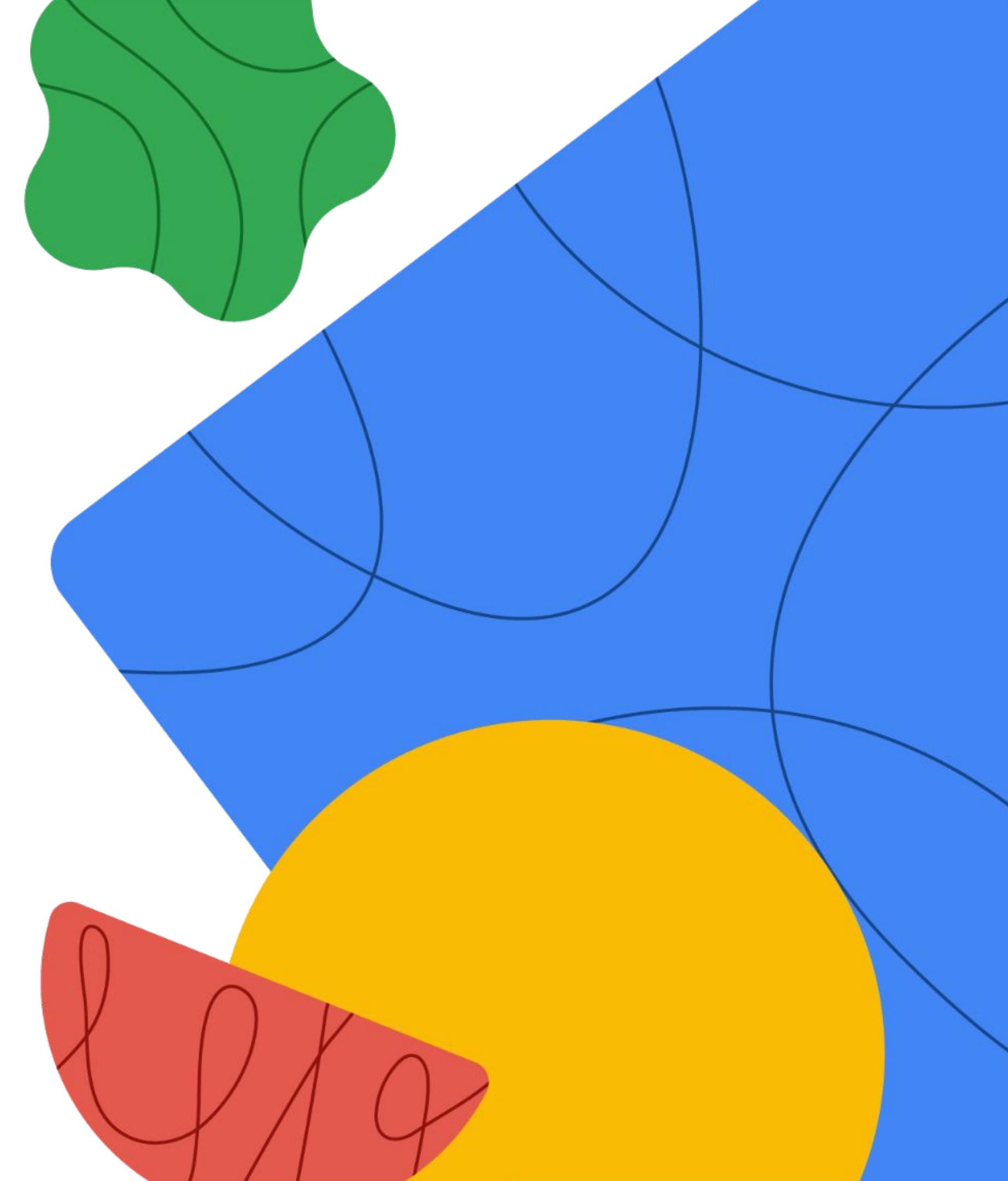
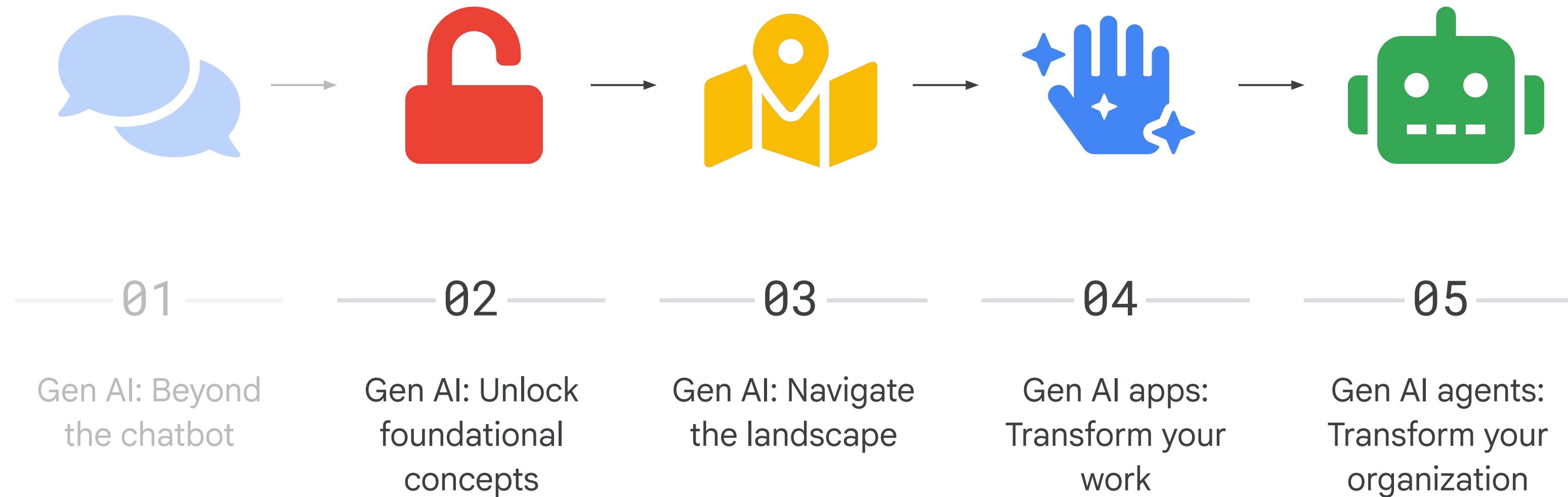


Module 02

Gen AI: Unlock foundational concepts



Generative AI Leader learning path



Module objectives

01

Define core gen AI concepts.

02

Explain how data types are used in gen AI for business impact.

03

Explain the role of foundation models in gen AI.

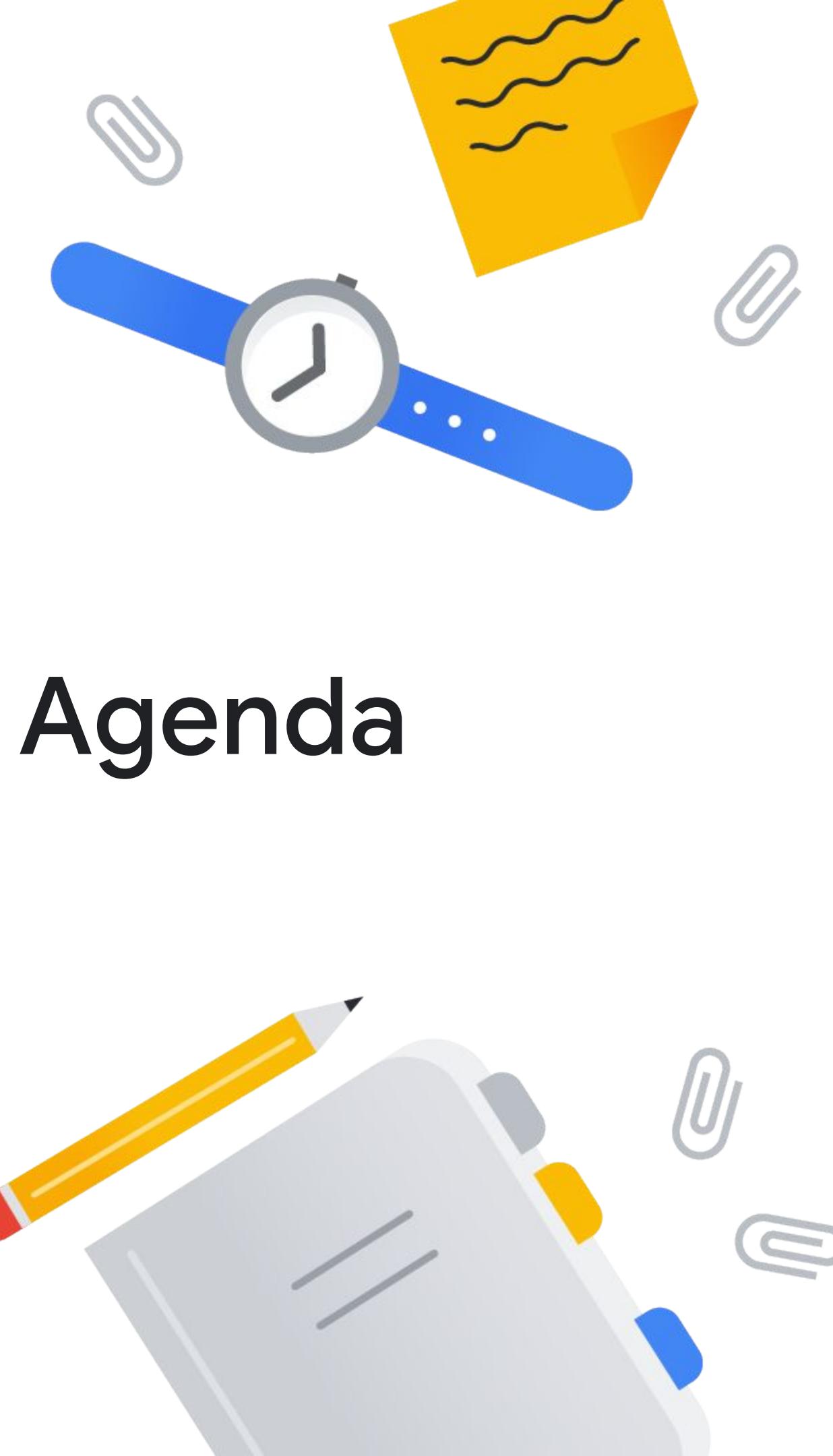
04

Describe Google Cloud's strategies for handling LLM limitations.

05

Describe the challenges for responsible and secure AI development and deployment.



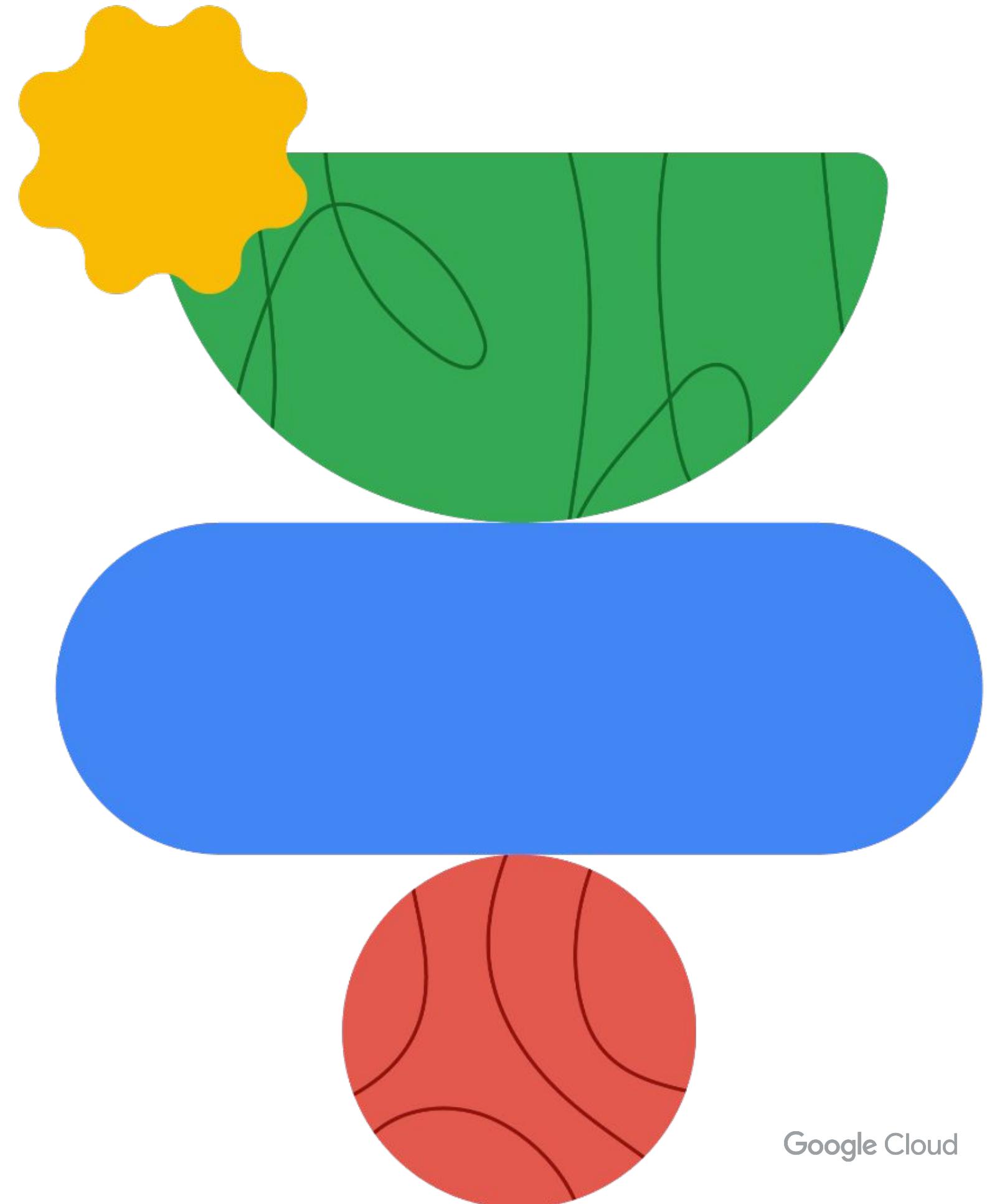


Agenda

- 01 Core gen AI concepts
- 02 Foundation models
- 03 Building AI securely and responsibly

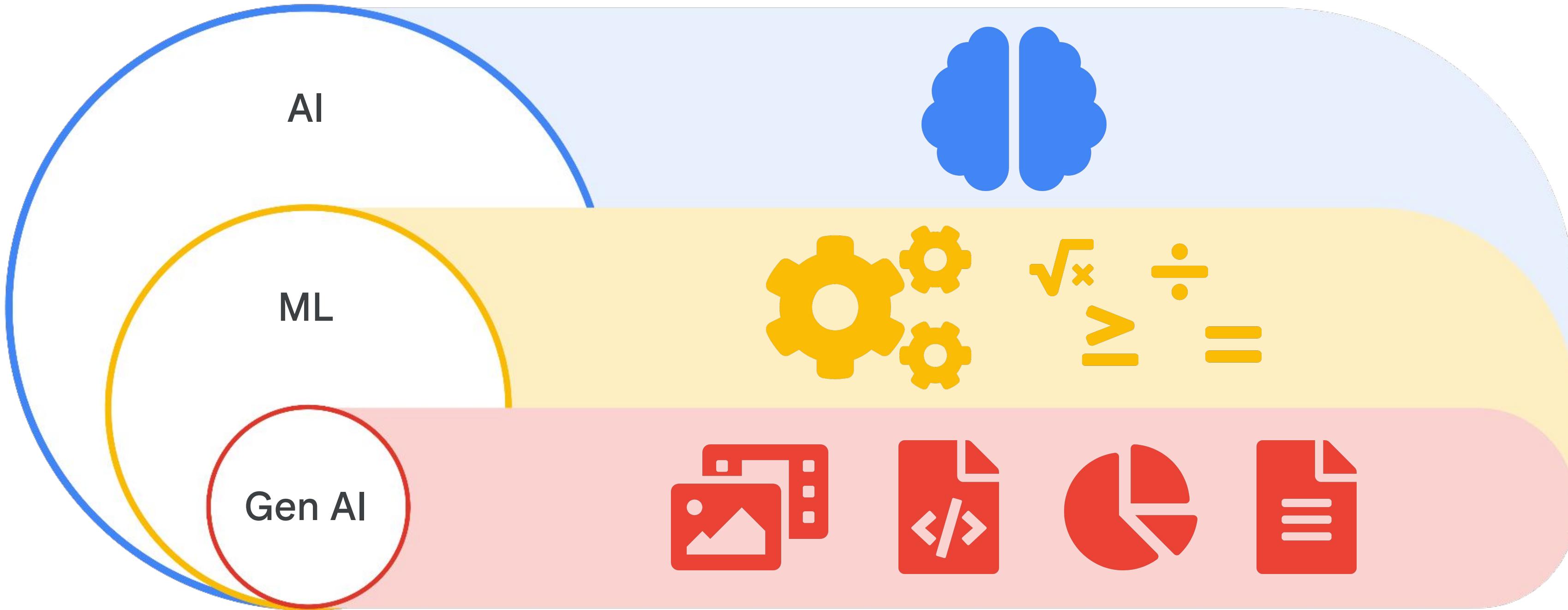
Core gen AI concepts

Understanding AI, ML, and gen AI

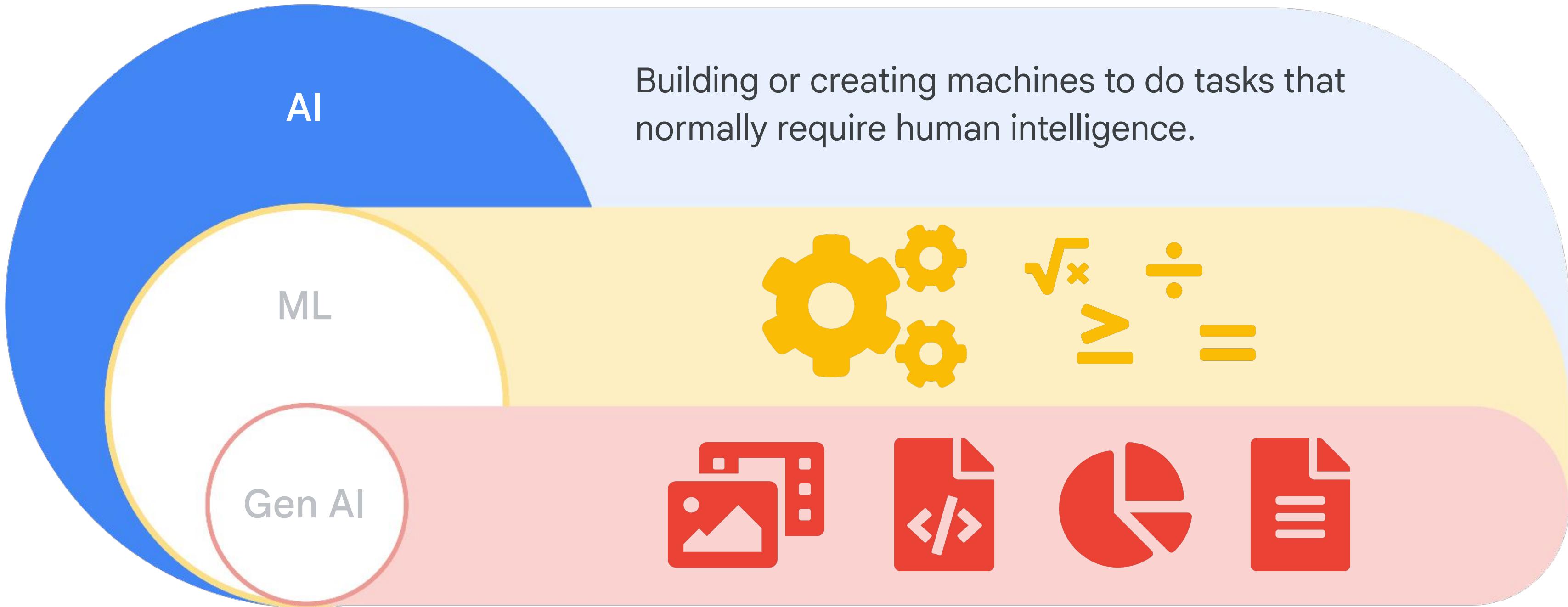


Google Cloud

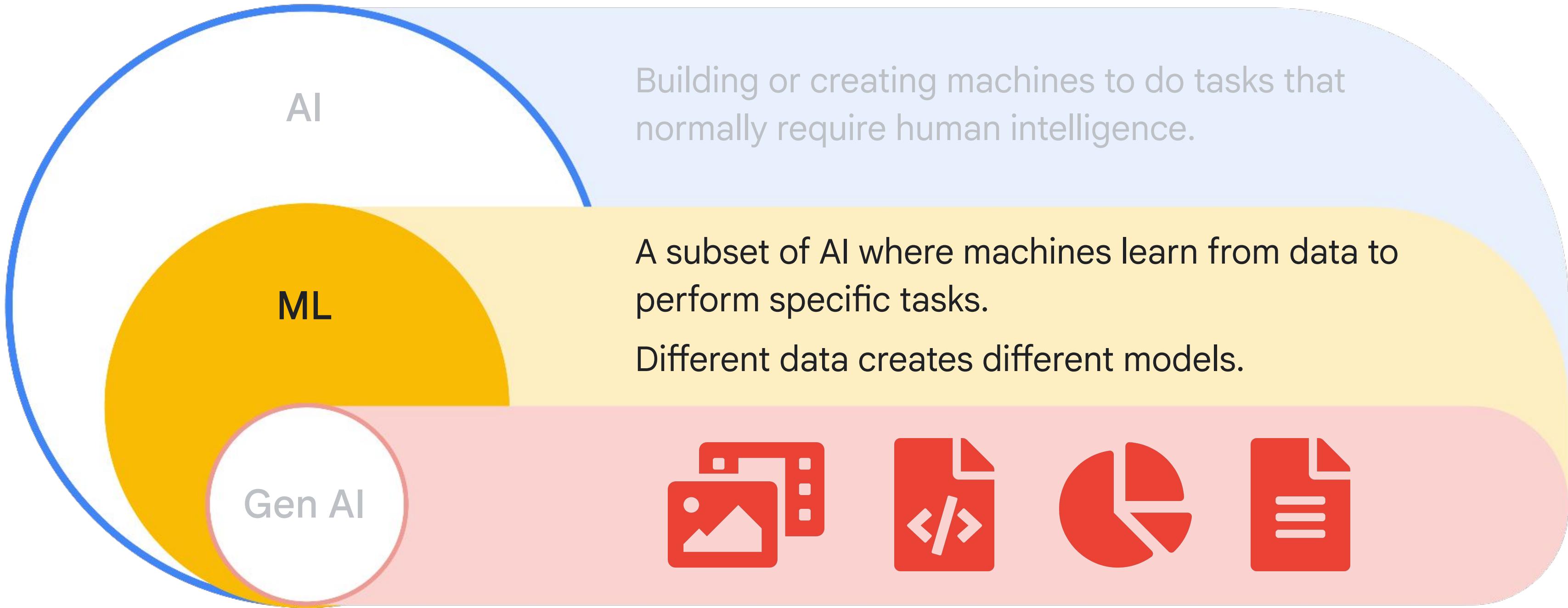
Defining AI, ML and gen AI



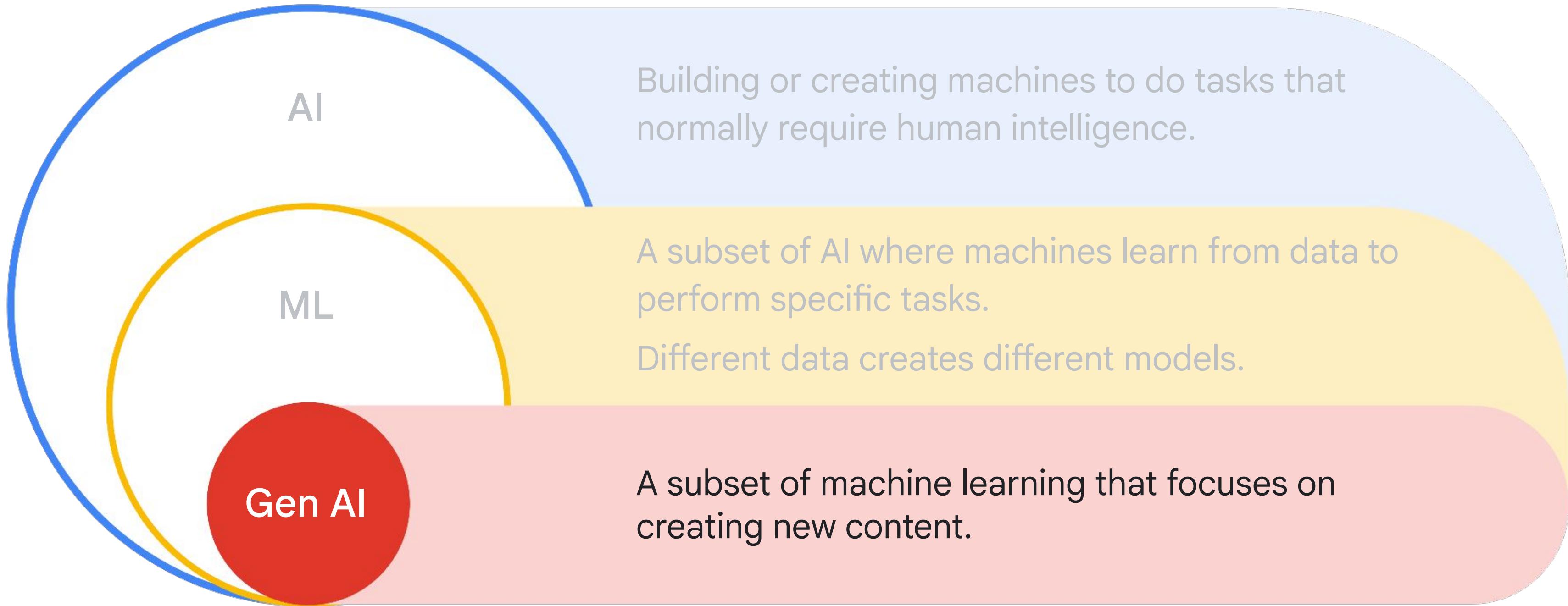
Defining AI, ML and gen AI



Defining AI, ML and gen AI



Defining AI, ML and gen AI

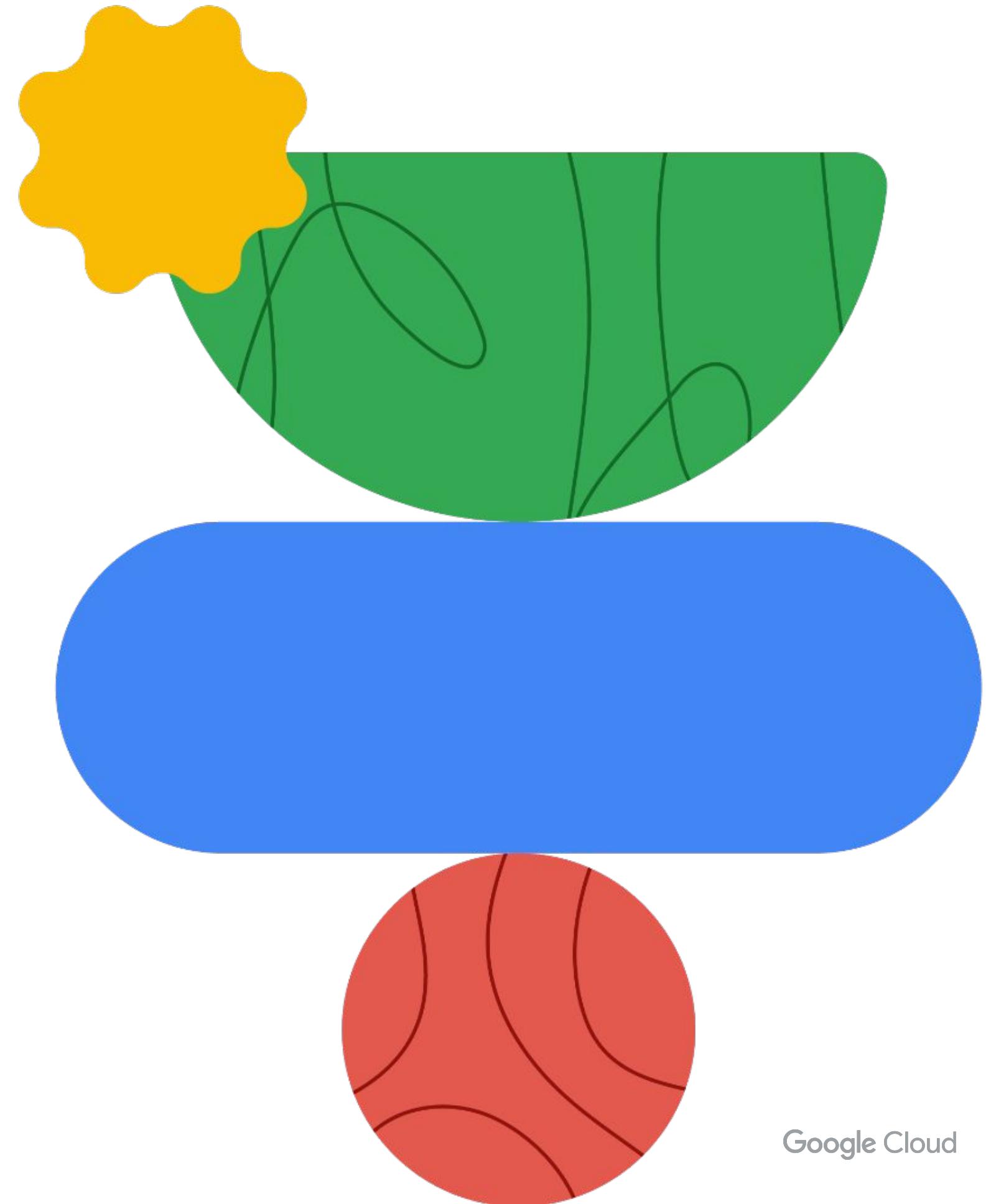


Traditional AI versus generative AI

Imagine you're teaching a child to paint.



What is data?



Google Cloud

The importance of data in ML

- We try to get machines to recognize patterns and make predictions based on data.
- Data is information, and it can come in many forms.
- A machine learning model needs the right kind of data to learn effectively.



ML models **analyze the data they've been given**, identify patterns, and then calculate the likelihood of different outcomes when presented with new information. So, **data quality and accessibility are crucial** for accurate predictions.

Data quality

-  Accuracy
-  Completeness
-  Representative
-  Consistency
-  Relevance

Data quality



Accuracy



Completeness



Representative



Consistency



Relevance

If the data is inaccurate, the model will learn incorrect patterns and make faulty predictions.

Data quality

-  Accuracy
-  Completeness
 - Completeness refers to the size of a dataset and representation within the dataset.
-  Representative
 - The model needs enough to make an accurate prediction.
-  Consistency
-  Relevance

Data quality



Accuracy



Completeness



Representative



Consistency



Relevance

Data needs to be representative and inclusive, otherwise it can lead to skewed samples and biased outcomes.

Data quality



Accuracy



Completeness



Representative



Consistency



Relevance

Inconsistent data formats or labeling can confuse the model and hinder its ability to learn effectively.

Data quality



Accuracy



Completeness



Representative



Consistency



Relevance

Data must be relevant to the task the AI is designed to perform.

Data accessibility

Availability

Cost

Format

Data accessibility

Availability

If the necessary data simply isn't available, the AI model can't be trained.

Cost

Format

Data accessibility

Availability

If the necessary data simply isn't available, the AI model can't be trained.

Cost

The cost of acquiring high-quality data can be a significant barrier to AI development.

Format

Data accessibility

Availability

If the necessary data simply isn't available, the AI model can't be trained.

Cost

The cost of acquiring high-quality data can be a significant barrier to AI development.

Format

Data needs to be in a format that the AI model can understand and process.

Understanding your company's data, its **quality**, **availability**, and **form** is essential for understanding the **realm of what will be possible** in terms of using that data for AI.

Data types

01

Structured data

02

Unstructured data

Cymbal Cleaning

Data stored for customer orders includes:

- Customer ID
- Customer name
- Delivery address
- Purchase date
- Order cost
- Product image
- Feedback (on a 1-5 star scale)
- Feedback (free-form text)
- Email content

Data types

01

Structured data

02

Unstructured data

Organized; easy to search for and find the information you need

Which of these are structured data?

- Customer ID
- Customer name
- Delivery address
- Purchase date
- Order cost
- Product image
- Feedback (on a 1-5 star scale)
- Feedback (free-form text)
- Email content

Data types

01

Structured data

02

Unstructured data

Organized; easy to search for and find the information you need

Cymbal Cleaning's **structured data** includes:

- Customer ID
- Customer name
- Delivery address
- Purchase date
- Order cost
- Product image
- Feedback (on a 1-5 star scale)
- Feedback (free-form text)
- Email content

Data types

01

Structured data

02

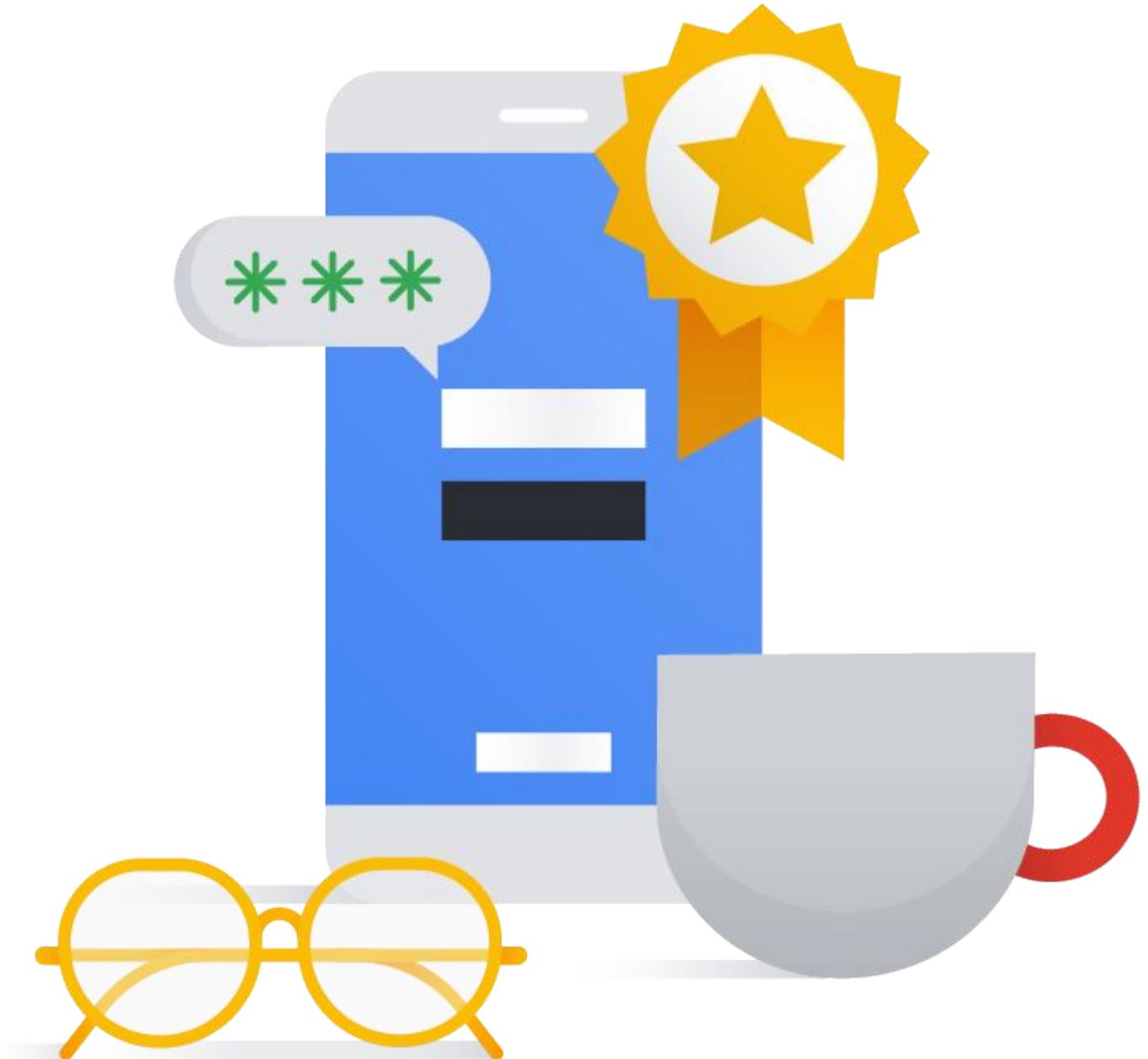
Unstructured data

Lacks a predefined structure; messy and complex, can't be easily organized

Cymbal Cleaning's **unstructured data** includes:

- Customer ID
- Customer name
- **Delivery address**
- Purchase date
- Order cost
- **Product image**
- Feedback (on a 1-5 star scale)
- **Feedback (free-form text)**
- **Email content**

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



Quiz | Question 01

Question

An AI model is trained on data containing numerous factual errors (e.g., incorrect historical dates, mislabeled customer details). In this scenario, which data quality factor is primarily compromised, leading to potentially flawed predictions?

- A. Completeness
- B. Consistency
- C. Accuracy
- D. Relevance

Quiz | Question 01

Answer

An AI model is trained on data containing numerous factual errors (e.g., incorrect historical dates, mislabeled customer details). In this scenario, which data quality factor is primarily compromised, leading to potentially flawed predictions?

- A. Completeness
- B. Consistency
- C. Accuracy
- D. Relevance



Quiz | Question 02

Question

Relating to data accessibility, what is the key nuance presented regarding large datasets?

- A. Larger datasets always guarantee better model performance.
- B. Data volume is irrelevant for complex generative AI models.
- C. Acquiring large datasets is always less expensive than acquiring smaller, specialized ones.
- D. While often beneficial, large volume isn't the only data aspect influencing performance.

Quiz | Question 02

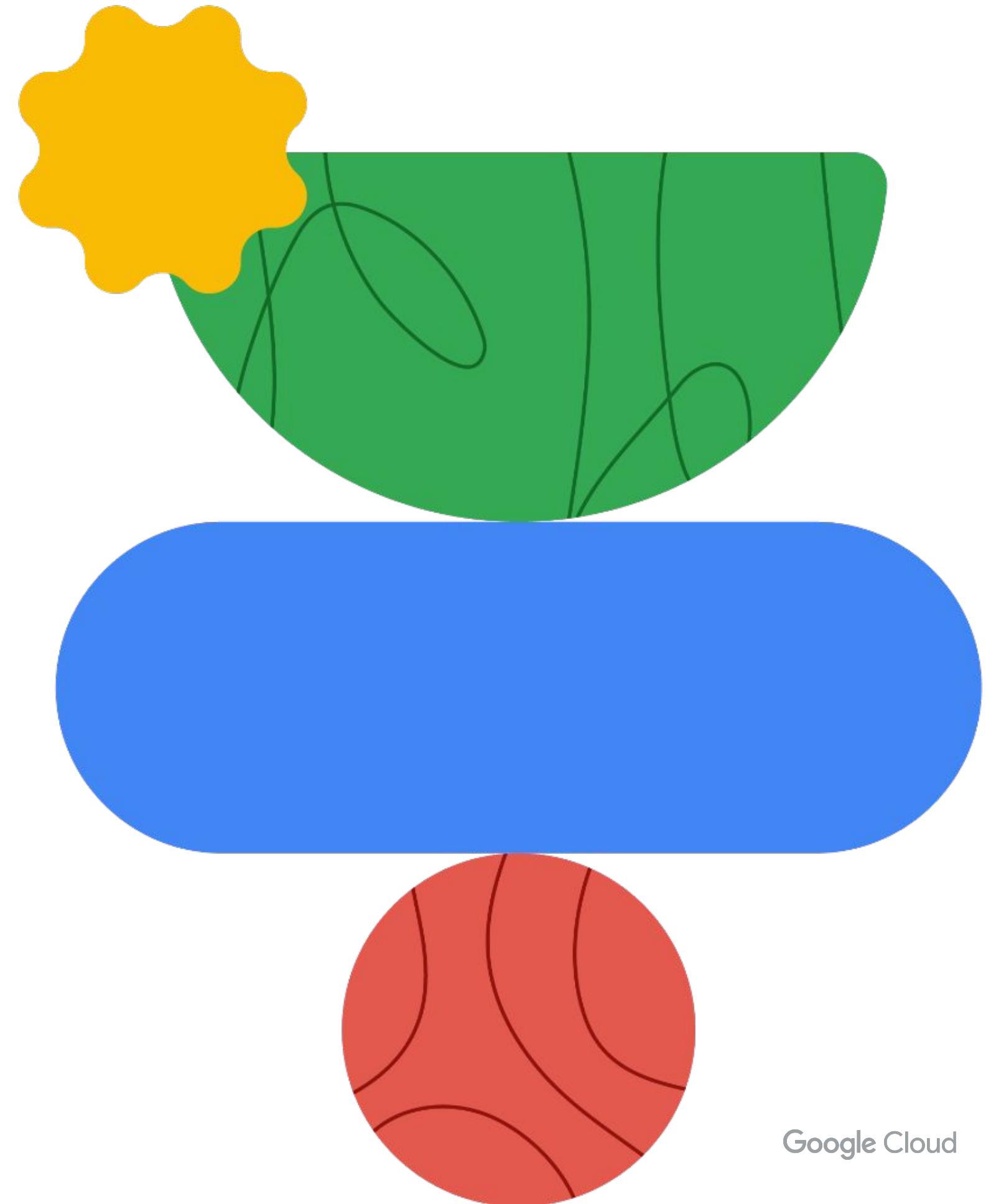
Answer

Relating to data accessibility, what is the key nuance presented regarding large datasets?

- A. Larger datasets always guarantee better model performance.
- B. Data volume is irrelevant for complex generative AI models.
- C. Acquiring large datasets is always less expensive than acquiring smaller, specialized ones.
- D. While often beneficial, large volume isn't the only data aspect influencing performance.

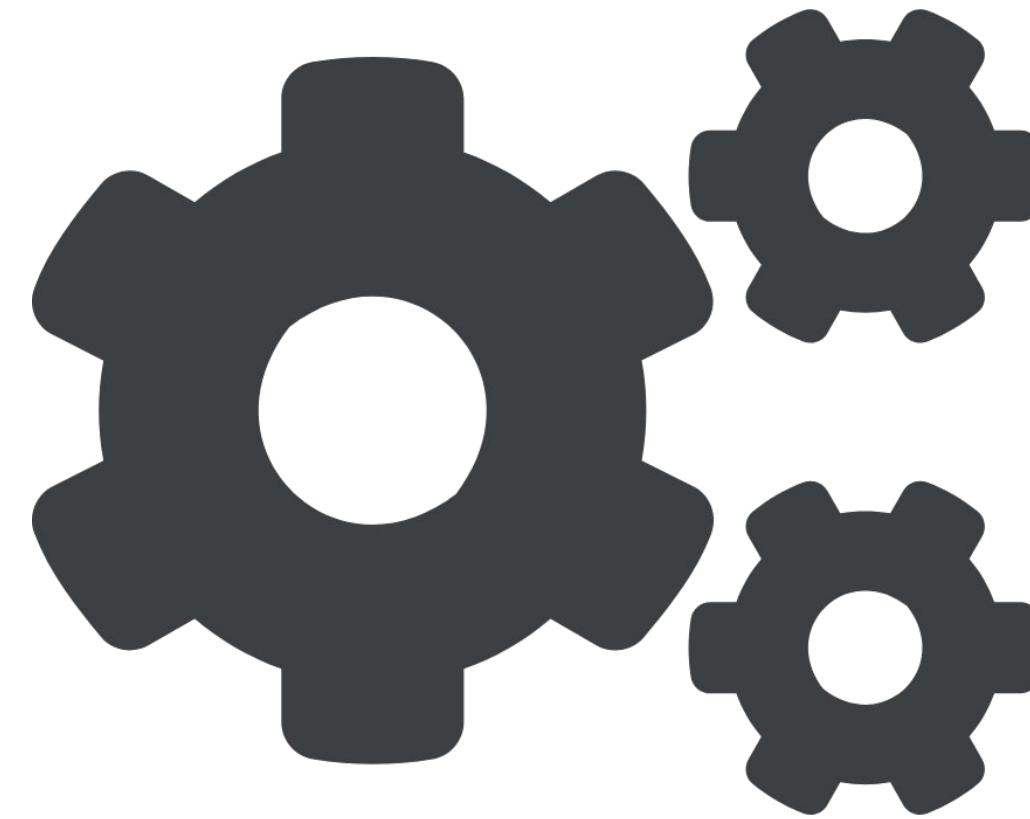


Types of learning

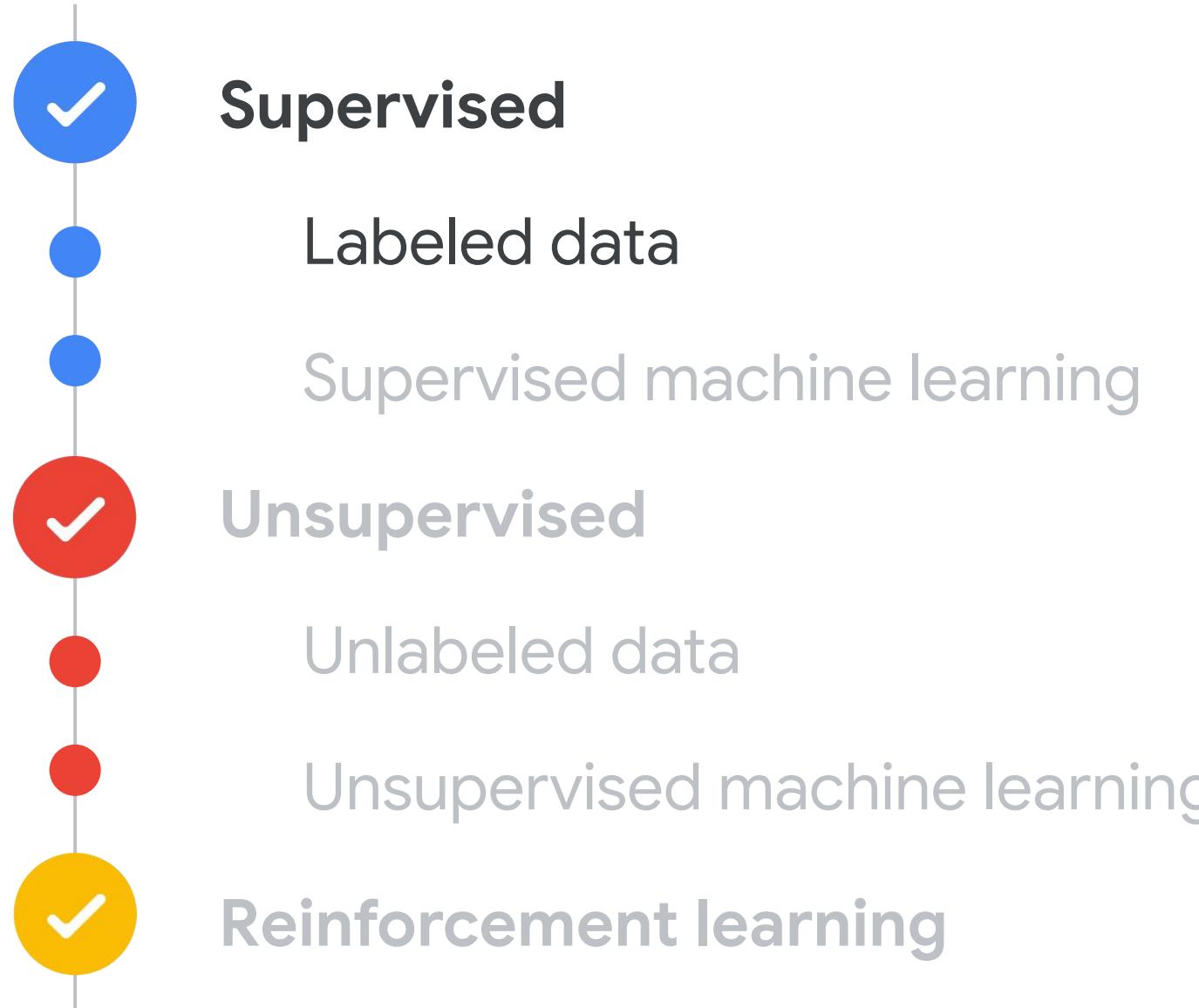


Machine learning approaches and data requirements

-  **Supervised**
 - Labeled data
 - Supervised machine learning
-  **Unsupervised**
 - Unlabeled data
 - Unsupervised machine learning
-  **Reinforcement learning**

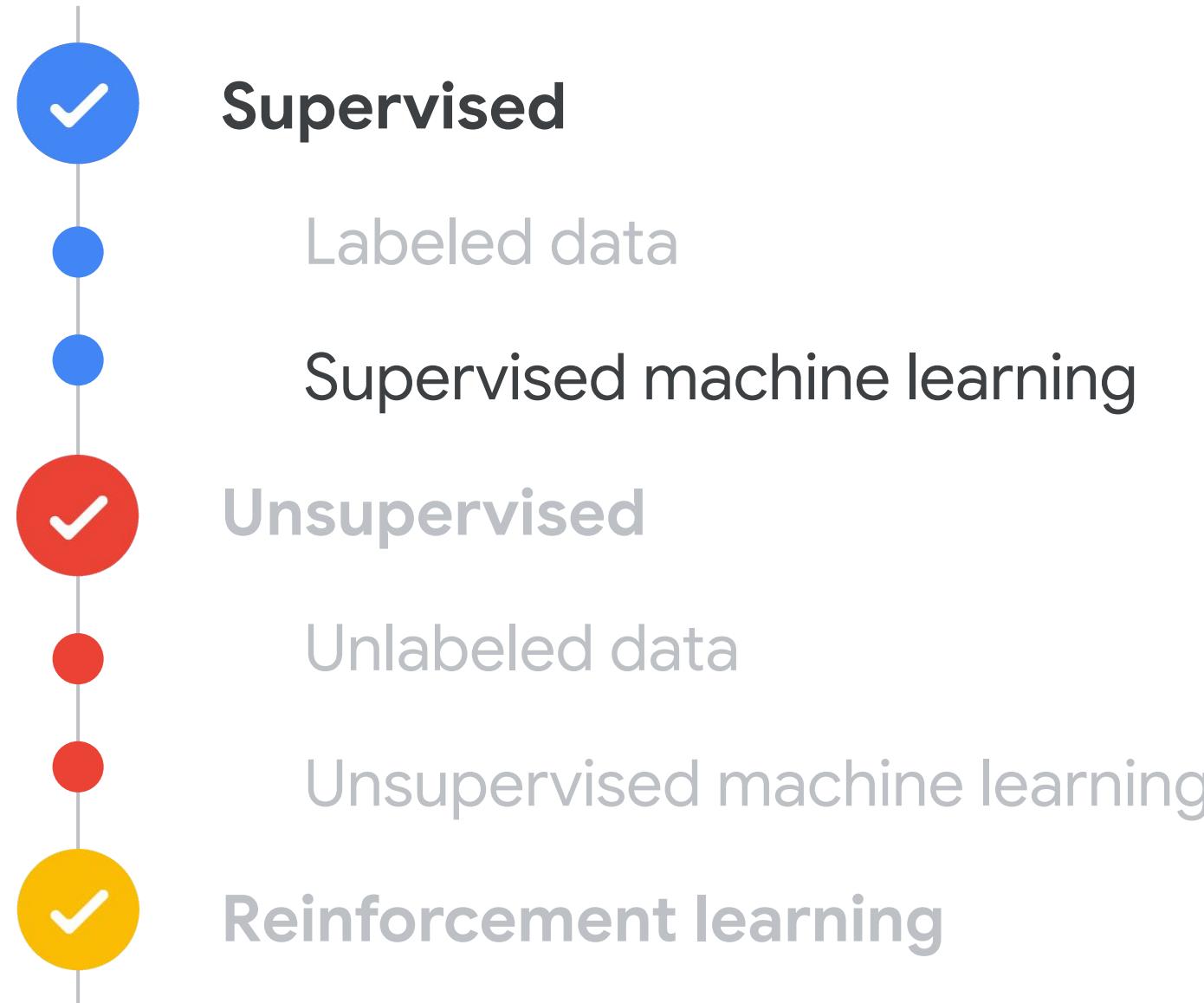


Machine learning approaches and data requirements



- Labeled data has tags e.g. name, type, number, which assign meaning to the data.
- Labels enable algorithms to learn relationships and make accurate predictions.

Machine learning approaches and data requirements

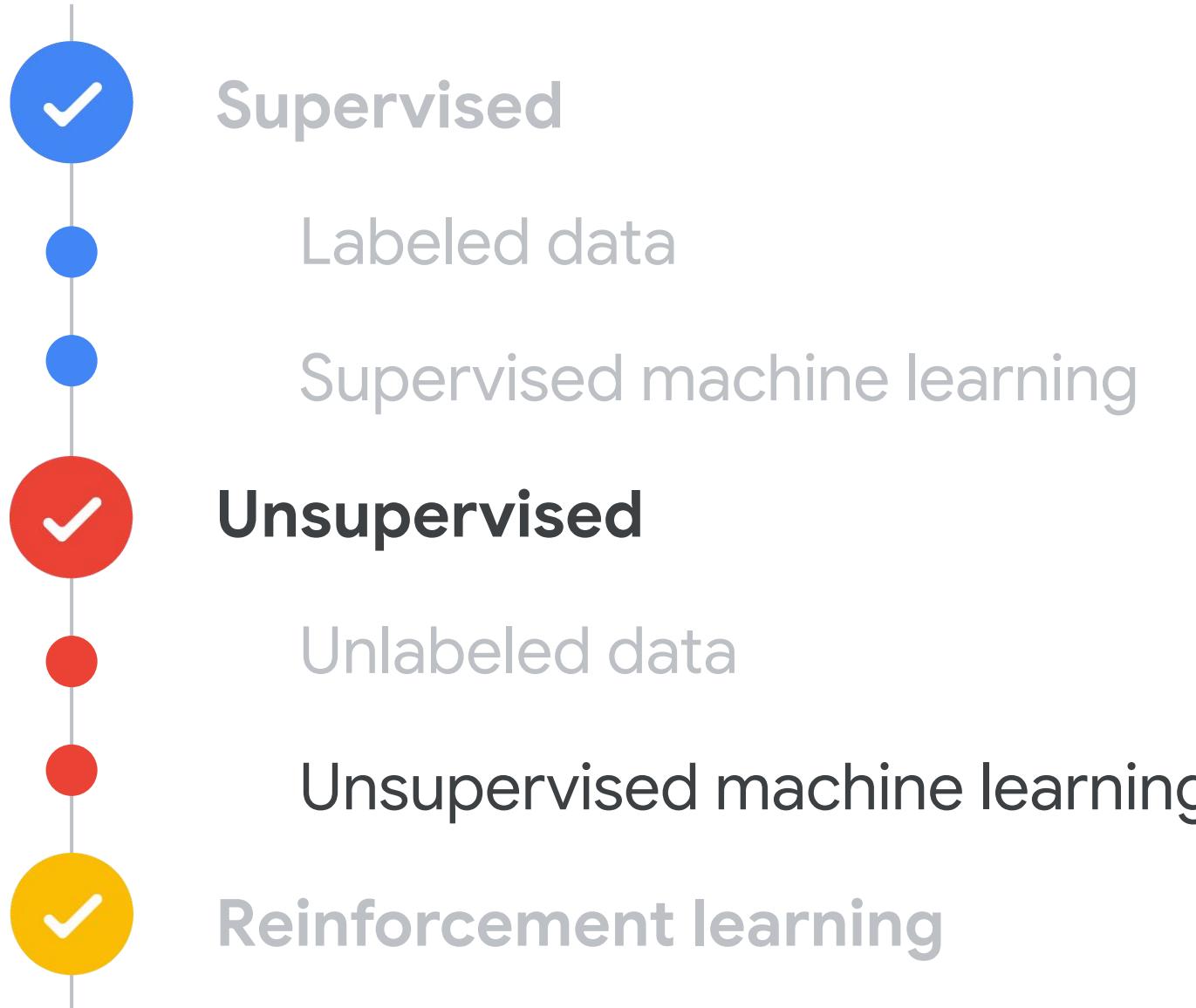


- The model's goal is to identify patterns and relationships within the labeled data, enabling it to accurately predict outputs for new, unseen inputs.

Machine learning approaches and data requirements

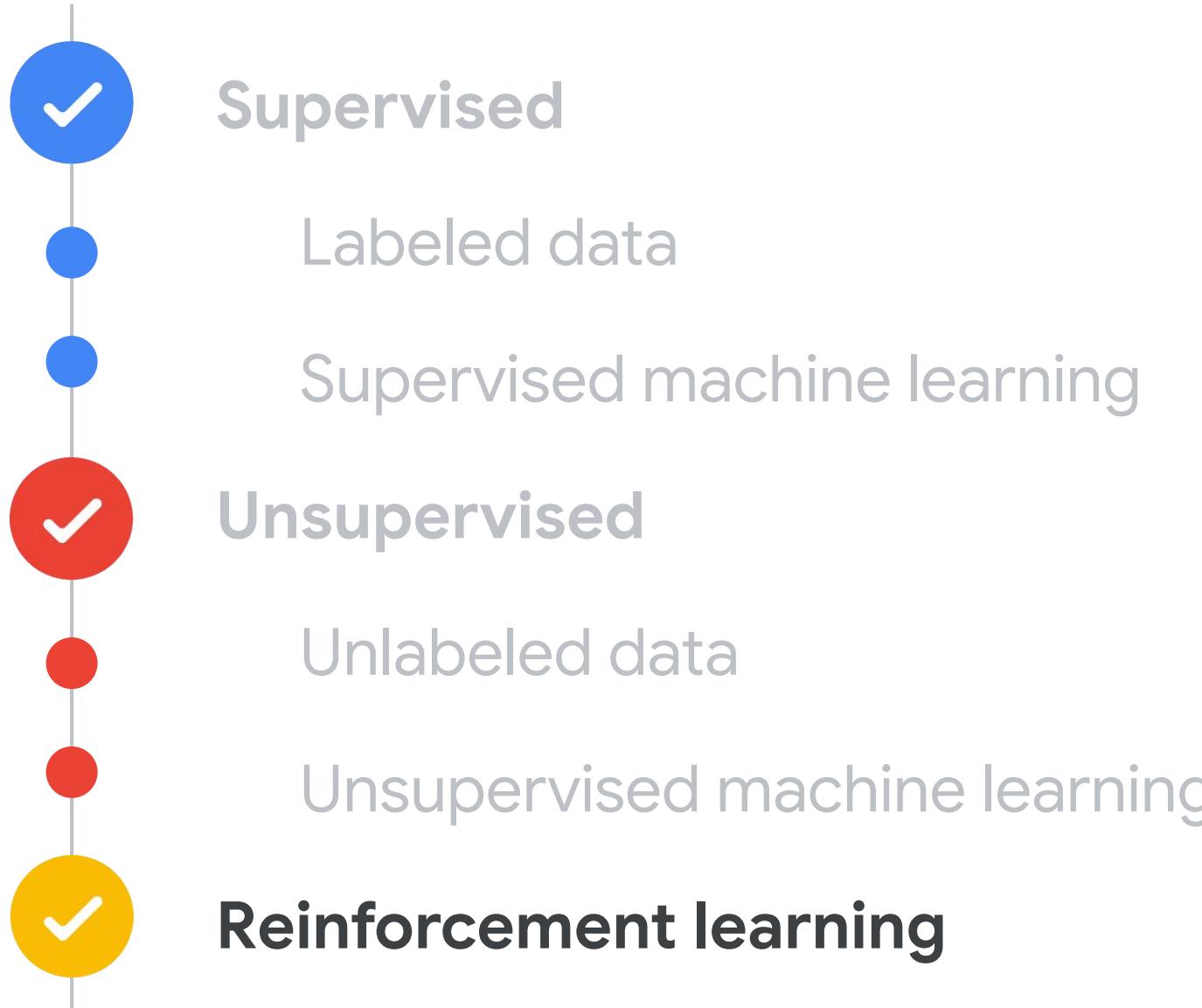
-  **Supervised**
 -  Labeled data
 -  Supervised machine learning
-  **Unsupervised**
 -  Unlabeled data
 -  Unsupervised machine learning
-  **Reinforcement learning**

Machine learning approaches and data requirements



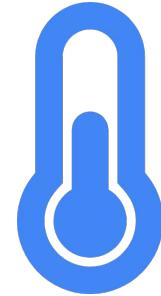
- The model deals with raw, unlabeled data to find natural groupings.
- Exploratory analysis that helps you understand the underlying structure of your data and uncover insights.

Machine learning approaches and data requirements



- The model learns through interaction and feedback.
- The algorithm learns which actions lead to the best outcomes.
- It is useful in situations where you can't provide explicit instructions or labeled data.

Machine learning approaches on Google Cloud: Examples



Predictive maintenance with
Vertex AI (Supervised learning)



Anomaly Detection with
BigQuery ML (Unsupervised
learning)



Product recommendations with
Vertex AI (Reinforcement learning)

Machine learning approaches on Google Cloud: Examples



Predictive maintenance with Vertex AI (supervised learning)



Anomaly detection with BigQuery ML (unsupervised learning)



Product recommendations with Vertex AI (reinforcement learning)

How it works

By training a model on sensor data from machines, Vertex AI can predict when a machine is likely to fail, enabling proactive maintenance and reducing downtime.

Machine learning approaches on Google Cloud: Examples



Predictive maintenance with Vertex AI (supervised learning)



Anomaly detection with BigQuery ML (unsupervised learning)

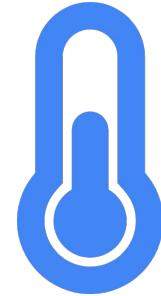


Product recommendations with Vertex AI (reinforcement learning)

How it works

BigQuery ML analyzes historical transaction data to identify patterns and flags unusual transactions that deviate significantly from the norm.

Machine learning approaches on Google Cloud: Examples



Predictive maintenance with Vertex AI (supervised learning)



Anomaly detection with BigQuery ML (unsupervised learning)



Product recommendations with Vertex AI (reinforcement learning)

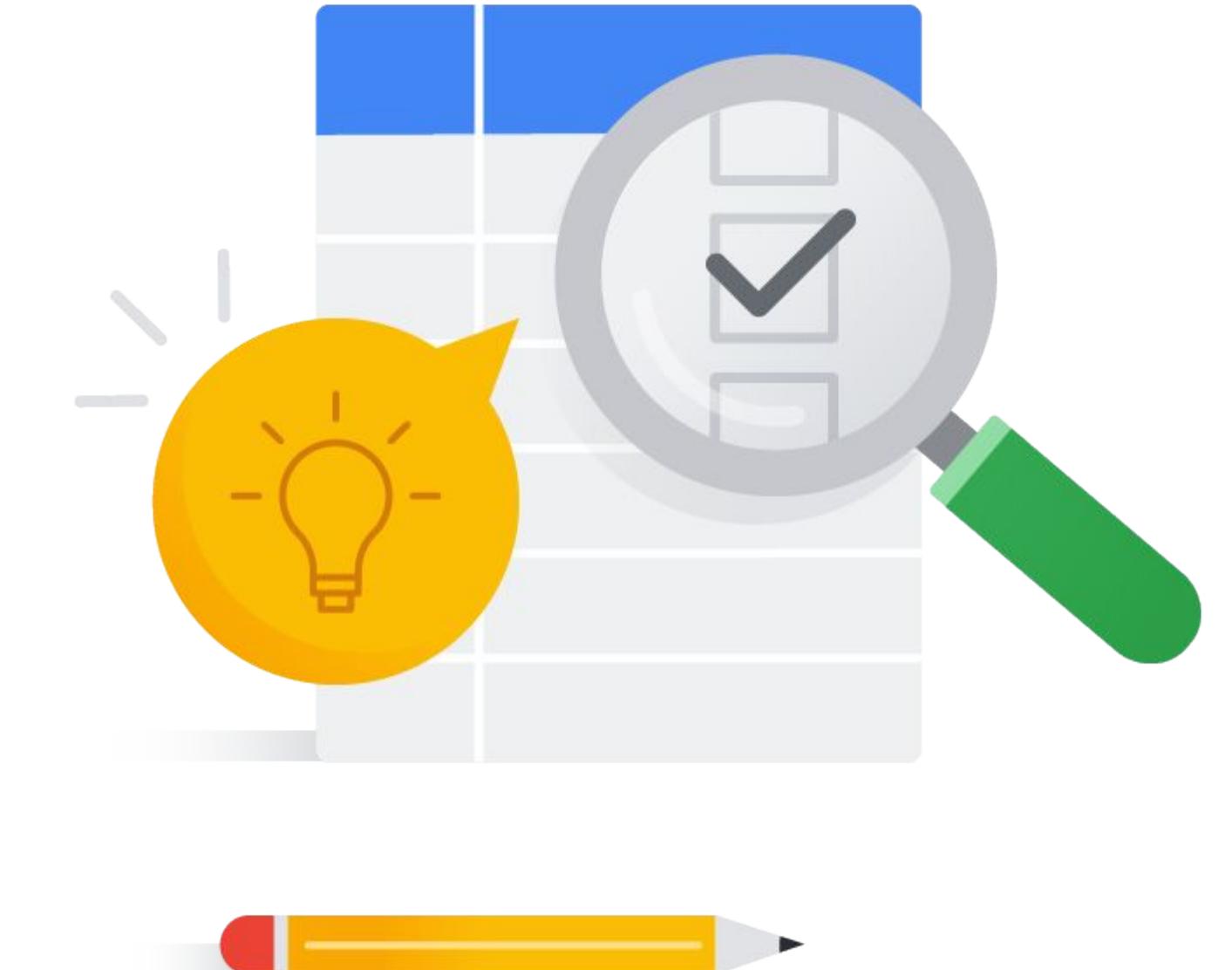
How it works

The model learns to maximize user engagement and sales by continuously refining its recommendations.

Activity: Supervised, unsupervised, or reinforcement learning?

⌚ 5 min

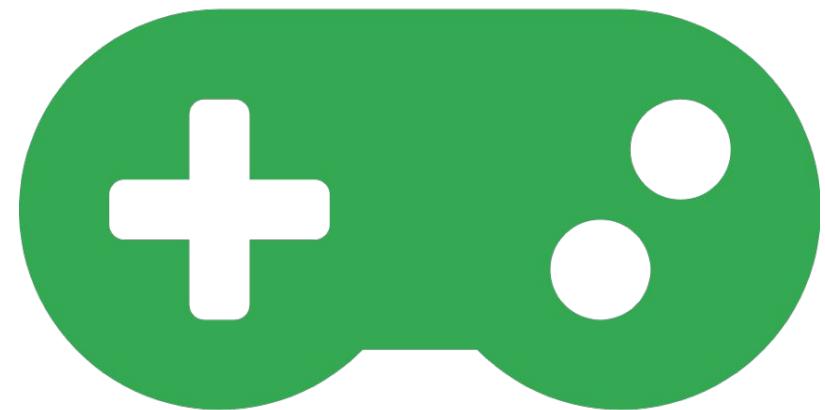
1. Read the scenario.
2. Identify if it is an example of supervised, unsupervised, or reinforcement learning.
Provide a reason for your choice.
3. Put your answer in the chat.



Scenario: Video game

Supervised, unsupervised, or reinforcement learning?

A developer is adding an AI component to a complex video game. The AI interacts with the game environment, taking actions and receiving rewards or penalties based on the outcomes of those actions. Through trial and error, the AI learns which actions lead to the highest cumulative reward, effectively learning to play the game.



Scenario: Video game | Feedback

Reinforcement learning

A developer is adding an AI component to a complex video game. The AI interacts with the game environment, taking actions and receiving rewards or penalties based on the outcomes of those actions. Through trial and error, the AI learns which actions lead to the highest cumulative reward, effectively learning to play the game.

Reason

The AI agent learns through interaction with an environment, receiving feedback in the form of rewards or penalties.

The goal is to maximize the cumulative reward over time, which is the core principle of reinforcement learning.

Scenario: Spam or not spam

Supervised, unsupervised, or reinforcement learning?

An email provider uses this learning technique to classify incoming emails as either spam or not spam. The model is trained on a dataset of emails that have been manually labeled as spam or not spam.



Scenario: Spam or not spam | Feedback

Supervised learning

An email provider uses this learning technique to classify incoming emails as either spam or not spam. The model is trained on a dataset of emails that have been manually labeled as spam or not spam.

Reason

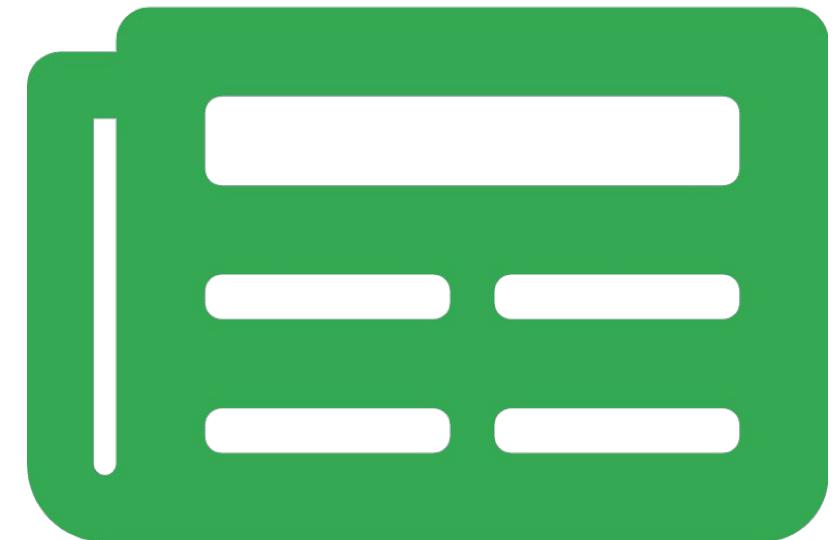
The model is trained on a labeled dataset (emails labeled as spam or not spam), and the goal is to learn a mapping between the input (email content) and the output (spam or not spam classification).

The presence of labeled data is the defining characteristic of supervised learning.

Scenario: Topics

Supervised, unsupervised, or reinforcement learning?

Discovering the underlying topics in a collection of documents. For example, analyzing a set of news articles to identify the main topics being discussed.



Scenario: Topics | Feedback

Unsupervised learning

Discovering the underlying topics in a collection of documents. For example, analyzing a set of news articles to identify the main topics being discussed.

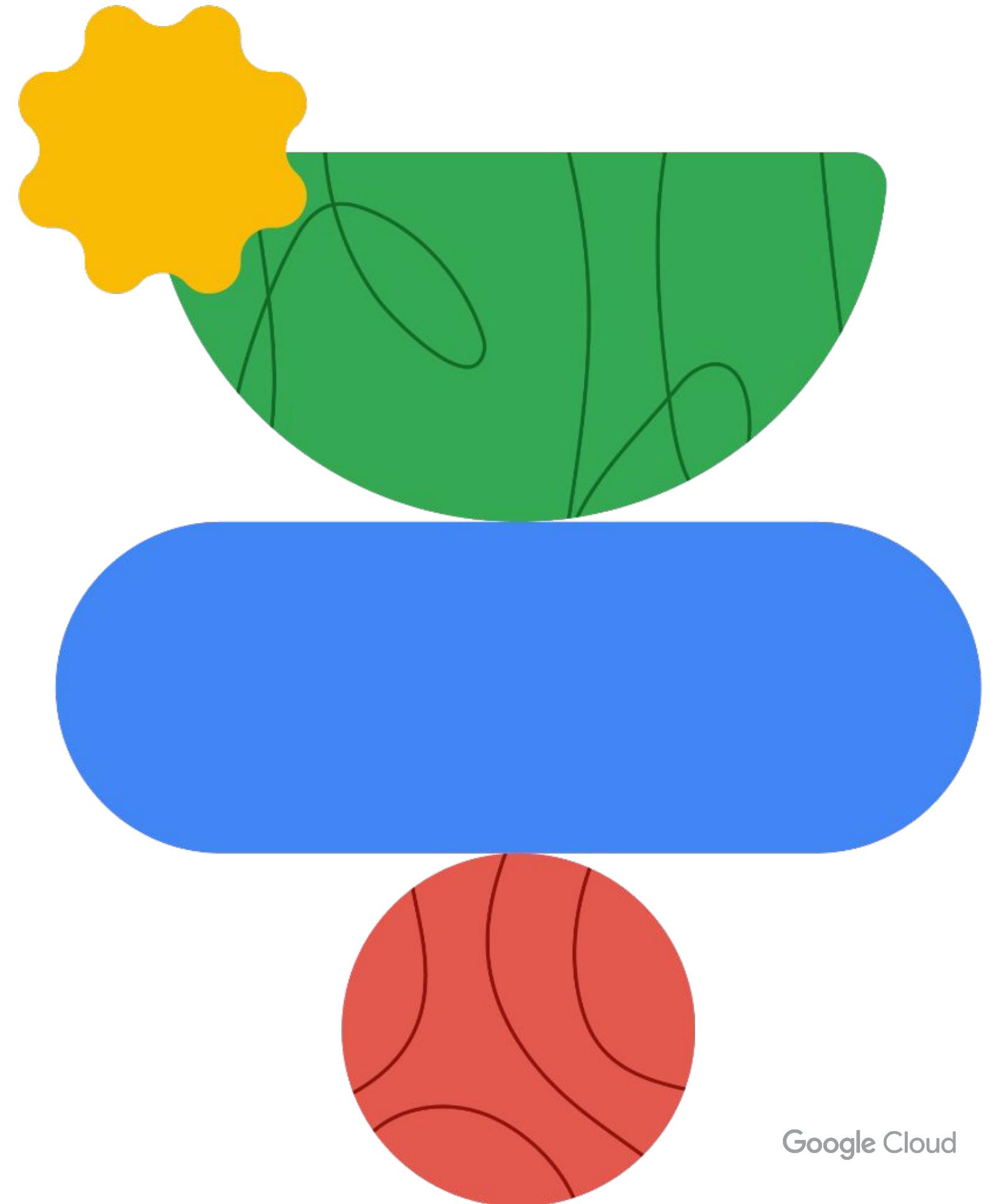
Reason

The task is to find patterns or structures in unlabeled data (the news articles).

Topic modeling, which aims to discover underlying topics, is a common application of unsupervised learning.

Since there are no predefined labels indicating the topics, the model must discover them on its own.

Turning data into learning using Google Cloud



Google Cloud

Key stages of making data accessible for AI

- 1 Gather your data
- 2 Prepare your data
- 3 Train your model
- 4 Deploy and predict
- 5 Manage your model



Key stages of making data accessible for AI

- 1** Gather your data
Determine the data you need based on the outcome you want to achieve.
- 2** Prepare your data
Google Cloud tools:
 - Pub/Sub (real-time streaming)
 - Cloud Storage (unstructured)
 - Cloud SQL and Cloud Spanner (structured)
- 3** Train your model
- 4** Deploy and predict
- 5** Manage your model

Key stages of making data accessible for AI

- 1** Gather your data
The process of cleaning and transforming raw data into a usable format, including proper formatting and labeling.
- 2** Prepare your data
Google Cloud tools:
 - BigQuery (data analysis)
 - BigQuery universal catalog (data governance)
- 3** Train your model
- 4** Deploy and predict
- 5** Manage your model

Key stages of making data accessible for AI

- 1** Gather your data

The process of creating your ML model using data is called **model training**.
- 2** Prepare your data

Google tools:

 - Vertex AI platform (managed environment for training ML models)
- 3** Train your model
- 4** Deploy and predict
- 5** Manage your model

Key stages of making data accessible for AI

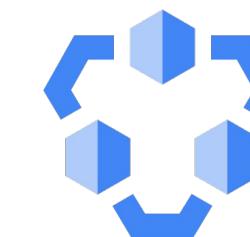
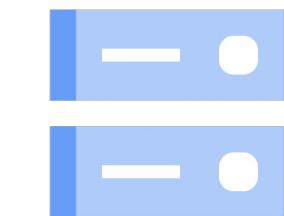
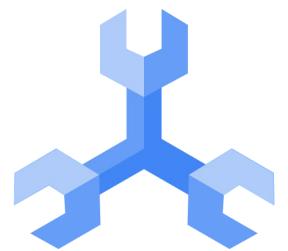
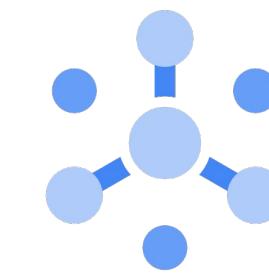
- 1** Gather your data
The process of making a trained model available for use.
- 2** Prepare your data
Google tools:
 - Vertex AI
- 3** Train your model
- 4** Deploy and predict
- 5** Manage your model

Key stages of making data accessible for AI

- 1** Gather your data
The process of managing and maintaining your models over time.
- 2** Prepare your data
Google tools:
 - Versioning
 - Performance tracking
 - Drift monitoring
 - Data management (Vertex AI Feature Store)
 - Storage (Vertex AI Model Garden)
 - Automate (Vertex AI Pipelines)
- 3** Train your model
- 4** Deploy and predict
- 5** Manage your model

Key stages of making data accessible for AI: Google tools

- 1 Gather your data
- 2 Prepare your data
- 3 Train your model
- 4 Manage your model
- 5 Deploy and predict

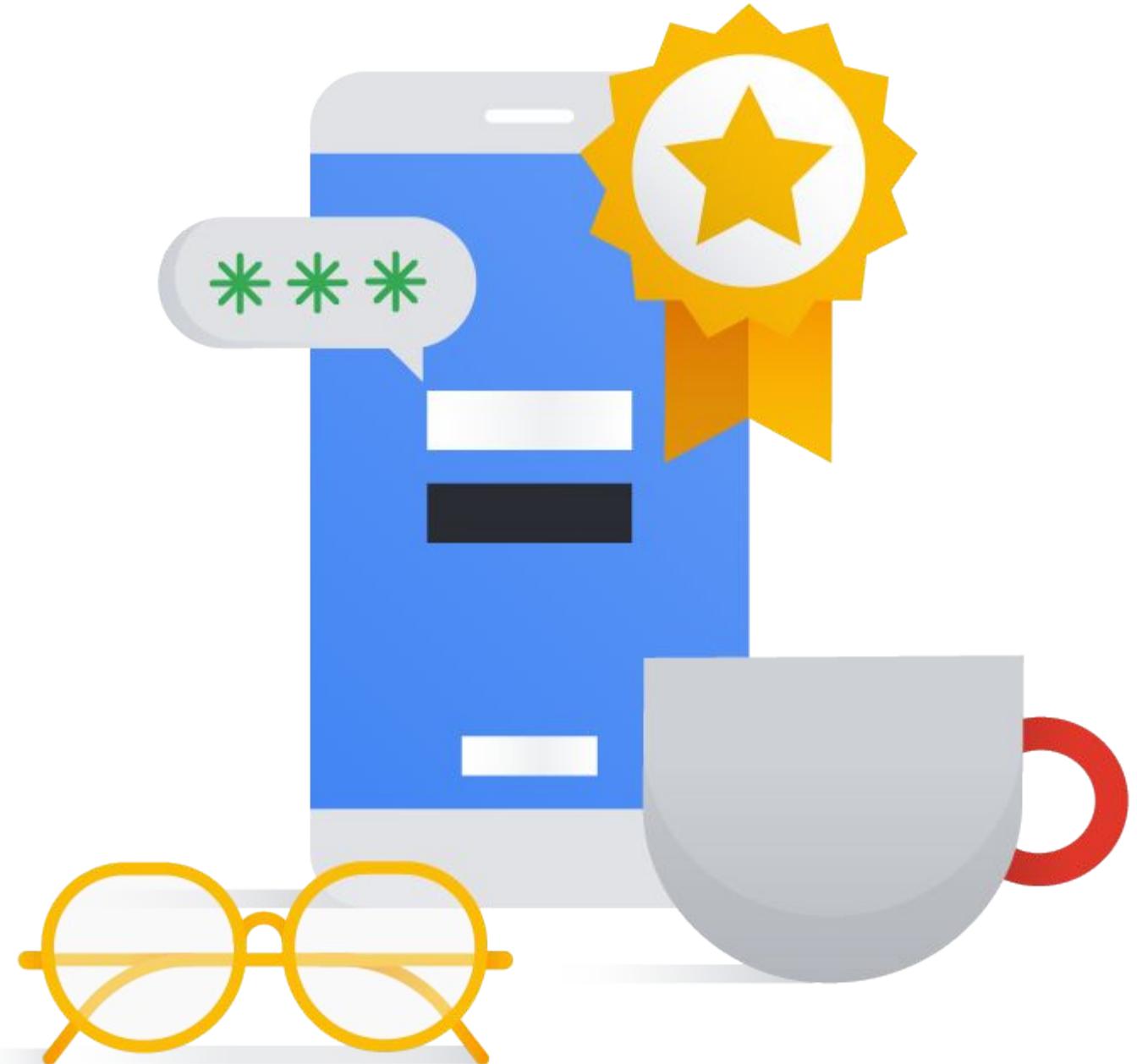


Identity and Access Management (IAM)

- ✓ Create and manage user accounts
- ✓ Assign roles to users
- ✓ Grant and revoke permissions to resources
- ✓ Audit user activity
- ✓ Monitor your security position



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



Quiz | Question 01

Question

How does consistency impact AI model training?

- A. It increases the speed of data retrieval.
- B. Inconsistent formats and labeling can confuse the model and hinder learning.
- C. It reduces the need for data storage.
- D. It ensures data is relevant to the task

Quiz | Question 01

Answer

How does consistency impact AI model training?

- A. It increases the speed of data retrieval.
- B. Inconsistent formats and labeling can confuse the model and hinder learning.
- C. It reduces the need for data storage.
- D. It ensures data is relevant to the task



Quiz | Question 02

Question

What is a model in the context of machine learning?

- A. A type of computer hardware
- B. A set of pre-defined rules for decision-making
- C. A visual representation of data
- D. A complex mathematical structure that processes inputs to generate outputs

Quiz | Question 02

Answer

What is a model in the context of machine learning?

- A. A type of computer hardware
- B. A set of pre-defined rules for decision-making
- C. A visual representation of data
- D. A complex mathematical structure that processes inputs to generate outputs



Quiz | Question 03

Question

What is the primary way that agents learn in reinforcement learning?

- A. By being explicitly programmed with the correct actions
- B. By observing and imitating expert demonstrations
- C. By interacting with their environment and receiving feedback
- D. By analyzing large datasets of labeled examples

Quiz | Question 03

Question

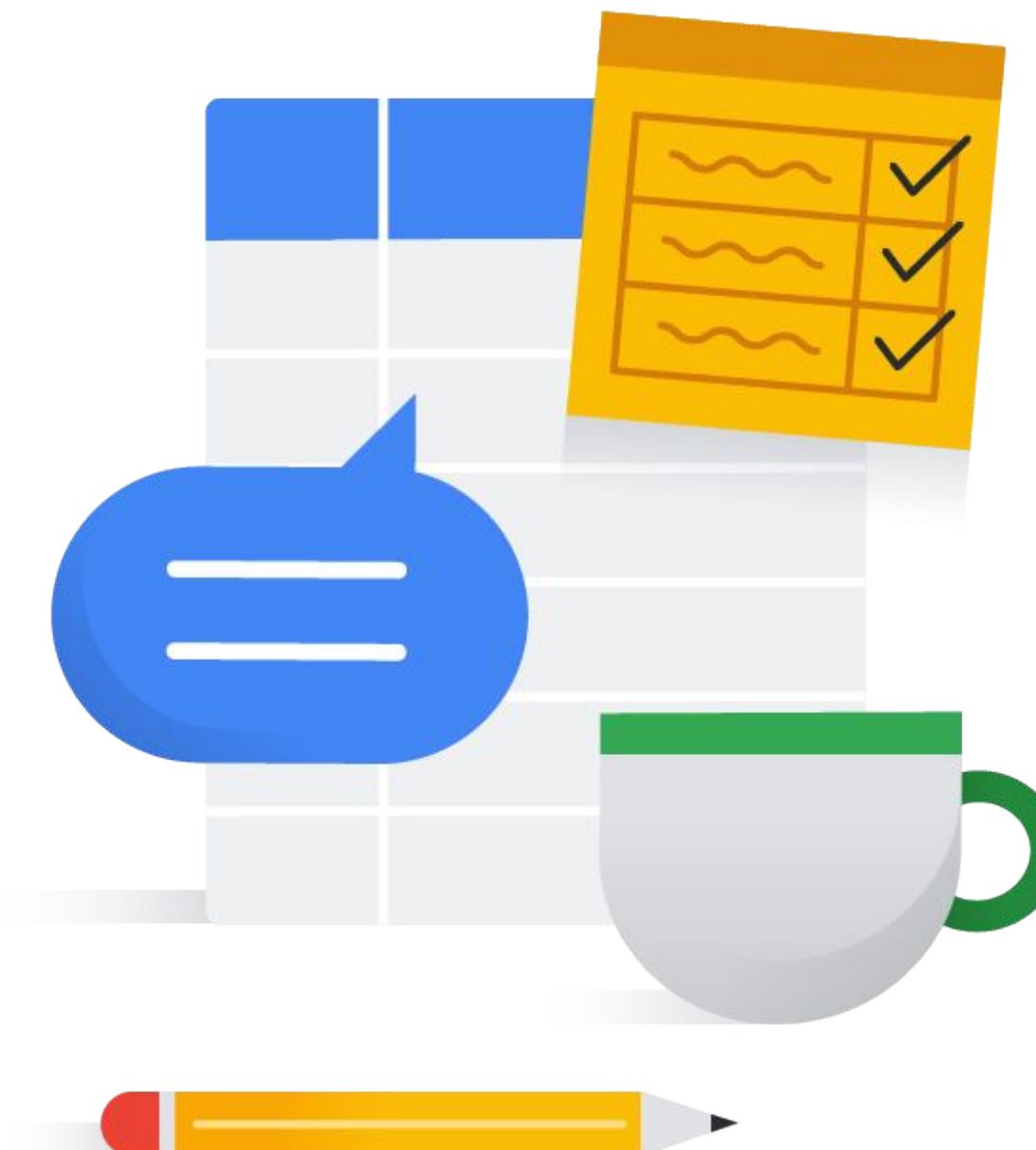
What is the primary way that agents learn in reinforcement learning?

- A. By being explicitly programmed with the correct actions
- B. By observing and imitating expert demonstrations
- C. By interacting with their environment and receiving feedback
- D. By analyzing large datasets of labeled examples



Key takeaways

- AI is the broad field, machine learning is a method or approach within AI, and generative AI is an application of AI that creates new content.
- Data quality and accessibility are crucial for AI. Understand data types (structured/unstructured) and quality for successful AI.
- Supervised, unsupervised, or reinforcement learning train ML models based on task and data.
- The ML lifecycle has several key stages supported by Google Cloud tools like Vertex AI for effective AI initiatives.



-
- 01 Core gen AI concepts

 - 02 Foundation models

 - 03 Building AI securely and responsibly
-

Agenda

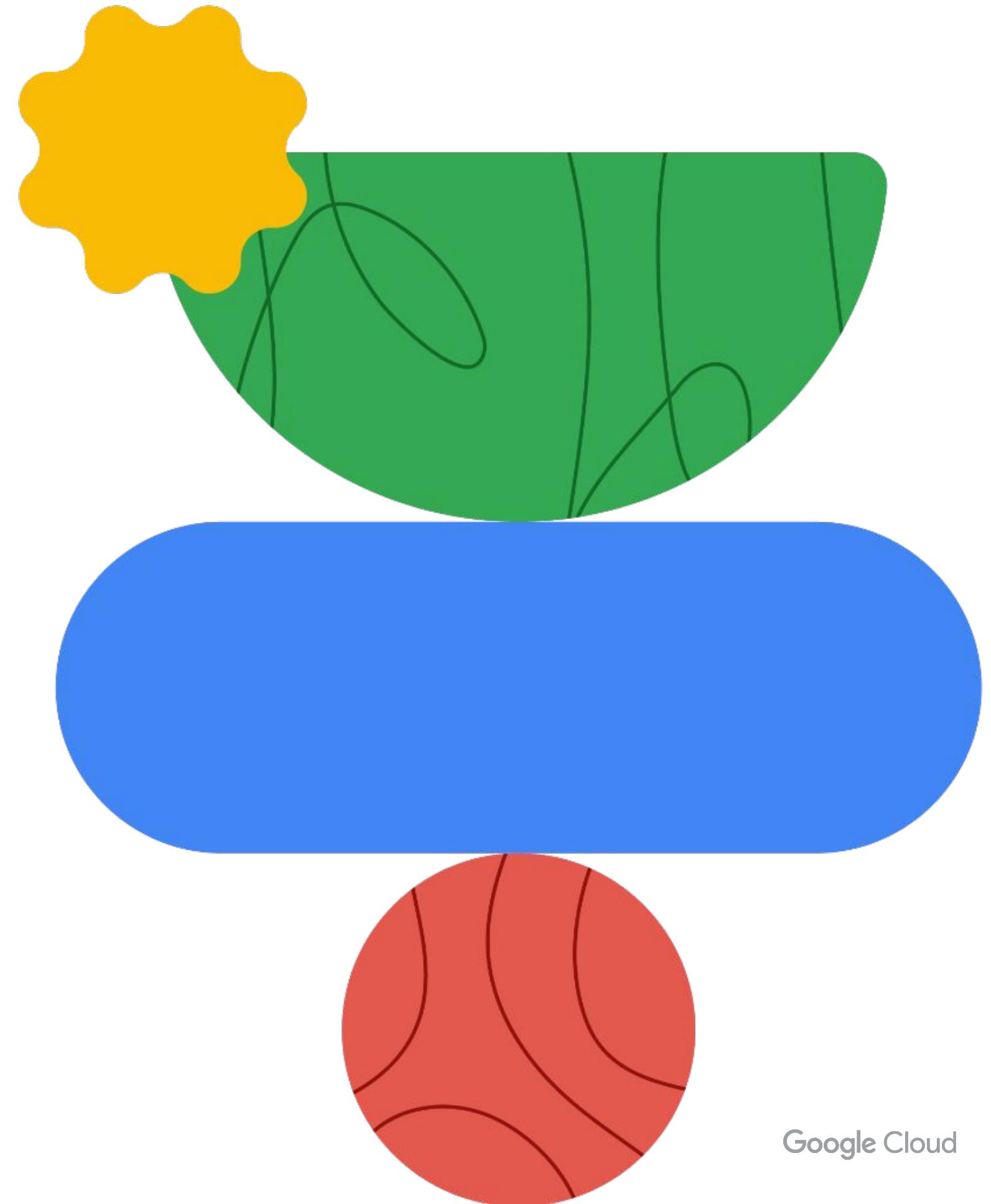




02

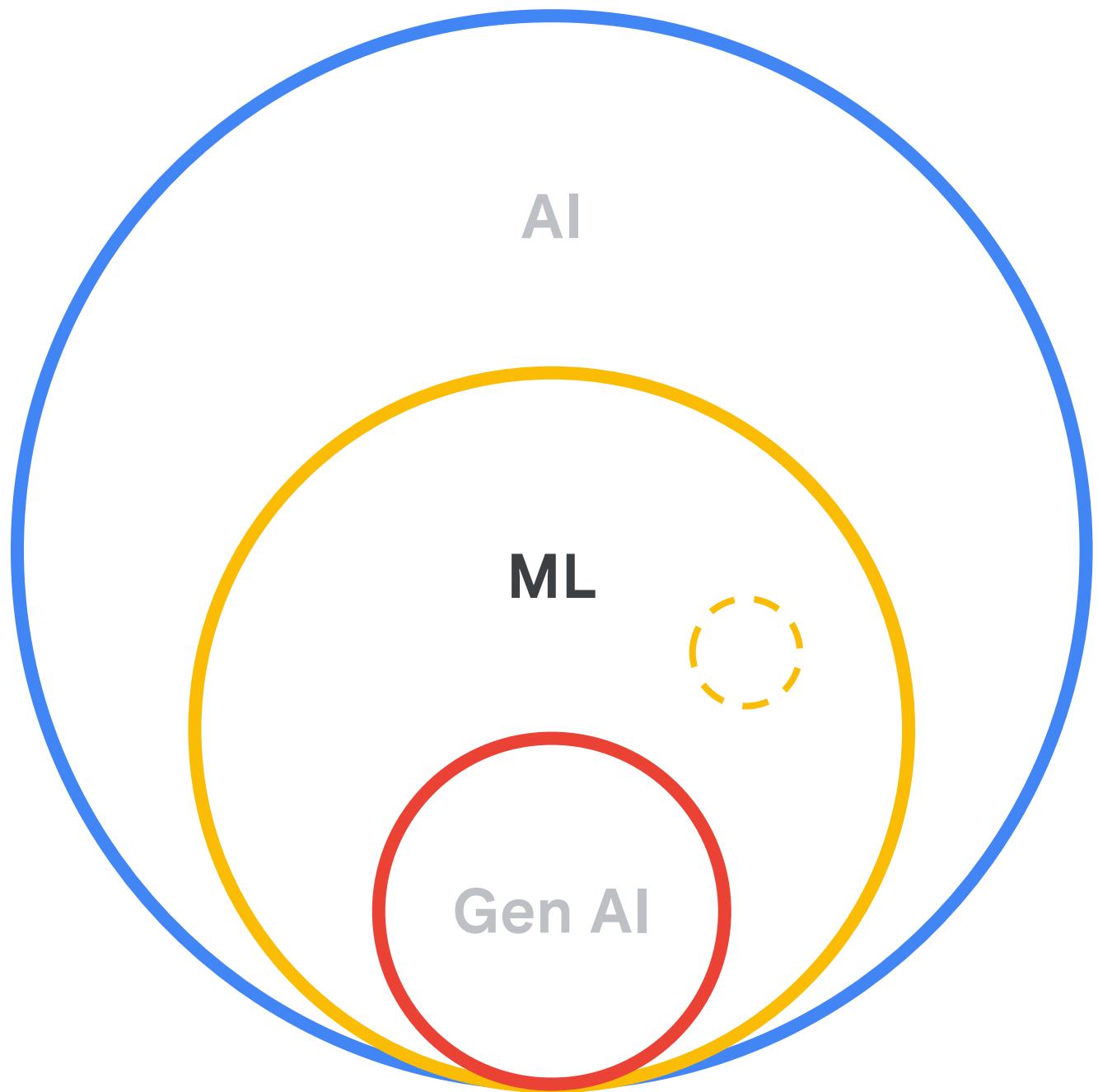
Foundation models

Deep learning



Google Cloud

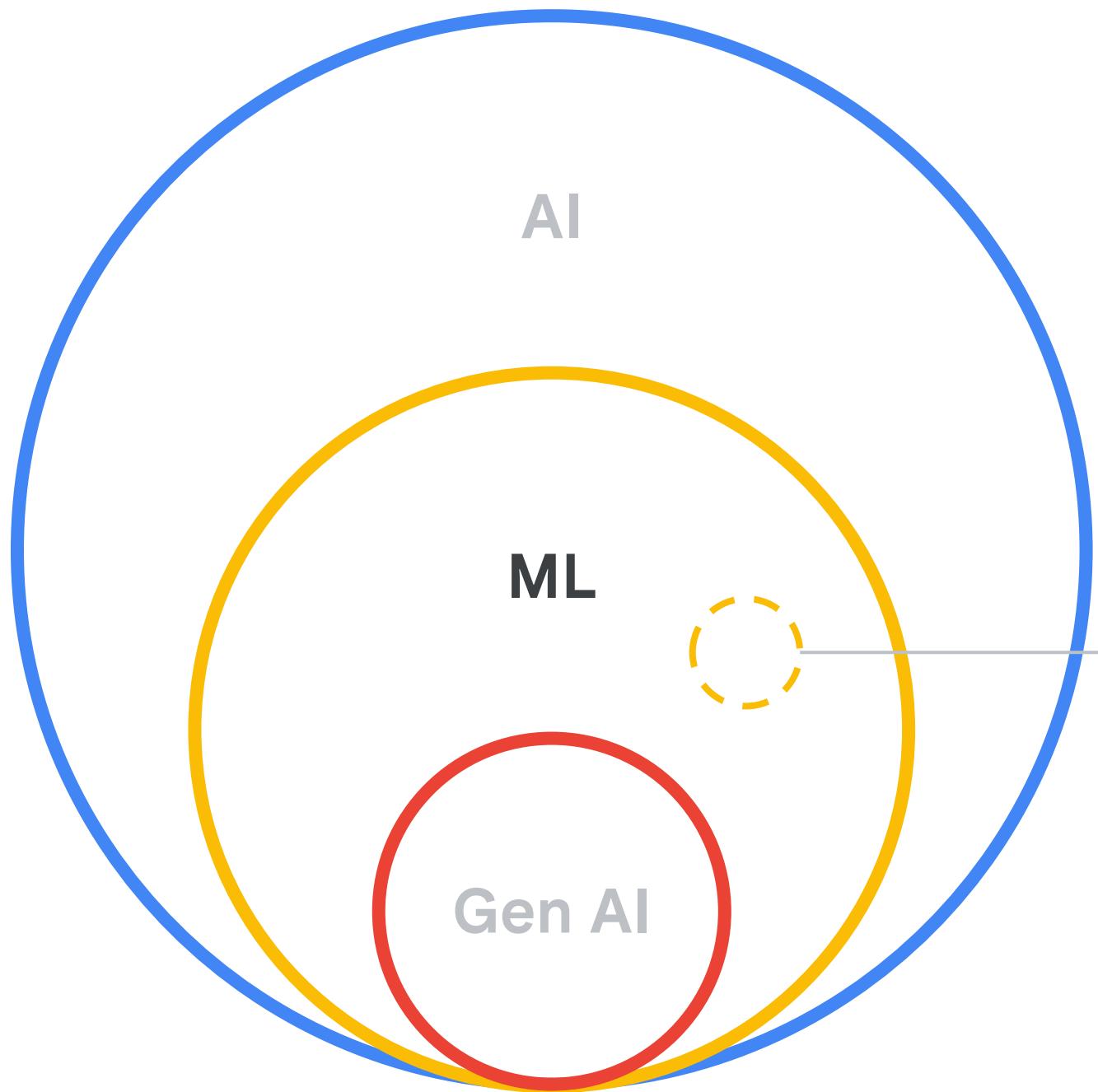
Deep learning



Machine learning

A broad field that encompasses many different techniques to teach computers to learn. One of these techniques is deep learning (DL).

Deep learning



Machine learning

A broad field that encompasses many different techniques to teach computers to learn. One of these techniques is deep learning (DL).

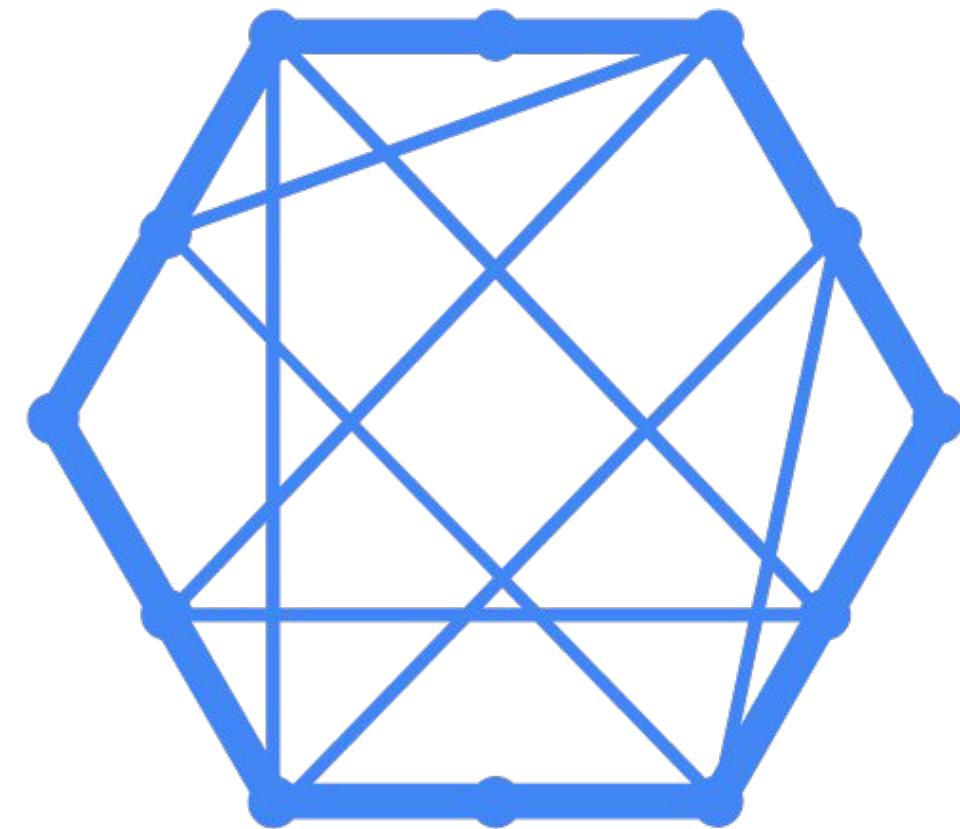
Deep learning

A specific way of teaching computers to learn from data by using artificial neural networks. Neural networks leverage labeled and unlabeled data (semi-supervised learning).

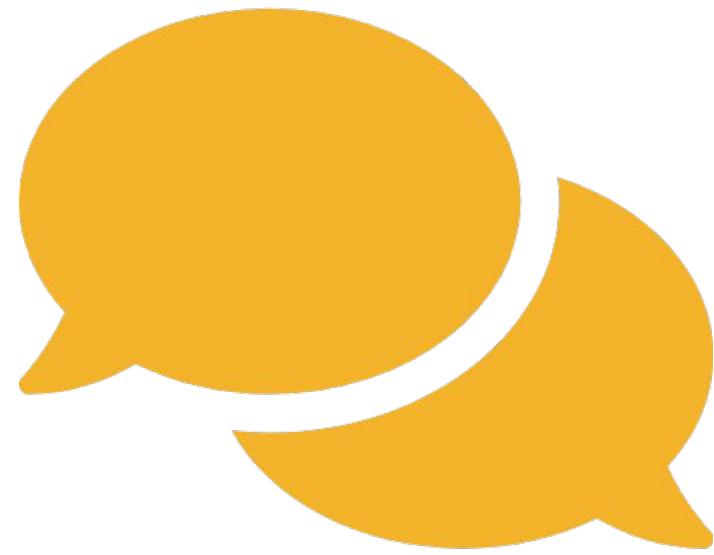
Generative AI uses the power of deep learning to create new content spanning text, images, audio, and beyond. **Deep learning techniques**, particularly those centered on neural networks, are the engine behind these generative models.

Foundation models

- They use deep learning. They are trained on massive datasets that allow them to learn complex patterns and perform a variety of tasks across different domains.
- They develop a broad understanding of the world, capturing intricate patterns and relationships within the data they consume.



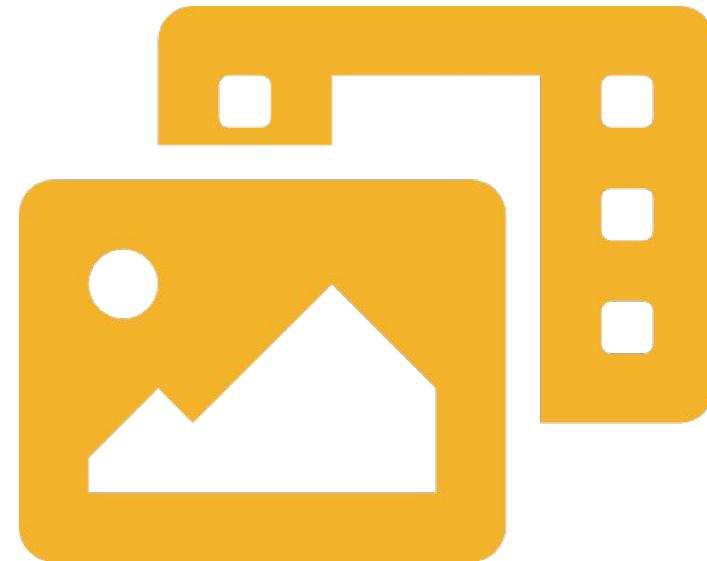
Types of foundation models



Large language models (LLMs)

They understand and generate human language.

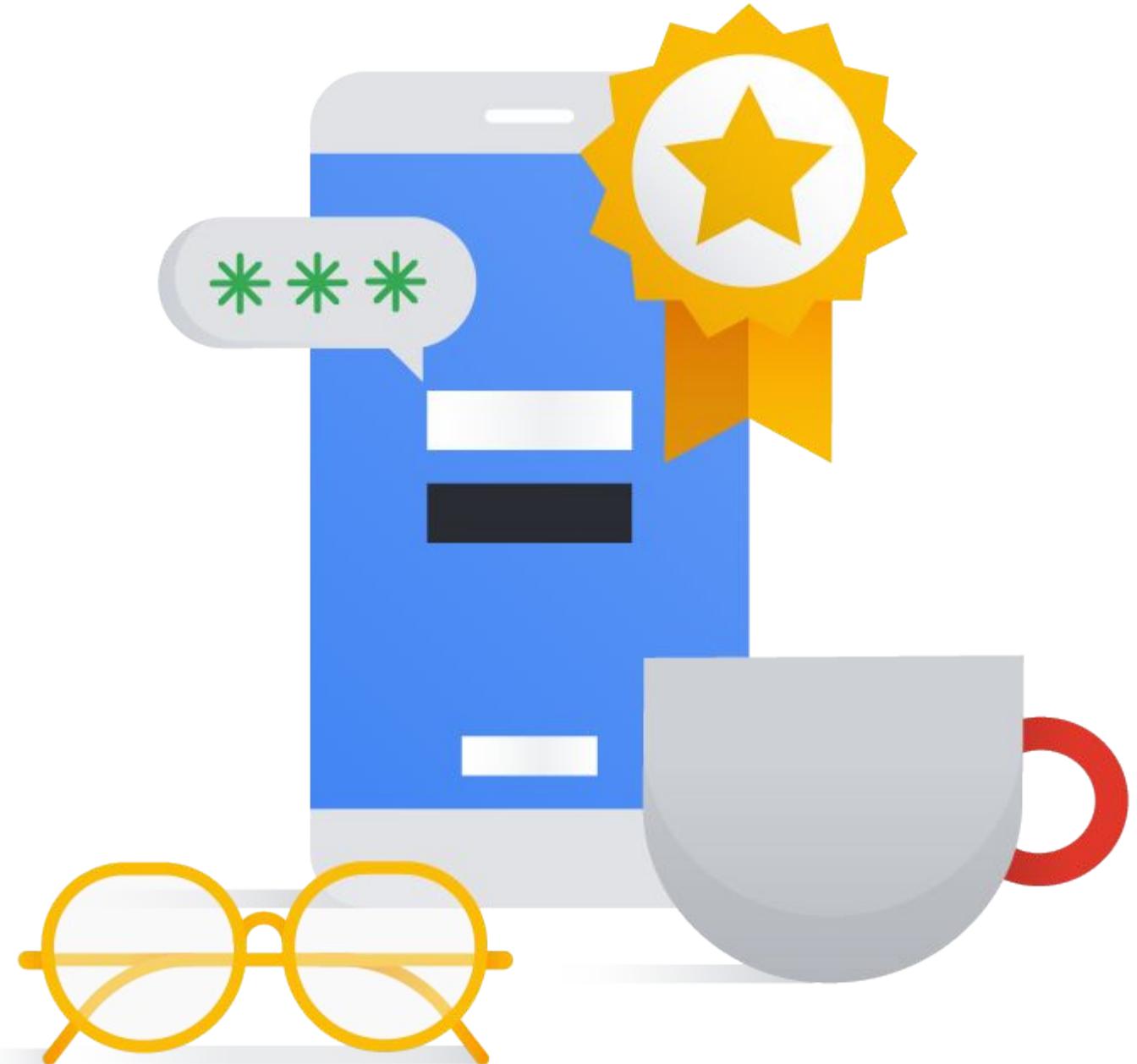
They translate, write content, answer questions.



Diffusion models

They generate high-quality images, audio, and video through iterative refining of data and patterns.

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



Quiz | Question 01

Question

What is the correct term for a subset of AI that enables machines to learn from data without explicit programming, improving their performance over time?

- A. Artificial intelligence
- B. Deep learning
- C. Prompting
- D. Machine learning

Quiz | Question 01

Answer

What is the correct term for a subset of AI that enables machines to learn from data without explicit programming, improving their performance over time?

- A. Artificial intelligence
- B. Deep learning
- C. Prompting
- D. Machine learning



Quiz | Question 02

Question

What is the correct term for a broad field of computer science focused on creating machines capable of performing tasks that typically require human intelligence?

- A. Artificial intelligence
- B. Deep learning
- C. Prompting
- D. Machine learning

Quiz | Question 02

Answer

What is the correct term for a broad field of computer science focused on creating machines capable of performing tasks that typically require human intelligence?

- A. Artificial intelligence
- B. Deep learning
- C. Prompting
- D. Machine learning



Quiz | Question 03

Question

What is the correct term for a specialized subset of machine learning that utilizes artificial neural networks with multiple layers to analyze complex data patterns?

- A. Artificial intelligence
- B. Deep learning
- C. Prompting
- D. Machine learning

Quiz | Question 03

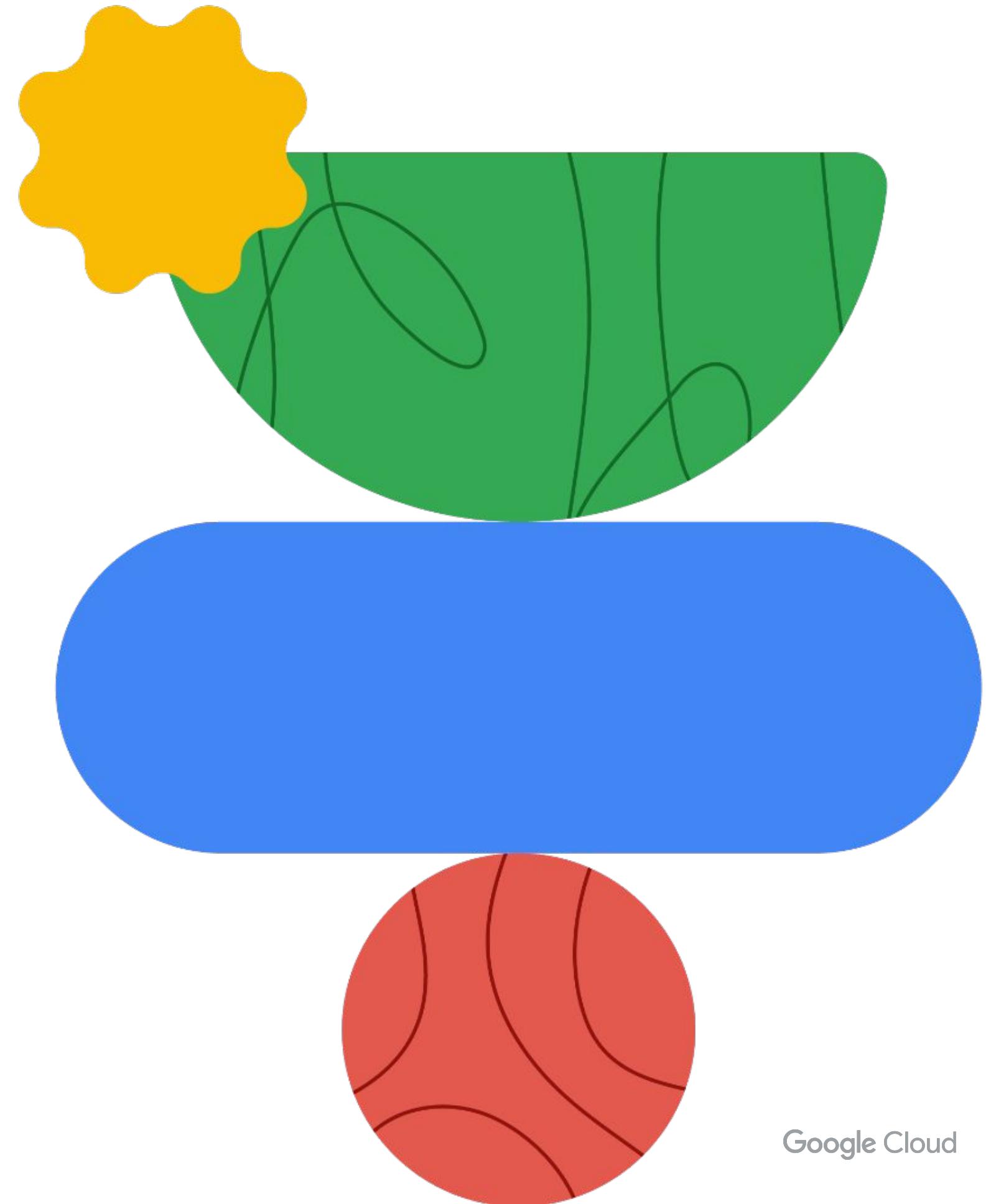
Answer

What is the correct term for a specialized subset of machine learning that utilizes artificial neural networks with multiple layers to analyze complex data patterns?

- A. Artificial intelligence
- B. Deep learning
- C. Prompting
- D. Machine learning

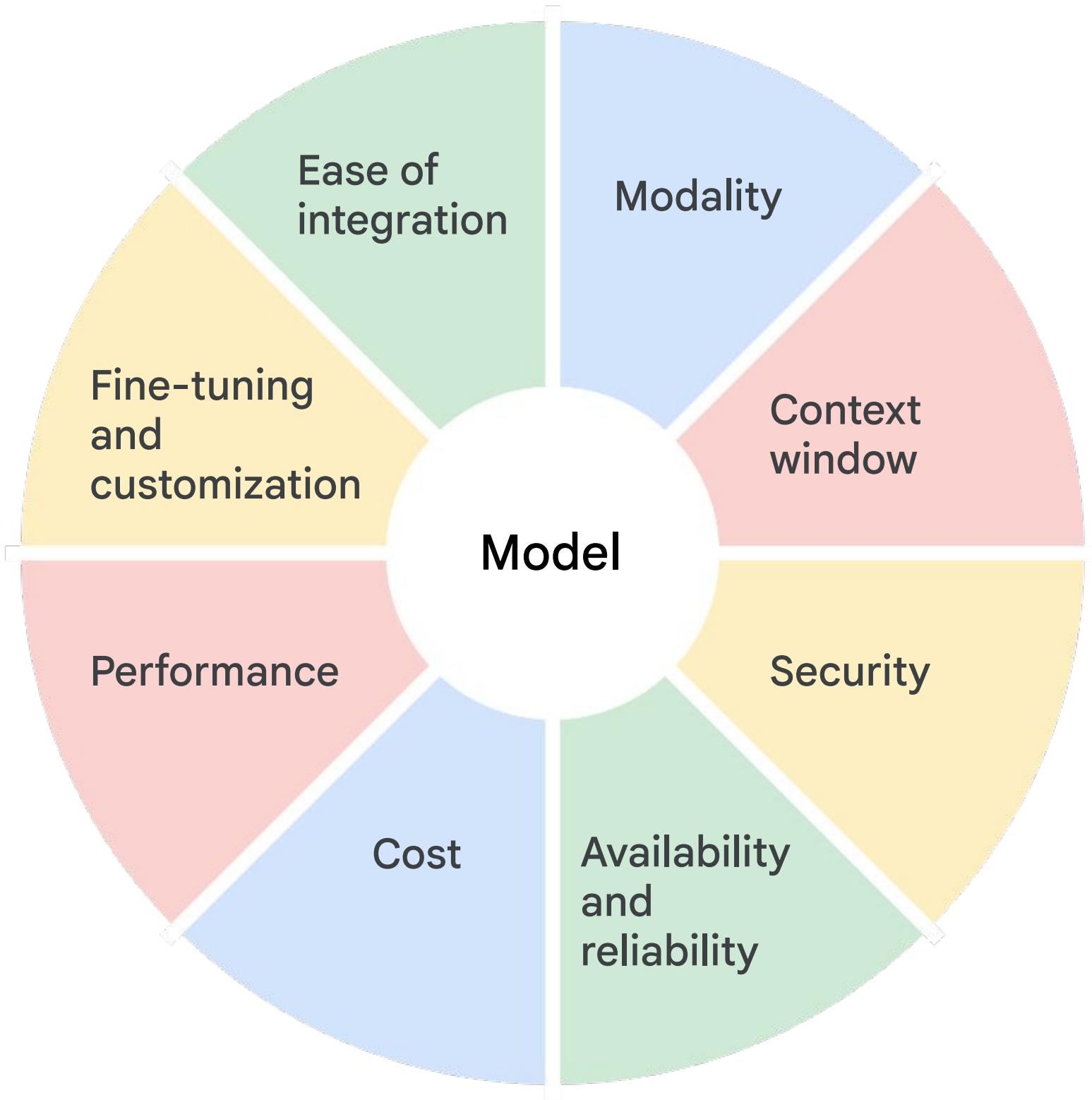


Choosing a model

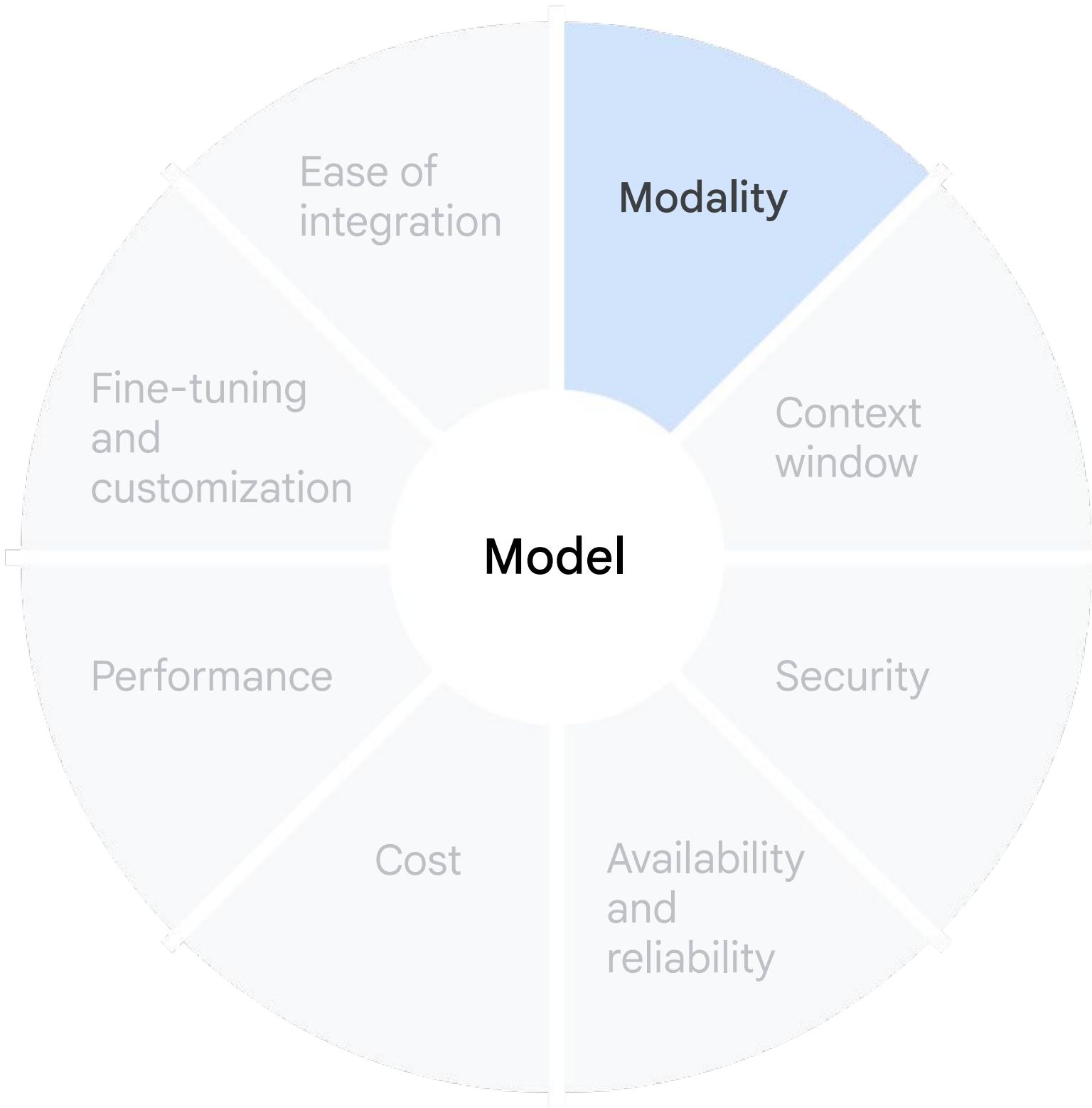


Google Cloud

Factors when choosing a model for your use case

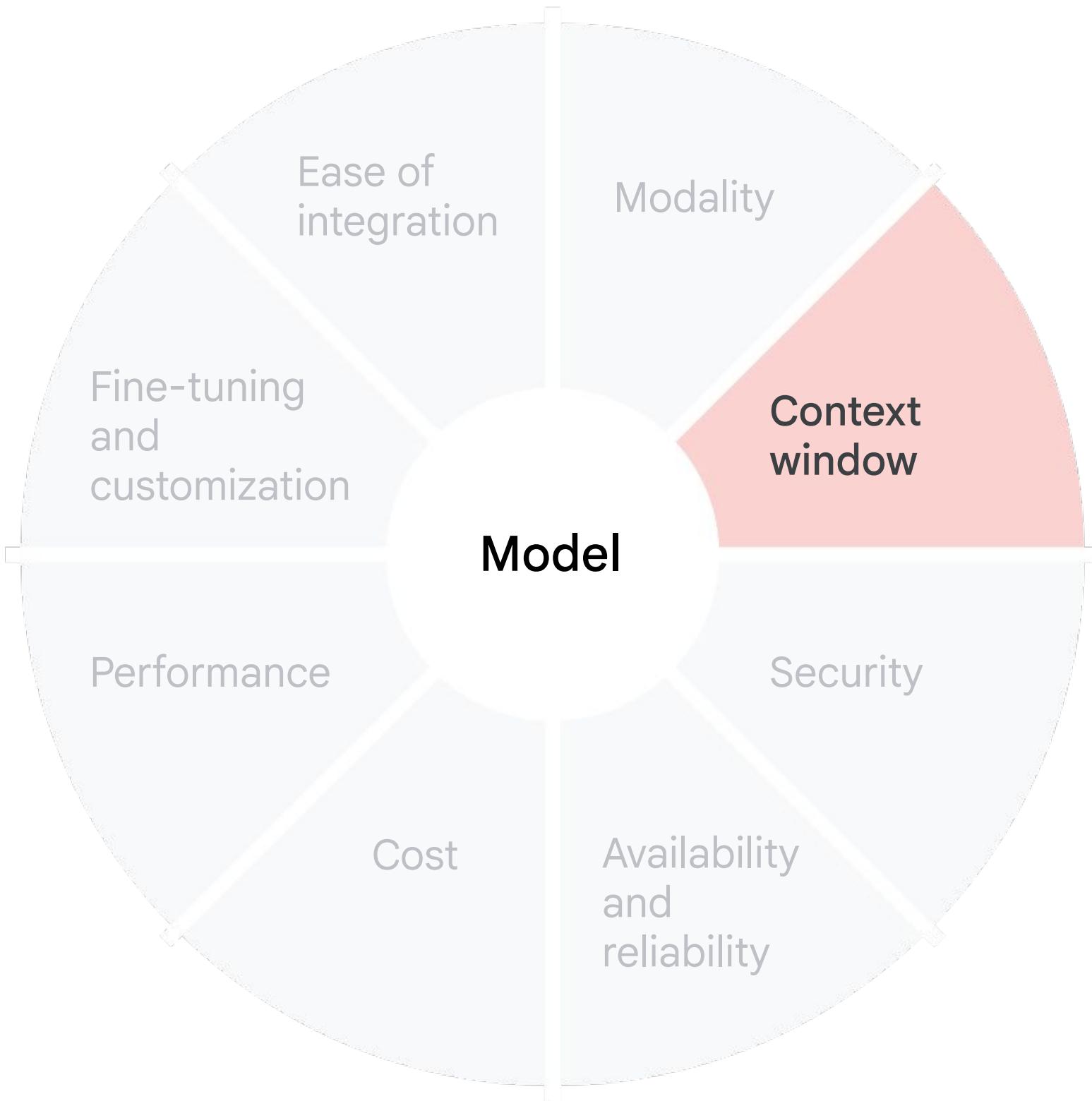


Factors when choosing a model for your use case



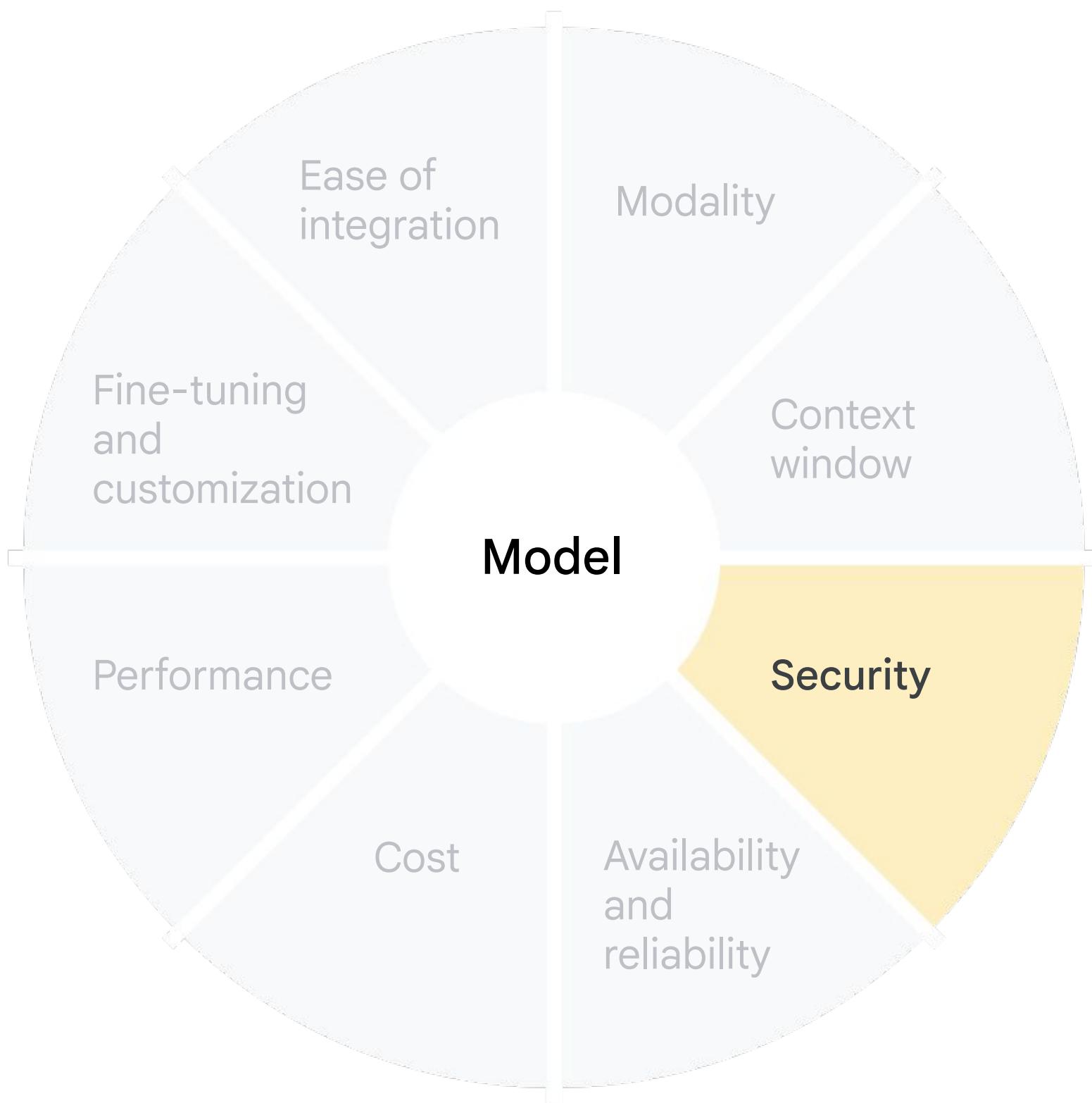
The type of data the model can process and generate e.g. text, images, video, audio.

Factors when choosing a model for your use case



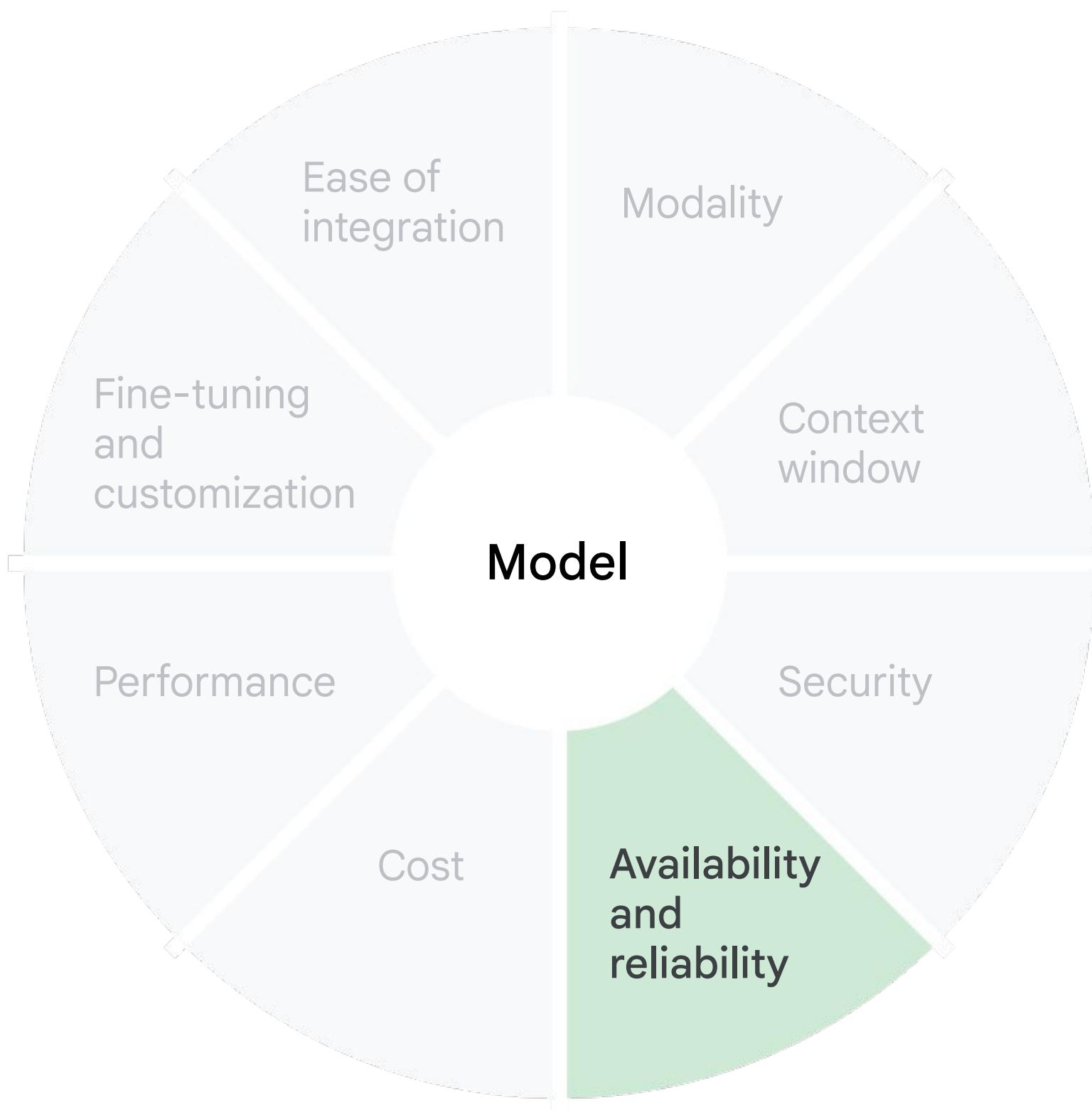
The amount of information a model can consider at one time when generating a response.

Factors when choosing a model for your use case



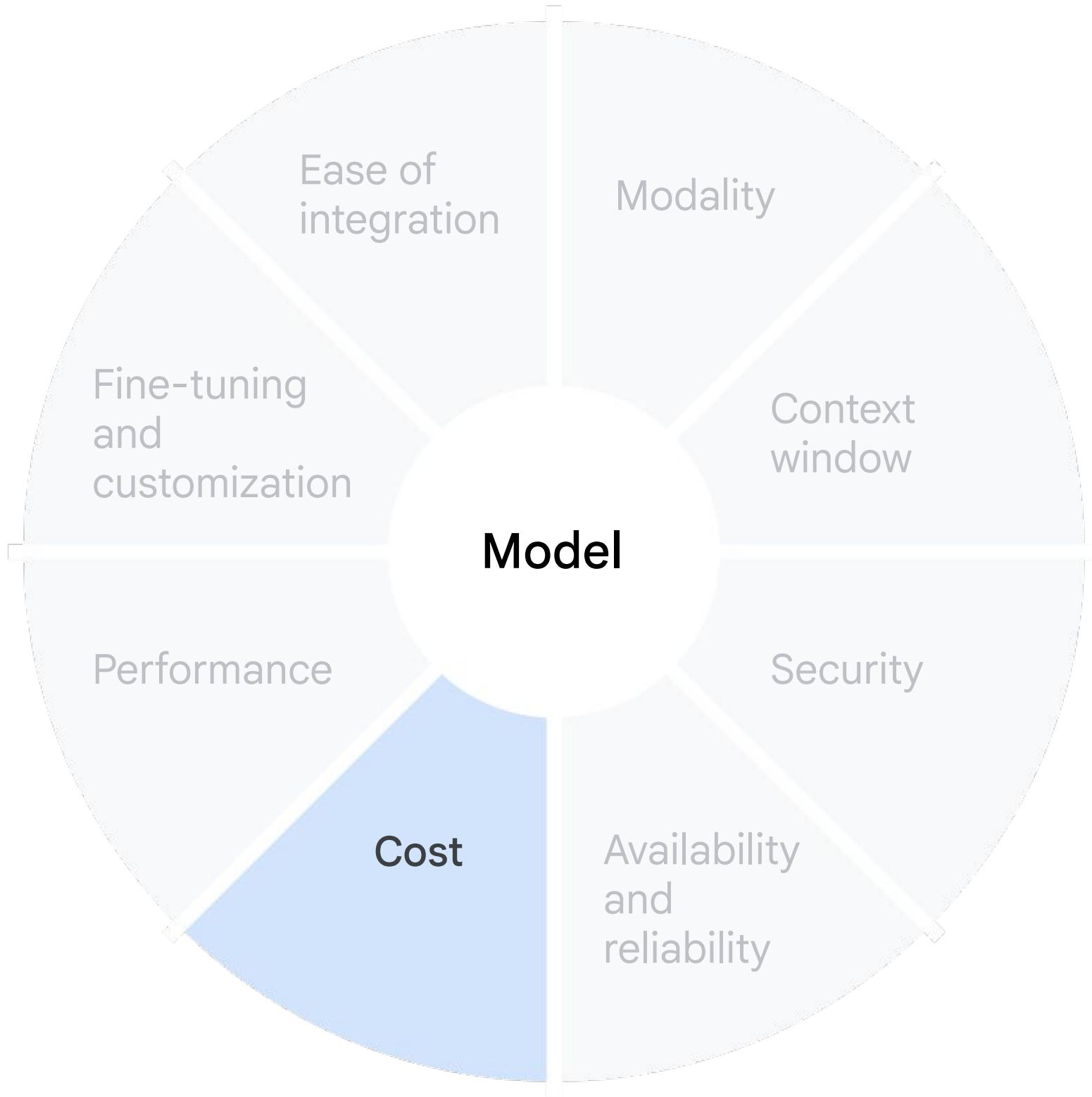
Security is paramount, especially with sensitive data.
Consider model security features, industry standards.

Factors when choosing a model for your use case



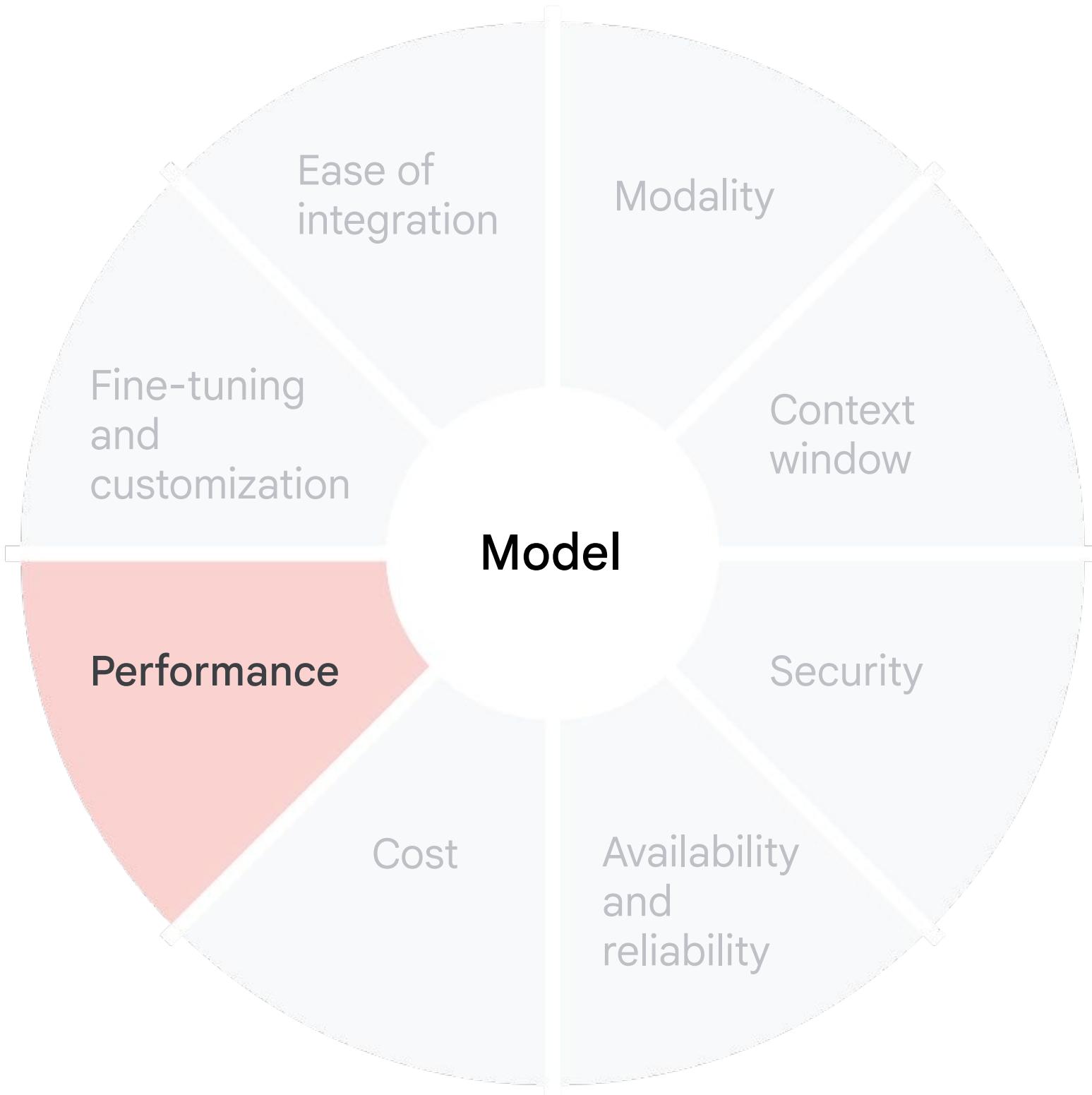
It is crucial for production applications.
Consider uptime guarantees, redundancy, and disaster recovery mechanisms.

Factors when choosing a model for your use case



Consider the pricing model and cost effectiveness.
Match the model to the task.

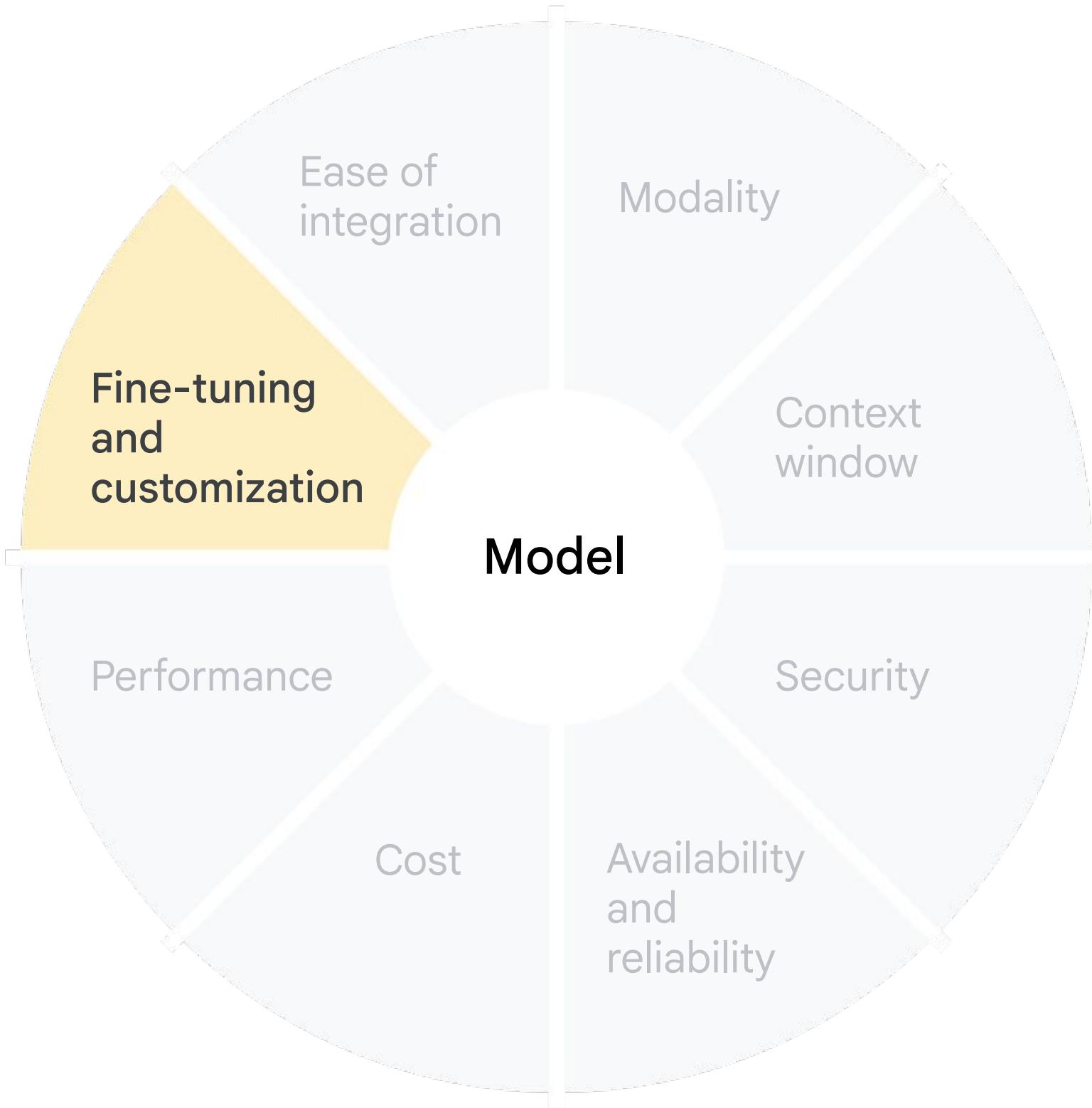
Factors when choosing a model for your use case



Accuracy, speed, and efficiency.

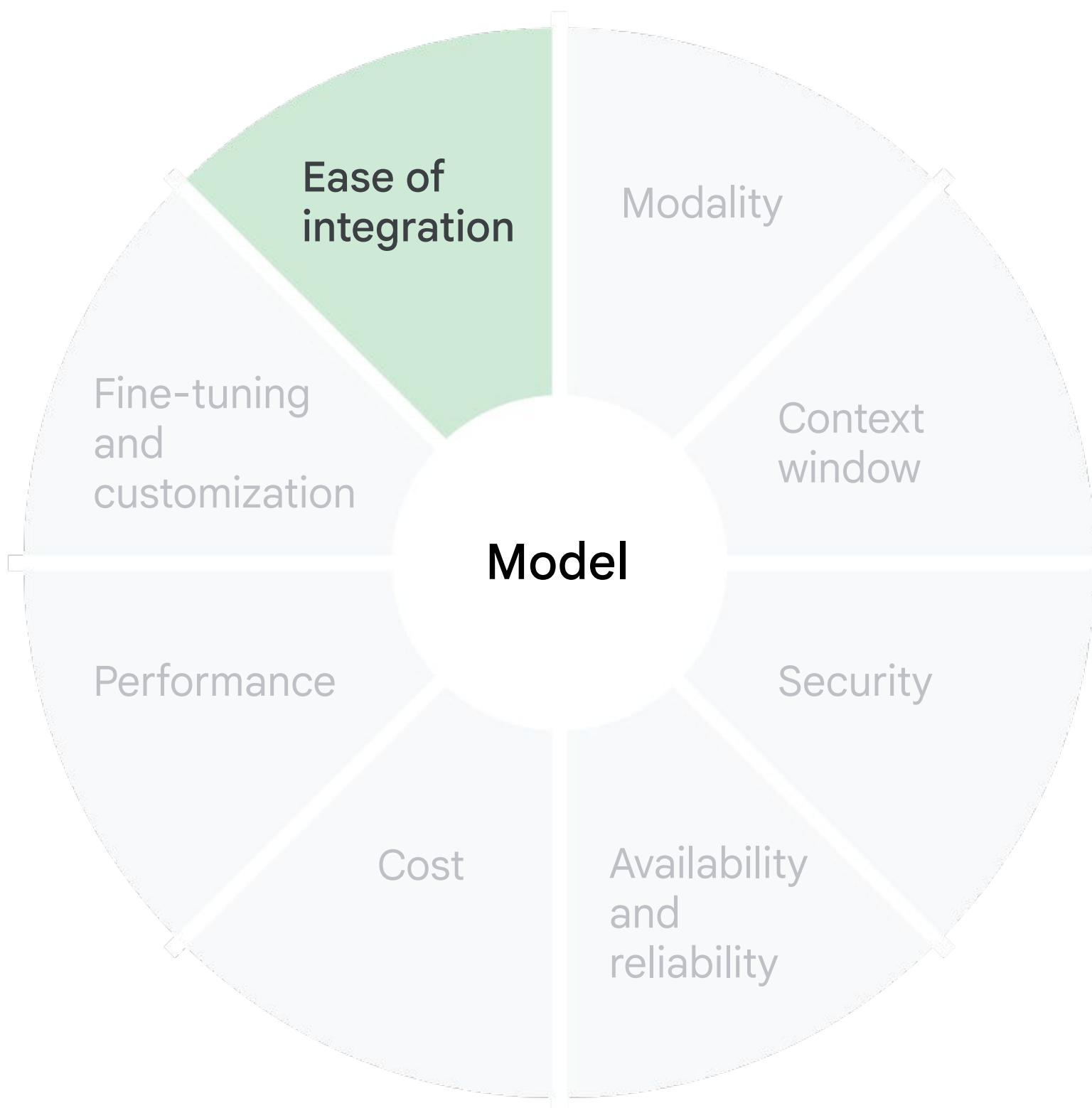
Evaluate performance on relevant benchmarks and datasets.

Factors when choosing a model for your use case



For specialized use cases, consider models that offer fine-tuning capabilities.

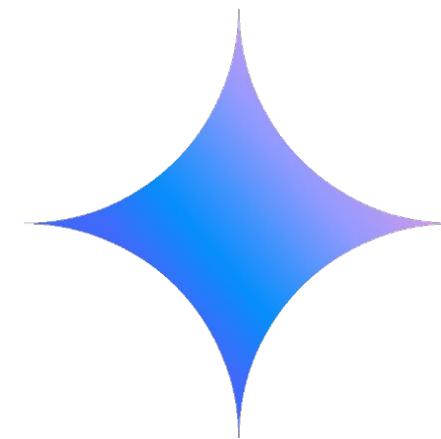
Factors when choosing a model for your use case



Integration into existing systems and workflows.
Well-documented APIs and SDKs.

Google Cloud's gen AI models

Gemini



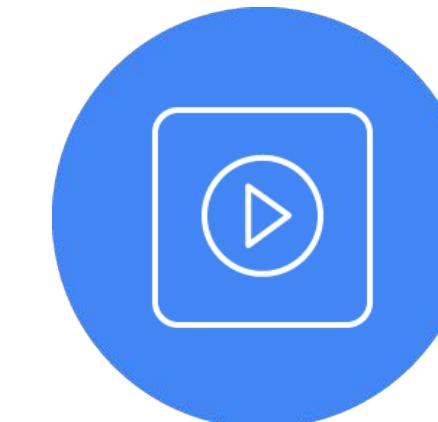
Imagen



Gemma



Veo



Google Cloud's gen AI models

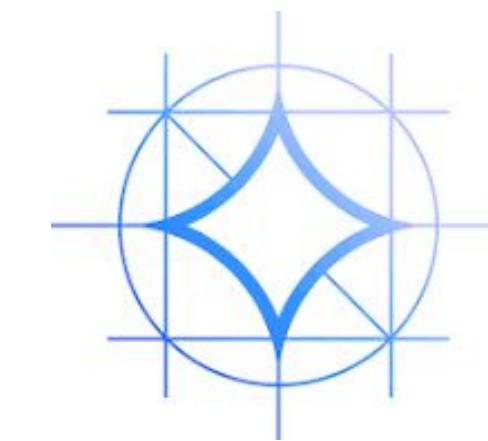
Gemini

Supports multimodal understanding, advanced conversational AI, content creation, and question answering.

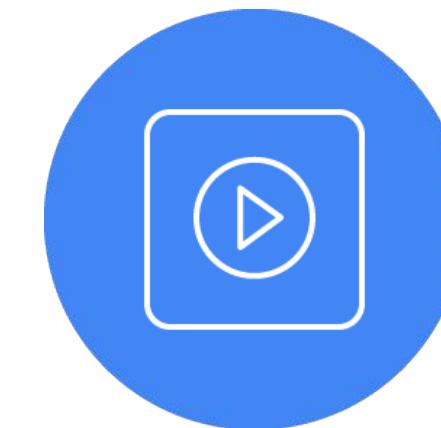
Imagen



Gemma

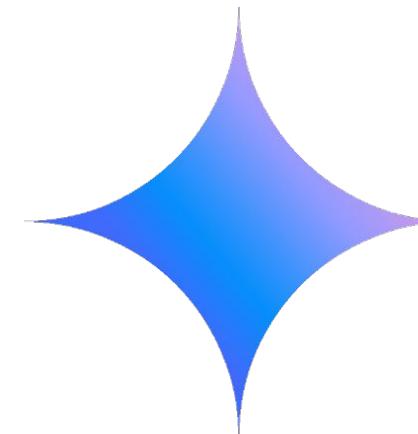


Veo



Google Cloud's gen AI models

Gemini



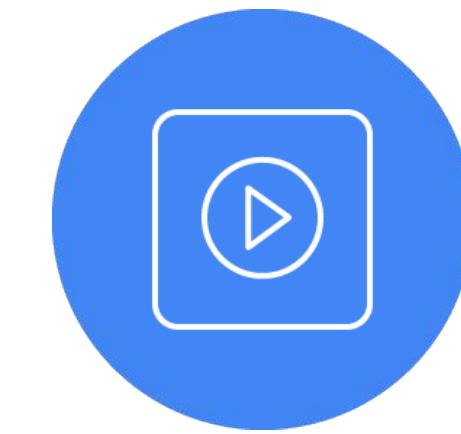
Imagen



Gemma

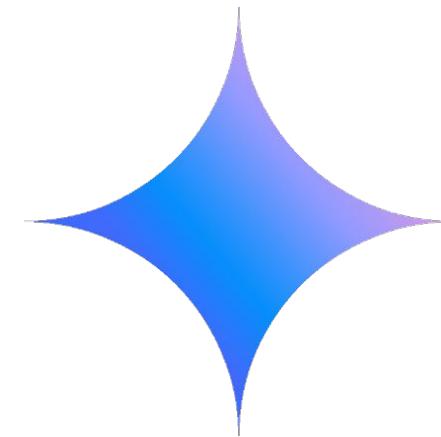
Offers developers a user-friendly, customizable solution for local deployments and specialized AI applications.

Veo



Google Cloud's gen AI models

Gemini



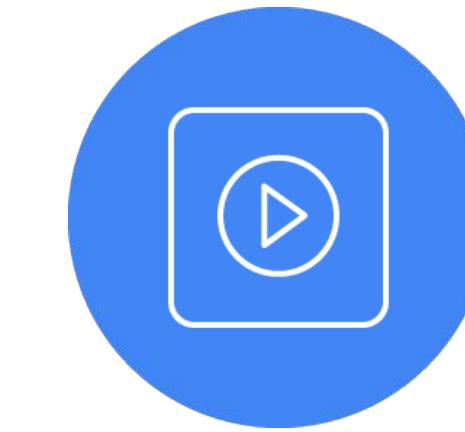
Gemma



Imagen

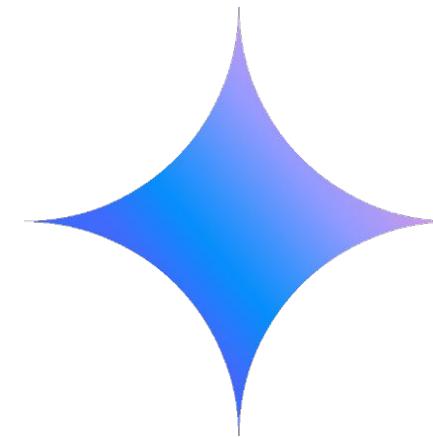
Generates high-quality images
from textual descriptions.

Veo



Google Cloud's gen AI models

Gemini



Imagen



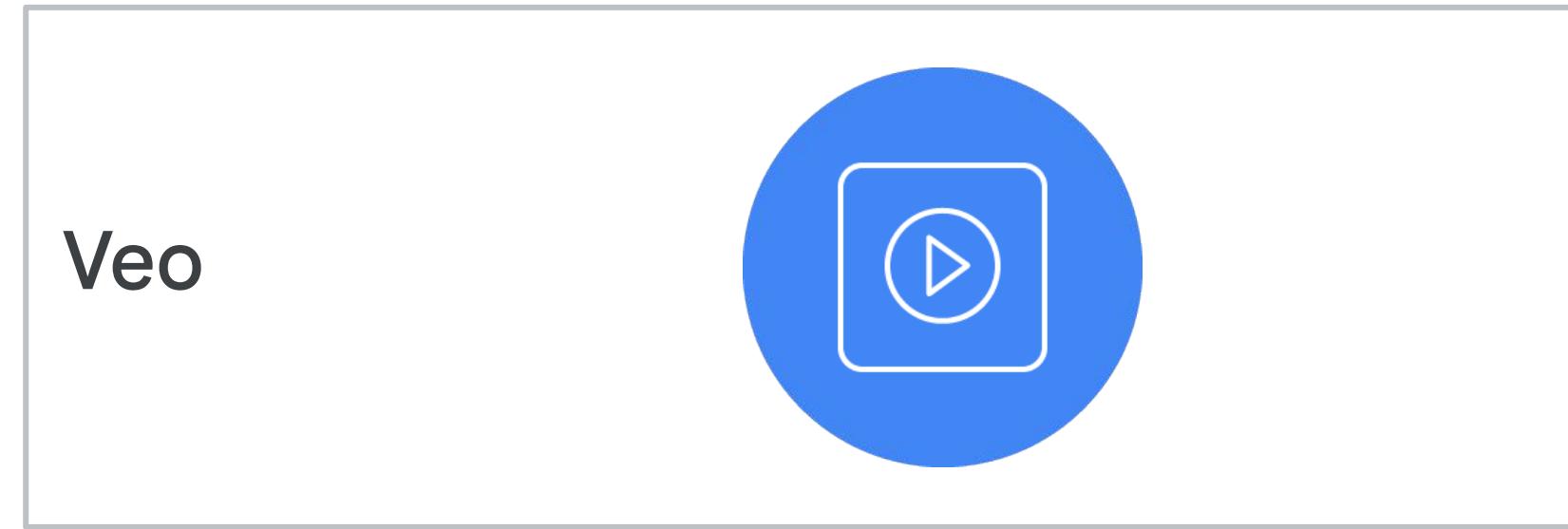
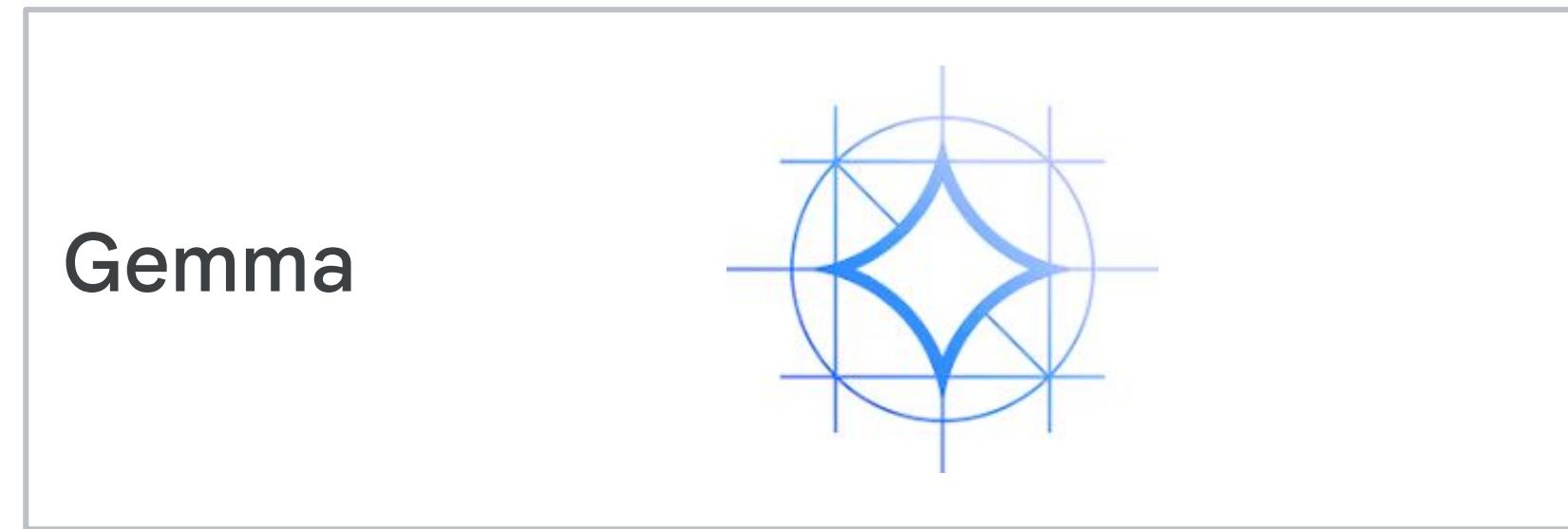
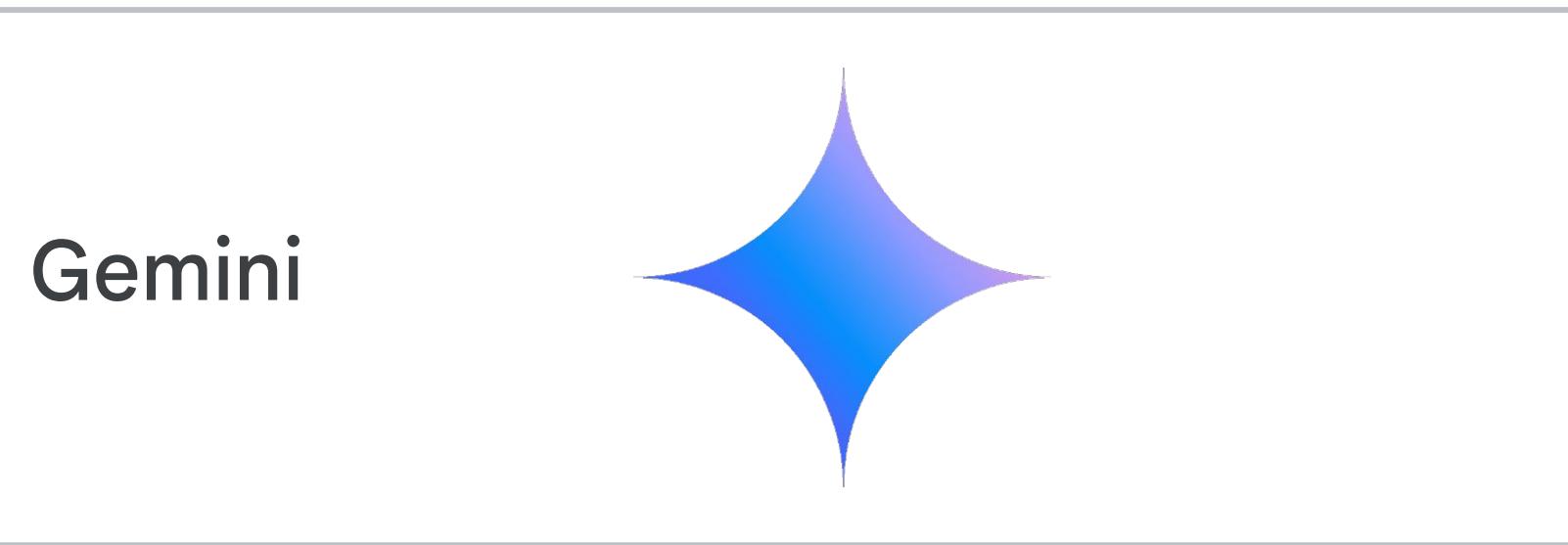
Gemma



Veo

Generates video content based on text descriptions or still images.

Google Cloud's gen AI models



Discussion: Model matchmaker

⌚ 5 min

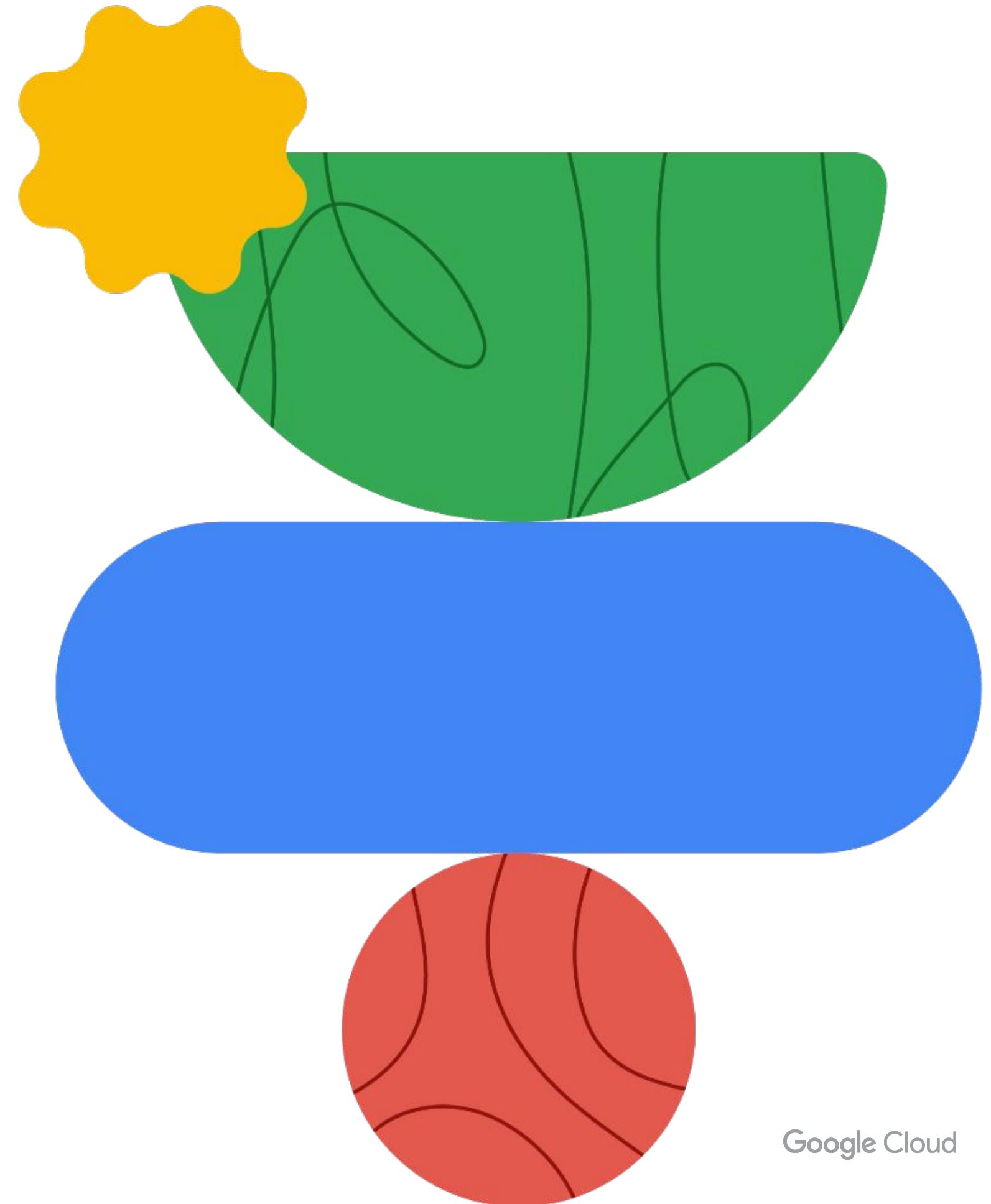
👥 Group

Gemini, Gemma, Imagen, and Veo

- Which specific business challenges or opportunities in your organizations do you see as most immediately addressable using these types of models?
- How can the strategic application of these models help your organization gain a competitive advantage in the market?



Google strategies for foundation model limitations



Foundation model limitations

-  Data dependency
-  Knowledge cut-off
-  Bias
-  Fairness
-  Hallucinations
-  Edge cases



Foundation model limitations

-  Data dependency
-  Knowledge cut-off
-  Bias
-  Fairness
-  Hallucinations
-  Edge cases

- The performance of foundation models is heavily data-dependent.
- Any biases or incompleteness will seep into their outputs.

Foundation model limitations

-  Data dependency
-  Knowledge cut-off
-  Bias
-  Fairness
-  Hallucinations
-  Edge cases

- This is the last date that an AI model was trained on new information.

Foundation model limitations

-  Data dependency
 - An LLM learns from large amounts of data, which may contain biases.
 - Subtle biases can be magnified in the outputs.
-  Knowledge cut-off
-  Bias
-  Fairness
-  Hallucinations
-  Edge cases

Foundation model limitations

-  Data dependency
 - Fairness can be interpreted differently.
 - Fairness assessments for gen AI models have inherent limitations. These evaluations typically focus on specific categories of bias, potentially overlooking other forms of prejudice.
-  Knowledge cut-off
-  Bias
-  Fairness
-  Hallucinations
-  Edge cases

Foundation model limitations

-  Data dependency
-  Knowledge cut-off
-  Bias
-  Fairness
-  Hallucinations
-  Edge cases

- Foundation models sometimes produce outputs that aren't accurate or based on real information.
- To fix this, we can ground the AI, connecting it to specific, reliable data.

Foundation model limitations

-  Data dependency
-  Knowledge cut-off
-  Bias
-  Fairness
-  Hallucinations
-  Edge cases

- Rare and atypical scenarios can expose a model's weaknesses, leading to errors, misinterpretations, and unexpected results.

Techniques to overcome limitations

- Prompt engineering
- Grounding
- Retrieval-Augmented Generation (RAG)
- Fine-tuning



Techniques to overcome limitations



Prompt engineering



Grounding



Retrieval-Augmented Generation (RAG)



Fine-tuning

Prompt engineering involves crafting precise prompts to guide the model towards desired outputs.

Techniques to overcome limitations

-  Prompt engineering
 -  Grounding
 -  Retrieval-Augmented Generation (RAG)
 -  Fine-tuning
- Grounding AI uses specific data, like company documents, to provide accurate and relevant enterprise-specific responses, increasing trustworthiness.

Techniques to overcome limitations



Prompt engineering



Grounding



Retrieval-Augmented Generation (RAG)



Fine-tuning

RAG is a grounding method that uses search to find relevant information from a knowledge base. Then, it provides that information to the LLM, giving it necessary context.

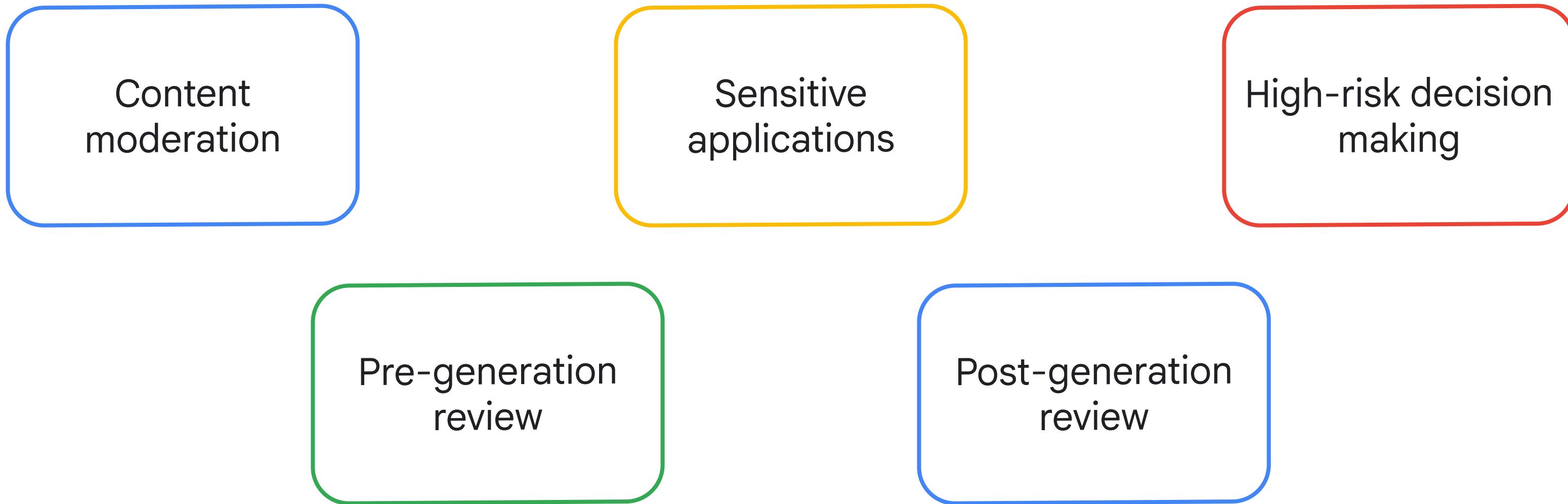
1. Retrieves information based on meaning.
2. Augments the prompt.
3. Generates a response.

Techniques to overcome limitations

-  Prompt engineering
-  Grounding
-  Retrieval-Augmented Generation (RAG)
-  Fine-tuning
- Tuning helps models excel in specific areas and is particularly useful for specific tasks or output formats.
- Tuning involves further training a pre-trained or foundation model on a new dataset specific to your task.

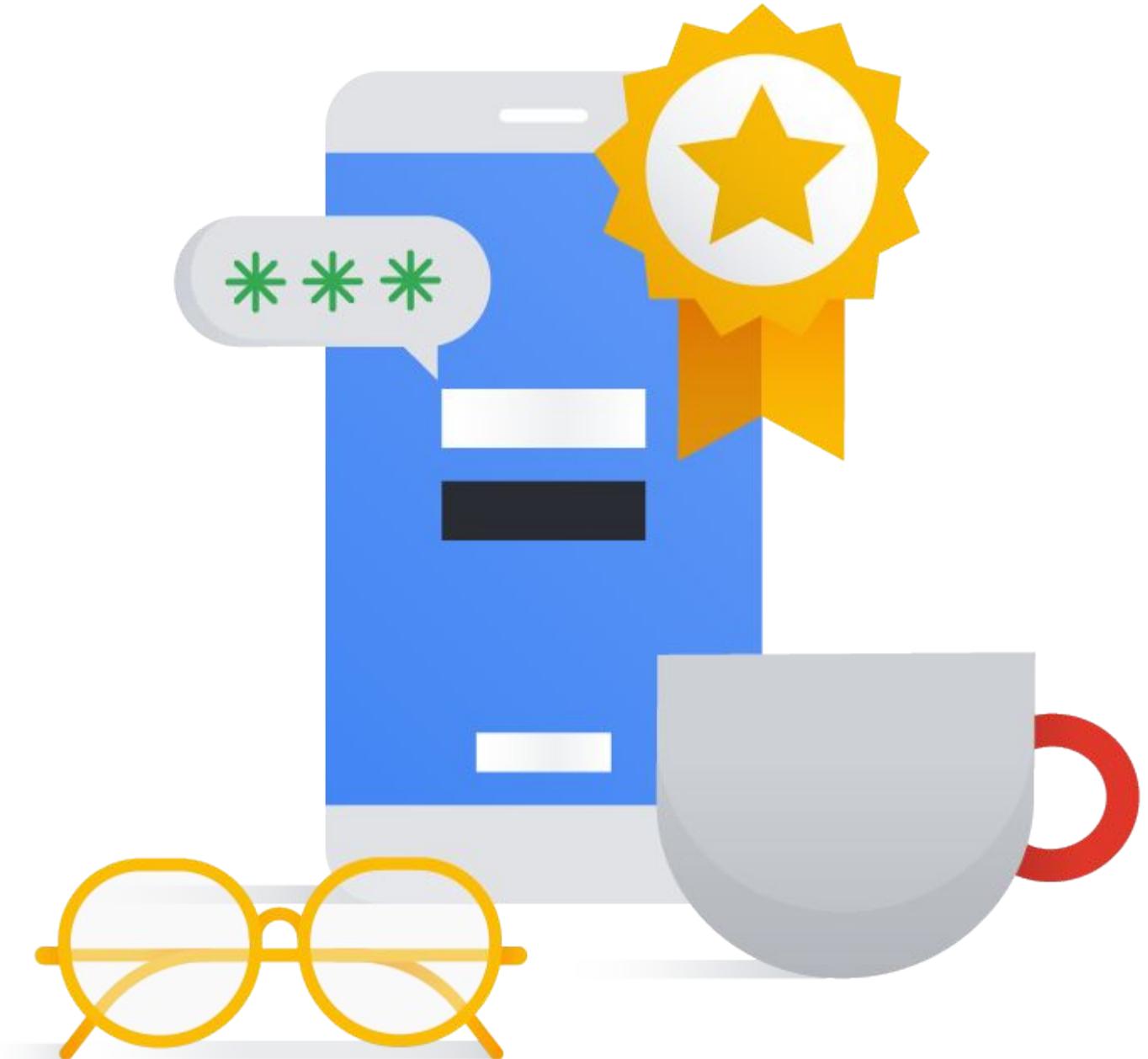


Humans in the loop (HITL)



Additional resource: [Summary of techniques to overcome limitations of foundation and pre-trained models](#)

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



Quiz | Question 01

Question

Which type of foundation model would be most suitable for generating photorealistic images from textual descriptions?

- A. Multimodal model
- B. Large language model (LLM)
- C. Diffusion model
- D. Classification model

Quiz | Question 01

Question

Which type of foundation model would be most suitable for generating photorealistic images from textual descriptions?

- A. Multimodal model
- B. Large language model (LLM)
- C. Diffusion model
- D. Classification model



Quiz | Question 02

Question

What is the primary role of humans in the loop (HITL) in machine learning?

- A. To replace AI algorithms entirely.
- B. To integrate human expertise into the ML process, especially for tasks requiring judgment or context.
- C. To automate all decision-making processes.
- D. To eliminate the need for data collection and model training.

Quiz | Question 02

Answer

What is the primary role of humans in the loop (HITL) in machine learning?

- A. To replace AI algorithms entirely.
- B. To integrate human expertise into the ML process, especially for tasks requiring judgment or context.
- C. To automate all decision-making processes.
- D. To eliminate the need for data collection and model training.



Quiz | Question 03

Question

Which techniques can be used to overcome the limitations of foundation model performance?

- A. Increased processing power and faster hardware
- B. Preventing hallucinations by restricting the AI model's access to external knowledge sources
- C. Grounding, prompt engineering, fine-tuning, and humans in the loop (HITL)
- D. Advanced algorithms and data structures

Quiz | Question 03

Answer

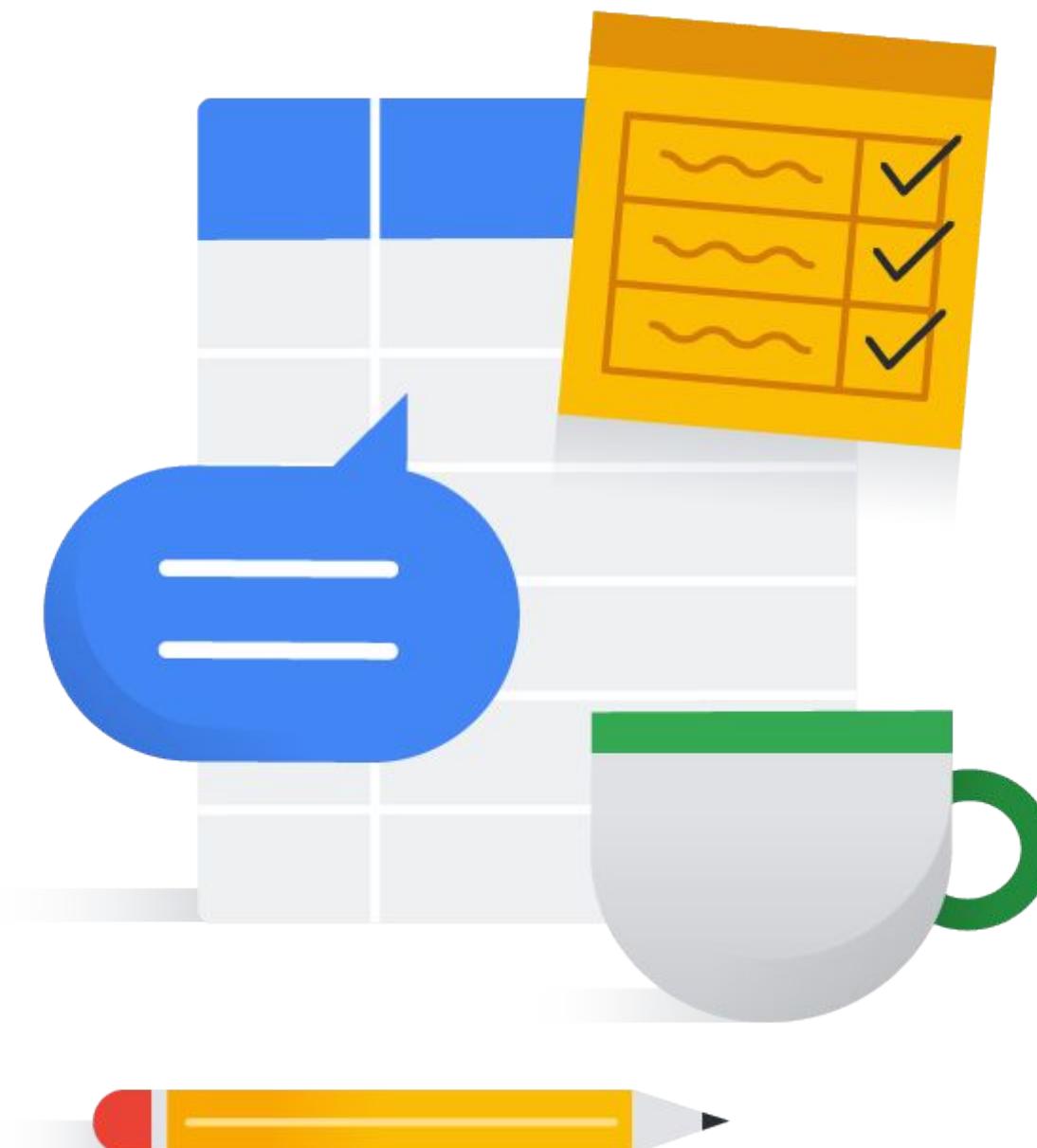
Which techniques can be used to overcome the limitations of foundation model performance?

- A. Increased processing power and faster hardware
- B. Preventing hallucinations by restricting the AI model's access to external knowledge sources
- C. Grounding, prompt engineering, fine-tuning, and humans in the loop (HITL)
- D. Advanced algorithms and data structures



Key takeaways

- Deep learning powers foundation models, enabling generative AI to create new content.
- Consider all relevant factors when choosing a generative AI model. Google Cloud's Vertex AI offers diverse models like Gemini, Gemma, Imagen, and Veo.
- Foundation models have limitations like bias and hallucinations. Grounding, prompt engineering, fine-tuning, and human-in-the-loop systems can address these.





Agenda

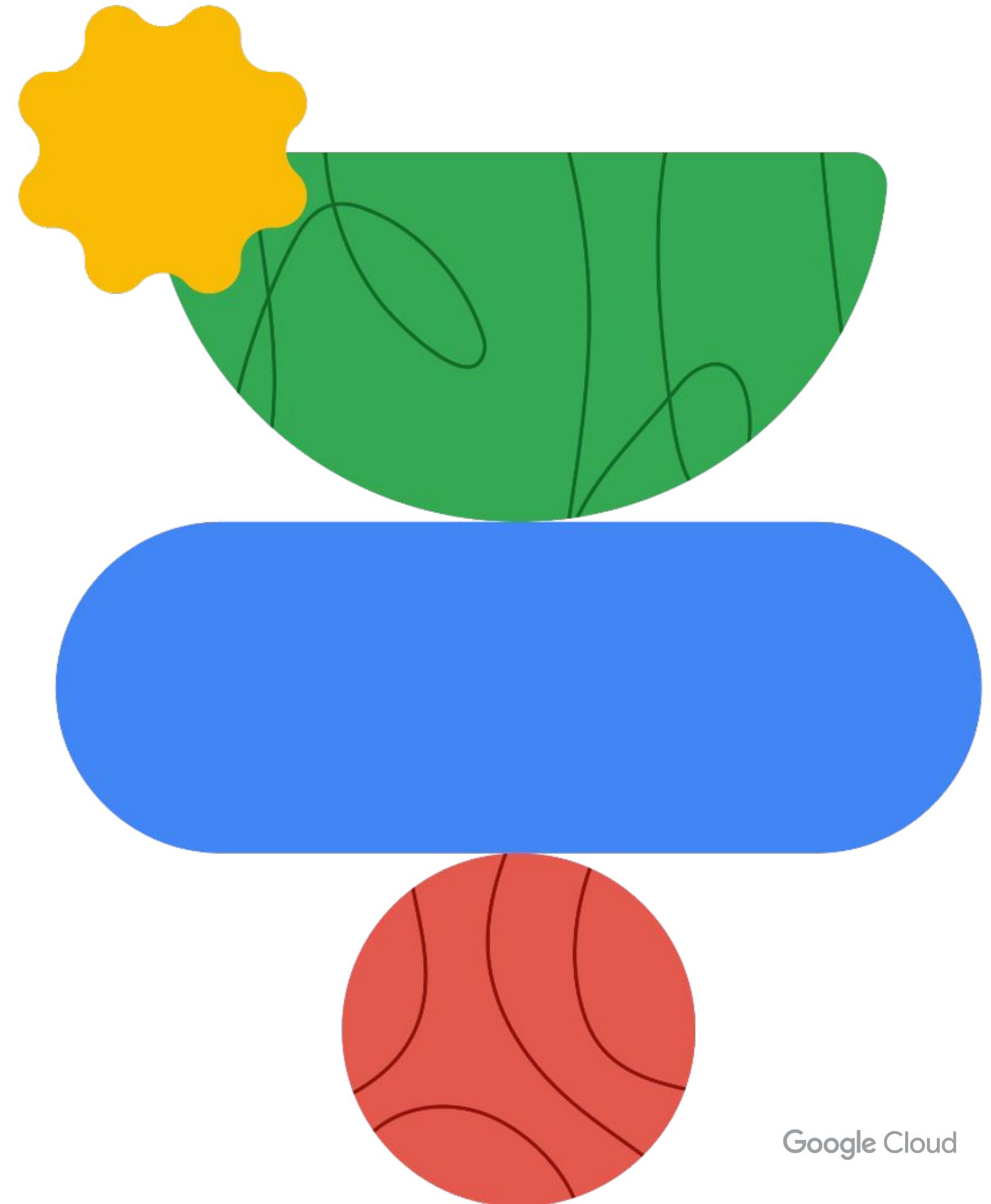


- 01 Core gen AI concepts
- 02 Foundation models
- 03 Building AI securely and responsibly

03

**Building AI securely
and responsibly**

Secure AI



Google Cloud

Secure AI principles applied to the ML lifecycle

- 1 Gather your data
- 2 Prepare your data
- 3 Train your model
- 4 Deploy and predict
- 5 Manage your model



Secure AI principles applied to the ML lifecycle

- 1 Gather your data
 - The foundation of any robust AI system is secure data, which must therefore be protected and secured at all times.
 - Access, addition, and input to data must be controlled.
- 2 Prepare your data
- 3 Train your model
- 4 Deploy and predict
- 5 Manage your model

Secure AI principles applied to the ML lifecycle

- 1 Gather your data
 - Special attention should be paid to confidential or sensitive data that might be present in the training data.
- 2 Prepare your data
 - Anonymization
 - Validation
 - Secure processing
 - Logging and real-time monitoring
- 3 Train your model
- 4 Deploy and predict
- 5 Manage your model

Secure AI principles applied to the ML lifecycle

- 1 Gather your data
 - Safeguarding both the training data and model parameters from unauthorized access or modification is paramount.
- 2 Prepare your data
- 3 Train your model
- 4 Deploy and predict
- 5 Manage your model

Secure AI principles applied to the ML lifecycle

- 1** Gather your data
 - Control access to the model.
 - Verify sources, and check for potential vulnerabilities for pre-built models.
- 2** Prepare your data
- 3** Train your model
- 4** Deploy and predict
- 5** Manage your model

Secure AI principles applied to the ML lifecycle

- 1 Gather your data
 - Stay up-to-date on the latest security best practices.
- 2 Prepare your data
 - Make updates regularly.
 - Monitor model performance and outputs for anomalies or signs of tampering.
 - Review access permissions regularly.
- 3 Train your model
- 4 Deploy and predict
- 5 Manage your model

Secure AI principles applied to the ML lifecycle

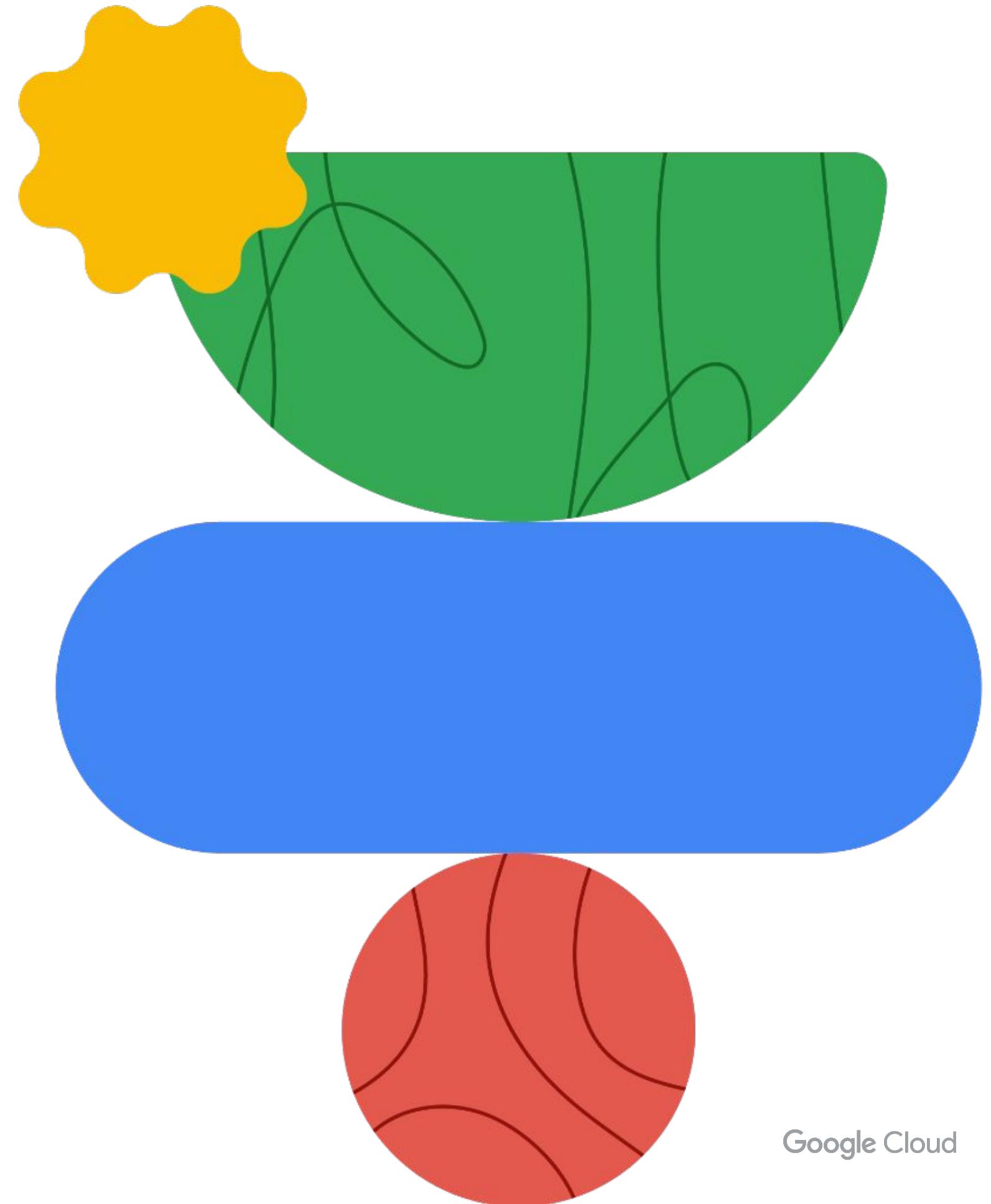
- Guard against adversarial attacks.
- Monitor AI outputs to prevent leaks and harmful content.



The **Secure AI Framework (SAIF)** establishes security standards for **building and deploying AI systems responsibly**.

Google Cloud enables secure development with its secure-by-design **infrastructure, encryption, IAM, Security Command Center, and tools for monitoring**, for comprehensive security.

Responsible AI



Google Cloud

Foundations of responsible AI



Transparency is key



Privacy in the age of AI



Data quality, bias, and fairness



Accountability and explainability

Foundations of responsible AI



Transparency is key

Users need to understand how their information is being used and how the AI system works.



Privacy in the age of AI



Data quality, bias, and fairness



Accountability and explainability

Foundations of responsible AI



Transparency is key

Protect privacy by anonymizing or pseudonymizing data.



Privacy in the age of AI

Safeguard against models inadvertently leaking sensitive information from training data.



Data quality, bias, and fairness



Accountability and explainability

Foundations of responsible AI



Transparency is key



Privacy in the age of AI



Data quality, bias, and fairness



Accountability and explainability

Ethical AI requires high quality data and responsible use of data itself.

AI inherits societal biases, causing unfair outcomes. Fairness must be central to AI development.

Foundations of responsible AI

-  Transparency is key
-  Privacy in the age of AI
-  Data quality, bias, and fairness
-  Accountability and explainability
- Fairness requires accountability.
Explainable AI makes the decision-making processes of AI models transparent and understandable.
- Understand how your application uses and interprets the AI's output.

Legal implications

Key legal areas for AI development

- Data privacy
- Non-discrimination
- Intellectual property
- Product liability



Legal aspects of AI

- AI laws require responsible data, bias mitigation, transparency, and model compliance.
- The evolving legal landscape demands vigilance and counsel for trustworthy AI.

Discussion: Organizational roadblocks to responsible AI

⌚ 5 min

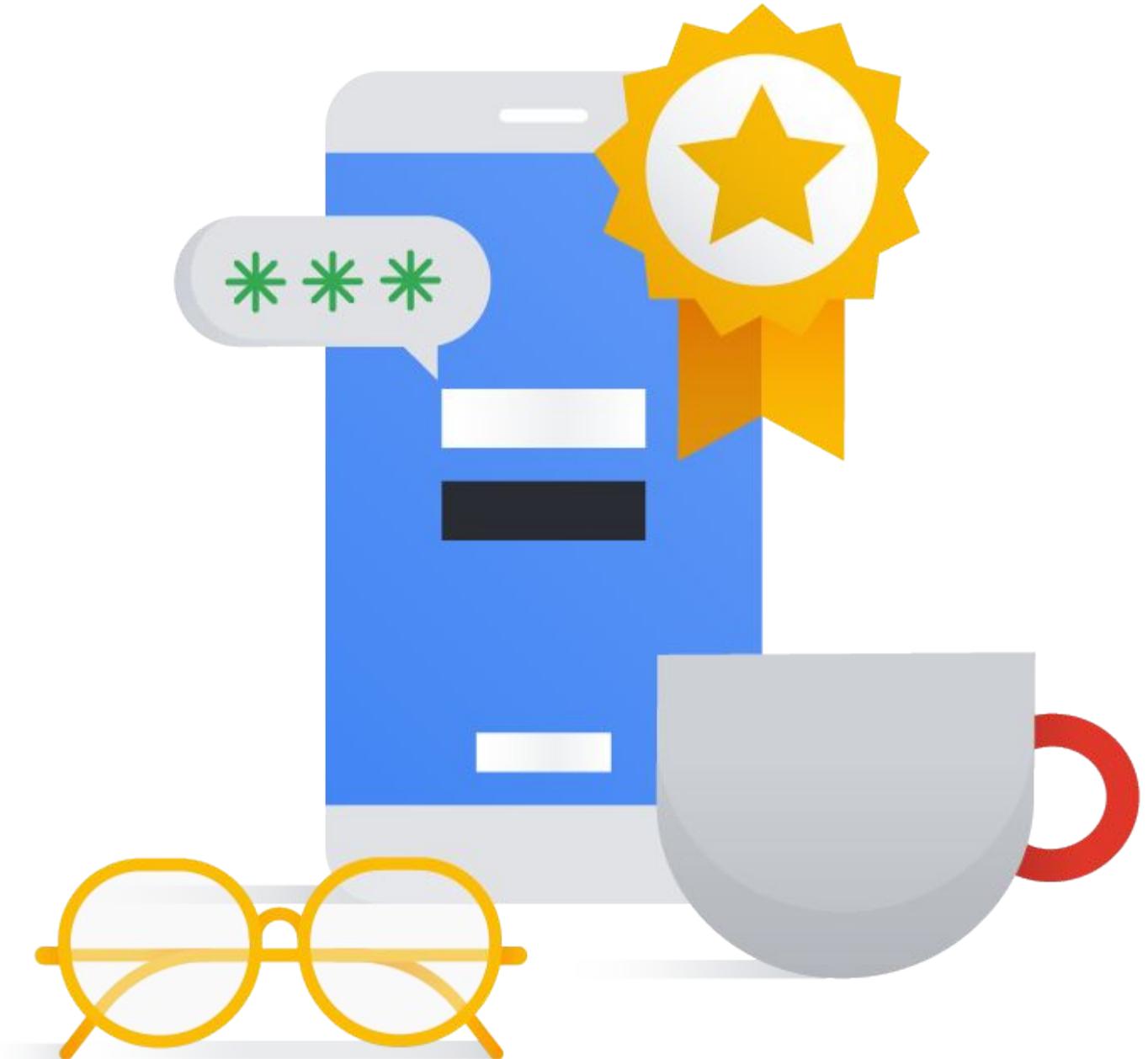
👥 Group

Imagine your organization is committed to developing and deploying AI responsibly.

- What is one significant internal challenge that you foresee hindering this commitment? For example, data governance, lack of expertise, or conflicting priorities.
- Briefly explain why.



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



Quiz | Question 01

Question

You're developing an AI-powered loan application assessment system. Which steps would you take to ensure your system is ethically sound? Select two.

- A. Prioritize speed and efficiency over fairness and transparency. These factors are more important for business success.
- B. Use a complex "black box" model that makes highly accurate predictions. Even if its decision-making process is difficult to understand.
- C. Regularly audit the model's performance to identify and mitigate any biases that may emerge over time.
- D. Deploy the model without any human oversight. Trust its ability to make fair and unbiased decisions.
- E. Train the model on a dataset that includes a diverse range of applicants. Ensure representation across different demographics and socioeconomic backgrounds.

Quiz | Question 01

Answers

You're developing an AI-powered loan application assessment system. Which steps would you take to ensure your system is ethically sound? Select two.

- A. Prioritize speed and efficiency over fairness and transparency. These factors are more important for business success.
- B. Use a complex "black box" model that makes highly accurate predictions. Even if its decision-making process is difficult to understand.
- C. Regularly audit the model's performance to identify and mitigate any biases that may emerge over time. 
- D. Deploy the model without any human oversight. Trust its ability to make fair and unbiased decisions.
- E. Train the model on a dataset that includes a diverse range of applicants. Ensure representation across different demographics and socioeconomic backgrounds. 

Quiz | Question 02

Question

What is the primary goal of the Secure AI Framework (SAIF)?

- A. To promote innovation and accelerate AI development without considering security risks.
- B. To restrict AI development and deployment due to concerns about potential security risks.
- C. To establish security standards for building and deploying AI responsibly, addressing the unique challenges and threats in the AI landscape.
- D. To focus solely on preventing external attacks on AI systems.

Quiz | Question 02

Answer

What is the primary goal of the Secure AI Framework (SAIF)?

- A. To promote innovation and accelerate AI development without considering security risks.
- B. To restrict AI development and deployment due to concerns about potential security risks.
- C. To establish security standards for building and deploying AI responsibly, addressing the unique challenges and threats in the AI landscape.
- D. To focus solely on preventing external attacks on AI systems.



Quiz | Question 03

Question

What is the primary goal of ethical AI development?

- A. To comply with all legal and regulatory requirements.
- B. To maximize AI capabilities regardless of societal impact.
- C. To promote transparency and accountability in AI systems.
- D. To ensure AI systems are used responsibly and do not cause harm.

Quiz | Question 03

Answer

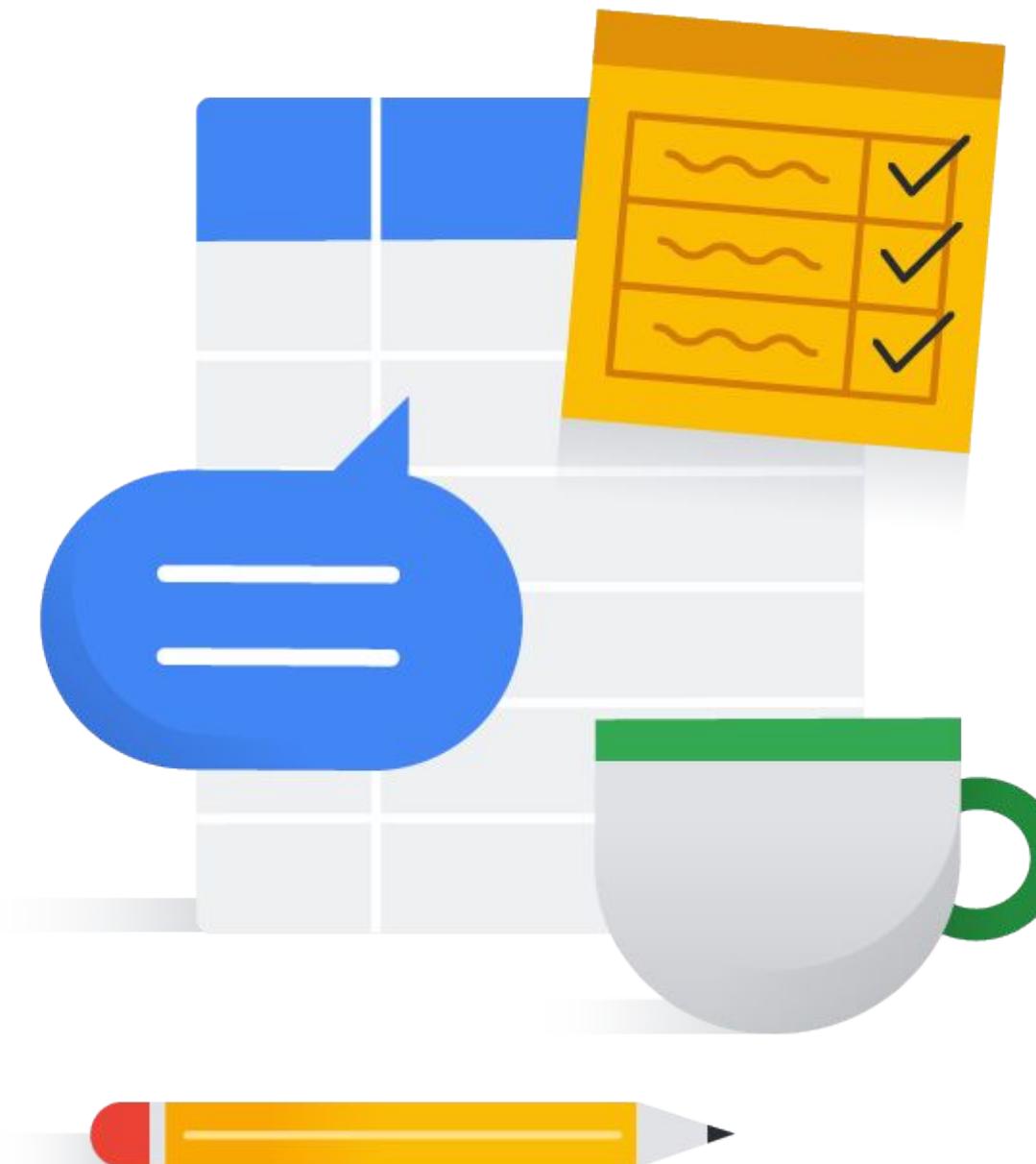
What is the primary goal of ethical AI development?

- A. To comply with all legal and regulatory requirements
- B. To maximize AI capabilities regardless of societal impact
- C. To promote transparency and accountability in AI systems
- D. To ensure AI systems are used responsibly and do not cause harm

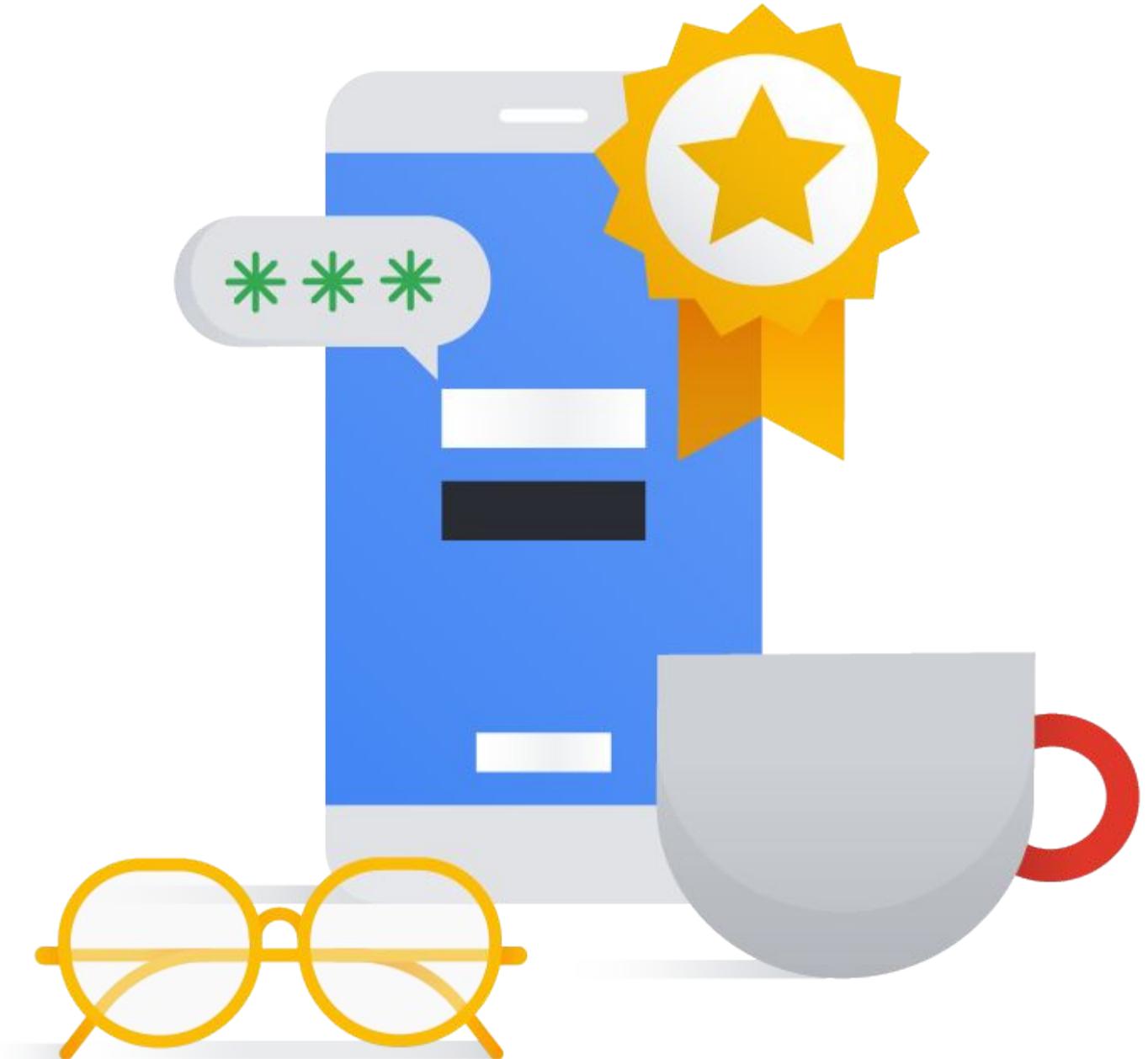


Key takeaways

- AI offers significant benefits, but it also introduces security risks. Google Cloud's SAIF and tools help build secure AI.
- Responsible AI development requires understanding potential issues and mitigating safety, security, and ethical implications throughout the AI lifecycle to benefit society.



Now let's wrap up
with a quiz to check
your knowledge on
Module 02.



Quiz | Question 01

Question

Which of the following is a key aspect of securing the model training phase?

- A. Safeguarding training data and model parameters from unauthorized access.
- B. Optimizing the model for faster processing.
- C. Minimizing the amount of training data used.
- D. Ensuring the model's output is visually appealing.

Quiz | Question 01

Answer

Which of the following is a key aspect of securing the model training phase?

- A. Safeguarding training data and model parameters from unauthorized access.
- B. Optimizing the model for faster processing.
- C. Minimizing the amount of training data used.
- D. Ensuring the model's output is visually appealing.



Quiz | Question 02

Question

What is a potential consequence of using inaccurate or incomplete data in AI training?

- A. It eliminates the need for fine-tuning and further model training.
- B. It results in data storage inefficiency.
- C. It introduces biased outcomes and unfair results.
- D. It results in an inefficient user interface.

Quiz | Question 02

Answer

What is a potential consequence of using inaccurate or incomplete data in AI training?

- A. It eliminates the need for fine-tuning and further model training.
- B. It results in data storage inefficiency.
- C. It introduces biased outcomes and unfair results.
- D. It results in an inefficient user interface.



Quiz | Question 03

Question

When is fine-tuning a particularly useful technique for enhancing a foundation model's performance?

- A. When the model needs to be connected to verifiable, real-time data sources to prevent hallucinations in its responses.
- B. When the primary goal is to quickly guide the model towards desired outputs by crafting precise prompts without altering the model's existing knowledge base.
- C. When prompt engineering alone doesn't achieve the desired outcomes, and the model needs to be specialized for specific tasks or output formats using a new, task-specific dataset.
- D. When the model frequently encounters edge cases and needs to be restricted from accessing any external information to avoid errors

Quiz | Question 03

Answer

When is fine-tuning a particularly useful technique for enhancing a foundation model's performance?

- A. When the model needs to be connected to verifiable, real-time data sources to prevent hallucinations in its responses.
- B. When the primary goal is to quickly guide the model towards desired outputs by crafting precise prompts without altering the model's existing knowledge base.
- C. When prompt engineering alone doesn't achieve the desired outcomes, and the model needs to be specialized for specific tasks or output formats using a new, task-specific dataset.
- D. When the model frequently encounters edge cases and needs to be restricted from accessing any external information to avoid errors



Quiz | Question 04

Question

What is a key legal responsibility for organizations developing or deploying AI?

- A. To ensure adherence to data privacy laws, non-discrimination principles, and the specific licensing terms of AI models.
- B. To prioritize rapid innovation over adherence to evolving legal standards for AI.
- C. To assume that once initial legal compliance is met, no further legal review or counsel is necessary.
- D. To consider legal compliance as solely a regulatory hurdle that does not contribute to the trustworthiness of AI systems.

Quiz | Question 04

Answer

What is a key legal responsibility for organizations developing or deploying AI?

- A. To ensure adherence to data privacy laws, non-discrimination principles, and the specific licensing terms of AI models.
- B. To prioritize rapid innovation over adherence to evolving legal standards for AI.
- C. To assume that once initial legal compliance is met, no further legal review or counsel is necessary.
- D. To consider legal compliance as solely a regulatory hurdle that does not contribute to the trustworthiness of AI systems.



Module objectives

01

Define core gen AI concepts.

02

Explain how data types are used in gen AI for business impact.

03

Explain the role of foundation models in gen AI.

04

Describe Google Cloud's strategies for handling LLM limitations.

05

Describe the challenges for responsible and secure AI development and deployment.



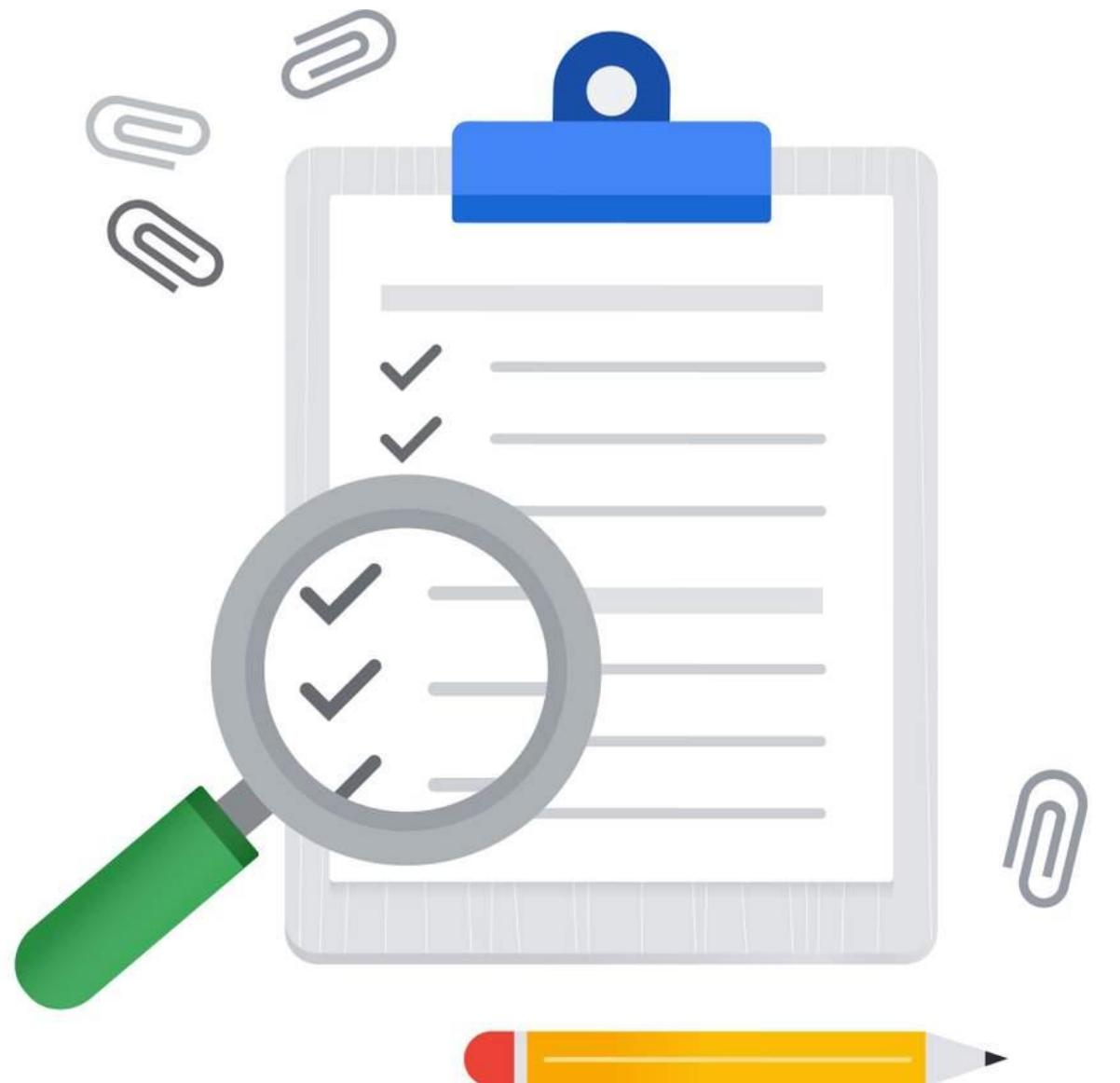
Additional resources

Lesson 02

- [Summary of techniques to overcome limitations of foundation and pre-trained models](#)

Lesson 03

- [Google's Secure AI Framework \(SAIF\)](#)



Module 02
complete!