Generative AI on Vertex AI

In this module, you learn to ...

01 Differentiate between machine learning in general and generative AI

02 Automate ML tasks using Vertex AI on Google Cloud

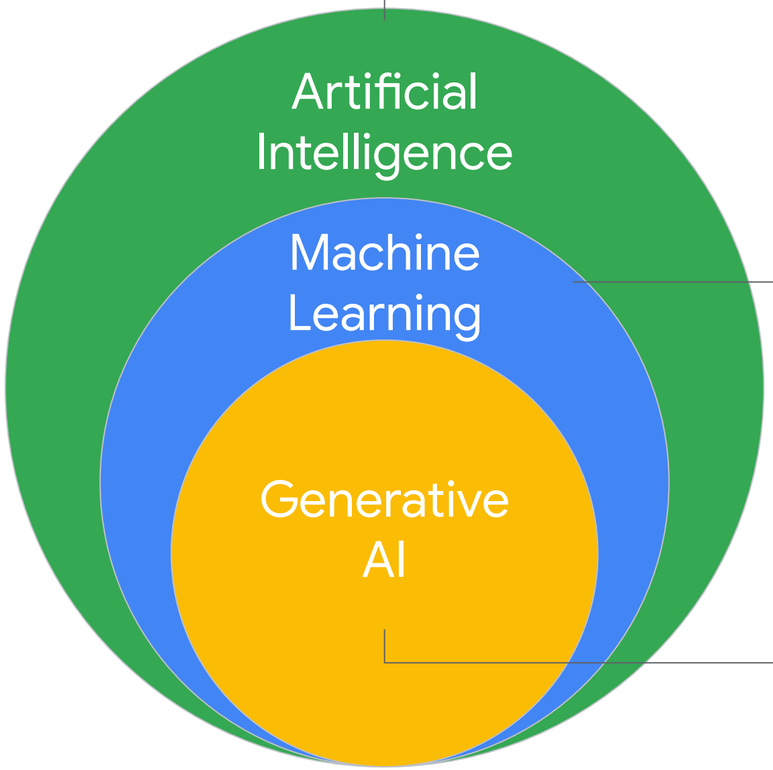
03 Choose from the available Generative AI options on Google Cloud

04 Explore text and code Chat AI use cases

Topics

01	What is Generative AI
02	Vertex AI on Google Cloud
03	Generative AI Options on Google Cloud
04	Introduction to the Course Use Case

Machine Learning is a type of AI, and Generative AI (GenAI) is a type of machine learning



Class of problems we can solve when **computers think/act like humans**

Make predictions based on relationships inferred from historical data

Generate a wide variety of content using complex models trained on enormous datasets

Machine learning allows computers to learn without explicit programming

- In traditional programming, the programmer must define the steps that the computer should take to perform a task
- In machine learning, algorithms are trained to make predictions using historical data
  - Computers iterate over the algorithm making adjustments to find the best solution

Machine learning use cases include:

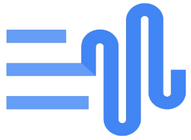
- Image recognition
- Sentiment analysis
- Speech recognition
- Fraud detection
- Customer segmentation
- Recommendation systems

Google has been a pioneer in machine learning for many years

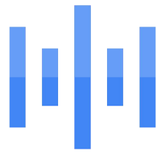
Cloud Vision API

Cloud Translation API

Natural Language API



Text to Speech

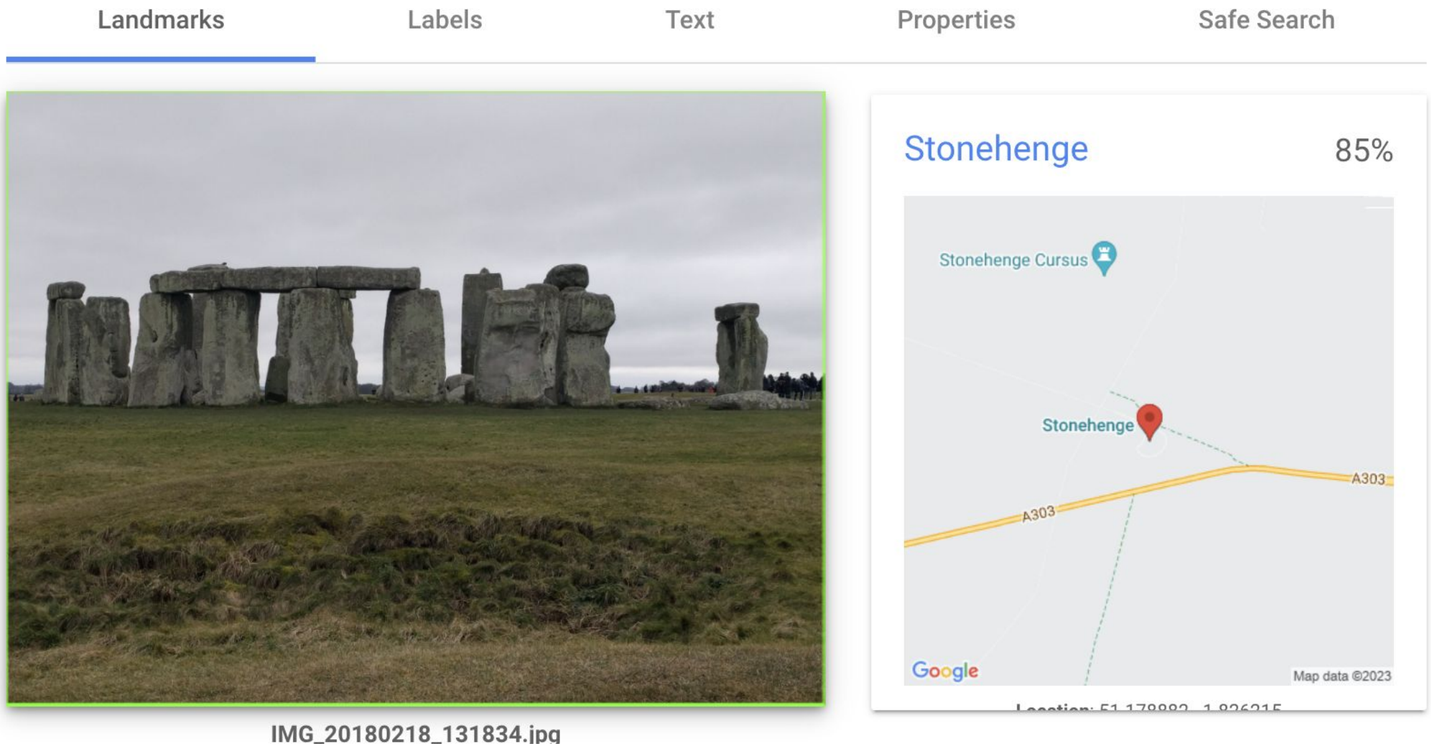


Speech to Text

Video Intelligence API

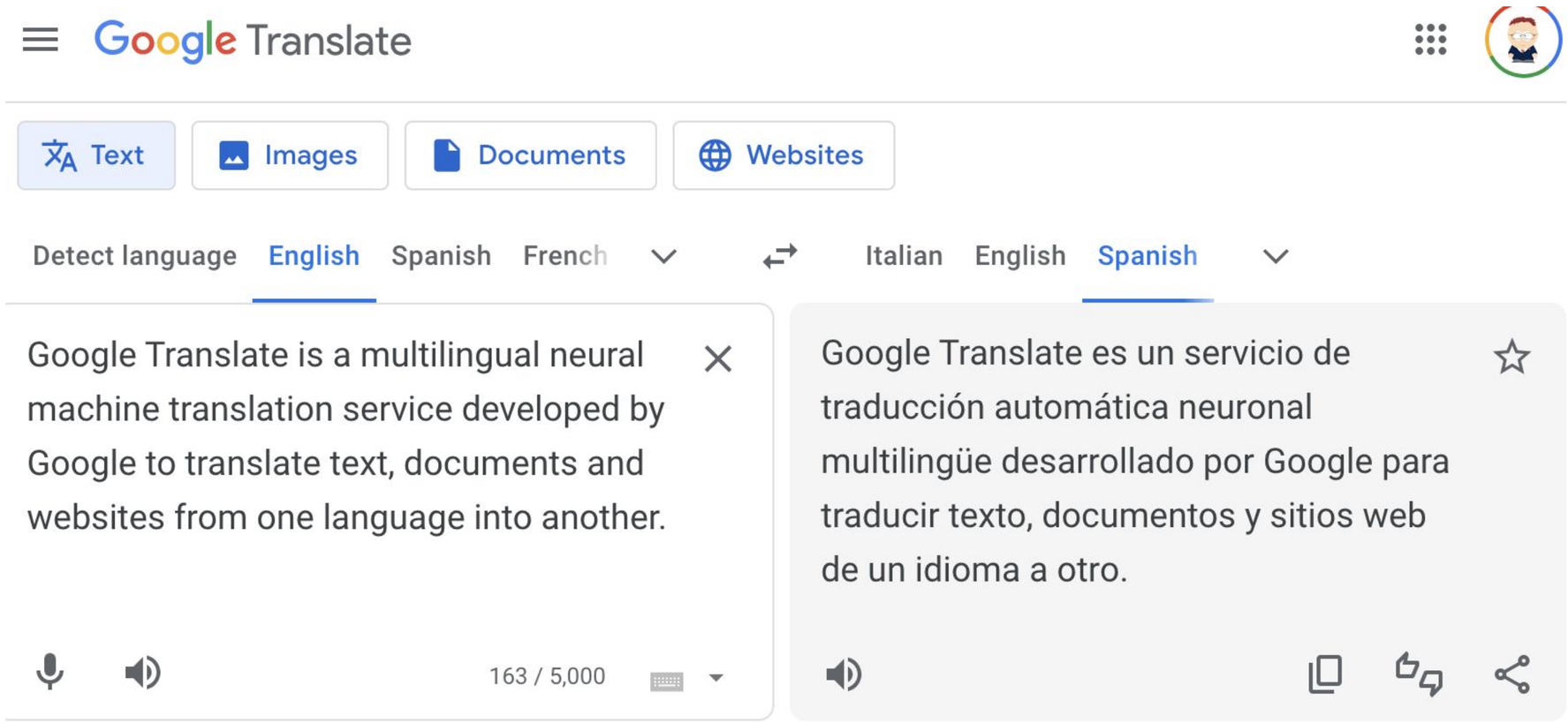
Google Vision API tells you what's in a picture

- Detects objects and landmarks
- Labels photos
- Recognizes text
- Detects inappropriate content



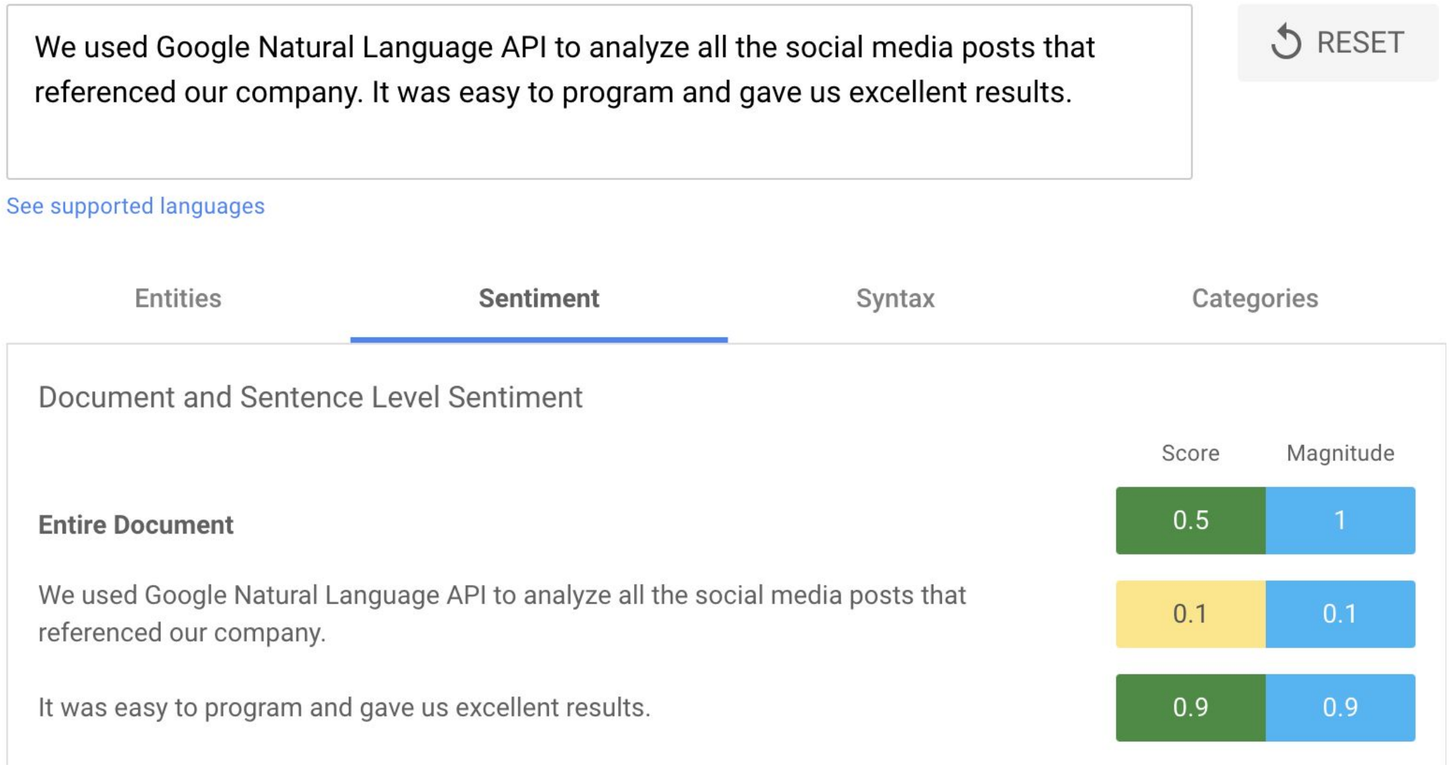
https://cloud.google.com/vision/docs/drag-and-drop

You can use Google ML to translate between different languages



https://translate.google.com/

Natural Language API can analyze text, extract entities, and perform sentiment analysis



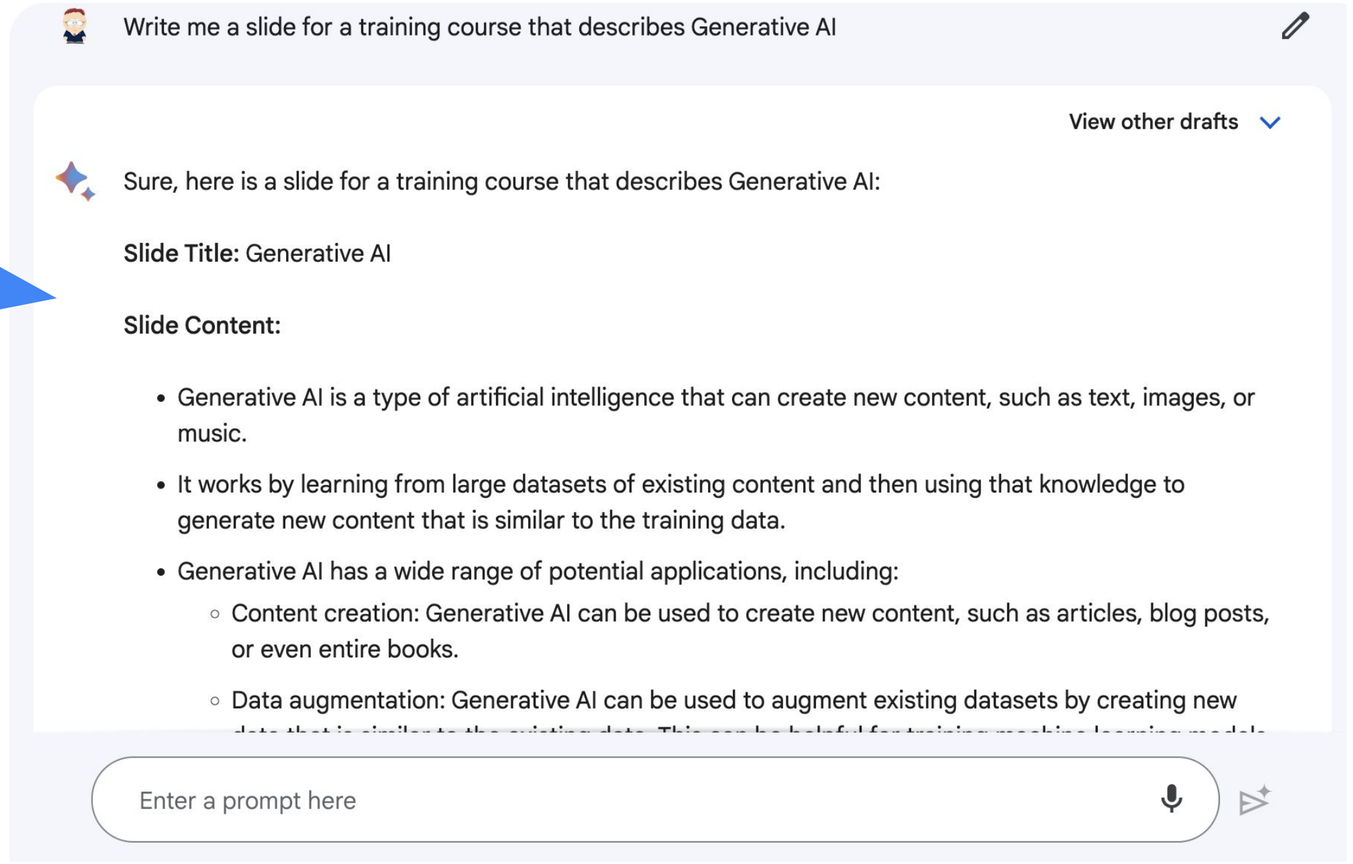
https://cloud.google.com/natural-language/

Generative AI

- Generative AI is a type of artificial intelligence that can create new content, such as text, images, or music
- It works by learning from large datasets of existing content and then using that knowledge to generate new content that is similar to the training data
- Generative AI has a wide range of potential applications, including:
  - Content creation: Generative AI can be used to create new content, such as articles, blog posts, or even entire books.
  - Data augmentation: Generative AI can be used to augment existing datasets by creating new data that is similar to the existing data. This can be helpful for training machine learning models.
  - Creative applications: Generative AI can be used to create new and innovative art, music, and other creative content.

Bard is a Google consumer-oriented generative AI tool for creating custom text-based content

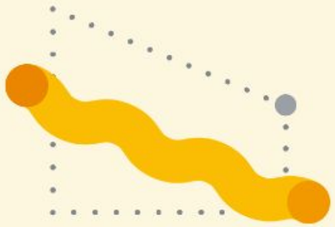
The previous slide!



https://bard.google.com/

There are huge benefits of generative AI

Increase in efficiency & productivity



Reduce costs for your organization

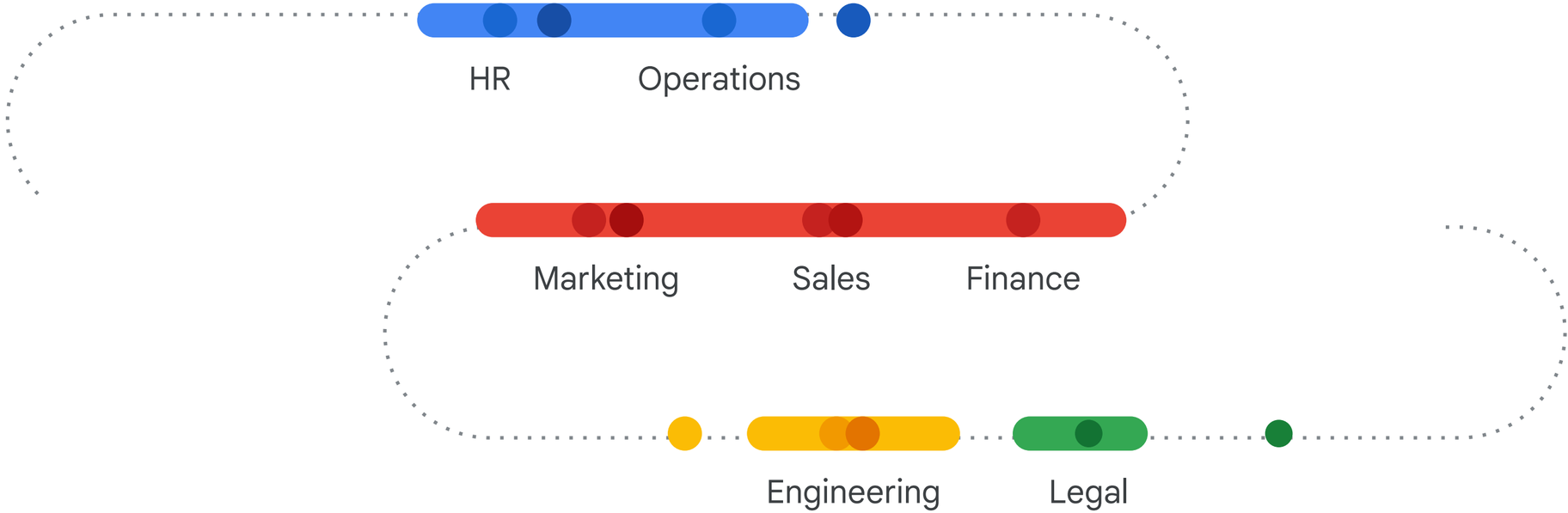
Automate monotonous tasks

There are also challenges with generative AI

- Can be difficult to control the quality of generated content
- Can be difficult to ensure that generated content is accurate
  - Untrue statements can be presented in a confident manner
  - These are known as hallucinations in generative AI terms
- Can be difficult to ensure that generated content is not offensive or harmful

There are many potential use-cases for generative AI across all aspects of an organization

- Content creation
- Marketing and advertising
- Customer service
- Education and research
- Interactive chat
- Code completion
- Many more...

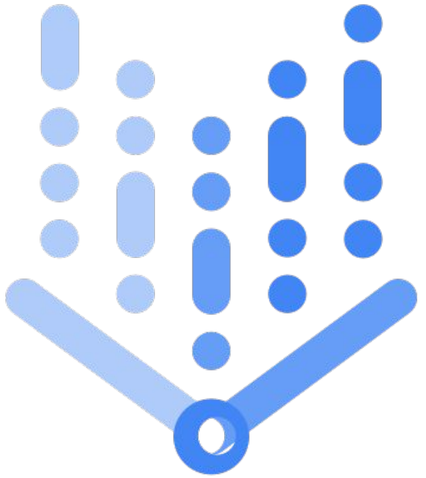


Topics

01	What is Generative AI
02	Vertex AI on Google Cloud
03	Generative AI Options on Google Cloud
04	Introduction to the Course Use Case

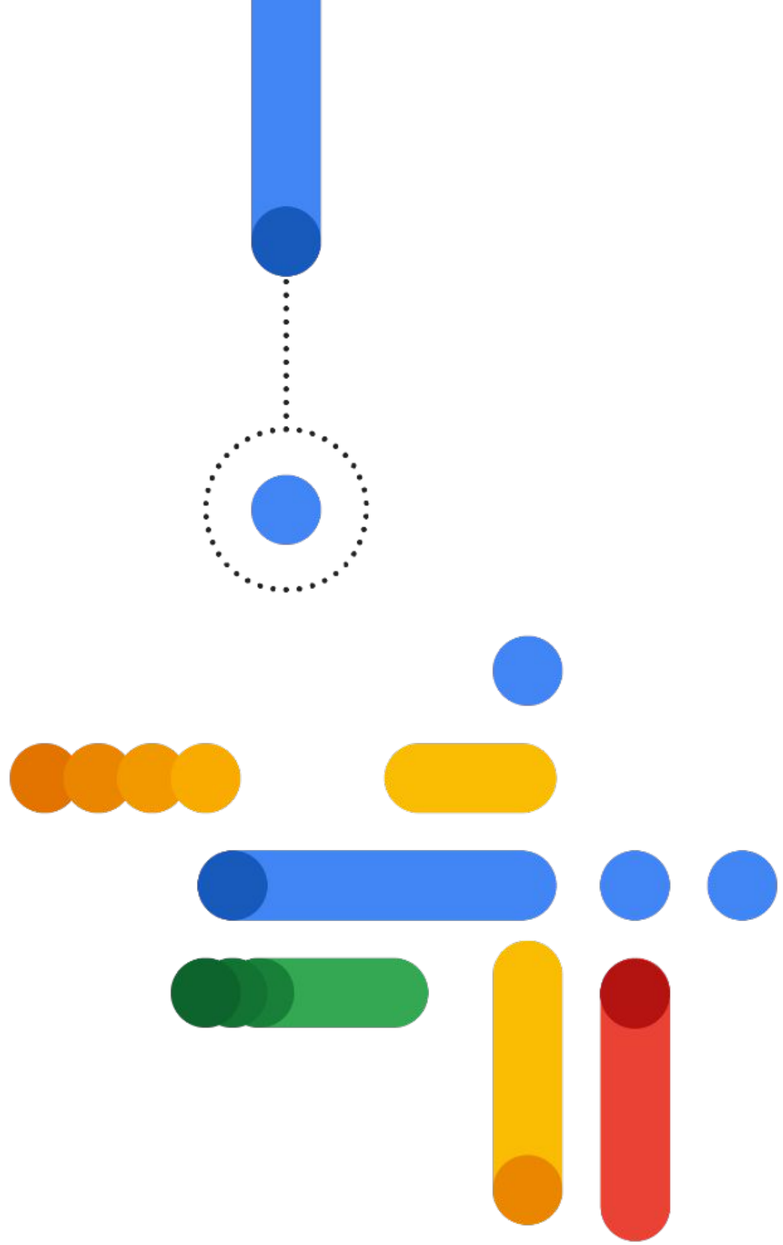
Vertex AI is a machine learning platform that helps you build, deploy, and manage ML models

- Combines data engineering, data science, and ML engineering workflows
- Provides several options for model training and deployment
  - AutoML
  - Custom Training
  - Model Garden
  - Vertex AI Studio
- Uses fully-managed infrastructure that you can customize based on your performance and budget needs
- Supports Python, the Google Cloud Console, the gcloud command line tool, client libraries, and Terraform



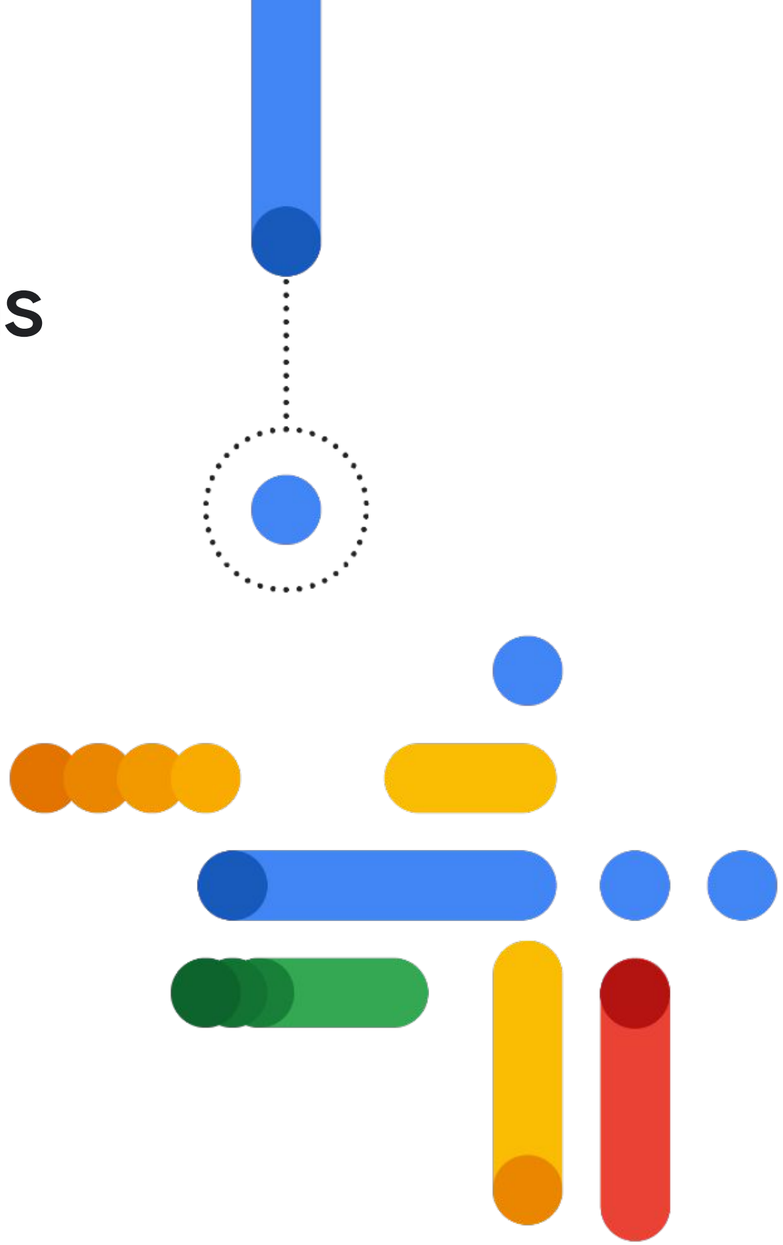
Vertex AI simplifies model training with AutoML

- Supports datasets created with image, text, tabular, and video data
- Zero-code, custom-model training for various use cases
  - Image detection and classification
  - Text classification, entity extraction, and sentiment analysis
  - Linear regression, classification, and forecasting from tabular datasets
  - Video action recognition and object tracking
- Automated deployment of models to service endpoints managed by Google



Model Garden provides a catalog of pre-trained models to build AI applications

- Foundation models are pre-trained multitask large models that can be tuned or customized for specific tasks
- Fine-tunable models are models that you can fine-tune using a custom notebook or pipeline
- Task-specific solutions are pre-built models that are ready to use and can be customized using your own data
- Model garden provides and easy to use interface for searching for models
  - Documentation and code samples are provided



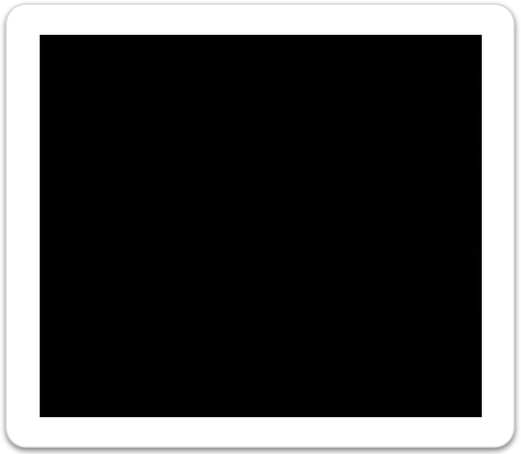
Vertex Model Garden

Choose a model that fits your needs

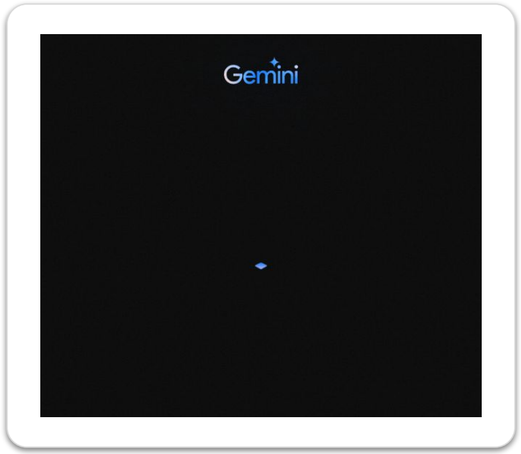
Model Type	Model Description	Model Details
First Party	**Foundation models** Leverage Google's multimodal models across vision, dialog, code generation/completion	• PaLM for Text | PaLM for Chat • Codey for code completion • Imagen for text-to-image • Chirp for speech-to-text • Gemini for multimodal
	Pre-trained APIs Build and deploy AI applications faster with our pre-trained APIs powered by the best Google AI research and technology	• Speech-to-Text • Natural Language Processing • Translation • Vision
Open Source	**Open Source** Access a wide variety of enterprise-ready open source models	• StableDiffusion • ViT • EfficientNet
Third-Party	**3rd Party** Over time Model Garden will support 3rd-party models from partners with foundation models	Coming soon

Source: Google Cloud

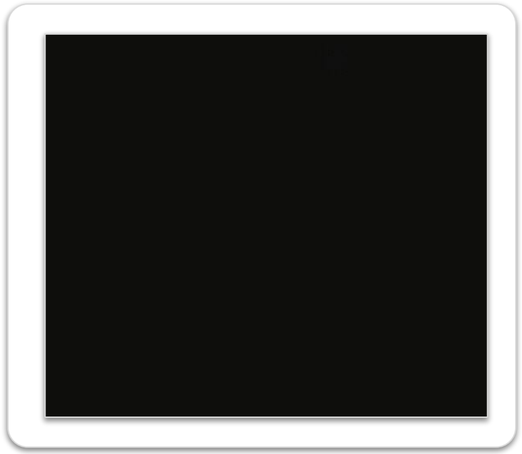
Gemini marks the next phase on our journey to making AI more helpful for everyone



State-of-the-art, natively multimodal reasoning capabilities



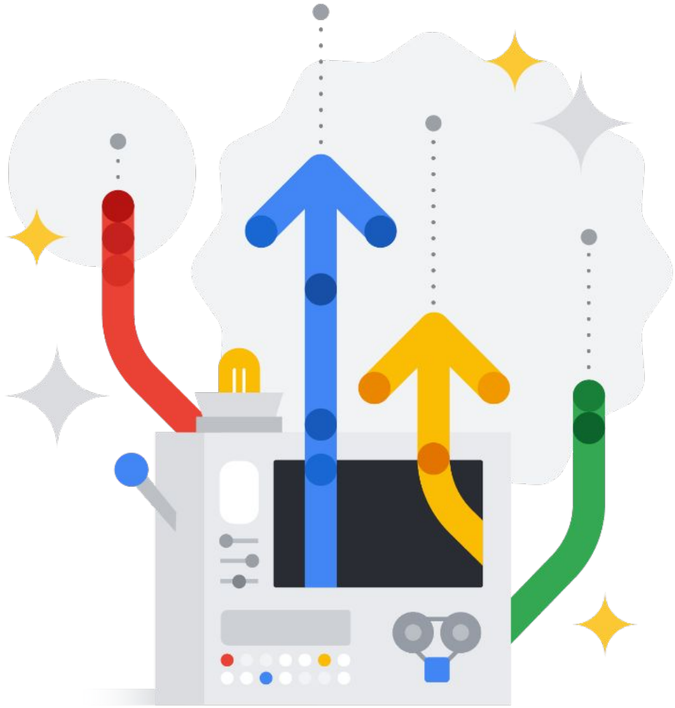
Highly optimized while preserving choice



Built with responsibility and safety at the core

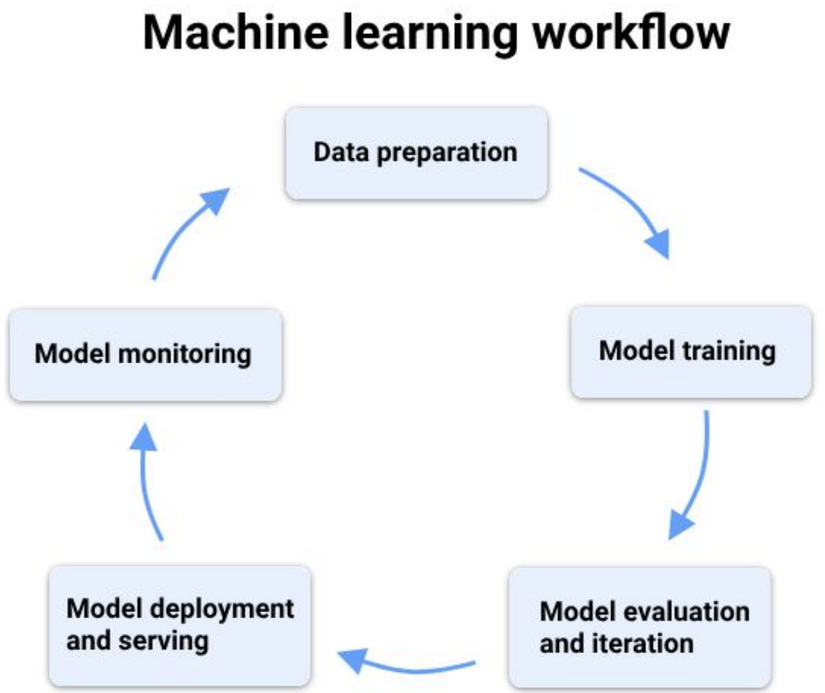
Vertex AI Studio simplifies prompting, tuning, and deploying Google foundational models

- Low-code, customizable solution for building generative AI solutions
- Easy access to Google foundational models
  - PaLM 2 (Pathways Language Model) for text and chat
  - Codey for code generation, completion, and chat
  - Chirp for speech
  - Imagen for text to image generation
  - Gemini for multimodal
  - Options for building question-answer and conversational chat applications



Vertex AI helps with all steps in an AI workflow

- Data preparation: clean and explore data, apply transformations and feature engineering
- Model training: choose a training method, train and tune the model
- Model evaluation and iteration: evaluate the model, make adjustments and iterate
- Model serving: deploy the model to production and get predictions
- Model monitoring: monitor the performance of the deployed model and retrain as needed



Vertex AI features by task

01

Data Preparation

Datasets
Labeling tasks
Feature store
AI Pipelines

02

Model Training

Auto ML
Model Garden
Vertex AI Studio
Online Training
Workbench

03

Model Evaluation

Built-in metrics
Experiments

04

Model Deployment

AI Pipelines
Endpoints
Batch predictions

05

Model Monitoring

Operations Suite
Monitoring
Logging

Topics

01	What is Generative AI
02	Vertex AI on Google Cloud
03	Generative AI Options on Google Cloud
04	Introduction to the Course Use Case

Generative AI use cases with Vertex AI

Language
Writing
Summarization
Ideation
Classification
Sentiment analysis
Extraction
Customer chat

Code
Code generation
Code completion
Code chat
Code conversion

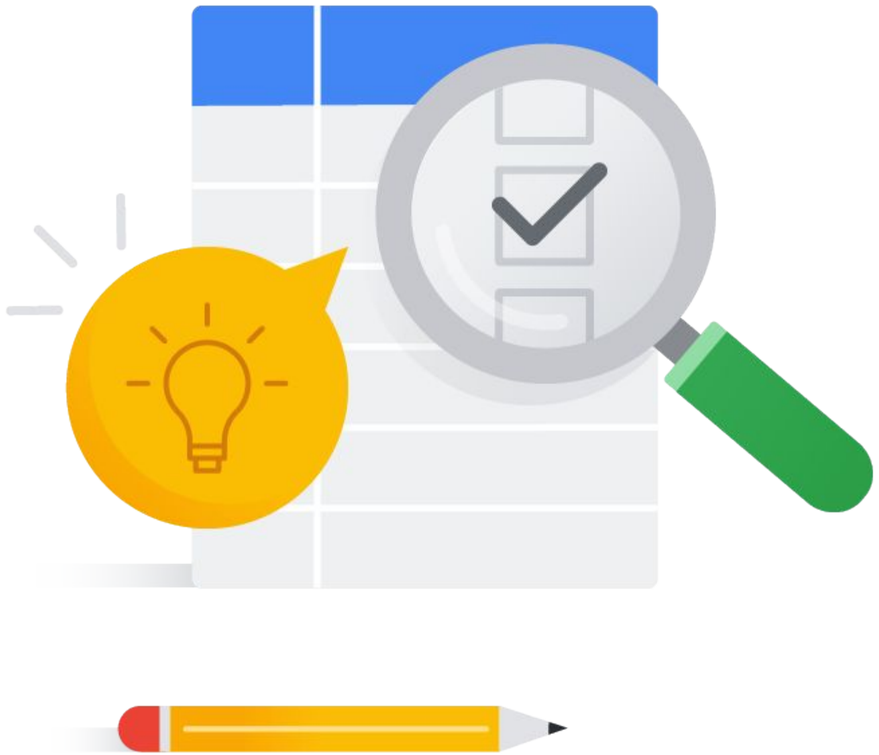
Speech
Speech to text
Text to speech

Group Discussion

10 min

Take a few minutes to come up with some use cases for generative AI that would benefit you or your organization

You will share your ideas with the class

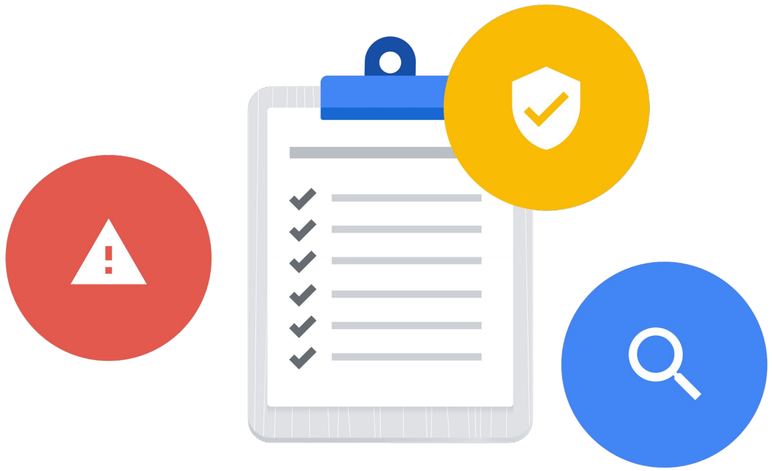


Topics

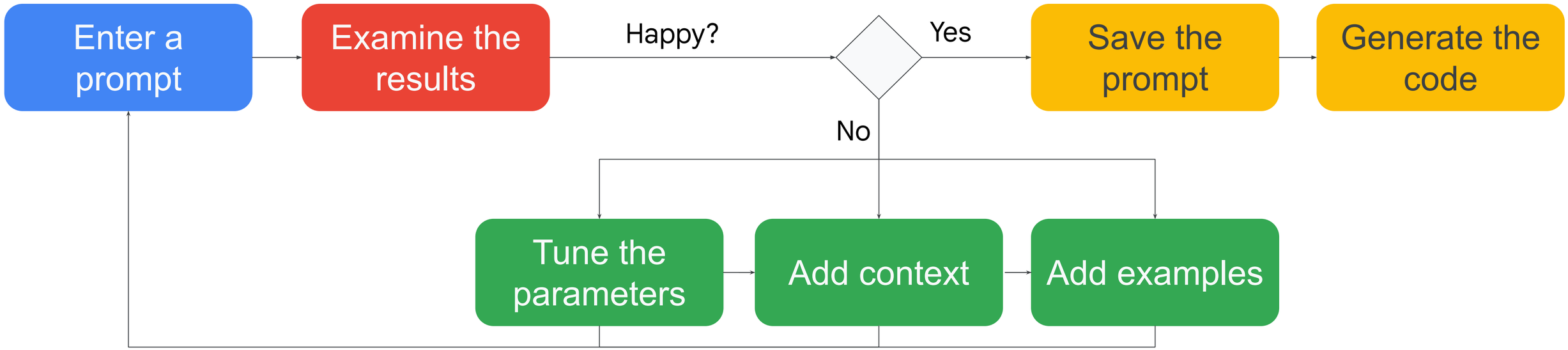
01	What is Generative AI
02	Vertex AI on Google Cloud
03	Generative AI Options on Google Cloud
04	Introduction to the Course Use Case

This course focuses on using generative AI for chat-based text and code use cases

- Q&A
- Code generation
- Content creation
- Ideation
- Summarization
- Classification
- etc.



You will learn the process of designing, tuning, and deploying prompts to generate content for each use case



In this module, you learned to ...

01 Differentiate between machine learning in general and generative AI

02 Automate ML tasks using Vertex AI on Google Cloud

03 Choose from the available Generative AI options on Google Cloud

04 Explore text and code Chat AI use cases

Questions and answers

Quiz question

What is a type of artificial intelligence that allows computers to learn based on data without having to be explicitly programmed?

A: Generative AI

B: Machine learning

C: Linear regression

D: Linear classification

Quiz question

What is a type of artificial intelligence that allows computers to learn based on data without having to be explicitly programmed?

A: Generative AI

B: Machine learning

C: Linear regression

D: Linear classification

Quiz question

What is a type of artificial intelligence that can create new content, such as text, images, or music?

A: Generative AI

B: Machine learning

C: Linear regression

D: Linear classification

Quiz question

What is a type of artificial intelligence that can create new content, such as text, images, or music?

A: Generative AI

B: Machine learning

C: Linear regression

D: Linear classification

Quiz question

When generative AI systems just make something up that isn’t true, it is called what?

A: A lie

B: A hallucination

C: A bug

D: It is not possible for that to happen

Quiz question

When generative AI systems just make something up that isn’t true, it is called what?

A: A lie

B: A hallucination

C: A bug

D: It is not possible for that to happen

Quiz question

List some use cases for generative AI:

Quiz question

List some use cases for generative AI:

Customer Service Chat

Writing content

Ideation

Summarization

Classification

Sentiment analysis

....