Google Cloud

Introduction to Generative AI for Healthcare

Module 1
April 2024

COURSE TITLE

V01
Introduction

Google Cloud

Disclaimer

The information in this course is for educational purposes only, and does not serve as Google's approval for use of its generative AI service for any diagnostic or therapeutic purpose, or for direct patient care.

COURSE TITLE

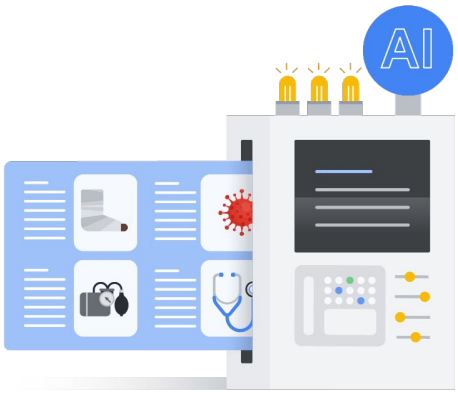
In the world today,

doctors,

researchers,

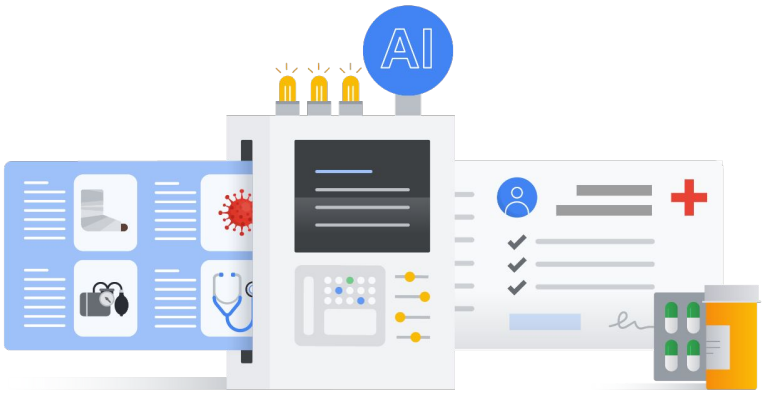
and other health professionals have a powerful tool at their disposal:

generative AI.



Medical data

With the ability to analyze vast amounts of medical data, generative AI can spot patterns invisible to the human eye, predict patient outcomes, and even design personalized medications.



Medical data

Personalized medications

With the ability to analyze vast amounts of medical data, generative AI can spot patterns invisible to the human eye, predict patient outcomes, and even design personalized medications.

In the first section of this course, titled “Introduction to Generative AI for Healthcare”, you’ll learn to:

Define artificial intelligence, machine learning, supervised learning, and unsupervised learning

- Define artificial intelligence, machine learning, supervised learning, and unsupervised learning.

- Define artificial intelligence, machine learning, supervised learning, and unsupervised learning
- Explain how generative AI and generative AI model types function

- Explain how Gen AI and Gen AI model types function.

- Define artificial intelligence, machine learning, supervised learning, and unsupervised learning
- Explain how generative AI and generative AI model types function
- Describe large language models and tune a model for healthcare-specific use cases

- Describe large language models and tune a model for healthcare-specific use cases.

- Define artificial intelligence, machine learning, supervised learning, and unsupervised learning
- Explain how generative AI and generative AI model types function
- Describe large language models and tune a model for healthcare-specific use cases
- Define prompt design and create effective prompts

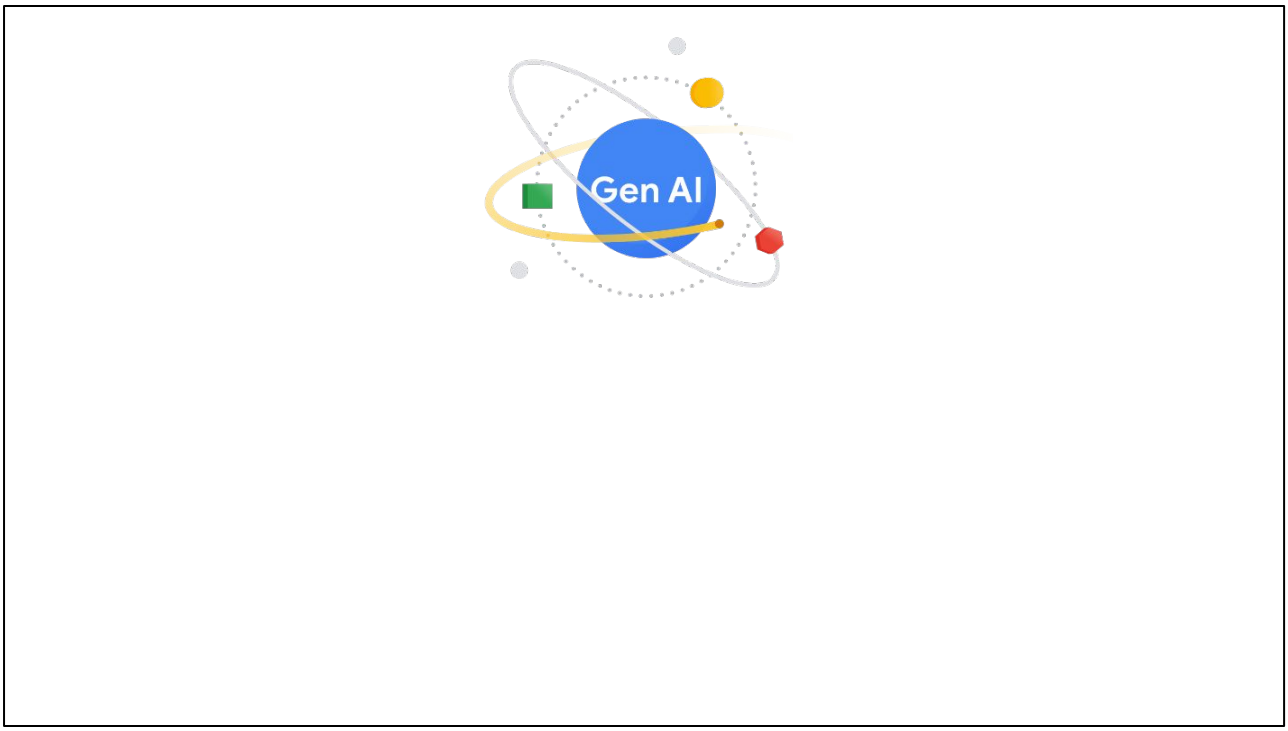
- Define prompt design and create effective prompts.

- Define artificial intelligence, machine learning, supervised learning, and unsupervised learning
- Explain how generative AI and generative AI model types function
- Describe large language models and tune a model for healthcare-specific use cases
- Define prompt design and create effective prompts
- Explore Google Cloud generative AI tools and customize generative AI models

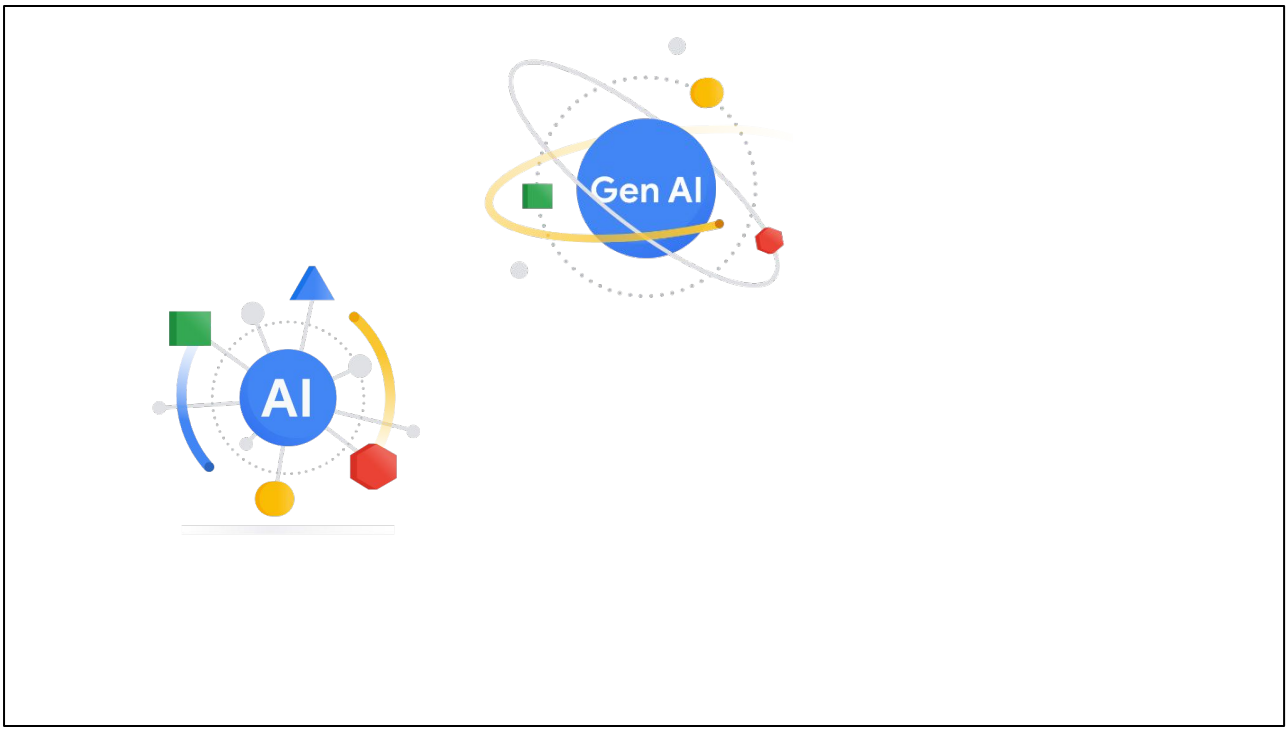
- And explore Google Cloud Gen AI tools and customize Gen AI models.

Deck unlocked

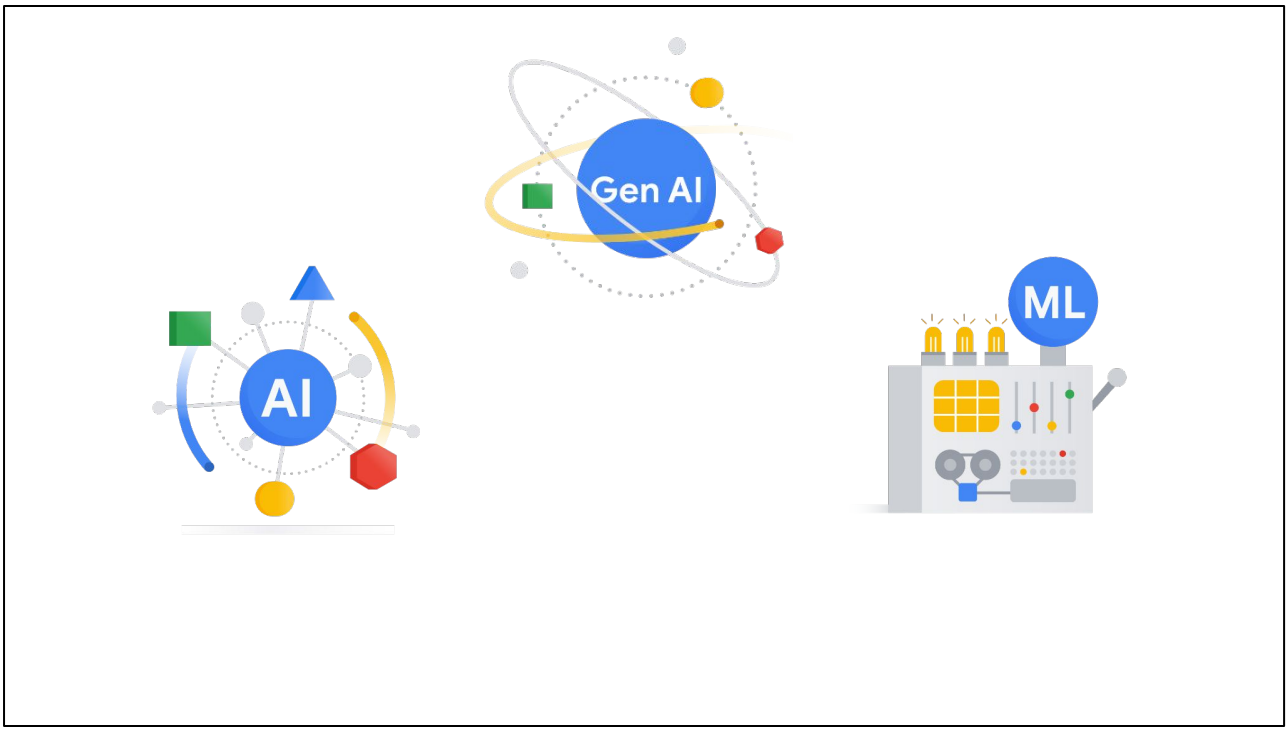
V02
Artificial Intelligence, Machine Learning, and Deep Learning



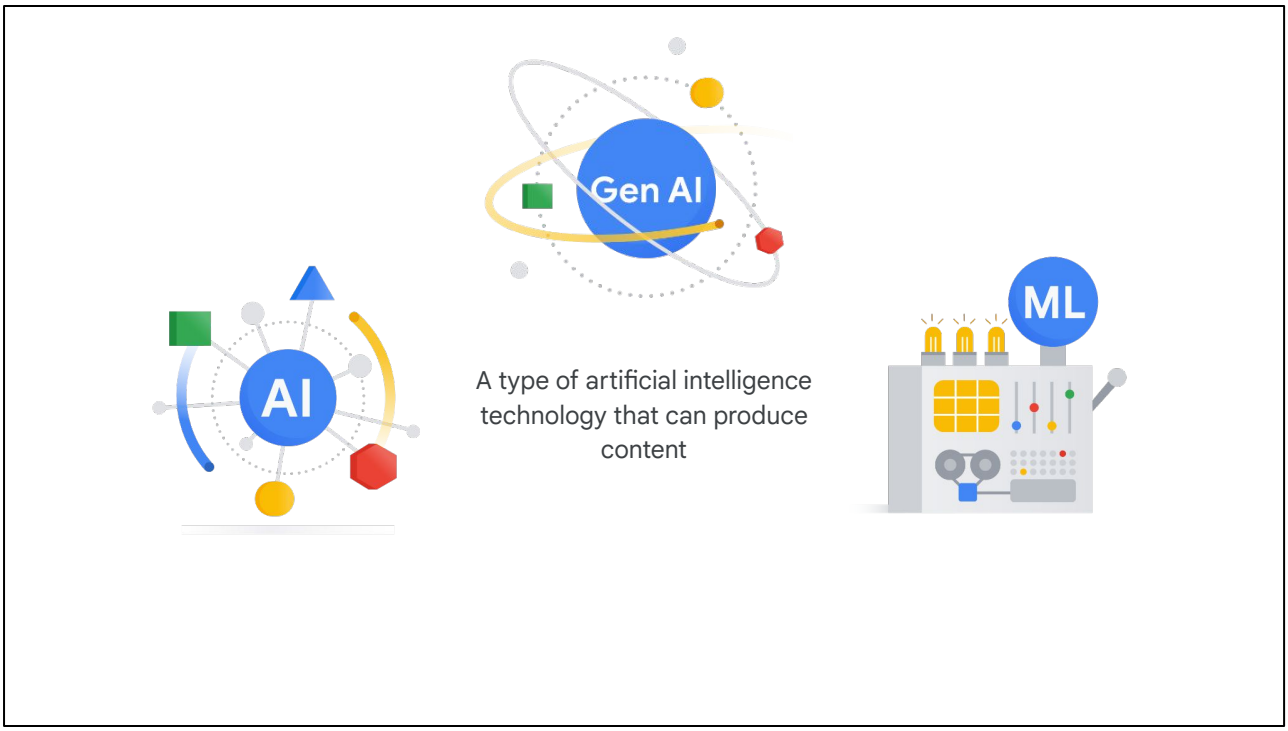
Let’s begin with a high-level explanation of generative AI



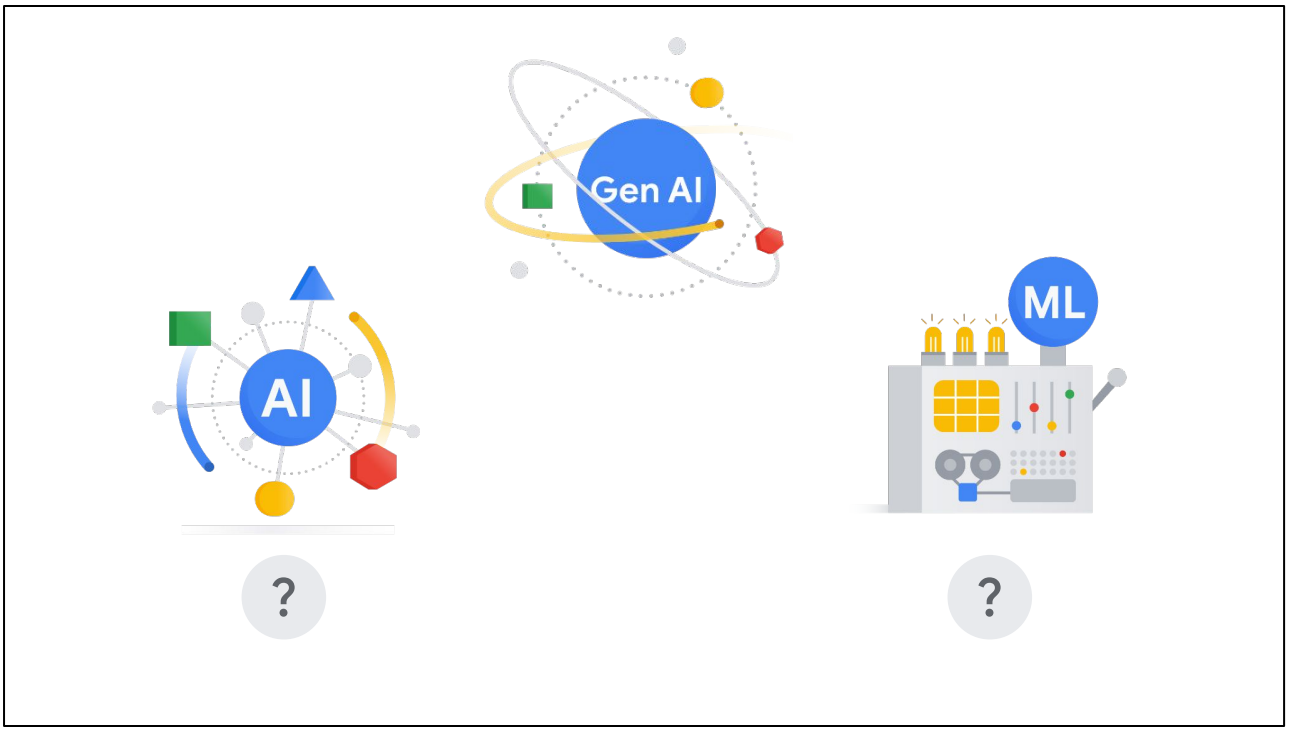
and explore how it fits into artificial intelligence



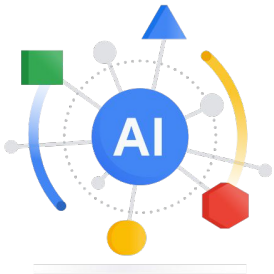
and machine learning. Understanding how generative AI works is essential before exploring healthcare applications, as it provides the foundation for how these complex tools function and generate results.

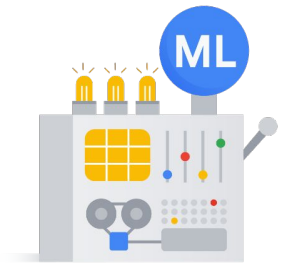


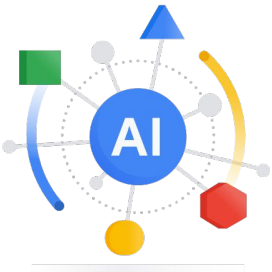
Generative AI is a type of artificial intelligence technology that can produce various types of content, including text, imagery, audio, and synthetic data.



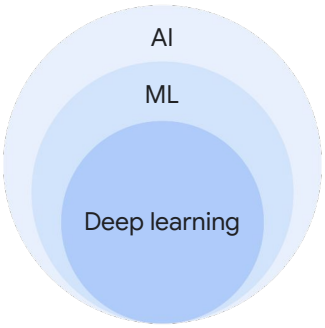
But what is artificial intelligence (or AI), and what's the difference between AI and machine learning?

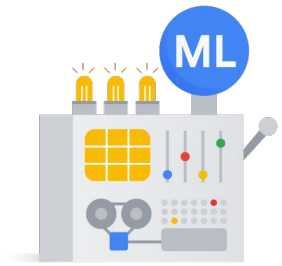

AI


ML

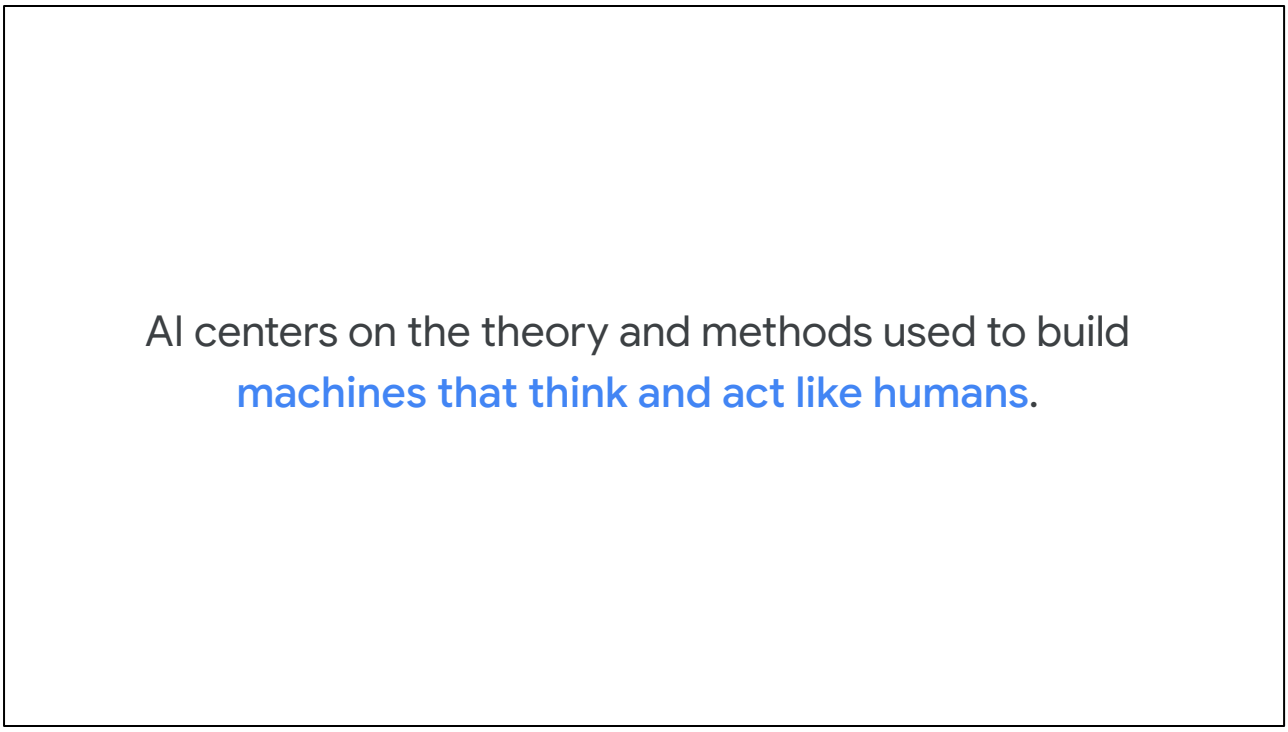


A discipline of computer science that's used to create intelligent agents

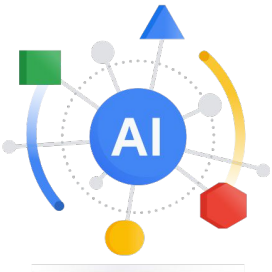




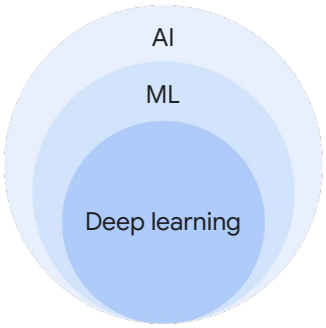
You can think of AI as a discipline, or a branch of computer science that's used to create intelligent agents, which are systems that can reason, learn, and act autonomously.

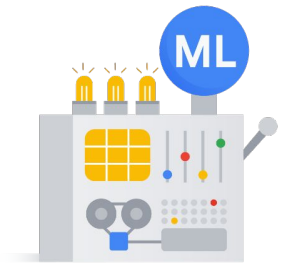


AI centers on the theory and methods used to build machines that think and act like humans.



A discipline of computer science that's used to create intelligent agents

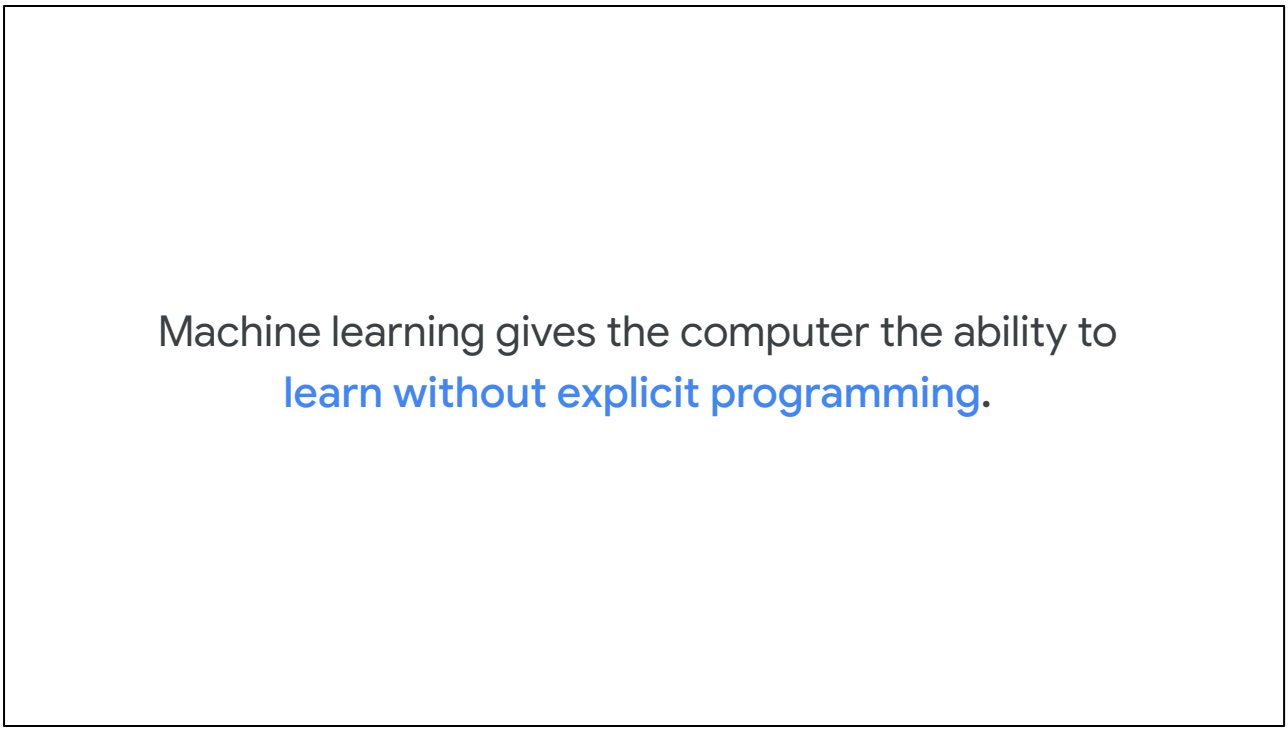




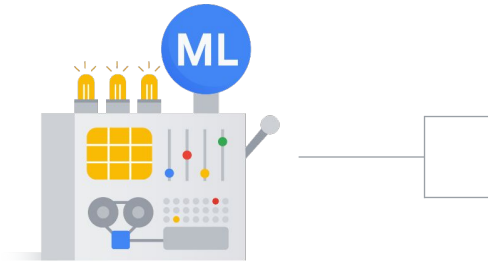
A subfield of AI that's used to train a model from input data

Machine learning is a subfield of AI, and is a program or system that trains a model from input data.

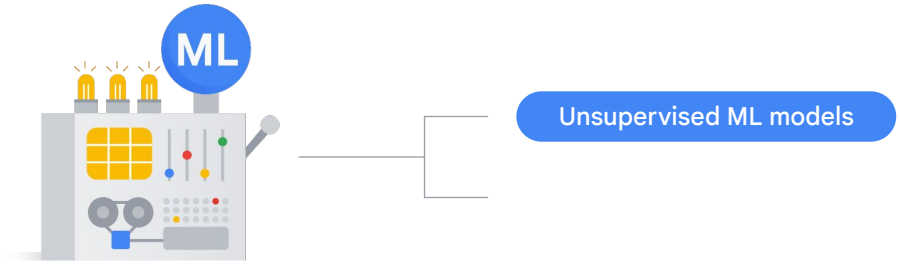
Once the model has been trained, it can then make meaningful predictions on new, unseen data similar to what it was trained on.



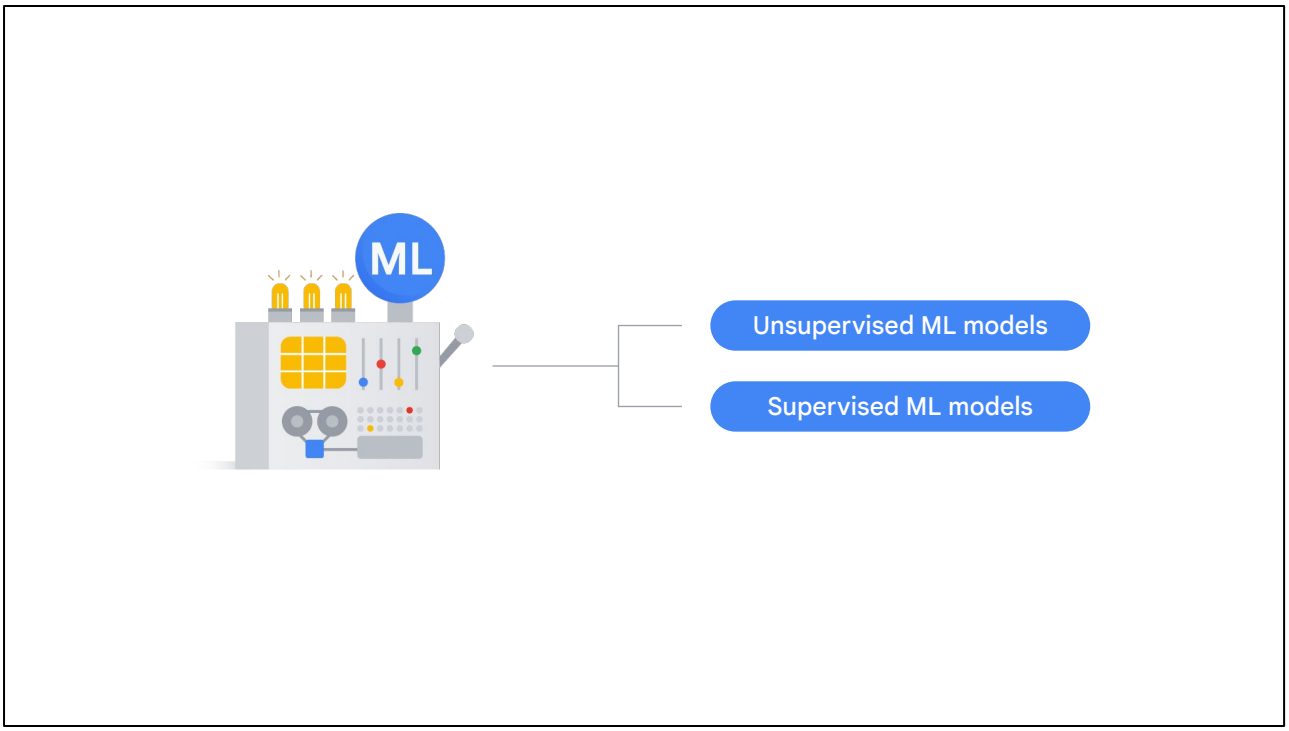
Machine learning gives the computer the ability to learn without explicit programming.



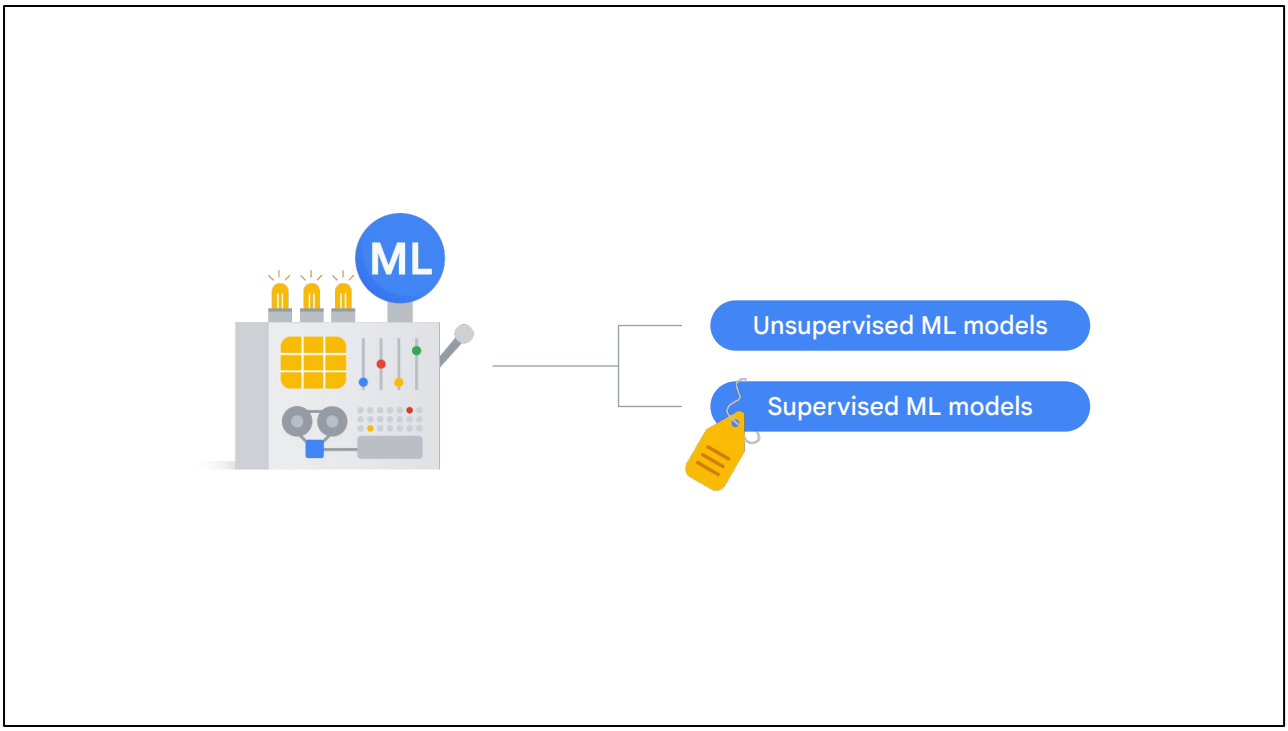
Two of the most common classes of machine learning models are



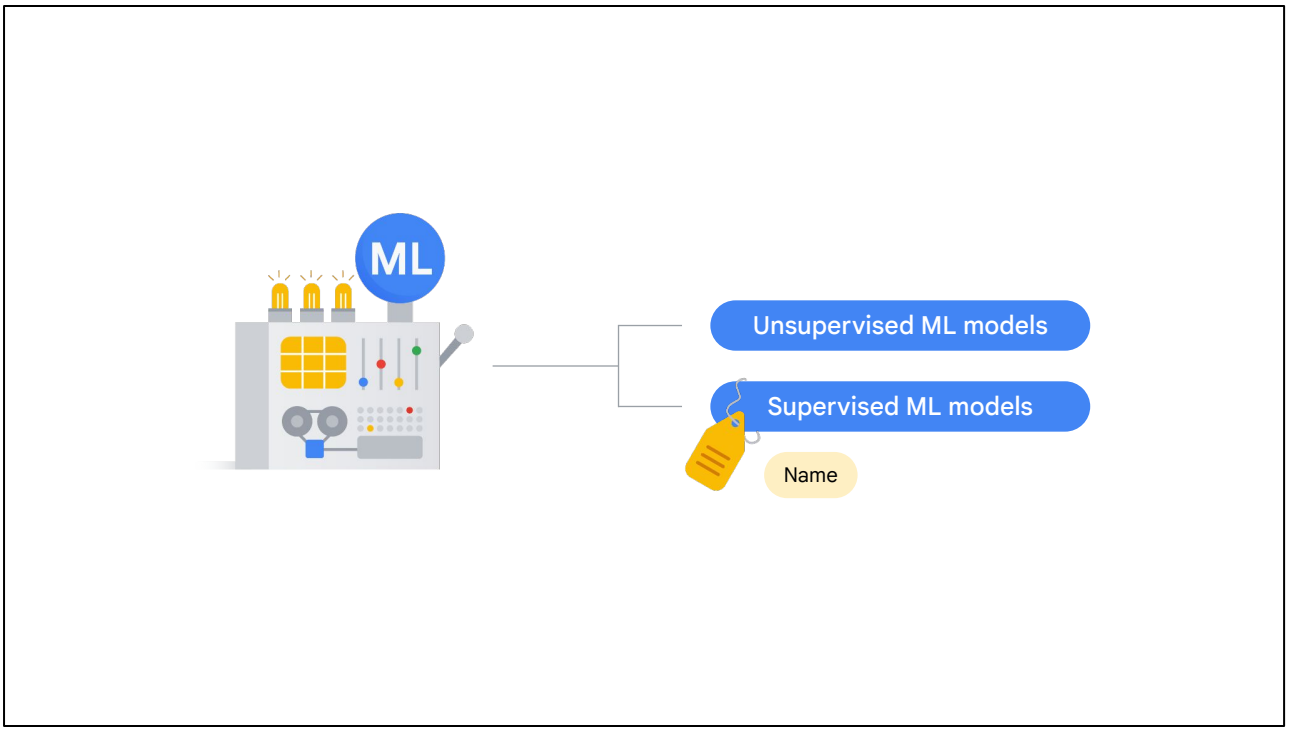
unsupervised



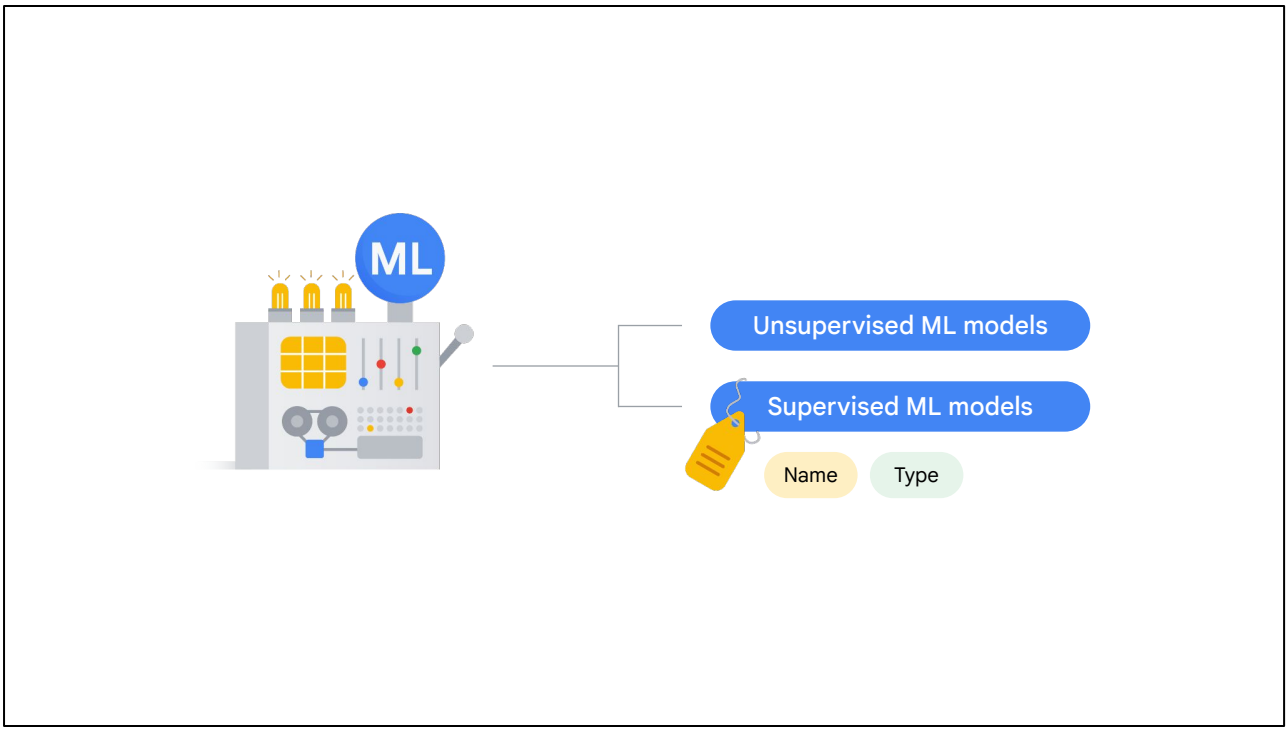
and **supervised**.



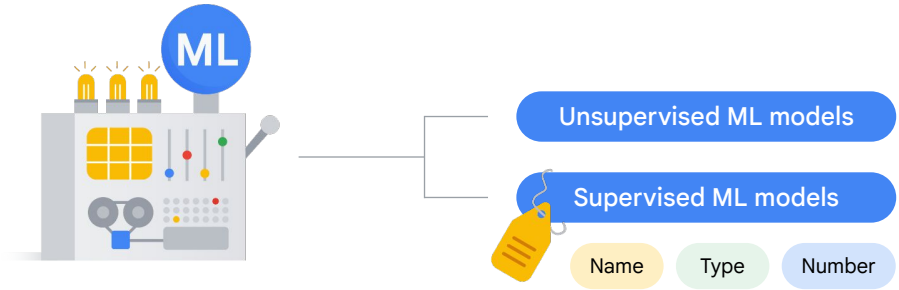
The key difference between the two is that supervised models have labels. Labeled data is data that comes with a tag, like



a name,

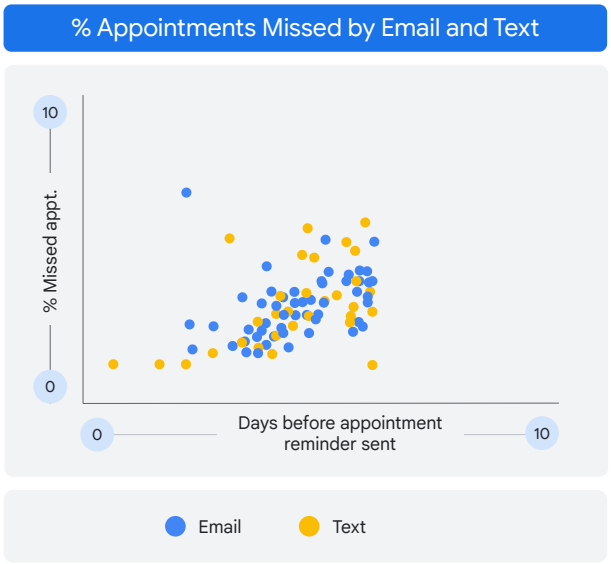


a type,



or a number. Labels are similar to column headers in a spreadsheet. Unlabeled data does not have a tag.

A supervised model learns from past examples to predict future values.

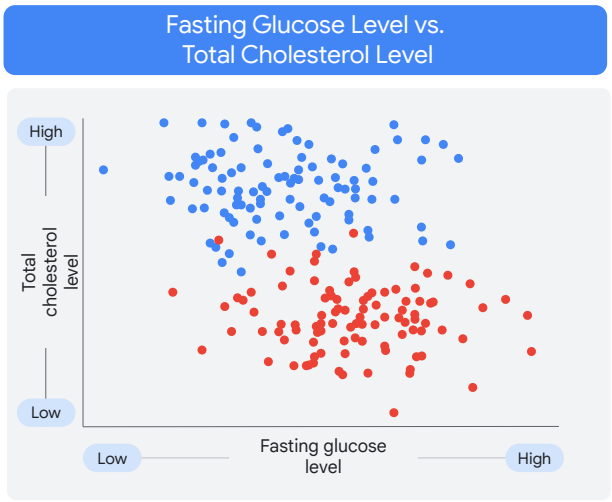


A **supervised** model learns from past examples to predict future values.

Let's say you own a clinic. You have historical data comparing text and email reminders, including the number of days before an appointment the reminder was sent and the corresponding missed appointment rates.

In this example the model predicts the likelihood of future missed appointments based on the timing and type (text or email) of past reminders.

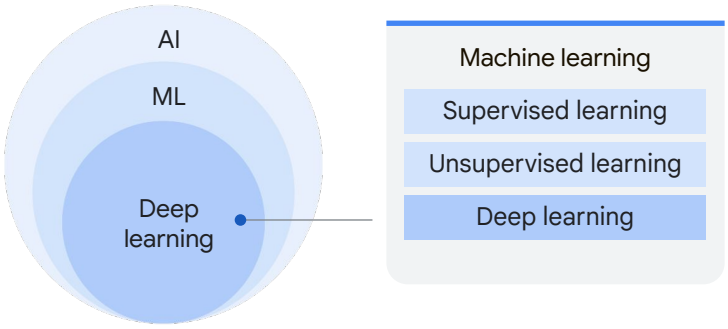
Unsupervised models can aid with discovery and identify patterns in raw data.



Example model - for educational purposes

If you need to analyze blood glucose and cholesterol levels to group patients with similar health profiles, an unsupervised model could identify those clusters without any labels.

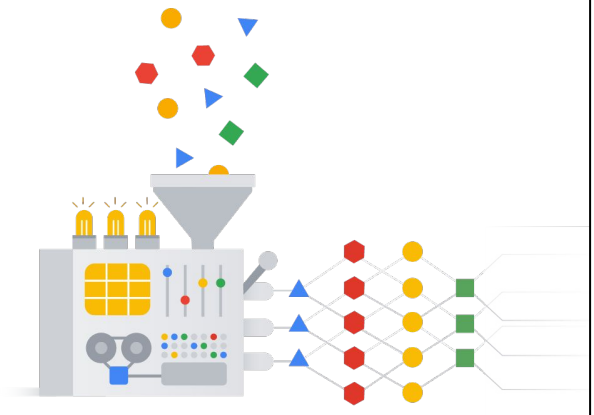
Deep learning is a subset of ML.



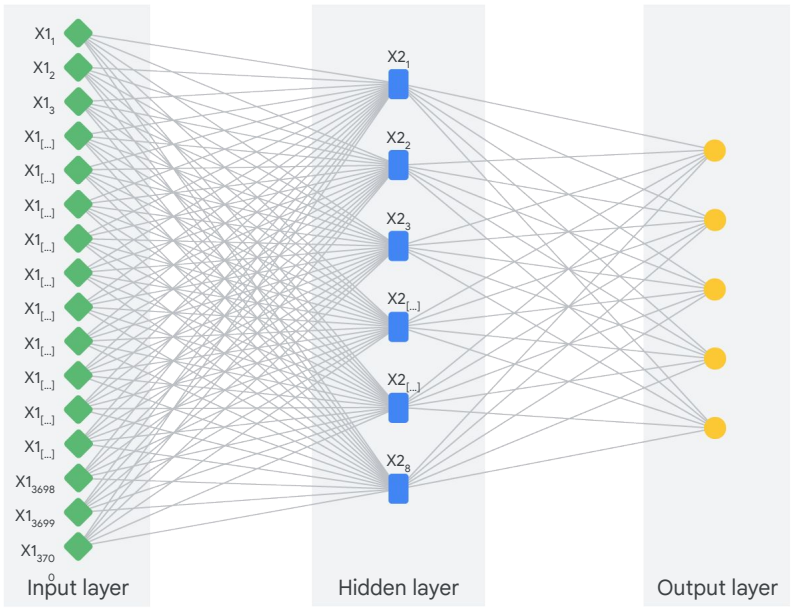
So, we've covered artificial intelligence, machine learning, and supervised and unsupervised learning.

Next, let's briefly explore deep learning.

Deep learning uses artificial neural networks to process more complex patterns than traditional machine learning.



While machine learning is a broad field with many techniques, deep learning is a subset that uses artificial neural networks to process more complex patterns than traditional machine learning models.



Artificial neural networks are inspired by the human brain. They're made up of many interconnected nodes, or neurons, that can learn to perform tasks by processing data and making predictions.

Deep learning models typically have many layers of neurons, which allows them to learn more complex patterns. Neural networks use both labeled and unlabeled data. This is called semi-supervised learning.

Deep learning model types

Predictive

Generative

Now, there are two types of deep learning models: generative and predictive.

Deep learning model types

Predictive

- Typically trained on a dataset of labeled data
- Used to classify or predict labels
- Learns the relationship between the features of the data points and the labels

Generative

A **predictive model** is typically trained on labeled data and is used to classify or predict labels for data points. Then, the model learns the relationship between the features of the data points and the labels.

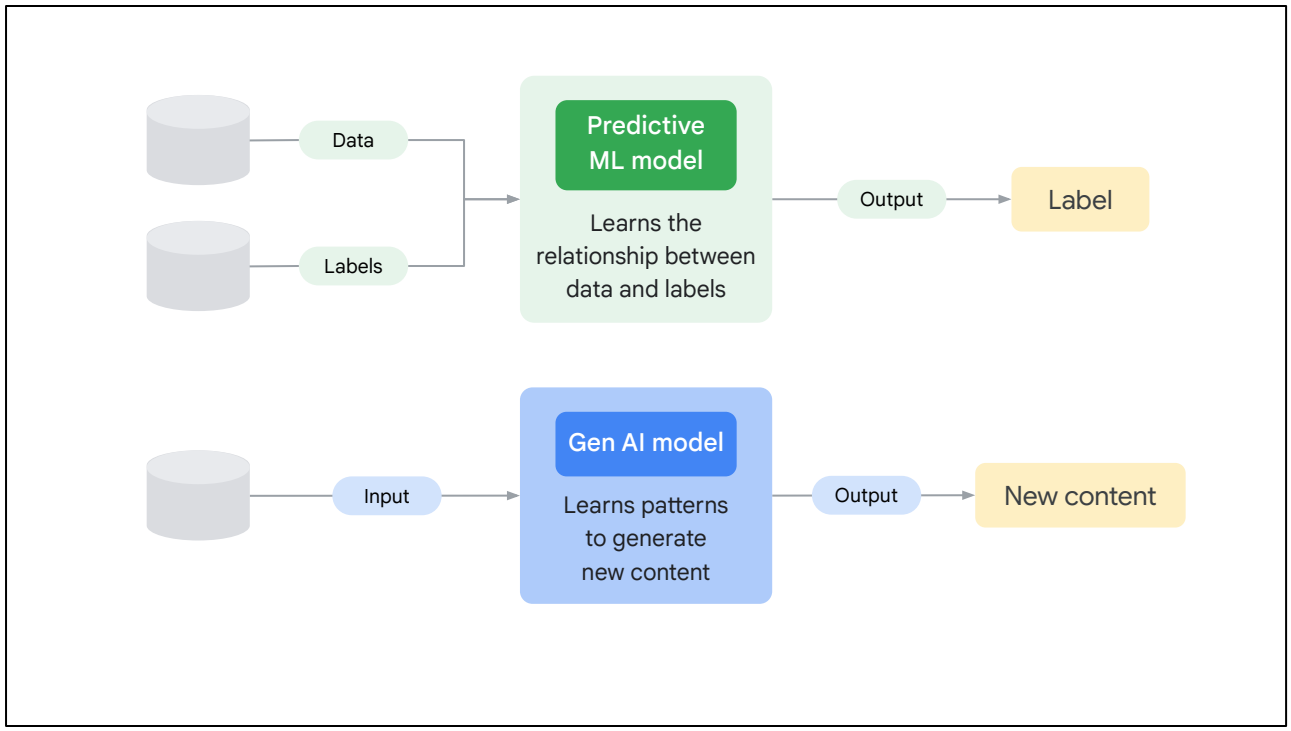
Deep learning model types

Predictive

Generative

- Generates new data that is similar to the data it was trained on
- Learns the probability distribution of existing data

A **generative model** generates new data instances based on a learned probability distribution of existing data.



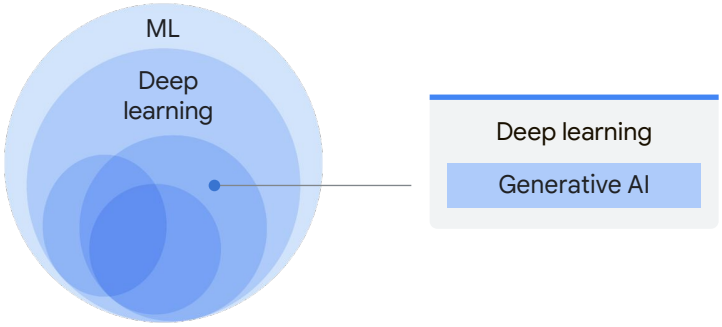
In simple terms, traditional “predictive” machine learning models attempt to learn the relationship between the data and what you want to predict.

A generative AI model attempts to learn patterns so that it can generate new content.

Deck unlocked

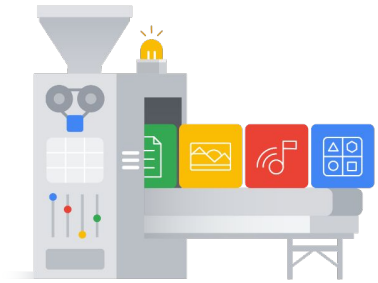
V03
Introduction to Generative AI

Generative AI is a subset of deep learning.



Now, we get to where generative AI fits into the AI discipline.

Gen AI is a subset of deep learning, and it uses artificial neural networks to process labeled and unlabeled data using supervised, unsupervised, and semi-supervised methods.



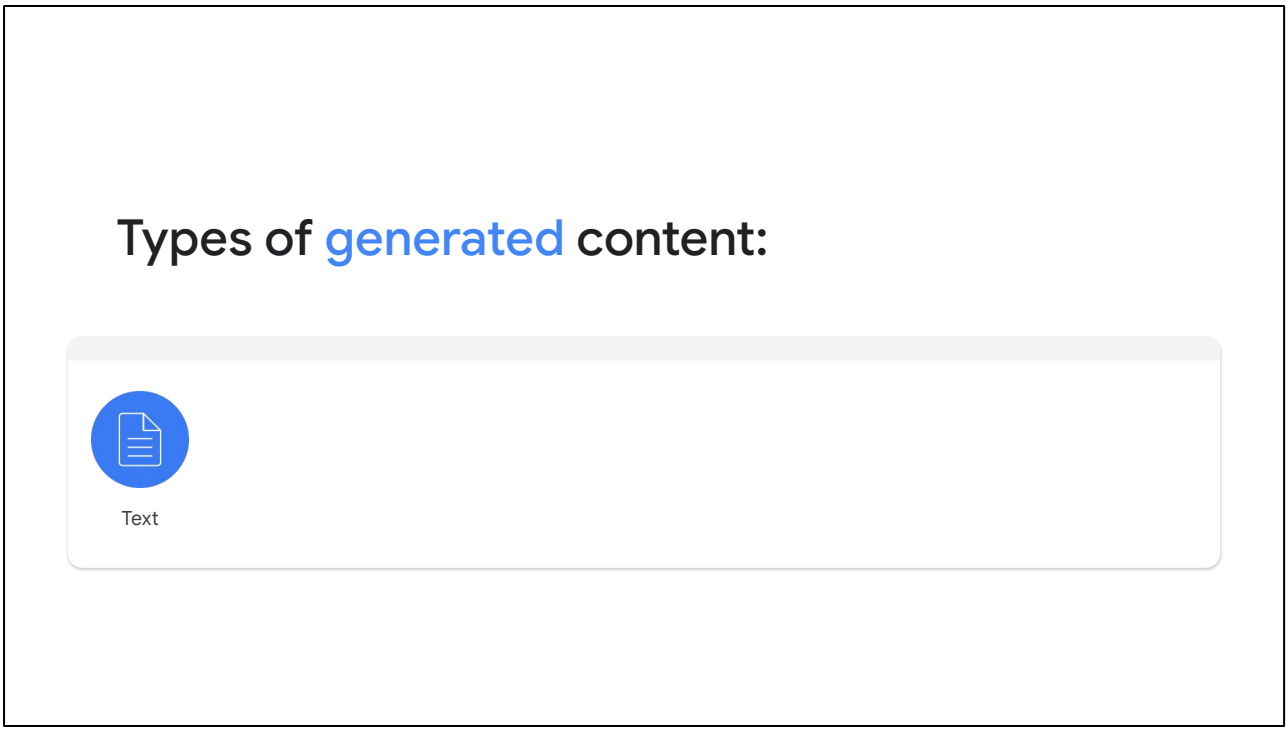
What is generative AI?

- Gen AI is a type of artificial intelligence that creates new content based on what it has learned from existing content.
- The process of learning from existing content is called training and results in the creation of a statistical model.
- When given a set of instructions, Gen AI uses this statistical model to predict an expected response, resulting in newly generated content.

Gen AI creates new content based on what it has learned from existing content.

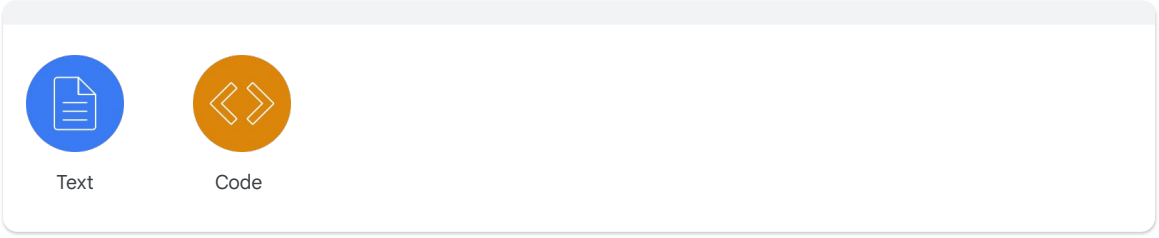
The process of learning from existing content is called “training” and results in the creation of a statistical model.

When given a set of instructions, Gen AI uses this statistical model to predict what an expected response might be, and this results in newly generated content.

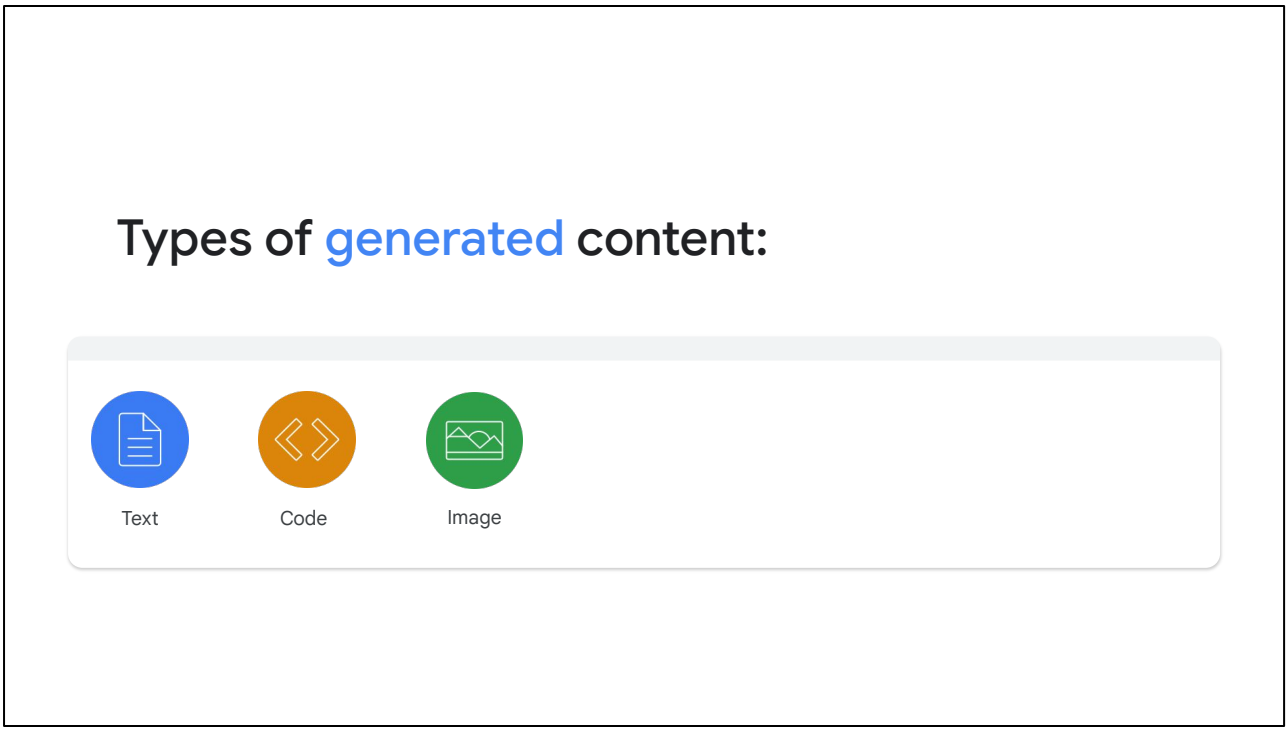


Generated content can be multimodal, including text,

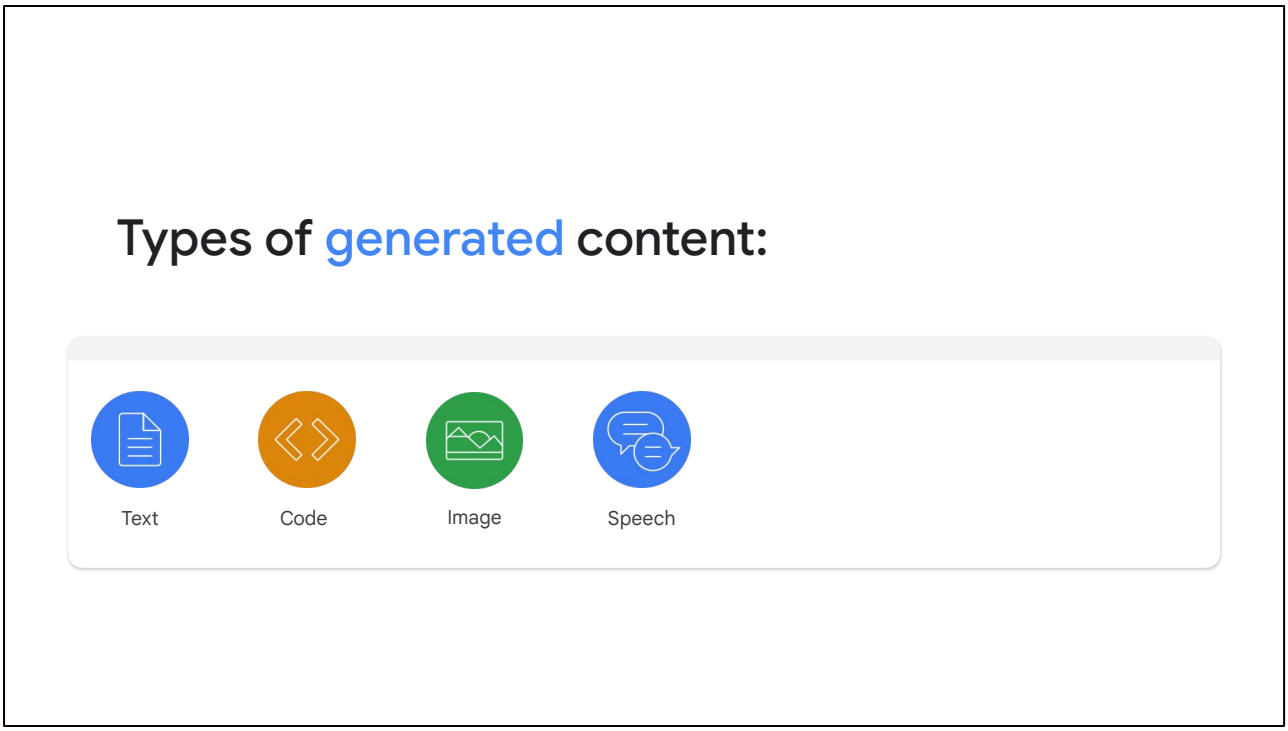
Types of generated content:



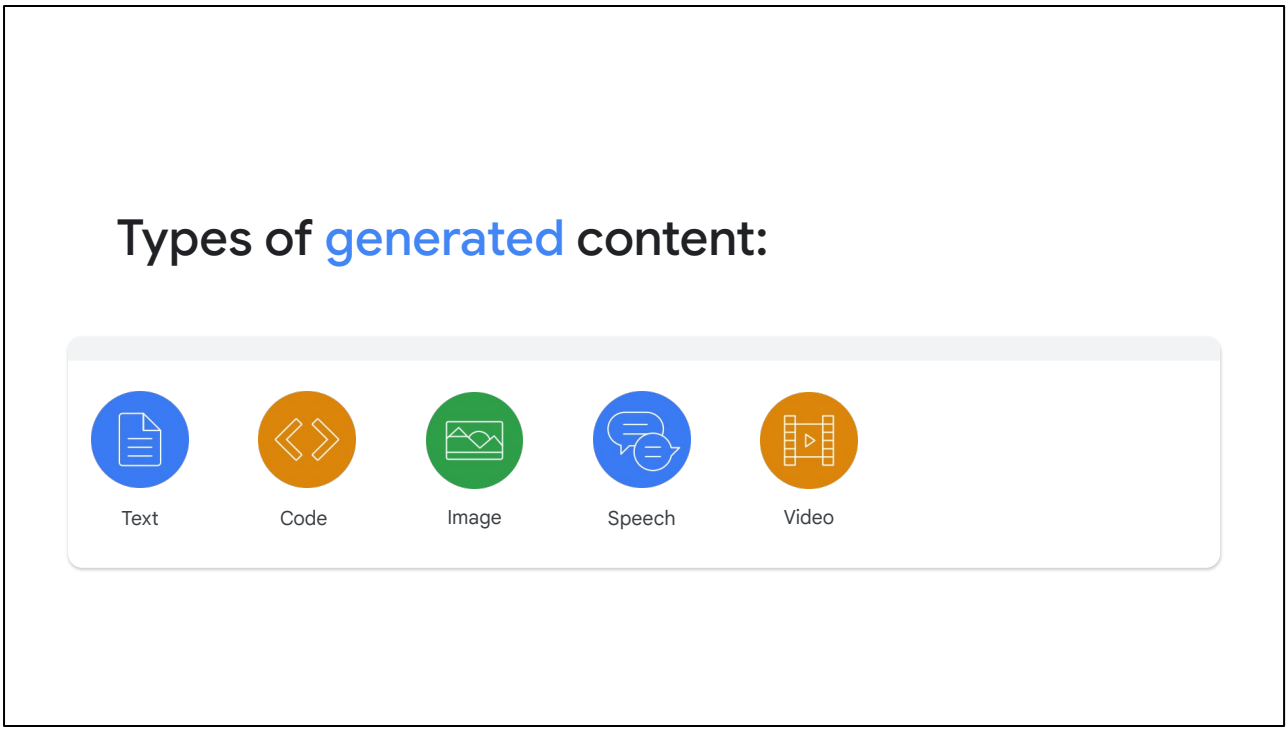
code,



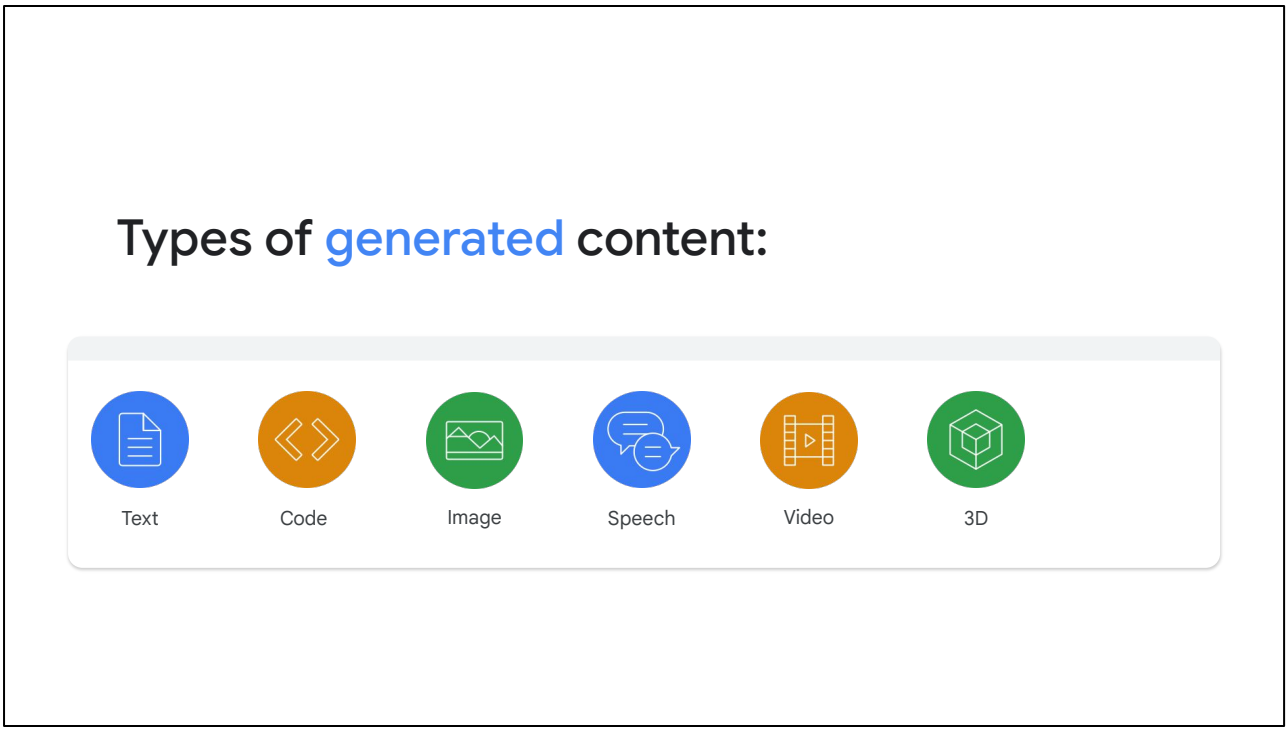
images,



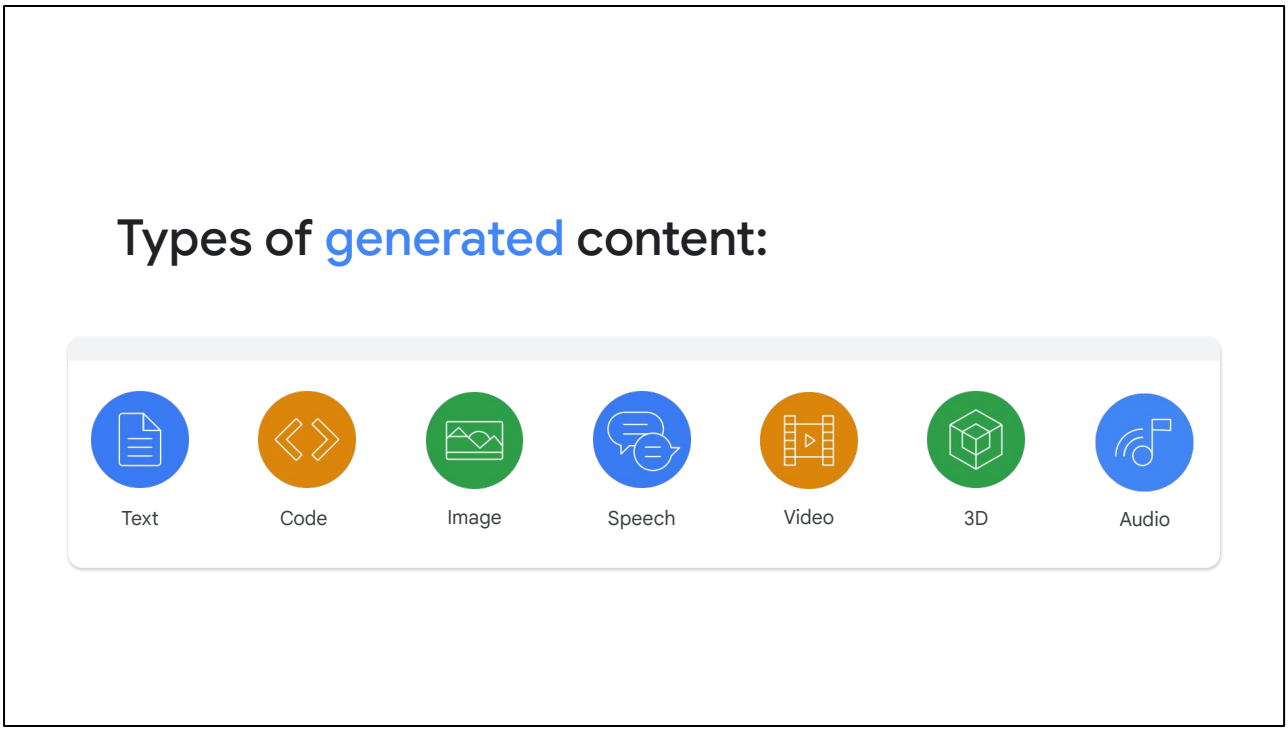
speech,



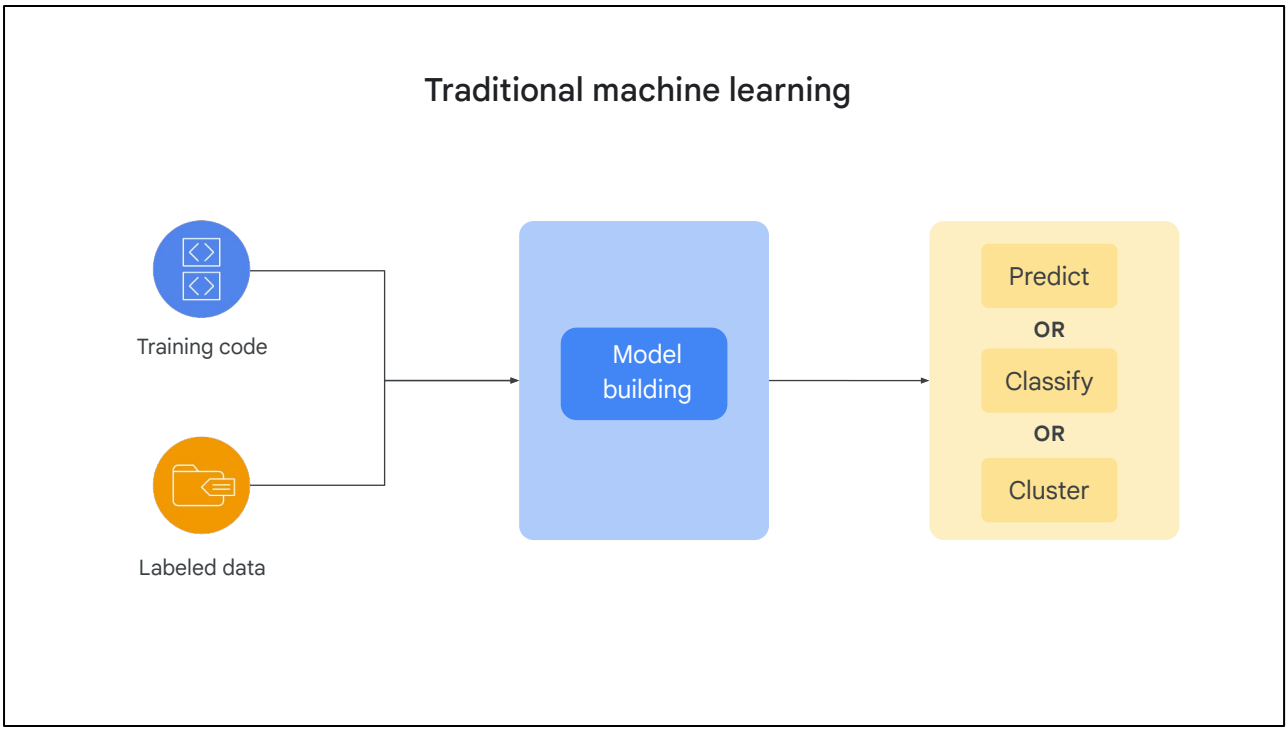
video,



three dimensional entities,


Types of generated content:
Text
Code
Image
Speech
Video
3D
Audio

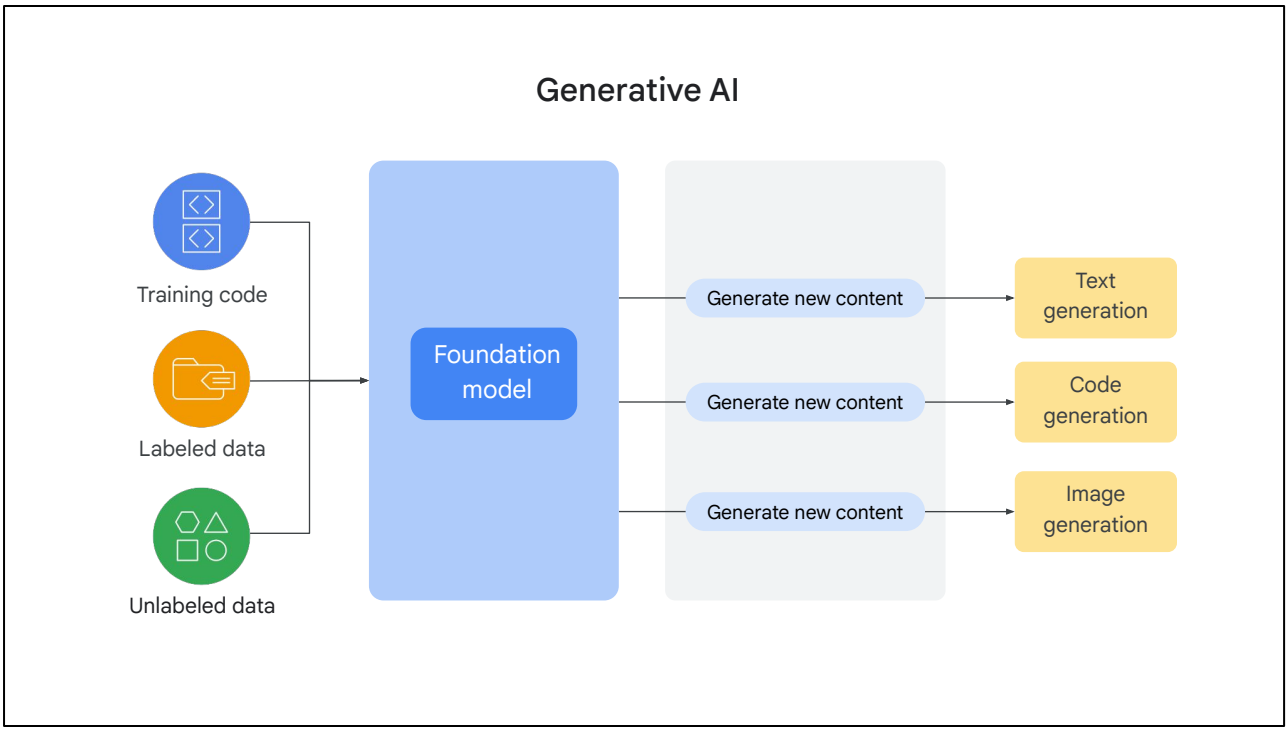
and audio.



The traditional machine learning process uses training code and labeled data to build a model.

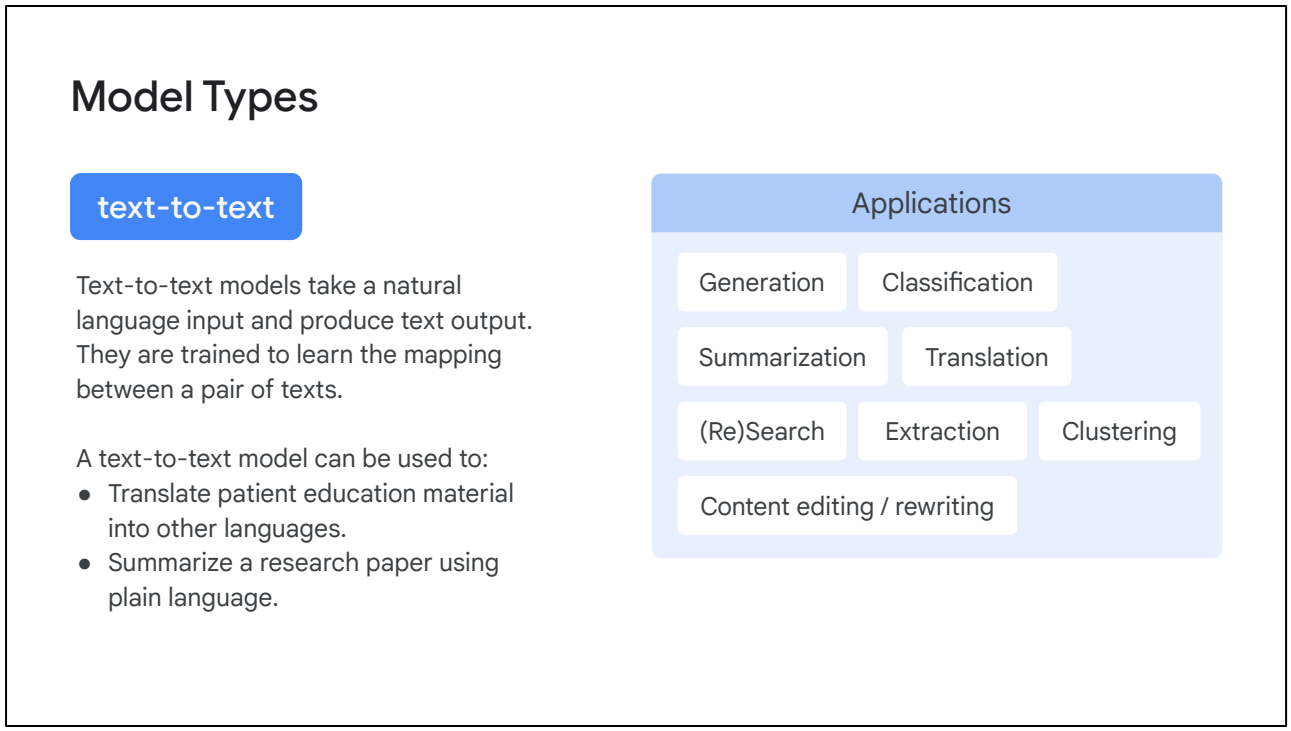
Depending on your needs, the model can be used for prediction, classification, or grouping data points.

But the generative AI process is much more robust.



You can build a foundation model, which is a model that is trained on a large amount of labeled or unlabeled data, and can be adapted to perform a wide range of tasks.

Then, the foundation model can be used to generate new content.



There are several types of Gen AI models.

Text-to-text models take a natural language input and produce text outputs.

These models are trained to learn the mapping between a pair of texts. For example translating from one language to another or summarizing a research paper using plain language.

Model Types

text-to-image

Text-to-image models are trained on a large set of images, each captioned with a short text description. Diffusion is one method used to achieve this.

A text-to-image model can be used to:

- Generate and edit an image explaining a medical condition or procedure.

Applications

- Image generation
- Image editing

Text-to-image models are trained on a large set of images, each captioned with a short text description. Diffusion is one method used to achieve this.

A text-to-image model can be used to generate and edit medical images.

Model Types

text-to-video | text-to-3D

Text-to-video models aim to generate a video representation from text input. The input text can be anything from a single sentence to a full script, and the output is a video that corresponds to the input text. Similarly, text-to-3D models generate three-dimensional objects that correspond to a user's text description.

A text-to-video model can be be used to:

- Generate a video for a patient demonstrating how to perform an exercise at home.

Applications

- Video generation
- Video editing
- Game assets

Text-to-video models aim to generate a video representation from text input. The input text can be anything from a single sentence to a full script, and the output is a video that corresponds to the input text.

Similarly, text-to-3D models generate three-dimensional objects that correspond to a user's text description.

Generating a video demonstrating how to perform an exercise at home is an example of a text-to-video model use case.

Model Types

text-to-task

Text-to-task models are trained to perform a specific task or action based on text input. Examples of tasks include answering a question, performing a search, making a prediction, or taking some sort of action.

A text-to-task model can be used to:

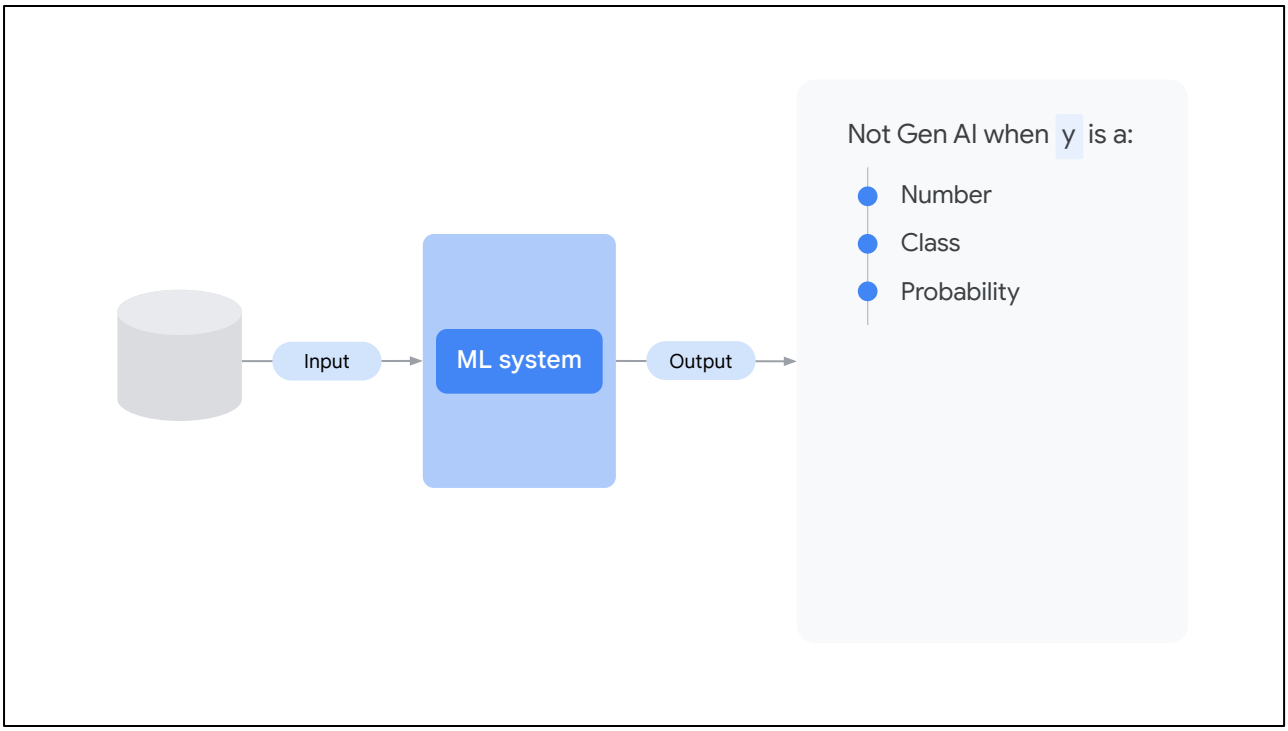
- Prioritize or flag patient messages that require urgent attention.
- Evaluate documentation to suggest appropriate medical codes.

Applications

- Software agents
- Virtual assistants
- Automation

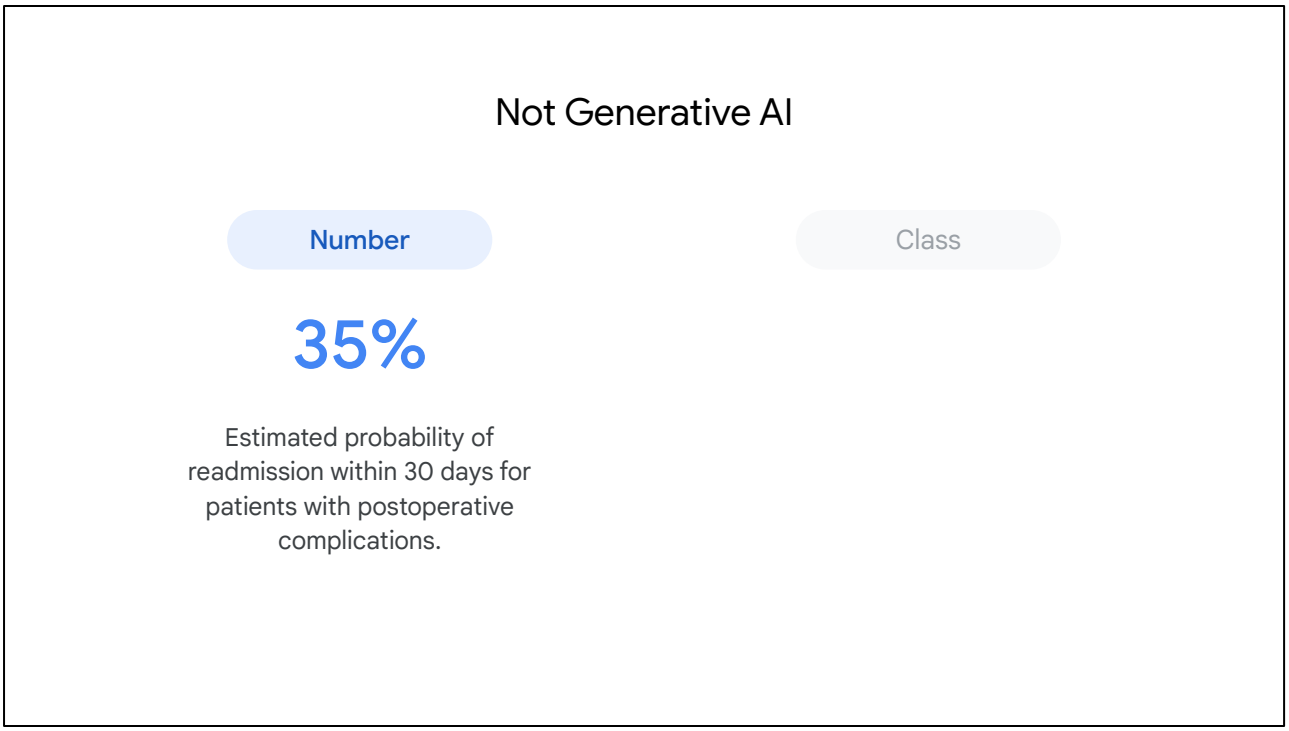
Text-to-task models are trained to perform a defined task or action based on text input.

An action could be flagging patient messages that require urgent attention, or evaluating documentation to suggest appropriate medical codes.



So how can you distinguish between Gen AI and other machine learning methods?

You can examine the output, or “y” label. If the output is a number, a class, or a probability, then it’s not Gen AI.



For example, estimating the probability of a patient being readmitted to the hospital within 30 days of operation (due to complications) results in a numerical value.

Not Generative AI

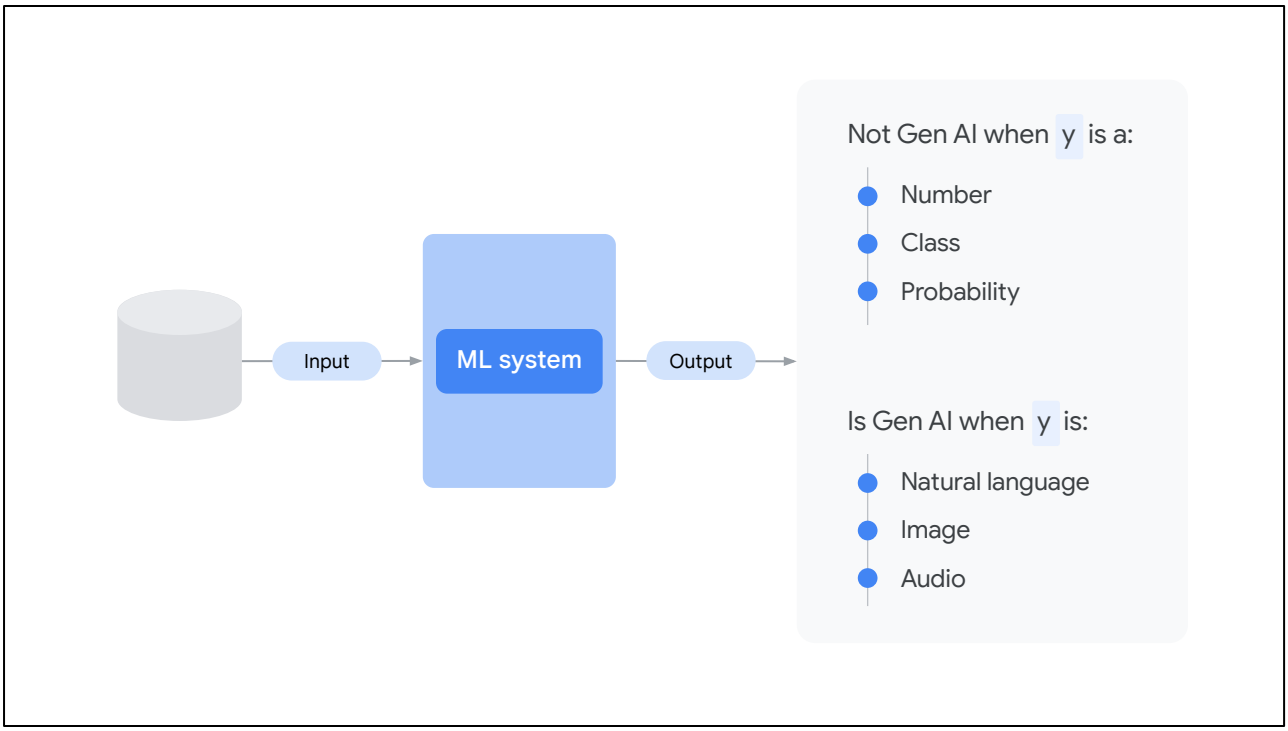
Number

35%

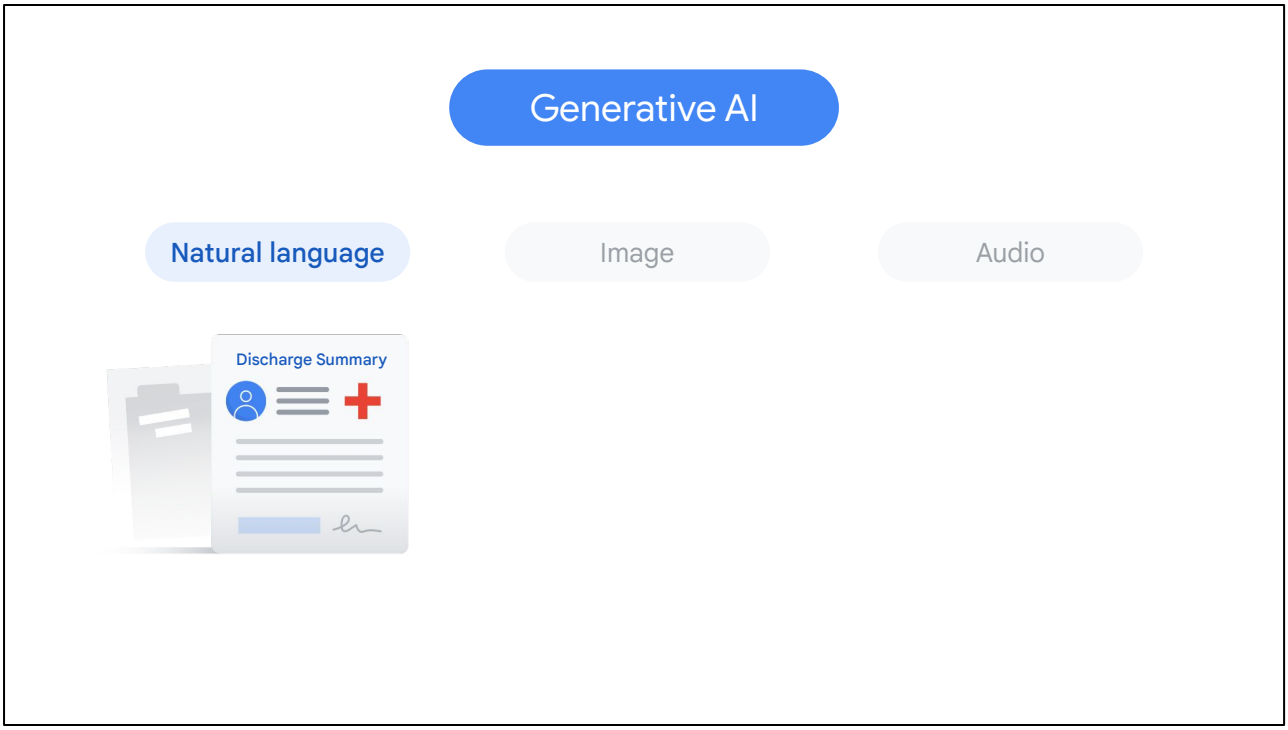
Estimated probability of readmission within 30 days for patients with postoperative complications.

Class

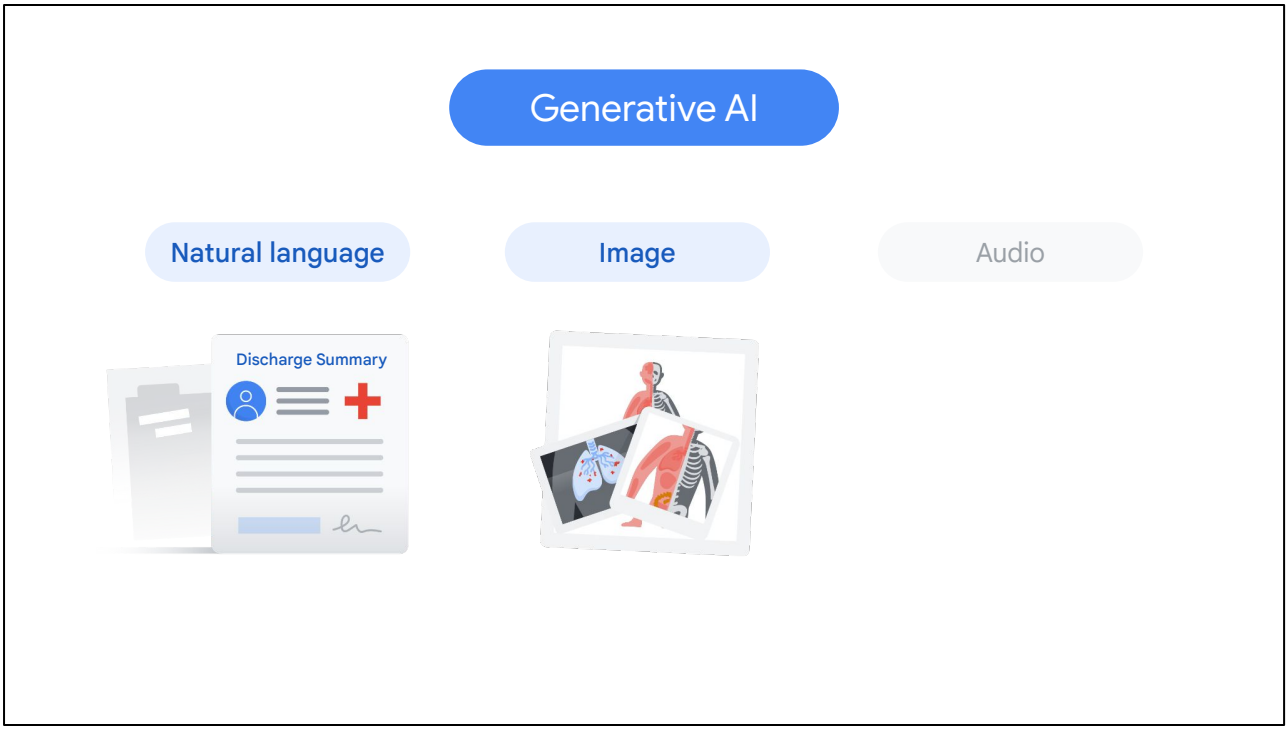
An example of a classification task could be determining if a chest x-ray shows signs of pneumonia.



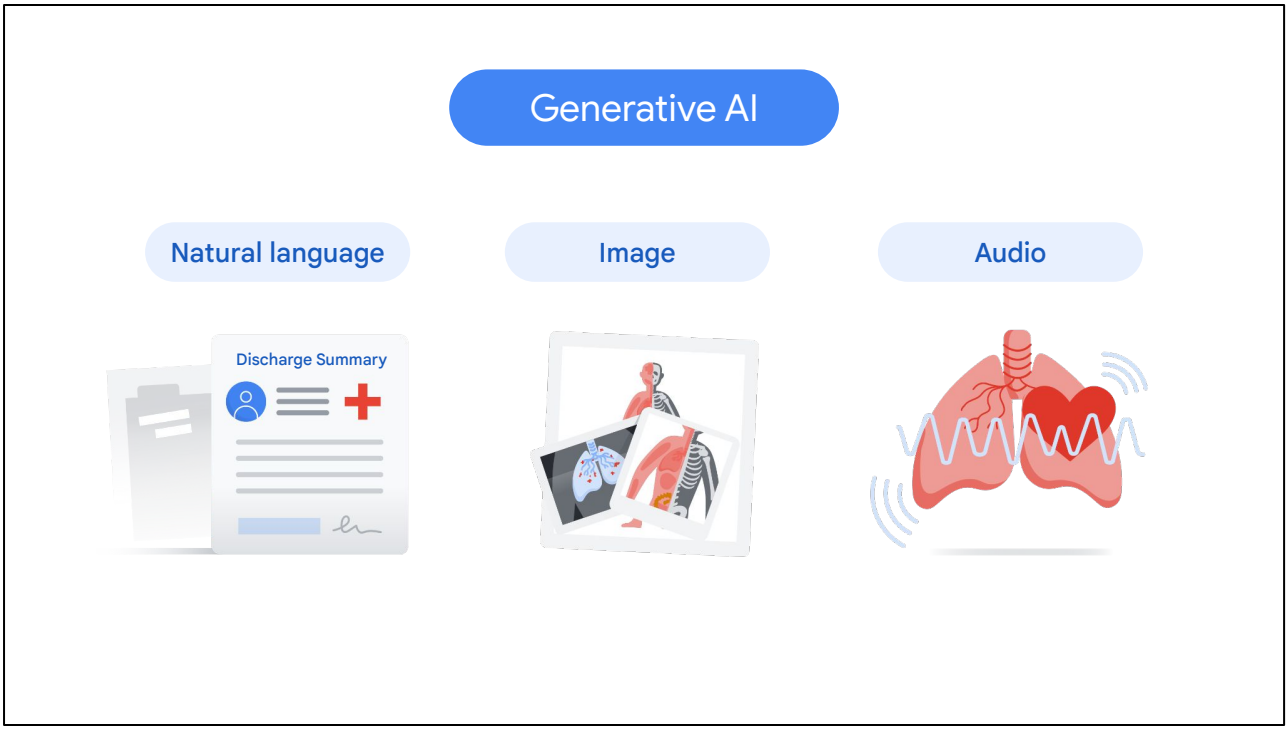
If the output is natural language (like speech or text), an image, or audio, then it is Gen AI.



Drafting a discharge summary from multiple medical notes is an example of natural language.



Creating anatomic models for surgical planning simulation cases results in an image.

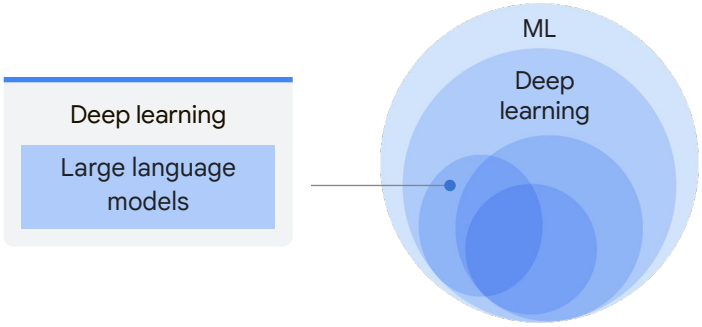


And generating heart and lung sounds to augment medical training is an example of audio.

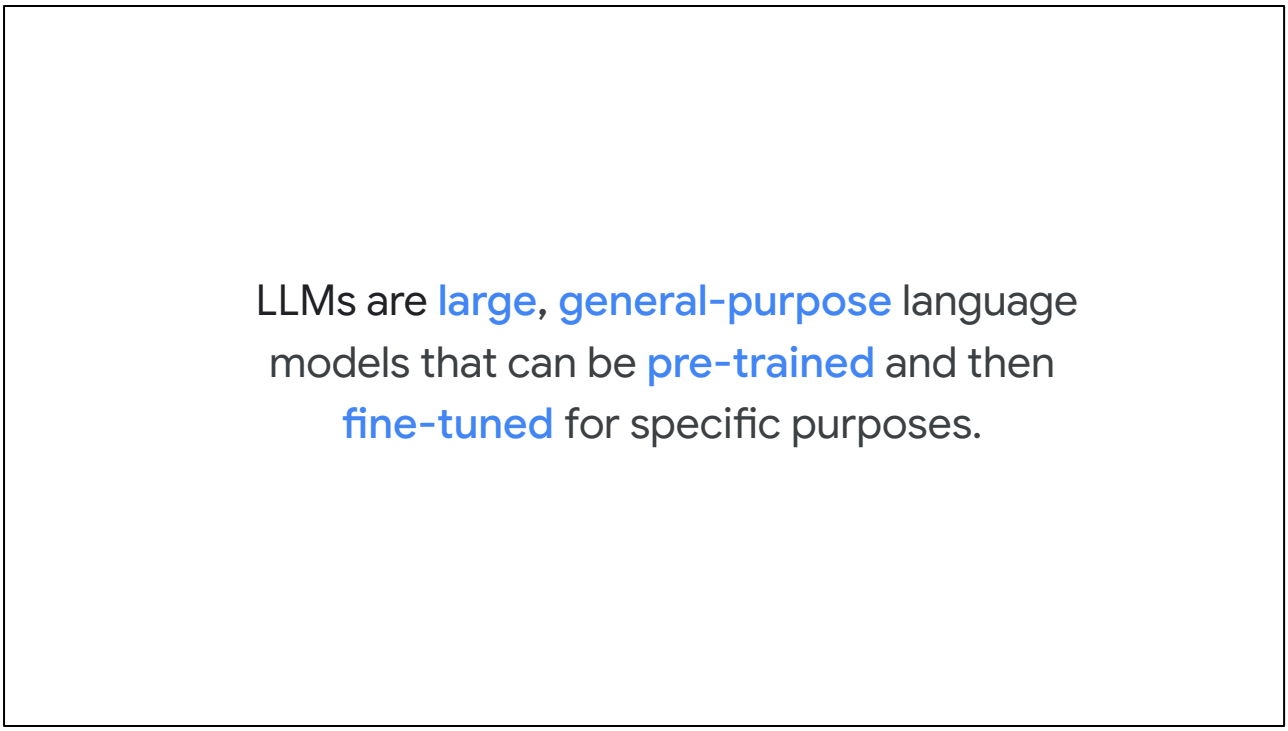
Deck unlocked

V04
Introduction to Large Language Models

Large language models (LLMs) are a specialized type of deep learning.



Large language models – or LLMs – are a specialized type of deep learning, but they aren't the same thing as deep learning in general.



As their name suggests, these are large, general-purpose language models that can be pre-trained and then fine-tuned for specific purposes.

Large language models

Large

- The size of the training dataset
- The number of parameters

What makes these language models so “large”?

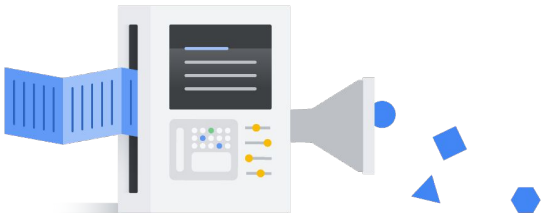
Large language models

Large

- The size of the training dataset
- The number of parameters

First is the enormous size of the training dataset, sometimes at the petabyte scale. To put it into perspective, a single petabyte can store approximately 20 million 4-drawer filing cabinets filled with text.

Large language models



Large

- The size of the training dataset
- The number of parameters

And second is the amount of parameters. Parameters are the memories and knowledge that the machine learned when it was trained.

They define the skill of a model in solving a problem, such as predicting text.

Large language models

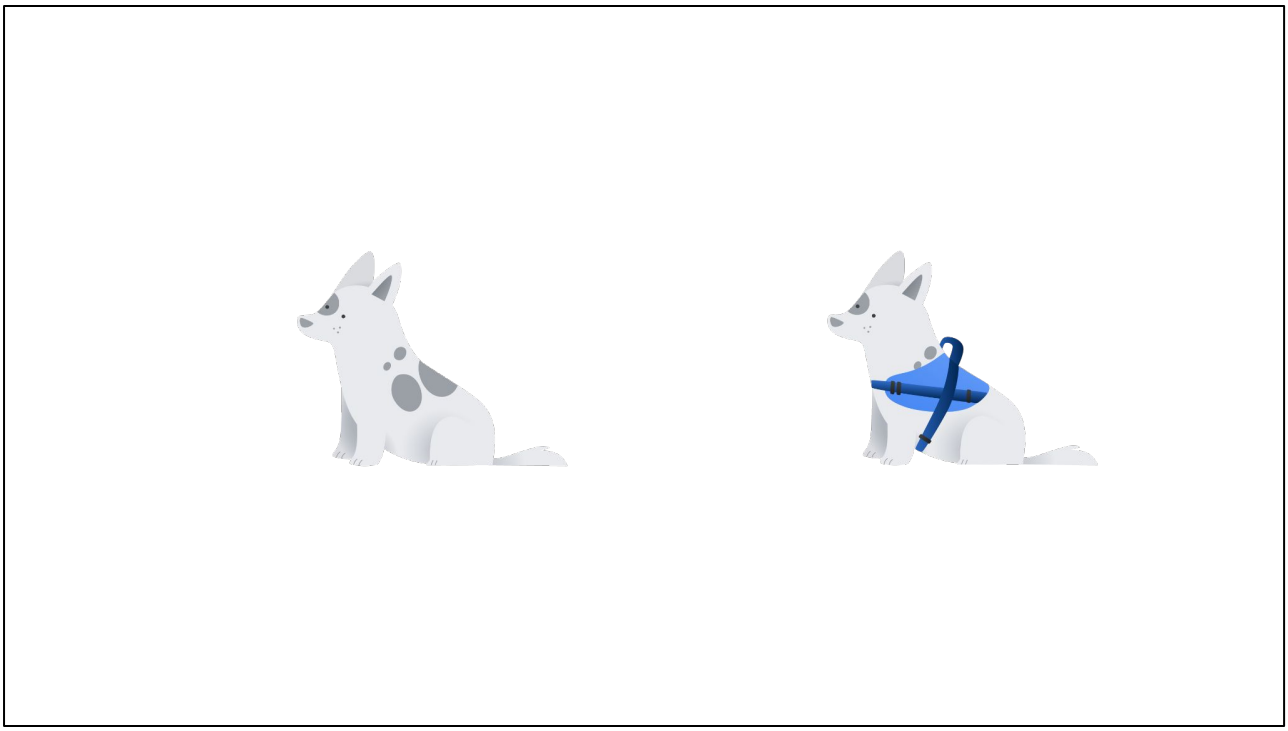
Large

- The size of the training dataset
- The number of parameters

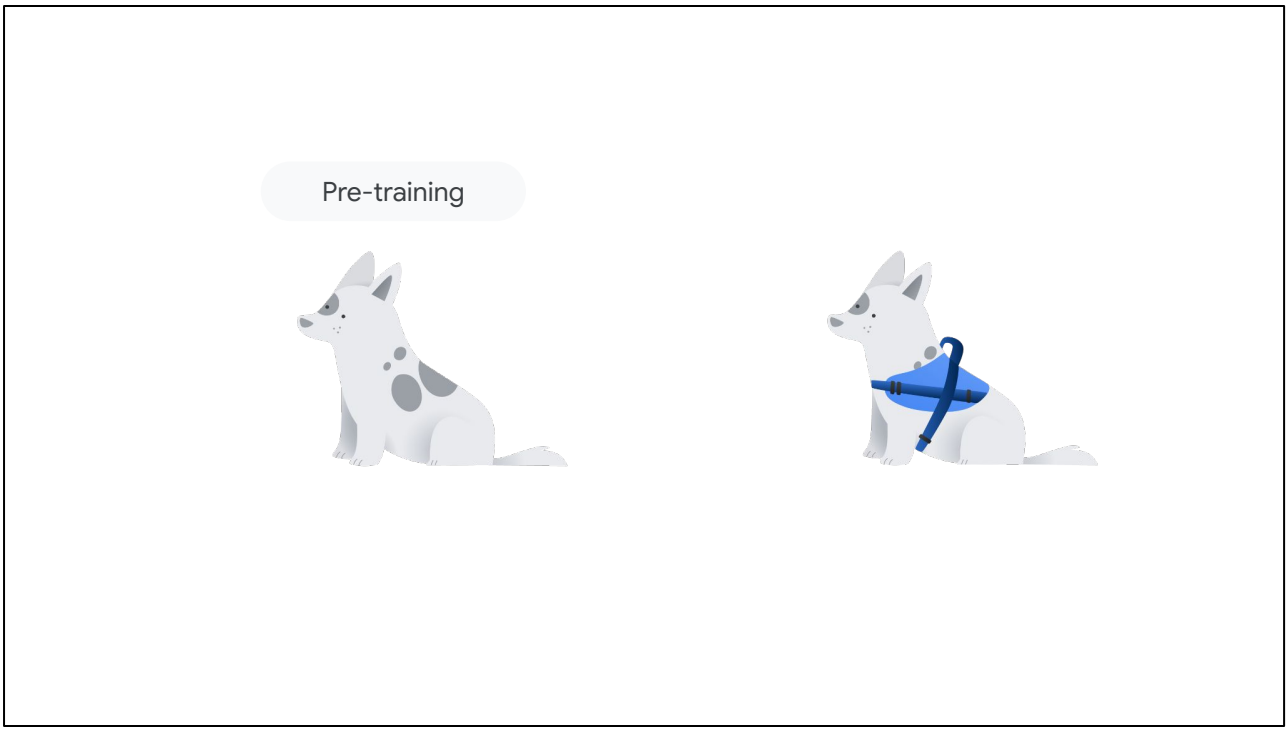
General purpose

- Can solve common problems

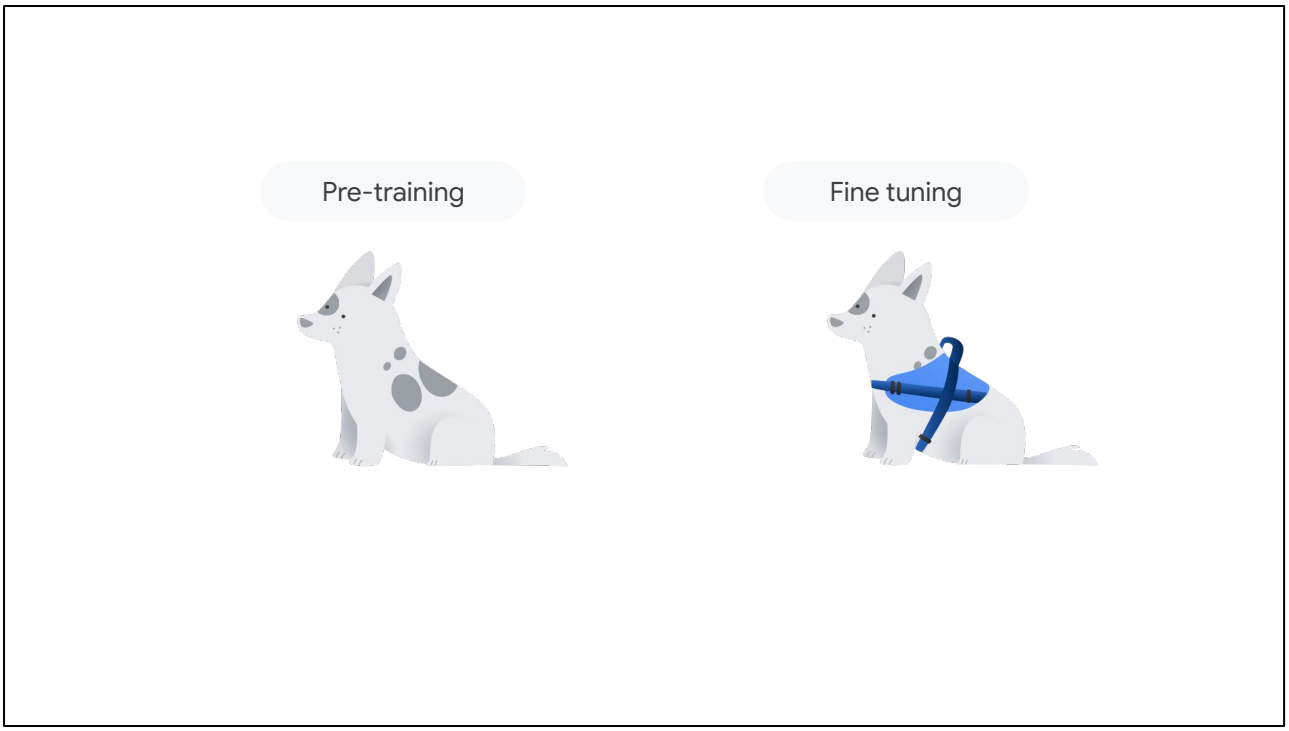
General-purpose means that the models are sufficient to solve common problems.



Now, let’s take a moment to define pre-training and fine tuning.



pre-training



and fine tuning.

Training a dog

Imagine training a dog.

Training a dog

sit

Often you train your dog basic commands such as sit,

Training a dog

come

come,

Training a dog

down

down,

Training a dog

stay

and stay.

A good canine citizen

sit

come

down

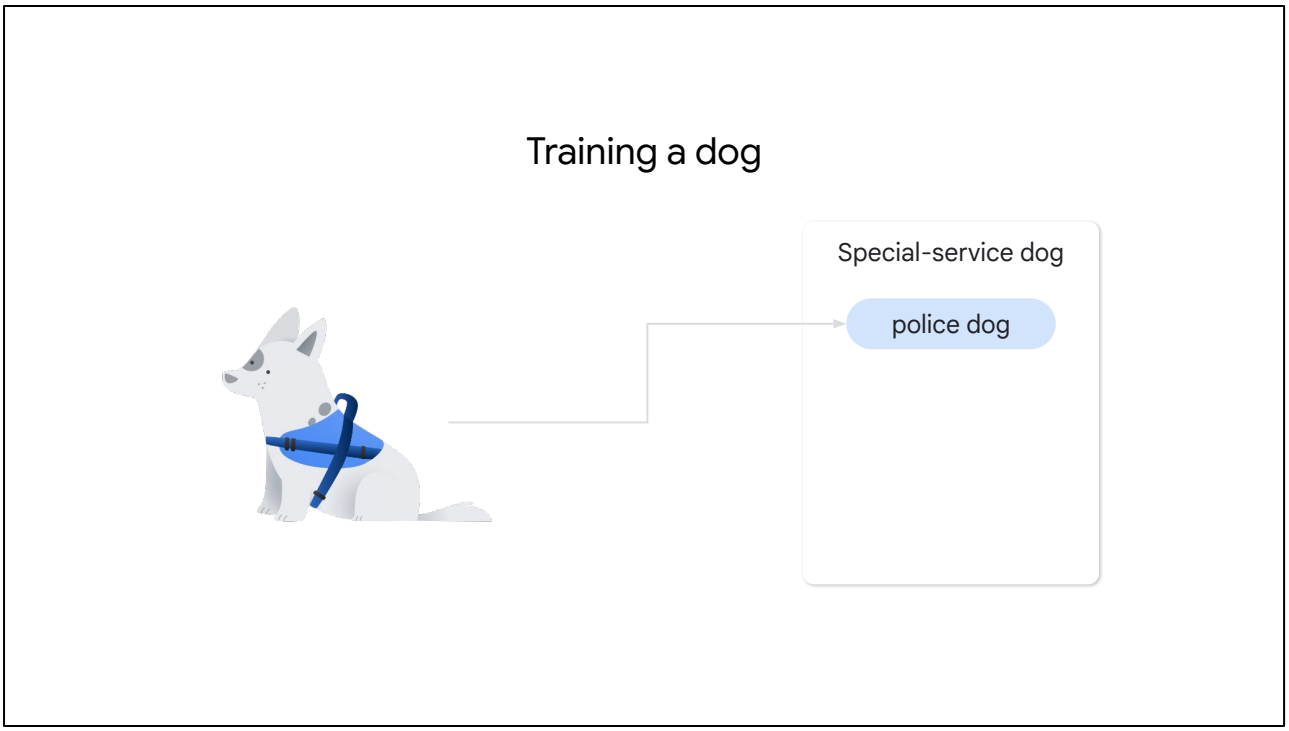
stay

These commands are normally sufficient for everyday life and help your dog become a good canine citizen.

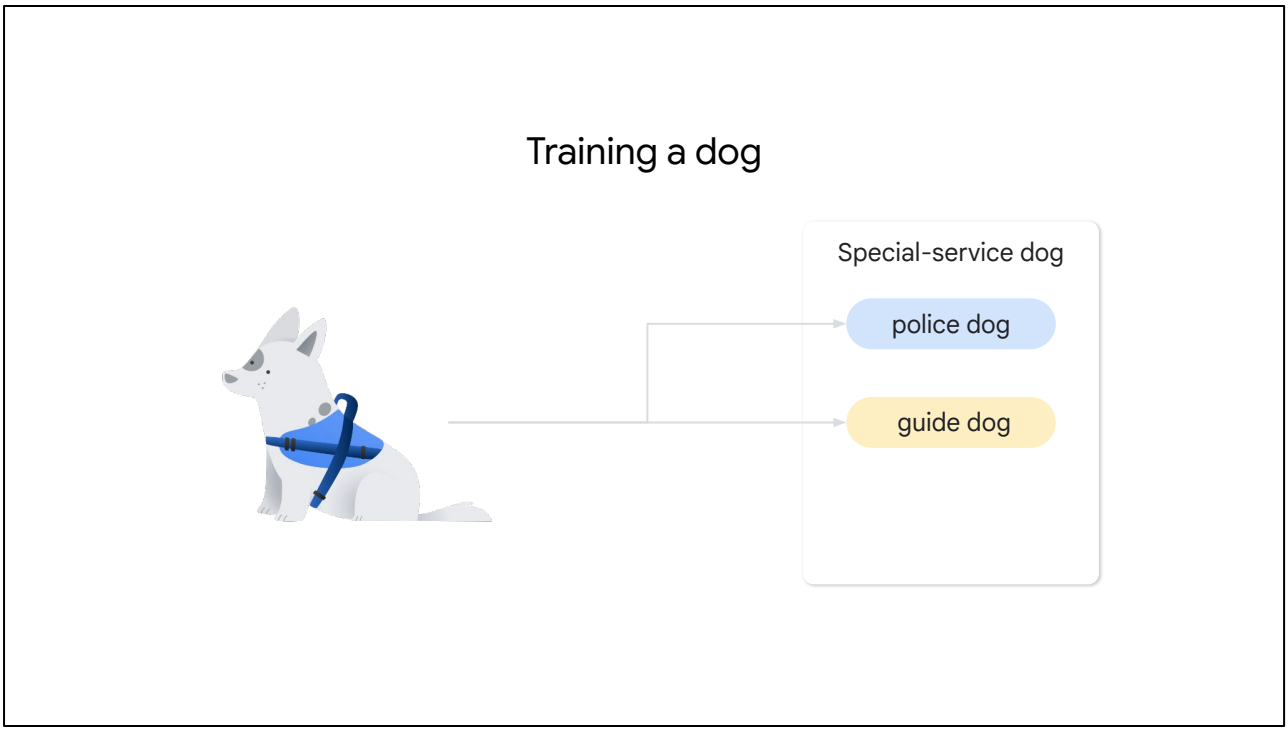
Similar to the basic commands you would teach a dog, pre-training is like giving an LLM a general education. Pre-training typically uses a large dataset.

Training a dog

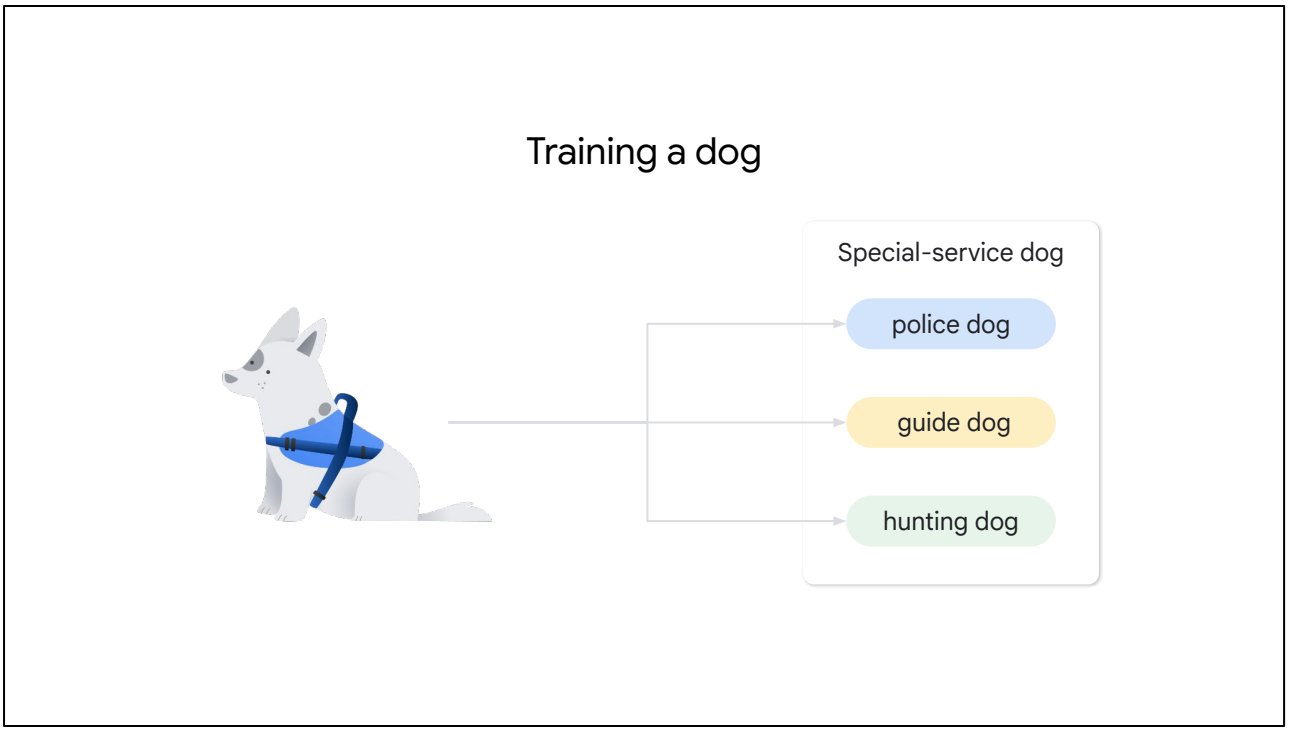
Now, if you need a special-service dog



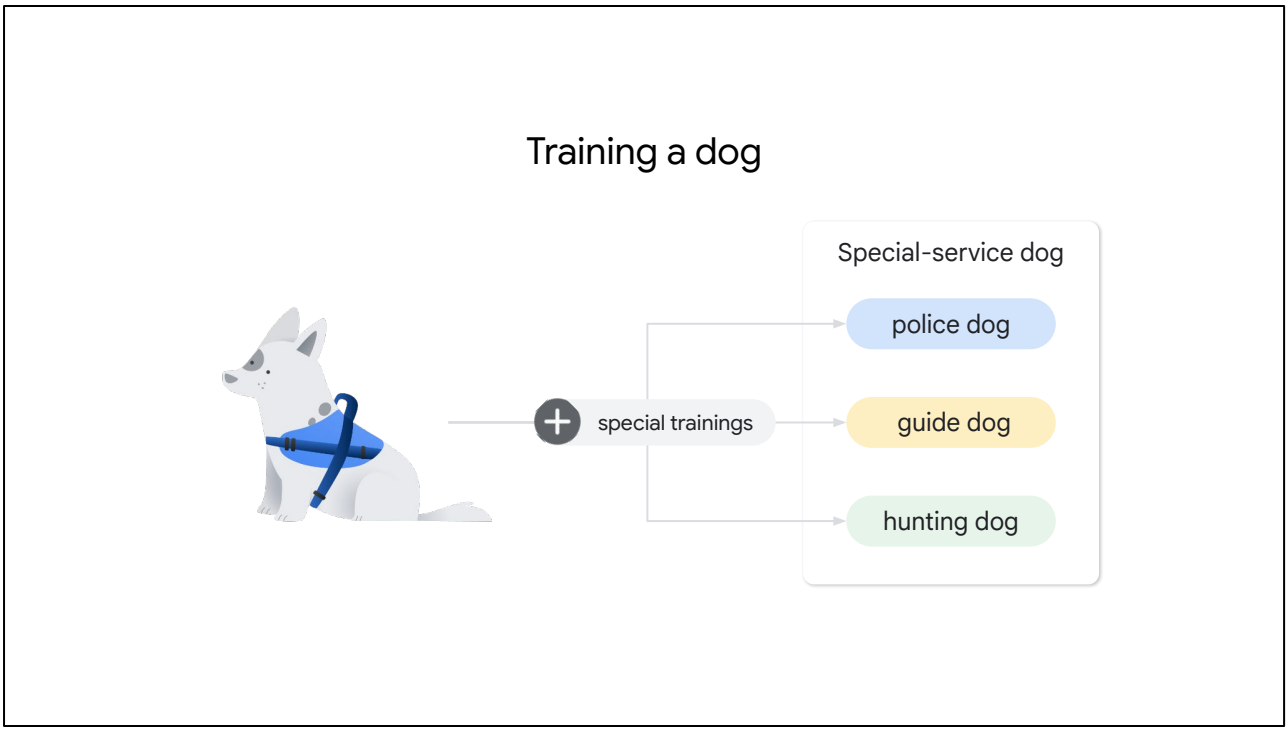
such as a police dog,



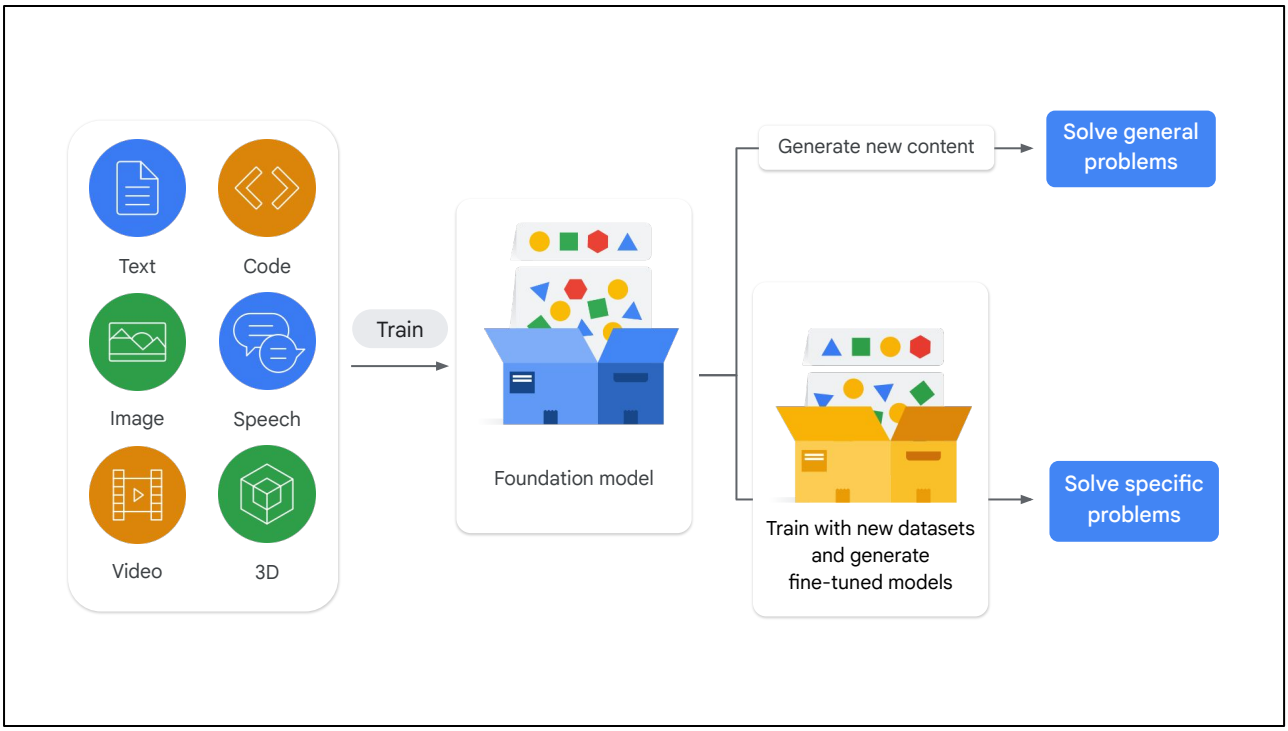
a guide dog,



or a hunting dog,



you add special training. Fine-tuning is like special training. It uses a smaller dataset to adapt a pre-trained LLM to a specific task.



Gen AI relies on the capabilities of large language models to generate text, audio, images, and synthetic data to solve specific and general problems.

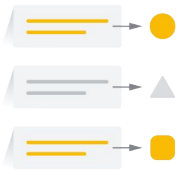
These large language models can be trained further with field-specific datasets to solve specific challenges, like optimizing medication management for patients with complex chronic conditions.

This results in the creation of a new model that's tailored to your needs.

Large language models are trained to solve
common language problems, like...

LLMs are trained for general purposes to solve common language problems

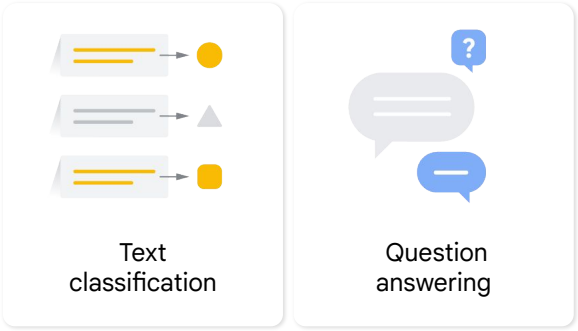
Large language models are trained to solve common language problems, like...



Text classification

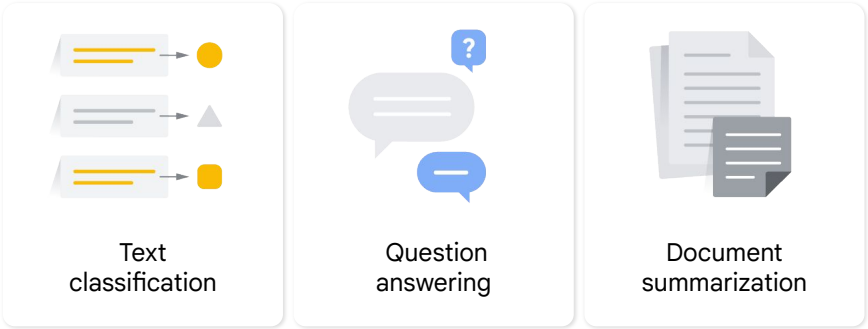
such as text classification,

Large language models are trained to solve common language problems, like...

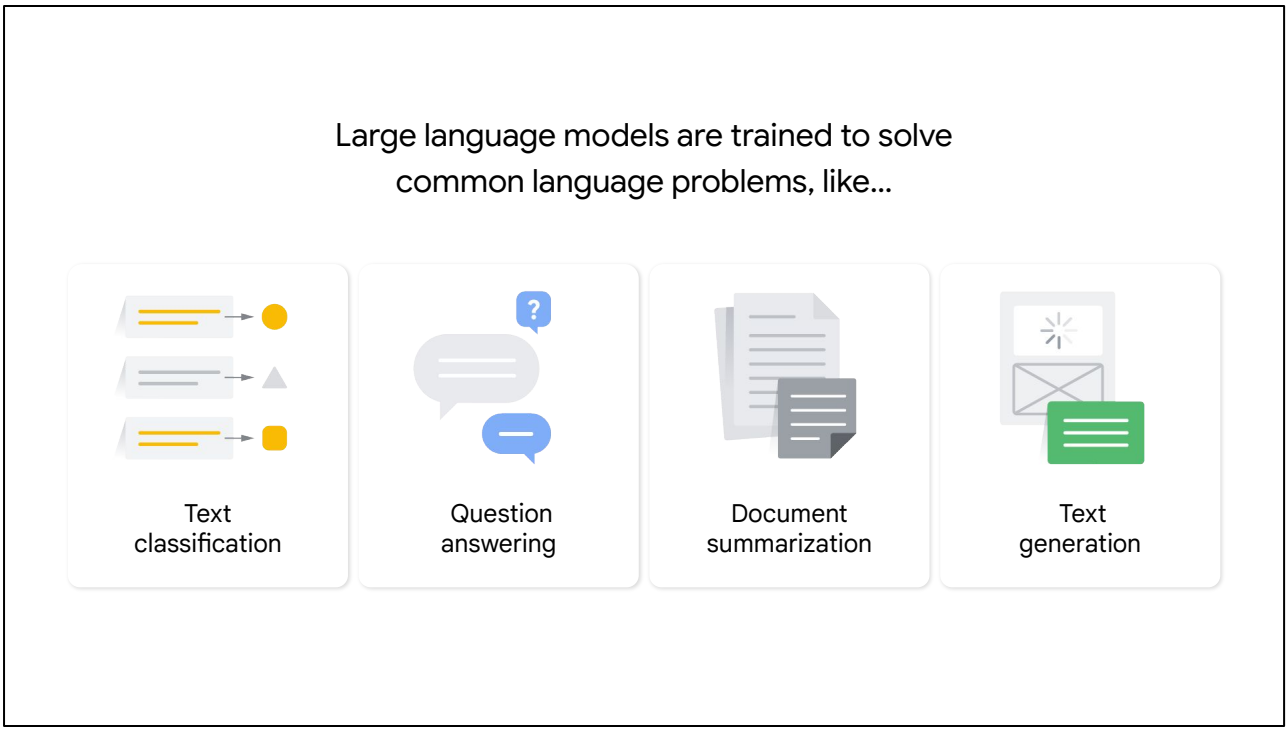


question answering,

Large language models are trained to solve common language problems, like...



document summarization,



and text generation.

...then they can be tailored to solve specific problems in different fields, like...

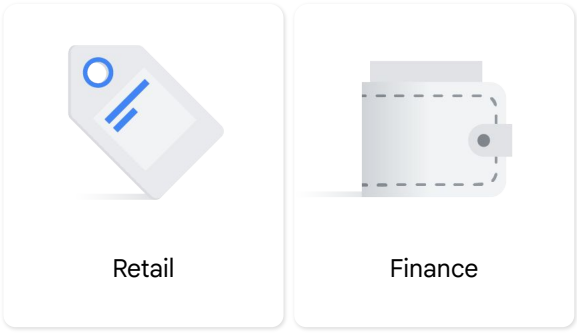
And then they can be tailored to solve specific problems in different fields such as

...then they can be tailored to solve specific problems in different fields, like...

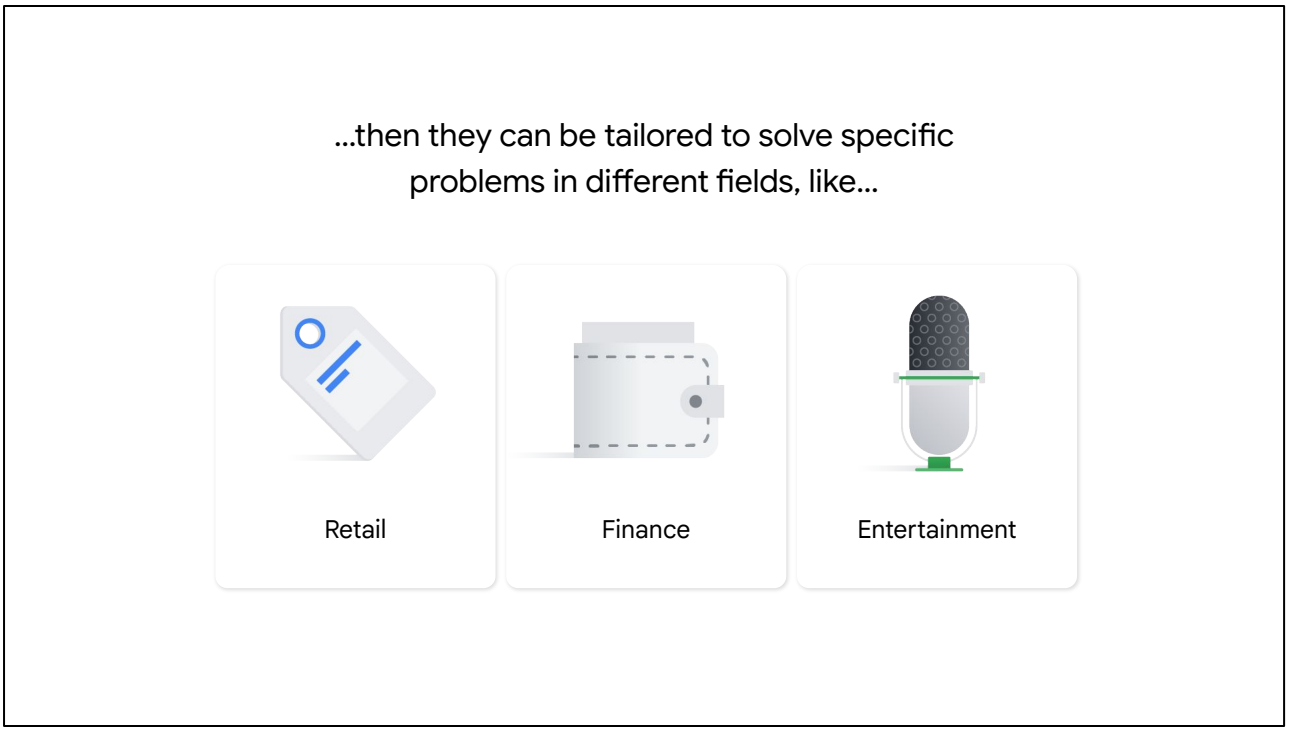
Retail

retail,

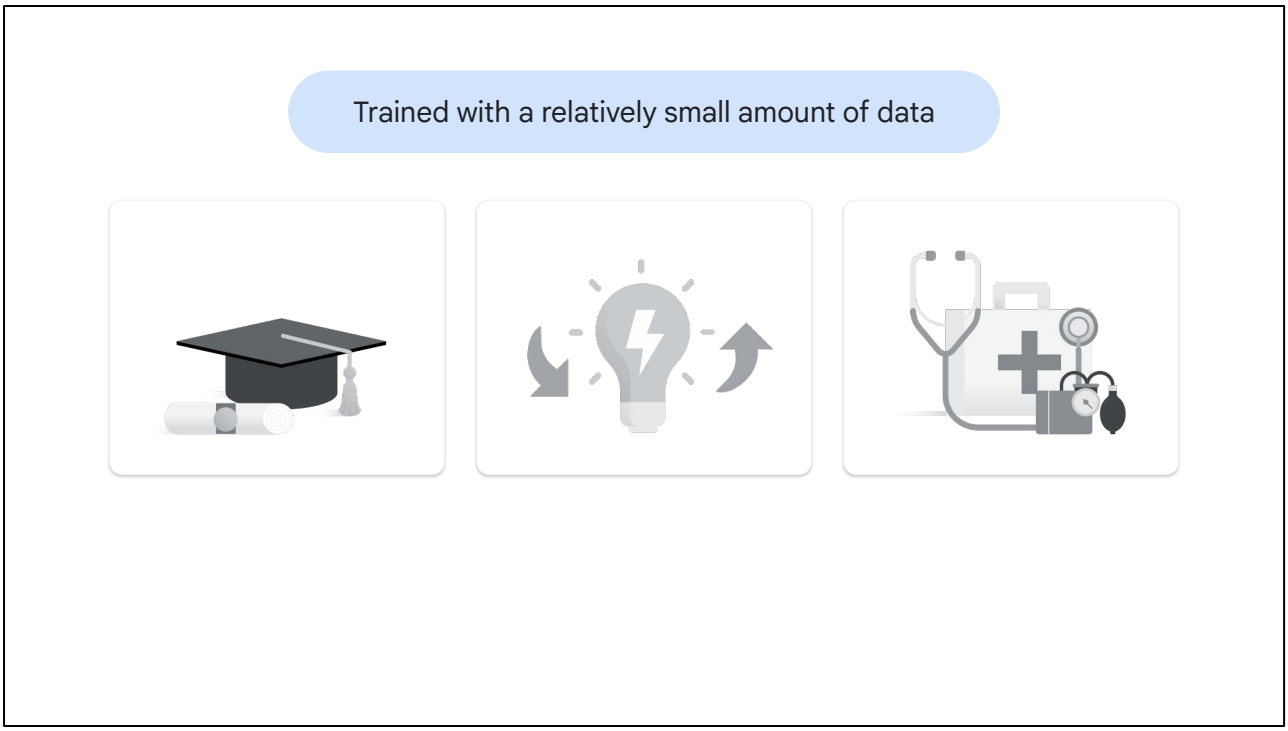
...then they can be tailored to solve specific problems in different fields, like...



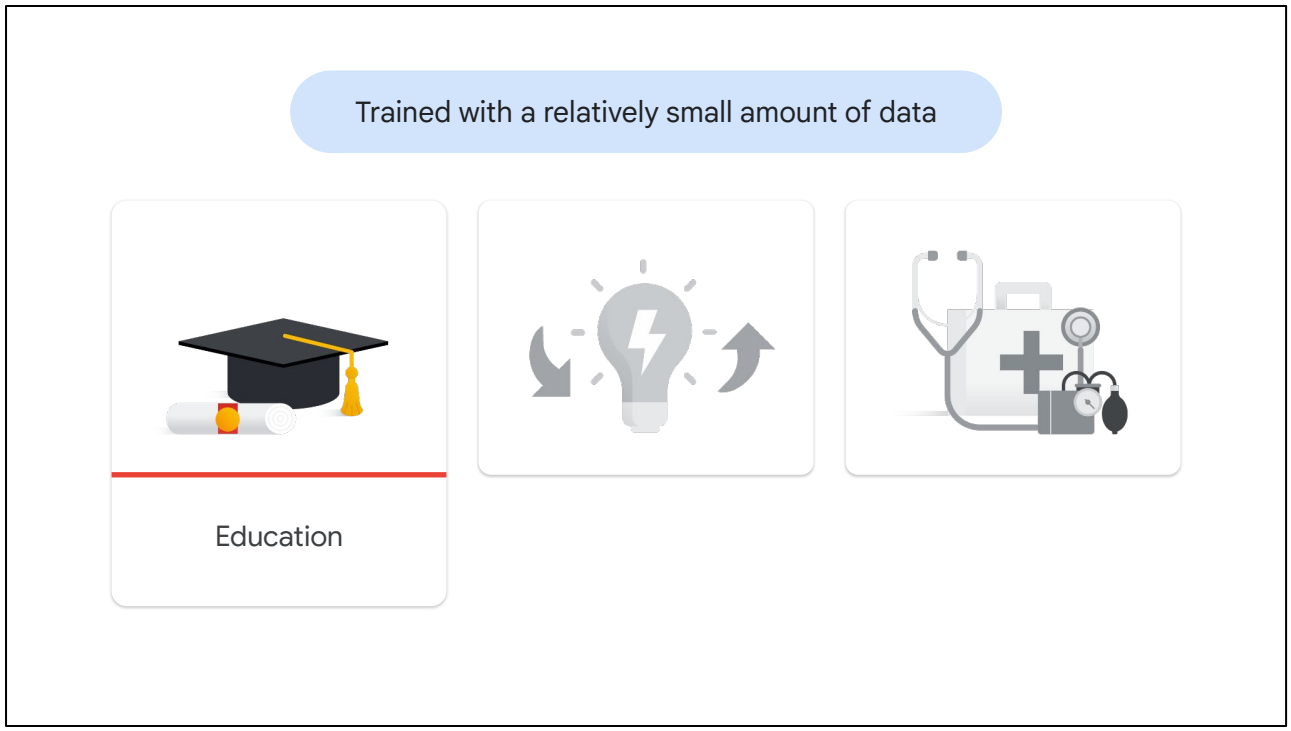
finance,



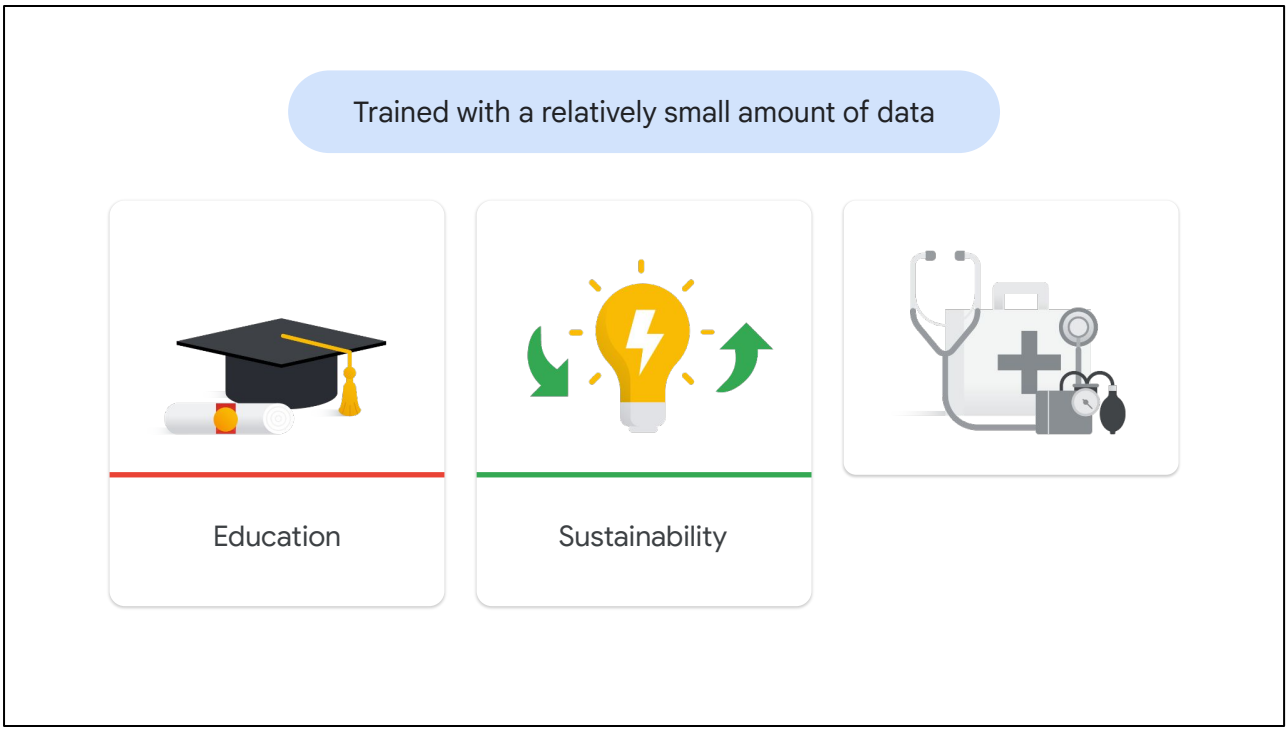
and entertainment.



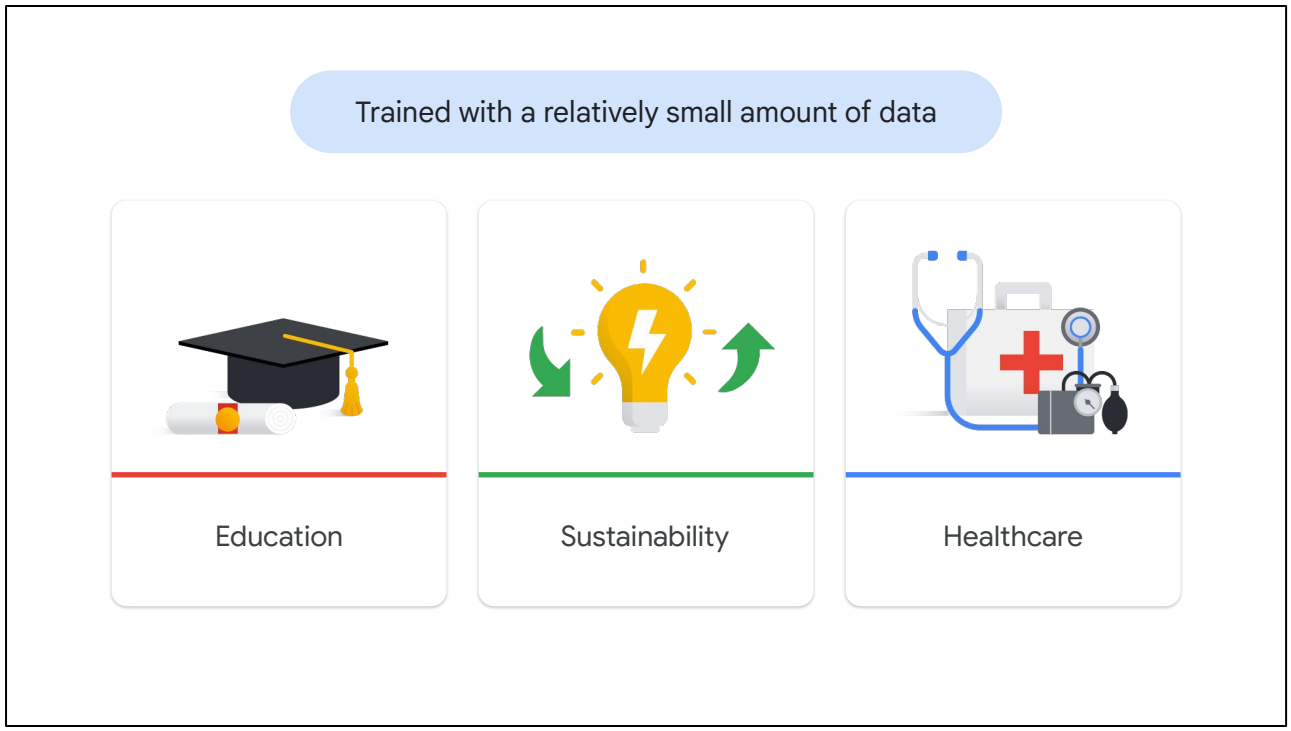
Using a relatively small amount of data,



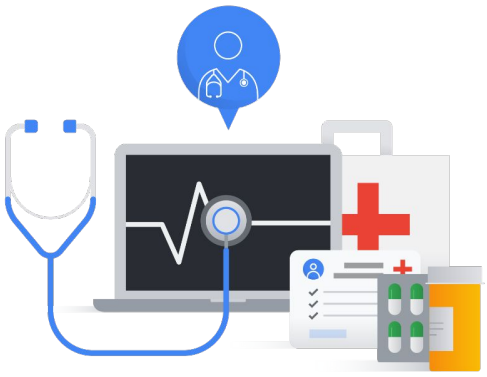
they can also be tailored for education,



sustainability,



and, of course, healthcare.



Problem

To make informed diagnosis and treatment decisions for rare diseases, clinicians need to quickly and easily access the latest research summaries.

Let’s explore a specific medical domain problem that an LLM could potentially help solve.

To make informed diagnoses and treatment decisions for rare diseases, clinicians need to quickly and easily access the latest research summaries.

This is a challenge because:

This can often be a challenge because:

This is a challenge because:

Research is rapidly evolving

- Rare diseases are, well, rare. Research is rapidly evolving

This is a challenge because:

- Research is rapidly evolving
- Doctors might have limited experience with rare diseases

and a doctor might have limited experience with a particular condition.

This is a challenge because:

- Research is rapidly evolving
- Doctors might have limited experience with rare diseases
- Scientific literature is vast

- Scientific literature is vast. Keeping up with the latest research across hundreds or thousands of rare diseases is time-consuming and overwhelming. Or the answers to medical questions might be buried in dense medical journals.

This is a challenge because:

- Research is rapidly evolving.
- Doctors might have limited experience with rare diseases.
- Scientific literature is vast.
- Doctors need reliable summaries quickly.

- Finally, doctors need reliable summaries quickly, and they shouldn't need to set aside time to read full research papers.

So how can a large language model help clinicians overcome this challenge?

Solution: Train an LLM to search and identify relevant studies

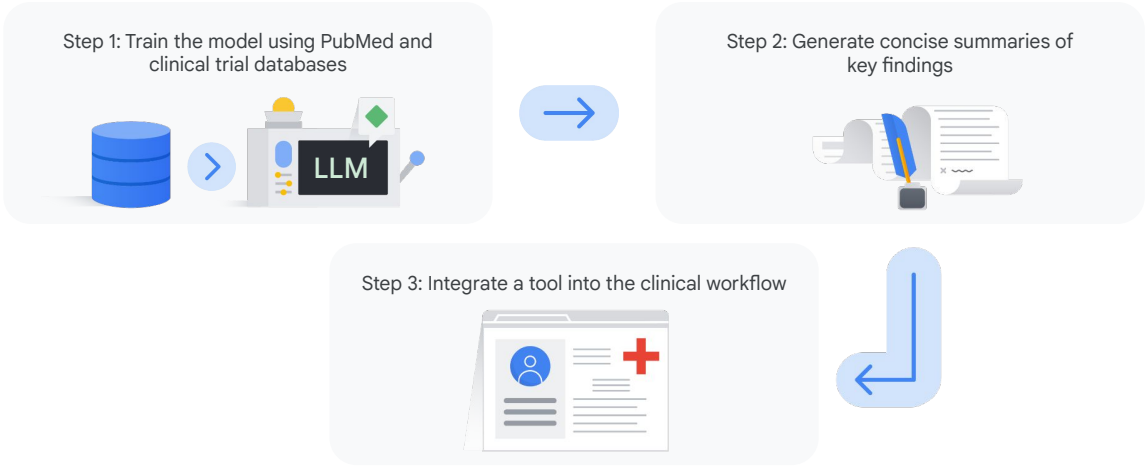
First, an LLM can be trained to search through massive repositories of medical literature (like PubMed and clinical trial databases) and identify information relevant to a specific rare disease.

Solution: Train an LLM to search and identify relevant studies

Then, using its natural language processing capabilities, the LLM can generate concise summaries of key findings from relevant research papers or reports, highlighting:

- Diagnostic criteria
- Latest treatment recommendations
- And ongoing clinical trials

Solution: Train an LLM to search and identify relevant studies



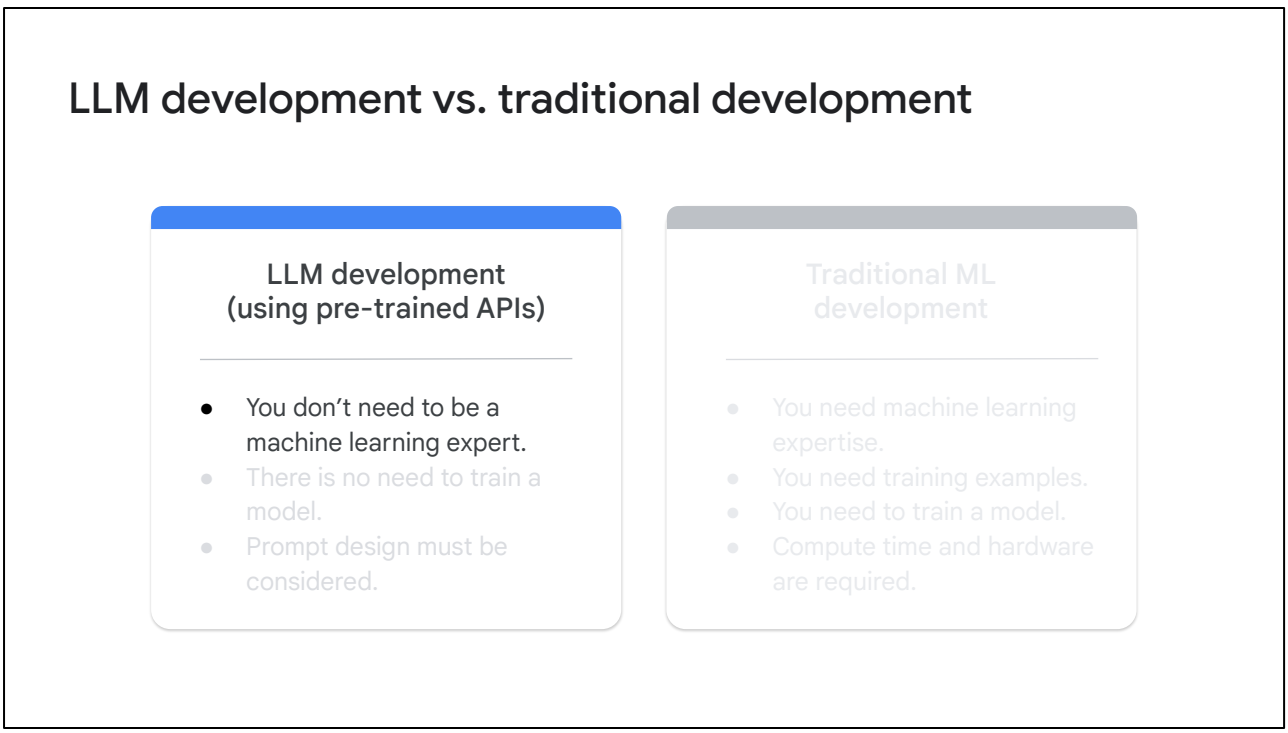
And finally, a tool can be integrated into the clinical workflow, with a simple interface where clinicians can enter a suspected rare disease and instantly access tailored summaries generated by the LLM.

LLM development vs. traditional development

LLM development (using pre-trained APIs)	Traditional ML development
• You don't need to be a machine learning expert. • There is no need to train a model. • Prompt design must be considered.	• You need machine learning expertise. • You need training examples. • You need to train a model. • Compute time and hardware are required.

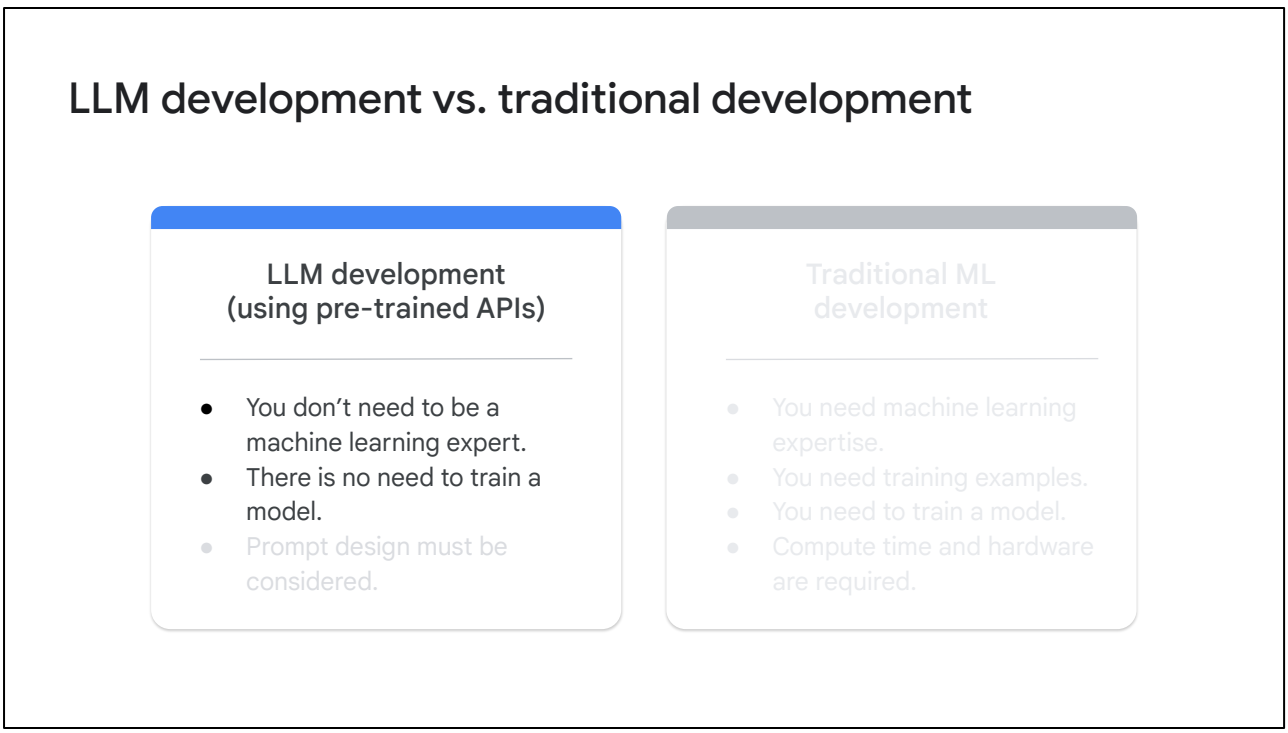
Next, let's explore the benefits of using pre-trained large language models in comparison to traditional development methods.

LLM development vs. traditional development



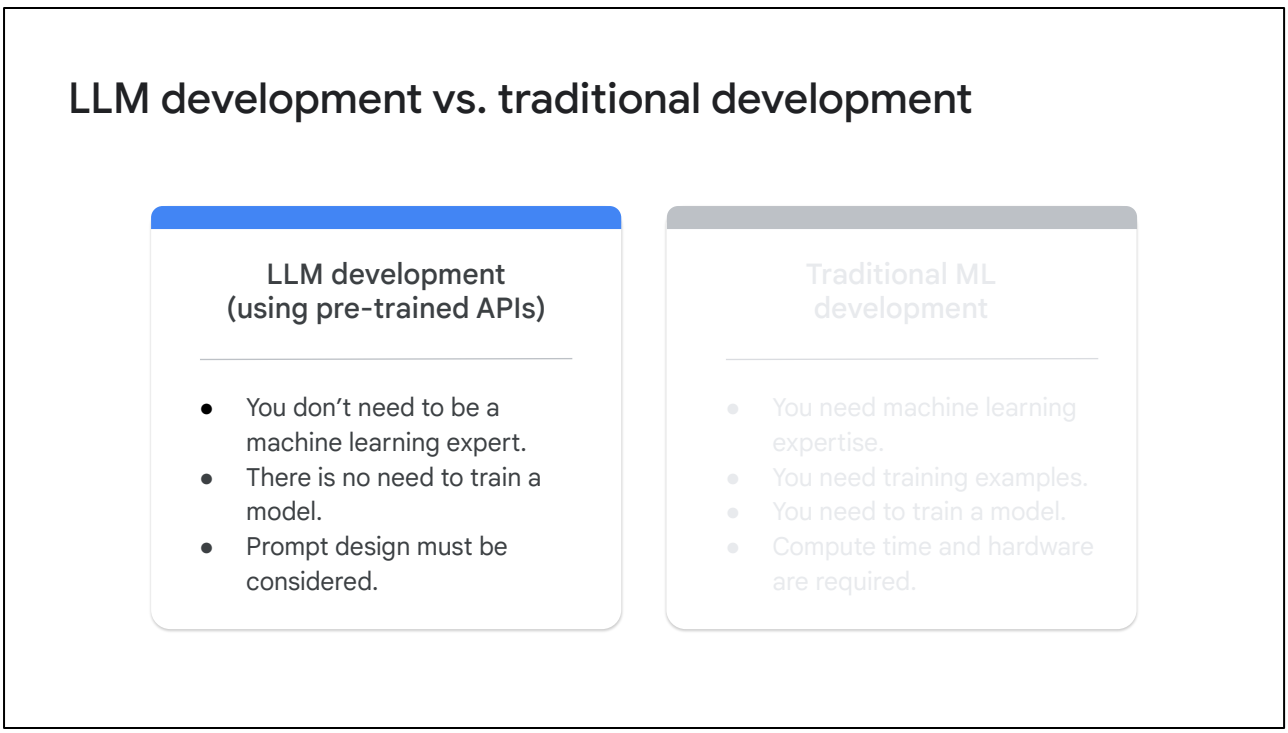
You don't need to be an expert to develop LLMs.

LLM development vs. traditional development



You don’t need training examples, and there’s no need to train a model.

LLM development vs. traditional development



However, it's important to consider prompt design, which is the process of creating a prompt that's clear, concise, and informative.

LLM development vs. traditional development

LLM development (using pre-trained APIs)	Traditional ML development
• You don't need to be a machine learning expert. • There is no need to train a model. • Prompt design must be considered.	• You need machine learning expertise. • You need training examples. • You need to train a model. • Compute time and hardware are required.

Traditional machine learning requires expertise,

LLM development vs. traditional development

LLM development (using pre-trained APIs)	Traditional ML development
• You don't need to be a machine learning expert. • There is no need to train a model. • Prompt design must be considered.	• You need machine learning expertise. • You need training examples. • You need to train a model. • Compute time and hardware are required.

training examples,

LLM development vs. traditional development

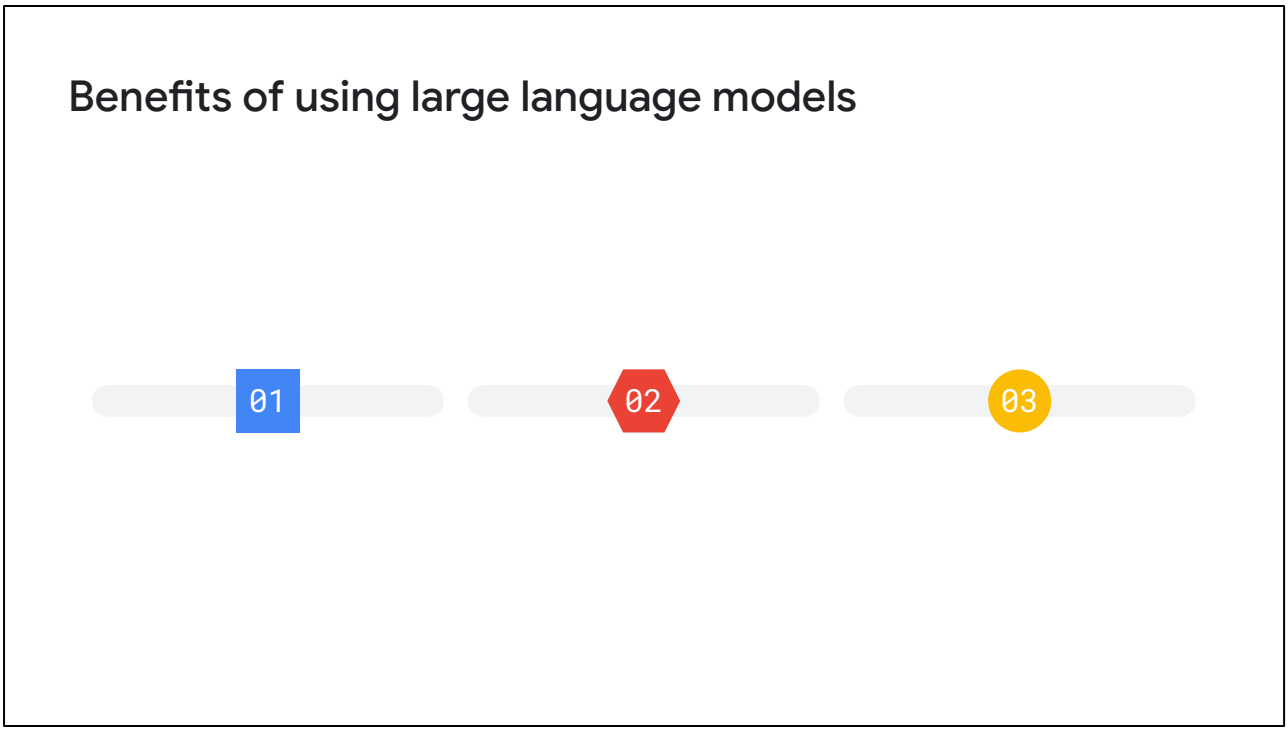
LLM development (using pre-trained APIs)	Traditional ML development
• You don't need to be a machine learning expert. • There is no need to train a model. • Prompt design must be considered.	• You need machine learning expertise. • You need training examples. • You need to train a model. • Compute time and hardware are required.

model training,

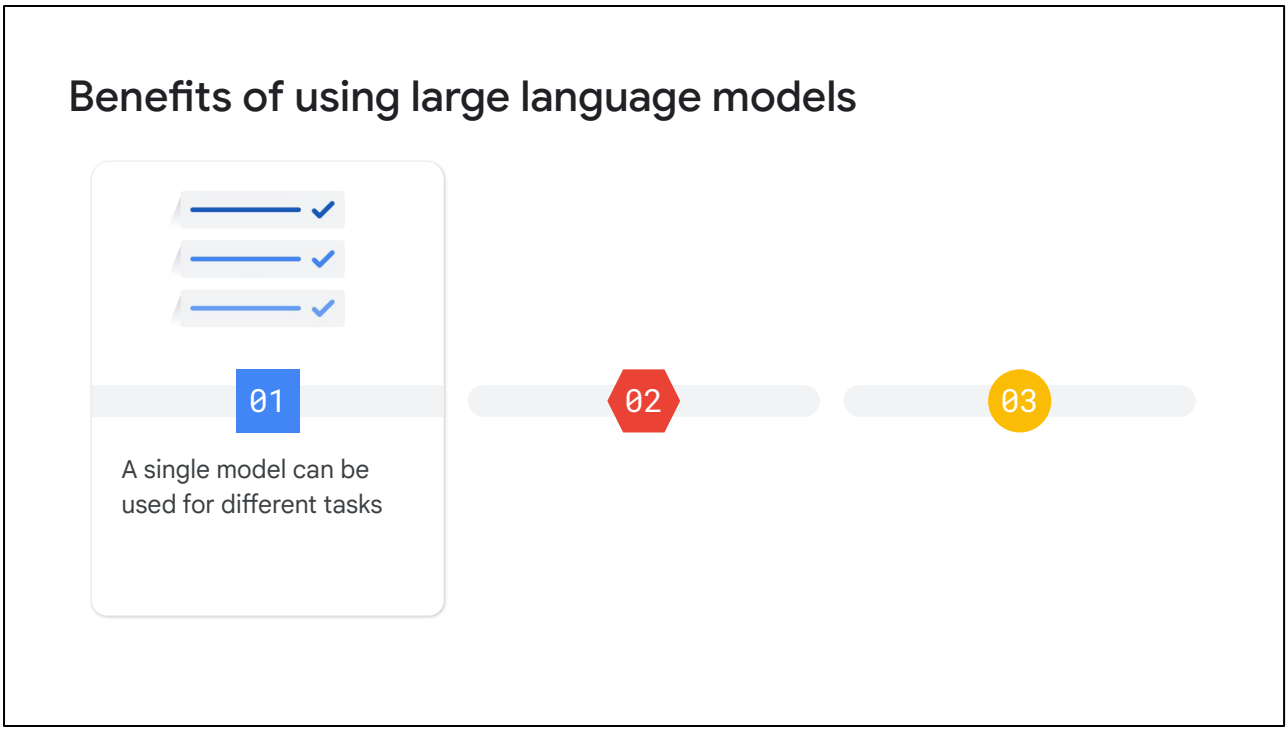
LLM development vs. traditional development

LLM development (using pre-trained APIs)	Traditional ML development
• You don't need to be a machine learning expert. • There is no need to train a model. • Prompt design must be considered.	• You need machine learning expertise. • You need training examples. • You need to train a model. • Compute time and hardware are required.

compute time, and hardware.

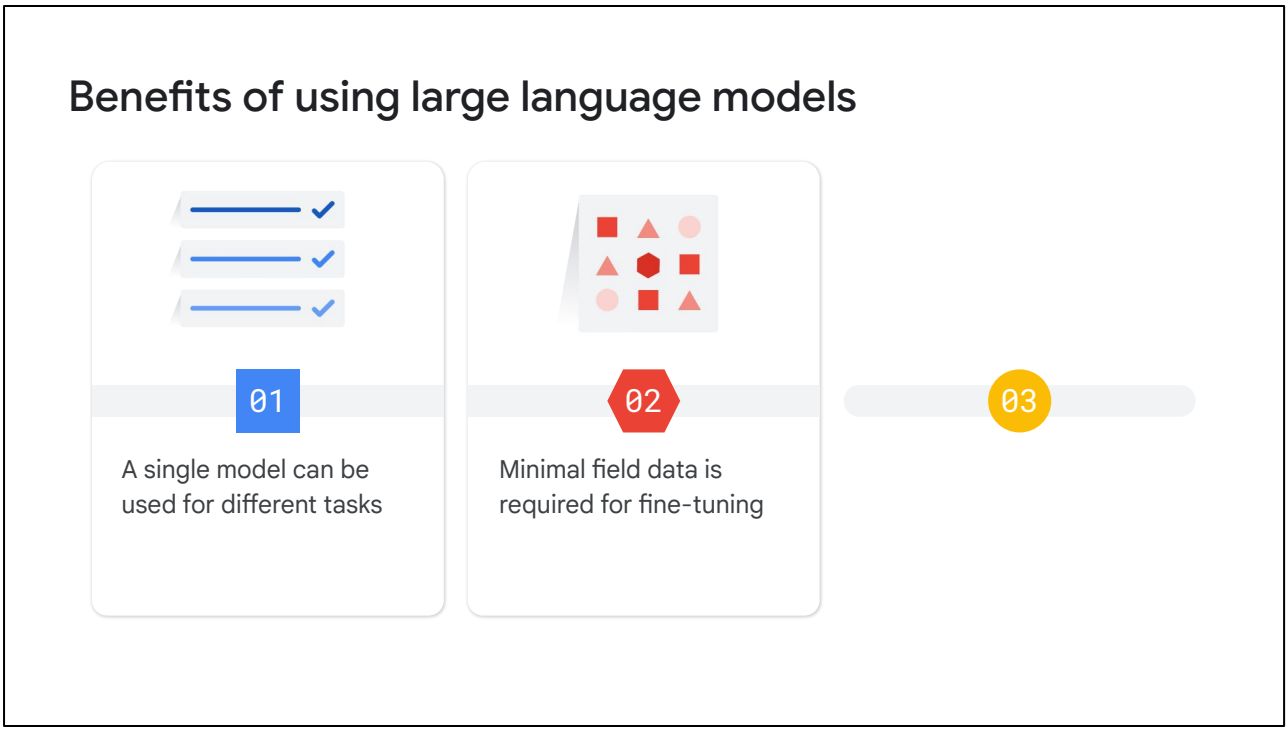


The benefits of using large language models are straightforward:



First, a single model can be used for different tasks. This is a dream come true.

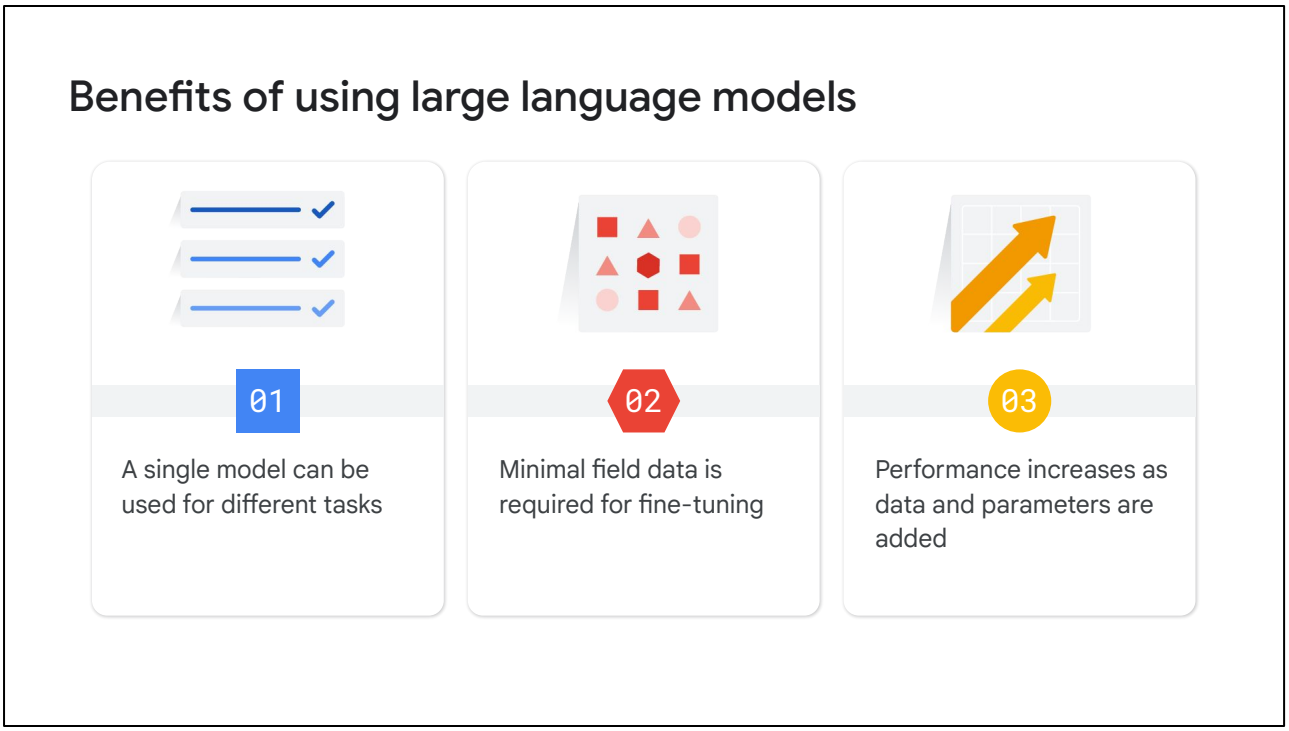
Large language models that are trained with petabytes of data and generate billions of parameters are smart enough to solve specific tasks like language translation, sentence completion, text classification, question answering, and more.



Second, tailoring large language models requires minimal field training data.

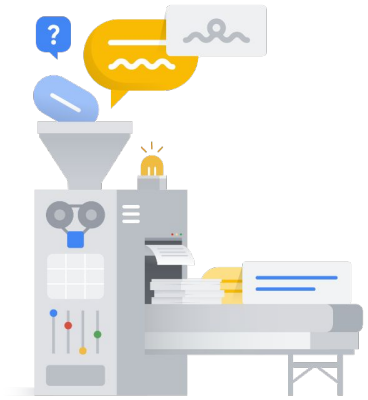
Even with little domain data, large language models achieve acceptable performance, and they can be used for few-shot or even zero-shot scenarios.

In machine learning, “few-shot” refers to training a model with minimal data and “zero-shot” implies that a model can recognize things that have not explicitly been taught in the training before.



Third, the performance of large language models continuously grows when you add more data and parameters.

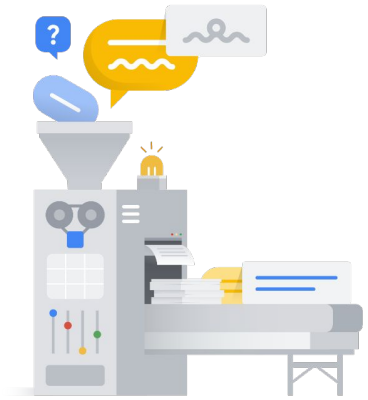
Large language models are:



Let’s take a moment to briefly summarize what you learned about LLMs.

Large language models are:

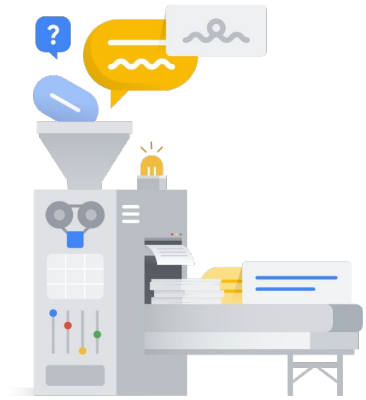
- Machine learning algorithms that can recognize, predict, and generate human languages.



LLMs are machine learning algorithms that can recognize, predict, and generate human languages.

Large language models are:

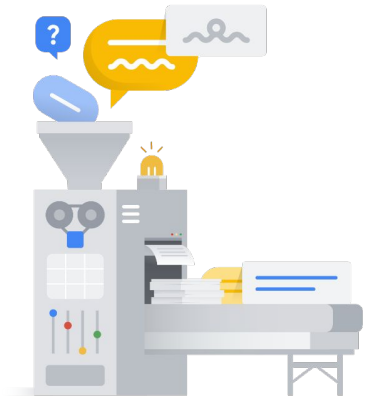
- Machine learning algorithms that can recognize, predict, and generate human languages.
- Pre-trained on large, text-based datasets, resulting in large models.



They're pre-trained on large, text-based datasets, resulting in large models,

Large language models are:

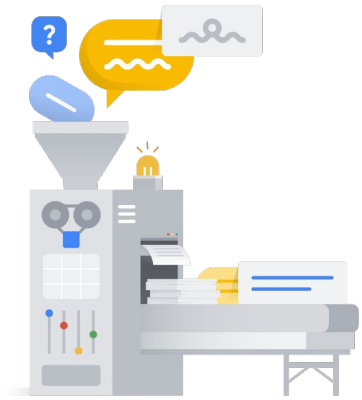
- Machine learning algorithms that can recognize, predict, and generate human languages.
- Pre-trained on large, text-based datasets, resulting in large models.
- Fine-tuned on a specific task.



and can then be fine-tuned on a specific task.

Large language models are:

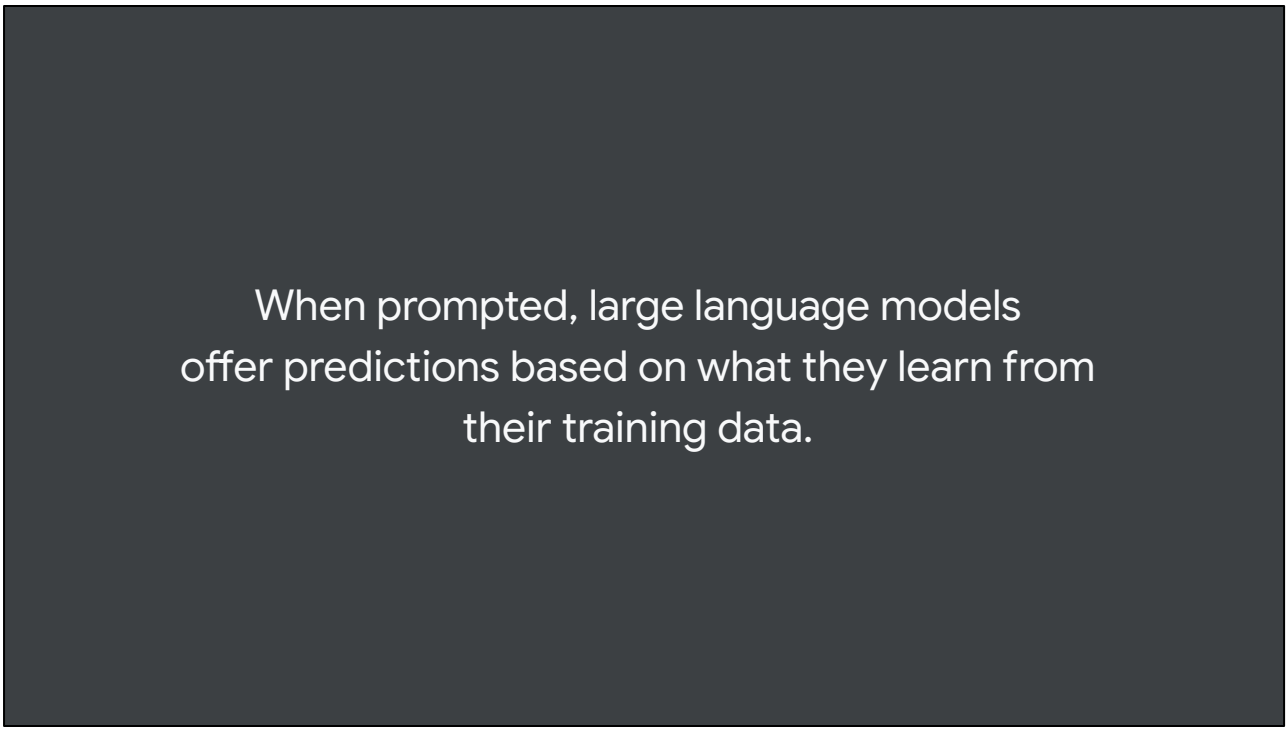
- Machine learning algorithms that can recognize, predict, and generate human languages.
- Pre-trained on large, text-based datasets, resulting in large models.
- Fine-tuned on a specific task.
- Sometimes referred to as large models (which process various data types) and generative AI (models that produce content).



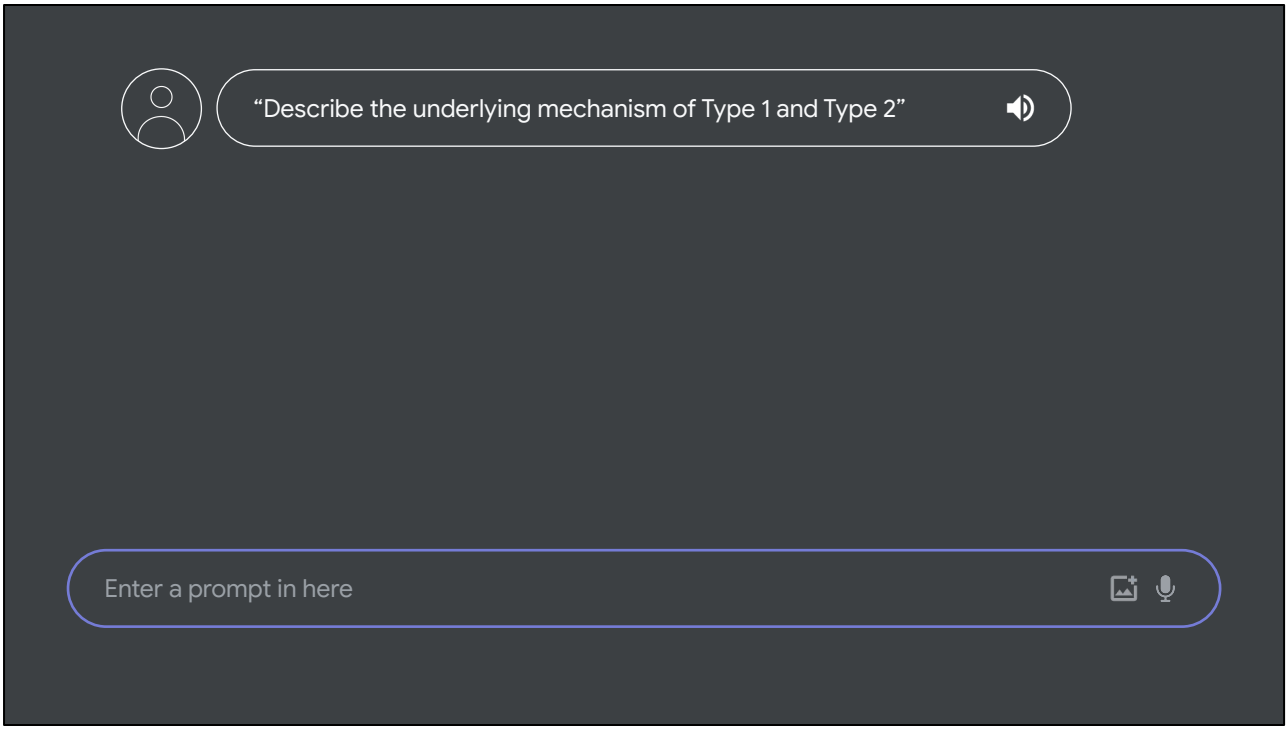
LLMs are sometimes referred to as large models (which process various data types) and generative AI (models that produce content).

Deck unlocked

V05
Prompts



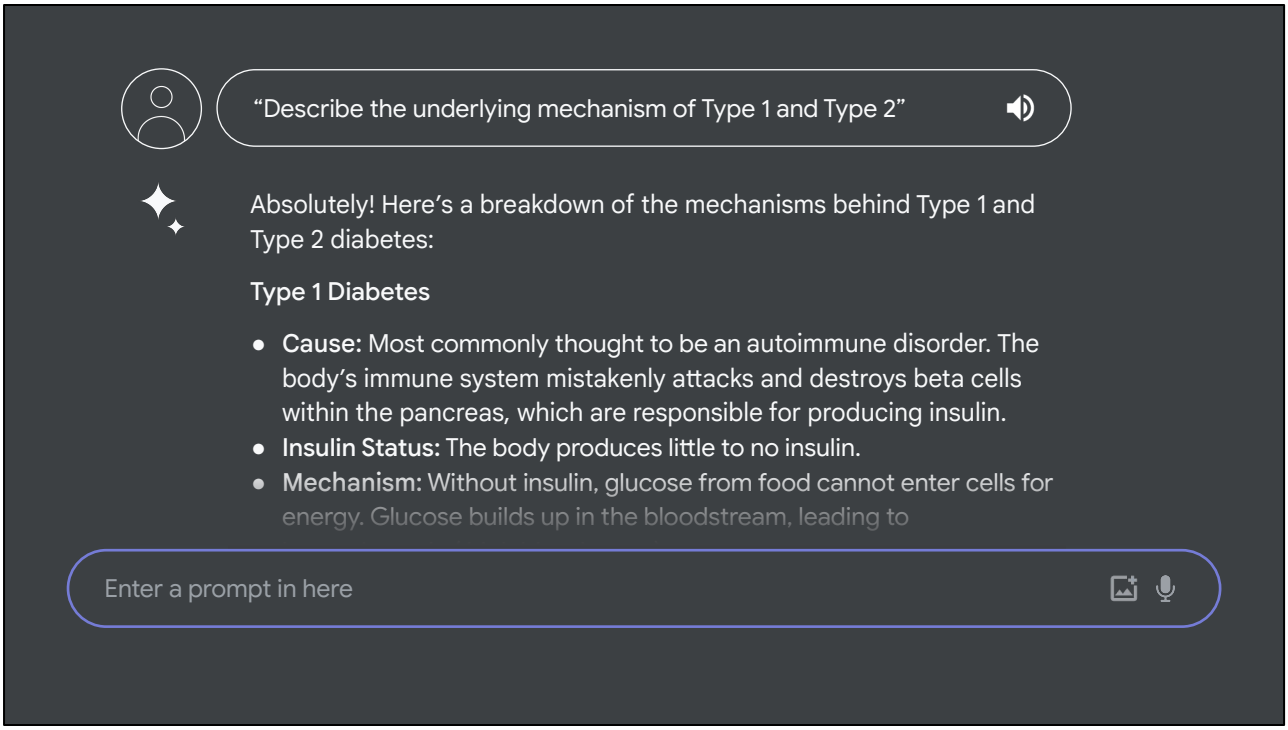
When prompted, large language models offer predictions based on what they learn from their training data. Let’s explore a few examples of this concept.



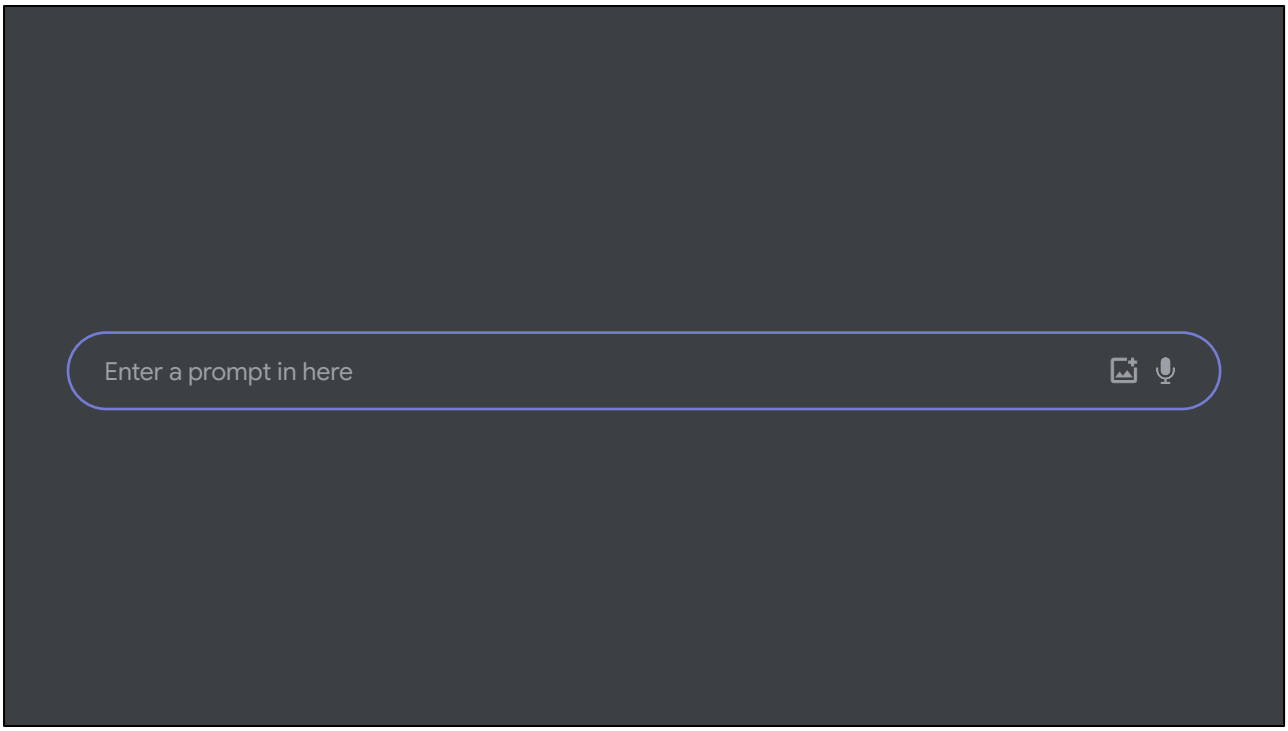
When prompted, large language models offer predictions based on what they learn from their training data. Let’s explore a few examples of this concept.

In this prompt, the sentence “Describe the underlying mechanism of Type 1 and Type 2” has been entered.

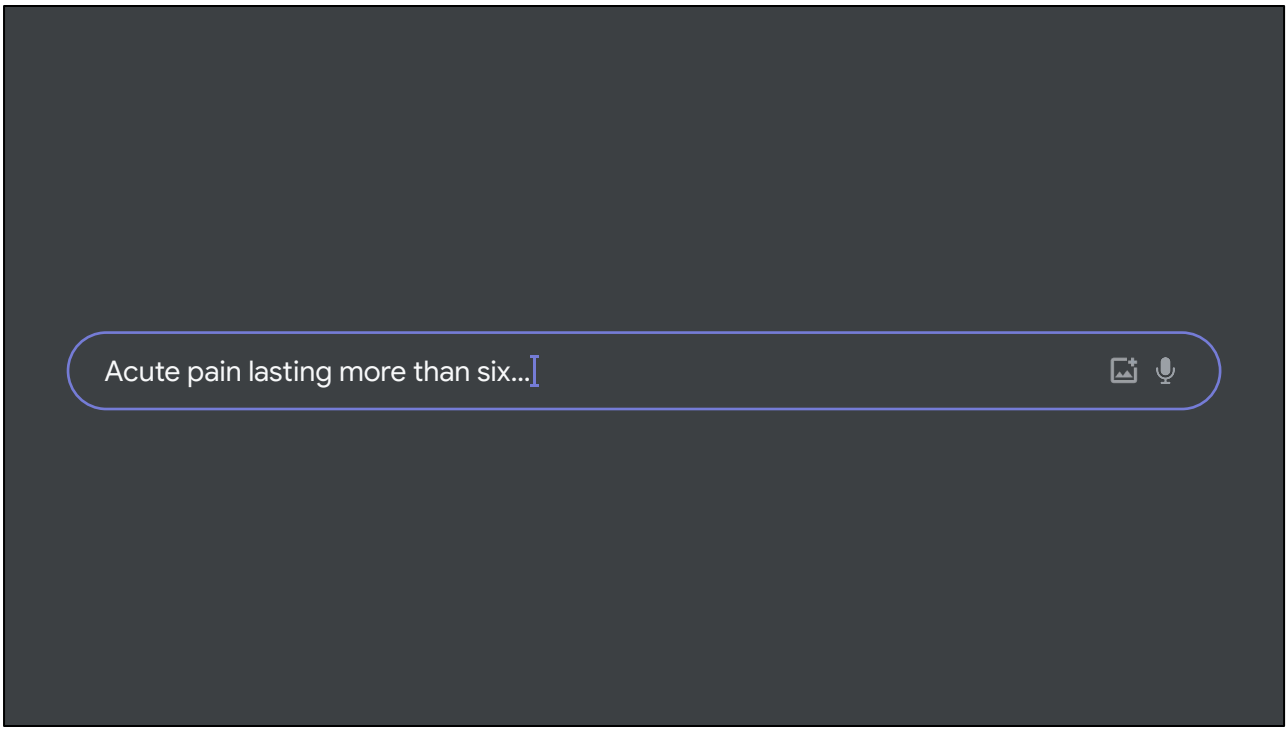
Based on what the LLM has learned from its training data, it offers predictions of how to complete the sentence.



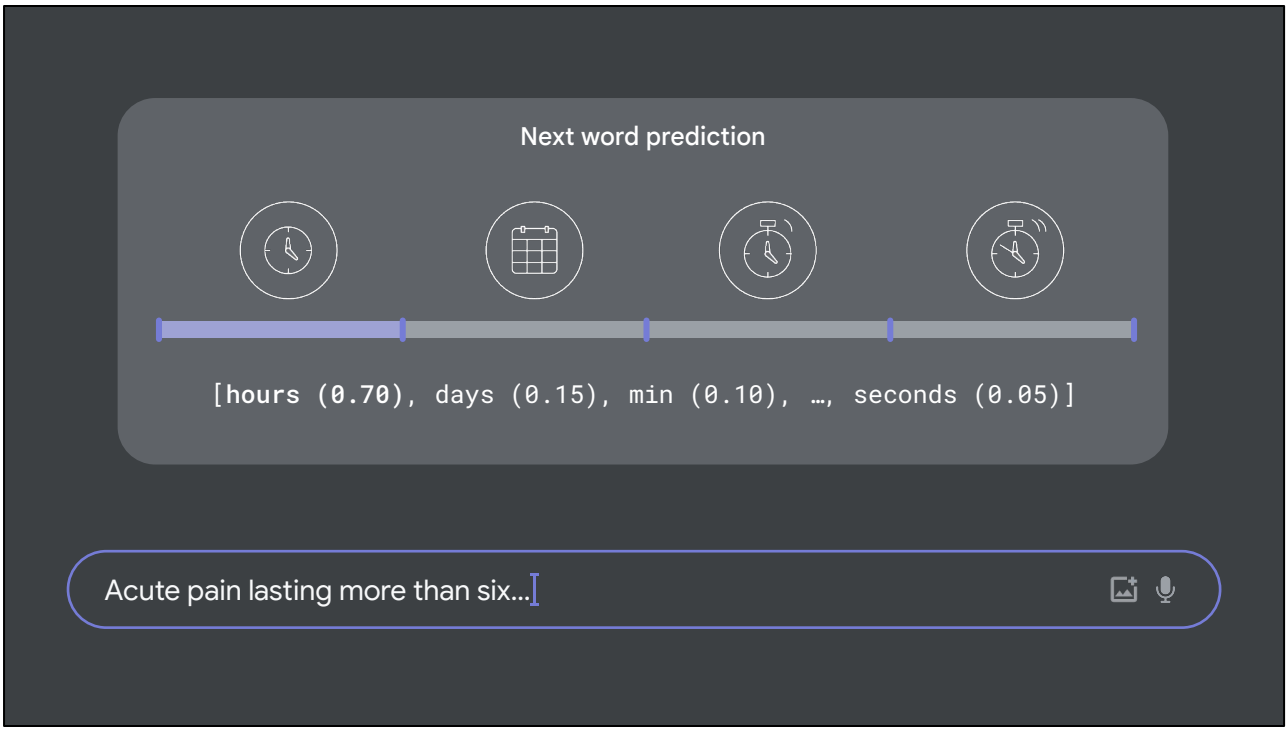
Not only does the LLM predict the next word, it goes on to explain the differences between type 1 and type 2 diabetes.



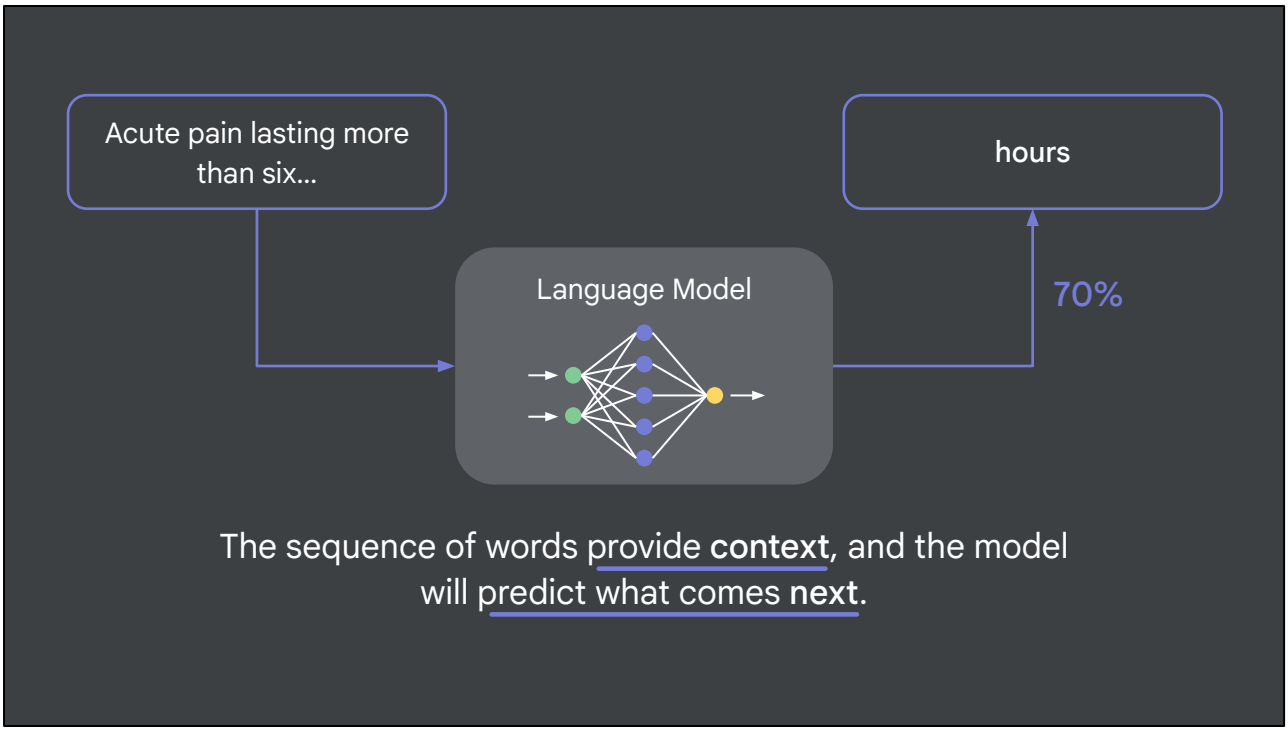
In this next example,



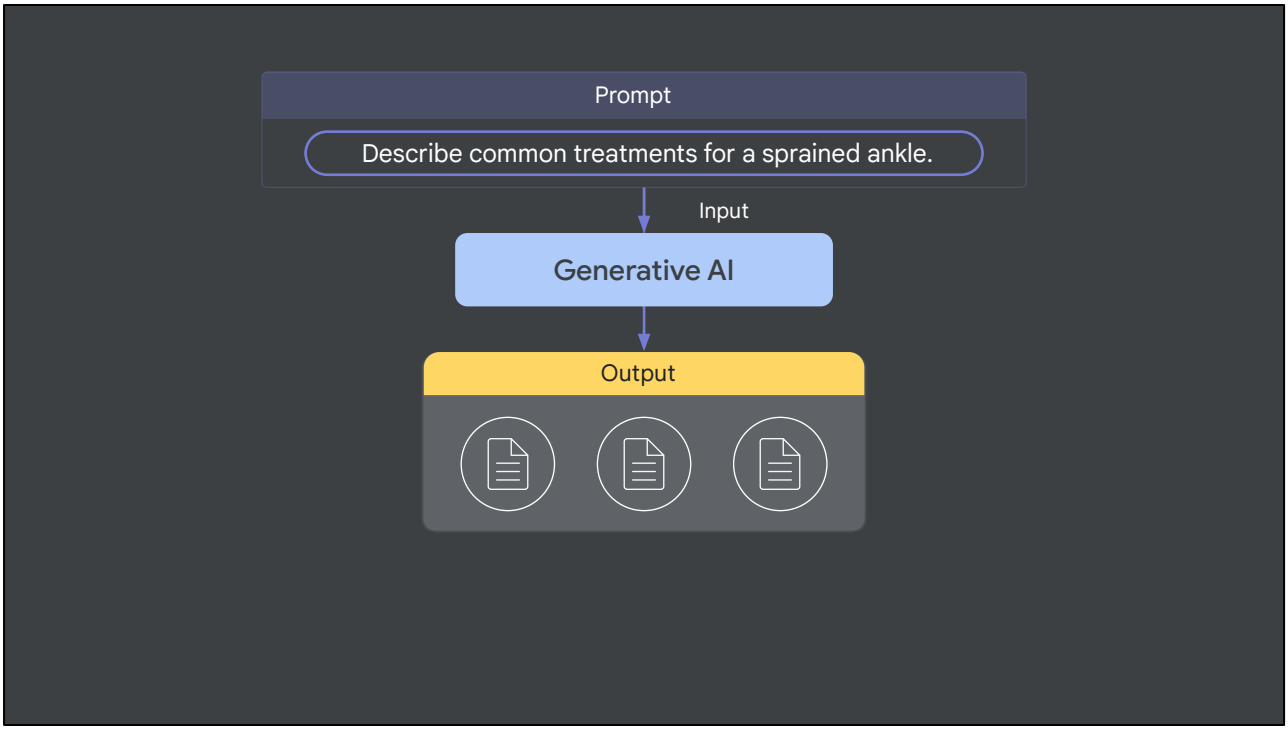
the prompt “acute pain lasting more than six –” is input into the model.



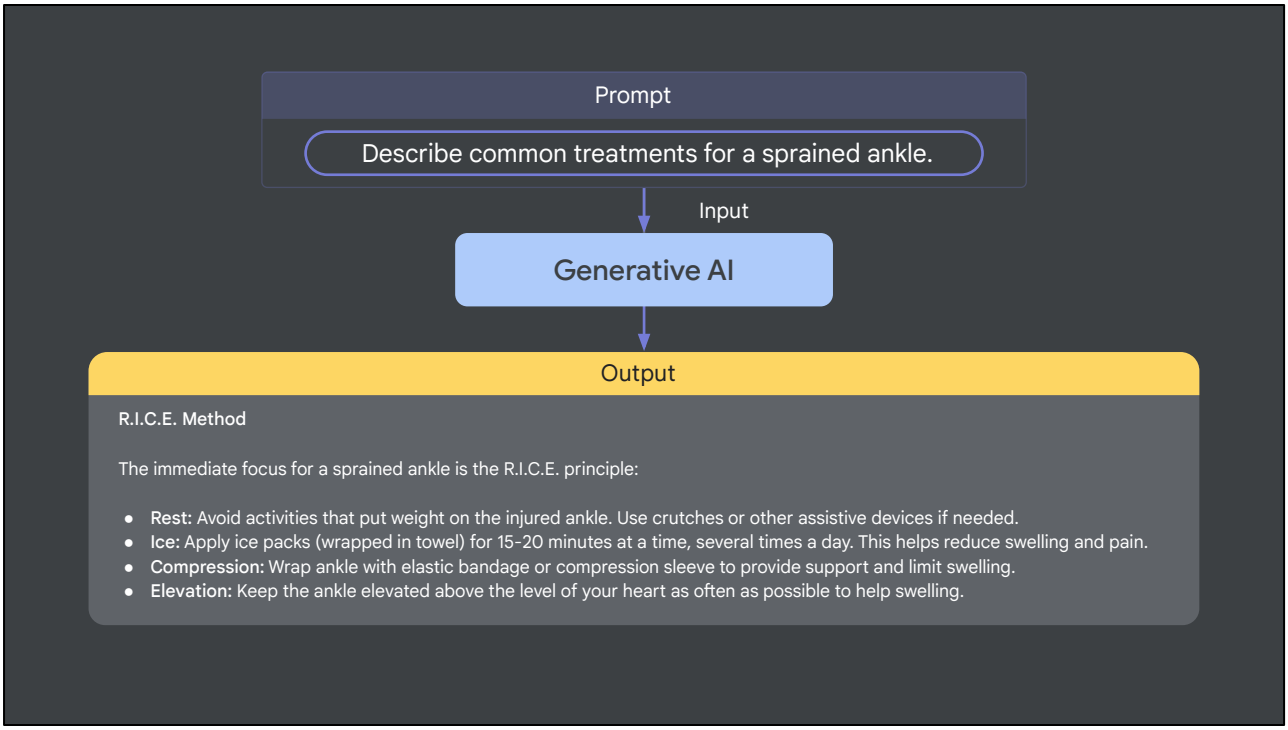
The probability that the word “hours” is selected is greater than the probability that “days” is selected.



How does the model know how to do this? Well, the sequence of words entered so far provide context, and the model will predict what comes next.

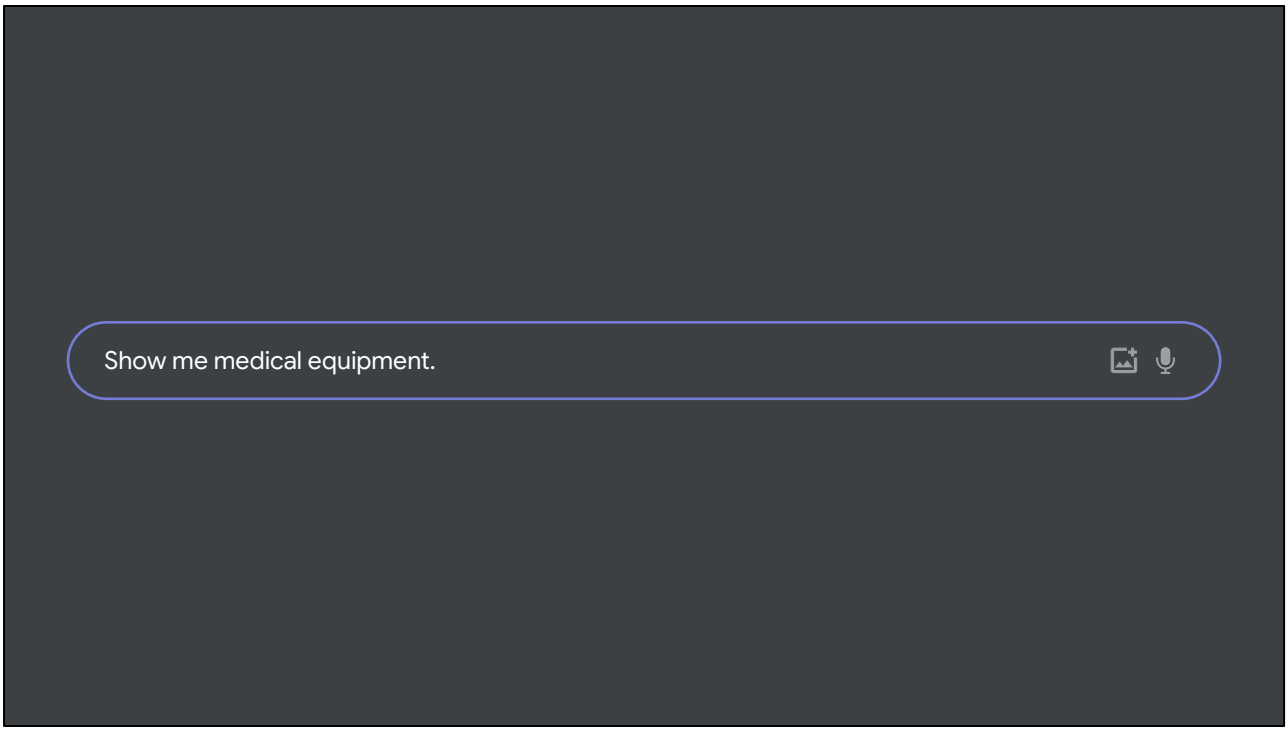


In addition to next-word prediction, large language models can also generate text based on your instructions. In this example, the model is prompted to describe common treatments for a sprained ankle.

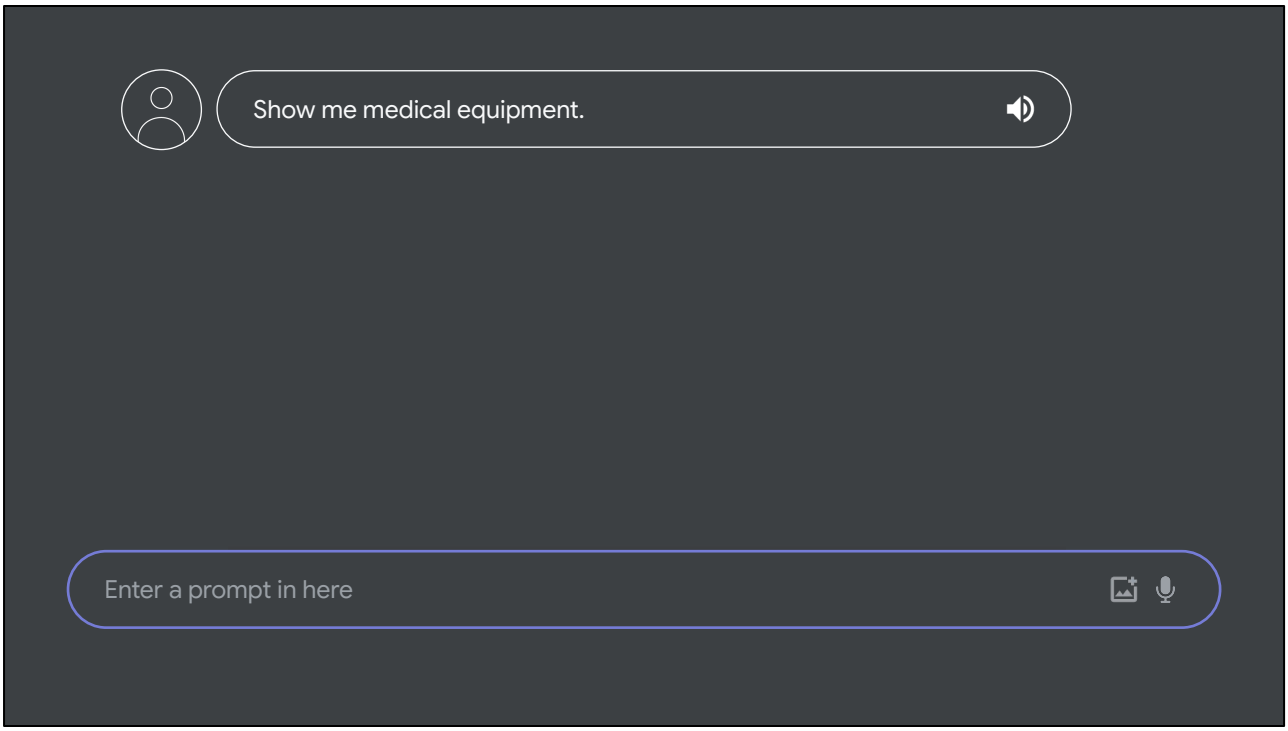


The model responds very quickly with information on the RICE method (or rest, ice, compression, and elevation).

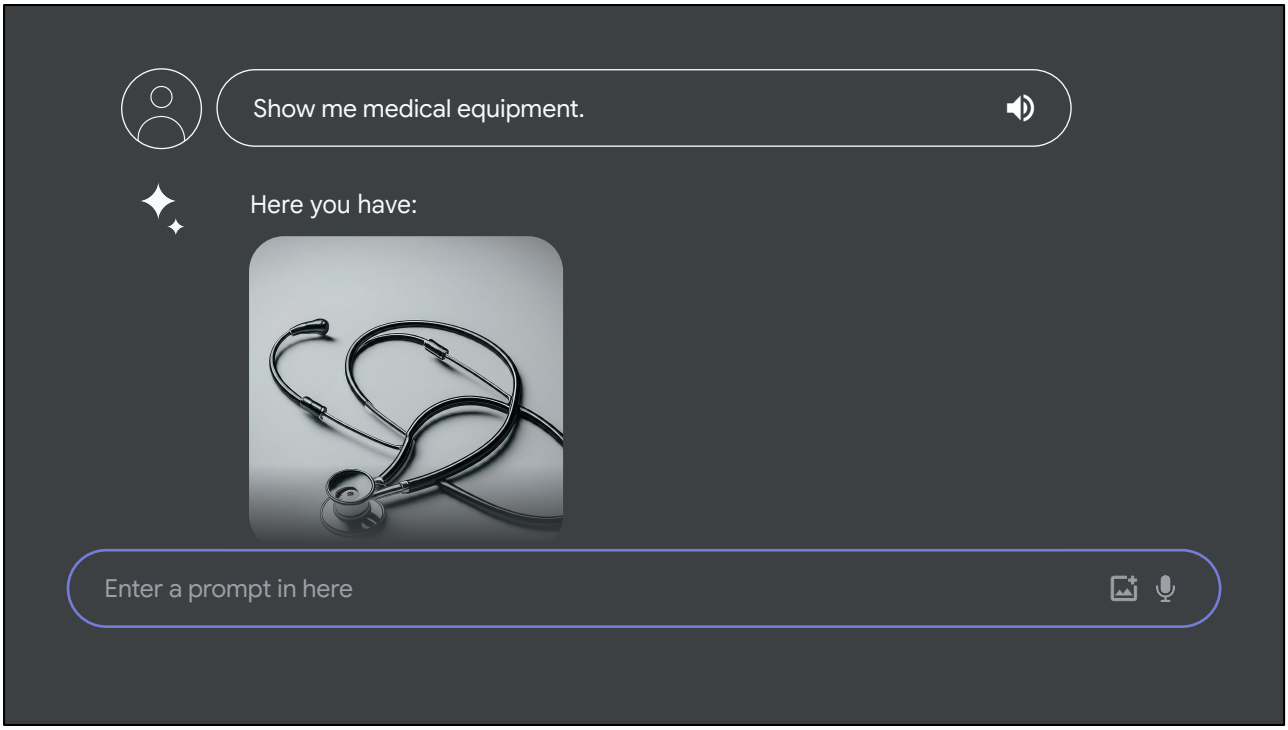
These instructions are similar to the ones patients receive in a hospital printout.



LLMs can also output images



in response to a text query.

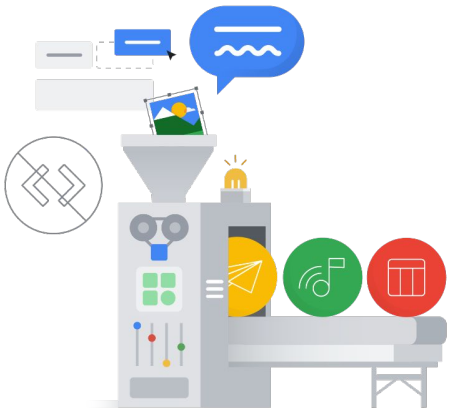


In this example the model is being prompted to show an example of medical equipment.

Deck unlocked

V06
Generative AI Tools

Vertex AI Agent Builder



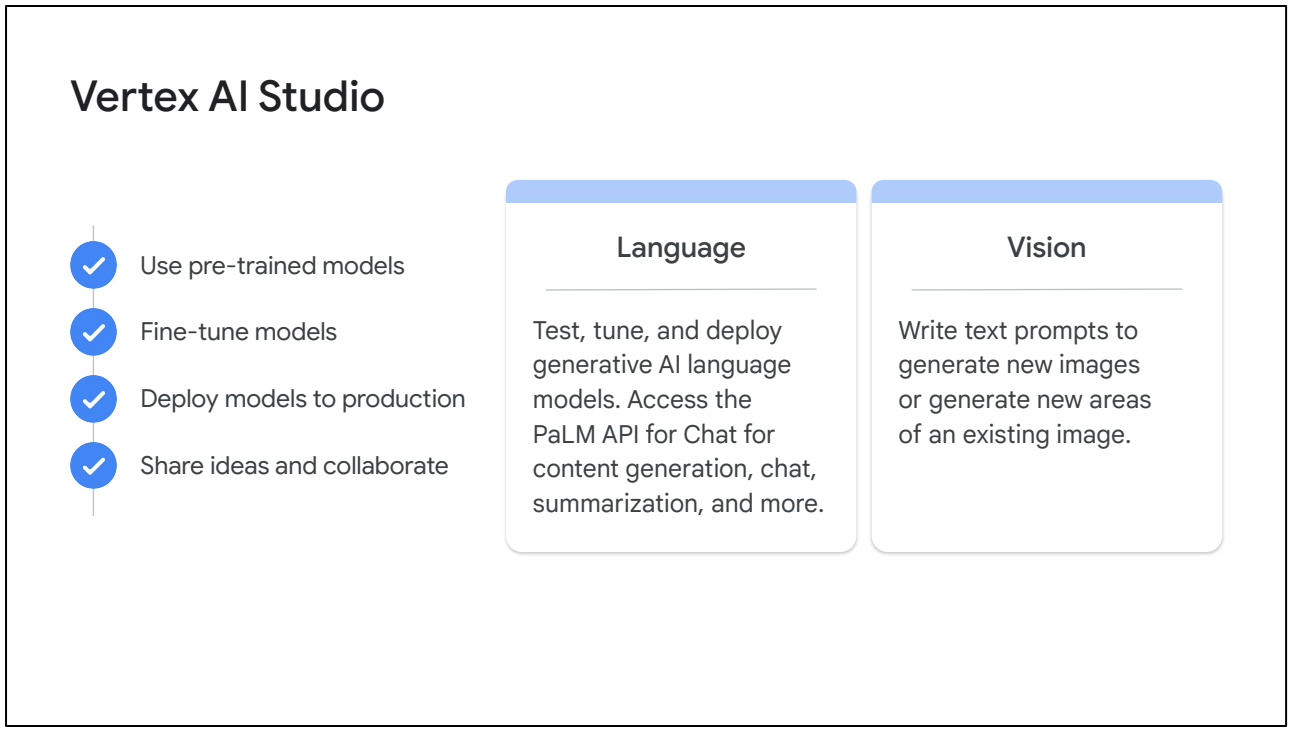
- No coding or machine learning experience required
- Build your own:
  - Chatbots.
  - Digital assistants.
  - Custom search engines.
  - Knowledge bases.
  - Training applications.
  - And more.

There are several Google Cloud tools available for you to use to build and customize Gen AI models.

You can build Gen AI search and conversations for customers and employees with Vertex AI Agent Builder without needing coding or machine learning experience.

You can use Vertex AI Agent Builder to build your own:

- Chatbots.
- Digital assistants.
- Custom search engines.
- Knowledge bases.
- Training applications.
- And more.



You can also use Vertex AI Studio to quickly explore and customize Gen AI models that can be leveraged in your applications on Google Cloud.

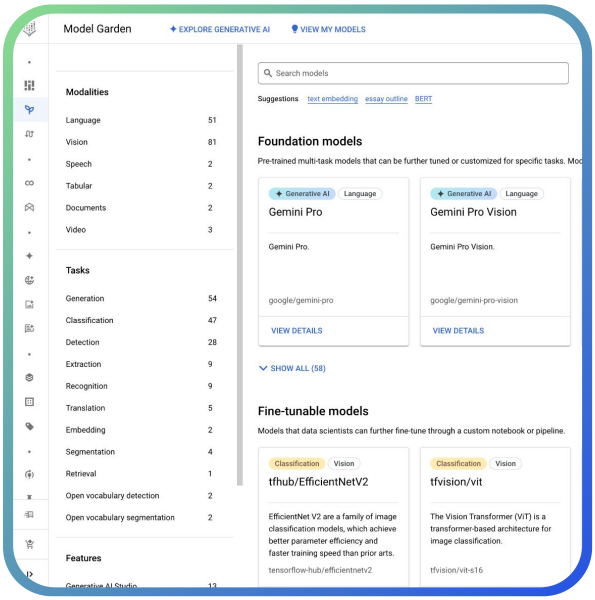
Vertex AI Studio provides a variety of tools and resources to help developers create and deploy Gen AI models, including:

- A library of pre-trained models.
- A tool for fine-tuning models.
- A tool for deploying models to production.
- And a community forum for developers to share ideas and collaborate.

Vertex AI Model Garden features a library of machine learning models that are consistently updated.

Gemini

Discover Gemini in Vertex Model Garden



Last but not least, Gemini is a multimodal AI model. Unlike traditional language models, it's not limited to understanding text alone. It can analyze images, understand the nuances of audio, and even interpret programming code.

This allows Gemini to perform complex tasks that were previously impossible for AI. Due to its advanced architecture, Gemini is incredibly adaptable and scalable, making it suitable for diverse applications.