1. how to Define generative AI
2. explain how generative AI Works
3. describe generative AI model types
4. describe generative AI applications

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

AI is a discipline like how physics is a discipline of science AI is a branch of computer science that deals with the creation of intelligent agents and our system systems that can reason learn and act autonomously

AI has to do with the theory and methods to build machines that think and act like humans pretty

Machine learning is a subfield of AI

It is a program or system that trains a model from input data .

The trained model can make useful predictions from new never-before seen data drawn from the same one used to train the model

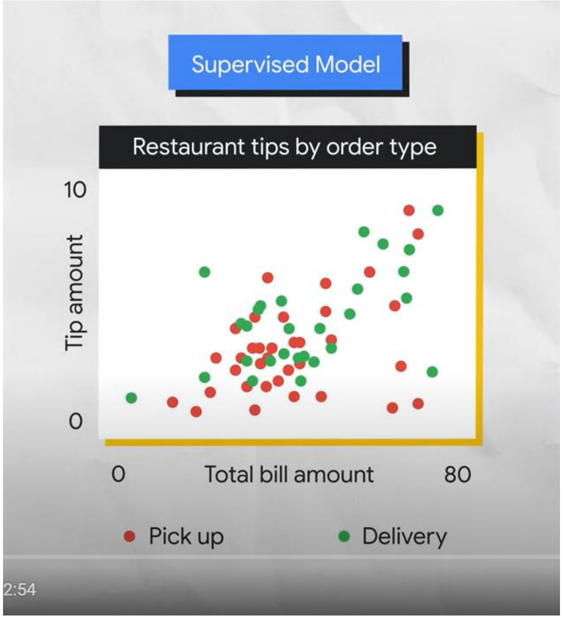
This means that machine learning gives the computer the ability to learn without explicit programming

Two most common classes of machine learning models are

1. unsupervised -- labelled data
2. supervised -- unlabelled data

The key difference between the two is that with supervised models we have labels. labelled data is data that comes with a tag like a name a type or a number.

unlabelled data is data that comes with no tag

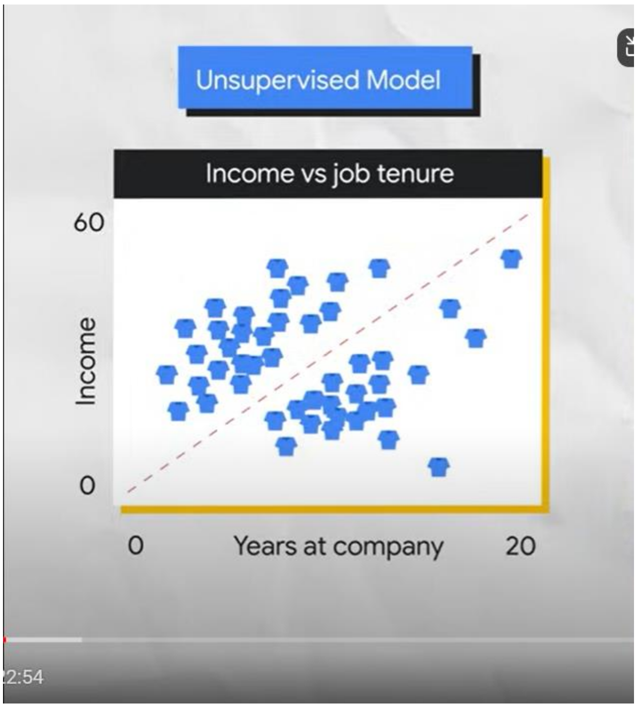


this graph is an example of the sort of problem a supervised model

let's say you're the owner of a restaurant what type of food do they serve .

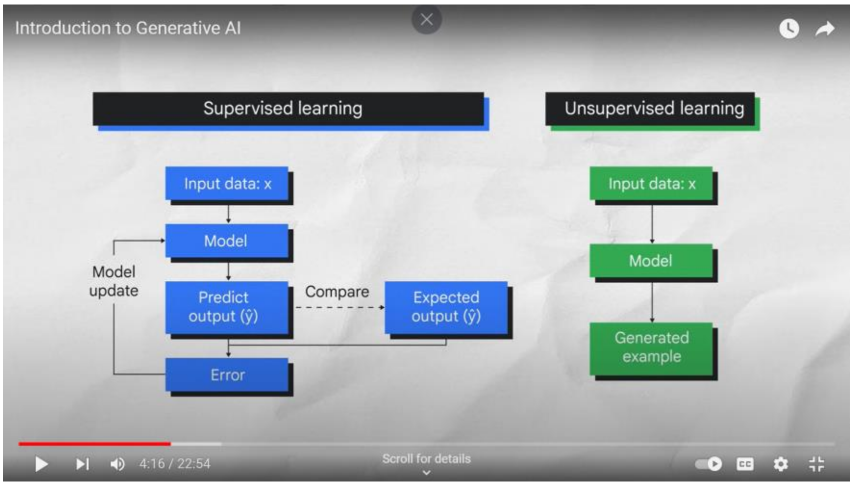
you have historical data of the bill amount and how much different people tipped based on the order type pickup or delivery in supervised learning.

the model learns from past examples to predict future values here. the model uses a total bill amount data to predict the future tip amount based on whether an order was picked up or delivered .



this is an example of the sort of problem that an unsupervised

unsupervised problems are all about discovery about looking at the raw data and seeing if it naturally falls into groups this is a good start but let's go a little deeper to show this difference graphically because understanding these Concepts is the foundation for your understanding of generative AI



in supervised learning testing data

values X our input into the model the

model outputs a prediction and Compares

it to the training data used to train

the model

if the predicted test data

values and actual training data values

are far apart that is called error

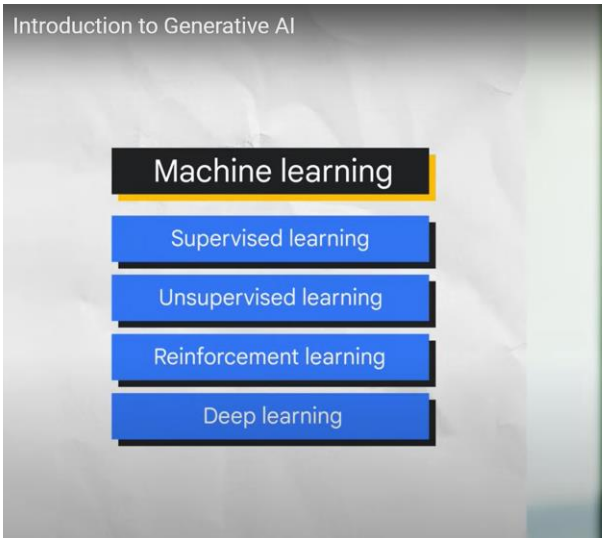
the

model tries to reduce this error until

the predicted and actual values are

closer together this is a classic

optimization



Introduction to Generative AI

Machine learning

Supervised learning

Unsupervised learning

Reinforcement learning

Deep learning

let's briefly explore where deep

learning fits as a subset of machine

learning methods

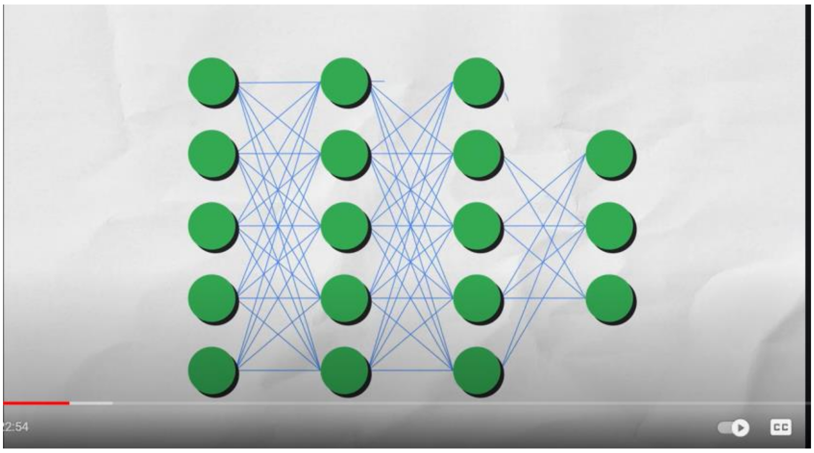
deep learning is a type of

machine learning that uses artificial

neural networks allowing them to process

more complex patterns than machine

learning



artificial neural networks are

inspired by the human brain

like your brain they are made up of

many interconnected nodes or neurons

that can learn to perform tasks by

processing data and making

predictions

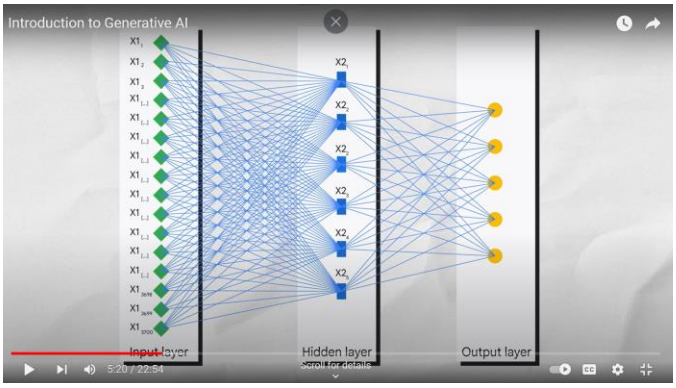
deep learning models

typically have many layers of neurons

which allows them to learn more complex

patterns than traditional machine

learning models



neural networks can use both

labeled and unlabeled data this is

called semi-supervised learning

in semi

supervised learning a neural network is

trained on a small amount of labeled

data

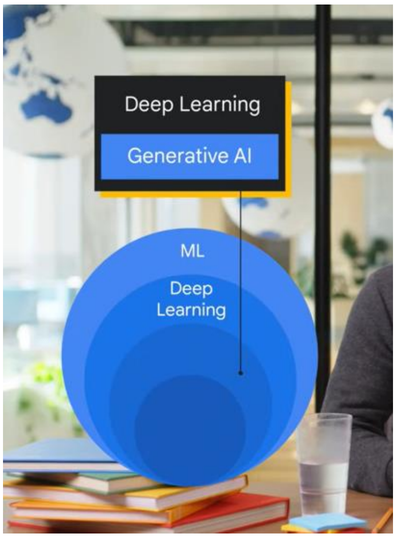
and a large amount of unlabeled

data

the labeled data helps the neural network to learn the basic concepts of the tasks

while the unlabeled data helps the neural network to generalize to new examples

gen AI is a subset of deep learning which means it uses artificial neural networks can process both labeled and unlabeled data using supervised unsupervised and semi-supervised methods



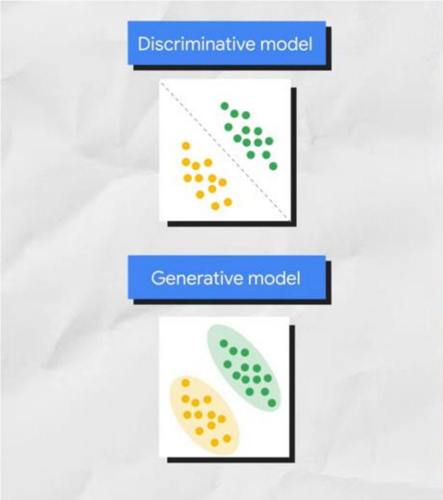
large language models are also a subset of deep learning deep learning models or machine learning

models in general can be divided into

two types

generative

discriminative



a discriminative model is

a type of model that is used to classify

or predict labels for data points

discriminative models are typically

trained on the data set of labeled data

points

they learn the relationship

between the features of the data points

and the labels once a discriminative model is

trained it can be used to predict the

label for new data

points

a generative model generates new

data instances based on a learned

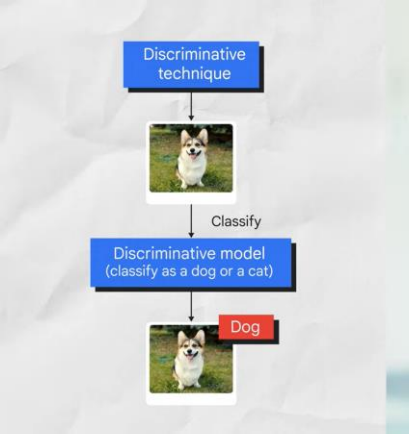
probability distribution of existing

data

generative models generate new

contents

take this example



here the

discriminative model learns the

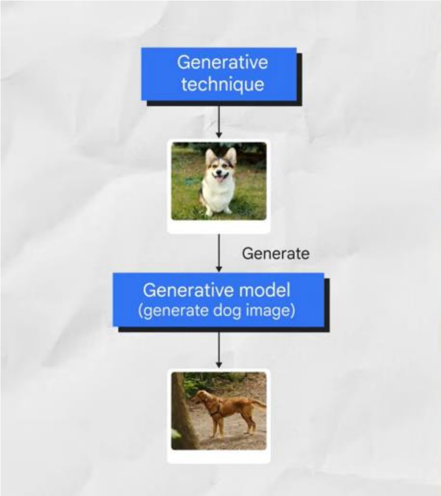
conditional probability distribution or

the probability of Y

our output given X

our input that this is a dog and

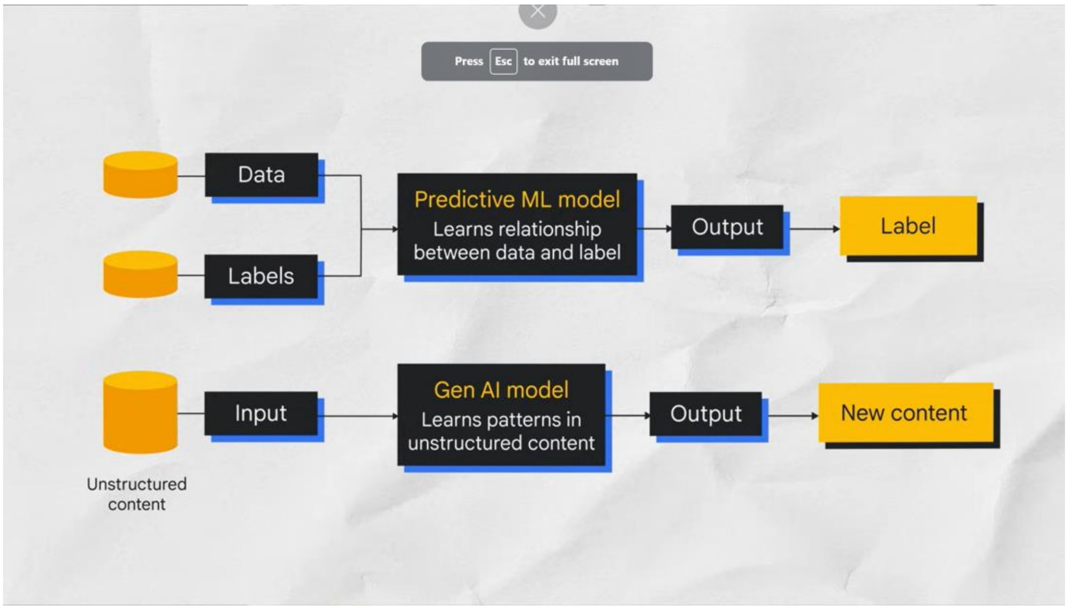
classifies it as a dog and not a cat



the generative model learns The Joint probability distribution or the probability of X and Y P of x y and predicts the conditional probability that this is a dog and can then generate a picture of a dog

to summarize generative models can generate new data instances and discriminative models discriminate between different kinds of data instances

one more quick example



Data

Labels

Predictive ML model
Learns relationship between data and label

Output

Label

Input

Unstructured content

Gen AI model
Learns patterns in unstructured content

Output

New content

the top

image shows a traditional machine

learning model which attempts to learn

the relationship between the data and

the label or what you want to predict

the bottom image shows a generative AI

model which attempts to learn patterns

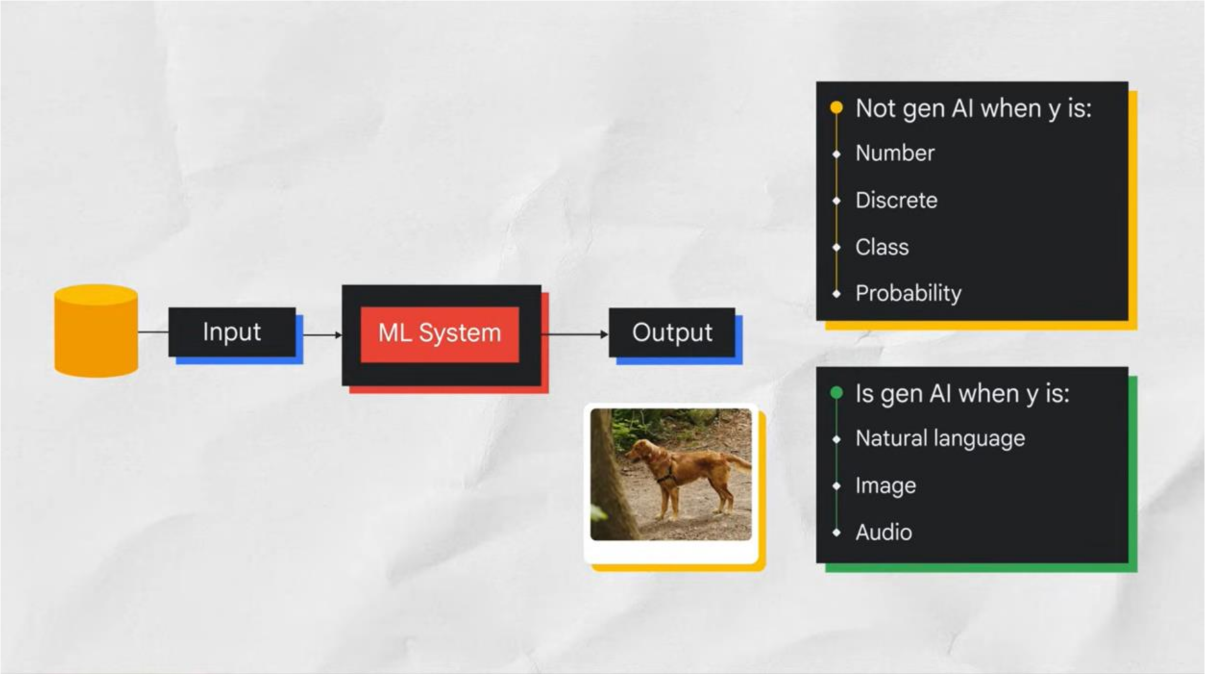
on content so that it can generate new content

a good way

to distinguish between what is Gen and

what is

not



it is not gen when the output or Y

or label is a number or a class for

example spam or not spam or a

probability it is Gen when the output is

natural language like speech or text

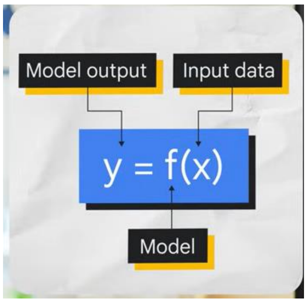
audio or an image l

for

example let's get a little mathy to

really show the difference visualizing

this mathematically would look like this



equation calculates the

dependent output of a process given

different inputs

inputs are the

data value files

text files audio files or image files

like Fred

so the model output is a

function of all the inputs

if the Y is a

number like predicted sales it is not

generative AI

if Y is a sentence like

Define sales it is generative as the

question would elicit a text

response

the response will be based on

all the massive large data the model was

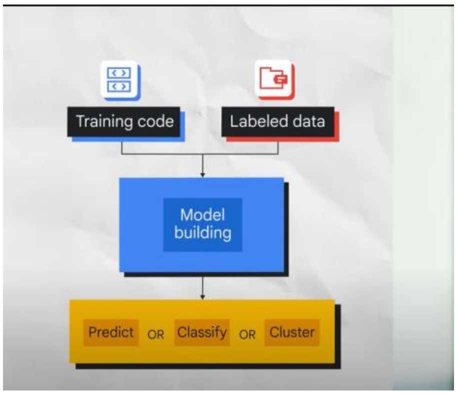
already trained on

so the traditional ml

supervised learning process takes

training code and label data to build a

model



depending on the use case or

problem the model can give you a

prediction classify something or cluster

something

the generative AI process can

take training code labeled data and

unlabeled data of all data types and

build a foundation model

the foundation

model can then generate new content

it

can generate text code images audio

video

we've come a long way

from traditional programming to neural

networks to generative

models

in traditional programming we
used to have to hardcode the rules for
distinguishing

in the wave of neural networks we
could give the networks pictures of cats
and dogs and ask is this a cat and it
would predict a cat or not a cat

what's
really cool is that in the generative
wave we as users can generate our own
content whether it be text images audio
video or more

for example models like
Palm or Pathways language model or
Lambda language model for dialogue
applications inset very very large data
from multiple sources across the
internet and build Foundation language
models
we can use simply by asking a
question whether typing it into a prompt
or verbally talking into the prompt
itself

so when you ask it what's a cat

it can give you everything it's learned

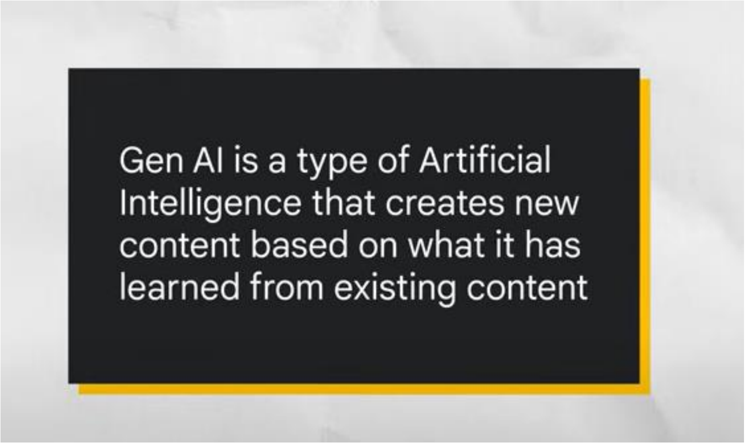
about a

cat

now let's make things a little more

formal with an official definition what

is generative



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 prompt gen uses a

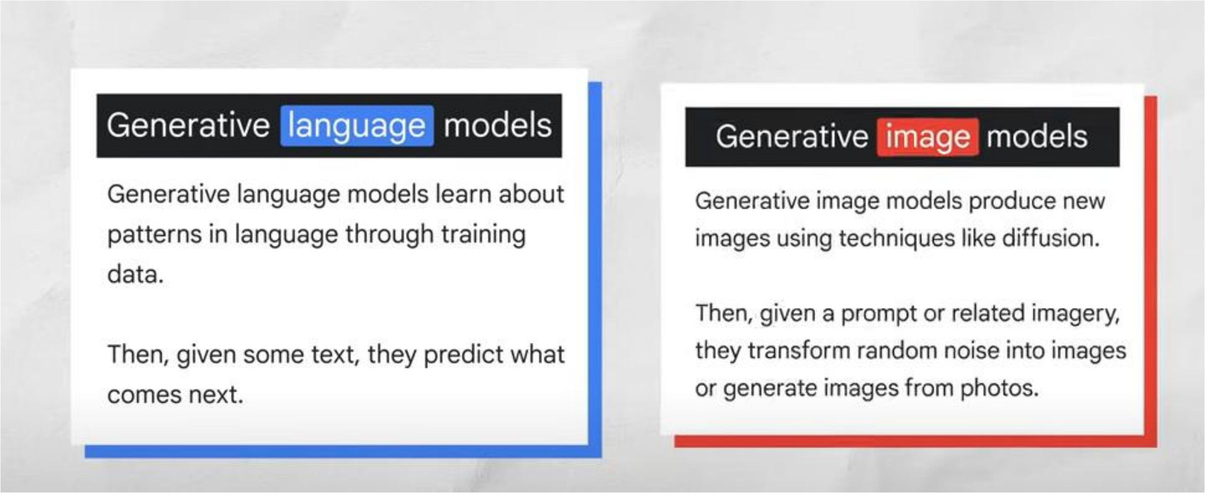
statistical model to predict what an

expected response might be

and this

generates new content

it learns the underlying structure of the data and can then generate new samples that are similar to the data it was trained on



Generative language models

Generative language models learn about patterns in language through training data.

Then, given some text, they predict what comes next.

Generative image models

Generative image models produce new images using techniques like diffusion.

Then, given a prompt or related imagery, they transform random noise into images or generate images from photos.

like I mentioned earlier a generative language model can take what has learned from the examples it's been shown and creat something entirely new based on that information

that's why we use the word generative but large language models which generate novel combinations of texts in the form of natural sounding language are only one type of generative AI a generative image model takes an

image as input and can output text

another image or video for example under

the output text you can get visual

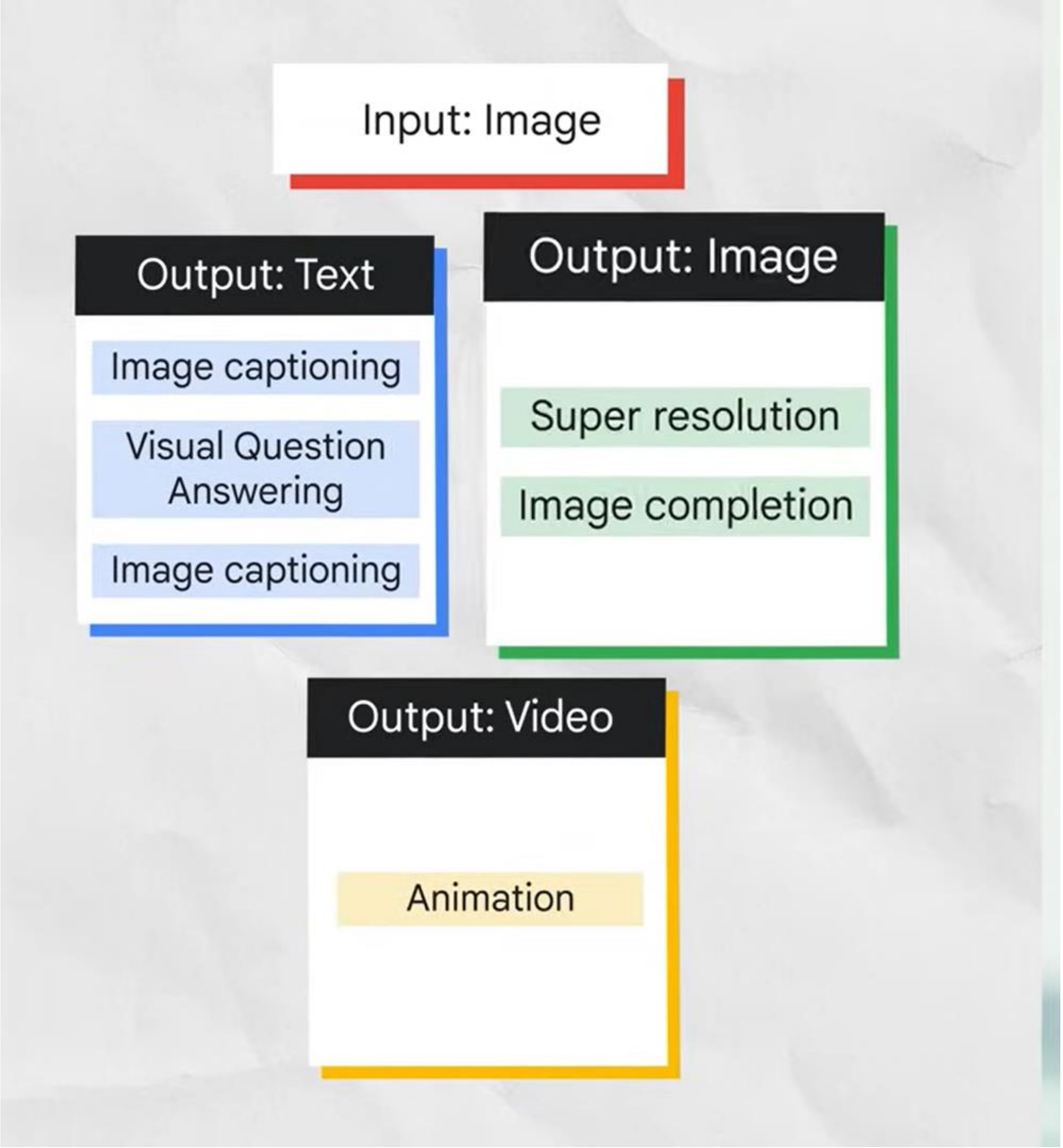
question and answering while under

output image an image completion is

generated and under output video

animation is

generated



a generative language model

takes text as input and can output more

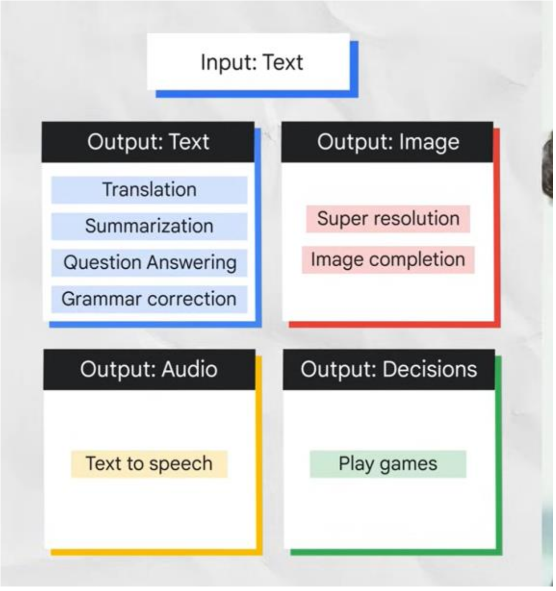
text an image audio or decisions for

example under the output text question

and answering is generated and under

output image a video is

generated



generative language models are

pattern matching systems they learn

about patterns based on the data that

Gemini which is trained on a

massive amount of Text data and it's

able to communicate and generate

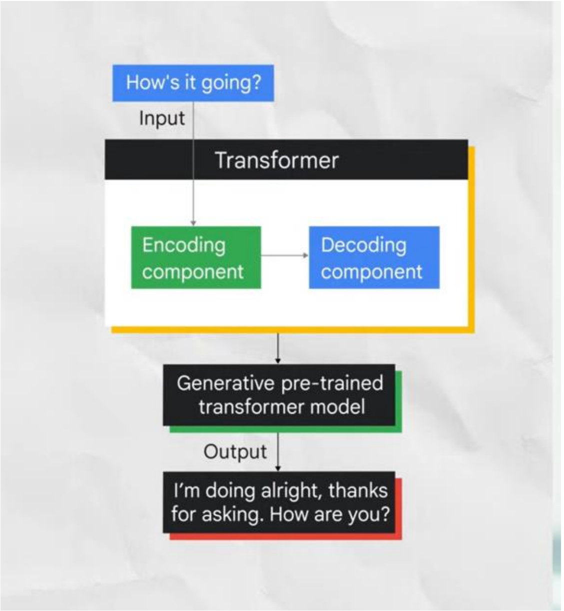
humanlike text in response to a wide

range of prompts and questions

the power of generative AI

comes from the use of

Transformers


How's it going?
Input
Transformer
Encoding component
Decoding component
Generative pre-trained transformer model
Output
I'm doing alright, thanks for asking. How are you?

Transformers produced the

2018 revolution in natural language

processing at a high level

a Transformer

model consists of an encoder and a

decoder

the encoder encodes the input

sequence and passes it to the decoder

which learns how to decode the

representations for a relevant

task

sometimes Transformers run into

issues though hallucinations are words

or phrases that are generated by the

model that are often nonsensical or

grammatically incorrect

hallucinations can be caused by a number

of factors like when the model is not

trained on enough data it's trained on

noisy or dirty data is not given enough

context or is not given enough

constraints hallucinations can be a

problem for Transformers because they

can make the output text difficult to

understand they can also make the model

more likely to generate incorrect or

misleading information so put simply

hallucinations are

bad

Hallucinations make the output text difficult to understand and make the model more likely to generate incorrect or misleading information

Hallucinations are bad

let's pivot slightly and talk about

prompts a prompt is a short piece of

text that is given to a large language

model or llm as input

and it can be used

to control the output of the model in a

variety of ways

prompted design is the

process of creating a prompt that will

generate the desired output from an

llm

like I mentioned earlier generative

AI depends a lot on the training data

that you have fed into it

it analyzes

the patterns and structures of the input

data and thus

learns

but with access to a browser

based prompt you the user can generate

your own

content so let's talk a little bit about

the model types available to us when

text is our input and how they can be

helpful in solving problems

the first is text to text text to

text models take a natural language

input and produce text output

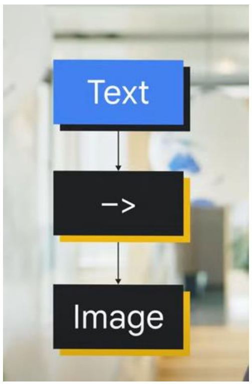
these

models are trained to learn the mapping

between a pair of text for example

translating from one language to

others



next we have 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

there's also text to video and text to 3D text to

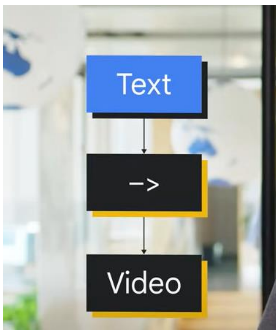
video models aim to generate a video

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

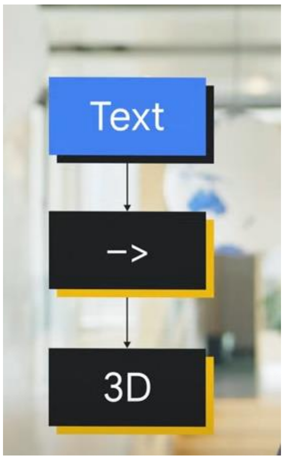


similarly text of 3D models

generate threedimensional objects that

correspond to a user's text description

for use in games or other 3D worlds

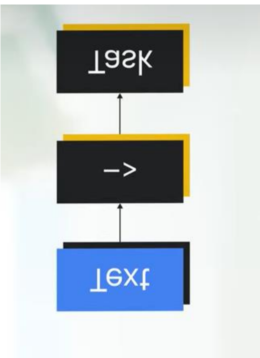


finally there's text to task text to

task models are trained to perform a

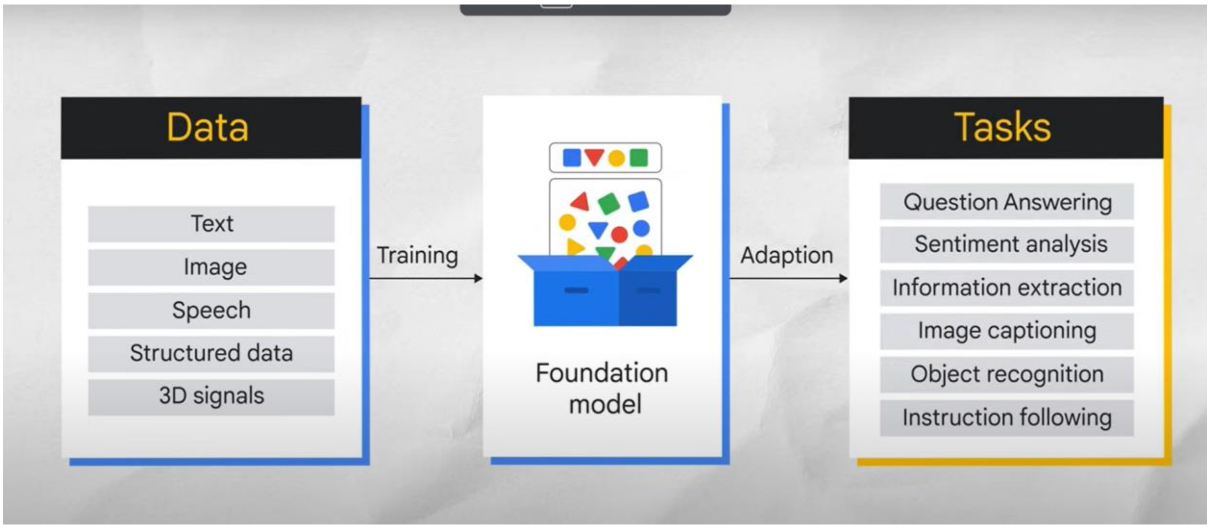
defined task or action based on text

input



this task can be a wide range of actions such as answering a question performing a search making a prediction or taking some sort of action for example a text to taxt model could be trained to navigate a web user interface or make changes to a doc through a graphical user interface

another model that's larger than those I mentioned is a foundation model which is a large AI model pre-trained on a vast quantity of data designed to be adapted or fine-tuned to a wide range of Downstream tasks such as sentiment analysis image captioning and object recognition



Data

Text

Image

Speech

Structured data

3D signals

Training

Foundation model

Adaption

Tasks

Question Answering

Sentiment analysis

Information extraction

Image captioning

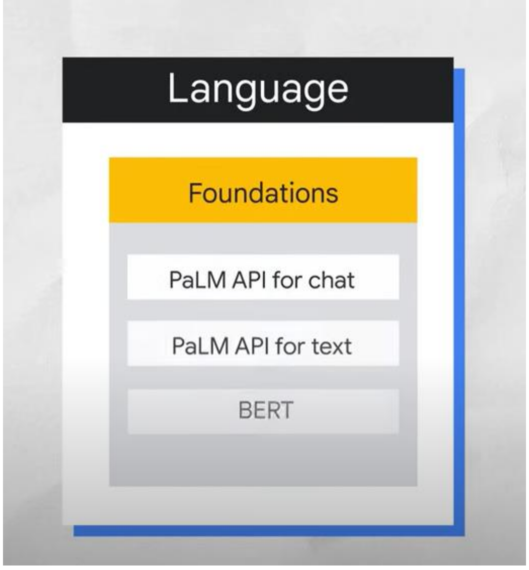
Object recognition

Instruction following

Foundation models have the potential to revolutionize many Industries including Healthcare finance and customer service they can even be used to detect fraud and provide personalized customer support

if you're looking for foundation models

vertex AI offers a model Garden that includes Foundation models



Language

Foundations

PaLM API for chat

PaLM API for text

BERT

the

language Foundation models include Palm

API for chat and text

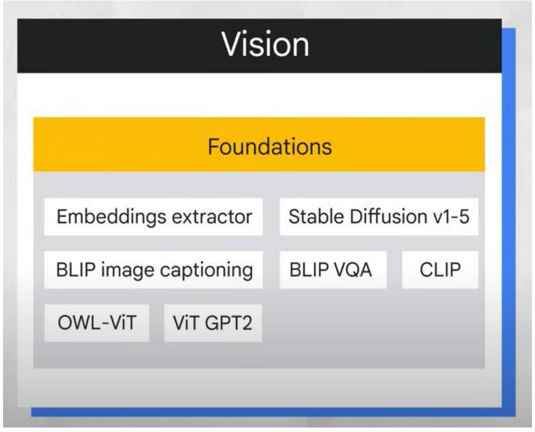
the vision

Foundation models include stable

diffusion which have been shown to be

effective at generating high quality

images from text



Vision

Foundations

Embeddings extractor

Stable Diffusion v1-5

BLIP image captioning

BLIP VQA

CLIP

OWL-ViT

ViT GPT2

you have a use

case where you need to gather sentiments

about how your customers feel about your

product or service you can use the

classification task sentiment analys

task model same for vision tasks if you

need to perform occupancy analytics

some examples of

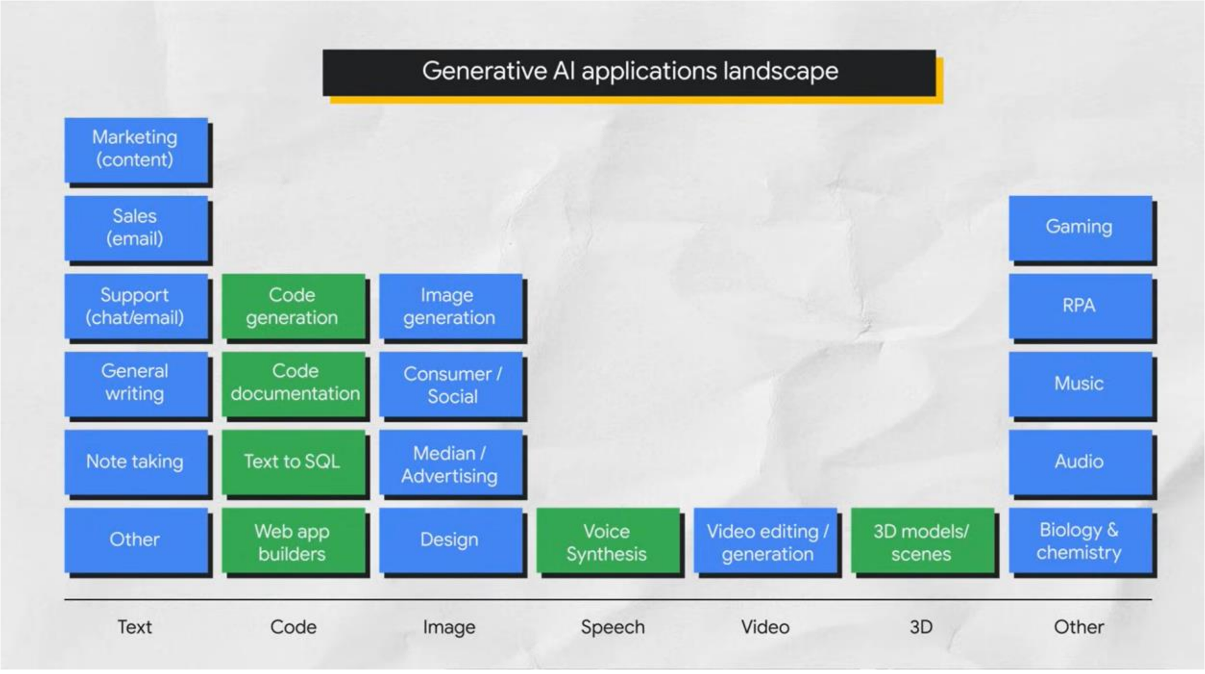
foundation models we can use

shown here are generative AI

applications

you can see there's quite a

lot



Generative AI applications landscape

Text	Code	Image	Speech	Video	3D	Other
Marketing (content)						
Sales (email)						Gaming
Support (chat/email)	Code generation	Image generation				RPA
General writing	Code documentation	Consumer / Social				Music
Note taking	Text to SQL	Median / Advertising				Audio
Other	Web app builders	Design	Voice Synthesis	Video editing / generation	3D models/ scenes	Biology & chemistry

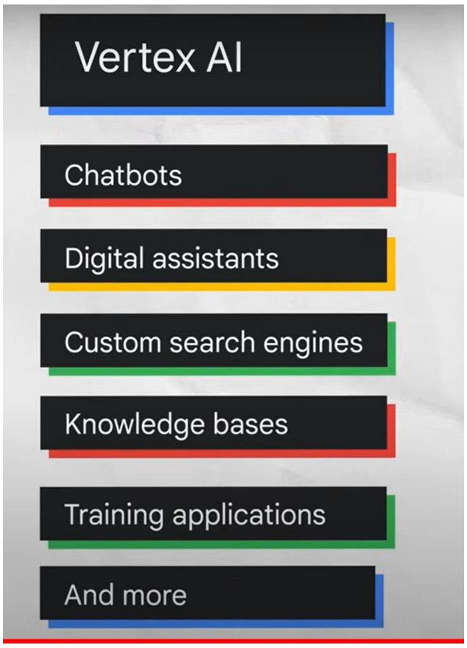
I'm going to
tell you about three other ways Google
Cloud can help you get more out of
generative AI

the first is vertex AI
Studio
vertex AI Studio lets you quickly
explore and customize generative AI
models that you can leverage in your
applications on Google Cloud
vertex AI
Studio helps developers create and
deploy generative AI models by providing
a variety of tools and resources that
make it easy to get
started

for example there is a library
of pre-trained models tool for
fine-tuning models tool for deploying
models

production and Community forum
for developers to share ideas and
collaborate

vertex AI can help



Vertex AI

Chatbots

Digital assistants

Custom search engines

Knowledge bases

Training applications

And more

lastly we have Palm API

Palm API lets

you test and experiment with Google's

large language models and gen tools to

make prototyping quick and more

accessible

developers can integrate Palm

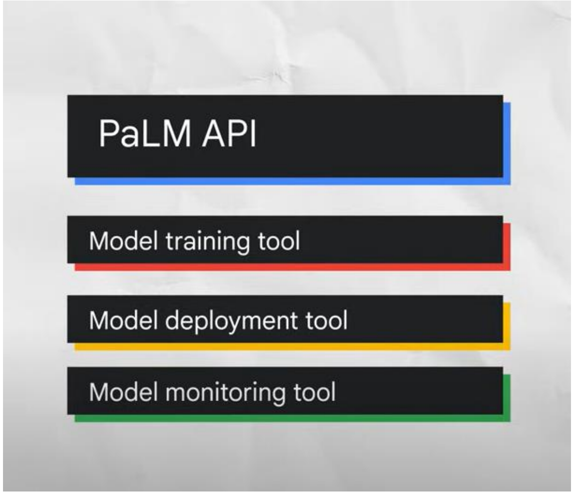
API with maker suite and use it to

access the API using graphical user

interface

the suite includes a number of

different tools such



PaLM API

Model training tool

Model deployment tool

Model monitoring tool

what do these tools

do I'm so glad you asked the model

training tool helps developers train ml

models on their data using different

algorithms

the model deployment tool

helps developers deploy ml models to

production with a number of different

deployment options

the model monitoring

tool helps developers monitor the

performance of their ml models in

production using a dashboard and a

number of different

metrics

lastly there is Gemini a

multimodal AI model unlike traditional

language models it's not limited to

understanding text alone

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

model Garden is
continuously updated to include new
models