Getting started with Gemini in Vertex AI

Introduction

- Experience in Analytics, MLOPs and data engineering
- Enthusiastic learner and tech community contributor

Navaneethan Gopal

Data Science Manager @ Ingram Micro

LinkedIn – https://www.linkedin.com/in/navaneethan-gopal-9b101612/
Twitter - https://twitter.com/NavaneethhGopal

IN GRAM MICRO ®

AI WILL NOT REPLACE YOU IN YOUR JOB, SOMEONE WHO USES AI WILL

Agenda

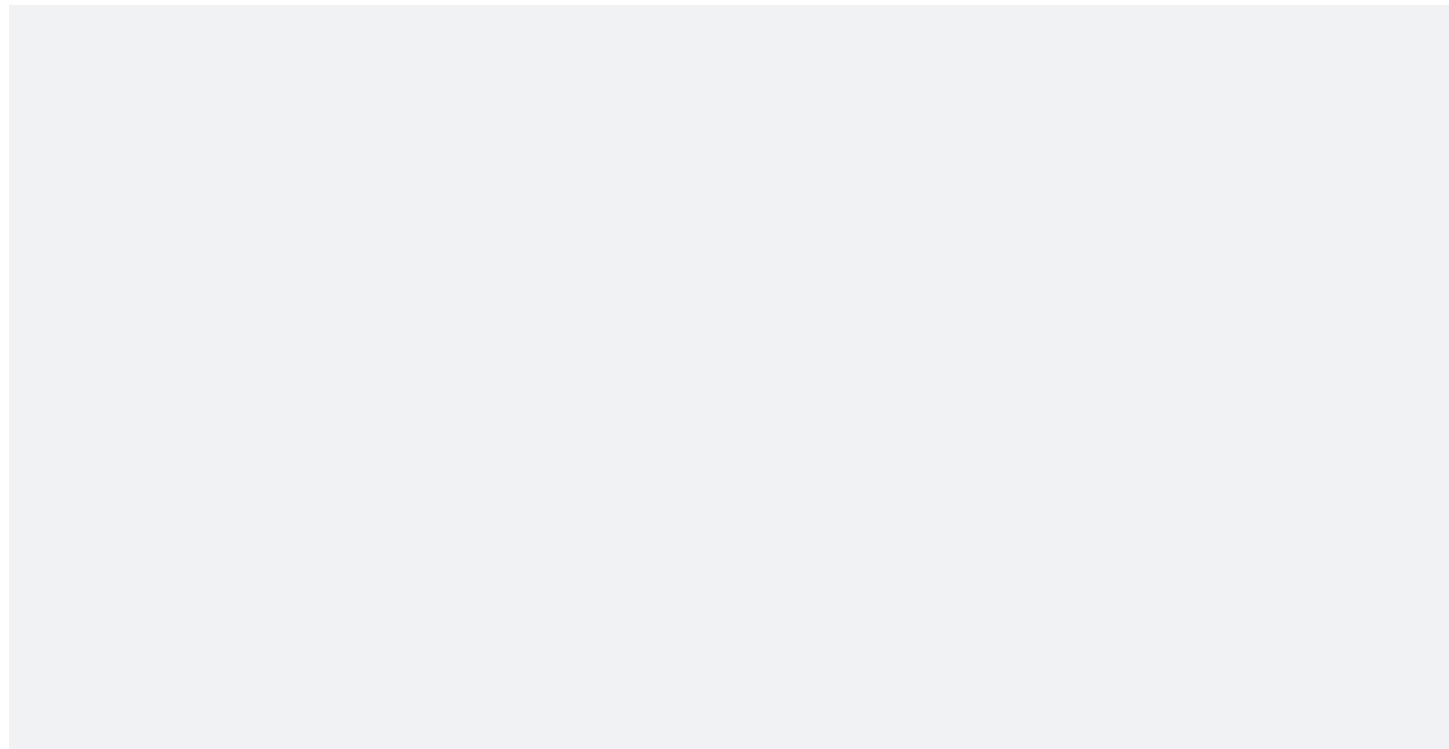Vertex AI 100   Gemini Era   Use cases   Next Steps

Target Audience:

Level: Beginner

The session is intended for all and beginners who are new to Vertex AI.

The Era of Gen AI
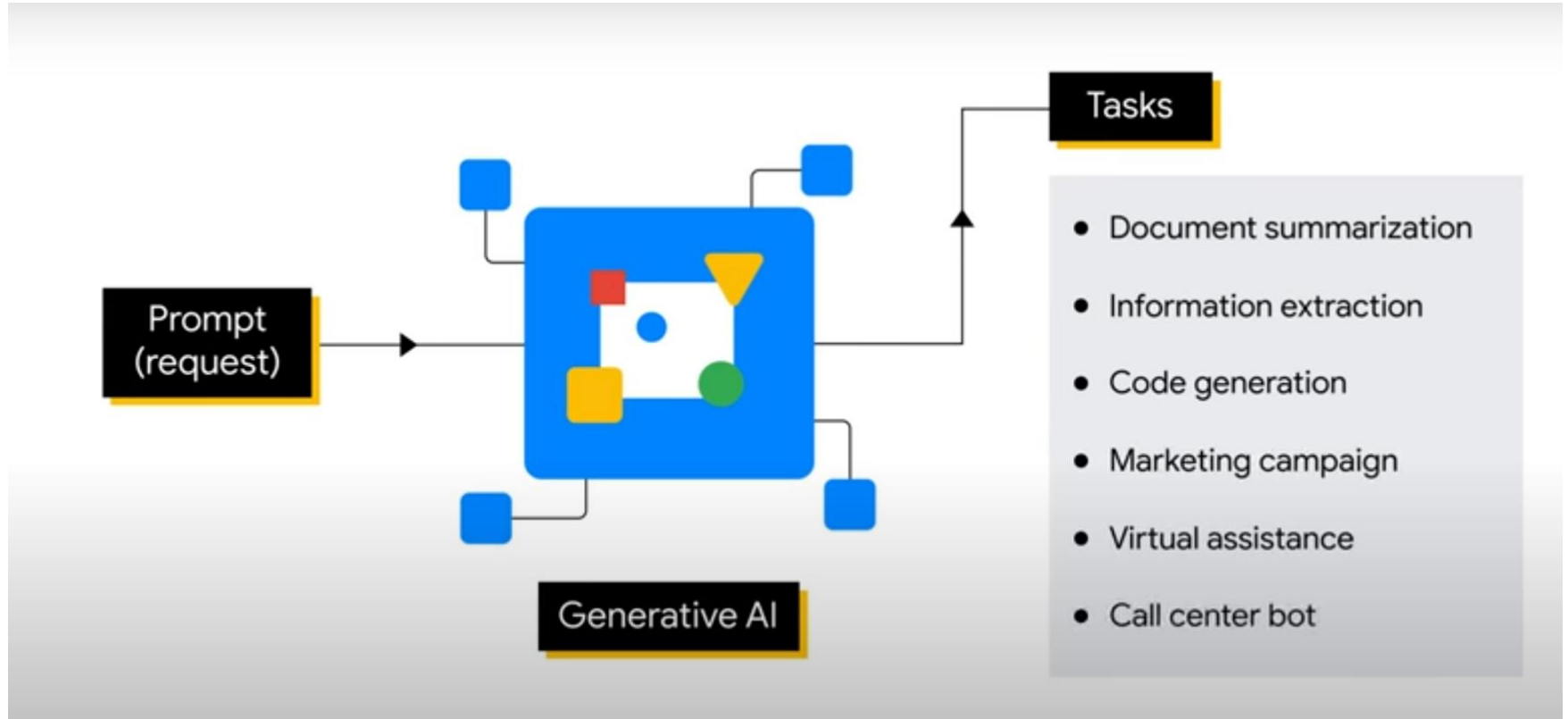
Generative AI
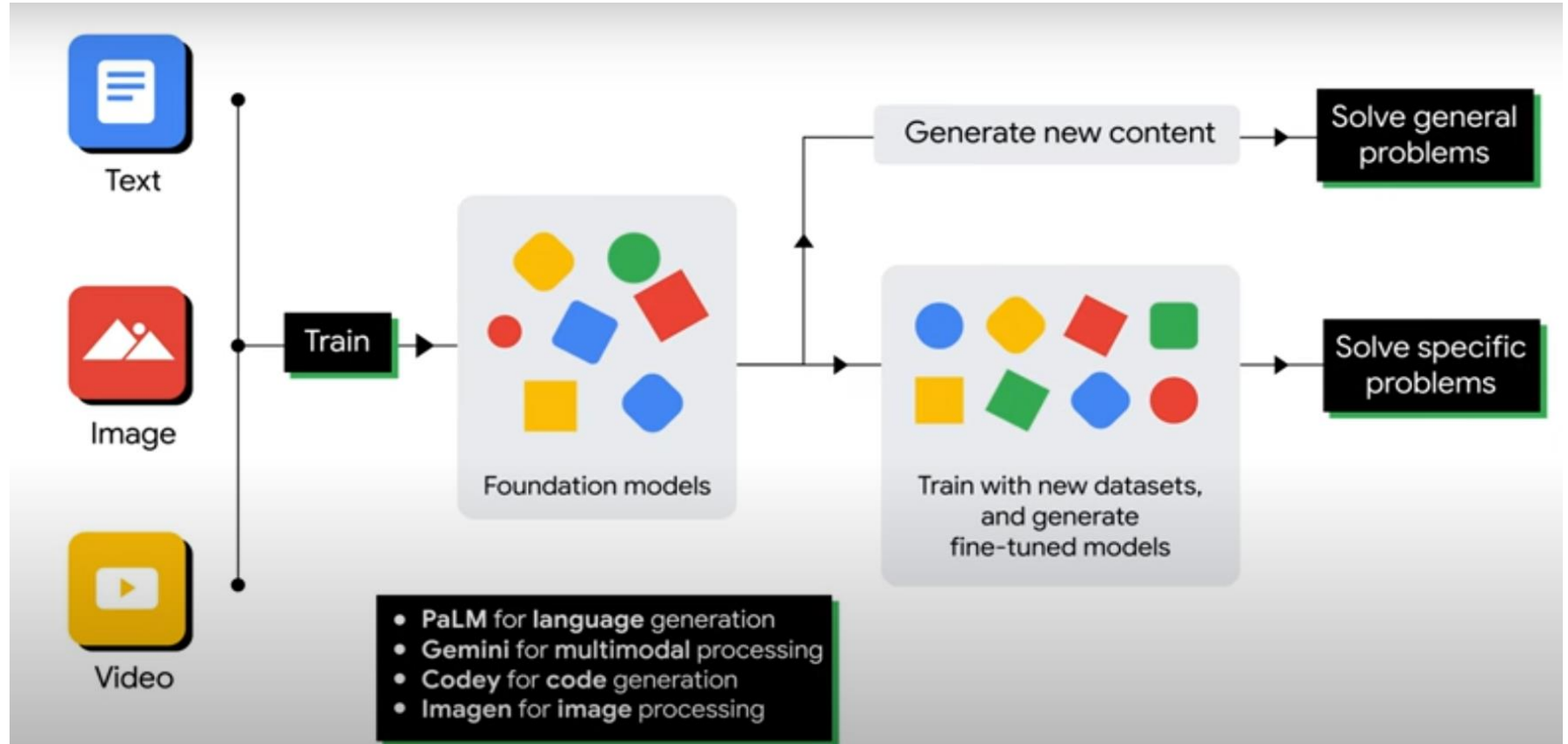
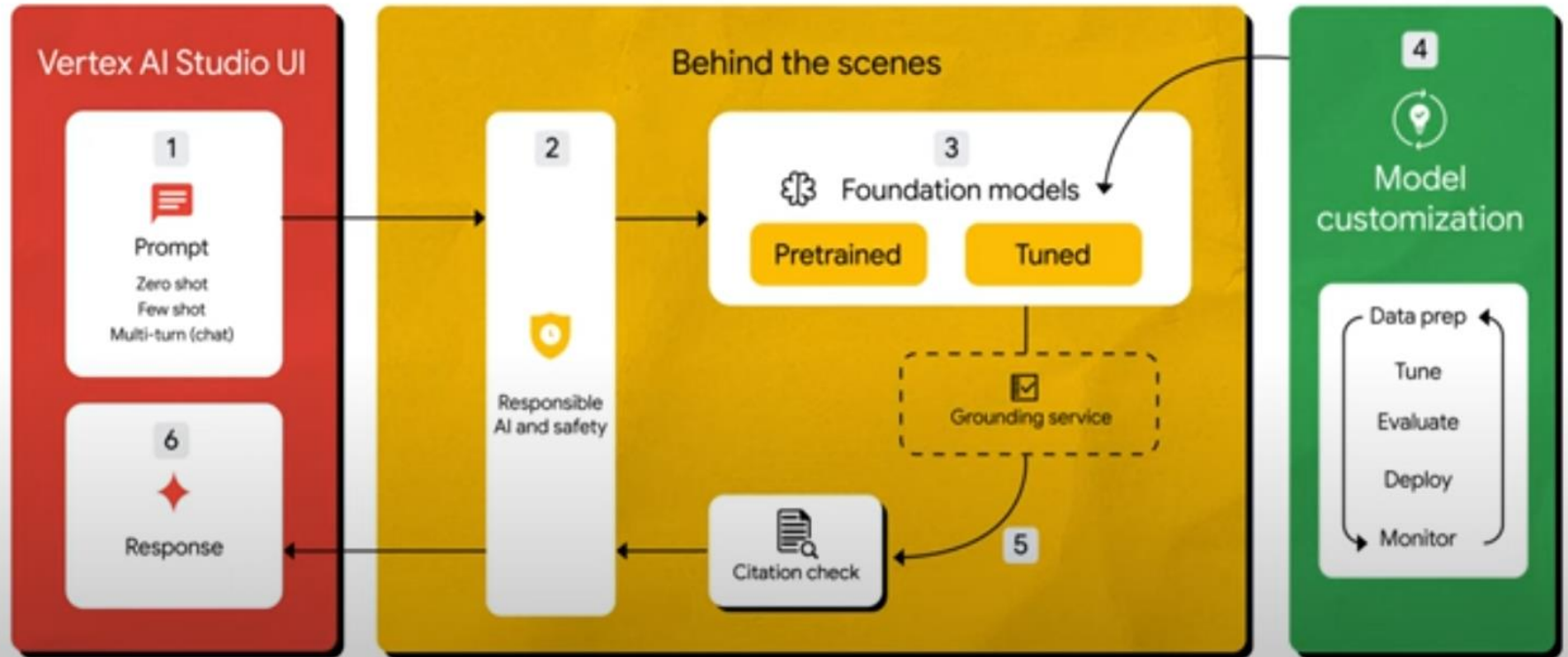AI that generates content for you

Text

Code

Image

Speech

Video

Overview

Architecture

Text

Image

Video

Train

Foundation models

Generate new content

Solve general problems

Train with new datasets, and generate fine-tuned models

Solve specific problems

- **PaLM** for **language** generation
- **Gemini** for **multimodal** processing
- **Codey** for **code** generation
- **Imagen** for **image** processing

Generative AI workflow

Vertex AI Platform

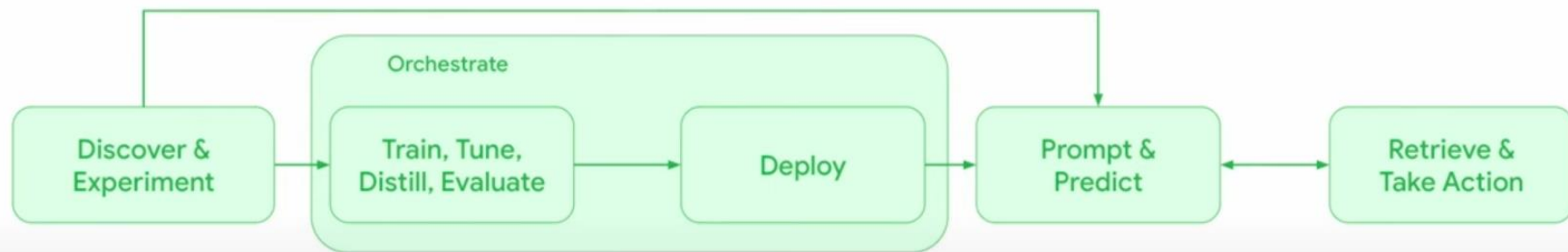Vertex AI: Discover, build, experiment, deploy, and manage ML models

GenAI

Model Garden	Generative AI Studio
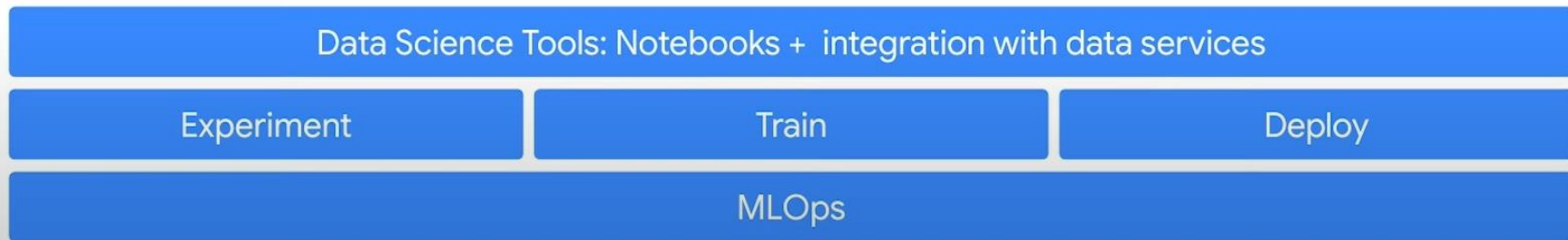
| Open Source models | Task-specific AutoML & APIs | Foundation models | Prompt design | Tuning adapter layers |

ML Platform

Data Science Tools: Notebooks + integration with data services

| Experiment | Train | Deploy |

MLOps

MLOps for Gen AI

Vertex AI Studio

Primary tool to access Google's Generative AI models

Test

Tune

Augment

Deploy

Vertex AI Studio

Multimodal

Process data in multiple modalities like image, video, and text

- Perform multimodal tasks

Language

Generate text and tune models

- Design prompts for different tasks
- Create conversations
- Tune and deploy a language model
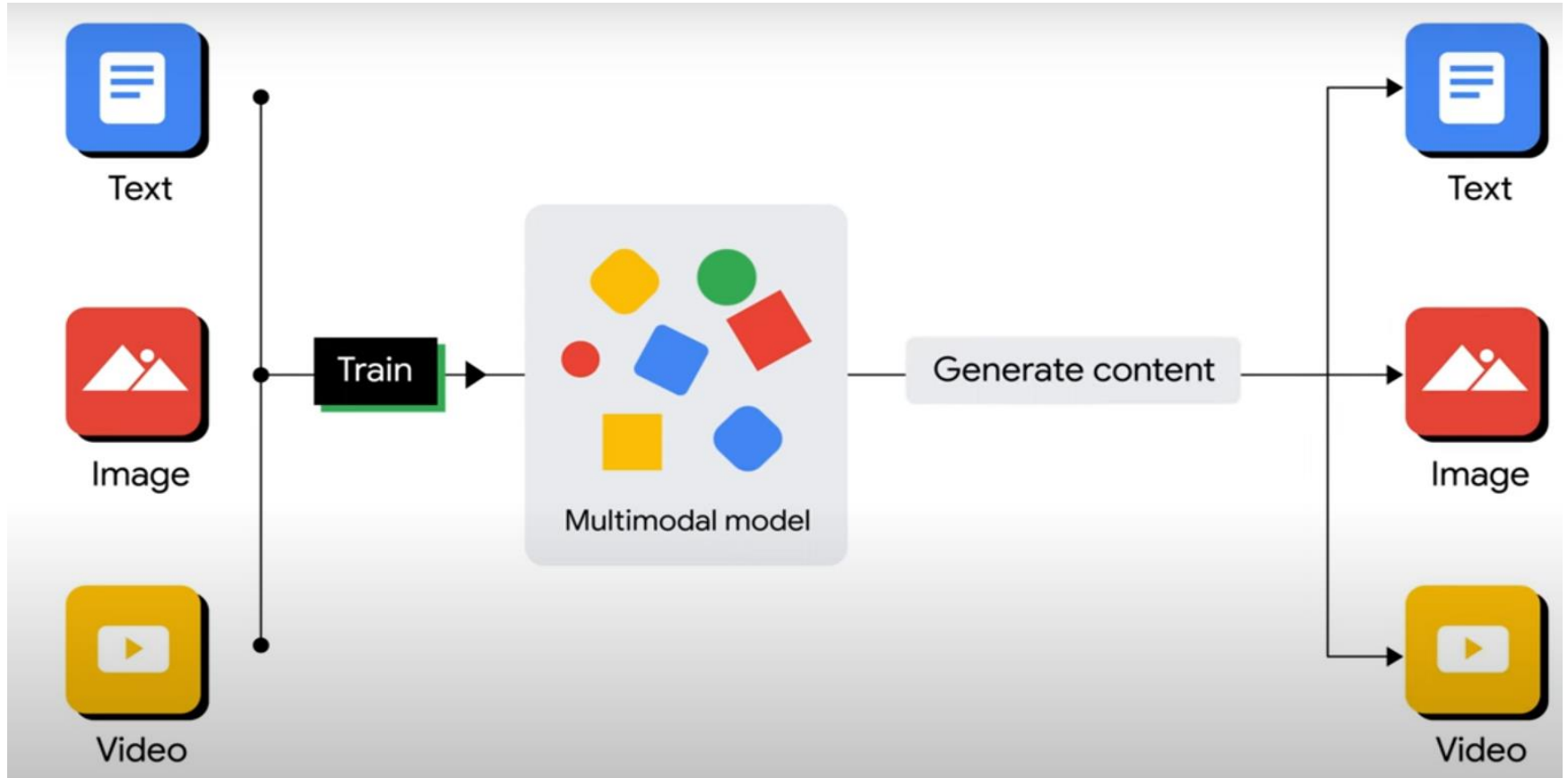
Vision

Generate images based on prompts

- Generate images
- Edit images

Speech

Generate text or speech

- Generate text from speech
- Generate speech from text

Multi Modal

Use cases

Description and captioning

Identifying the objects and describing them

Information extraction

Reading text from images and videos

Information analysis

Analyzing the images and videos based on prompts

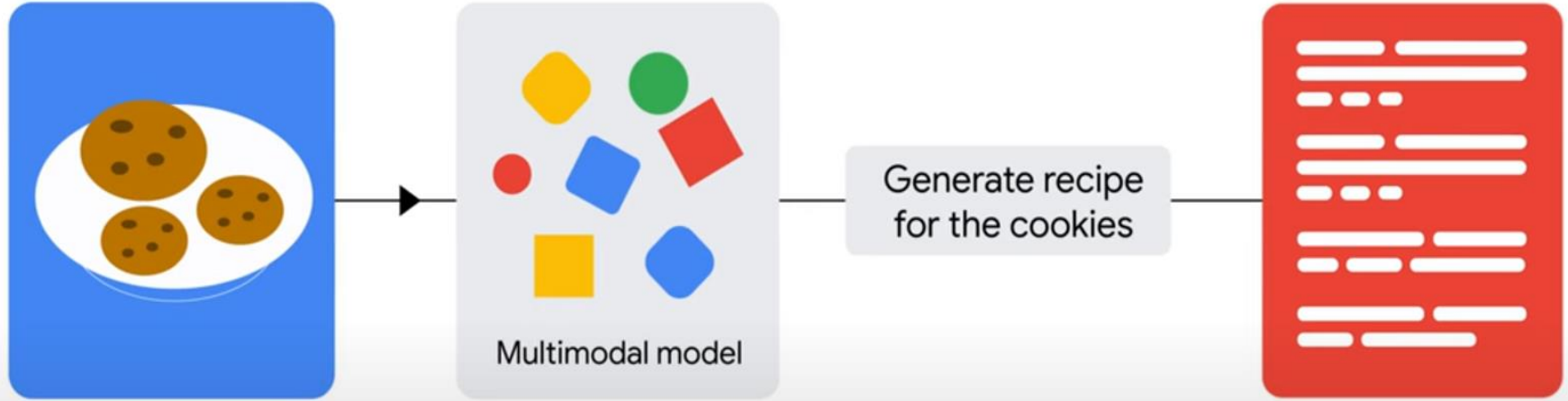**Information seeking**

Answering questions or generating Q&A

Content creation

Generating a story or advertisement

Data conversion

Converting text responses in formats like JSON

Multi Modal

Multimodal model

Generate recipe for the cookies

Gemini

Gemini – Ways to Interreact

User interface (UI)	SDKs	APIs
Using a UI with Google Cloud console (no-code)	Using SDKs with a notebook (in Python/Java)	Using APIs with command-line tools

Prompt Design

The response from the model

The answers you get depend on the questions you ask

The prompt you designed
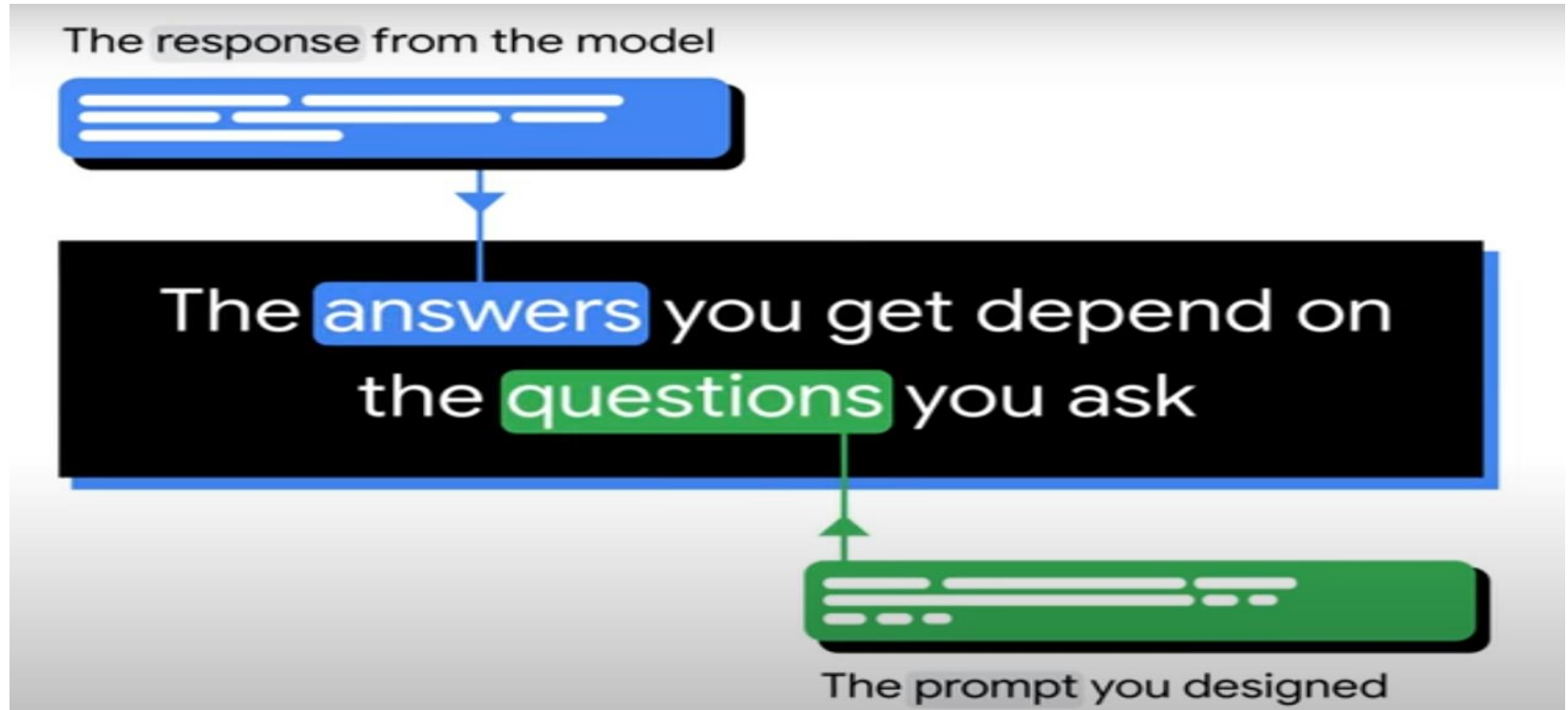
Customize & Tune a Gen AI model
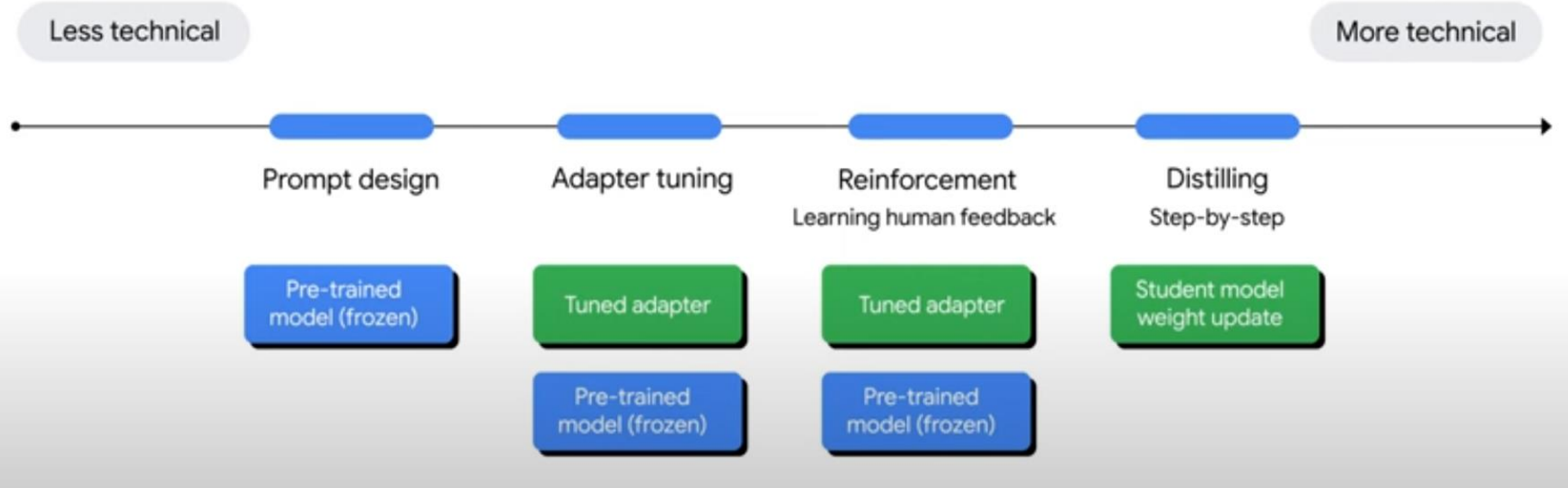
Enterprise Readiness

Consumer

Plan a 3-day trip to Patagonia

Create a valentine poem

Generate a picture of a panda playing yahtzee

Enterprise

How do we **control** our data?

How will we manage **costs?**

How to ensure data **accuracy** and **security**?

Data governance and privacy
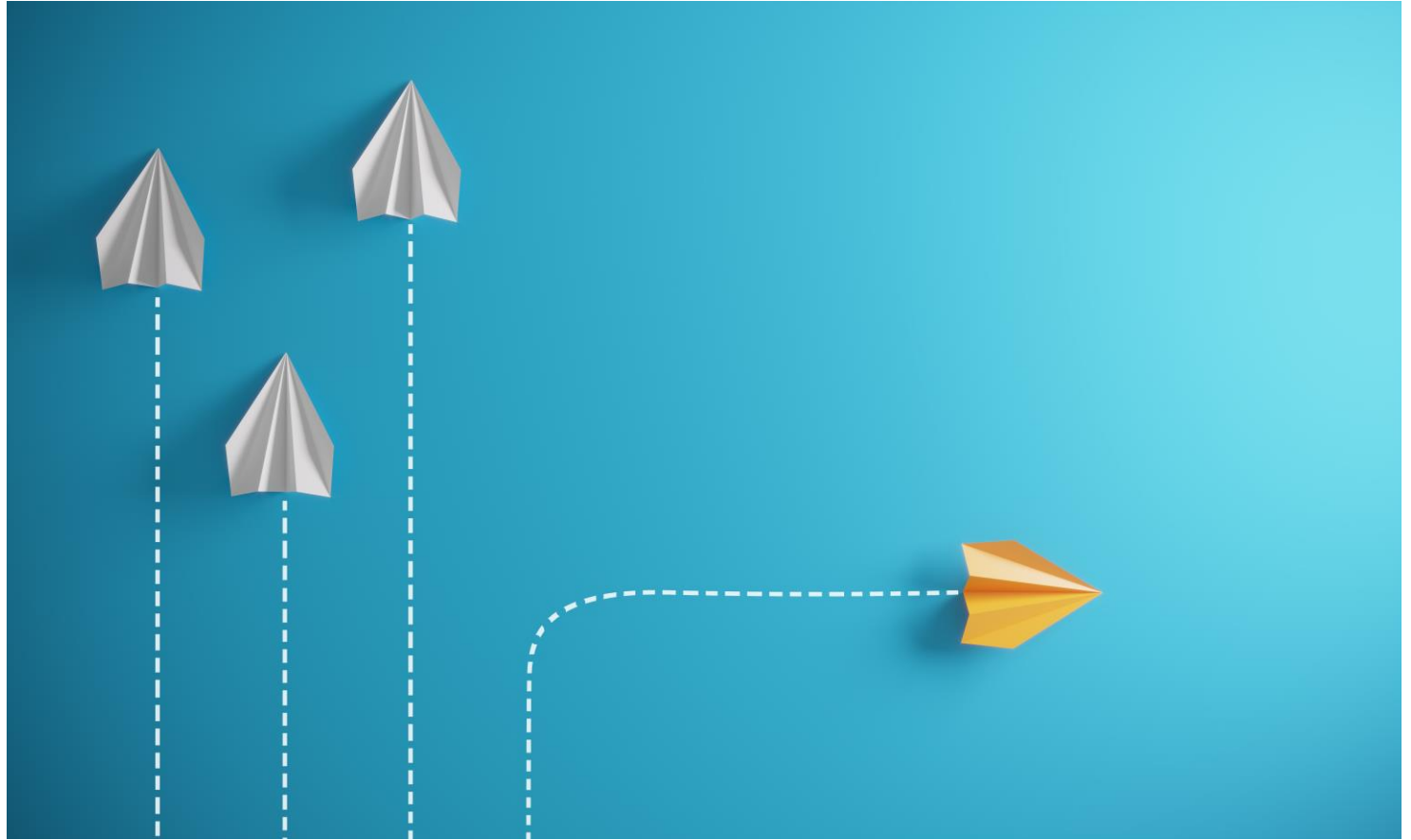
Security and compliance support

Reliability and sustainability

Responsible AI

You own and control your data, not Google
Security Controls with VPC-SC, CMEK, access transparency
Responsible AI tooling, enablement & support

Demo

Learning Resources

- https://deepmind.google/technologies/gemini/

- https://blog.google/technology/ai/google-gemini-next-generation-model-february-2024

- https://github.com/GoogleCloudPlatform/generative-ai/

Source:

Open for Q&A

INGRAM MICRO®

Thank you!!

Navaneethan Gopal