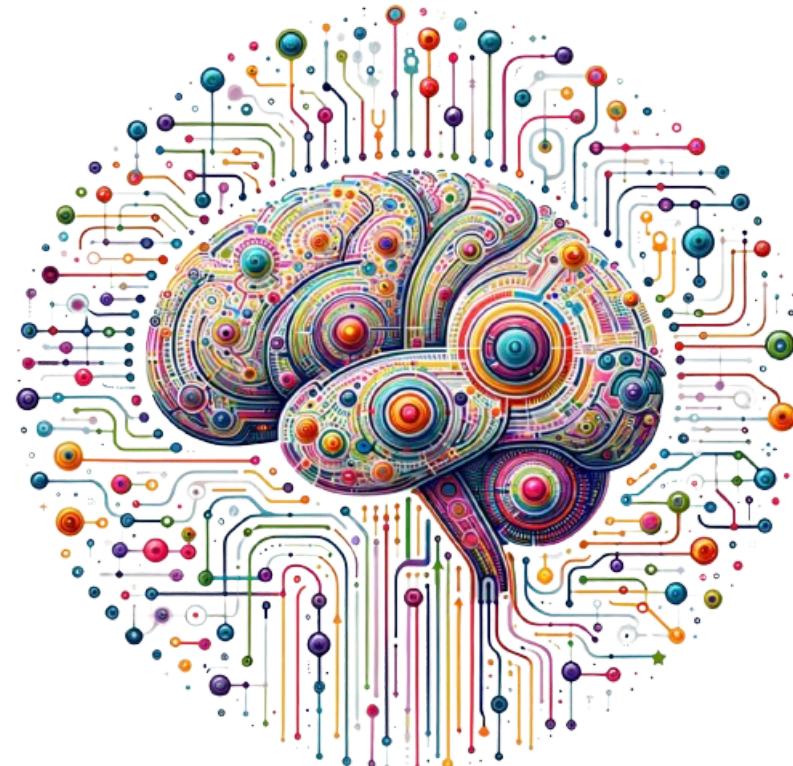


Artificial Intelligence

Artificial Intelligence - All around you

-  **Self-driving cars:** Navigate roads and make real-time decisions
-  **Spam Filters:** Keep unwanted emails out of your inbox
-  **Email Grouping:** Automatically sort emails into categories
-  **Fraud Detection:** Identify suspicious credit card transactions
-  **Recommendation Systems:** Movies, songs or product recommendations based on your preferences



What is AI? (Oxford Dictionary)

The theory and development of computer systems able to perform tasks normally requiring human intelligence, such as visual perception, speech recognition, decision-making, and translation between languages

What is AI? (in Simple Terms)

- Goal of AI: Create machines that can simulate human-like intelligence and behavior
 - 🕸️ Play Chess
 - 🛒 Make Purchase Decisions
 - 🚗 Drive a Car
 - 📝 Write an Essay!
 - 🎵 Compose Music
 - 🎨 Generate Art



Understanding Types of AI

- **Strong AI (General AI):** Machine intelligence = Human intelligence
 - A machine that can solve problems, learn, & plan for future
 - An expert at everything — even all sports and games!
 - Learns like a child, building on its own experiences
 - We are a little far away from this
 - Estimates: few decades to never
- **Narrow AI (Weak AI):** Designed for a specific task
 - Example: 🚗 Self-driving car
 - Example: ♟ Playing Chess
 - Example: 🏠 Predicting House Price

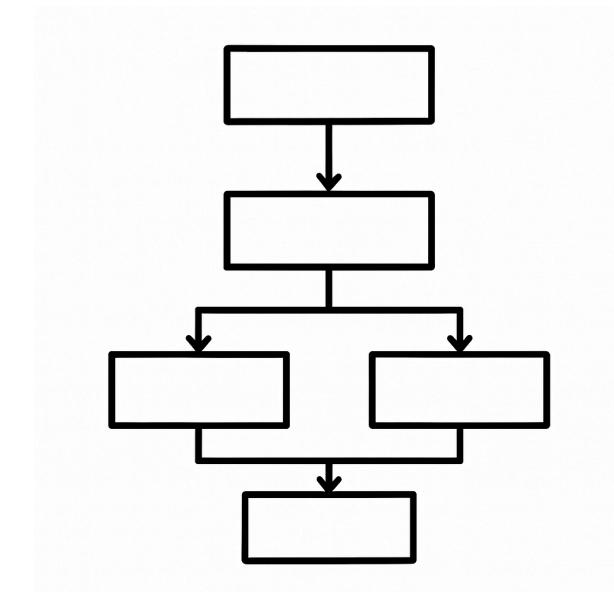


AI Fundamentals - Scenarios

Scenario	Solution
Categorize: Building a computer system as intelligent as a human. An expert at everything (all sports and games!)	Strong AI
Categorize: Building a computer system that focuses on specific task (Self-driving cars, virtual assistants, object detection from images)	Narrow AI (or weak AI)

Traditional Programming is based on Rules

- **Traditional Programming:** Based on Rules
 - IF this THEN do that
- **Example:** Predict price of a home
 - Design an algorithm with predefined rules considering:
 - Location
 - Home size
 - Age
 - Condition
 - Market Trends
 - Economic indicators etc..



Machine Learning Learns From Examples

- Machine Learning: Learns from examples (instead of rules)
 - 1: Provide millions of examples
 - 2: Train an algorithm to create a model
 - 3: Use the model to make predictions on new data
 - **Example - House Price Prediction**
 - Data: Location, Size, Bedrooms, Age, Condition and Price of the home
 - ML learns the relationship from past sales data
 - **Example - Handwriting Recognition**
 - Data: Images of handwritten digits (0-9) along with the correct digit
 - ML learns to identify digits based on pixel patterns from the image data

Home size (Square Yds)	Age	Condition (1-10)	Price \$\$\$
300	10	5	XYZ
200	15	9	ABC
250	1	10	DEF
150	2	34	GHI

A 4x4 grid of handwritten digits (0-9) used for training a handwriting recognition machine learning model. The digits are arranged in a 4x4 pattern. Some digits are bolded to represent the target output for each input image.

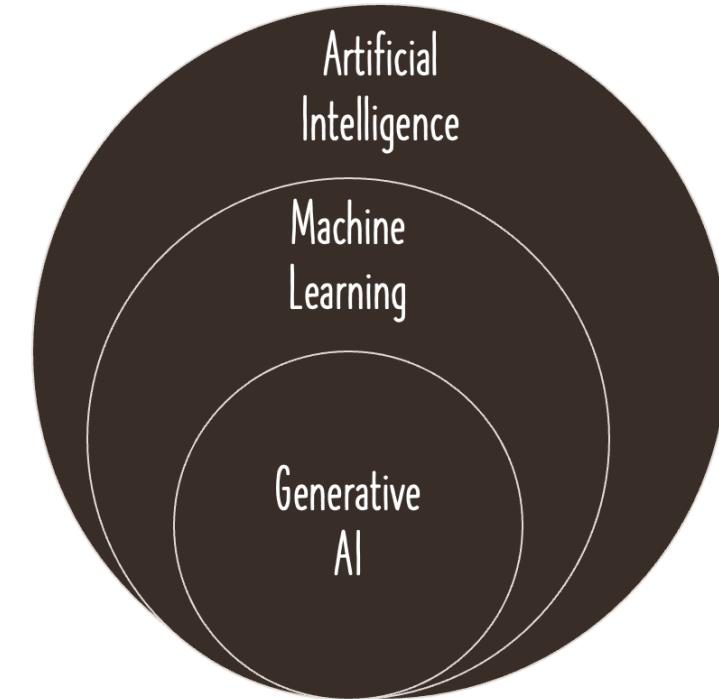
0	0	0	0
1	1	1	1
2	2	2	2
3	3	3	3
4	4	4	4
5	5	5	5
6	6	6	6
7	7	7	7
8	8	8	8
9	9	9	9

Machine Learning Fundamentals - Scenarios

Scenario	Solution
Category of AI that focuses on learning from data (examples)	Machine learning
How is ML different from traditional programming?	Traditional Programming: Rules. Machine Learning: Examples

Generative AI - How is it different?

- **Artificial intelligence (AI)**: Create machines that can simulate human-like intelligence and behavior
- **Machine learning (ML)**: A subset of AI where machines learn from data.
- **Generative AI**: An subset of ML that creates new content.



Generative AI - Generates New Content

- **Goal:** Generate New Content
 - Instead of making predictions, Generative AI focuses on creating new content
 - **Examples:**
 - **Text Generation:** Writing e-mails, essays & poems. Generating ideas. Generate responses for customer service interactions. Draft professional communication for clients.
 - **Code Generation:** Generate code snippets. Create unit tests.
 - **Media Creation:** Design images. Create instructional videos.
- How else is Generative AI different?
 - Let's find out!

≡ Gemini ▾



Hello, In28Minutes
How can I help you
today?

Give me ways to add
certain foods to my
diet

Generate unit tests for
the following C#
function



Enter a prompt here



Generative AI uses Foundation Models

- Traditional models are **narrow** — trained for a specific task
 -  Example: A model that detects spam emails can't generate a marketing campaign
- What if we had models trained on **large, diverse datasets** across domains?
 - Enter **Foundation Models**
 -  Pretrained on **massive datasets**
 - REMEMBER: Needs complex training with huge infrastructure!
 -  Adaptable to a wide variety of tasks
 -  Foundation for many Generative AI apps



Foundation Models Have Complex Training Processes

- **Complex Training Process:** Training is computationally intensive and complex
- **Huge Volumes of Data:** Requires learning from billions of text tokens or millions of labeled images
 - **Text Models:** Trained on Wikipedia, books, code, etc.
 - **Image Models:** Trained on datasets with image-caption pairs
- **Specialized Hardware:** Training needs huge volumes of specialized hardware - GPUs (Graphics Processing Units), TPUs (Tensor Processing Units) and Distributed Storage



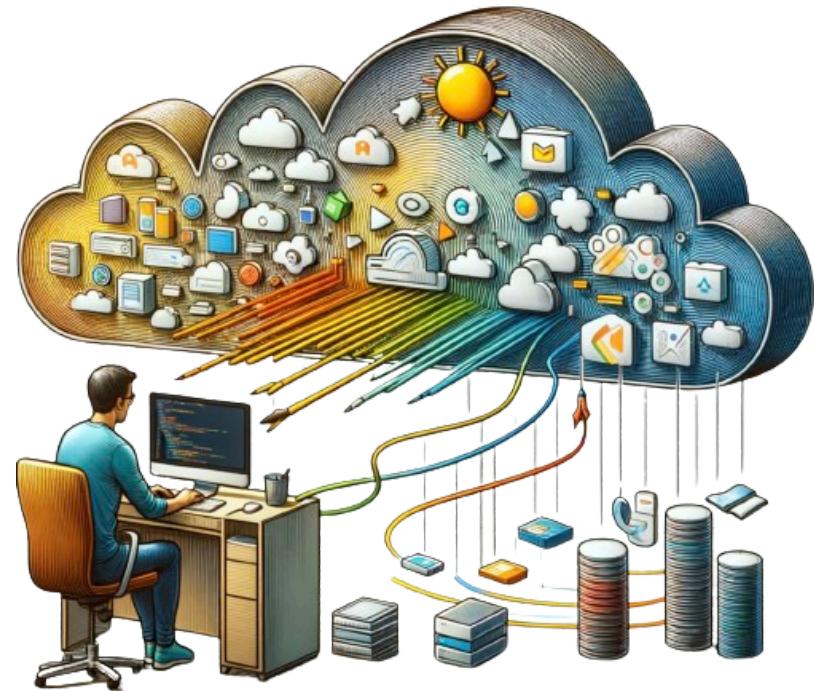
Large Language Models: Text Foundation Models

- Trained on **huge volumes** of diverse text data
 - Books, articles, websites, code, and more
- Learn to **predict the next token** in a sequence
 - Example: "The cat sat on the ..."
 - Model gives probabilities:
 - "mat" (40%), "table" (20%), "chair" (20%), "moon" (10%)
 - It usually picks the most likely token — but we can tune settings to make it more creative
 - e.g., using the temperature parameter
- **Usecases:** Chat, summarization, translation, Q&A, ..



Play with Foundation Models - Google AI Studio

- **Easy experimentation:** Ideal for learners and experimentation on a small scale
- **Needs Google Account:** Requires login with your Google account
- **Quick Prototyping:** Useful for prototyping ideas and designing effective prompts
- **Key Features:**
 - Easy-to-use, code-free interface
 - Built-in model playground with foundation models
 - Generate code to execute Gemini APIs
 - Build apps with Gemini



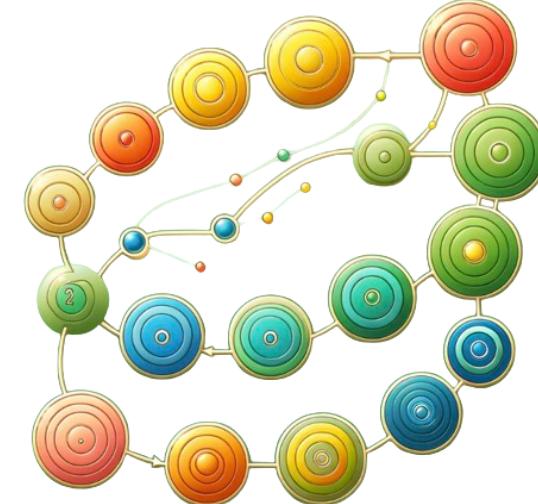
What is Grounding?

- **Grounding** = Connecting an AI's output to **verifiable sources**
- Enhances **trust, accuracy, and relevance**
- In Gemini:
 - Add links, documents, or guidelines in your prompt
 - Gemini uses them to generate informed and tailored results
- *Example:* Using your company training documentation to help answer trainee questions



RAG: Expanding Model Capabilities

- Before RAG: LLMs could only use static training data
- RAG = Retrieval + Augmentation + Generation
- Empowers LLMs to access external knowledge in real-time
- Improves accuracy, relevance, and transparency



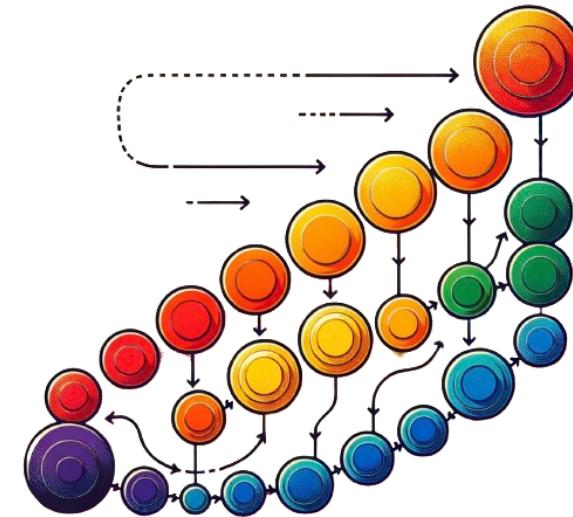
How RAG Works?

- **Retrieval:**
 - Search tools fetch context from data stores, APIs, or vector DBs
- **Augmentation:**
 - Retrieved info is added to the prompt
- **Generation:**
 - LLM responds using both query + retrieved context
- **Iteration (Optional):**
 - Model refines its search if context isn't useful



Why Use RAG?

- **Improves Accuracy:** Grounds answers in real, verifiable documents
- **Reduces Hallucinations:** Constrains responses to authorized knowledge only
- **Increases Transparency:** Supports source citations
- **Constrain LLMs:** Constrain model to only respond with *your* knowledge base
 - Example: Only answer using internal policy documents

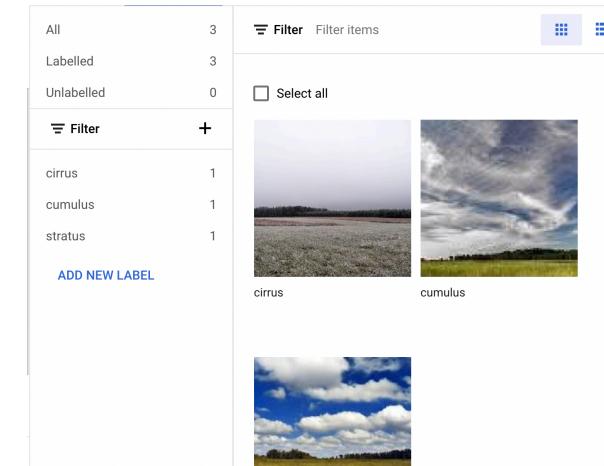


ML in Google Cloud

Before Generative AI

ML in Google Cloud - Traditional Landscape

- This is the ML landscape before emergence of Generative AI
- **Machine Learning based API**
 - Natural Language, Vision, Speech etc
- **Custom Models** without needing ML expertise
 - Vertex AI > Auto ML
- **Build Complex Custom Models**
 - Vertex AI > Custom Training



Google Cloud - 1 - Using Machine Learning API

- **Usecase:** Derive insights from unstructured text
 - Natural Language API - <https://cloud.google.com/natural-language>
- **Usecase:** Speech to Text
 - Speech to Text API - <https://cloud.google.com/speech-to-text>
- **Usecase:** Convert text into speech
 - Text to Speech API - <https://cloud.google.com/text-to-speech>
- **Usecase:** Extract insights from images, documents, and videos
 - Vision API - <https://cloud.google.com/vision>
- **Usecase:** Content moderation for Video, Object detection and tracking
 - <https://cloud.google.com/video-intelligence>



Google Cloud - 2 - Custom Models without ML expertise

- What if your team **DOES NOT** have ML expertise but you want to build custom machine learning models?
- **Solution:** Vertex AI AutoML
- You want to build a custom image classification solution without ML expertise!
- **Example:** Identify the specific type of cloud
- **1:** Provide examples - Example images and categorization
- **2:** AutoML creates the model for you!



AI in Google Cloud - 3 - Build Complex Custom Models

In 28
Minutes

- You have a complex ML problem to solve
- You have a team with the skills needed
 - Data Scientists, ..
- You want to make use of ML Frameworks
 - TensorFlow
 - PyTorch
 - scikit-learn
 - ...
- Solution: Vertex AI Custom Training

Model training method

AutoML

Train high quality models with minimal effort and machine learning expertise. Just specify how long you want to train. [Learn more](#)

Custom training (advanced)

Run your TensorFlow, scikit-learn and XGBoost training applications in the cloud. Train with one of Google Cloud's pre-built containers or use your own. [Learn more](#)

[CONTINUE](#)

ML in Google Cloud

After Generative AI

ML in Google Cloud – After Generative AI

- Generative AI brought **new capabilities**:
 - 🧠 Foundation models trained on vast, diverse datasets
 - 💡 Generate text, images, code, and audio
- Google Cloud has adapted by:
 - **Upgraded Machine Learning API** with generative capabilities
 - e.g., Natural Language API now includes summarization and content generation
 - **Vertex AI Studio**: Test, tune and deploy enterprise-ready generative AI
 - **Vertex AI Model Garden**: Access and finetune Foundation Models: Gemini, Imagen, and other models
 - **Vertex AI Agent Builder**: Build multi-step AI workflows using Agents



Flexible ML Pathways in Google Cloud - Basic

- 1. Use Pretrained APIs
 - Natural Language API, Vision API, Speech-to-Text, etc.
 - Fastest way to integrate AI into applications — just call the API
- 2. Use Foundation Models (as-is via API)
 - Vertex AI Model Garden: Experiment with first-party models like Gemini, or Partner models like Llama
- 3. Customize Foundation Models
 - Vertex AI Model Garden + Fine Tuning: Fine-tune base models using your domain-specific data
 - Adjust tone, behavior, or style (e.g., legal, support)



Flexible ML Pathways in Google Cloud - Advanced

In 28
Minutes

- **4. Train Your Own Models with AutoML**

- High-quality models without extensive ML expertise
- For teams without deep ML expertise
 - BUT you want to build your own custom model from scratch!
- Automatically handles all machine learning complexity

- **5. Build Complex Custom Models**

- You want full control over model creation
- For teams with deep ML expertise
- Use frameworks like TensorFlow, PyTorch,..
- Solution: Vertex AI Custom Training



Flexible ML Pathways in Google Cloud – Scenarios 1

In 28
Minutes

Scenario	Solution
A company wants to build a chatbot using a foundation model without any fine-tuning	Use Foundation Models from Model Garden
A healthcare company wants their AI assistant to respond in a medical tone with accurate terminology	Fine Tune Foundation Model
An e-commerce platform wants an image generator that matches their brand aesthetic	Fine Tune a Foundation Model (with brand-specific data)
A support team wants to use a model that knows the company's internal tone and product line	Fine Tune Foundation Model (using support transcripts)

Flexible ML Pathways in Google Cloud – Scenarios 2

In 28
Minutes

Scenario	Solution
A retail team wants to build a model to categorize images but doesn't have ML engineers	Vertex AI AutoML
A telecom firm has a team of data scientists building a complex churn prediction model using TensorFlow	Build Complex Custom Models (Vertex AI Custom Training)
A logistics company wants to train a route optimization model using PyTorch	Build Complex Custom Models (Vertex AI Custom Training)
A bank wants to build and train a fraud detection model on GPU-based infrastructure with full control over training logic	Build Complex Custom Models (Vertex AI Custom Training)

Vertex AI

Vertex AI: Google Cloud's AI Platform

- Build, deploy, and scale AI – all in one place
- Combines infrastructure, tools, and API for:
 - Traditional ML Lifecycle AND
 - Generative AI usage and tuning



Vertex AI – Build Your Models From Zero

- **Two Approaches To Build Models From Zero**
 - **AutoML:** Build high-quality models without deep ML expertise (Just provide the dataset)
 - **Custom Training:** Create and train models at scale using any ML framework - TensorFlow, PyTorch,.. (Take complete control)
- **End-to-end MLOps:** From data to predictions
 - **Fully managed infrastructure:** Scale infrastructure on demand for training and deployment
 - **Datasets:** Manage your training data
 - **Experiments:** Track and compare your ML experiments
 - **Model Registry:** Maintain model versions with complete tracking
 - **Model Monitoring:** Monitor your model; auto-trigger retraining if needed



Vertex AI Studio - Gen AI Made Easy

- **Vertex AI Studio:** Rapid prototyping and testing of generative AI models
 - **Model Garden:** Hundreds of models from Google, partners, and open-source
 - *First Party Google Models:* Gemini, Imagen, Veo, ..
 - *Open Models:* Gemma (Lightweight, state-of-the-art open models from Google), CodeGemma, PaliGemma, Llama, Mistral
 - *Partner Models:* Claude (Anthropic), and more
 - **Prompt Gallery:** Explore ready-to-use prompts for common use cases
 - **Tuning:** Adapt foundation models to your domain with custom data



Google AI Studio vs Vertex AI Studio

Feature	Google AI Studio	Vertex AI Studio
Audience	Beginners and Experimenters	Enterprise teams
Access	Google account	Google Cloud
Use Case	Try prompts and models quickly	Build, tune, and deploy production-grade models
Governance & Security	Basic	Enterprise-grade access control and compliance