

# INTRODUCTION TO GENERATIVE AI

Yogesh Haribhau Kulkarni



# Outline

## ① OVERVIEW

## ② CONCLUSIONS

## ③ REFERENCES

## About Me

# Yogesh Haribhau Kulkarni

## Bio:

- ▶ 20+ years in CAD/Engineering software development
- ▶ Got Bachelors, Masters and Doctoral degrees in Mechanical Engineering (specialization: Geometric Modeling Algorithms).
- ▶ Currently doing Coaching in fields such as Data Science, Artificial Intelligence Machine-Deep Learning (ML/DL) and Natural Language Processing (NLP).
- ▶ Feel free to follow me at:
  - ▶ Github ([github.com/yogeshhk](https://github.com/yogeshhk))
  - ▶ LinkedIn ([www.linkedin.com/in/yogeshkulkarni/](https://www.linkedin.com/in/yogeshkulkarni/))
  - ▶ Medium ([yogeshharibhaukulkarni.medium.com](https://yogeshharibhaukulkarni.medium.com))
  - ▶ Send email to [yogeshkulkarni at yahoo dot com](mailto:yogeshkulkarni@yahoo.com)



Office Hours:  
Saturdays, 2 to 5pm  
(IST); Free-Open to all;  
email for appointment.

# Introduction to Generative AI

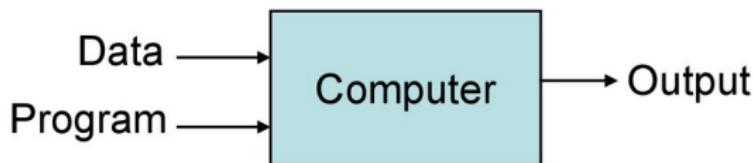
YHK

# Introduction

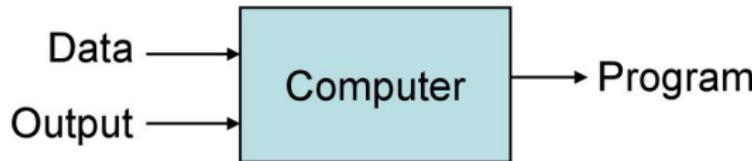
- ▶ Generative AI rapidly generates new data using machine learning algorithms.
- ▶ But What is Generative AI?
- ▶ How is it related to AI-ML-DL?

# Traditional vs. Machine Learning?

## Traditional Programming



## Machine Learning



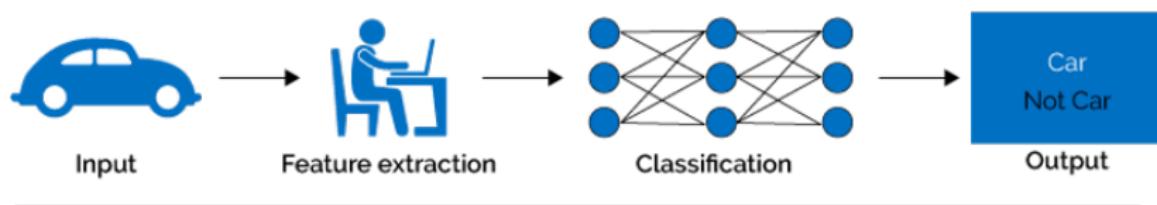
# Why Machine Learning?

- ▶ Problems with High Dimensionality
- ▶ Hard/Expensive to program manually
- ▶ Job \$\$\$

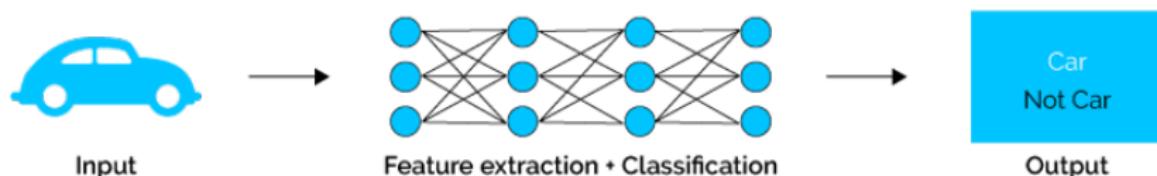
## ML vs DL: What's the difference?

Deep learning algorithms attempt to learn (multiple levels of) representation by using a hierarchy of multiple layers

### Machine Learning



### Deep Learning



(Reference: <https://www.xenonstack.com/blog/static/public/uploads/media/machine-learning-vs-deep-learning.png>)

## Use Deep Learning When ...

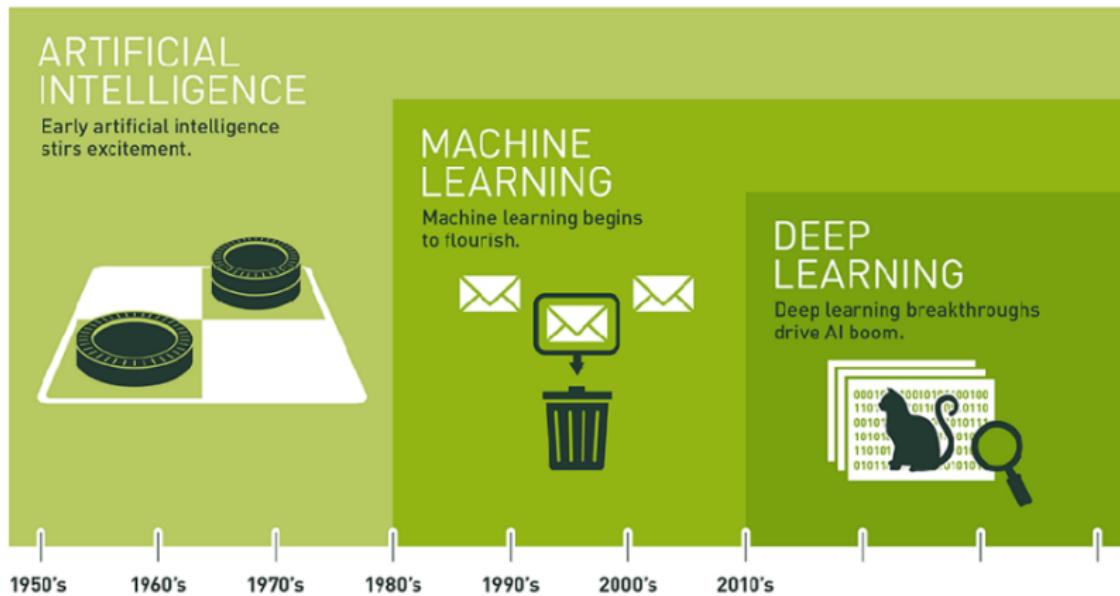
- ▶ You have lots of data (about 10k+ examples)
- ▶ The problem is “complex” - speech, vision, natural language
- ▶ The data is unstructured
- ▶ Techniques to model ‘ANY’ function given ‘ENOUGH’ data.

(Ref: Introduction to TensorFlow 2.0 - Brad Miro)



# Relationship between AI, ML, DL

First, let's see what's AI-ML-DL and relationship among them.



(Ref: <https://blogs.nvidia.com/blog/2016/07/29/whats-difference-artificial-intelligence-machine-learning-deep-learning-ai/>)

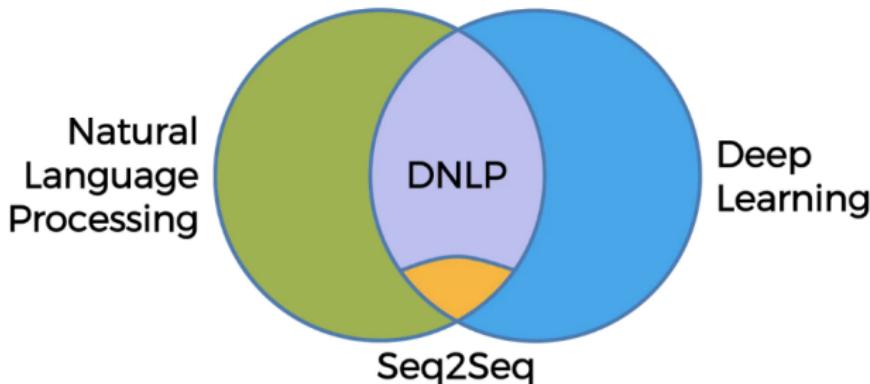
# What is Gen AI wrt AI, ML, DL

**Generative AI**  
is a **subset** of  
**Deep Learning**



(Ref: Introduction to Generative AI - Google Cloud Tech)

## What is Deep NLP



(Ref: Deep Learning and NLP A-Z - Kirill Eremenko)  
(Note: Size is not indicative of importance)

Seq2Seq is heavily used technique of DNLP for sequence to sequence modeling, eg Translation, Q & A, etc. Thats the basis of Large Language Models (LLMs)

# Types of Approaches

## Deep Learning Model Types



### Discriminative

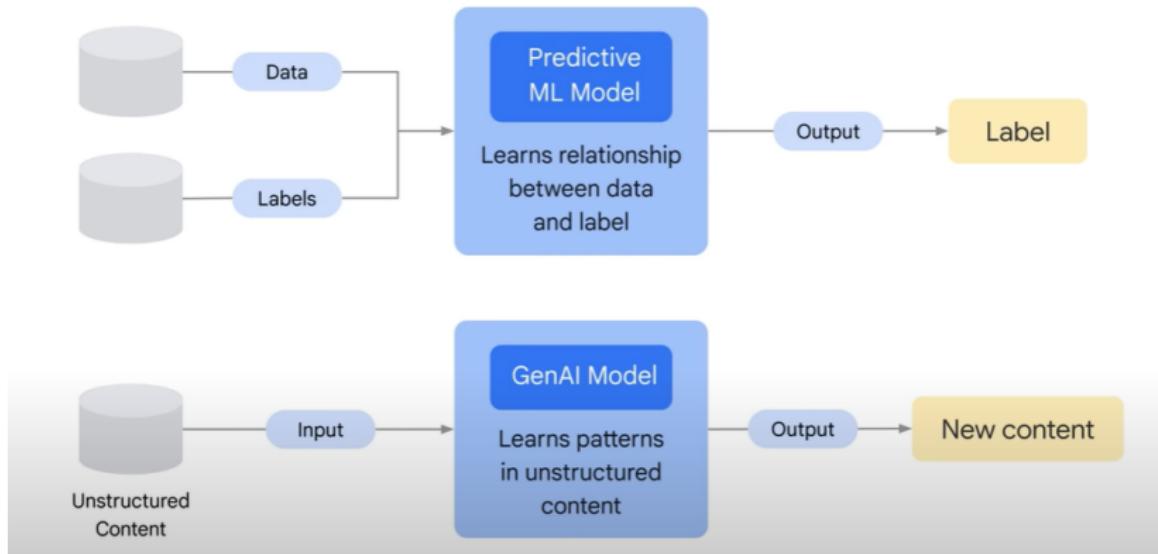
- Used to classify or predict
- Typically trained on a dataset of labeled data
- Learns the relationship between the features of the data points and the labels

### Generative

- Generates new data that is similar to data it was trained on
- Understands distribution of data and how likely a given example is
- Predict next word in a sequence

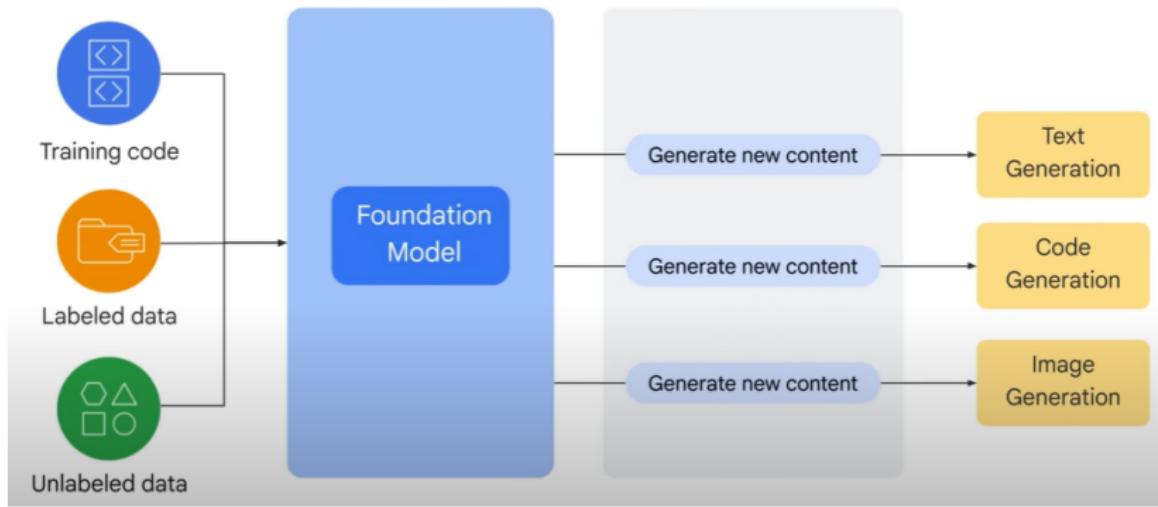
(Ref: Introduction to Generative AI - Google Cloud Tech)

## Types of Approaches



(Ref: Introduction to Generative AI - Google Cloud Tech)

# What is Foundation Model?



(Ref: Introduction to Generative AI - Google Cloud Tech)

## Same Problem, using different Technologies

## Difference across technologies, old to new

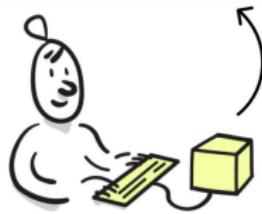
Lets see how the solutions to the problem of detecting a cat from images using traditional programming, deep learning, and generative AI, respectively.



## Traditional Programming

- ▶ Traditional programming involves writing explicit rules to detect a cat in images.
- ▶ Features like color, texture, and shape can be used to define these rules.
- ▶ However, designing accurate rules for complex patterns like cat detection can be challenging.
- ▶ It requires extensive domain knowledge and might not generalize well to different images.

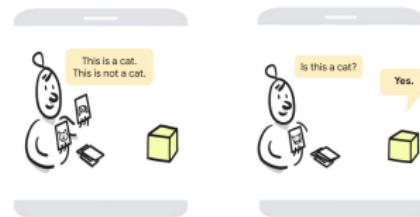
```
cat:  
  type: animal  
  legs: 4  
  ears: 2  
  fur: yes  
  likes: yarn, catnip
```



(Ref: Primer on LLM and Gen AI - Google Cloud)

# Deep Learning

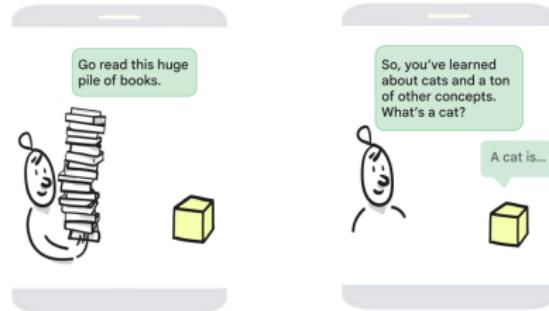
- ▶ Deep learning utilizes neural networks to automatically learn features for cat detection.
- ▶ Convolutional Neural Networks (CNNs) are particularly effective for image classification tasks.
- ▶ Large labeled datasets of cat images are used to train the network.
- ▶ The network learns to identify unique cat features and generalize them to detect cats in new images.
- ▶ Deep learning offers better accuracy and can handle complex patterns without explicit rule definition.



(Ref: Primer on LLM and Gen AI - Google Cloud)

# Generative AI

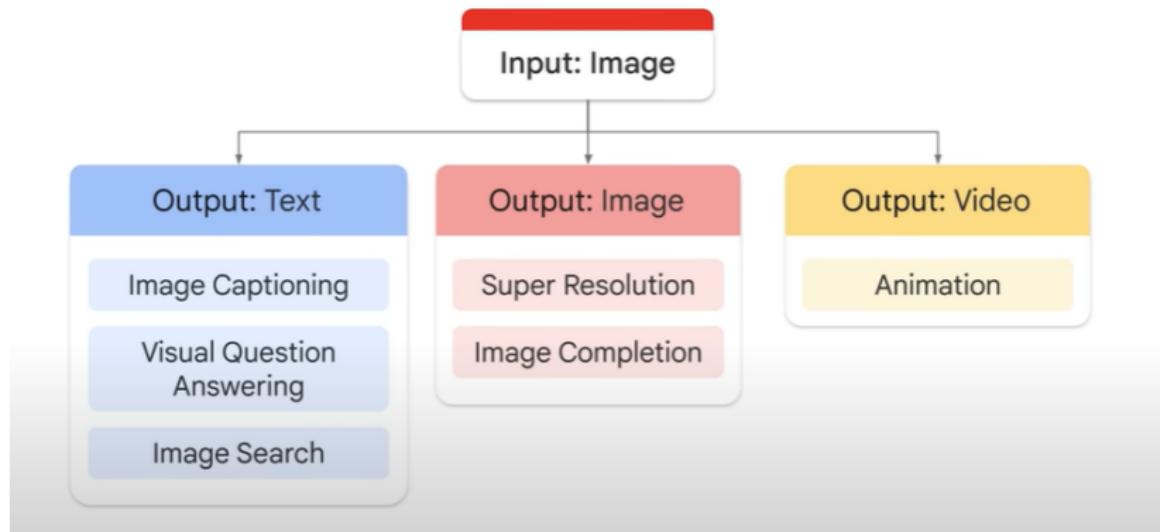
- ▶ Generative AI focuses on generating new data, including images of cats.
- ▶ Generative Adversarial Networks (GANs) are used to generate realistic cat images.
- ▶ The GAN consists of a generator and a discriminator that compete against each other.
- ▶ The generator learns to generate increasingly realistic cat images, while the discriminator learns to distinguish real from generated images.
- ▶ The generated cat images can be used to augment datasets for cat detection models.



(Ref: Primer on LLM and Gen AI - Google Cloud)

YHK

# Modalities in Generative AI

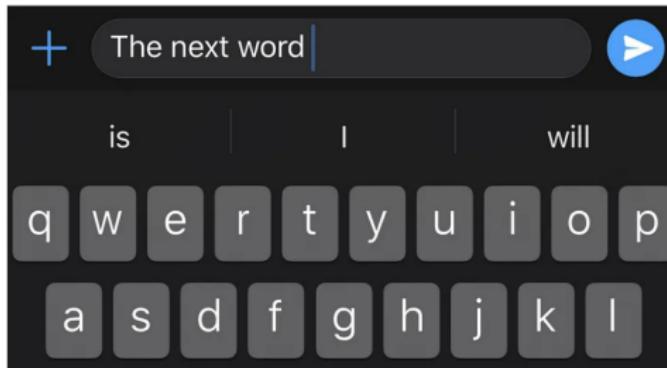


(Ref: Introduction to Generative AI - Google Cloud Tech)

Let's focus on the most popular modality ...

## What is a Language Models?

- ▶ While typing SMS, have you seen it suggests next word?
- ▶ While typing email, have you seen next few words are suggested?
- ▶ How does it suggest? (suggestions are not random, right?)
- ▶ In the past, for "Lets go for a ... ", if you have typed 'coffee' 15 times, 'movie' say 4 times, then it learns that. Machine/Statistical Learning.
- ▶ Next time, when you type "Lets go for a ", what will be suggested? why?
- ▶ This is called Language Model. Predicting the next word. When done continuously, one after other, it spits sentence, called Generative Model.



Next word prediction using language modeling in keyboards(Mandar Deshpande)

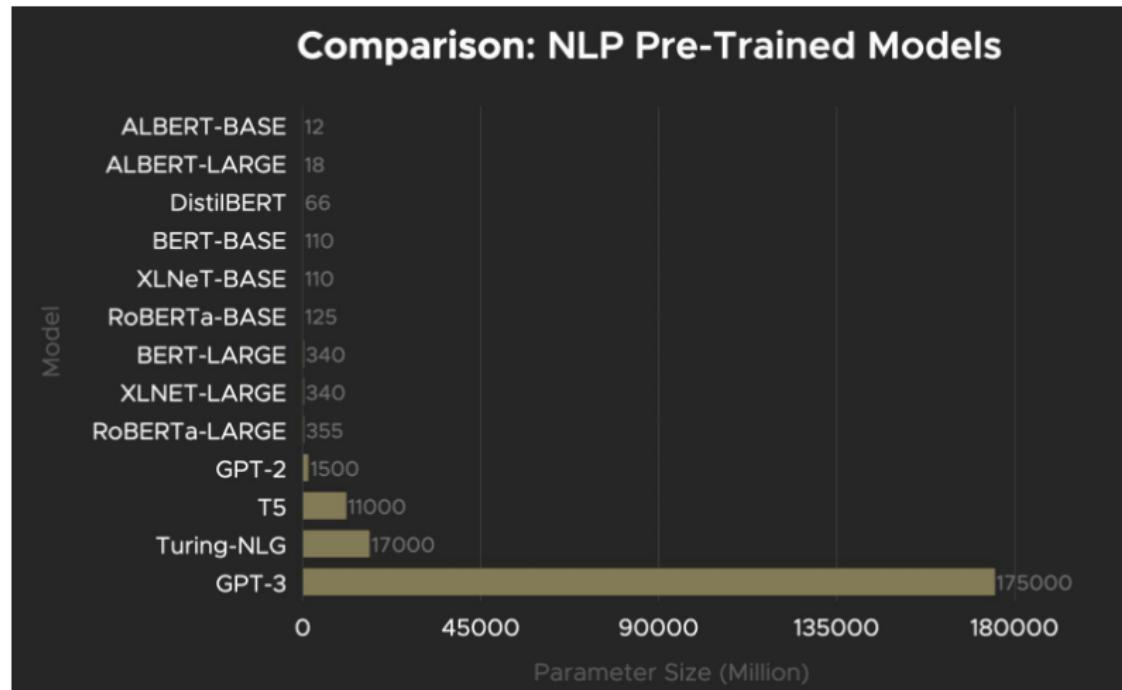
YHK

## Why they are called Large? Corpus

### GPT: Generative Pre-trained Transformers

- ▶ GPT-1 is pre-trained on the BooksCorpus dataset, containing 7000 books amounting to 5GB of data
- ▶ GPT-2 is pre-trained using the WebText dataset which is a more diverse set of internet data containing 8M documents for about 40 GB of data
- ▶ GPT-3 uses an expanded version of the WebText dataset, two internet-based books corpora that are not disclosed and the English-language Wikipedia which constituted 600 GB of data

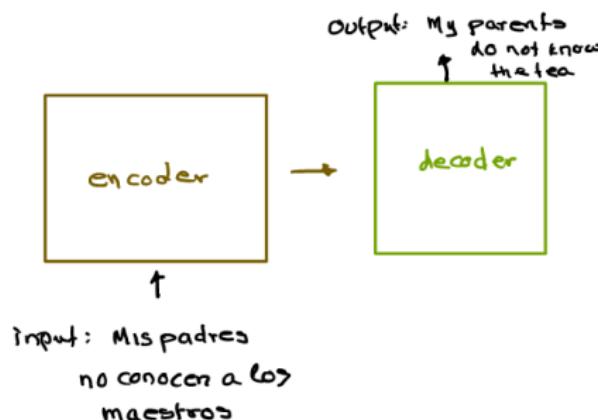
## Why they are called Large? Parameters



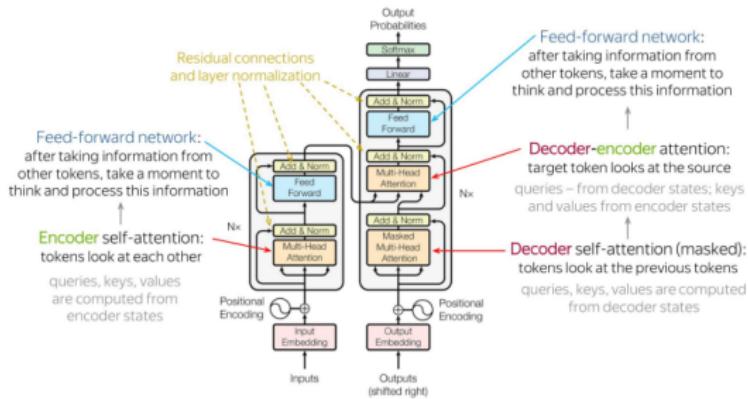
(Ref: Deus.ai <https://www.deus.ai/post/gpt-3-what-is-all-the-excitement-about>)

## Underlying Architecture: Transformers

- ▶ In its heart it contains an encoding component, a decoding component, and connections between them.
- ▶ The Transformer is a model that uses attention to boost the speed with which seq2seq with attention models can be trained.
- ▶ The biggest benefit, however, comes from how The Transformer lends itself to parallelization. How?



# Transformer Models

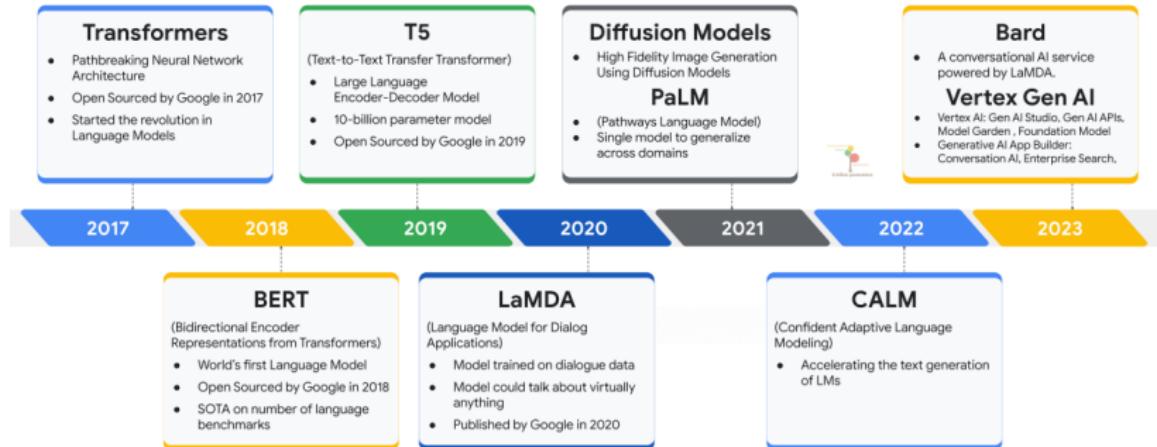


(Ref: The Complete Prompt Engineering for AI Bootcamp (2023))

- ▶ No recurrence, so parallelization possible
- ▶ Context information captured via attention and positional encodings
- ▶ Consists of stacks of layers with various sublayers

Transformers are basis of (the most) Large Language Models

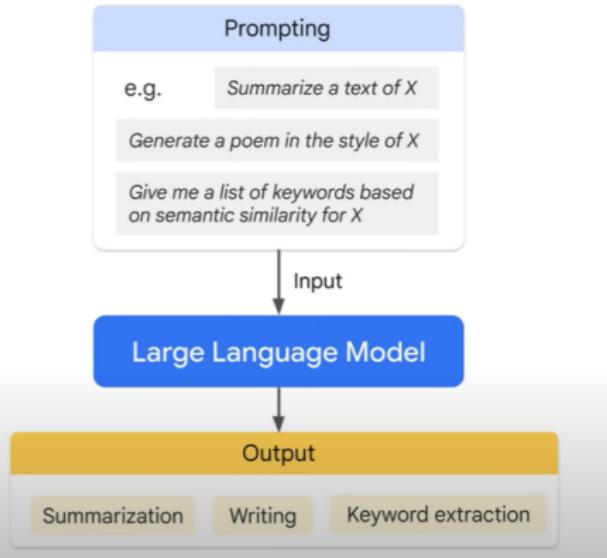
# The Progress of Models ...



(Ref: Primer on LLM and Gen AI - Google Cloud)

# Prompts driving Generative AI

**Prompt Design:**  
the quality of the  
input **determines** the  
quality of the output.



(Ref: Introduction to Generative AI - Google Cloud Tech)

# Model Types

## 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 texts (e.g. translation from one language to another).

## Applications

Generation

Classification

Summarization

Translation

(Re)Search

Extraction

Clustering

Content editing / rewriting

(Ref: Introduction to Generative AI - Google Cloud Tech)

# Model Types

## text-to-image

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

## Applications

Image generation

Image editing

(Ref: Introduction to Generative AI - Google Cloud Tech)

# 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 (for use in games or other 3D worlds).

Applications

Video generation

Video editing

Game assets

(Ref: Introduction to Generative AI - Google Cloud Tech)

# Model Types

## text-to-task

Text-to-task models are trained to perform a specific 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-task model could be trained to navigate web UI or make changes to a doc through the GUI.

## Applications

Software agents

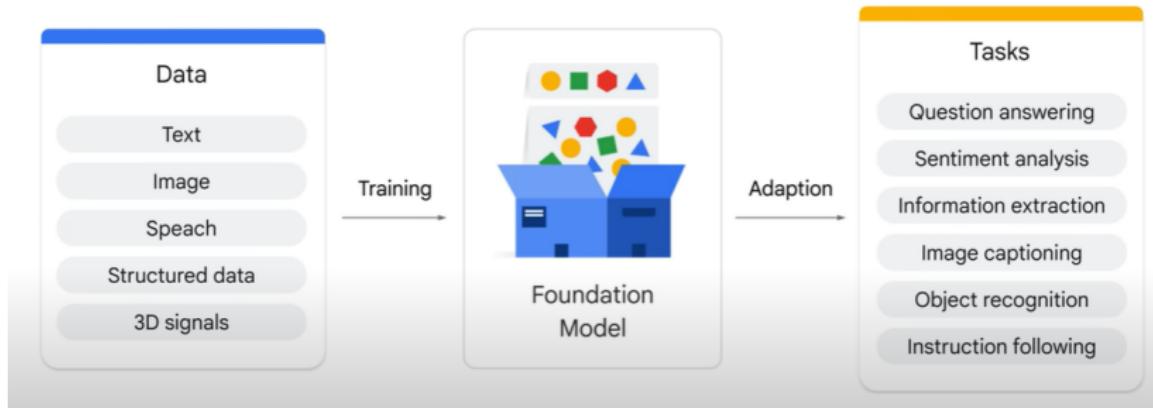
Virtual assistants

Automation

(Ref: Introduction to Generative AI - Google Cloud Tech)

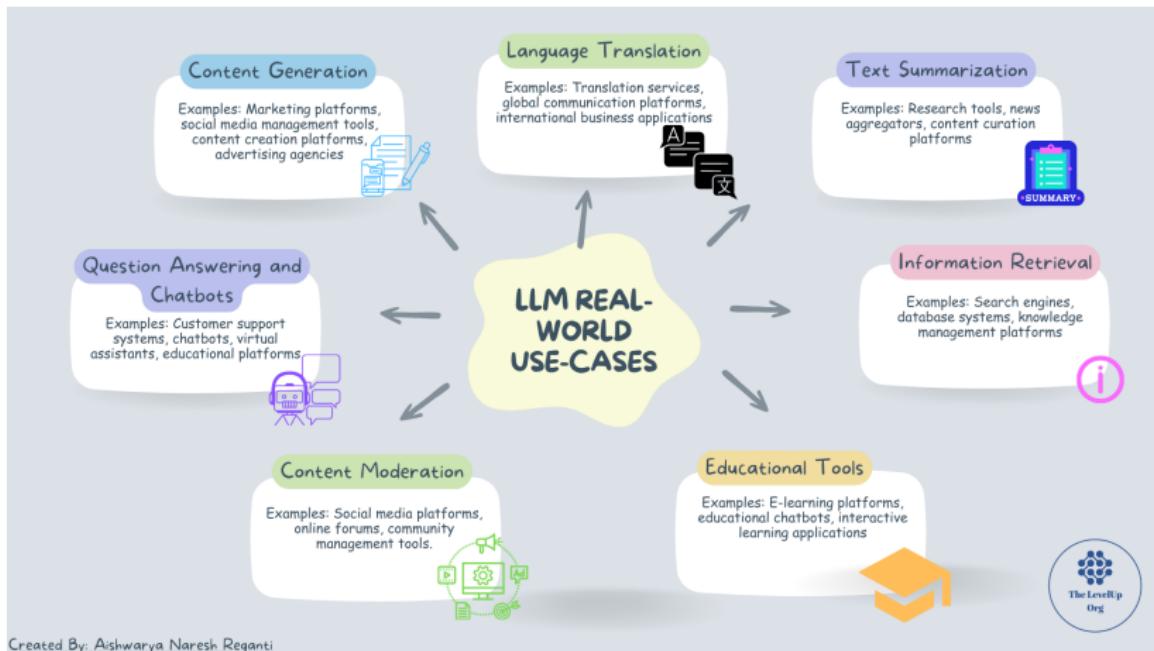
# Model Types

## Foundation Model:



(Ref: Introduction to Generative AI - Google Cloud Tech)

# LLM Real World Use Cases



Created By: Aishwarya Naresh Reganti

(Ref: Applied LLMs Mastery 2024 - Aishwarya Reganti)

# LLM Challenges



Created by: Aishwarya Naresh Reganti

(Ref: Applied LLMs Mastery 2024 - Aishwarya Reganti)

What's IN these days . . .

YHK

## 2021: Dall-E

### Zero-Shot Text-to-Image Generation

- ▶ OpenAI's 2021 release: DALL-E
- ▶ Text-to-image generation model
- ▶ Implementation of GPT-3
- ▶ Generates images from text descriptions
- ▶ Utilizes text-image dataset



"An astronaut riding a horse in a photorealistic style." Credit: OpenAI

(Ref: Example prompt "An astronaut riding a horse in photorealistic style" using DALL-E: Open AI)

## ChatGPT - A Tipping Point for Generative AI

- ▶ Released by OpenAI in November 2022
- ▶ Generative AI chatbot
- ▶ Rapid worldwide popularity
- ▶ 1 million users in 5 days
- ▶ Netflix took 3.5 years for same user count
- ▶ 100 million monthly active users by January 2023
- ▶ Fastest-growing application in history



# THE CHATGPT

## Versatile Capabilities of ChatGPT

- ▶ Generates human-like chat conversations
- ▶ Tasks: questions, creative content, explanations
- ▶ Suggestions, poems, code generation, and more

## Pivotal Moment in AI

- ▶ Demonstrated advanced language model sophistication
- ▶ Prompted activity in AI community
- ▶ Rush to develop similar language models
- ▶ Expectation of even more impressive models

# What's Inside?

## Technical Details

- ▶ Based on GPT3 architecture
- ▶ Estimated 175 billion parameters
- ▶ Fine-tuned on chat-specific task
- ▶ Curated dataset for fine-tuning
- ▶ Human AI trainers provide feedback
- ▶ Model adjusts responses based on feedback

## Specialization Through Fine-tuning

- ▶ Specializes in domains or tasks
- ▶ Incorporates human preferences and guidelines
- ▶ Examples: conversational data, customer support
- ▶ More suitable for interactive applications

# Improving ChatGPT with RLHF

- ▶ Key technique: Reinforcement Learning from Human Feedback (RLHF)
- ▶ Trains language model to align with human preferences
- ▶ Collects human feedback on model-generated text
- ▶ Updates model's parameters using feedback
- ▶ Enhances ChatGPT responses' quality
- ▶ Increases factual, informative, and creative output



## Midjourney: Image Generation Model

- ▶ Developed by Midjourney Inc.
- ▶ Released in July 2022
- ▶ Architecture details undisclosed
- ▶ High-quality image generation
- ▶ Wide variety of styles and genres

# Meta Releases LLaMA

## Open Source LLMs Explode!

- ▶ February 2023: Meta releases LLM "LLaMA"
- ▶ LLaMA: 65-billion parameter model
- ▶ Trained on extensive text and code dataset

## Significance of LLaMA Release

- ▶ One of the largest public LLMs
- ▶ Suited for complex and challenging tasks
- ▶ Open source, initially for research purposes
- ▶ Model weights leaked online, accessible to all
- ▶ Sparked development of numerous open source LLMs



# Anthropic Claude

The screenshot shows a user interface for generating product names. At the top, there's a header with a square icon, three dots, and a search bar containing 'Fit Shaker: Fast and Healthy'. Below the search bar are two buttons: 'Chat' and 'Edit Chat'. A light blue sidebar contains the following text:

- Product names: HomeShaker, Fit Shaker, QuickShake, Shake Maker
- Product description: A pair of shoes that can fit any foot size.
- Seed words: adaptable, fit, omni-fit
- Product names:

A small circular icon with a 'M' is positioned next to the 'Product names:' text. In the main area, a message says: 'Here are some suggested product names based on the seed words:' followed by a bulleted list:

- OmniFit
- AdaptiShoe
- UniSole
- One Size Fits All
- FlexiFit

At the bottom, there's a text input field with placeholder text 'Write a message...' and a blue send button with a white arrow icon. A small circular icon with a 'M' is also located near the bottom right of the main content area.

(Ref: The Complete Prompt Engineering for AI Bootcamp (2023))

Created by Anthropic  
<https://console.anthropic.com/>  
or API Uses Constitutional AI rather than RLHF  
Constitutional AI trains to follow a set of high-level principles or rules, such as a constitution, that specify the desired behavior and outcomes of the system. RLHF uses human feedback, such as ratings, preferences, or corrections, to optimize a language model or an agent's policy using reinforcement learning

## Github Copilot: Breakthrough Coding Assistant

- ▶ OpenAI introduced Github Copilot in 2021
- ▶ Built on GPT-3 architecture
- ▶ Fine-tuned on millions of public code lines
- ▶ Auto-completes and suggests code
- ▶ Supports multiple programming languages

# Is everything POSTIVE?

## Concerns and Considerations

- ▶ Raised intellectual property rights issues
- ▶ Trained on publicly available GitHub code
- ▶ Potential licensing conflicts
- ▶ Risk of suggesting code with vulnerabilities
- ▶ Biases and security flaws in training data

## Impact on Developer Community

- ▶ Enhanced productivity among developers
- ▶ Focus on higher-level tasks and problem-solving
- ▶ Reduced time spent on repetitive code
- ▶ Empowers creativity and innovation
- ▶ New perspectives in software development

# Google is not behind

## Google Releases Bard

- ▶ March 2023: Google introduces Bard chatbot
- ▶ Built on LLM framework
- ▶ Fine-tuned model based on PaLM 2
- ▶ PaLM 2: Around 340 billion parameters
- ▶ Bard trained on dataset 10 times larger than ChatGPT

## Bard's Unique Features

- ▶ Access to the internet
- ▶ Continuous learning and knowledge update
- ▶ More up-to-date and accurate than ChatGPT
- ▶ ChatGPT limited to training dataset info
- ▶ Bard surpasses ChatGPT's knowledge scope

# Bard — Palm 2/Gemini

 Product description: A home milkshake maker  
Seed words: fast, healthy, compact  
Product names: HomeShaker, Fit Shaker, QuickShake, Shake Maker

Product description: A pair of shoes that can fit any foot size.  
Seed words: adaptable, fit, omni-fit  
Product names:

 Here are some product names for a pair of shoes that can fit any foot size:

- Omni-Fit Shoes
- Adaptable Shoes
- Fit-All Shoes
- Size-Free Shoes
- Universal Shoes

(Ref: The Complete Prompt Engineering for AI Bootcamp (2023))

Created by Google Access it via chat <https://bard.google.com/> or encounter it in search results

# The BEST so far

## OpenAI Releases GPT4

- ▶ May 2023: OpenAI releases GPT4
- ▶ Multimodal: accepts image and text inputs
- ▶ Generates text outputs
- ▶ 6 months spent on safety and alignment
- ▶ GPT-4 82% less likely for disallowed content
- ▶ 40% more likely for factual responses than GPT-3.5

## GPT4's Performance

- ▶ Human-level performance on benchmarks
- ▶ Passed simulated bar exam in top 10%
- ▶ Impressive results on various professional exams

# Conclusions

## Advantages

- ▶ Conversational Abilities
- ▶ Solving Complex Problems
- ▶ Retaining Previous Information
- ▶ Will replace mundane language tasks, How to articles, homework, etc

## Dis-advantages

- ▶ Sensitive to Input Phrasing
- ▶ Cannot replace humans for innovation, for which data does not exist already
- ▶ Keeps “hallucinating”: Tends to write plausible but incorrect content with confidence
- ▶ May not get language structure right all the time, e.g try getting ghazal written

## Compared to humans?

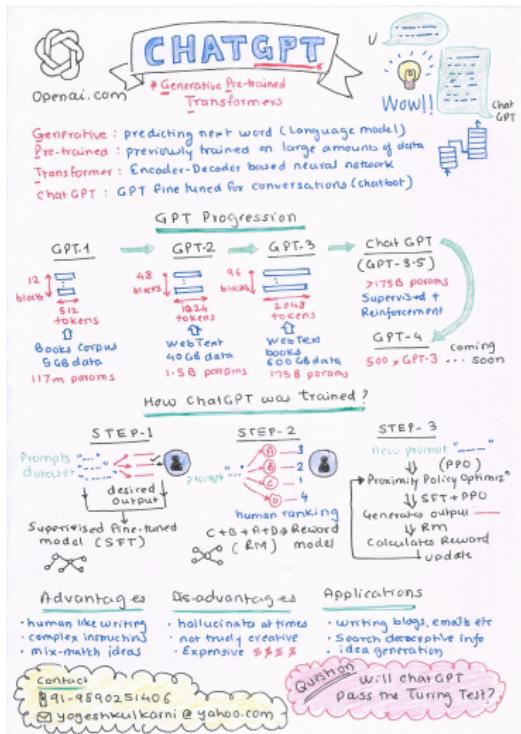
- ▶ Although GPT models are far more knowledgeable due to vast training (GPT-3 was trained on 499B tokens, GPT-4 on 1.4 T tokens whereas a human can read about 5.26B tokens in 80 years)
- ▶ It's not about the volume of data ...
- ▶ Children can talk long before they reach adulthood, and it's likely that < 1B tokens is enough for a human to master language. Need to figure out how nature has made her neural network so efficient.
- ▶ Imagine the pretraining done over billions of years at the time a baby comes ready with

## Future scope

- ▶ Planning and task loops (for handling multi hop queries)
- ▶ LLM based post processing (Entity Extractions)
- ▶ Temporal Querying (find abc from last 3 weeks)
- ▶ Resolving conflicting data (when two different values for capital of India)
- ▶ Improve explanability and interpret-ability

(Ref: Combining LLMs with Knowledge Bases to Prevent Hallucinations - Scott Mackie - LLMs in Prod Con 2 )

# My Sketchnote



(Ref: <https://medium.com/technology-hits/overview-of-chatgpt-95f4b43645c0>)

# The After-shocks

## ChatGPT vs Google: Who is better?



Google: almost latest data, gives source, not trainable, looks more accurate

ChatGPT: a bit old data, no source, trainable, hallucinates

(Ref: ChatGPT Explained: Complete A-Z Guide - Kripesh Advani)

# Will ChatGPT Kill Jobs?



Repetitive, boring and standard, language based jobs, for sure. Need to be more creative, experiential to stand against ChatGPT.

(Ref:ChatGPT Explained: Complete A-Z Guide - Kripesh Adwani)

## New Job Roles?

Prompt Engineer: Preparing input to AI effectively to get the desired answer.  
Will need to AI works in the background plus domain knowledge. Give context, examples etc to prime the model to give short specific answers than the usual page-long ones (davinci GPT3 in this case)

Playground Q&A Save View code Share ...

I am a highly intelligent question answering bot. If you ask me a question that is rooted in truth, I will give you the answer. If you ask me a question that is nonsense, trickery, or has no clear answer, I will respond with "Unknown".

Q: What is human life expectancy in the United States?  
A: Human life expectancy in the United States is 78 years.

Q: Who was president of the United States in 1955?  
A: Dwight D. Eisenhower was president of the United States in 1955.

Q: Which party did he belong to?  
A: He belonged to the Republican Party.

Q: What is the square root of banana?  
A: Unknown

Q: How does a telescope work?  
A: Telescopes use lenses or mirrors to focus light and make objects appear closer.

Q: Where were the 1992 Olympics held?  
A: The 1992 Olympics were held in Barcelona, Spain.

Mode: text-davinci-003  
Temperature: 0  
Maximum length: 100  
Stop sequences: Enter sequence and press Tab  
Top P: 1  
Frequency penalty: 0  
Presence penalty: 0

(Ref: Advanced ChatGPT Guide - How to build your own Chat GPT Site - Drian Twarog)

## Being 'AI Engineer'

- ▶ "AI Engineers" bridge software engineering and AI research.
- ▶ Strong engineering foundations, focus on AI applications.
- ▶ Connect AI building blocks, lack deep ML backgrounds.

## AI API and Open Source Models

- ▶ Proliferation of powerful AI APIs and open source models.
- ▶ Companies like OpenAI, Anthropic, Cohere, Stability AI.
- ▶ Abstracted model training, access to cutting-edge AI via APIs.
- ▶ "AI Engineers" create user experiences using these tools.

## Examples of AI Engineering

- ▶ Indie developers using LangChain, LlamaIndex for AI apps.
- ▶ Startups leverage Claude, GPT for rapid idea validation.
- ▶ Agile methodologies influenced by AI's emergence.

## Implications for the Future

- ▶ Democratization of AI: Any competent engineer can build AI products.
- ▶ Faster iteration: Quick testing accelerates innovation and user-focused designs.
- ▶ Job market evolution: AI Engineer roles in demand, similar to data engineers and scientists.

## Challenges for AI Engineering

- ▶ Specification and design: Rigorous design thinking for productization.
- ▶ UX and AI: Blend UX and ML for clear prompting, explainability, error handling.
- ▶ Robustness and safety: Rigorous testing, monitoring, and governance for model deployment.
- ▶ Privacy and security: Safeguards against abuse, increasing importance of AI security.

## Preparing for AI Engineering

- ▶ Engineering leaders must build AI capabilities.
- ▶ Companies level up through learning, strategic hiring, partnerships.
- ▶ AI expertise becoming essential for top-tier products.

# The Hype



Elon Musk ✅ @elonmusk · Dec 4

Replies to @sama

ChatGPT is scary good. We are not far from dangerously strong AI.

...



François Chollet ✅ @fchollet · 8h

With apps like Midjourney, Lensa, and ChatGPT, the age of consumer AI has finally arrived.

...



@jh@sigmoid.social (Mastodon) @jeremyphoward · 9h

...

I remember well when Google was first released. I felt confident that it would become a critical tool that we'd all rely on.

I haven't felt that way again, until now: ChatGPT.

In just a few days I've gotten so much out of it.

(Ref: ChatGPT - Intro & Potential Impact Sudalai Rajkumar, SRK)

## Finally, from horses mouth!!



Sam Altman  @sama · 10 dic.

ChatGPT is incredibly limited, but good enough at some things to create a misleading impression of greatness.

it's a mistake to be relying on it for anything important right now. it's a preview of progress; we have lots of work to do on robustness and truthfulness.

877 3.856 27,5 mil

(Ref: ChatGPT: training process, advantages, and limitations - By Sergio Soage, Machine Learning Engineer at Aivo)

## References

- ▶ Introduction to Generative AI - Google Cloud Tech
- ▶ Generative AI Presentation - Laura Worden

## Newsletters to subscribe

- ▶ **The Batch by DeepLearning.AI:**
  - ▶ Summarizes diverse AI news with nuanced viewpoints.
  - ▶ Andrew Ng's thought leadership adds significant value.
- ▶ **The Rundown AI by Rowan Cheung:**
  - ▶ Go-to for generative AI events and product innovations.
  - ▶ Quick rundown with bullet point details for easy comprehension.
- ▶ **AI Supremacy by Michael Spencer:**
  - ▶ Personal writing style with in-depth exploration.
  - ▶ Offers multiple perspectives on AI topics.
- ▶ **Ahead of AI by Sebastian Raschka, PhD:**
  - ▶ Technical focus covering applied deep learning and generative AI.
  - ▶ Valuable insights for those seeking in-depth technical content.
- ▶ **To Data and Beyond by Youssef Hosni:**
  - ▶ Resource hub for hands-on projects, learning roadmaps, and research papers.
  - ▶ Ideal for those looking to dive into practical aspects of AI.



Thanks ...

- ▶ Search "**Yogesh Haribhau Kulkarni**" on Google and follow me on LinkedIn and Medium
- ▶ Office Hours: Saturdays, 2 to 5pm (IST); Free-Open to all; email for appointment.
- ▶ Email: yogeshkulkarni at yahoo dot com



(Generated by Hugging Face QR-code-AI-art-generator,  
with prompt as "Follow me")