



---

# Demystifying Generative AI

FOCUSING ON NLP

MAY 16, 2023





## Abinaya Mahendiran

Program Manager, IITM

AI/ML Consultant (Freelancer)

M.Tech IT, International Institute of Information and Technology Bangalore

Mentor at TeamEpic, Springboard, Topmate.io

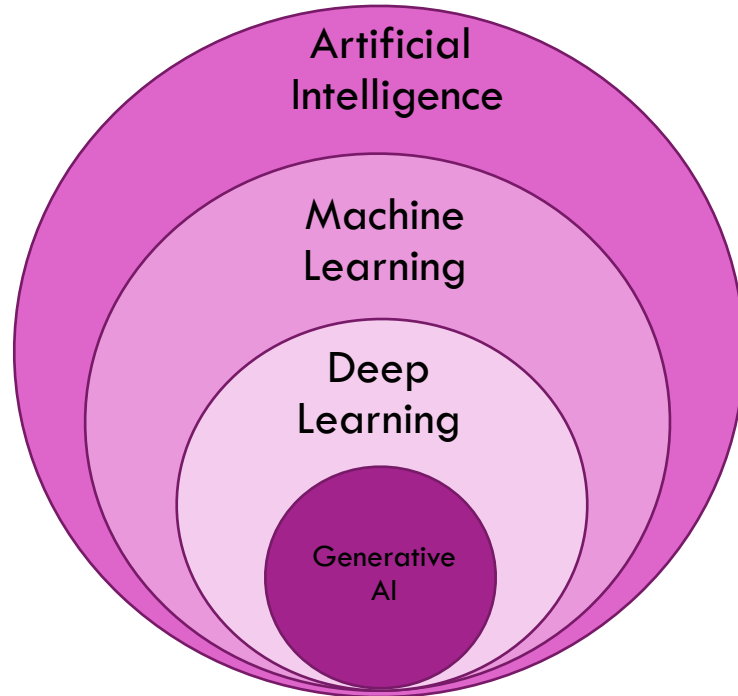
Volunteer at WTM, Data Conversations, GHCI, WAI, Women Who Code

Interests : Building NLP/NLU/NLG/MLOps systems, Open source, Applied Research

# Agenda

- What is Generative AI?
- Types of Models
- Traditional Approach
- Transfer Learning
- Types of Fine-tuning
- Foundation Models
- NLP Timeline
- Concepts
- Limitations of Generative AI

# What is Generative AI?



- AI:** Build intelligent agents that can act like humans autonomously.

- ML:** A machine learns the patterns in the data by training a model.

- Supervised learning** – Use labeled data, train models, predict on unseen data.

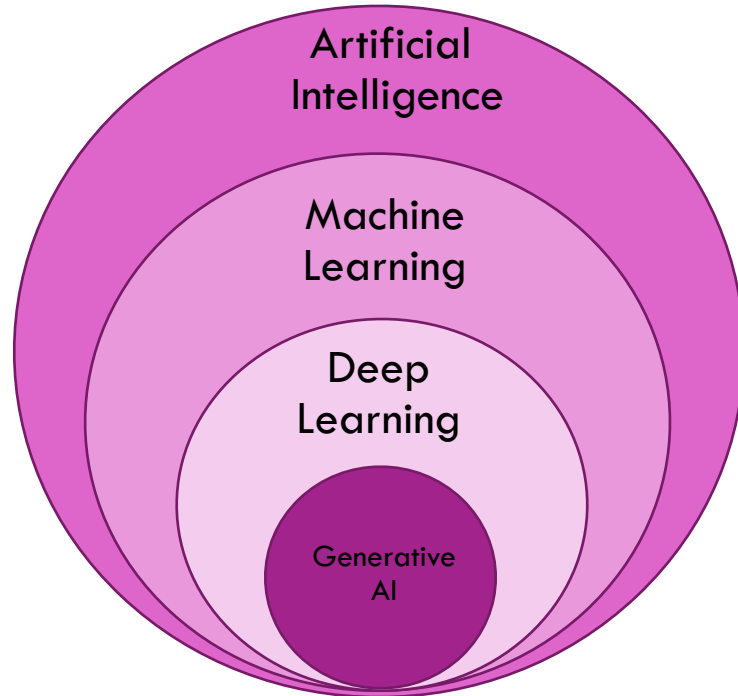
- Classification/Regression

- Unsupervised learning** – Use unlabeled data to identify groups or clusters.

- Semi supervised learning** – uses little labeled data and more unlabeled data to train models.

- Reinforcement Learning** – An agent performs actions based on the environment and learns through trial and error (either rewarded or punished).

# What is Generative AI?



• **DL:** A neural network with interconnected nodes and layers is trained to learn complex patterns in the data.

- Uses supervised, Unsupervised and Semi supervised methods of learning.

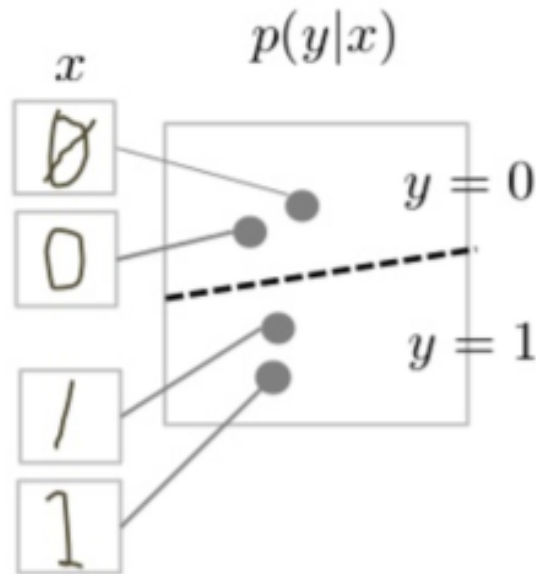
• **Generative AI:** It is a type of AI that can create new content, such as text, images, audio, and video.

- Learns from existing data and then uses that knowledge to generate new and unique outputs.

# Types of Models

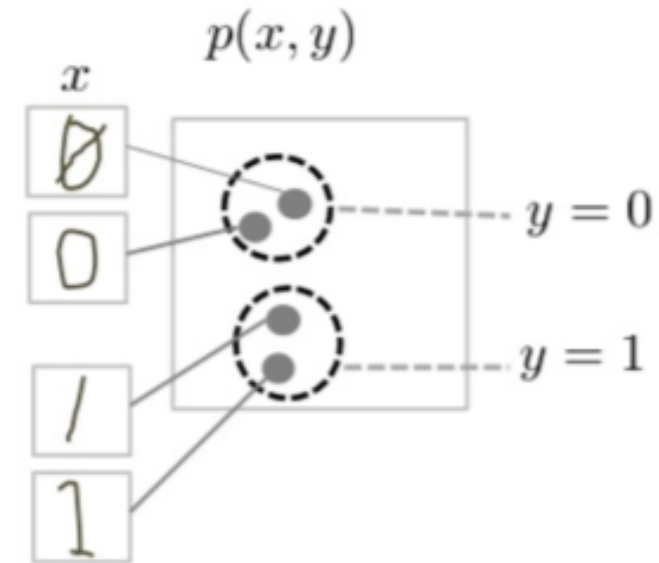
## Discriminative Models:

- Discriminates between different classes.
- Approximates the decision boundary or the distribution function, given the data points it predicts the labels.

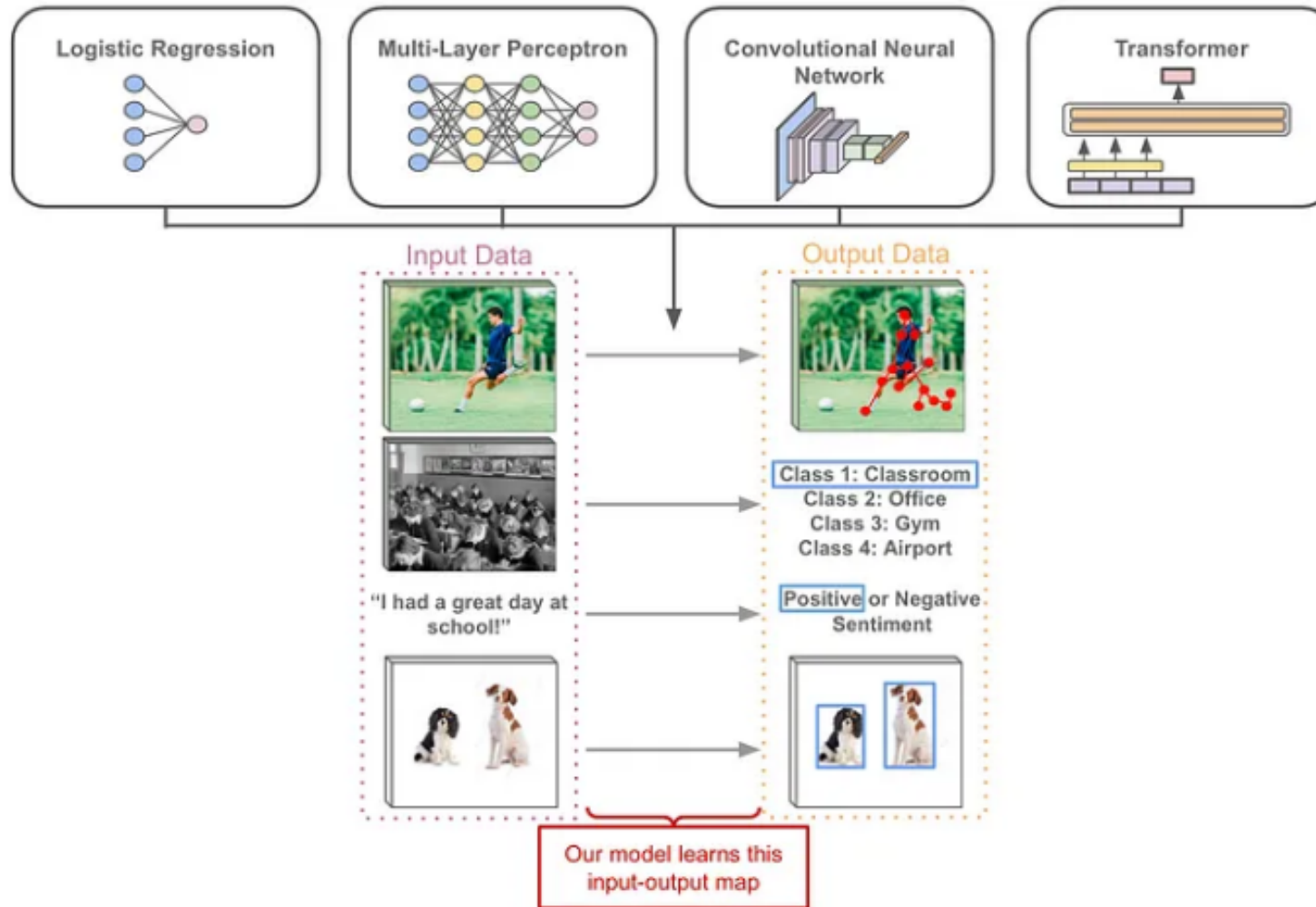


## Generative Models:

- Generates new data points.
- Assumes the data distribution and produces convincing data points that are close to its real counterparts in the space.



# Traditional Approach



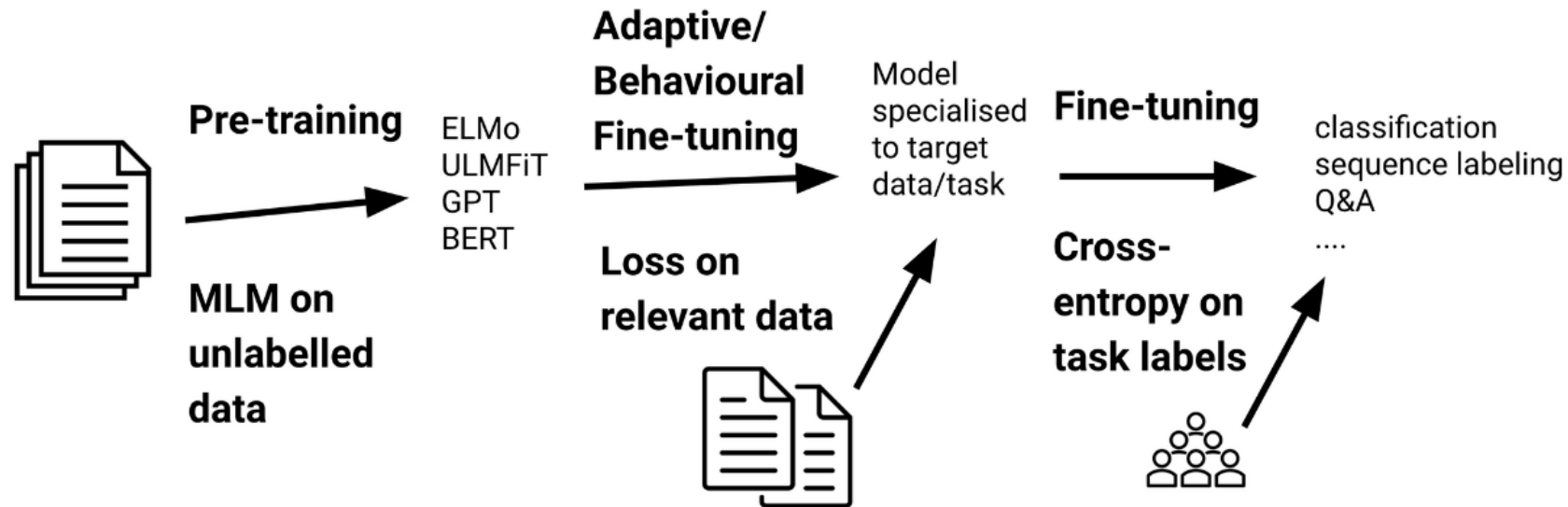
# Transfer Learning

## Step 1: Pre-training

- Use large amounts of generic data and train on a specific objective function.
- Unlabeled data is used to train on the language modelling objective like MLM.

## Step 2: Fine-tuning

- Fine-tuning is done using task-specific objective function.
- Labelled data is used to fine-tune model on the downstream tasks.



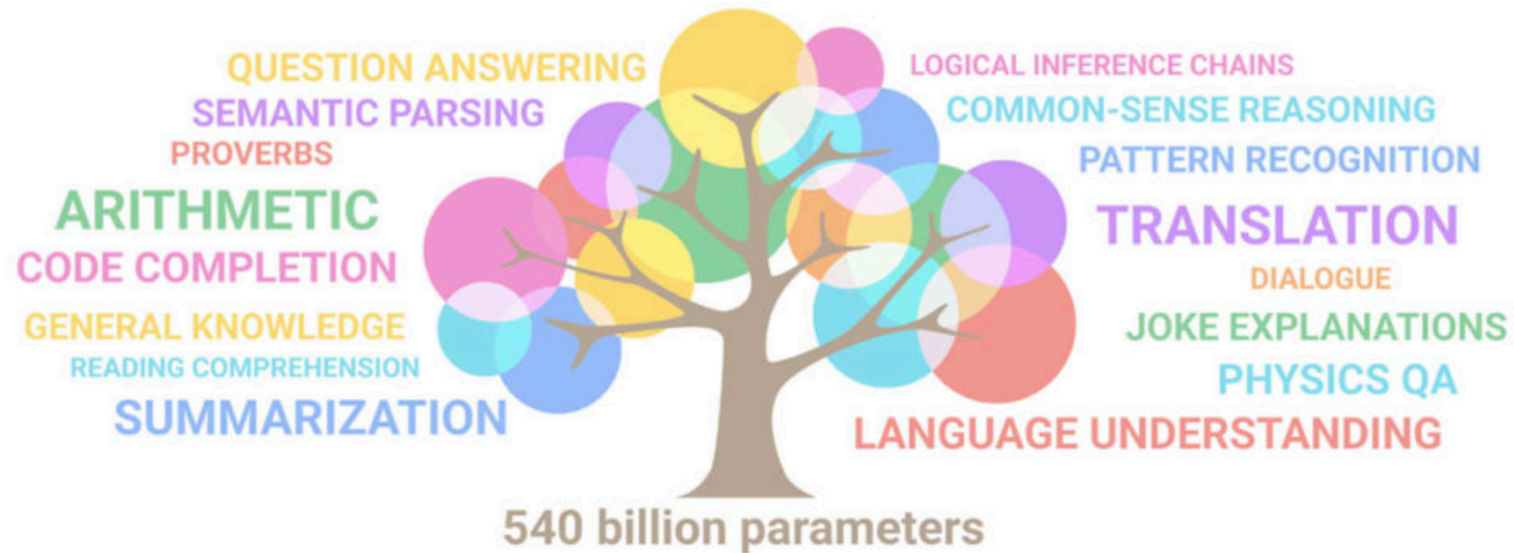


# Types of Fine Tuning

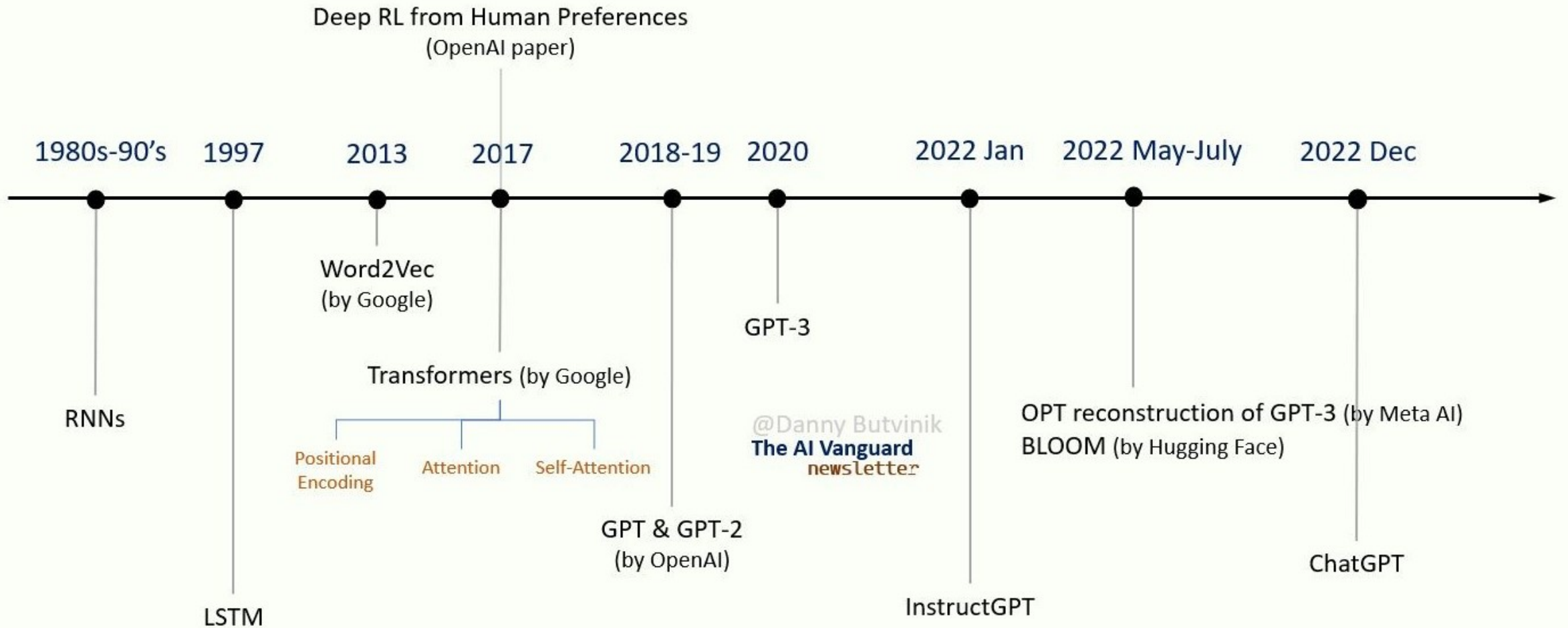
Category	Methods	Motivation
<b>Adaptive Fine-tuning</b>	Domain/task/ language adaptive fine-tuning	Specialise to target domain
<b>Behavioural Fine-tuning</b>	Intermediate-task training, self-supervised, frame as MLM	Specialise to target task
<b>Parameter-efficient Fine-tuning</b>	Adapters, sparse parameter permutations, pruning	Reduce space of fine-tuned models
<b>Text-to-text Fine-tuning</b>	Frame as text-to-text, prompt engineering, controllable NLG	Effectively use large autoregressive pre-trained LMs
<b>Mitigating Fine-tuning Instabilities</b>	Stop runs early, use a small lr, regularisation, avoid random init	Reduce variance of fine-tuning runs

# Foundation Models

- A foundation model is a large AI model pre-trained on a vast quantity of unlabelled data that was "designed to be adapted" (or fine-tuned) to a wide range of downstream tasks, such as sentiment analysis, image captioning, and object recognition.



# NLP Timeline



# Concepts: Language Modelling

Input  
Features

Trained Language Model

**Task:**  
Predict the next word

Output  
Prediction

Thou →

shalt →

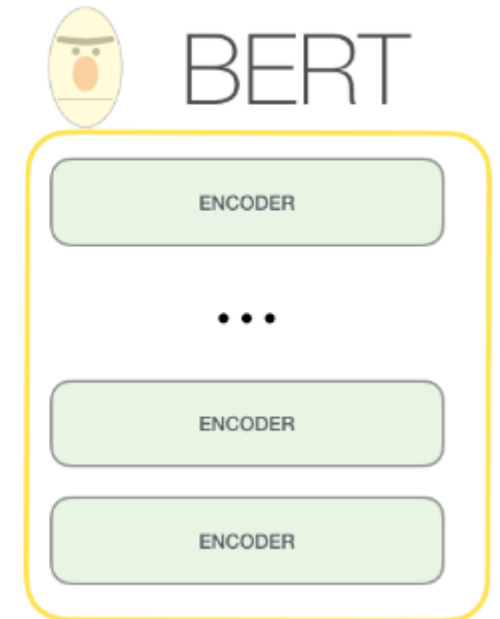
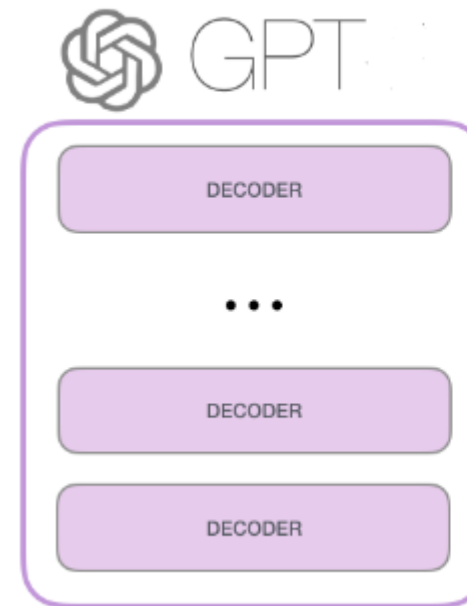
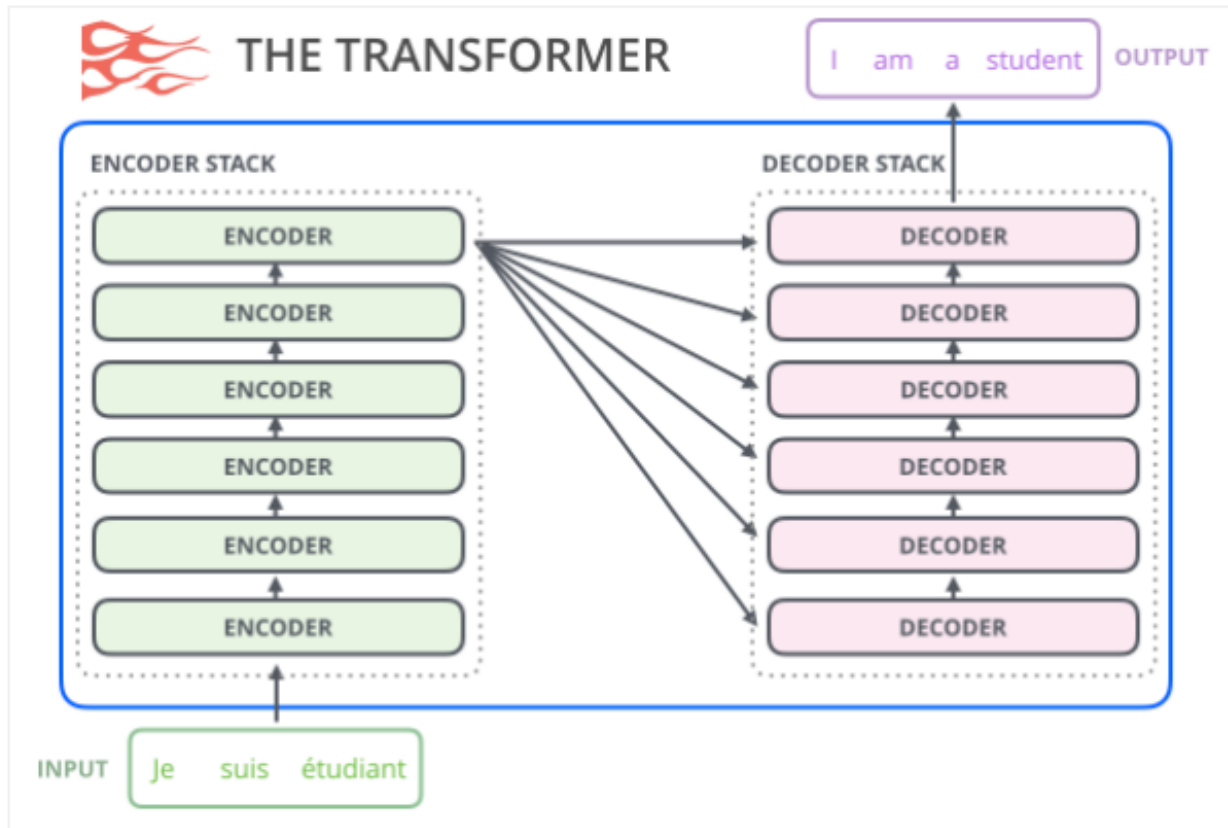
1) Look up  
embeddings

2) Calculate  
prediction

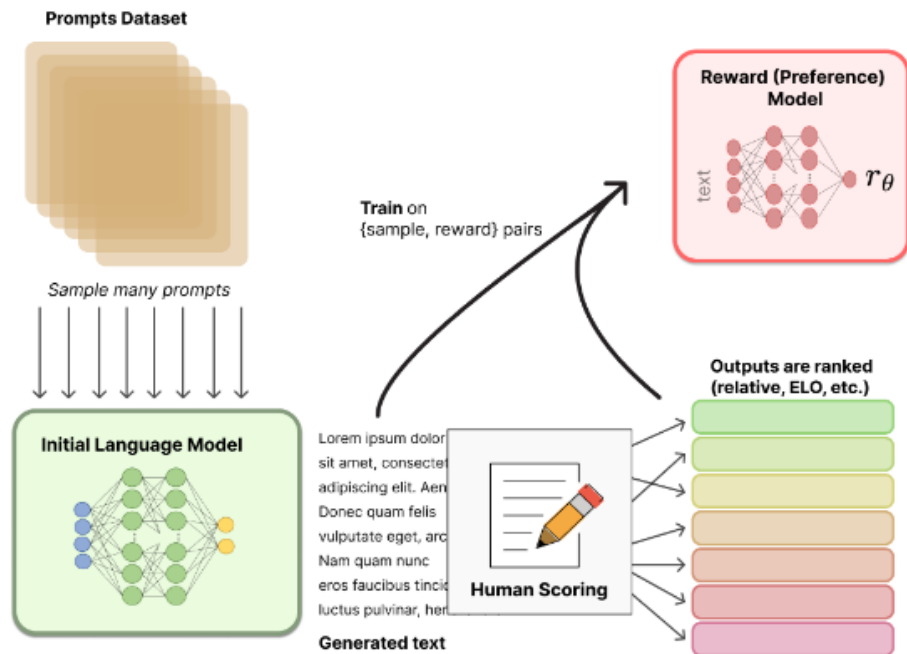
3) Project  
to output  
vocabulary

0	aardvark
0	aarhus
0.001	aaron
...	
0.4	not
...	
0.0001	zyzzyva

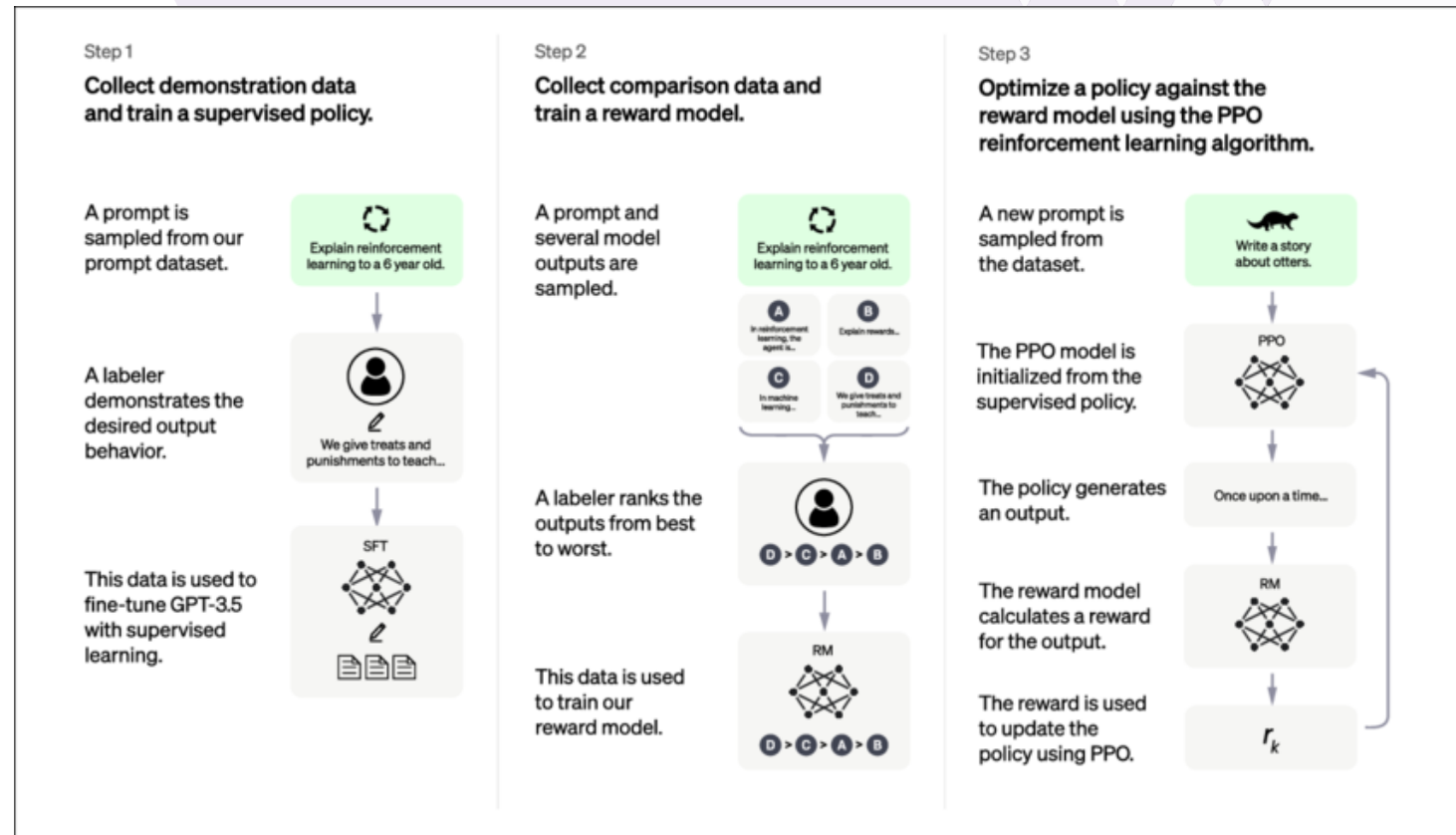
# Concepts: Encoder / Decoder



# Concepts: RLHF



<https://huggingface.co/blog/rlhf>



The ChatGPT training process. The figure is from OpenAI (2022a).

## Concepts: Prompt

A prompt is a short piece of text that is given to the large language model as input, and it can be used to control the output of the model in many ways.

A prompt contains any of the following elements:

**Instruction** - a specific task or instruction you want the model to perform

**Context** - external information or additional context that can steer the model to better responses

**Input Data** - the input or question that we are interested to find a response for

**Output Indicator** - the type or format of the output.

# Concepts: Prompt Engineering

## Zero-shot Prompting

*Prompt:*

```
Classify the text into neutral, negative or positive.  
Text: I think the vacation is okay.  
Sentiment:
```

*Output:*

```
Neutral
```

## Few-shot Prompting

*Prompt:*

```
A "whatpu" is a small, furry animal native to Tanzania. An example of a sentence that uses  
the word whatpu is:  
We were traveling in Africa and we saw these very cute whatpus.  
To do a "farduddle" means to jump up and down really fast. An example of a sentence that uses  
the word farduddle is:
```

*Output:*

```
When we won the game, we all started to farduddle in celebration.
```



# Concepts: Prompt Engineering

## Chain-of-Thought Prompting

### Standard Prompting

#### Model Input

Q: Roger has 5 tennis balls. He buys 2 more cans of tennis balls. Each can has 3 tennis balls. How many tennis balls does he have now?

A: The answer is 11.

Q: The cafeteria had 23 apples. If they used 20 to make lunch and bought 6 more, how many apples do they have?

#### Model Output

A: The answer is 27. ❌

### Chain-of-Thought Prompting

#### Model Input

Q: Roger has 5 tennis balls. He buys 2 more cans of tennis balls. Each can has 3 tennis balls. How many tennis balls does he have now?

A: Roger started with 5 balls. 2 cans of 3 tennis balls each is  $5 + 6 = 11$ . The answer is 11.

Q: The cafeteria had 23 apples. If they used 20 to make lunch and bought 6 more, how many apples do they have?

#### Model Output

A: The cafeteria had 23 apples originally. They used 20 to make lunch. So they had  $23 - 20 = 3$ . They bought 6 more apples, so they have  $3 + 6 = 9$ . The answer is 9. ✅

### (a) Few-shot

Q: Roger has 5 tennis balls. He buys 2 more cans of tennis balls. Each can has 3 tennis balls. How many tennis balls does he have now?

A: The answer is 11.

Q: A juggler can juggle 16 balls. Half of the balls are golf balls, and half of the golf balls are blue. How many blue golf balls are there?

A:

(Output) The answer is 8. ❌

### (b) Few-shot-CoT

Q: Roger has 5 tennis balls. He buys 2 more cans of tennis balls. Each can has 3 tennis balls. How many tennis balls does he have now?

A: Roger started with 5 balls. 2 cans of 3 tennis balls each is 6 tennis balls.  $5 + 6 = 11$ . The answer is 11.

Q: A juggler can juggle 16 balls. Half of the balls are golf balls, and half of the golf balls are blue. How many blue golf balls are there?

A:

(Output) The juggler can juggle 16 balls. Half of the balls are golf balls. So there are  $16 / 2 = 8$  golf balls. Half of the golf balls are blue. So there are  $8 / 2 = 4$  blue golf balls. The answer is 4. ✅

### (c) Zero-shot

Q: A juggler can juggle 16 balls. Half of the balls are golf balls, and half of the golf balls are blue. How many blue golf balls are there?

A: The answer (arabic numerals) is

(Output) 8 ❌

### (d) Zero-shot-CoT (Ours)

Q: A juggler can juggle 16 balls. Half of the balls are golf balls, and half of the golf balls are blue. How many blue golf balls are there?

A: Let's think step by step.

(Output) There are 16 balls in total. Half of the balls are golf balls. That means that there are 8 golf balls. Half of the golf balls are blue. That means that there are 4 blue golf balls. ✅

# Limitations of Generative AI

- Hallucinations are words or phrases that are generated by the model that are often nonsensical or grammatically and factually incorrect.
  - The model is not trained on enough data. Misleading information.
  - The model is trained on noisy or dirty data. Garbage in => Garbage out!
  - The model is not given enough context. Misleading information.
  - The model is not given enough constraints. Anyone can use it.
- Ethical concerns – what if the models are biased and are used for unintended purpose.
- Cost and Time – Costly and takes longer to build your own LLMs.
- Explainability is difficult.

Thank You :)

Questions?



---

At AnitaB.org, we envision a future where the people who imagine and build technology mirror the people and societies for whom they build it. We connect, inspire, and guide women and non-binary people in computing, and organizations that view technology innovation as a strategic imperative.

