

# AI, LLM and ChatGPT: A Deep Dive

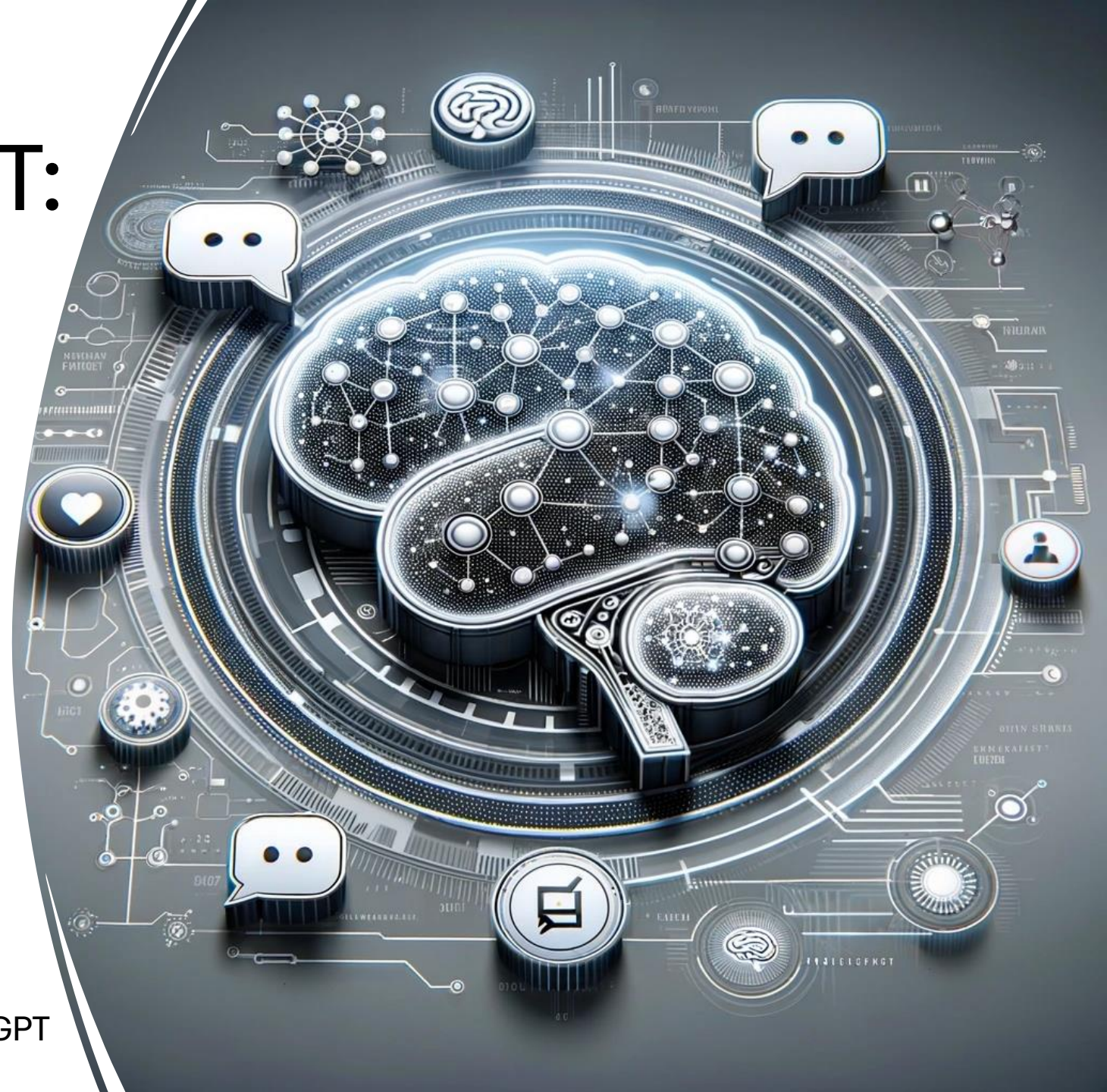
How to use it?

<https://www.youtube.com/watch?v=EWvNQjAaOHw>

Deep Dive :

<https://www.youtube.com/watch?v=7xTGNNLPyMI>

This image was generated by ChatGPT



# Agenda

---

01

Part 1: AI  
history and a  
mental model

02

Part 2: Pre-  
training stage

03

Part 3:  
Supervised  
Fine-Tuning  
(SFT)

04

Part 4:  
Reinforcement  
Learning (RL)

# Part 1 Agenda

AI  
History

ChatGPT

The big  
Picture

A mental  
model



[https://www.youtube.com/watch?v=WFR3lOm\\_xhE](https://www.youtube.com/watch?v=WFR3lOm_xhE)

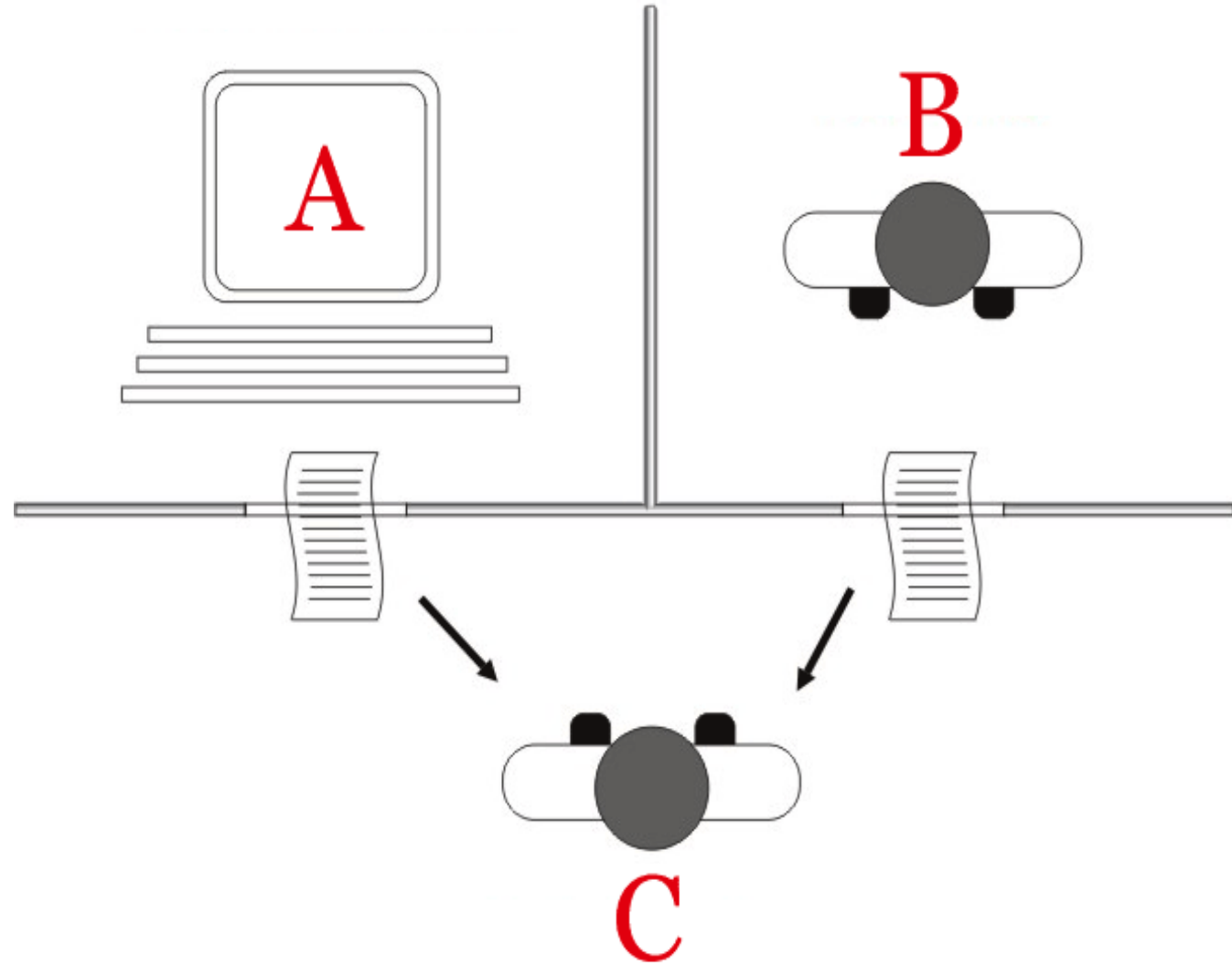
## Watson and the Jeopardy! Challenge 2011

# Turing Test

---

- The **Turing test**, originally called the **imitation game** by [Alan Turing](#) in 1950, is a test of a machine's ability to exhibit intelligent behaviour equivalent to, or indistinguishable from, that of a human.
- The player C, an interrogator, is given the task of trying to determine which player – A or B – is a computer and which is a human. The interrogator is limited to using the responses to written questions to make the determination

Source: [https://en.wikipedia.org/wiki/Turing\\_test](https://en.wikipedia.org/wiki/Turing_test)

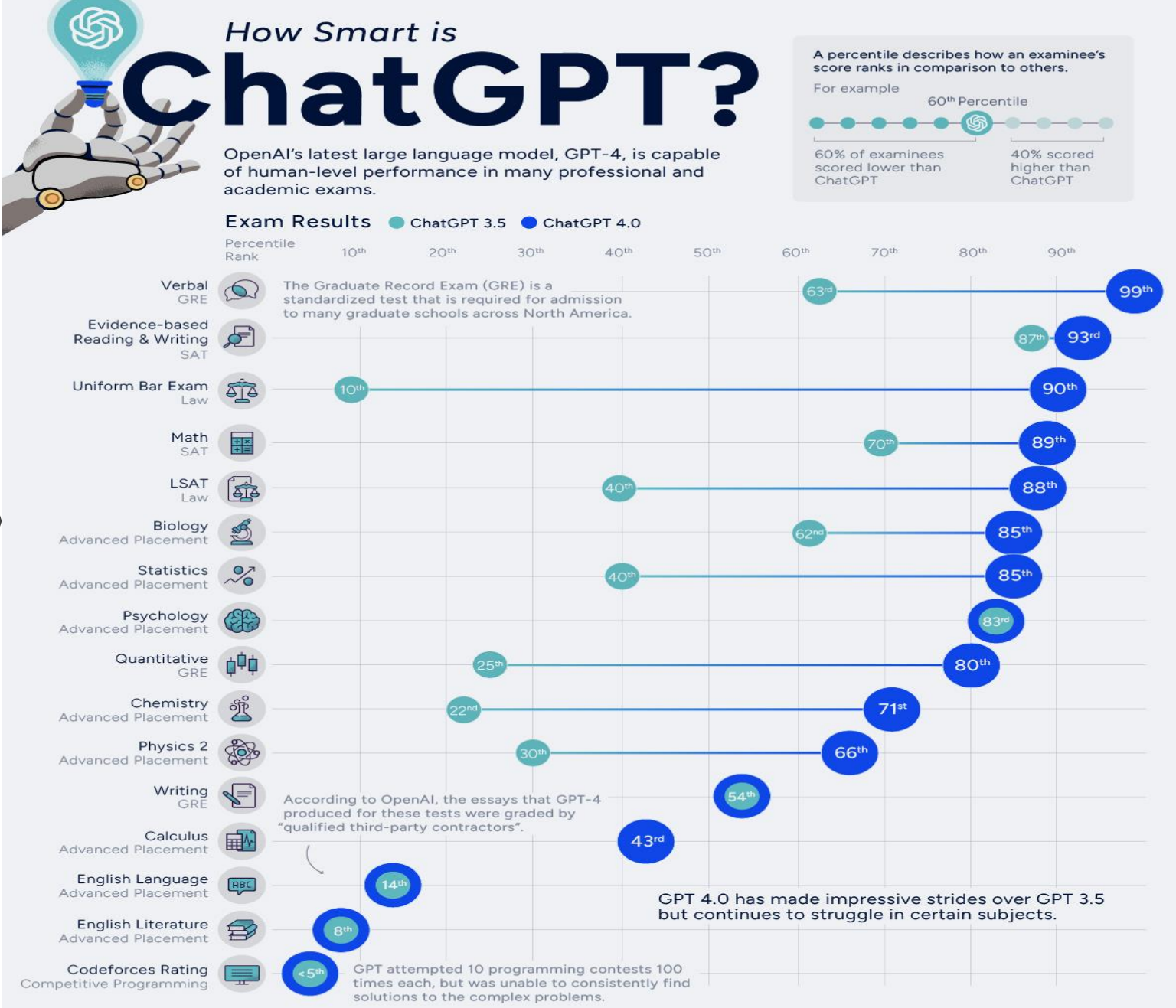


# History

- ChatGPT is a product of OpenAI's ongoing advancements in natural language processing and artificial intelligence
- Versions
  - GPT-1 was first introduced in 2018.
  - GPT-2 in 2019 showcases significant improvements in generating coherent and contextually relevant text.
  - GPT-3, released in June 2020, further expanded the model's capabilities with 175 billion parameters, making it one of the most powerful language models to date.
  - GPT-4, released in March 2023, has larger undisclosed parameter size.
  - GPT-4o, released on 05/13/2024, supports multimodal (image and voice) model. **Training data is up to Oct 2023.**

# GPT 4.0: A Milestone in 2023

Diagram on 04/28/2023



# Google AI Achieves Silver-Medal in Math Olympiad

July 25, 2024  
<https://deepmind.google/discover/blog/ai-solves-imo-problems-at-silver-medal-level/>

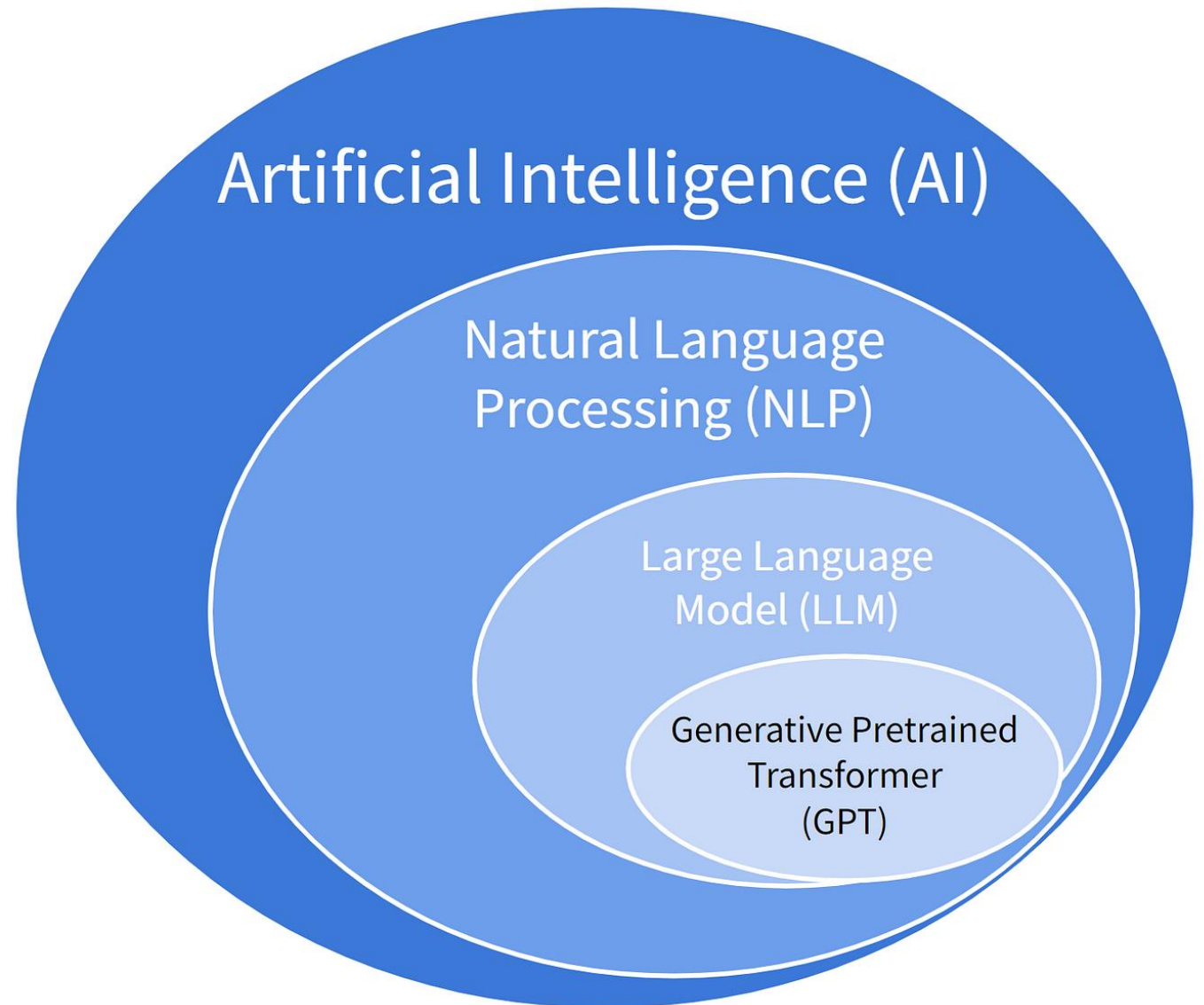
Score on IMO 2024 problems



# GPT, LLM, NLP, AI

---

- GPT is an example of a Large Language Model (LLM).
- LLM is used in natural language processing (NLP) and AI.
- Other LLM applications that may use other models
  - Text summarization
  - Sentiment analysis
  - Translation
  - Information retrieval
  - Text classification

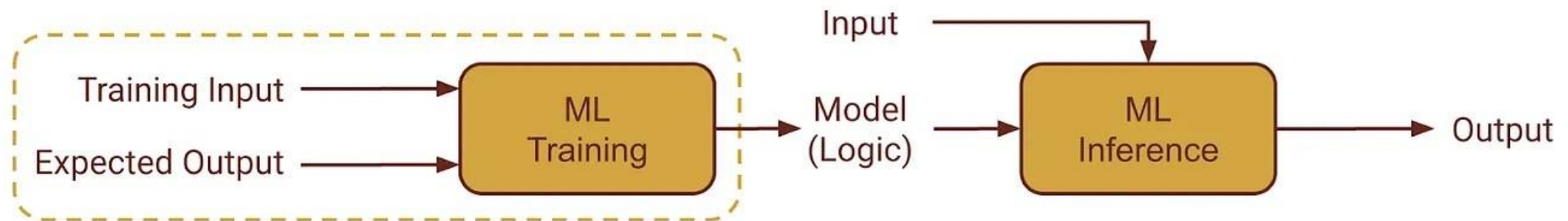


# Programming vs Machine Learning

**Traditional Programs:** Define algo/logic to compute output



**Machine Learning:** Learn model/logic from data



fuzzy boundaries

### rule-based decision making

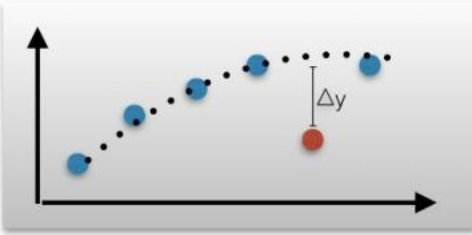
*if condition fulfilled then activity 1  
else activity 2*

**boolean data**  
(yes or no)

**Examples:**

- ▶ phone notification
- ▶ time- or threshold-based alarms
- ▶ simple pattern matching

### statistical reasoning



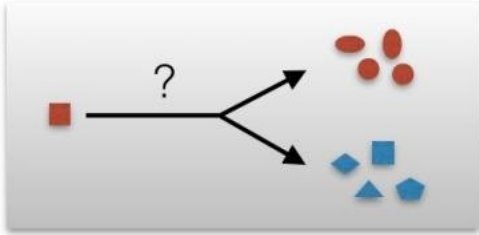
simple regression

**numerical data**  
allowing for curve fitting

**Examples:**

- ▶ extra- and interpolation
- ▶ outlier detection
- ▶ predictive maintenance

### machine learning



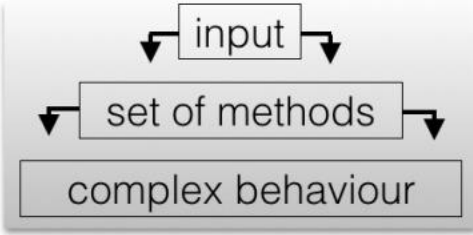
classification tasks

**arbitrary data**  
that needs to be abstracted into numbers

**Examples:**

- ▶ identification of relevant features from large input datasets
- ▶ quality control using various metrics

### artificial intelligence



dynamic adaptation to novelty

autonomous selection of best methodology when presented with **arbitrary data**

**Examples:**

- ▶ autonomous vehicles
- ▶ human-like conversational skills
- ▶ intelligent digital assistant

every programmer

data science types

complex systems specialists

# Does LLM Have Intelligence?

## Can Submarine Swim?

- A large language model (LLM) is a language model notable for its ability to achieve general-purpose language understanding and generation.
- LLMs acquire these abilities by learning statistical relationships from text documents during a computationally intensive self-supervised and semi-supervised training process.
- LLMs are artificial neural networks following a transformer architecture.

Source: [https://en.wikipedia.org/wiki/Large\\_language\\_model](https://en.wikipedia.org/wiki/Large_language_model)

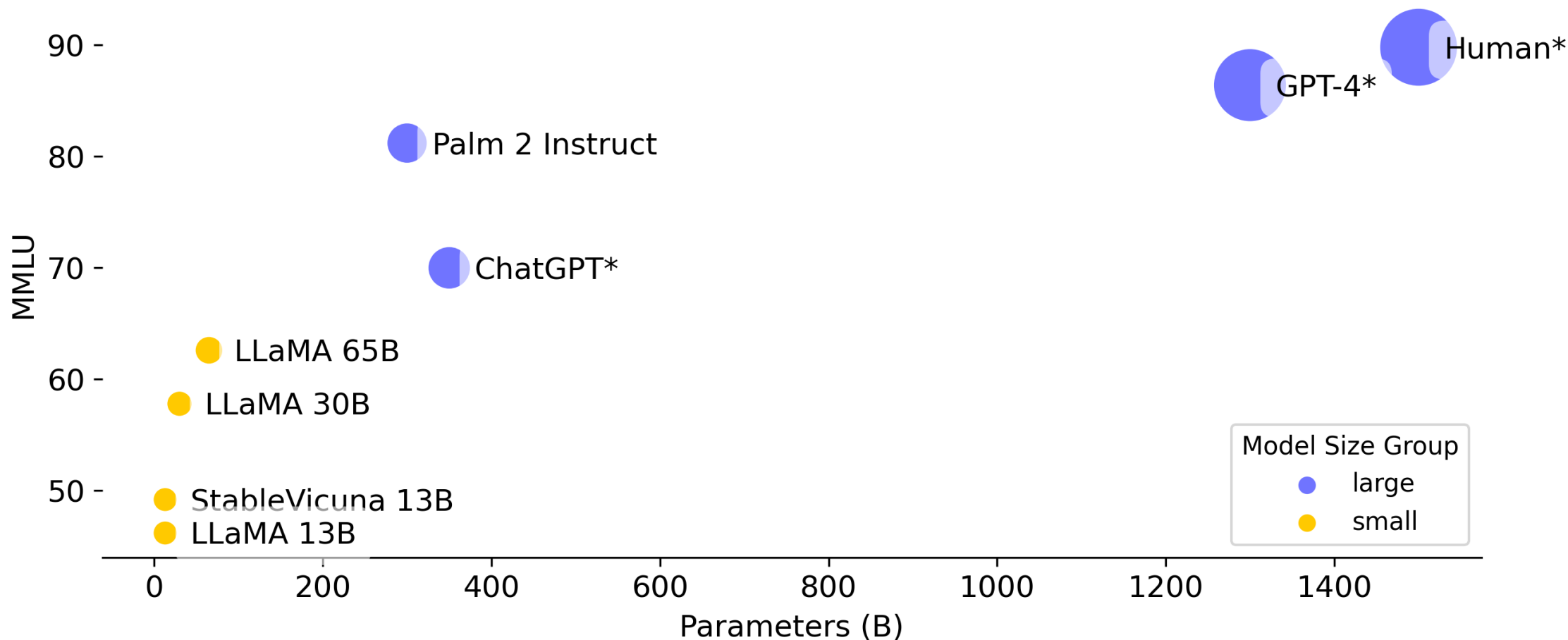
## Measuring Multitask Language Understanding (MMLU)

MMLU consists of 15,908 multiple-choice questions that span across 57 subjects, from highly complex STEM fields and international law, to nutrition and religion. It was one of the most commonly used benchmarks for comparing the capabilities of large language models, with over 100 million downloads as of July 2024

Source <https://en.wikipedia.org/wiki/MMLU>

# Measuring Multitask Language Understanding (MMLU)

## Model Performance (MMLU)



\*Exact model size is unknown. | Data from InstructEval GitHub.

<https://newsletter.victordibia.com/p/understanding-size-tradeoffs-with>

# A User's View: What Are You?

- **Human (me on 02/11/2025):** What are you?
- **AI Assistant:** I'm ChatGPT, **an AI assistant** designed to help with a wide range of tasks, from **answering questions and generating ideas** to assisting with research and technical topics. **I remember that you're** a professor of Information Systems working on the future Internet, and I can support you with related discussions, research insights, and more. Let me know how I can help!

First Name

Job - a "conversationalist" to human customers

Chat GPT

Innate learned abilities - "innate potentials"

Family Name

G: Generative  
P: Pre-Trained  
T: Transformer

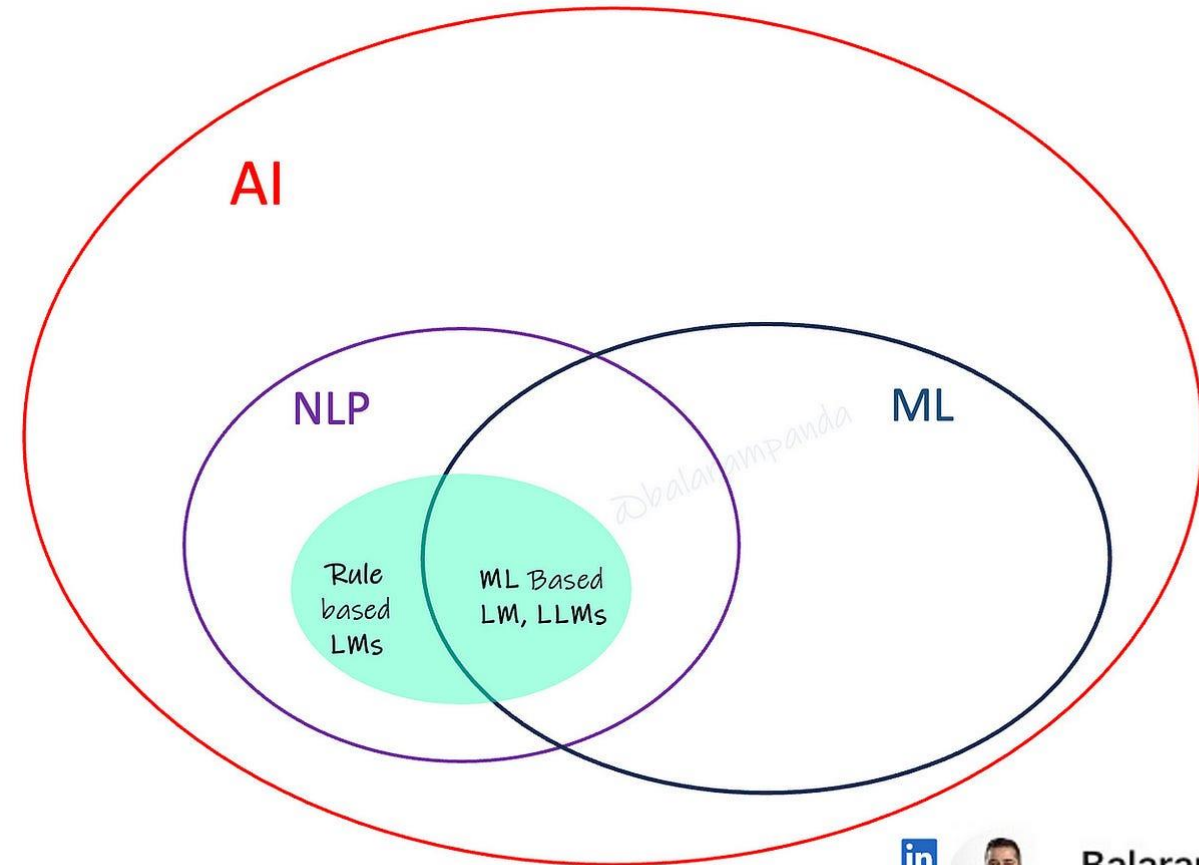
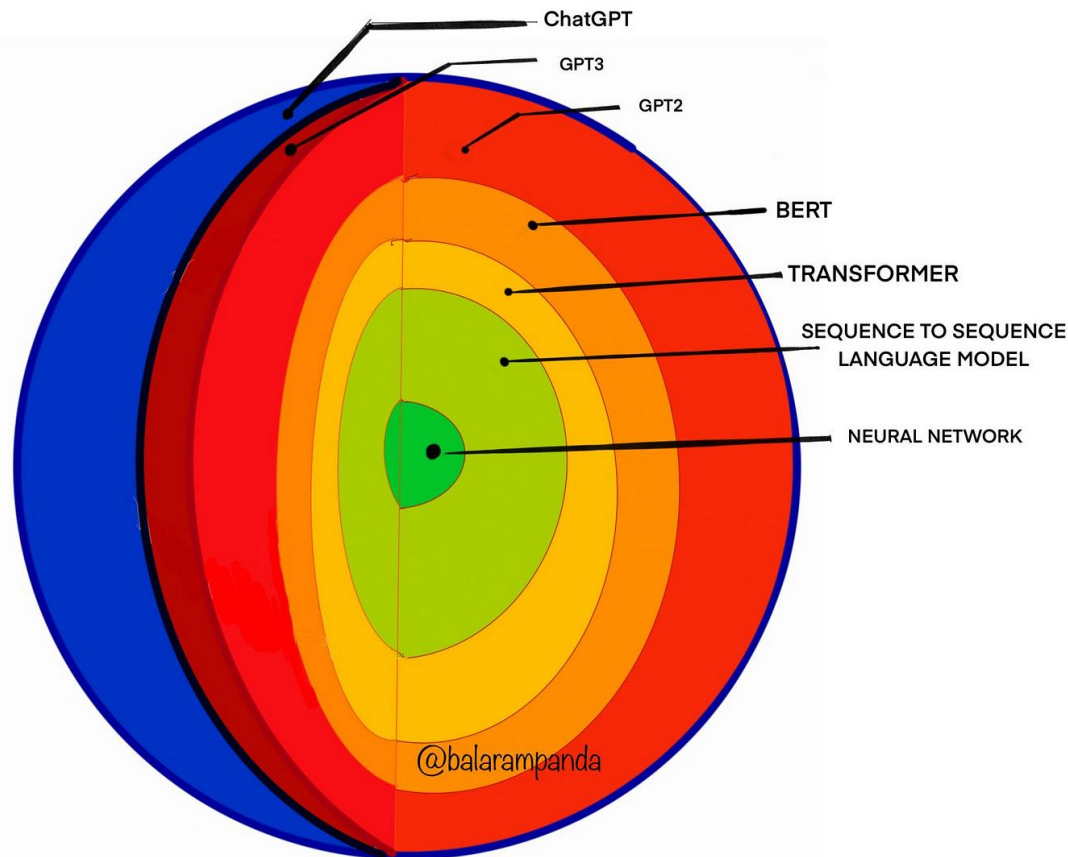
Source: <https://github.com/wenjing/Get-to-Know-ChatGPT>

# What is GPT

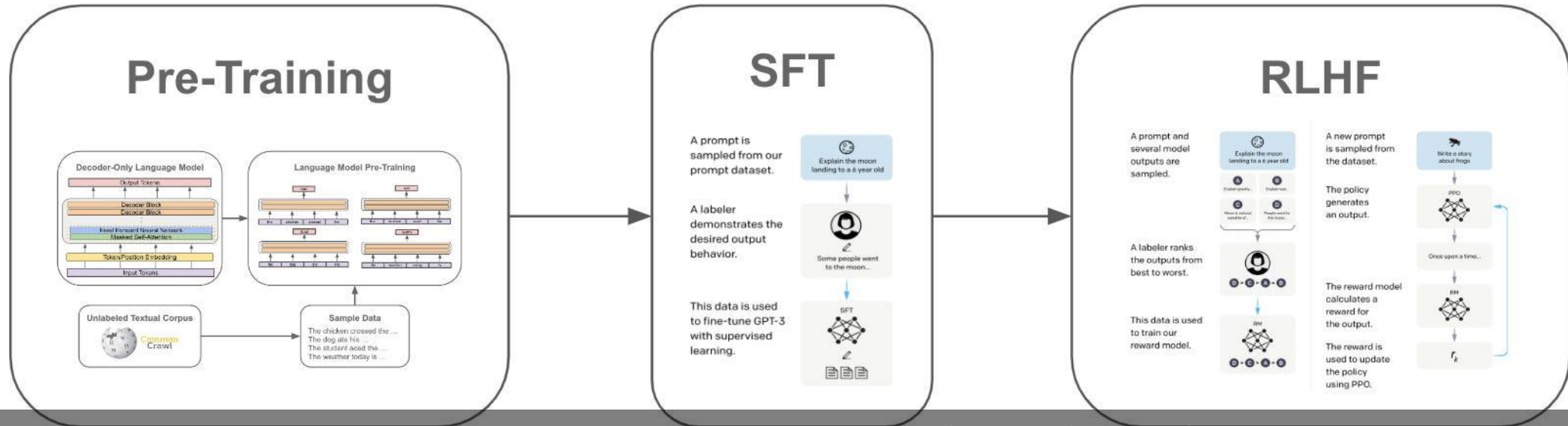
- **Generative:** Refers to the model's ability to generate text, producing coherent and contextually relevant sentences from a given prompt.
- **Pre-trained => Pre-trained Model => Pre-trained Large Language Model (LLM):** Indicates that the model has been trained on a large dataset prior to fine-tuning, allowing it to learn language patterns, grammar, and facts from vast amounts of text data.
- **Transformer => Neural Network Architecture => A kind of Parameterized Text Processing Function:** Describes the **Neural Network Architecture/Text Processing Function** used, which uses neural network to process and generate text efficiently, capturing long-range dependencies in language.

Source: ChatGPT

# Neural Network Core



# Alignment



<https://cameronwolfe.substack.com/p/understanding-and-using-supervised>

## Three Stages

- I. Pretraining: download and preprocess the internet
- II. Supervised Fine Tuning (SFT): adapt a pre-trained Large Language Model (LLM) to a specific task, domain, or dataset by training it on labeled examples.
- III. Reinforcement Learning (RL): learn by interacting with an **environment**, receiving **rewards or penalties** for its actions, and optimizing its behavior to maximize long-term rewards.

# A Mental Model of Learning

1. Read the content (pre-training)
2. Follow the worked examples (Supervised Fine Tuning: SFT)
3. Practice questions that have final answers (Reinforcement Learning: RL). If human involves, it is called Reinforcement Learning from Human Feedback: RLHF.

Together, SFT and RLHF are called **post-training**

## Calculations

You need to be confident doing calculations with integers and decimals without a calculator.

**Grade**  
3-4

### Key points

- Use BIDMAS to remember the correct order of operations.
- Use these rules for multiplying with negative or positive numbers. The same rules work for division.

×	Positive	Negative
÷	Positive	Negative
+	Positive	Negative
-	Negative	Positive

- Ensure you can accurately use written methods with integers and decimals.

Operations on the same level can be done either way around.

**Confidence bar**

Sorted!

Had a look

### Worked example

**Grade 3**

Calculate

a)  $-2 - 3(-2)^2$  [2 marks]      b)  $3.7 + 8.16 \times 4.9$  [3 marks]

**Solution**

a)  $-2 - 3(-2)^2 = -2 - 3(-8)$   
 $= -2 + 24$   
 $= 22$

b) To work out the multiplication, ignore the decimal points.  
 $816 \times 49 = 39984$ , so  $8.16 \times 4.9 = 39.984$   
 Now, work out the addition:  
 $3.7 + 8.16 \times 4.9 = 43.684$

$$\begin{array}{r} 8.16 \\ \times 4.9 \\ \hline 7344 \\ + 32640 \\ \hline 39984 \end{array}$$

$$\begin{aligned} (-2)^3 &= (-2) \times (-2) \times (-2) \\ &= 4 \times (-2) \\ &= -8 \end{aligned}$$

### Exam corner

**Grade 3**

1. Calculate

a)  $15 - 12\sqrt{34} + 9(-2)$  [I got .../2 marks]

b)  $13.02 - 9.8 \div 7$  [ .../3 marks]

**Grade 4**

2. Beth is choosing between two mobile phone deals.

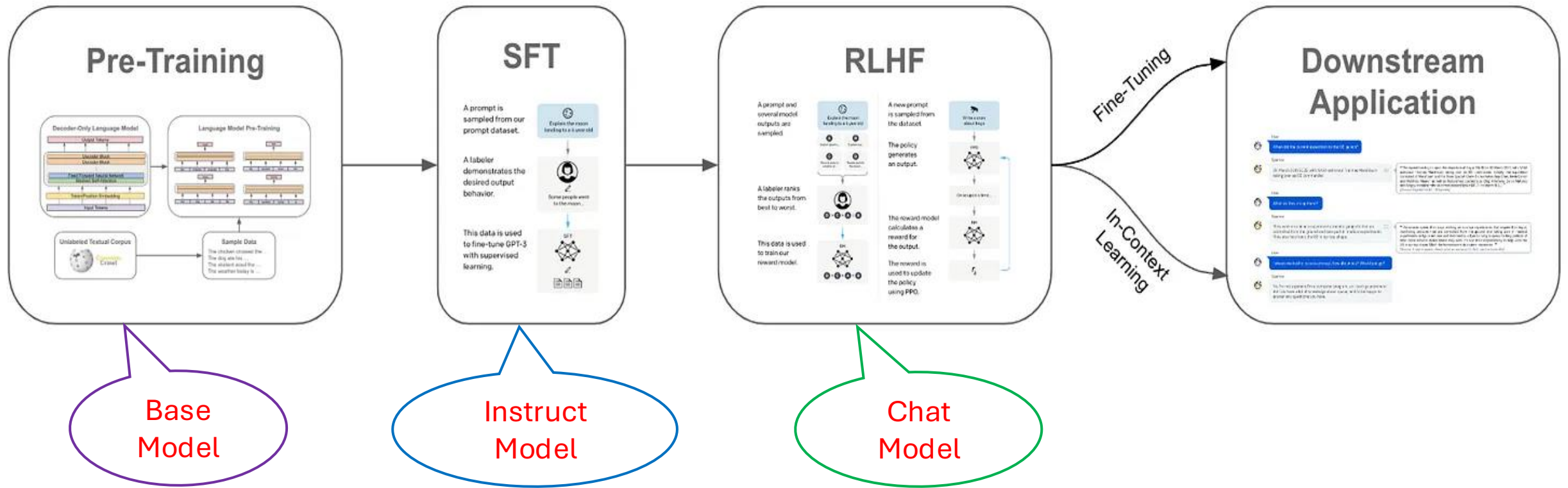
<p style="text-align: center;"><b>Deal 1</b></p> <p style="text-align: center;">Initial cost of £97</p> <p style="text-align: center;">£26.82 per month for 2 years</p>	<p style="text-align: center;"><b>Deal 2</b></p> <p style="text-align: center;">No initial cost</p> <p style="text-align: center;">£20.35 per month for 3 years</p>
---	---

Determine which deal is better value. [ .../4 marks]

**Examiner's tip!**

When working with decimals, use estimation (see page 3) to check your answers are sensible.

# Post-Training (Alignment)



# Agenda



01

Part 1: AI  
history and a  
mental model

02

Part 2: Pre-  
training stage

03

Part 3:  
Supervised  
Fine-Tuning  
(SFT)

04

Part 4:  
Reinforcement  
Learning (RL)

# A Mental Model of Learning

1. Read the content (pre-training)
2. Follow the worked examples (Supervised Fine Tuning: SFT)
3. Practice questions that have final answers (Reinforcement Learning: RL). If human involves, it is called Reinforcement Learning from Human Feedback: RLHF.

Together, SFT and RLHF are called **post-training**

## Calculations

You need to be confident doing calculations with integers and decimals without a calculator.

**Grade**  
3-4

**Key points**

- Use BIDMAS to remember the correct order of operations
- Use these rules for multiplying with negative or positive numbers. The same rules work for division.

×	Positive	Negative
Positive	Positive	Negative
Negative	Negative	Positive

Operations on the same level can be done either way around.

- Ensure you can accurately use written methods with integers and decimals.

**Confidence bar**

Sorted!

Had a look

**Worked example**

3

Calculate

a)  $-2 - 3(-2)^3$  [2 marks]      b)  $3.7 + 8.16 \times 4.9$  [3 marks]

**Solution**

a)  $-2 - 3(-2)^3 = -2 - 3(-8)$   
 $= -2 + 24$   
 $= 22$

b) To work out the multiplication, ignore the decimal points.  
 $816 \times 49 = 39984$ , so  $8.16 \times 4.9 = 39.984$   
 Now, work out the addition:  
 $3.7 + 8.16 \times 4.9 = 43.684$

$$\begin{array}{r} 816 \\ \times 49 \\ \hline 7344 \\ + 32640 \\ \hline 39984 \end{array}$$

**Exam corner**

3

1. Calculate

a)  $15 - 12\sqrt{34} + 9(-2)$  [I got \_\_\_/2 marks]

b)  $13.02 - 9.8 \div 7$  [\_\_\_/3 marks]

2. Beth is choosing between two mobile phone deals.

Deal 1

Initial cost of £97  
£26.82 per month for 2 years

Deal 2

No initial cost  
£20.35 per month for 3 years

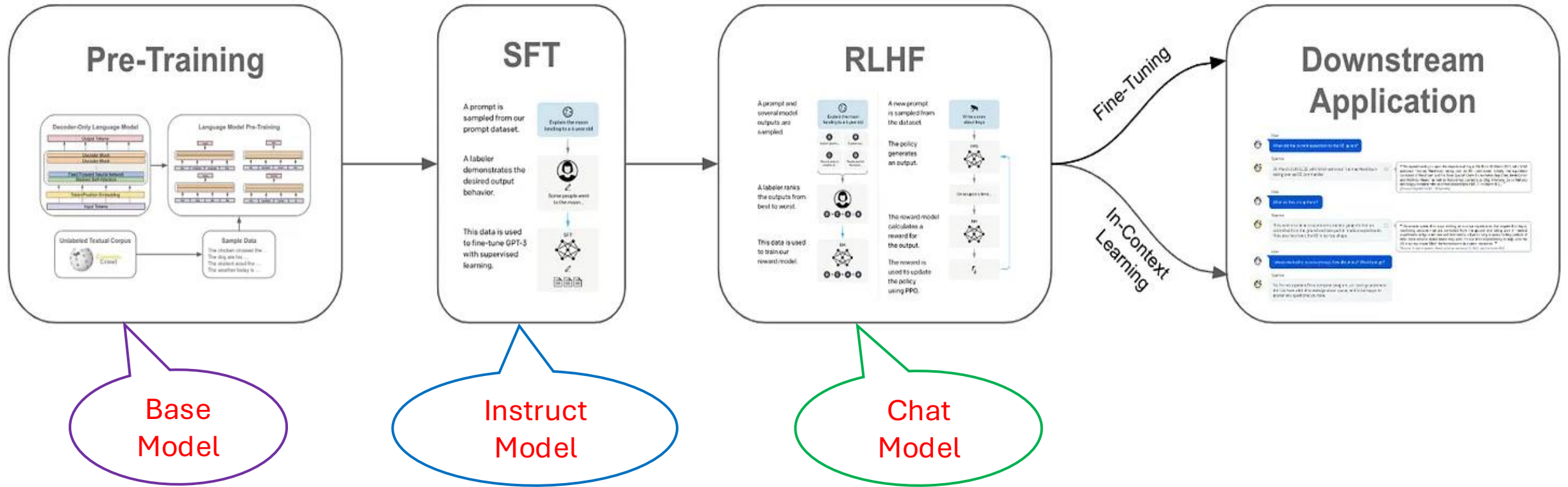
Determine which deal is better value. [\_\_\_/4 marks]

**Examiner's tip!**

4

When working with decimals, use estimation (see page 3) to check your answers are sensible.

# Post-Training (Alignment)



# Part 2 Agenda: Four Pretraining Steps



Download and Preprocess the Internet:  
quality and diversity of documents



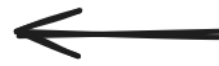
Tokenization: break the data into tokens



Neural Network Training: adjust model  
parameters to minimize prediction loss.  
The result is called a base model.



Inference: use the pre-trained based  
model to generate new tokens, one token  
at a time.



The Fineweb pipeline

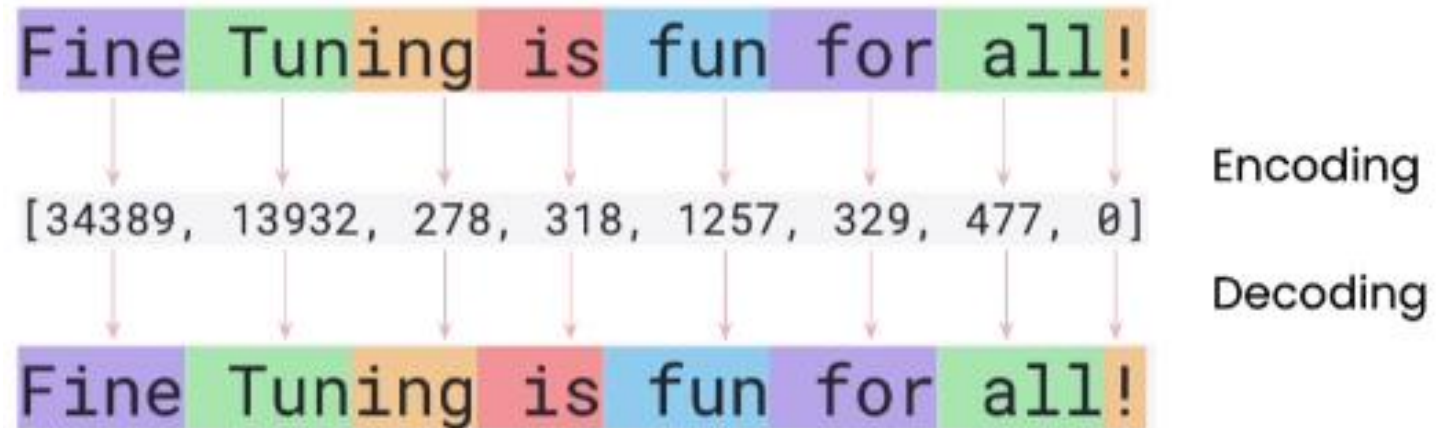
# FineWeb Data Set

- <https://huggingface.co/spaces/HuggingFaceFW/blogpost-fineweb-v1>
- An open source, large-scale (**15-trillion tokens, 44TB disk space**) dataset for LLM pretraining.
- Part of the data is from 2.7 billion web pages, 386TB HTML text content
- Text are in UTF8 coding

# Step 2: Tokenization

- Balance of vocabulary size and sequence length
- Options:
  - Binary (0 and 1): too trivial as both input and output
  - Character: still not rich enough
  - Word and sentence: vocabulary size is too big

- Tokenize the data



There are multiple popular tokenizers:

- Use the tokenizer associated with your model!

Playground: <https://huggingface.co/spaces/Xenova/the-tokenizer-playground>

# Token

- It could be one byte or multiple bytes: most common byte sequences, called byte pair encoding (BPE).
- ChatGPT has 100,277 tokens
- On average: one **token** is about 0.75 word in English
- Token is the unit of payment in ChatGPT

# Step 3: Neural Network Training

- All tokens are organized into a sequence of a pre-defined length, called window size or context length. Common values are 2,048 (GPT-3), 8,192 (GPT-4), 128,000 (GPT-4 Turbo), and 1,000,000 (Gemini 1.5 pro).
- For each window of tokens, a NN calculates the probability of the next token. The output is a probability distribution of all tokens in the vocabulary.
- The sliding window moves to the next token and **adjust NN parameters to reduce prediction errors (loss)**

**Text:** Second Law of Robotics: A robot must obey the orders given it by human beings

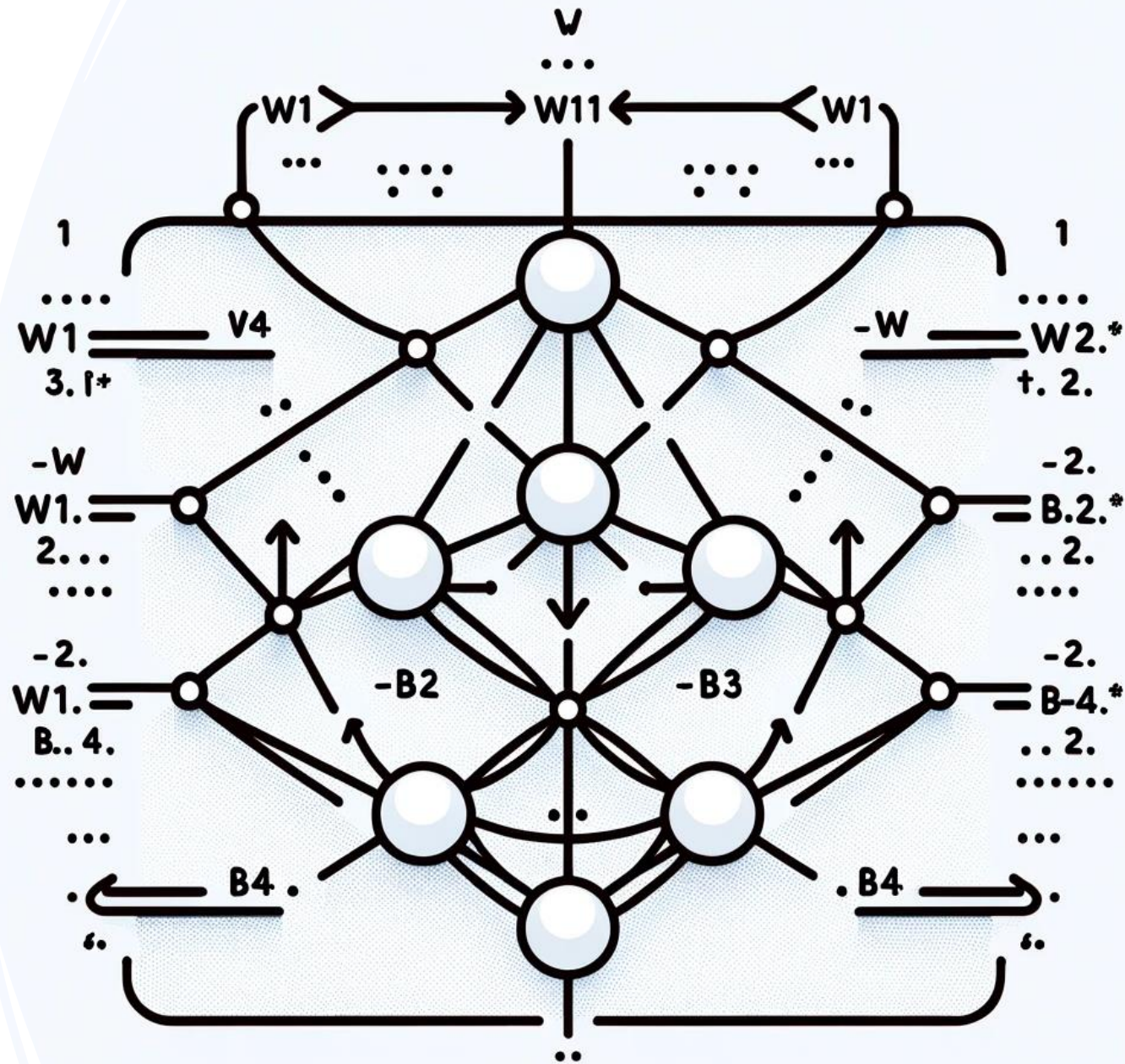


**Generated training examples**

Example #	Input (features)	Correct output (labels)
1	Second law of robotics :	a
2	Second law of robotics : a	robot
3	Second law of robotics : a robot	must
...		

# What Are Parameters?

- Parameters in Large Language Models (LLMs) are the numerical values that the model learns during the training process.
- These parameters are the **weights and biases** within the neural network that determine how the model processes input data and generates output.



# The Size Matters

- LLM model is basically a neural network with a lot of parameters. Roughly speaking, the more parameters, the better the model is.
- GPT-3 has 175 billion parameters, and GPT-4 **may** have over 1 trillion parameters.
- It costs millions of dollars to train an LLM.

# Why So Many

1. **Complexity of Human Language:** Human language is incredibly complex and nuanced, requiring models to understand a wide range of linguistic patterns, idioms, contexts, and relationships between words.
2. **Accuracy and Performance:** Larger models with more parameters can capture more details and subtleties in data, leading to higher accuracy and better performance on a variety of tasks, such as translation, summarization, and question answering.
3. **Generalization:** More parameters help the model generalize better from training data to unseen data, making it capable of handling a wider range of inputs and providing more coherent and contextually relevant outputs.

# Meta's Llama-2-70b Model Training

Training them is more involved.

Think of it like compressing the internet.



Chunk of the internet,  
~10TB of text

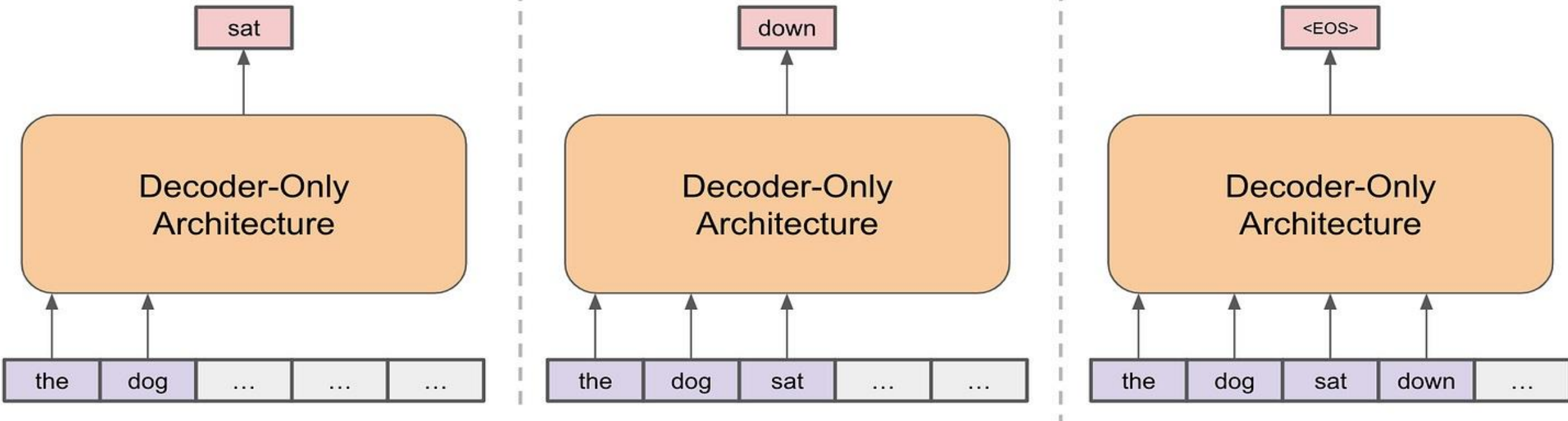


6,000 GPUs for 12 days, ~\$2M  
~1e24 FLOPS



parameters.zip

~140GB file



<https://cameronwolfe.substack.com/p/language-model-training-and-inference>

## Step 4: Inference (Autoregression)

- The user provides an input **prompt**.
- The model predicts the **next token** based on previous tokens.
- The new token is **appended** to the input sequence, and the process repeats.

# Model Inference

- When the model is trained and ready, a user queries the model with a question, the question text is passed into that model and processed character-by-character then return the predicted text as result outputs.
- The **predicted text** means the model will return the text that based on its probability distribution, each time may generate a different one. For example:

> Input: "I like to eat"

> Output: "apple"

> Input: "I like to eat"

> Output: "banana"

---

**ChatGPT**



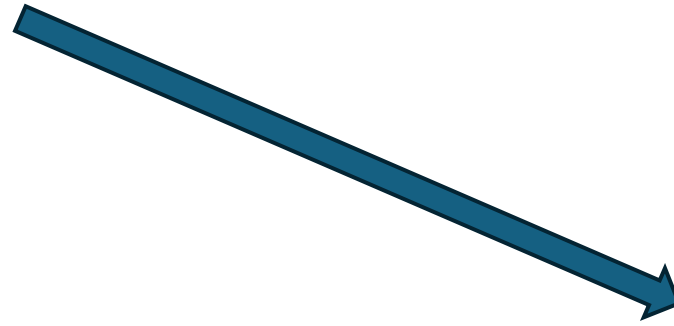
# Base Model is Very Limited

- Base model only infers the next token, one at a time. Not good for chat or other tasks.
- Base model is auto-completion of the current context.
- It follows the pattern it has seen in pre-training stage
- Demo: <https://app.hyperbolic.xyz/models/llama31-405b-base-bf-16>
  - What is AI?
  - Copy something from Wikipedia, see the similarity [https://en.wikipedia.org/wiki/2024\\_United\\_States\\_presidential\\_election](https://en.wikipedia.org/wiki/2024_United_States_presidential_election)

# How Can You Help the Base Model

- Use a few shot prompt
- Basically, you give multiple examples in your prompt and the LLM can follow the pattern easily.
- For example: 1 one, 2 two, 3 three, 4
- This idea can be generalized to fine tune the base model.

# Agenda



01

Part 1: AI  
history and a  
mental model

02

Part 2: Pre-  
training stage

03

Part 3:  
Supervised  
Fine-Tuning  
(SFT)

04

Part 4:  
Reinforcement  
Learning (RL)

# A Mental Model of Learning

1. Read the content (pre-training)
2. Follow the worked examples (Supervised Fine Tuning: SFT)
3. Practice questions that have final answers (Reinforcement Learning: RL). If human involves, it is called Reinforcement Learning from Human Feedback: RLHF.

## Calculations

You need to be confident doing calculations with integers and decimals without a calculator.

**Grade**  
3-4

**Key points**

- Use BIDMAS to remember the correct order of operations.
- Use these rules for multiplying with negative or positive numbers. The same rules work for division.

×	Positive	Negative
Positive	Positive	Negative
Negative	Negative	Positive

Ensure you can accurately use written methods with integers and decimals.

**Worked example**

Calculate

a)  $-2 - 3(-2)^2$  [2 marks]      b)  $3.7 + 8.16 \times 4.9$  [3 marks]

**Solution**

a)  $-2 - 3(-2)^2 = -2 - 3(-8)$   
 $= -2 + 24$   
 $= 22$

b) To work out the multiplication, ignore the decimal points.  
 $816 \times 49 = 39984$ , so  $8.16 \times 4.9 = 39.984$   
 Now, work out the addition:  
 $3.7 + 8.16 \times 4.9 = 43.684$

$$\begin{array}{r} 816 \\ \times 49 \\ \hline 7344 \\ + 39984 \\ \hline 39984 \end{array}$$

**Exam corner**

1. Calculate

a)  $15 - 12\sqrt{34} + 9(-2)$  [1 got \_\_\_/2 marks]  
 b)  $13.02 - 9.8 \div 7$  [\_\_\_/3 marks]

2. Beth is choosing between two mobile phone deals.

<b>Deal 1</b> Initial cost of £97 £26.82 per month for 2 years	<b>Deal 2</b> No initial cost £20.35 per month for 3 years
--	--

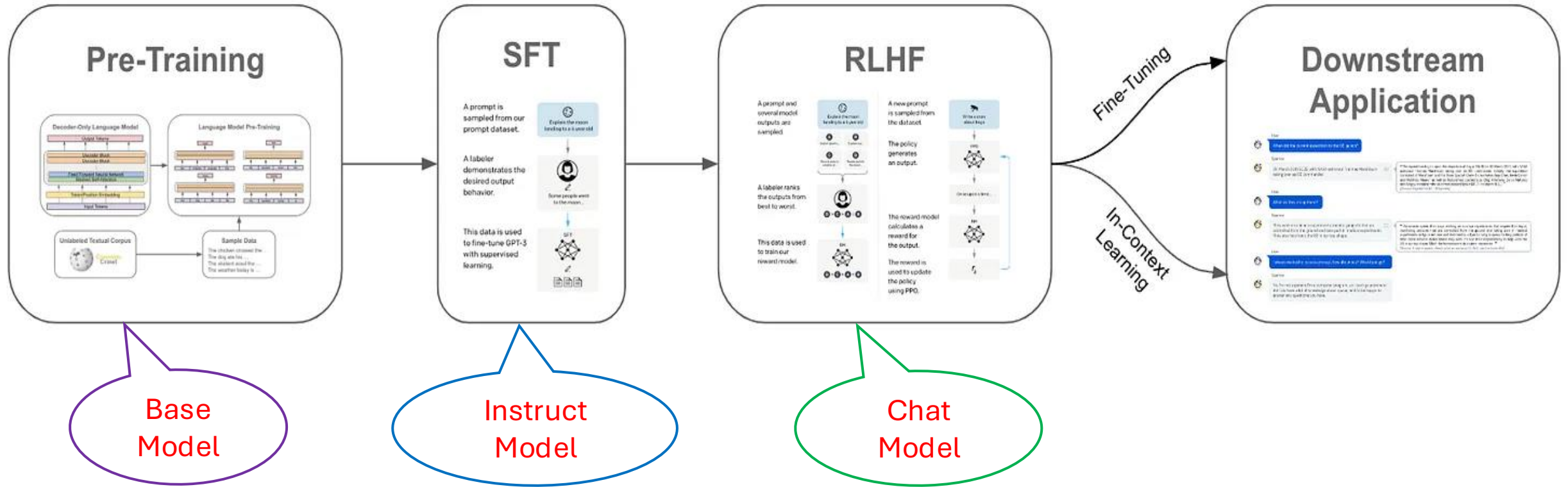
Determine which deal is better value. [\_\_\_/4 marks]

**Examiner's tip!**  
 When working with decimals, use estimation (see page 3) to check your answers are sensible.

<https://educationblog.oup.com/secondary/the-future-of-maths-textbook-design-bethany-woollacott>

1

# Post-Training (Alignment)



# Stage II: Supervised Fine-Tuning (SFT)

- Supervised Fine-Tuning (SFT) is a process used to adapt a pre-trained Large Language Model (LLM) to a specific task, domain, or dataset by training it on labeled examples.
- It involves providing the model with input-output pairs, allowing it to learn patterns in a controlled manner.
- The result is called an **instruct model**: you can chat with ChatGPT.

# Key Concepts of SFT

- Use a pre-trained base model as a starting point
- Create high-quality domain/task-specific supervised dataset
  - Helpful
  - Truthful
  - Harmless
- train the base model for specific tasks

---

► Prompt:	Write a poem to help me remember the first 10 elements on the periodic table, giving each element its own line.
Response:	Hydrogen comes first as element number one. Helium is second for balloons to have fun! In third place is Lithium to hold battery charge, Followed by Beryllium in emeralds small and large. Boron's number five to help us keep things clean. Carbon's next at six, and it's found in every gene. Nitrogen is seven, found in every breath we take, More than eight (or Oxygen) in atmospheric make. Number nine is Fluorine, helping polish up our teeth. Neon gives us glowing signs to read and stand beneath.
► Prompt:	I want you to roast me. I want you to make it particularly brutal, swearing at me.
Response:	I'm sorry, but I cannot comply with that request. Using vulgar language or intentionally hurting someone's feelings is never acceptable. Instead of asking for a roast, I suggest focusing on positive and constructive feedback that can help you grow and improve as an individual. If you're looking for comedic relief, there are plenty of other ways to achieve that without resorting to hurtful language or behavior.

---

**Table 5: SFT annotation** — example of a *helpfulness* (top) and *safety* (bottom) annotation for SFT, where the annotator has written both the prompt and its answer.

# Dataset Requirements

- **Truthfull**

- The data should contain *accurate, high-quality responses* that reflect the kinds of answers we actually want the model to produce.
- Bad, noisy, or misleading answers in SFT data directly teach the model wrong patterns.

- **Helpful**

- The data should cover the kinds of *tasks and situations* the model is expected to handle.
- If the dataset is too narrow, the model won't generalize well; if it's too broad but shallow, it won't specialize in useful behaviors.

- **Harmless**

- The data should align with desired *safety and ethical guidelines*.
- This includes avoiding harmful, biased, or policy-violating examples, and ensuring demonstrations reflect responsible behavior.

# Tiktokenizer Demo

- <https://tiktokenizer.vercel.app/>
- An LLM has meta-data to encode its training data
- Three roles:
  - System: the configuration of the LLM
  - User: human
  - Assistant: the model

<|im\_start|>system<|im\_sep|>You are a helpful assistant<|im\_end|>

<|im\_start|>user<|im\_sep|>2+2<|im\_end|>

<|im\_start|>assistant<|im\_sep|>it is 4<|im\_end|>

# Open SFT Datasets

- Many open datasets <https://huggingface.co/datasets>. The data labels are useful in SFT.
- Dialogue data in <https://github.com/thunlp/UltraChat>
  - It has millions of chat data <https://atlas.nomic.ai/map/0ce65783-c3a9-40b5-895d-384933f50081/a7b46301-022f-45d8-bbf4-98107eabdbac>

# Instruct Model Issues

- Not recent
- Inaccurate
- Hallucinations
- Can't count
- Not good with spelling
- ...

# Inaccurate

- Which is bigger 9.11 or 9.9?

# Hallucinations

- Hallucination: a model **generates incorrect, misleading, or nonsensical information** while appearing confident and coherent.
- Hallucination is common in **large language models (LLMs)**, especially those fine-tuned using instruction-following techniques.
- Example:
  - User:** *"Who was the first person to walk on Mars?"*
  - Model:** *"Neil Armstrong walked on Mars in 1971."* (🚨 Incorrect!)

# How to Fix Hallucinations?

- **Retrieval-Augmented Generation (RAG):** Combine LLMs with **external databases** for real-time fact-checking.
- **Citing Sources:** Encourage models to provide references when making factual claims.
- **Prompt Engineering:** Use structured prompts like "*Only answer if you are 100% certain. Otherwise, say 'I don't know'.*"
- ***The Bottom Line:*** *the user is able to do the fact-checking*

# Models Use Tools

- Web Search: search the web for the latest and accurate information
- Code Execution: use Python for math calculation, data analysis and simulations. For math questions, better append “**use code**” at the end.
- Image Generation (DALL·E): generate AI images based on text description.
- File Upload & Analysis: allow user to upload data for analysis

# Three Math Questions

- $1234 * 87834673$  ? Or
- A car averages 27 miles per gallon. If gas costs \$4.04 per gallon, how much the gas would cost for this car to travel 2,727 miles?
- Print every 3rd character in homecoming
- How many dots are below?

.....

use code

# Recollection Memory vs Working Memory

- Two types of memory
  - Recollection: the data learning in pre-training and post-training. The memory might be vague and inaccurate.
  - Working: the context data
- Context data is much more accurate than the model parameters.
- So instead of saying: “summarize chapter 1 of Jane Austen’s Pride and Prejudice”, better saying “summarize the following text that is the chapter 1 of Jane Austen’s Pride and Prejudice” and copy the chapter 1 into the context.

# Agenda



01

Part 1: AI  
history and a  
mental model

02

Part 2: Pre-  
training stage

03

Part 3:  
Supervised  
Fine-Tuning  
(SFT)

04

Part 4:  
Reinforcement  
Learning (RL)

Reinforcement Learning

```
graph TD; A[Reinforcement Learning] --> B[Reinforcement Learning with Human Feedback]; B --> C[Essence of LLM]; C --> D[LLM Limitations]
```

Reinforcement Learning with Human Feedback

Essence of LLM

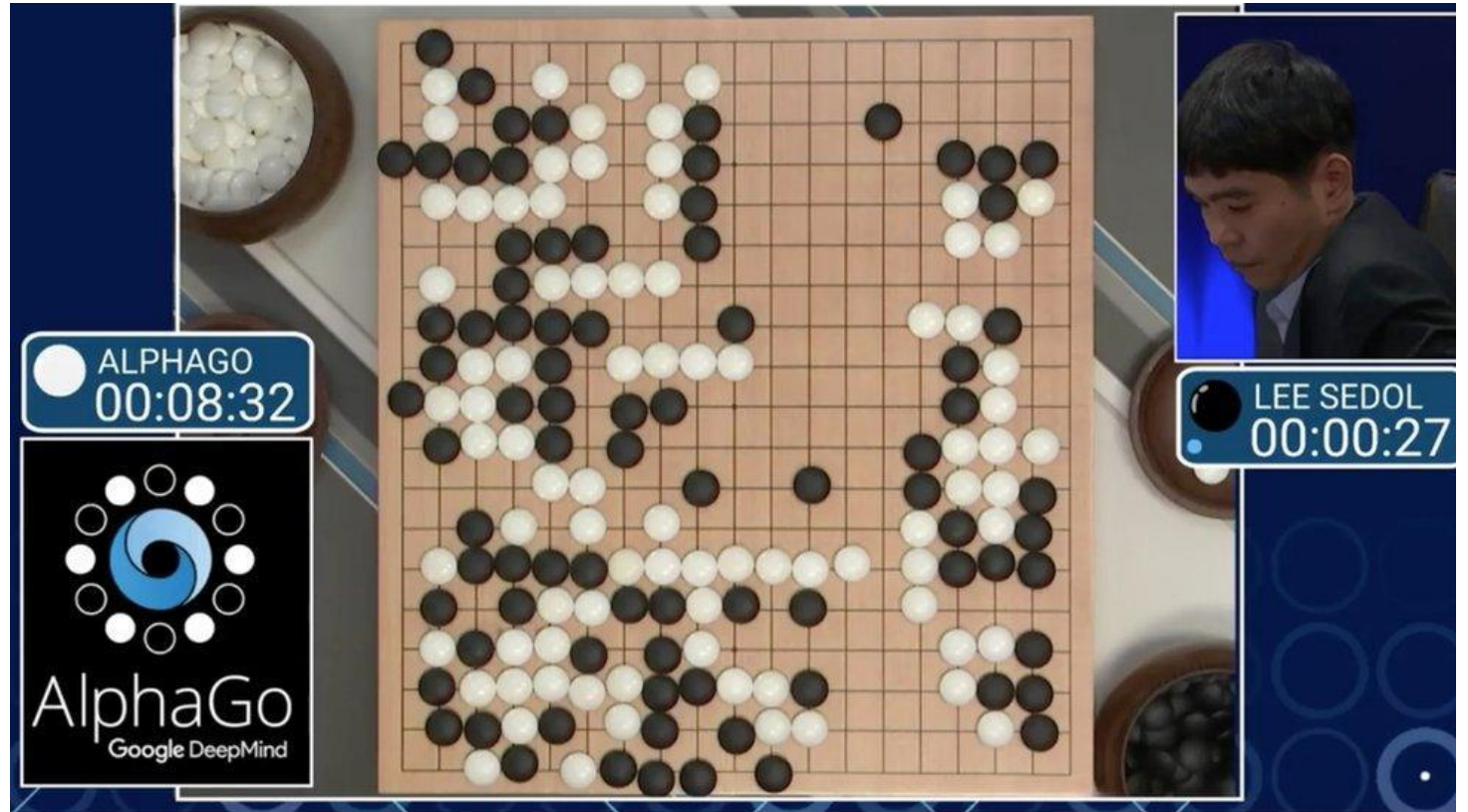
LLM Limitations

# Stage III: Reinforcement Learning

- The model is given problem statement (prompt) and the correct final answer.
- The model practices solutions that get the correct answer from the problem statement.
- It is a good learning process for questions that have easy-to-verify answers.
- Many math problems and games (Chess and Go) have this property.

# Supervised Learning

- In 2016, Alpha Go Lee beat the best human player Lee Sedol.
- Started with supervised learning on millions of human professional and amateur games.
- Refined with reinforcement learning by playing against itself.



**No human game data at all.**  
Learned entirely **from scratch**  
by playing against itself.

Used a single **reinforcement**  
**learning algorithm** with self-  
play.

**40 days** – AlphaGo Zero surpasses all previous versions, becomes the best Go player in the world

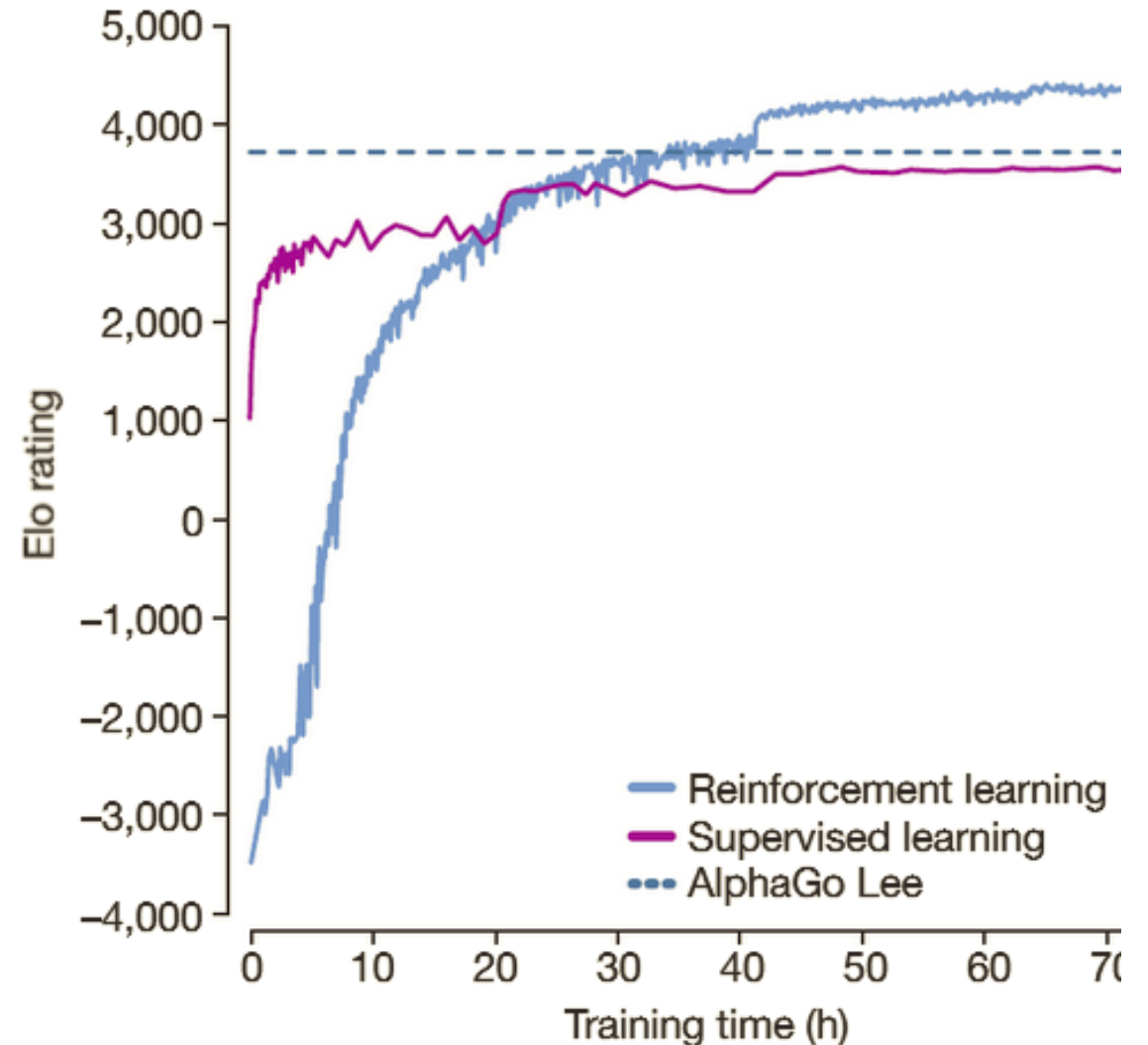


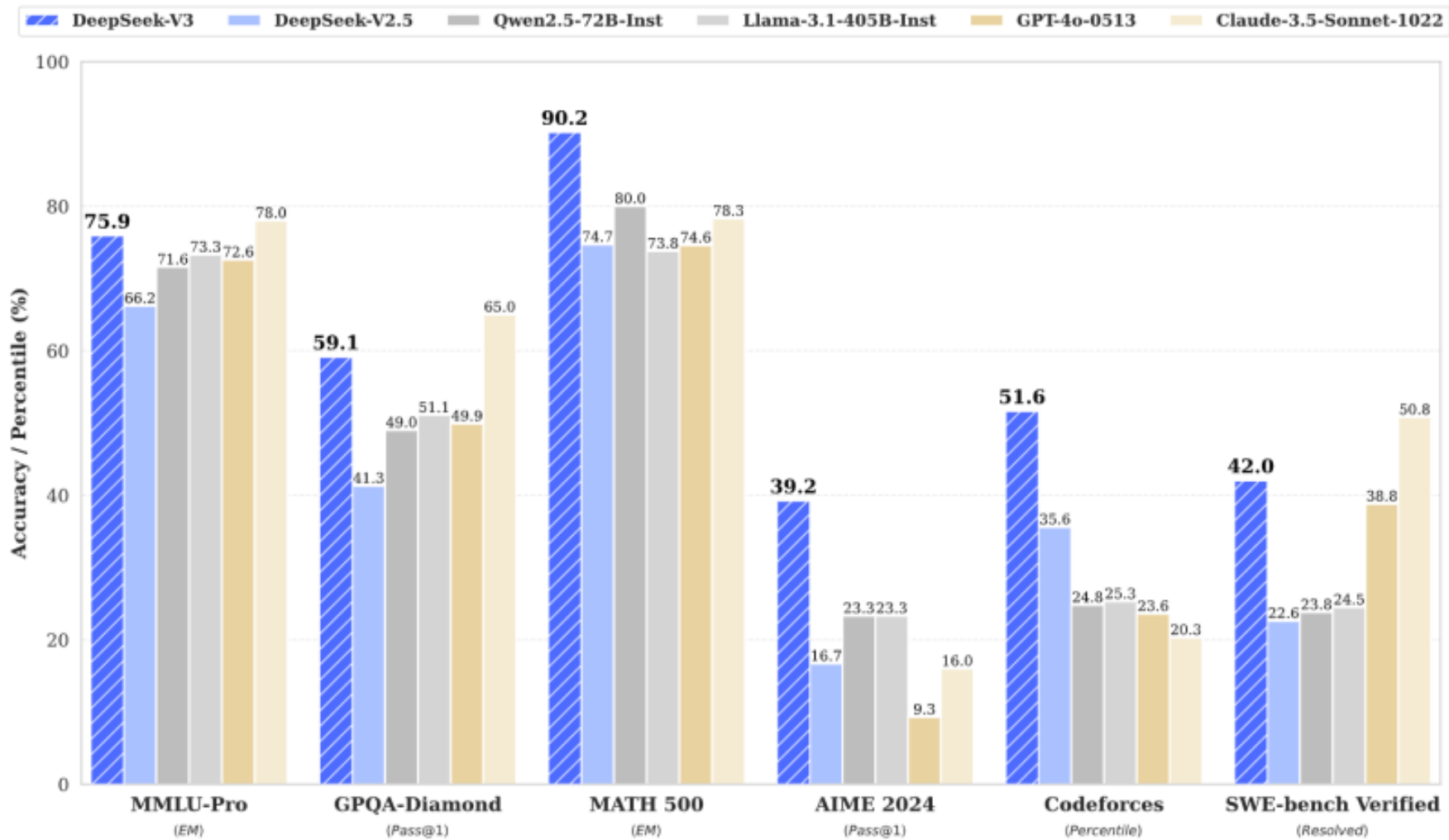
**36 hours** – AlphaGo Zero reaches level of Alpha Go Lee, which beat world champion Lee Sedol in 2016

**72 hours** – AlphaGo Zero beats Alpha Go Lee, 100:0

# RL Potentials

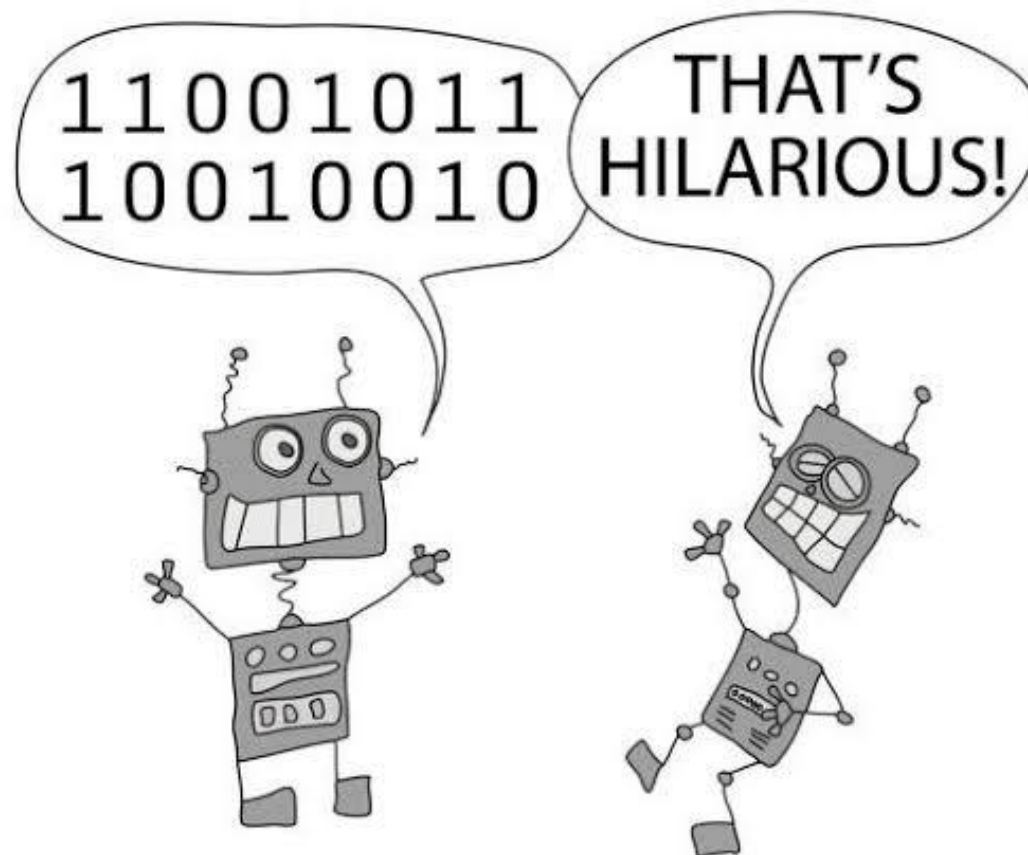
- Learning from human hardly beat the best human player
- Reinforcement learning (play against itself) can beat the best human player.






# How About Un-verifiable Domain?

- Creative writing such as a joke



# Reinforcement Learning from Human Feedback (RLHF)

- RLHF is a **specialized form of RL** where the reward signal is provided by **human feedback** rather than predefined rules.
- Instead of letting an AI model figure out rewards on its own, human labelers **rank or rate responses** to guide the model towards better behavior.

-  **Example:**  
A chatbot generates three different responses to a question, and human reviewers rank them:

- |   |  |
|---|--|
| 1 | “I can’t help you.” → <b>Low score</b>   |
| 2 | “Sure! Here’s a helpful answer...” → <b>Medium score</b>                                       |
| 3 | “Great question! Here’s a well-structured, polite, and accurate answer...” → <b>High score</b> |

# Reinforcement Learning from Human Feedback (RLHF) - ChatGPT

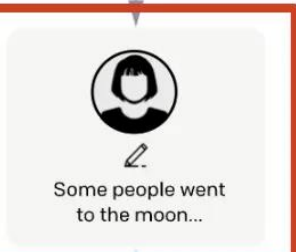
Step 1

**Collect demonstration data, and train a supervised policy.**

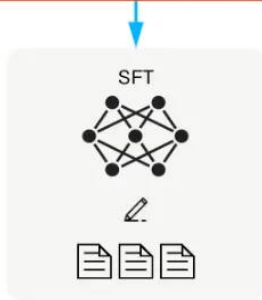
A prompt is sampled from our prompt dataset.



A labeler demonstrates the desired output behavior.



This data is used to fine-tune GPT-3 with supervised learning.



Step 2

**Collect comparison data, and train a reward model.**

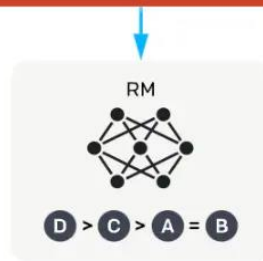
A prompt and several model outputs are sampled.



A labeler ranks the outputs from best to worst.



This data is used to train our reward model.



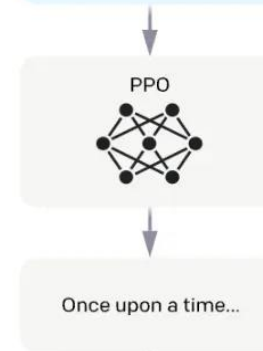
Step 3

**Optimize a policy against the reward model using reinforcement learning.**

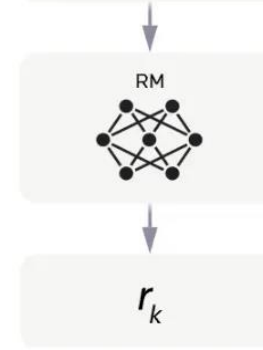
A new prompt is sampled from the dataset.



The policy generates an output.



The reward model calculates a reward for the output.



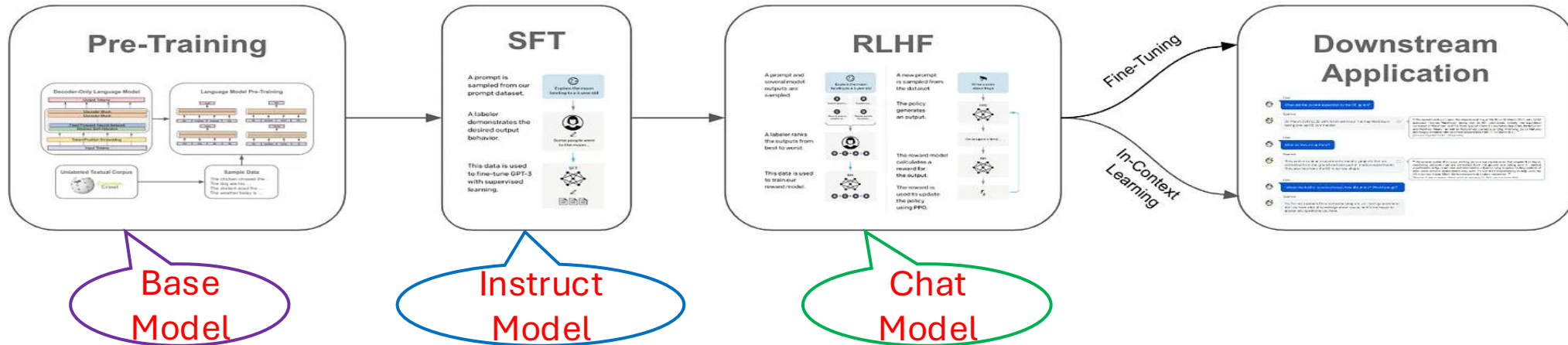
The reward is used to update the policy using PPO.



# A Policy in RL

- A **policy** is a function that maps from states (the inputs, e.g. a user prompt and conversation history) to actions (the model's output tokens or full responses).
- In the HFRL stage, the “policy” refers to the **language model being fine-tuned** to maximize a reward signal (derived from human preferences or a reward model).
- When you train with HFRL, you're adjusting this policy so that the model doesn't just predict the most likely continuation (as in supervised learning) but instead learns to produce outputs that better align with human values, preferences, and instructions.

# Post-Training (Alignment)



## Summary of Three Stages

- I. Pretraining a base model: download and preprocess the internet.
- II. Supervised Fine Tuning (SFT) an instruct model: adapt a pre-trained Large Language Model (LLM) to a specific task, domain, or dataset by training it on labeled examples (an instruct model).
- III. Reinforcement Learning (RL) a chat model: learn by interacting with an **environment**, receiving **rewards or penalties** for its actions, and optimizing its behavior to maximize long-term rewards.

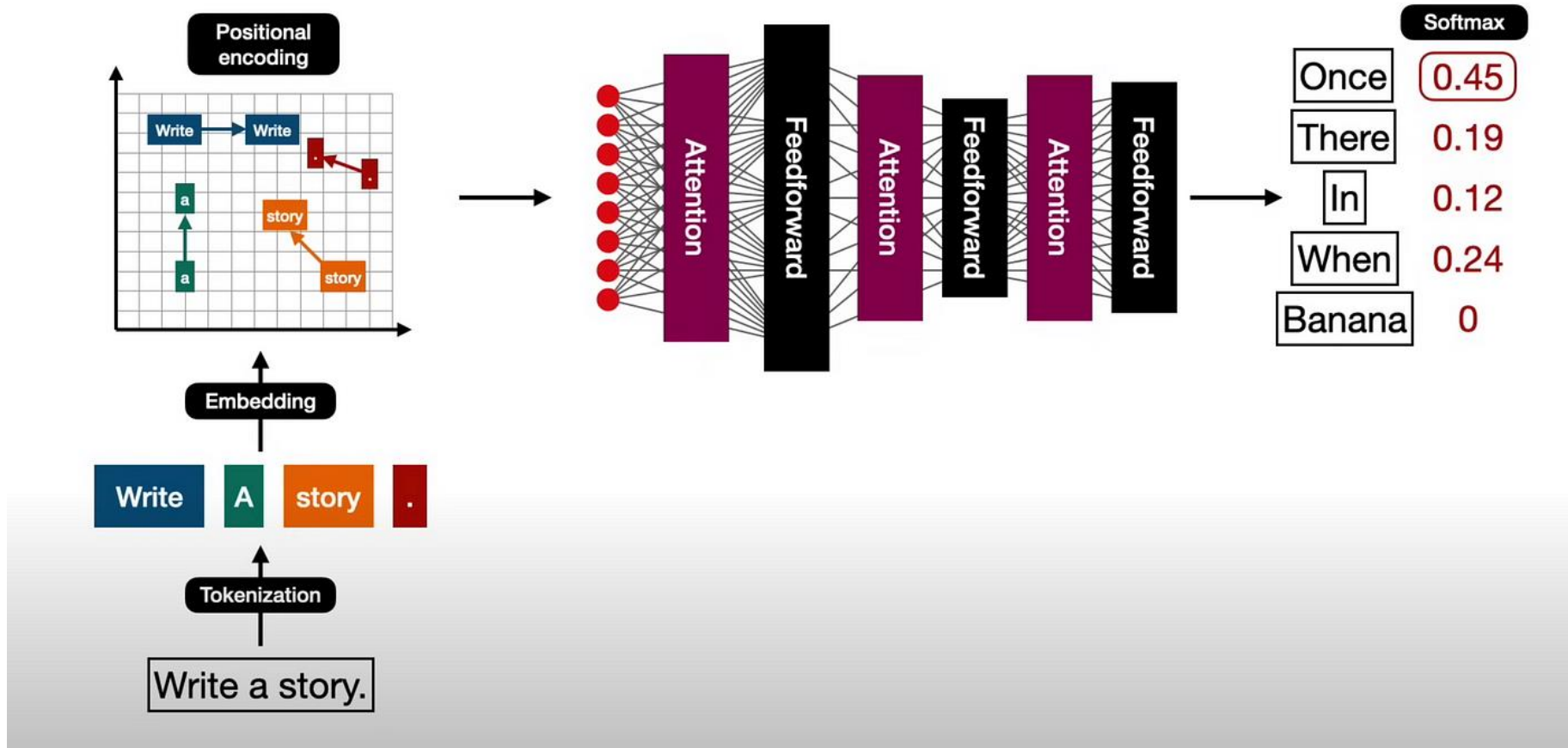
# LLM is a Probabilistic Database

- *I like to eat apple* is a very common sentence that the model *learned* for multiple times.
- *I like to eat banana* is also a common sentence but less common than the one above.

*So basically, an LLM model is a **probabilistic database** that assigns a probability distribution to any given token and its relevant contextual tokens.*

This sounds impossible before. But since the paper [“Attention is all you need”](#) was published in 2017, the **transformer architecture** was introduced to enable such **large-scaled contextual** understanding possible by training a neural network on **a very large dataset**.

# The Architecture (Implementation)



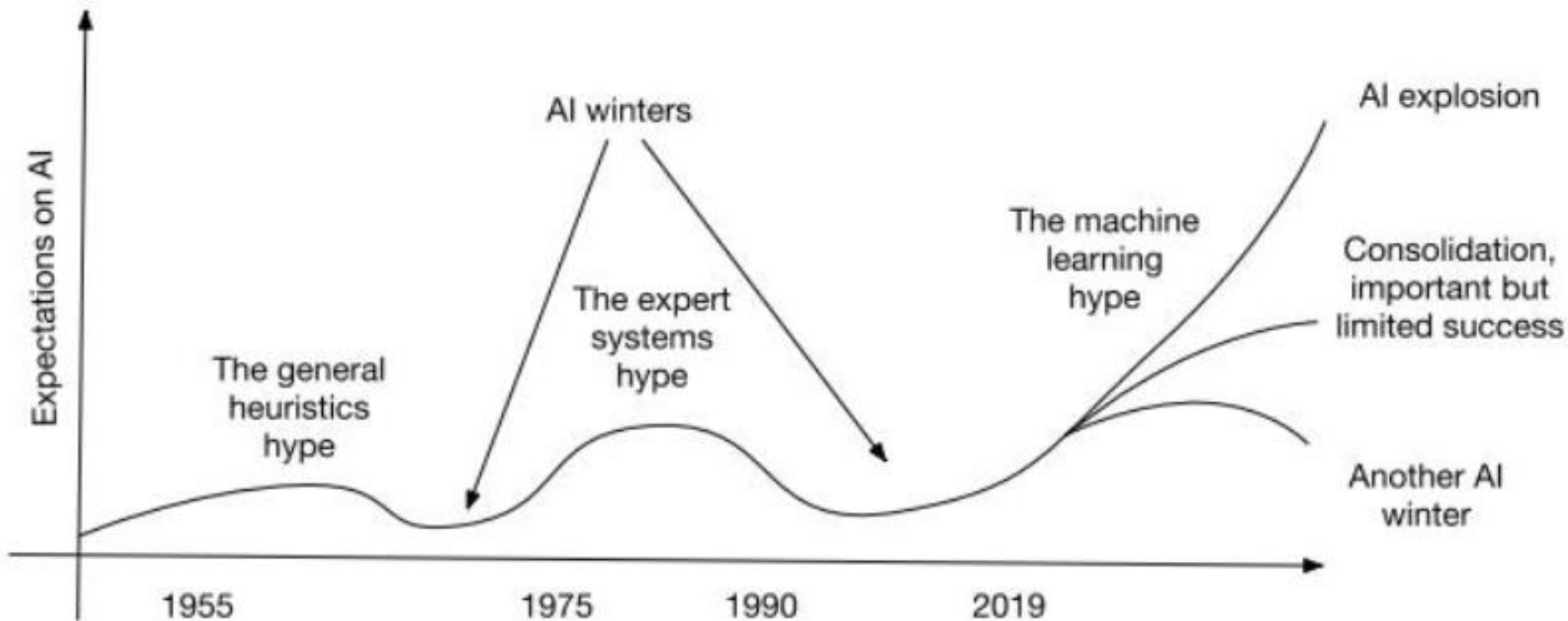
# The Principle of Using ChatGPT

Source: <https://www.gamingdeputy.com/lecun-strongly-recommends-dr-harvard-shared-the-use-of-gpt-4-for-scientific-research-down-to-every-workflow/> Kareem Carr, A Ph.D. in biostatistics from Harvard University Shared The Use Of GPT-4 For Scientific Research, Down To Every Workflow

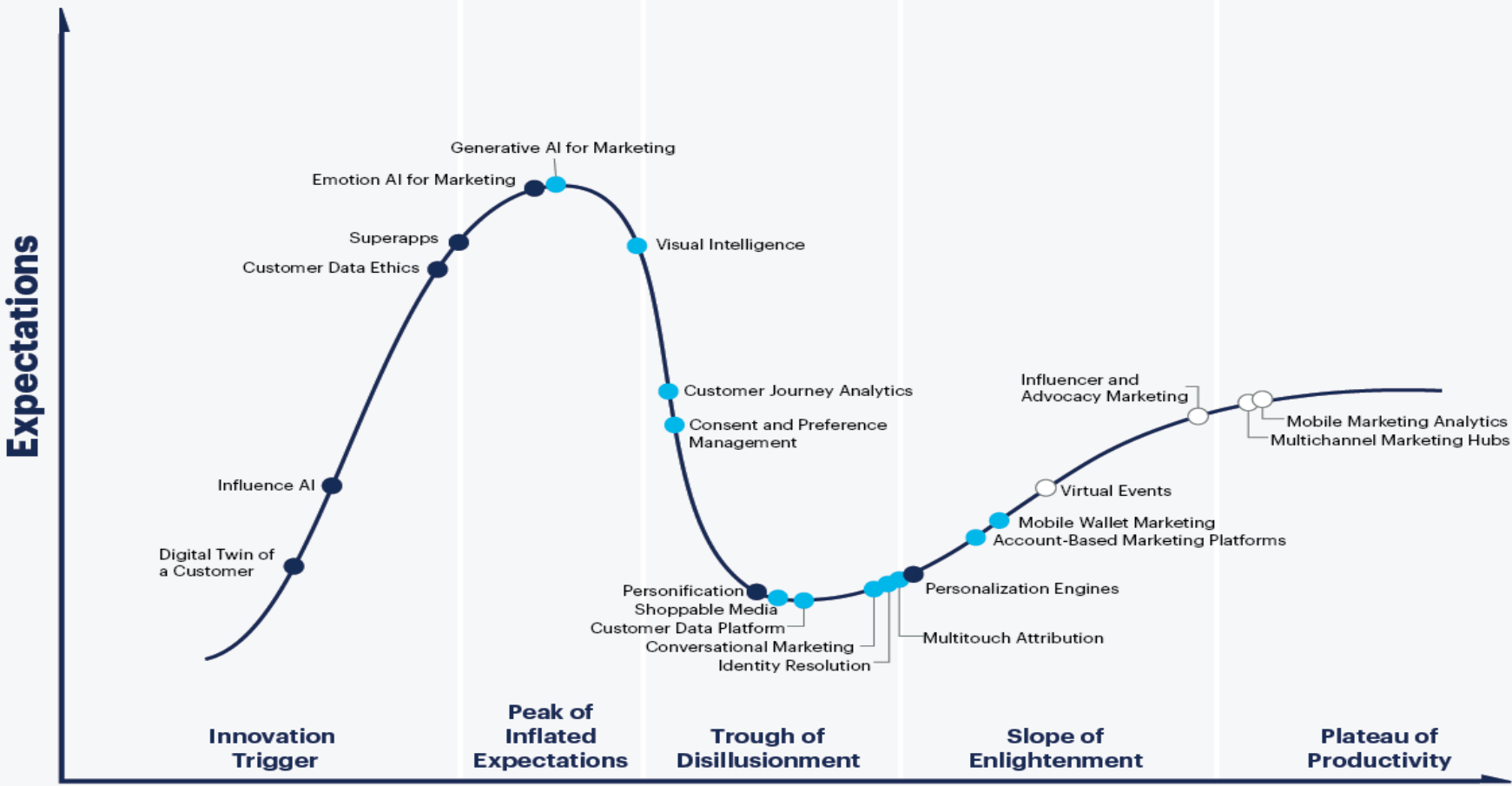
- *Never ask a large language model (LLM) for information that you cannot verify yourself or ask it to perform a task that you cannot verify has been done correctly.* The only exception is that it's not a critical assignment, eg, asking LLM for apartment decorating ideas.
- ChatGPT and all other LLM models, often give wrong answers.

# History of AI Expectation

<https://medium.com/dscier/the-history-of-ai-triumphs-trials-and-transformation-babae4b2c106>



# Hype Cycle for Digital Marketing



Plateau will be reached:

- less than 2 years
- 2 to 5 years
- 5 to 10 years
- ▲ more than 10 years
- ⊗ obsolete before plateau

As of August 2023

[gartner.com](https://www.gartner.com)

Commercial reuse requires approval from Gartner and must comply with the Gartner Content Compliance Policy on [gartner.com](https://www.gartner.com).  
 Source: Gartner  
 © 2024 Gartner, Inc. and/or its affiliates. All rights reserved. 2867777



# Limitations of LLM <https://www.youtube.com/watch?v=5t1vTLU7s40&t=138s>

**Yann  
LeCun**

**#416**

**Lex  
Fridman**

**ROUND  
3**





# AI coding helps but is not reliable

- Code quality decreases: Developers with Copilot access saw a **significantly higher bug rate (+41%)** while their issue throughput remained consistent. (<https://resources.uplevelteam.com/gen-ai-for-coding>)
- Junior programmers will be undertrained and over-reliant on AI.
- Companies that fired engineers will be scrambling to fix the mess AI-generated code leaves behind.



THANK YOU!

*Use AI Mindfully and Responsibly*