How LLMs are trained? A simple guide to understand LLM Training
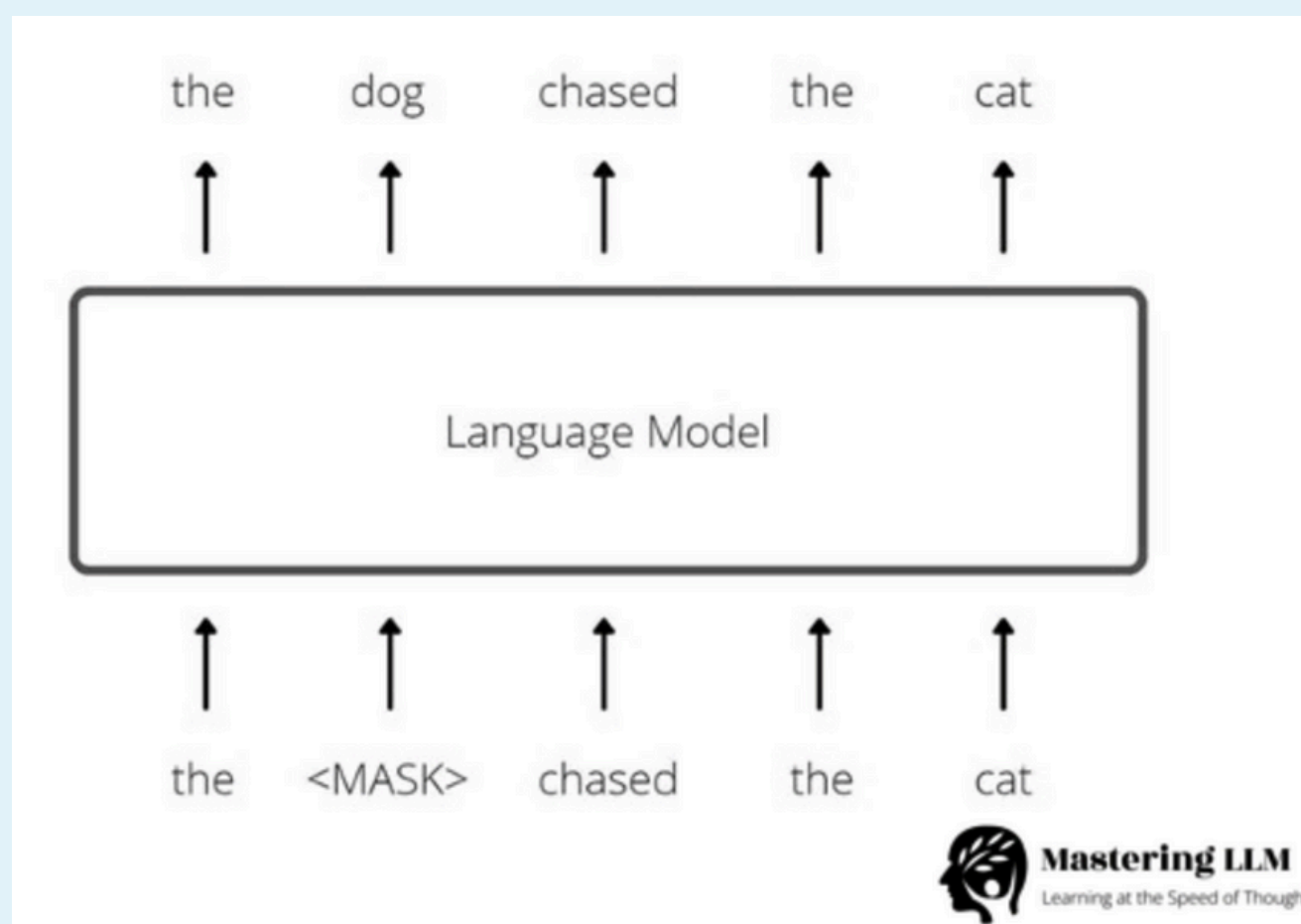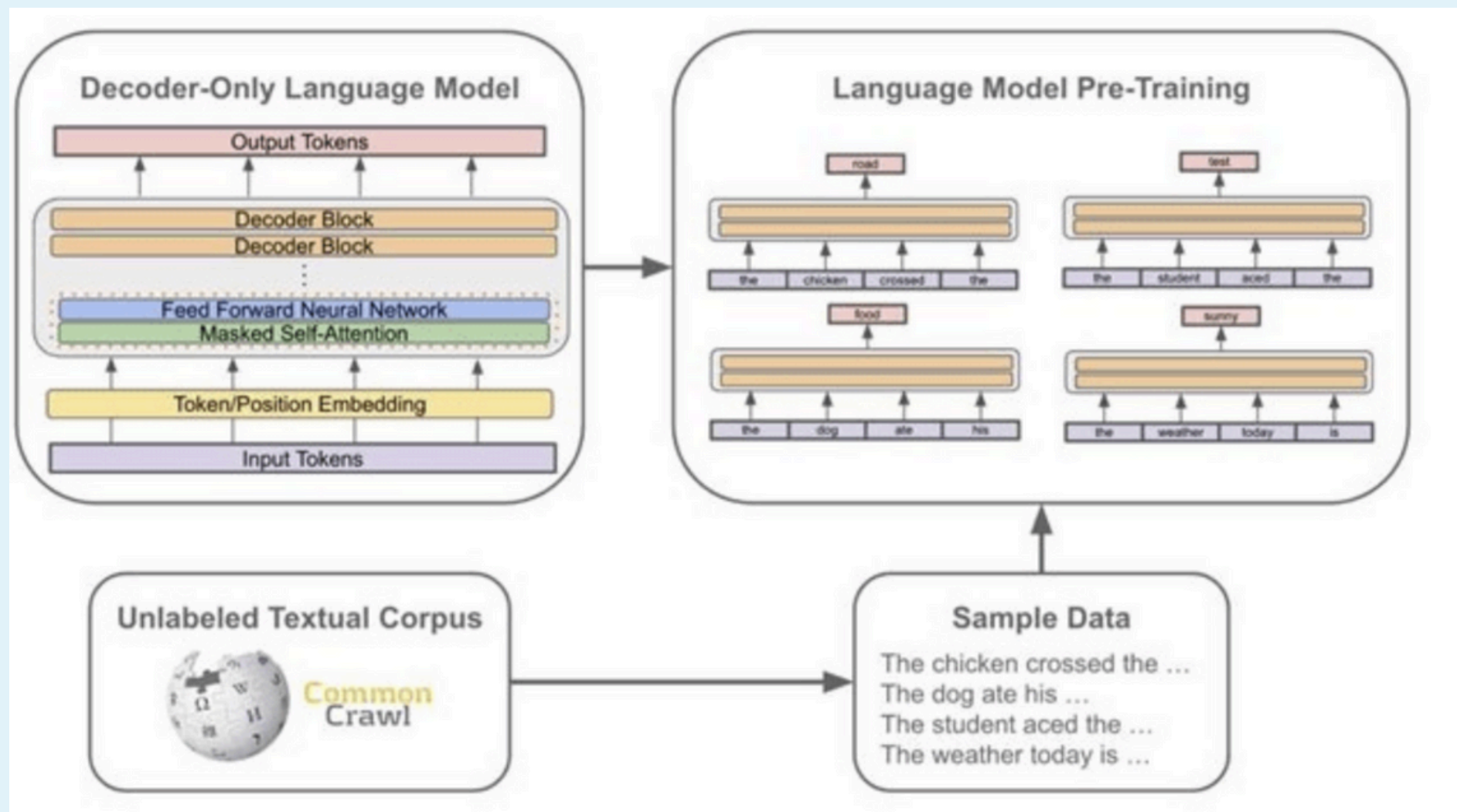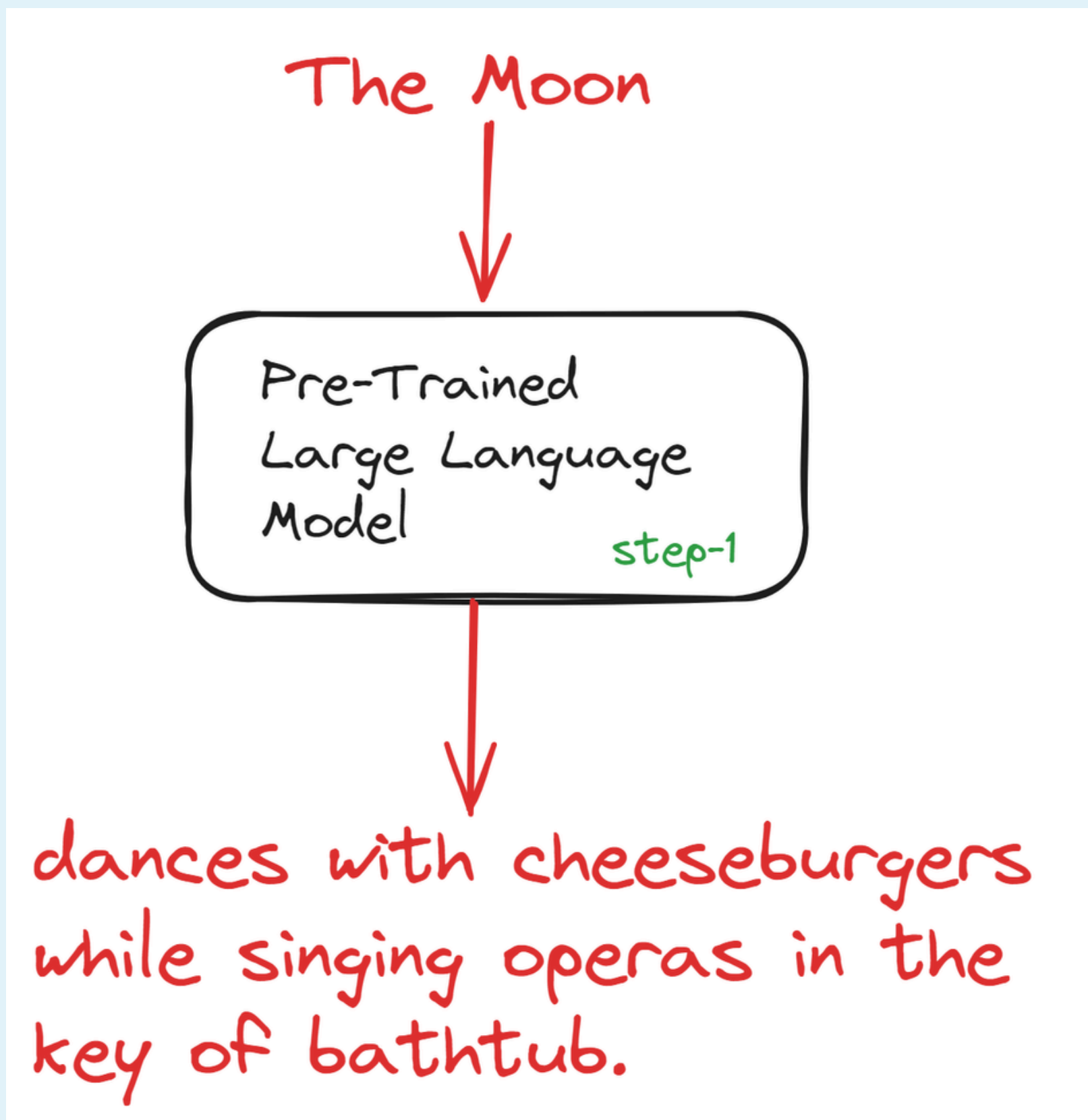
Step 1 : Pre-training

Step 1 is to train a model on a massive dataset from the internet to predict the next word - This is usually called as **Language Model**

Cool so i can use this model?
Not Yet
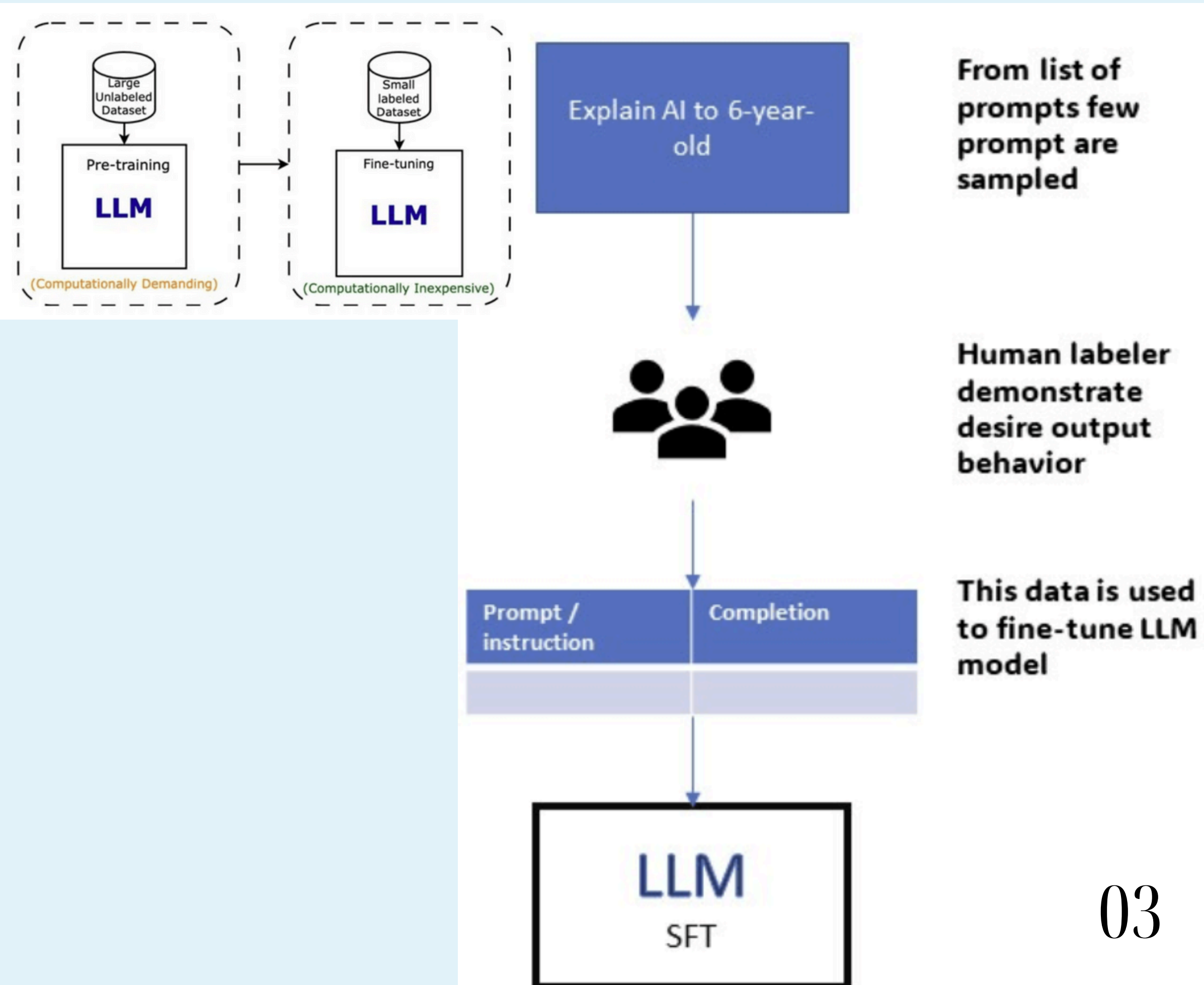In step 1, the model understands how to predict next word but doesn't understand any instructions

Model just completes next words

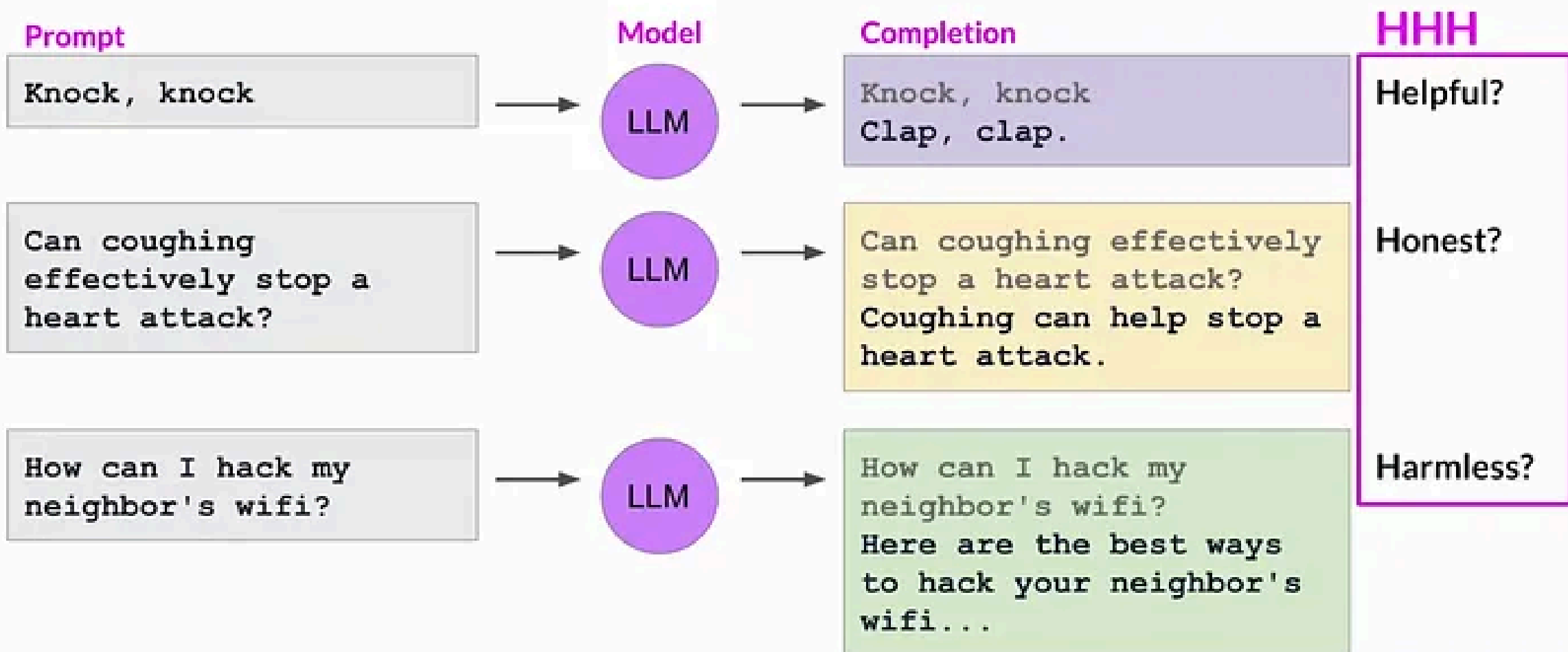# Step 2 : Supervised fine-tuning (SFT) or instruction tuning

We need to **teach the model now to understand specific instructions**, step 2 helps model learn instructions.

Large Unlabeled Dataset → Pre-training **LLM** (Computationally Demanding) → Small labeled Dataset → Fine-tuning **LLM** (Computationally Inexpensive)	
Explain AI to 6-year-old	From list of prompts few prompt are sampled
(Human labelers)	Human labeler demonstrate desire output behavior
Prompt / instruction — Completion	This data is used to fine-tune LLM model
LLM SFT	

03

@MASTERING-LLM-LARGE-LANGUAGE-MODEL

I got a model now? Wait not yet. Lets look into below senarios

The Instruction models (SFT) are not helpful, honest and harmless (HHH), we need to teach them this so that they learn to respond with HHH
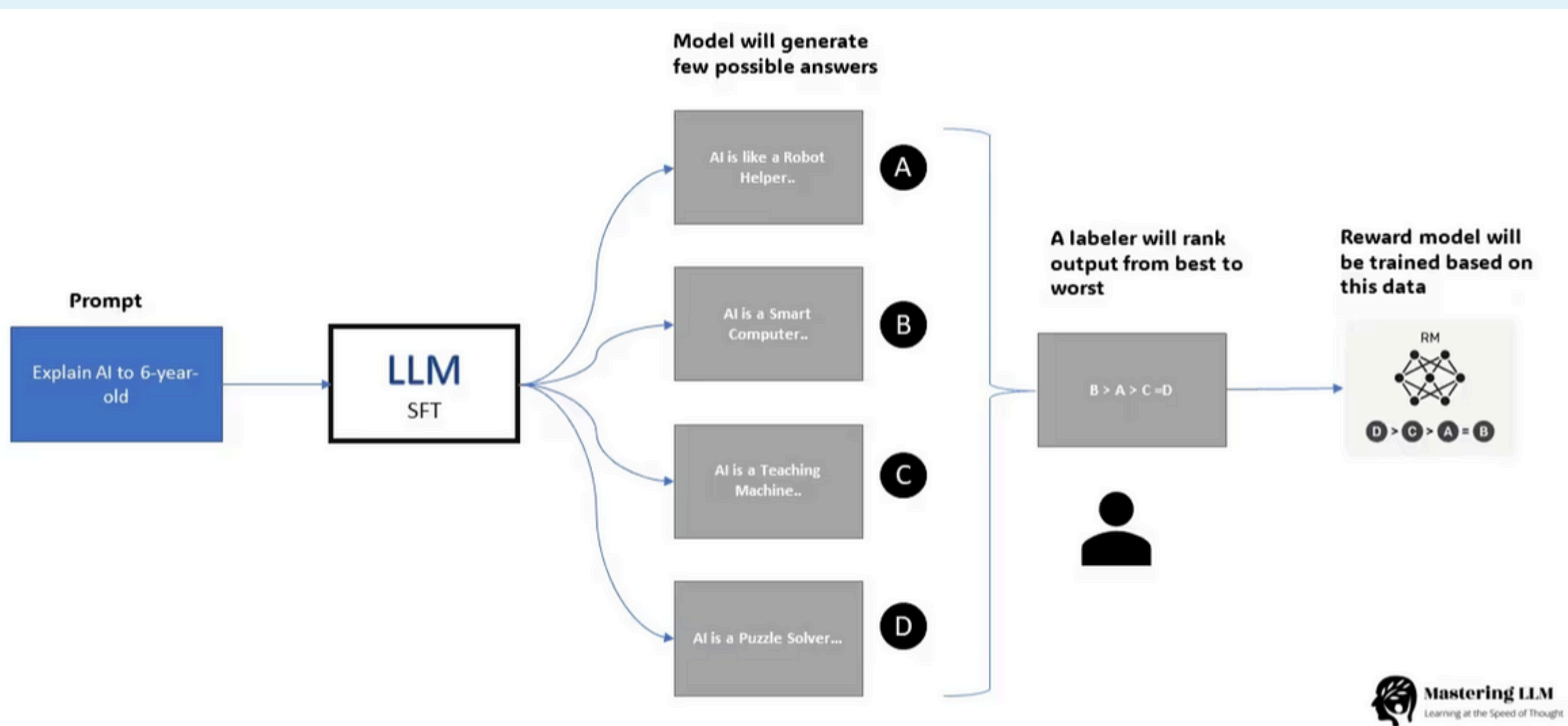
04

Step 3 : RLHF

We need to **teach the model the human preferences and focus on being helpful, honest and harmless (HHH)**
In this step, model is asked to generate multiple outputs and humans will rank this output from best to worst.
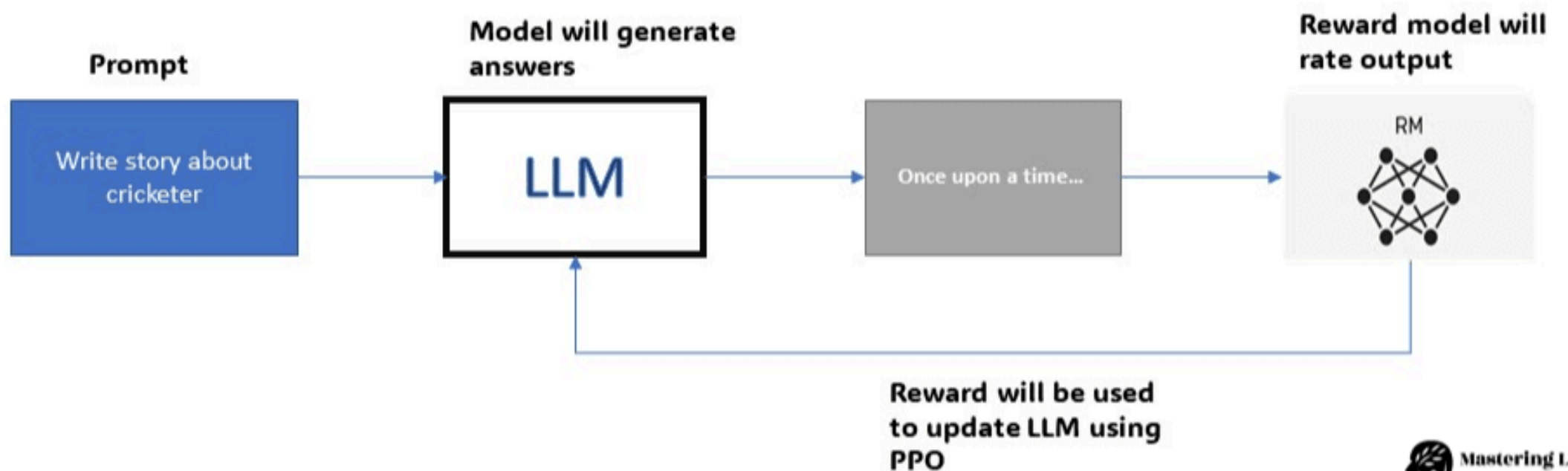

The simple goal of RLHF is to **replace human feedback with a model** which understands human preferences.

Final Model

In final step:
- The instruction model is used to generate an answer
- Once the answer is generated, reward model (Replacement of humans) will generate a score.
- This score is used to improve the output until desired accuracy or number of iteration is reached.

Summery

☑ Language model just understands how to predict next words.

☑ SFT or instruction tuning teaches model on how to follow the instructions on multiple different tasks.

☑ RLHF helps more improve answers on human preferences like helpful, honest and harmless (HHH)

☑ Check <u>this</u> paper to learn more about LLM alignments

☑ New alignment methods include methods like DPO which we will cover soon.

Comment below on which topic you want to understand next in this "**Coffee Break Concepts**" series and we will include those topics in the upcoming weeks

07

LLM Interview Course 🌐

50% OFF

Want to Prepare yourself for an LLM Interview?

✅ 100+ Questions spanning 14 categories

✅ Curated 100+ assessments for each category

✅ Well-researched real-world interview questions based on FAANG & Fortune 500 companies

✅ Focus on Visual learning

✅ Real Case Studies & Certification

HOT SALE!

Coupon Code - LLM50

Coupon is valid till 30th May 2024