

## What is Fine-tuning?

Models like GPT-3, which have been pre-trained on extensive datasets, possess a strong foundation in general language skills and knowledge. This allows them to perform well right from the start. However, their expertise remains broad. To tailor them for specific industries and tasks, we can further refine these models using smaller, targeted datasets. For instance, although GPT-3 is not explicitly trained to produce Python code, we can fine-tune it with Python-related data to make it proficient in that area.

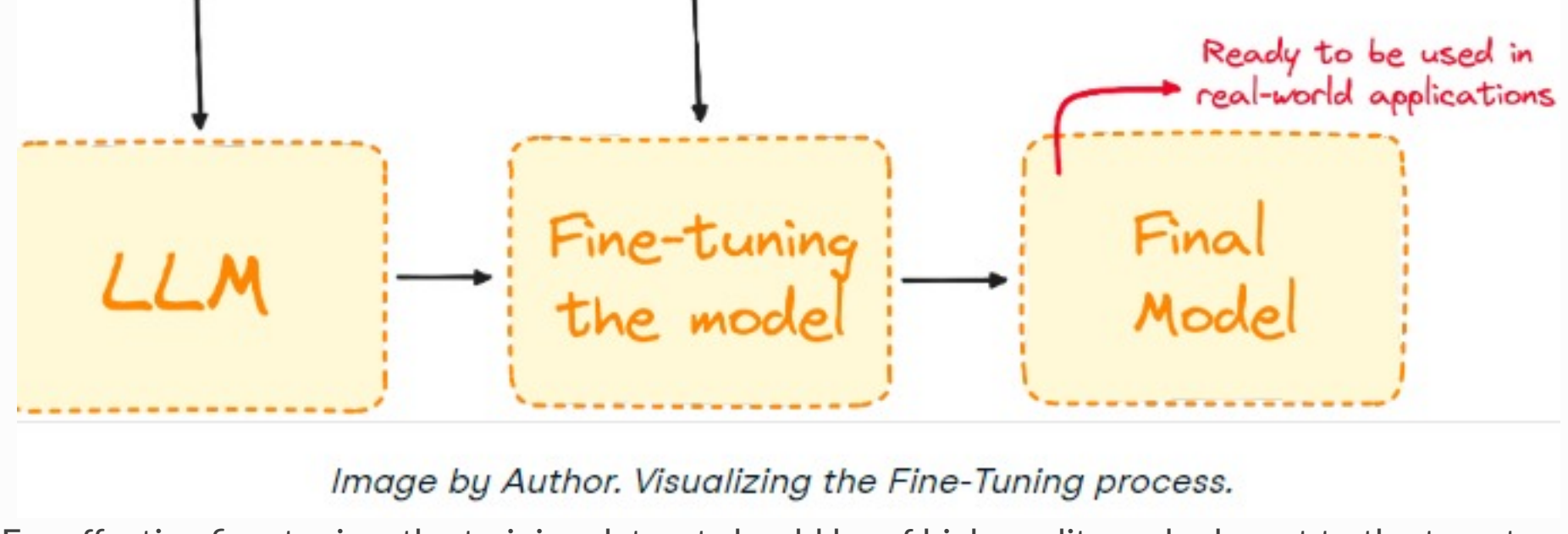
The process of fine-tuning modifies the model's internal parameters, skewing them towards the new information without erasing its previously acquired knowledge. This way, the model maintains its general abilities while acquiring new, specialized expertise.

If you're wondering how to do this, don't worry—we've got you covered.

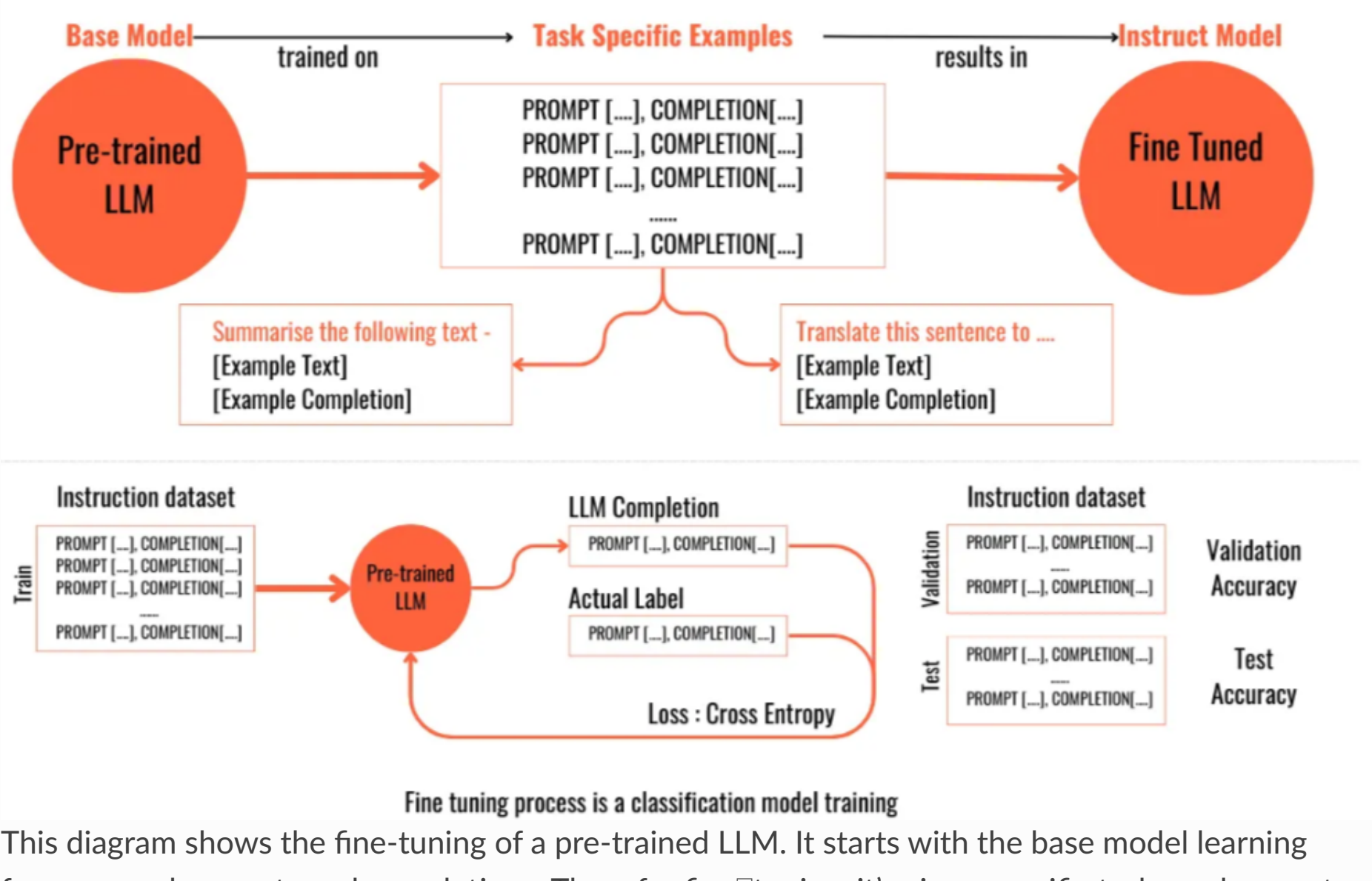
## How to Fine-tune a Large Language Model (LLM)

Here is an overview of the fine-tuning process for a model like GPT-3:

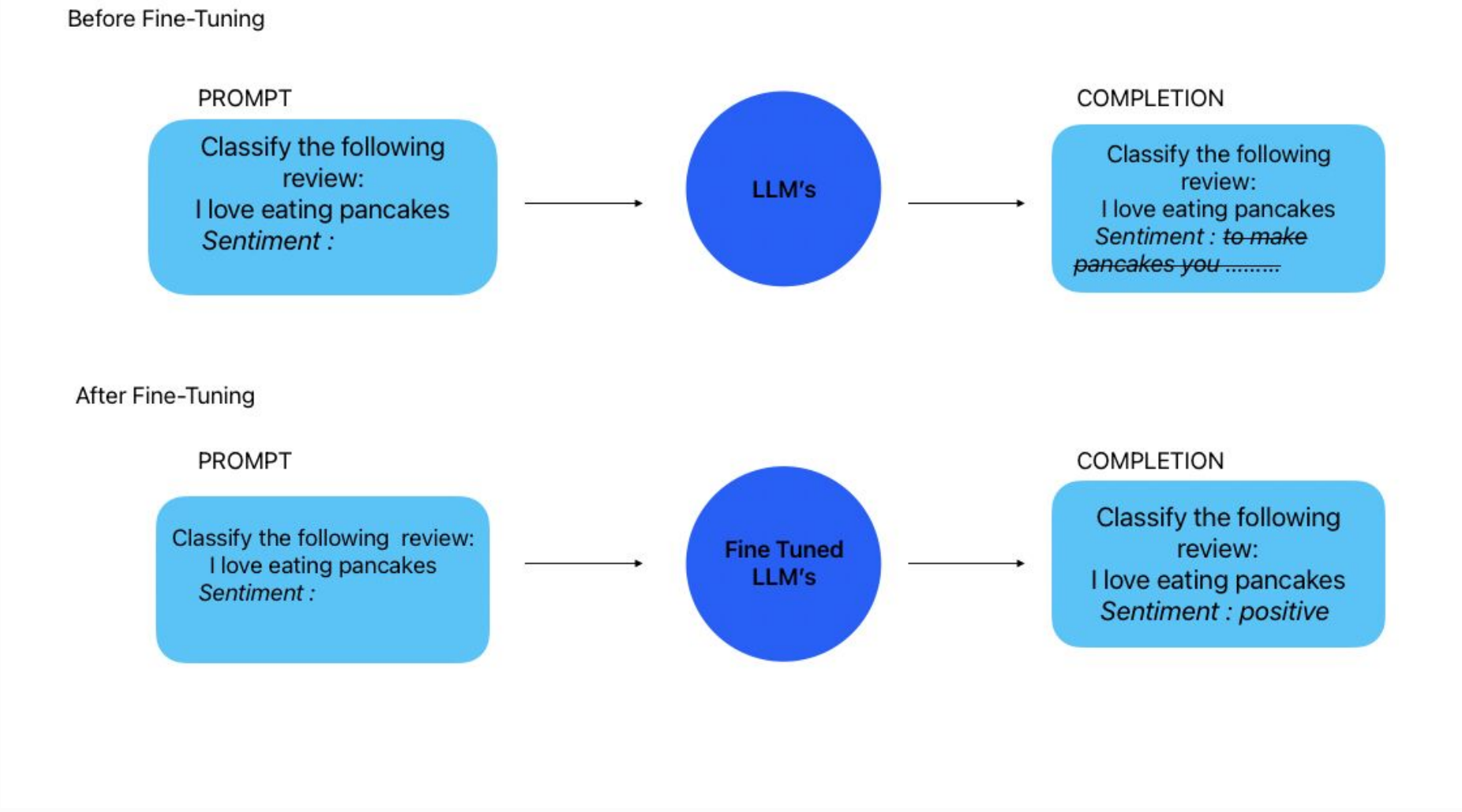
1. Begin with a pre-trained model, such as GPT-3.
2. Collect a dataset tailored to your specific task, known as the "fine-tuning set."
3. Feed examples from the fine-tuning set to the model and record its predictions.
4. Compute the loss by comparing the model's predictions to the expected outcomes.
5. Use gradient descent and backpropagation to adjust the model's parameters, aiming to minimize the loss.
6. Repeat steps 3-5 over multiple epochs until the model's performance stabilizes.
7. Once fine-tuned, the model is ready for deployment on new, unseen data.



For effective fine-tuning, the training dataset should be of high quality and relevant to the target task. Typically, a few hundred to a few thousand examples are required to fine-tune a large model successfully. Let's get into more details.



This diagram shows the fine-tuning of a pre-trained LLM. It starts with the base model learning from general prompts and completions. Then, for fine-tuning, it's given specific tasks and correct responses to learn from. It's trained and validated to improve its task accuracy, transforming it into a fine-tuned LLM specialized for particular functions. And we can simply put it this way :

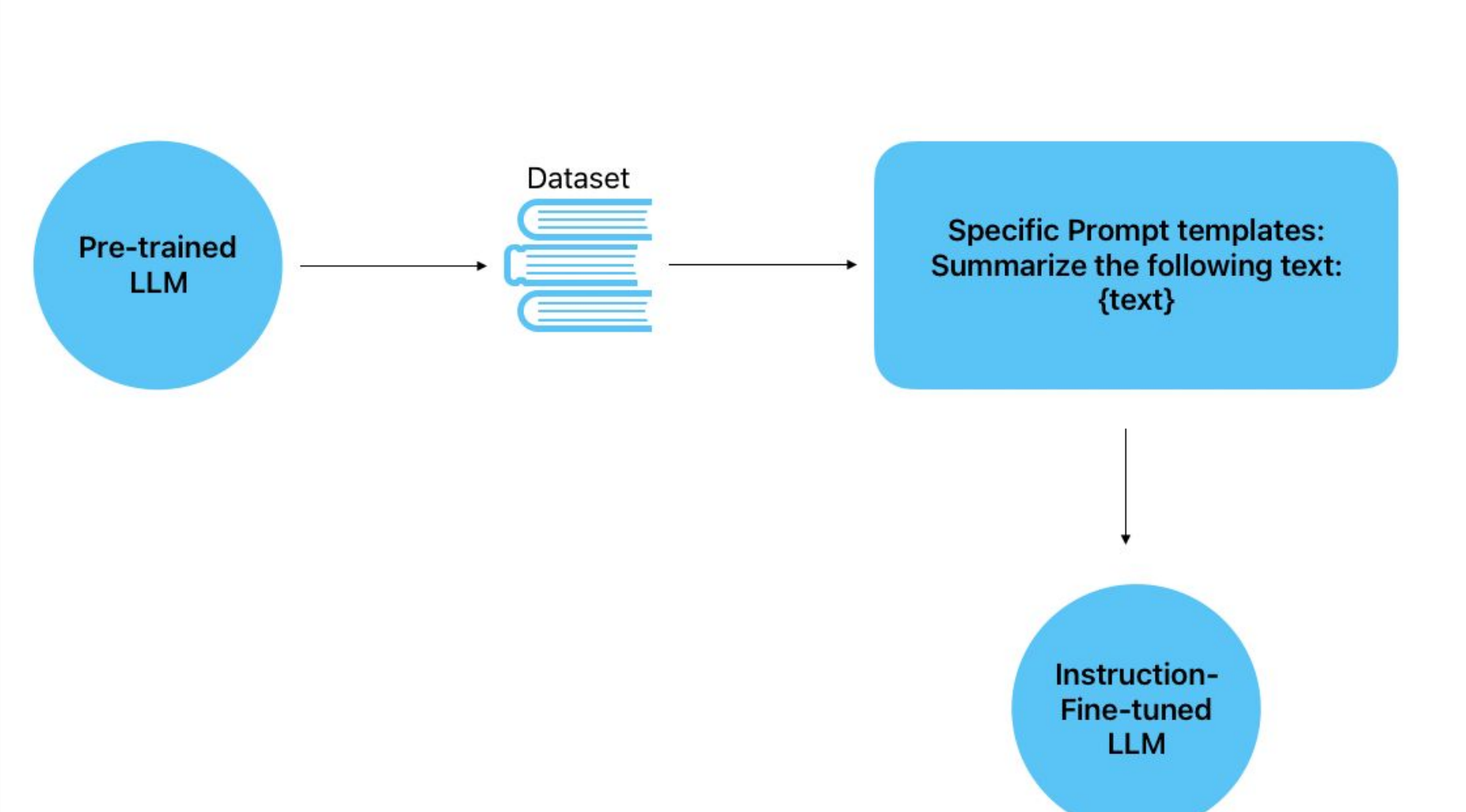


The image contrasts the output of a pre-trained LLM with that of a fine-tuned LLM. Before fine-tuning, the LLM gives an incorrect completion to a sentiment analysis prompt. After fine-tuning, the same prompt receives an accurate positive sentiment completion, demonstrating the effectiveness of the fine-tuning process.

## Instruction Fine-Tuning :

Instruction tuning or Instruction fine-tuning represents a specialized form of fine-tuning in which a model is trained using pairs of input-output instructions, enabling it to learn specific tasks guided by these instructions.

Instruction fine-tuning is particularly useful for training models to perform complex tasks that require precise, step-by-step instructions. By providing detailed guidance, the model can learn to follow the instructions accurately and produce the desired output.

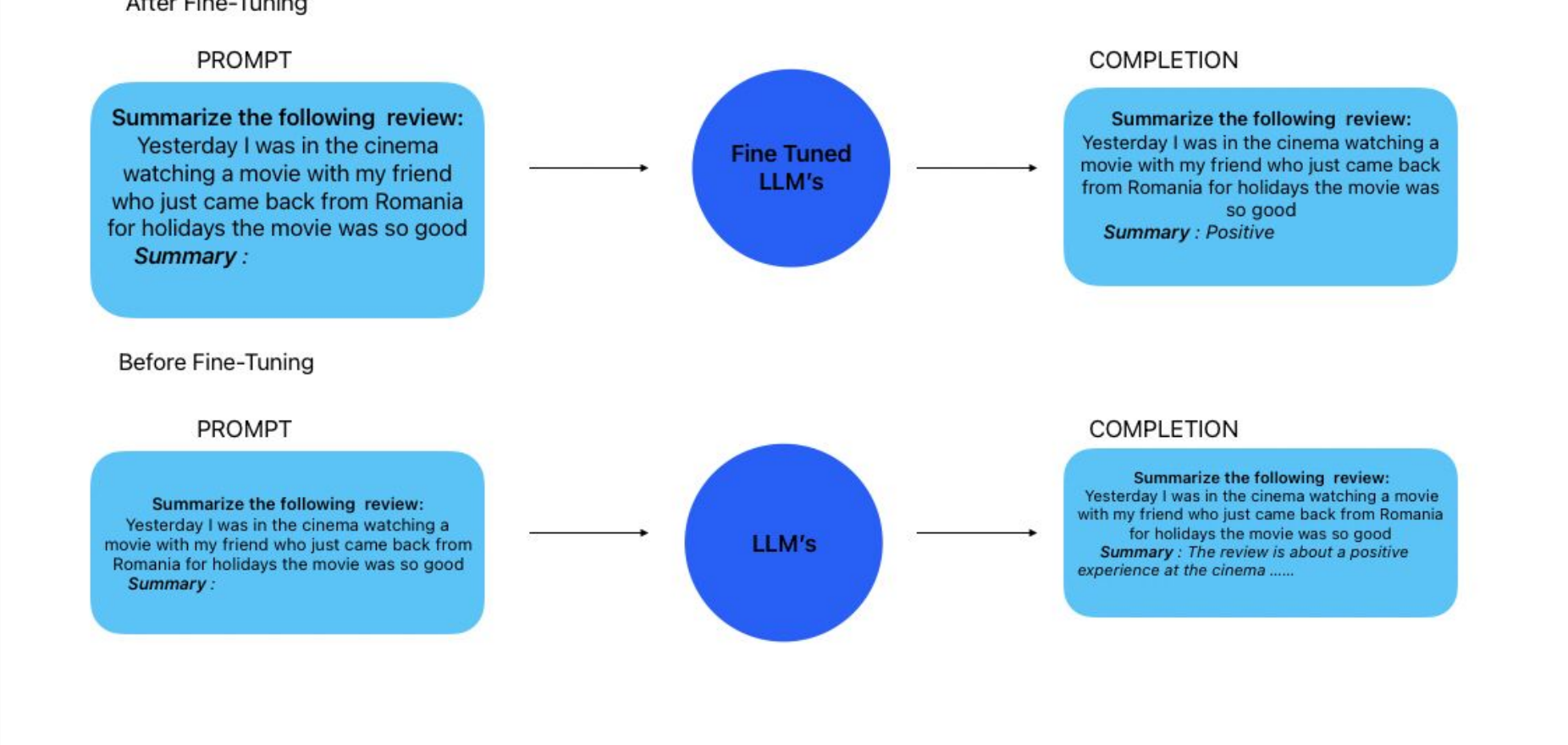


The image depicts the instruction fine-tuning process for a pre-trained LLM. The model is fine-tuned with a dataset and specific prompts that instruct it to summarize text. This process results in an LLM that has been specially trained to generate summaries, reflecting a targeted improvement in this particular task. Later on this article you will learn how to instruct fine-tune your own LLM using a framework called LUDWIG.

## Catasrophic Forgetting :

One of the challenges of fine-tuning is catastrophic forgetting. This phenomenon occurs when a model forgets previously learned information while adapting to new data. To mitigate this issue, researchers have developed various techniques, such as elastic weight consolidation (EWC) and synaptic intelligence (SI), which help models retain their original knowledge during fine-tuning.

Or we can say Catastrophic forgetting (CF) is a phenomenon that occurs in machine learning when a model forgets previously learned information as it learns new information.



This statement highlights the issue of catastrophic forgetting in fine-tuned LLMs. The same LLM, initially capable of both sentiment analysis and summarization, struggles with summarization after being fine-tuned for sentiment analysis. This exemplifies the problem where an LLM, post fine-tuning, may lose its ability to perform tasks it could handle before, emphasizing the need for strategies to retain previous knowledge while incorporating new capabilities.

## How to prevent Catastrophic Forgetting :

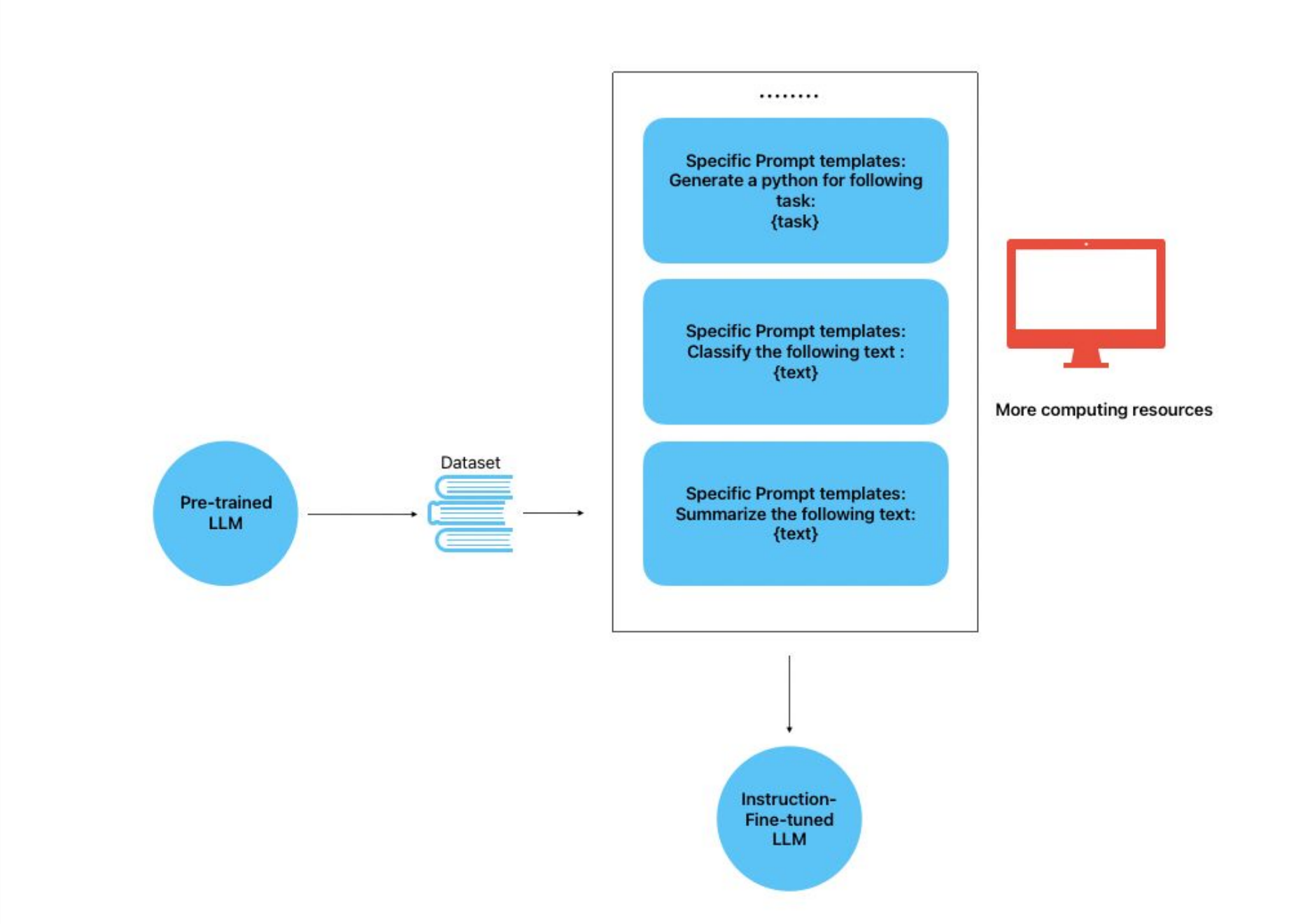
Researchers have developed several strategies to tackle catastrophic forgetting in LLMs:

1. **Elastic Weight Consolidation (EWC):** EWC is a regularization technique that helps LLMs retain previously learned information while fine-tuning on new tasks. By assigning importance weights to the model's parameters based on their relevance to the original task, EWC helps prevent catastrophic forgetting.
2. **Synaptic Intelligence (SI):** SI is another regularization method that addresses
3. **Multi-Task Learning:** Training an LLM on multiple tasks simultaneously can help prevent catastrophic forgetting. By exposing the model to diverse tasks during training, it learns to balance its knowledge across different domains, reducing the risk of forgetting.
4. **Full fine-tuning:** In some cases, full fine-tuning may be necessary to achieve optimal performance on a specific task. While this approach can lead to catastrophic forgetting, it may be the most effective way to adapt the model to the new task.
5. **Knowledge Distillation:** Knowledge distillation involves transferring the knowledge from a large, pre-trained model to a smaller, task-specific model. By distilling the information from the larger model, the smaller model can benefit from the pre-trained knowledge without the risk of catastrophic forgetting.
6. **PEFT** : Parameter Efficient Fine-Tuning (PEFT) is a technique that aims to minimize catastrophic forgetting by fine-tuning only a subset of the model's parameters. By identifying and updating the most relevant parameters for the new task, PEFT helps preserve the model's original knowledge while adapting to the new task.

## Instruction Multi-Fine-Tuning :

In order to prevent the (CF) phenomena, we want to make sure to train our model on a variety of tasks, So the model can no longer forget how to perform every task.

Instruction Multi-Fine-Tuning is a technique that involves fine-tuning a model on multiple tasks simultaneously using instruction-based training. By providing detailed instructions for each task, the model can learn to perform a diverse range of tasks without forgetting previously acquired knowledge.



This image outlines the instruction multi-task fine-tuning process. A pretrained LLM is further trained with a dataset and various specific prompt templates for different tasks, like generating Python code, classifying text, and summarizing text. This multi-task fine-tuning requires more computing resources and results in a more versatile