Fine-Tuning Pretrained

Large Language

Models

Training LLMs

Training for an LLM is a long and compute intensive task. The level of current models such as GPT5 with 1.8trillion. This initial training create a generalist model that struggles with specific information and queries. A more specifically trained model is needed to allow for accurate domain specific responses. Recreating the model for each task in this case is highly inefficient and would lead to bloated systems with many AIs. As every task would require its own fully trained model.

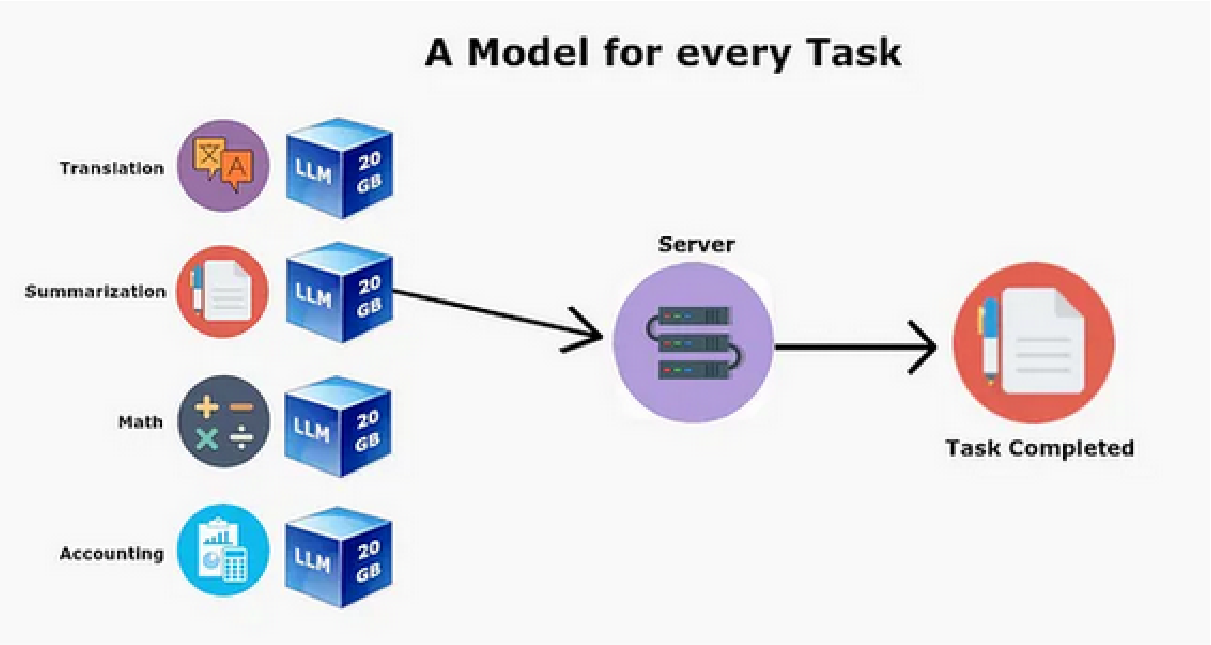

Figure 1: One model for every task

Adapter Architecture

This deficiency in base models opened the door for the adapter architecture to take place. This approach applies an additional layer to the AI adapter. This adapter is a created by freezing the current model weights, then training an additional set of weights to act upon the base model that allows the LLM to have its weights altered only by the adapter letting it be a plug and play solution, The matematical representation of this is at a high level:

min L(D;θ0 +Δθ)

- L : This represents the loss function used to calculate the gradient that needs to be minimized for the LLM.
- D : This represents the dataset the loss function is being trained on and what is being optimized for.
- Θ : θ0 represent the weights for base model and Δθ represent the weights from the fine tuning. This is how the adapter is swappable as the weights are not integrated into the base model.

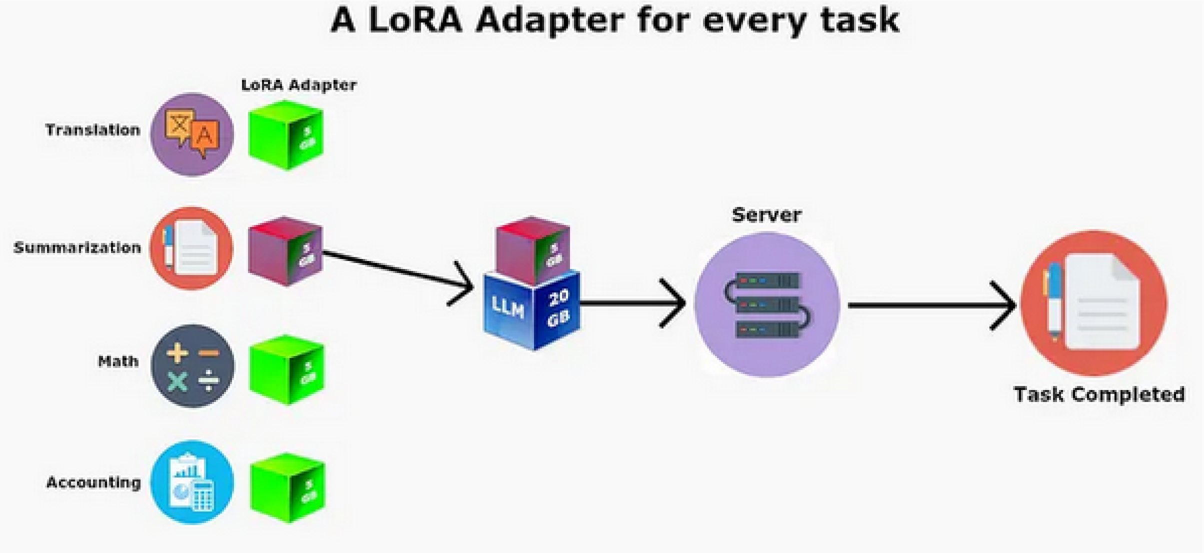

Figure 2: Adapter pattern

Fine Tuning approaches

Full Parameter Fine tuning

This was the first fine tuning approach proposed in 2018 (Jeremy Howar and Sebastian Ruder), which at the time was called ULMFiT. This is the simplest approach conceptually as it creates a mirror of the model weights and trains those to scale each weight individually to work towards the new goal. This is still a computationally heavy process as every single weight is modified but as a baseline it allows for the adapter architecture to be used.

Lora Fine Tuning

Vera Fine Tuning

QLora Fine Tuning