Lecture 6

Pre-training is done on unlabeled data. (Foundational model)

But fine tuning is done on labeled data.

LLMs are mainly based on the transformer architecture. The key idea behind transformers is the attention mechanism. Attention mechanism helps the LLM to understand the importance of words, not just the word in the current sentence, but also the words that came before. Because context is important for predicting the next word. So attention mechanism gives access to the entire context and select which words to put more weight on to predict the next one.