

What is the main architectural difference between BERT and GPT? Which uses an encoder and which uses a decoder?

Ans:

- BERT uses a Transformer-based encoder architecture designed for understanding context from both directions (bi-directional).
- GPT uses a Transformer-based decoder architecture, focusing on generating text sequentially from left to right (uni-directional).

Explain why using pre-trained models (like BERT or GPT) is beneficial for NLP applications instead of training from scratch.

Ans:

- Efficiency: Pre-trained models save significant computational resources and time since they have already learned general language patterns from large datasets.
- Performance: They typically achieve better accuracy because they capture extensive linguistic knowledge, which helps them perform effectively even on relatively small domain-specific datasets.
-