

Module 13

Compare BERT,GPT-2 and XLNET.Write down the differences between them.

Feature	BERT (Bidirectional Encoder Representations from Transformers)	GPT-2 (Generative Pre-trained Transformer 2)	XLNet (eXtreme Language Model)
Architecture	Encoder-only model	Decoder-only model	Encoder-decoder model
Training Objective	Masked Language Modeling (MLM) and Next Sentence Prediction (NSP)	Autoregressive language modeling (predicts next word)	Permutation Language Modeling (predicts words in various orders)
Context Handling	Considers both left and right context simultaneously (bidirectional)	Considers left context only (unidirectional)	Considers context in all possible permutations (bidirectional)
Pre-training Data	Trained on a large corpus including Wikipedia and BookCorpus	Trained on a large corpus from the internet (WebText)	Trained on the same data as BERT but using permutations of the sequences
Use Cases	Excellent for tasks requiring deep understanding of context like question answering and sentence classification	Effective for generating coherent and contextually relevant text	Better at capturing long-term dependencies and can outperform BERT on certain NLP tasks
Performance	Strong in tasks requiring understanding of bidirectional context	Strong in text generation tasks	Often surpasses BERT in performance on some NLP benchmarks

- ➔ BERT focuses on bidirectional context understanding with masked language modeling,
- ➔ GPT-2 is unidirectional and excels in text generation
- ➔ while XLNet combines bidirectional and autoregressive approaches for improved performance on various NLP tasks.