**LARGE LANGUAGE MODEL**

**Architecture:**

* Encoders - embed
* Decoders – generate text

1. Multiple Architectures focused on encoding and decoding that is embedding and text generation.
2. All Models built on the Transformer Architecture.

**Encoders:**

Models that convert a sequence of words to an embedding (vector representation).

**Ex:**

MiniLM, Embed-Light, BERT, RoBERTA, DistillBERT, SBERT,…

**Decoders: - produces single token at a time**

Models that take a sequence of word and output next word.

**Ex:**

GPT-4, Llama, BLOOM, Falcon,…