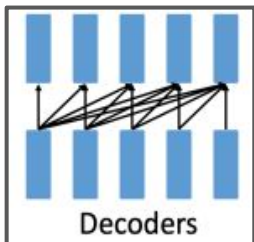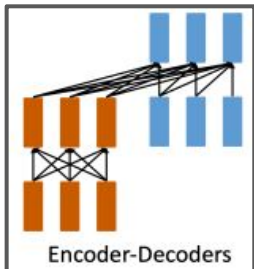Encoder-only (e.g., BERT): bidirectional contextual embeddings

Decoder-only (e.g., GPT-x): unidirectional contextual embeddings, generate one token at a time

Encoder-decoder (e.g., T5): encode input, decode output