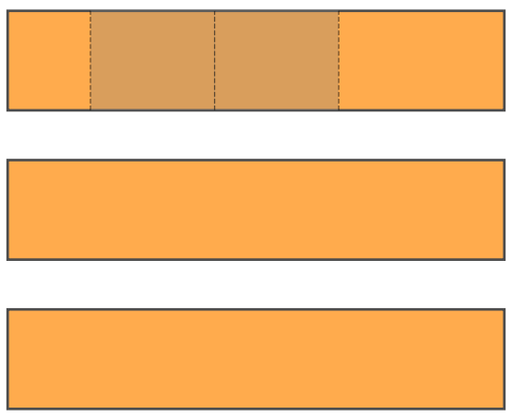
Transformer Architecture Paradigms

Encoder-Only (BERT)

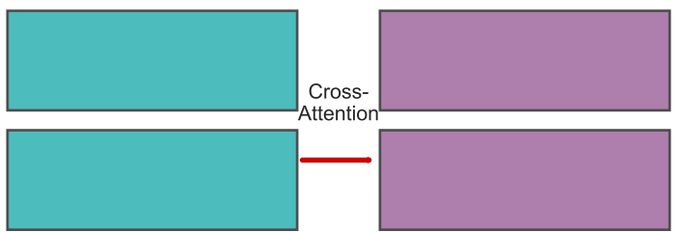
Bidirectional Attention

Decoder-Only (GPT)



Causal Attention

Encoder-Decoder (T5)



Encoder

Decoder