Architecture Comparison: Encoder vs Decoder

**BERT: Encoder Stack
(Bidirectional)**
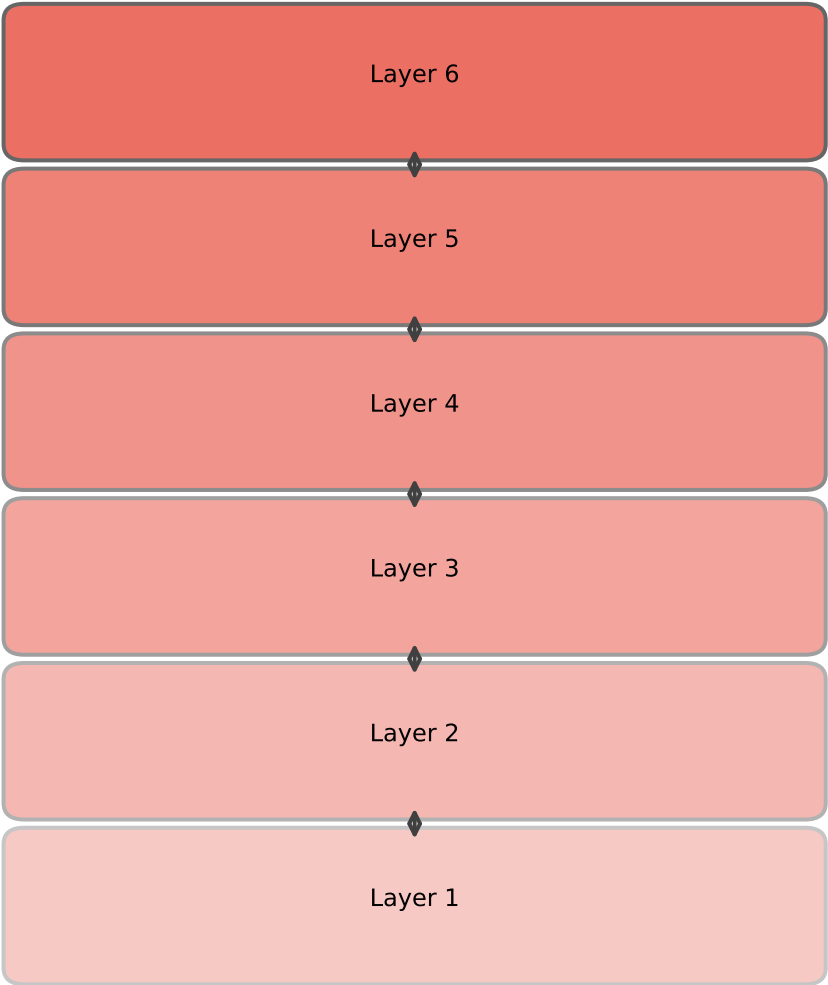
**GPT: Decoder Stack
(Causal)**

**Full Context
No Mask**

**Left-to-Right
Causal Mask**

BERT	GPT
Layer 6	Layer 6
Layer 5	Layer 5
Layer 4	Layer 4
Layer 3	Layer 3
Layer 2	Layer 2
Layer 1	Layer 1