RNN

Recurrent Neural Networks
Step-by-step sequential processing

→

Problems

⚠ Vanishing gradient
⚠ Sequential processing limits
⚠ Long sequence challenges

↓

LSTM & GRU

Addressed vanishing gradient with gates

→

Remaining Issues

⚠ Still sequential processing
⚠ Limited parallelization
⚠ Long-range dependencies

↓

Attention Mechanism

Improved encoder-decoder models
Direct connections to all positions

↓

Solution: Transformer Architecture