RNN Attention vs. Self-Attention

Aspect	RNN Attention	Self-Attention
📍 Source	Query from decoderKeys/Values fromencoder	All from thesame sequence
⚙️ Processing	Sequentialprocessing	Parallelprocessing
🔗 Relationships	Limited to current step	Captures relationships between all positions
⏱️ Complexity	$O(n)$per step	$O(n^2)$total

⚡ Key Advantage

Self-attention enables parallelization and better long-range dependency modeling