RNN vs Transformer Comparison Table (Landscape PDF)

Aspect	RNNs (LSTM/GRU)	Transformers
Architecture	Sequential recurrence Hidden state step-by-step	Parallel feed-forward layers Self-attention
Parallelization	Inherently sequential	Fully parallel over sequence
Long-Range Dependencies	Hard; vanishing gradients	Excellent via attention
Computational Cost	$O(T)$ sequential	$O(T^2)$ but parallel
Memory Scaling	$O(1)$ per step	$O(T^2)$
Training Speed	Slow for long seqs	Fast
Inference	Streaming $O(1)$ per token	Requires KV cache (grows with T)
Context Length	Limited	Long (4k–1M tokens)
Data Efficiency	Good on small data	Needs more data
Model Size	Compact models	Scales extremely well
Stability	Harder gradients	Very stable
Edge/Mobile	Excellent	Harder; memory-heavy
Streaming Tasks	Perfect fit	Not ideal
Best Use Cases	RL, IoT, time-series	NLP, vision, multimodal
SOTA Performance	Rare today	Dominant everywhere