LSTM (Long Short-Term Memory)

✓ Solves vanishing gradient problem with gating mechanism

Forget Gate

Decides what to remove from cell state

Input Gate

Controls new information addition

Output Gate

Determines hidden state output

Cell State

Long-term memory pathway

Hidden State

Short-term working memory

🎯 **Success**

Widely successful for various sequence tasks across domains

📊 RNN vs LSTM Comparison

🔴 RNN Issues

- Vanishing gradient problem
- Cannot capture long-term dependencies

✅ LSTM Solutions

- Gating mechanism prevents gradient vanishing
- Maintains long-term memory via cell state

- Simple architecture with single state
- Gradient diminishes exponentially over time

- Complex architecture with gates
- Controlled information flow preserves gradients

💡 Gradient Flow Example

Consider a sequence of length T with gradient backpropagation:

$$\text{RNN: } \partial L/\partial h_0 = \partial L/\partial h_t \times \left(\prod_{i=1}^{T} \partial h_i/\partial h_{i-1}\right) \rightarrow \text{vanishes when T is large}$$

LSTM gates control gradient flow, preventing exponential decay:
Forget gate + Input gate + Cell state = Stable gradient pathway