**Standard RNN
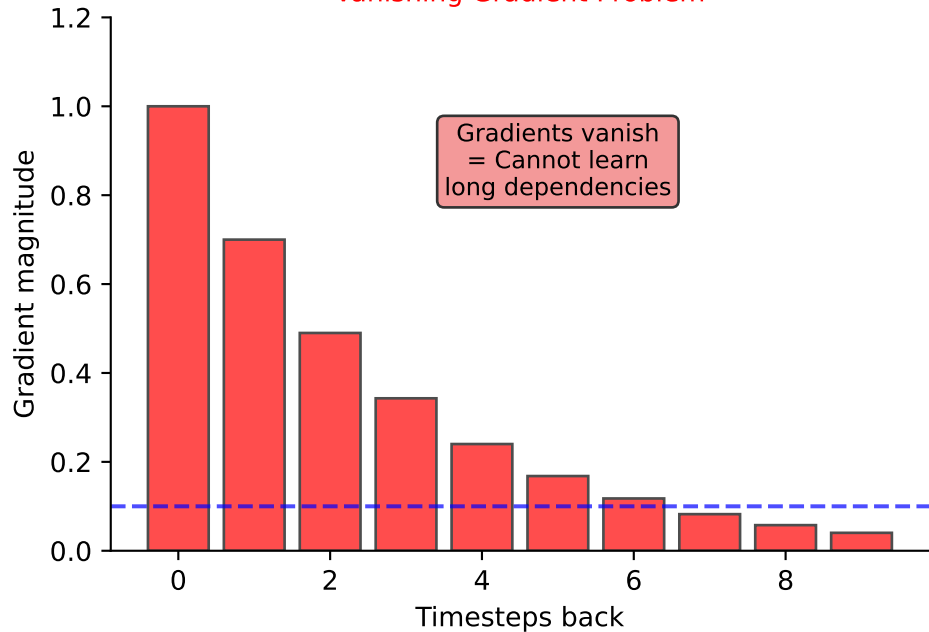Vanishing Gradient Problem**

Gradients vanish
= Cannot learn
long dependencies

Gradient magnitude

Timesteps back

**LSTM
Gradients Preserved**

Cell state allows
gradients to flow
= Learns long
dependencies

Gradient magnitude

Timesteps back