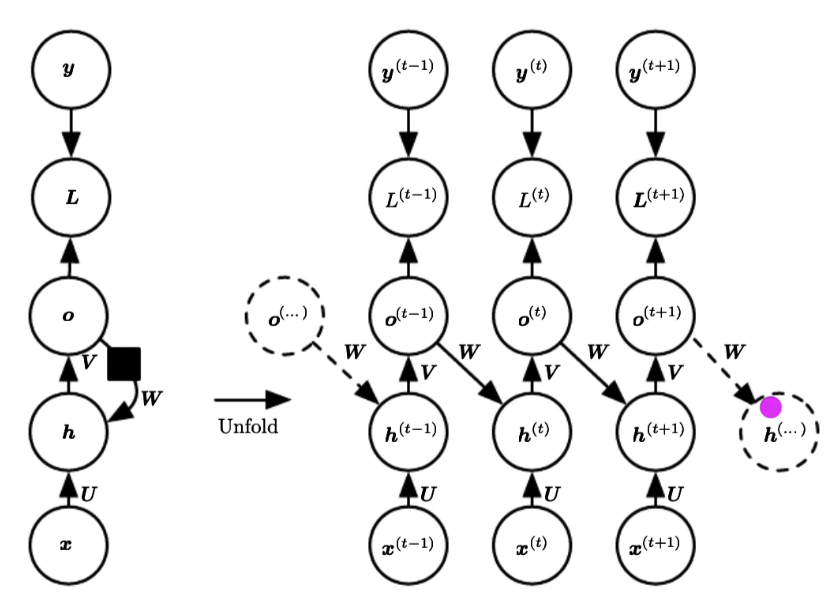
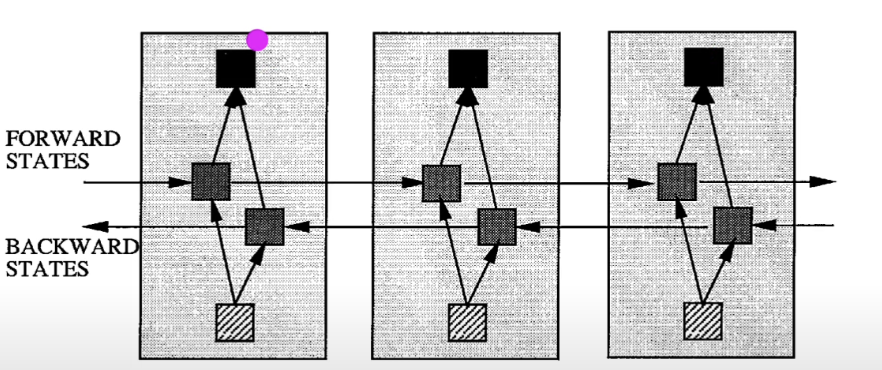
# Content

1. Motivation for attention
2. Transformer architecture
3. Self-attention [code, math, intuition]

# Recurrent Neural Networks Disadvantages

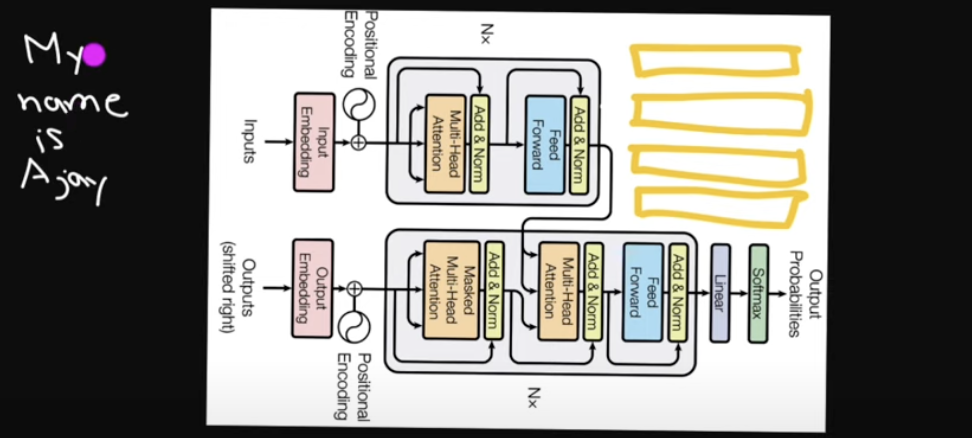


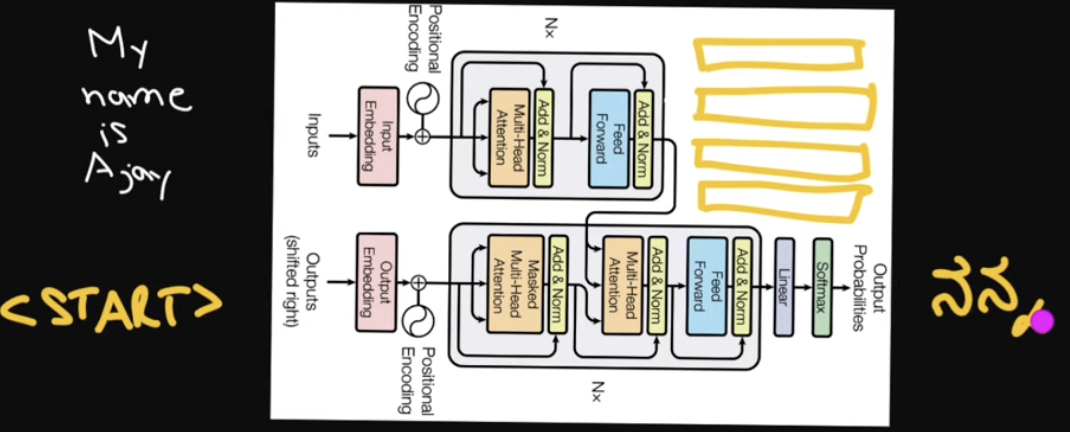


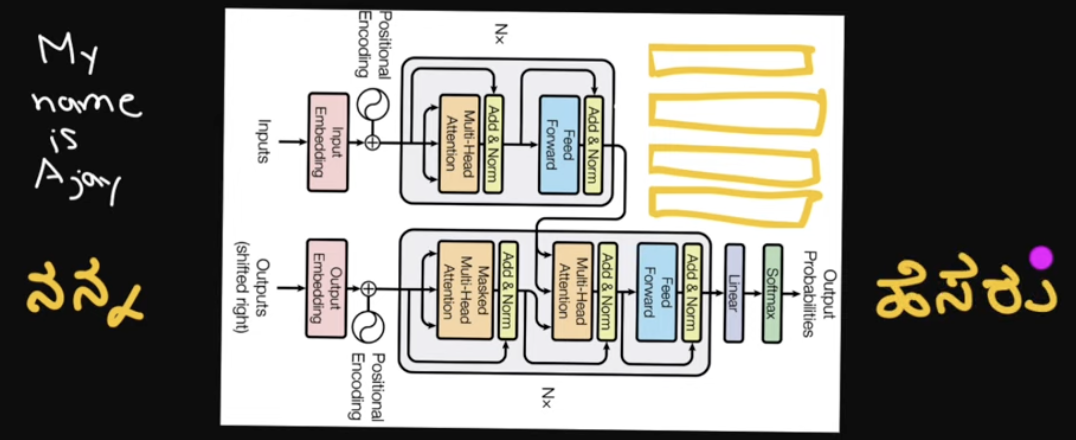
# Pass 1: Attention Mechanism



# Pass 2: Transformer Overview







Diagram

Description automatically generated

# Pass 3: Attention in Transformers

Diagram

Description automatically generated

Diagram

Description automatically generated