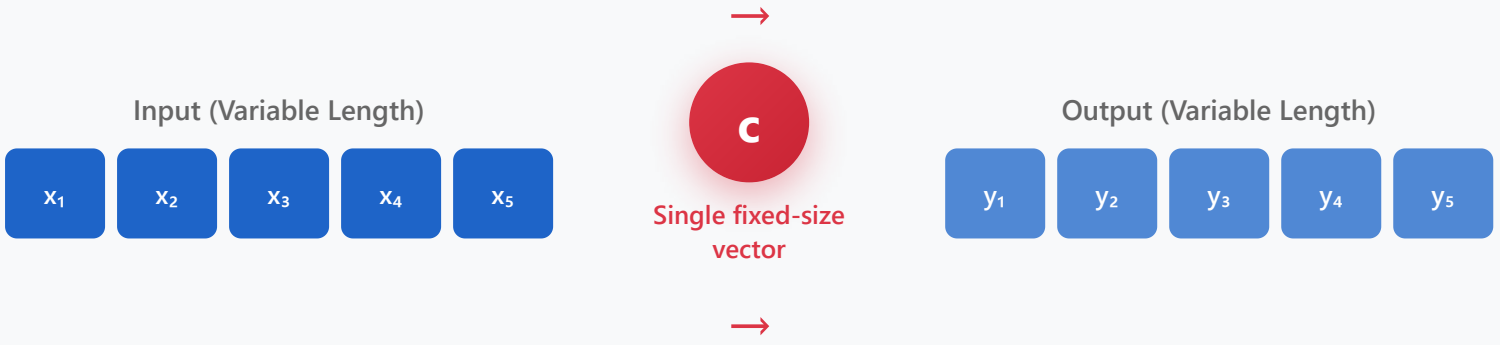
Limitations of Basic Seq2Seq Models

The Information Bottleneck Problem



1 Information Bottleneck

All input information must be compressed into a **single fixed-size context vector**, regardless of input length.

Impact:

Long sequences lose critical information. Model struggles with lengthy inputs.

2 Gradient Vanishing

Earlier tokens in long sequences have **vanishing gradients** as information must flow through many time steps.

Impact:

Poor performance on long-range dependencies and long documents.

3 Equal Weighting

All input tokens treated equally when creating context vector - **no selective focus** on important parts.

Impact:
Cannot emphasize relevant information for each output step.

4 No Alignment

Decoder has **no mechanism to look back** at specific input positions when generating each output token.

Impact:
Difficult tasks requiring input-output alignment (e.g., translation).

Solution needed: Attention Mechanism allows decoder to selectively focus on different parts of input for each output step!