Attention: Motivation

Let's revisit the Encoder-Decoder architecture

The Encoder

- Acts on input sequence $\left[\mathbf{x}_{(1)} \ldots \mathbf{x}_{(\bar{T})}\right]$
- Producing a sequence of latent states $\left[\bar{\mathbf{h}}_{(1)}, \ldots, \bar{\mathbf{h}}_{(\bar{T})}\right]$