

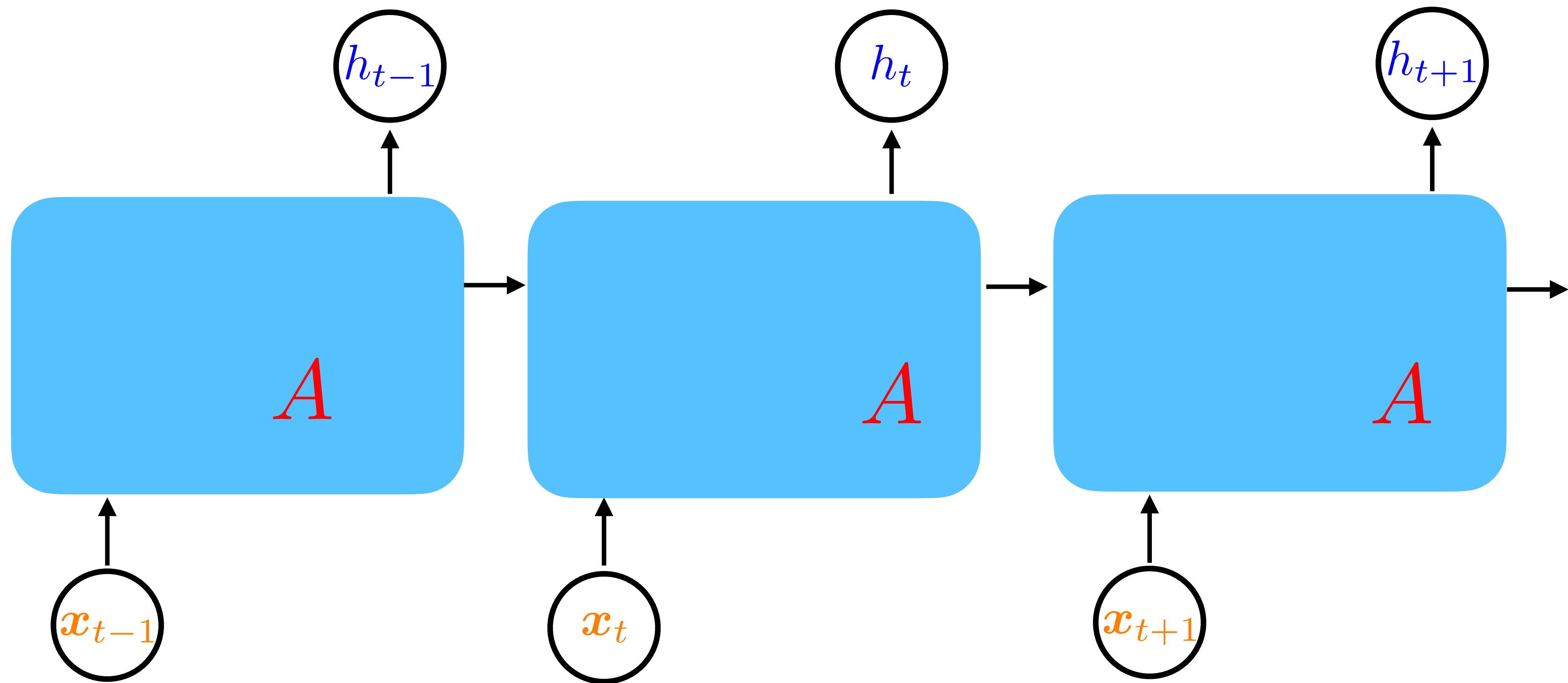
# Sequence to Sequence Model

COMP3361 — Week 3

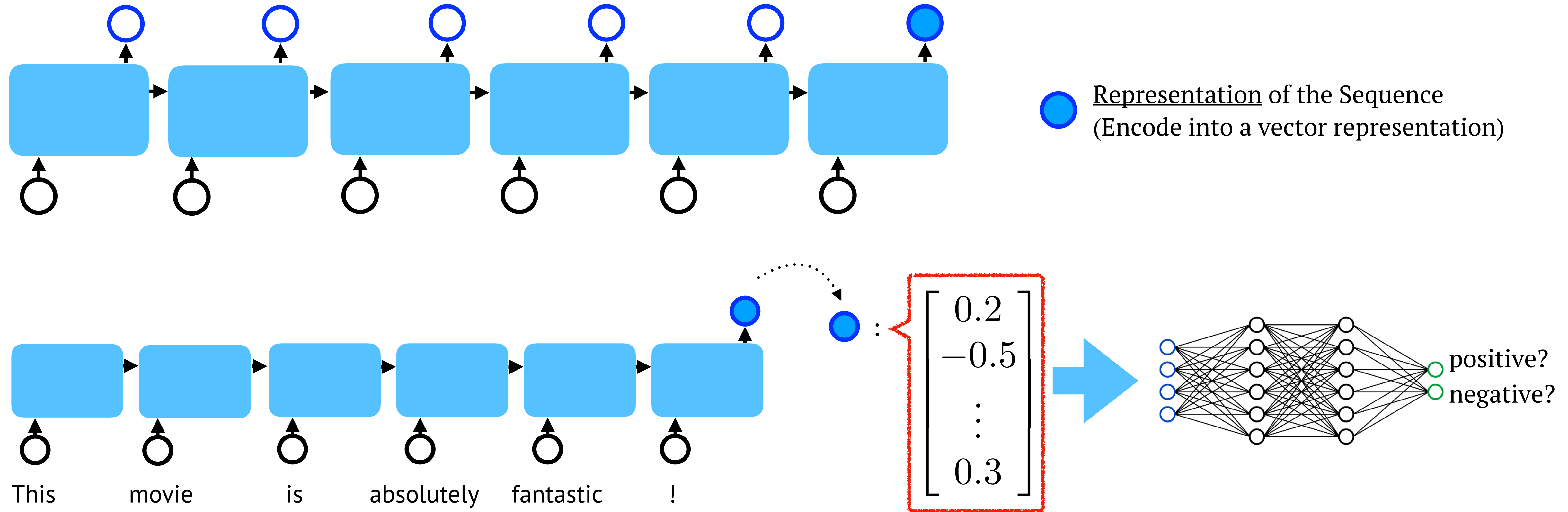
Lingpeng Kong

Department of Computer Science, The University of Hong Kong

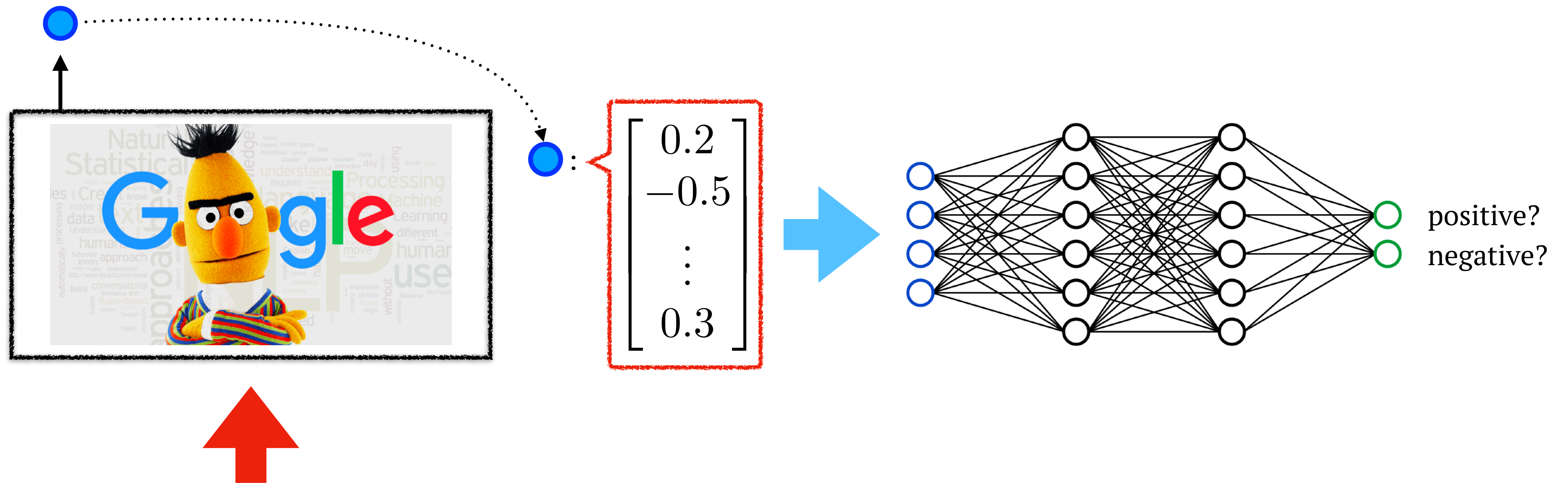
# Recurrent Neural Network



# RNN as Encoder

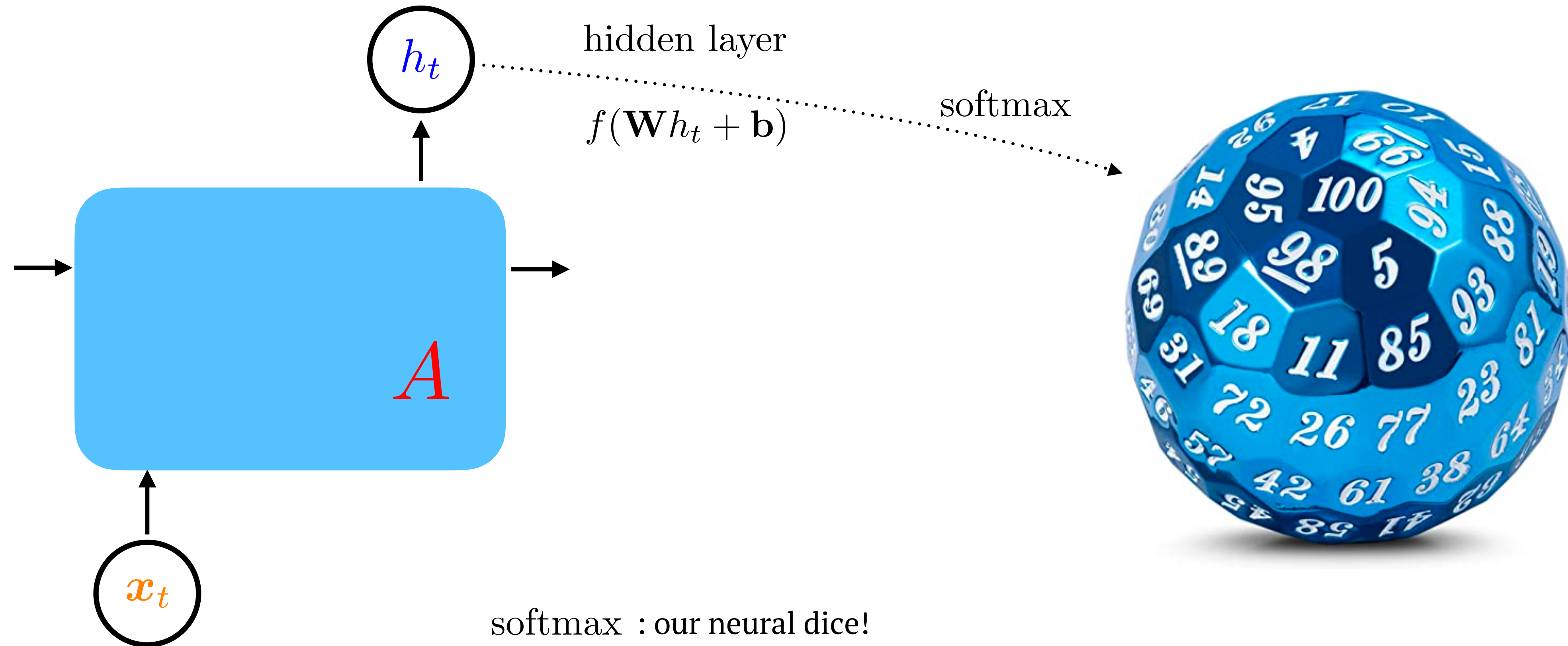


# Flashback: Pretraining and Fine-tuning



This movie is absolutely fantastic !

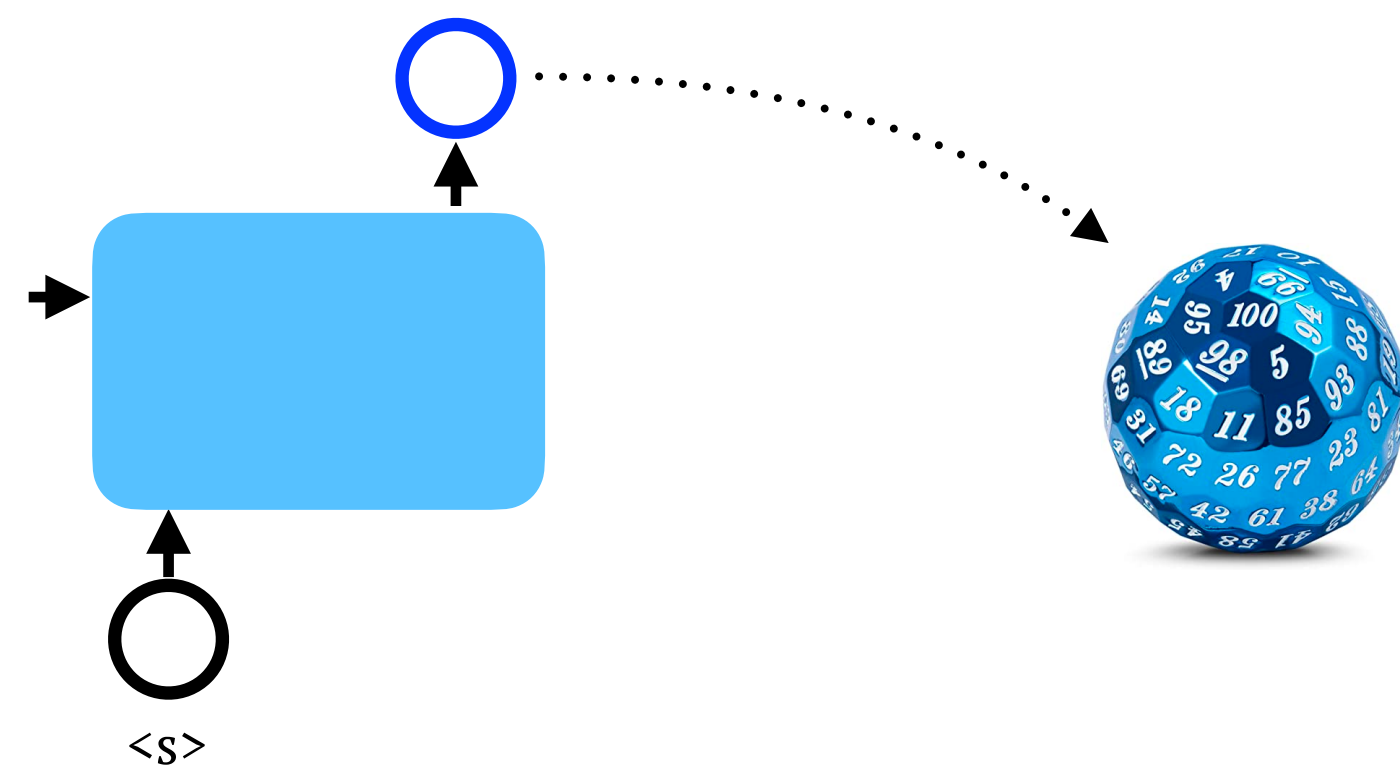
# Sample a sentence from RNNLMs





# Sample a sentence from RNNLMs

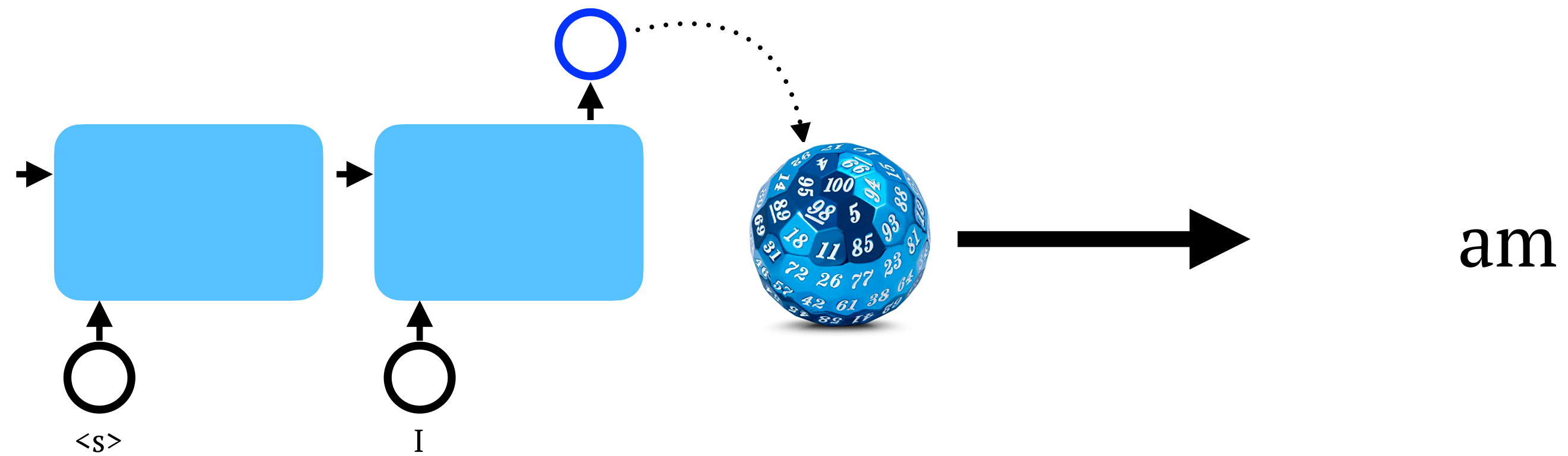
I



I

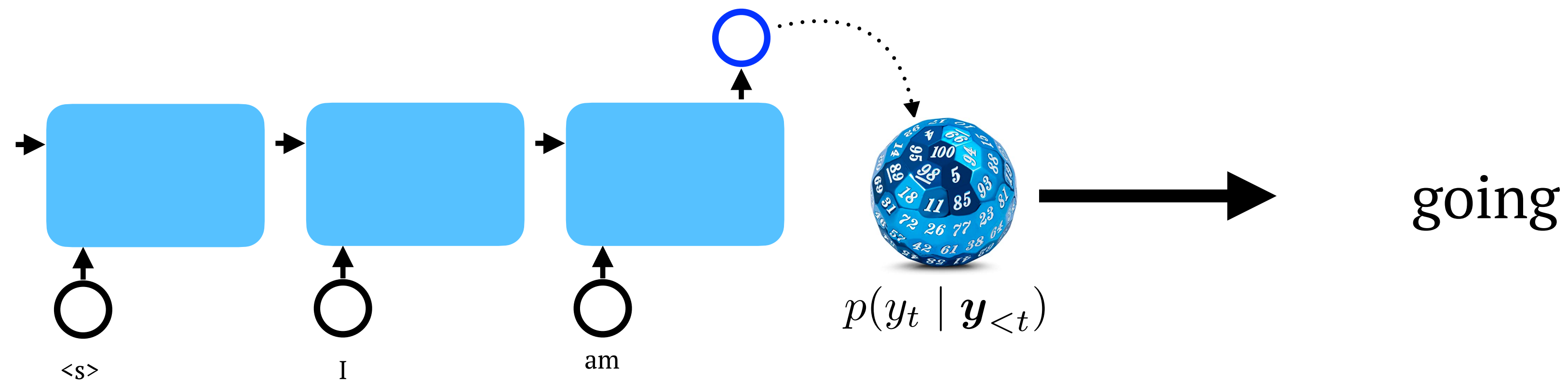
# Sample a sentence from RNNLMs

I am



# Sample a sentence from RNNLMs

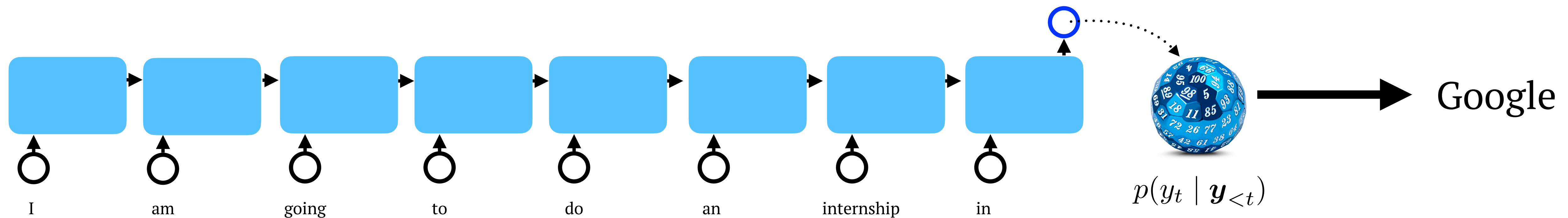
I am going



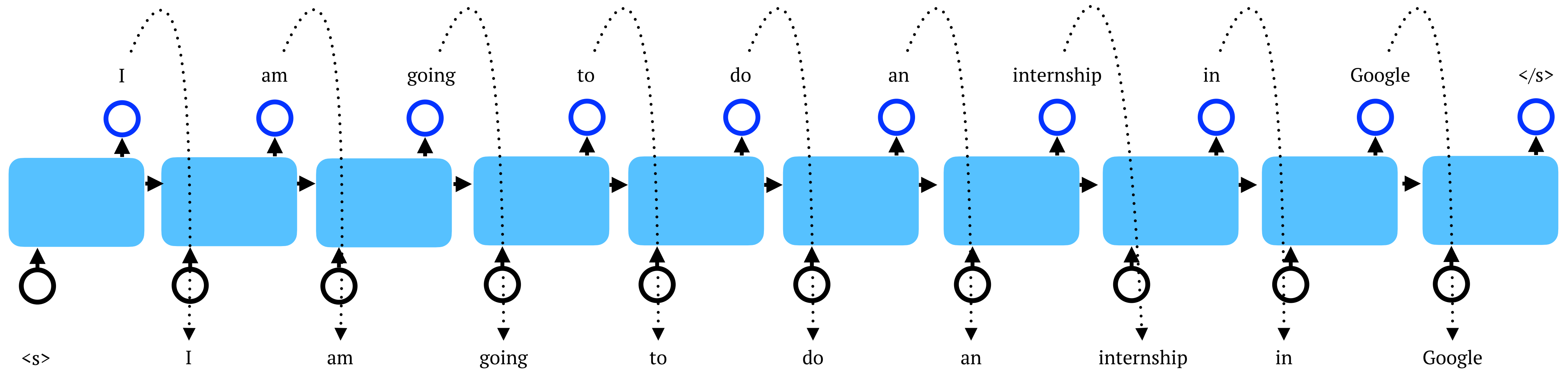


# Sample a sentence from RNNLMs

I am going to do an internship in Google



# RNN as Decoder (RNNLM)



$$p(y_t \mid \mathbf{y}_{<t})$$

# Machine Translation

中秋快樂！

$x$

Happy mid autumn festival !

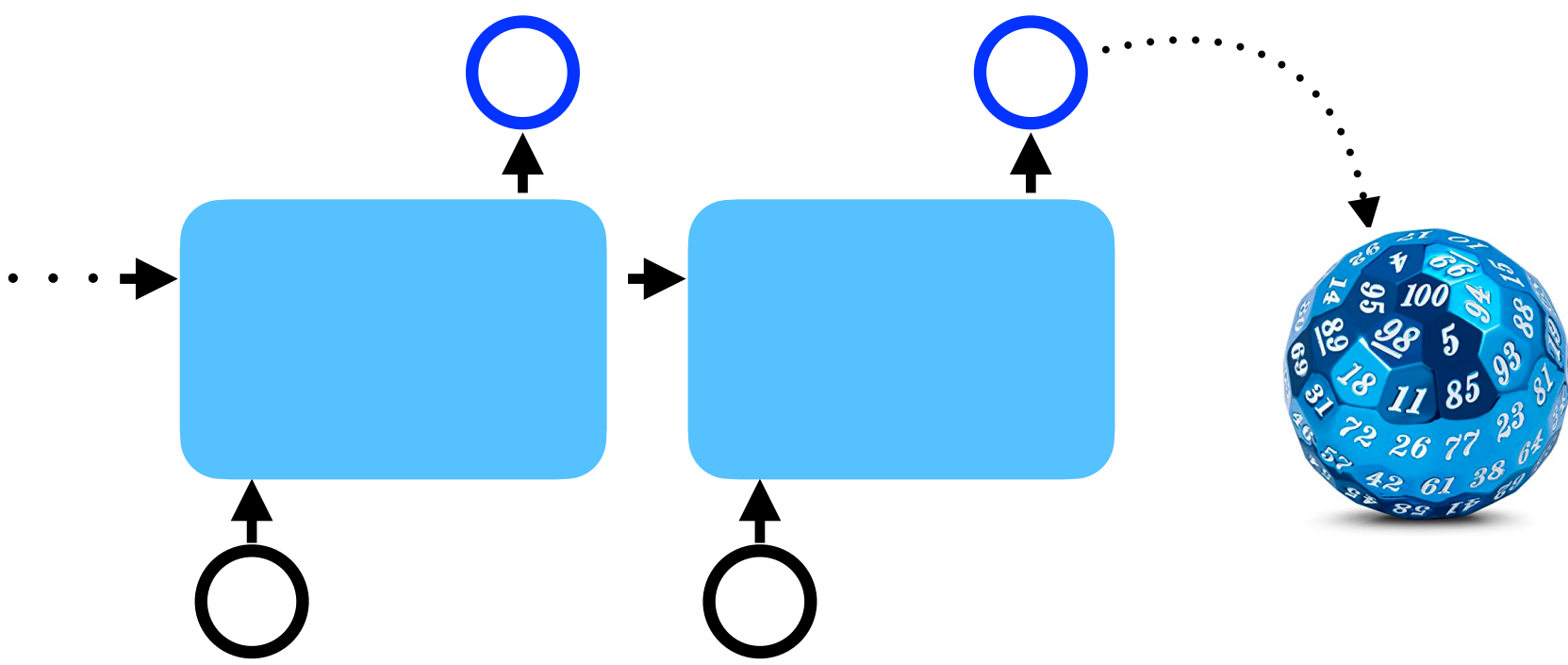
$y$

Happy mid autumn festival !

$$p(\mathbf{y}) = p(y_1 \dots y_n) = \prod_{t=1}^n p(y_t \mid \mathbf{y}_{<t})$$



$p(y_t \mid \mathbf{y}_{<t})$



# Machine Translation

中秋快樂！

$x$

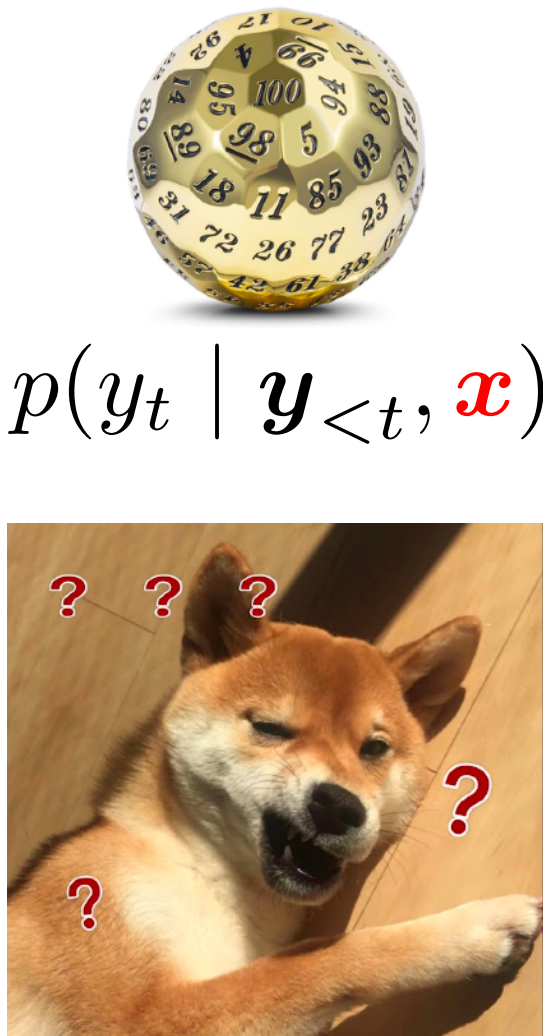
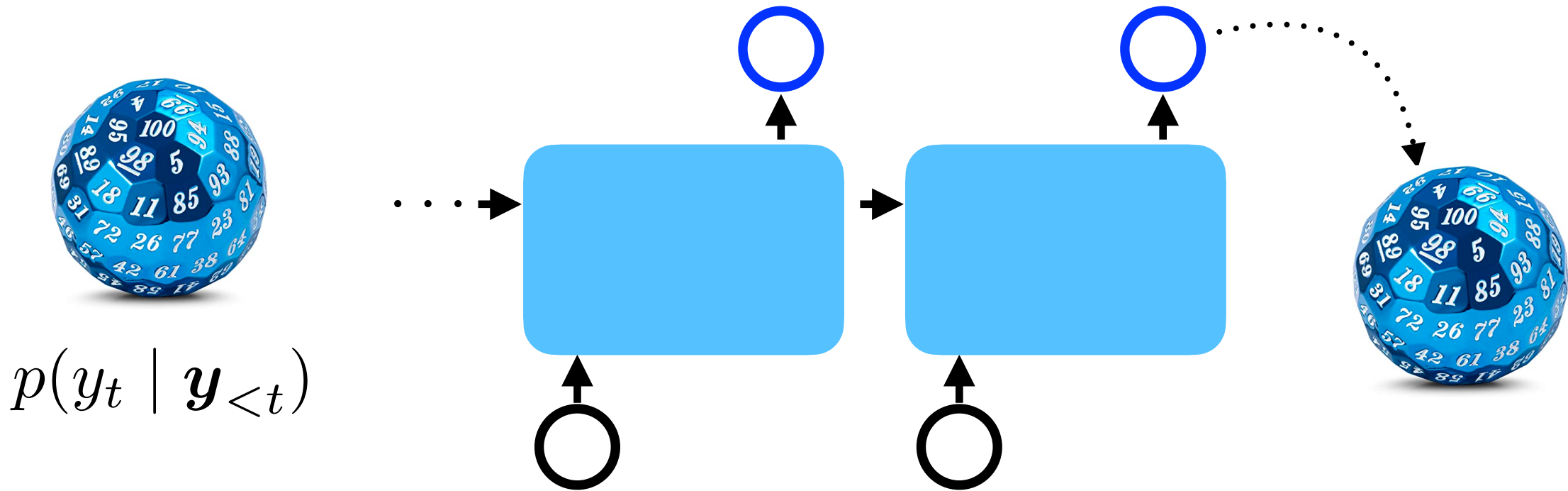
Happy mid autumn festival !

$y$

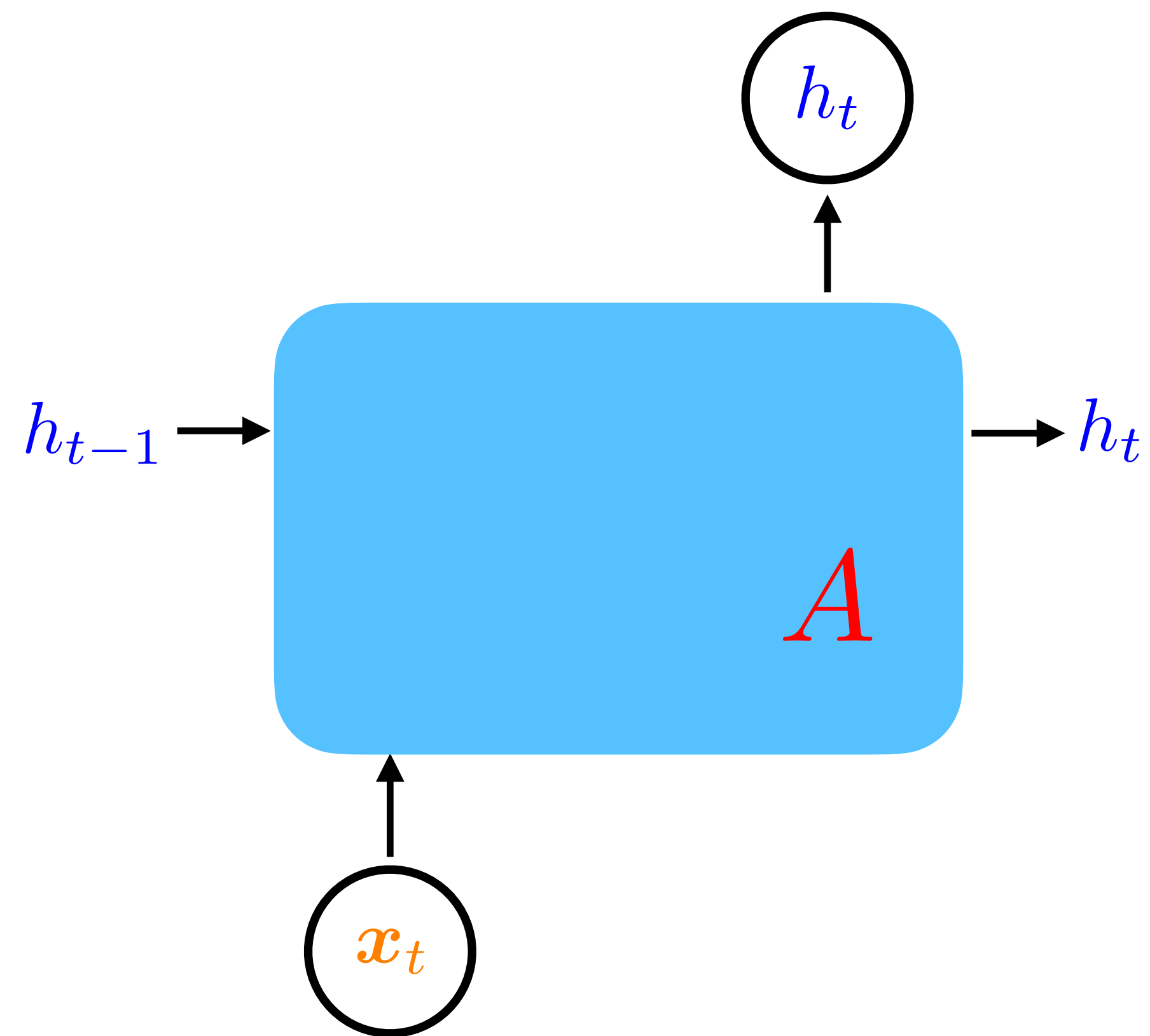
$$p(\mathbf{y} \mid \mathbf{x}) = p(y_1 \dots y_n \mid x_1 \dots x_m) = \prod_{t=1}^n p(y_t \mid \mathbf{y}_{<t}, \mathbf{x})$$

target source

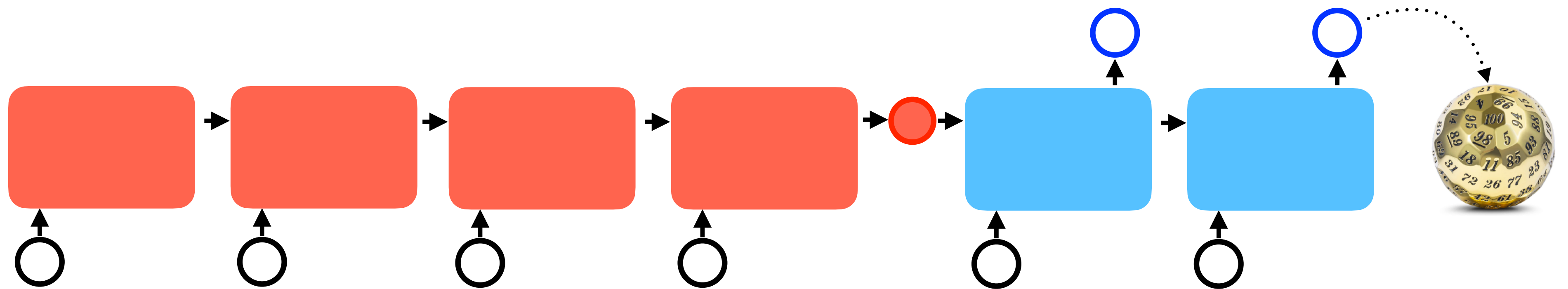
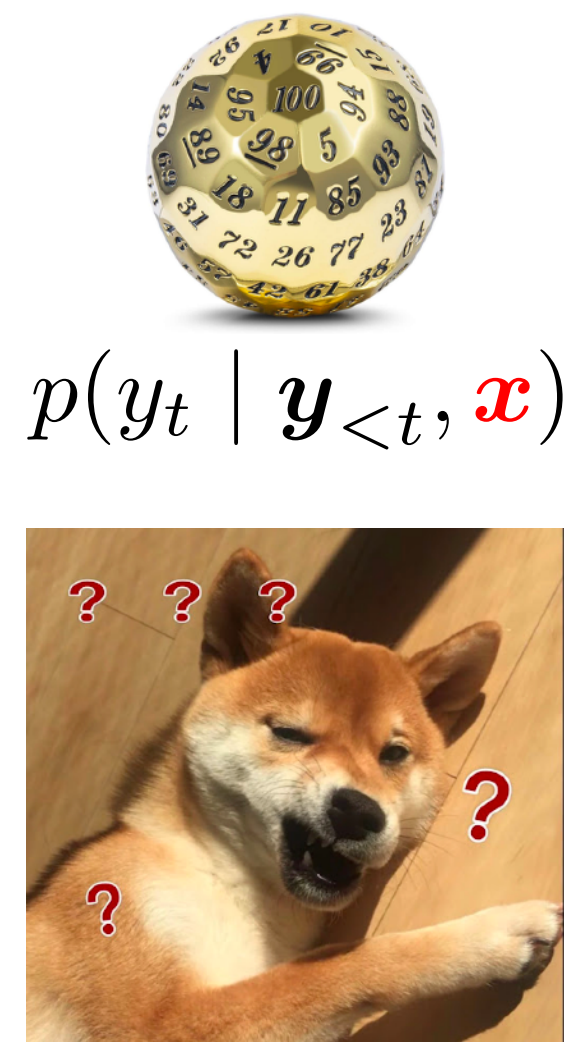
Conditional Language Model



# Recurrent Neural Network

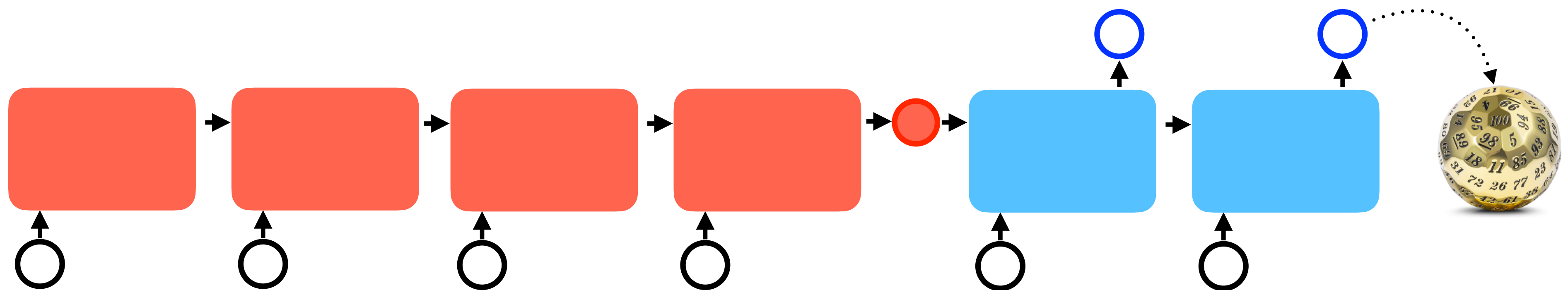
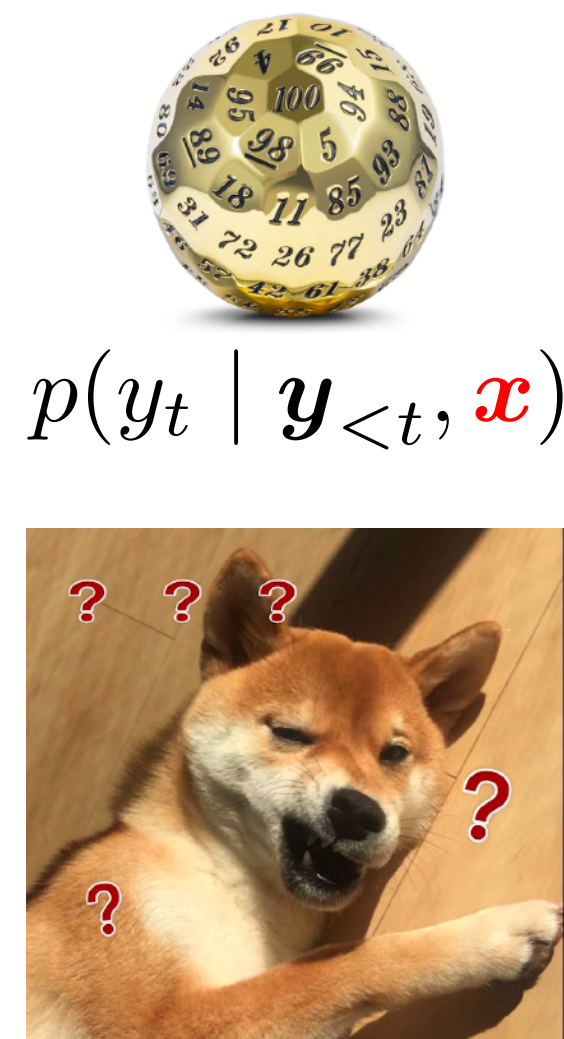


# Encoder + Decoder





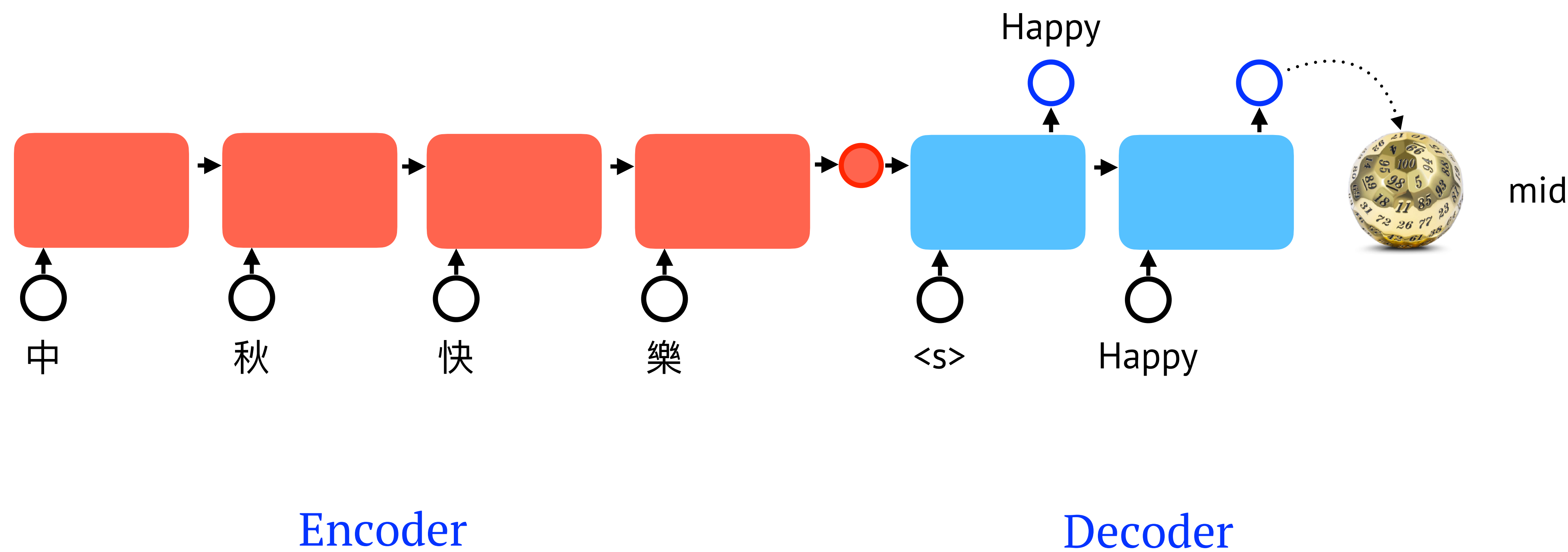
# Sequence to Sequence Model



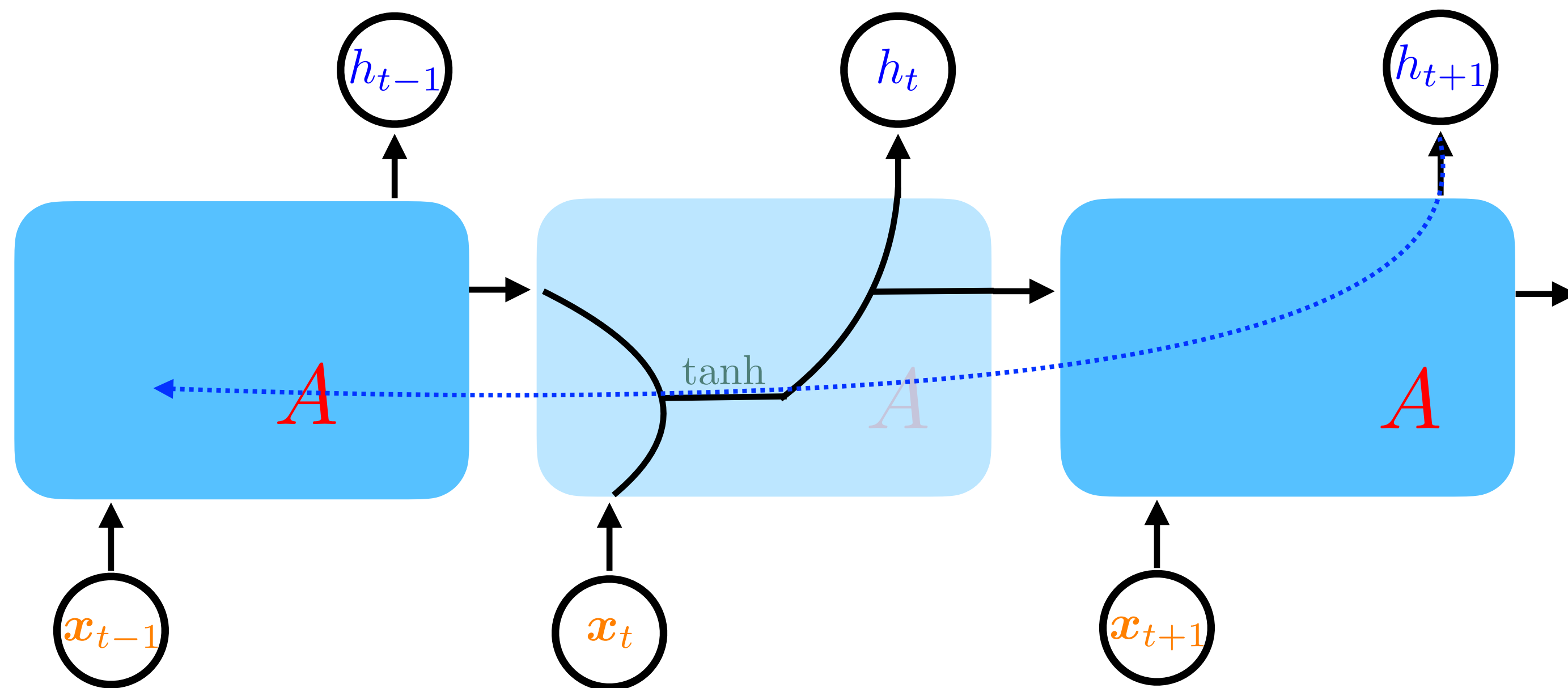
Encoder

Decoder

# Sequence to Sequence Model



# Vanishing Gradient in RNNs



$$h = f(z) \quad \begin{array}{ccc} \textcircled{z} & \xrightarrow{f} & \textcircled{h} \\ \frac{\partial s}{\partial z} = \frac{\partial s}{\partial h} \frac{\partial h}{\partial z} & & \frac{\partial h}{\partial z} \quad \frac{\partial s}{\partial h} \end{array}$$

← Gradient Flow Direction

In general, the longer the path, the smaller the gradient signal.