

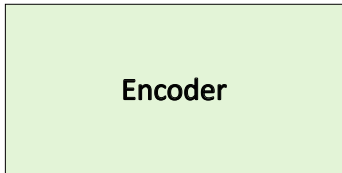
# Transformers (Part 1)

Dr. Alireza Aghamohammadi

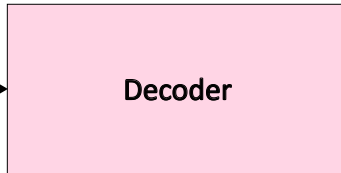
# Encoder-Decoder Architecture

- ▶ The Transformer model utilizes an encoder-decoder architecture.
- ▶ This pattern is widely used in deep learning for tasks such as machine translation, question answering, and image captioning.
- ▶ The encoder processes the input and converts it into a **context representation**, while the decoder uses this context to generate a meaningful output.

The encoder creates a meaningful representation of the source input and provides it to the decoder.



Sarah se promenait dans le parc avec son chat, Alex.  
Elle portait une belle montre en or.



Sarah was walking in the park with her cat, Alex.  
She was wearing a beautiful gold watch.

The decoder then uses this context to generate the target output

## Sequence to Sequence Model

- **Input:**  $x = (x_1, x_2, x_3, \dots, x_m)$
- **Output:**  $y = (y_1, y_2, y_3, \dots, y_n)$

$$\begin{aligned} y' &= \underset{y}{\operatorname{argmax}} P(y|x, \theta) \\ &= \prod_{t=1}^n P(y_t|x, \theta, y_1, y_2, \dots, y_{t-1}) \end{aligned}$$

- At time  $t$ , the model predicts a probability distribution  $P(y_t|x, \theta, y_1, y_2, \dots, y_{t-1})$

previous tokens

Sarah se promenait dans le parc avec son chat, Alex.

Elle portait une belle montre en or.

we want the model  
to predict this