



# Workshop 9

COMP90051 Statistical Machine Learning  
Semester 1, 2023

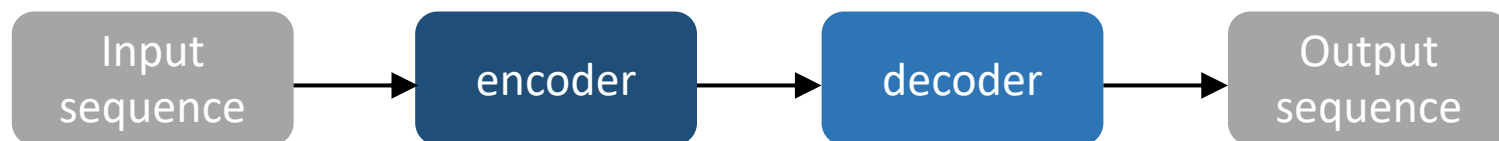
# Learning Outcomes

By the end of this workshop you should be able to:

1. explain **sequence-to-sequence** task, familiar with **encoder-decoder** architecture
2. Be able to implement **Transformer encoder** and **decoder**
3. Know how to train a translation model from scratch using **Transformer**

# Seq2seq task

- **Sequence-to-Sequence** (Seq2seq) is mapping an input sequence to an output sequence.
- It has been widely used in NLP tasks, such as **machine translation**, chatbots, text summarization.
- Encoder-decoder is most popular architecture for **seq2seq** task
  - \* Both can be implemented using RNN, GRU, LSTM, or **Transformer**



# Transformer

- The **Transformer** is a NN architecture introduced by paper “Attention is All You Need”
- Similar to the traditional Seq2Seq model, it also comprises an **encoder** and a **decoder**
- By leveraging a **self-attention** mechanism, model to capture long-term dependencies more effectively than RNN

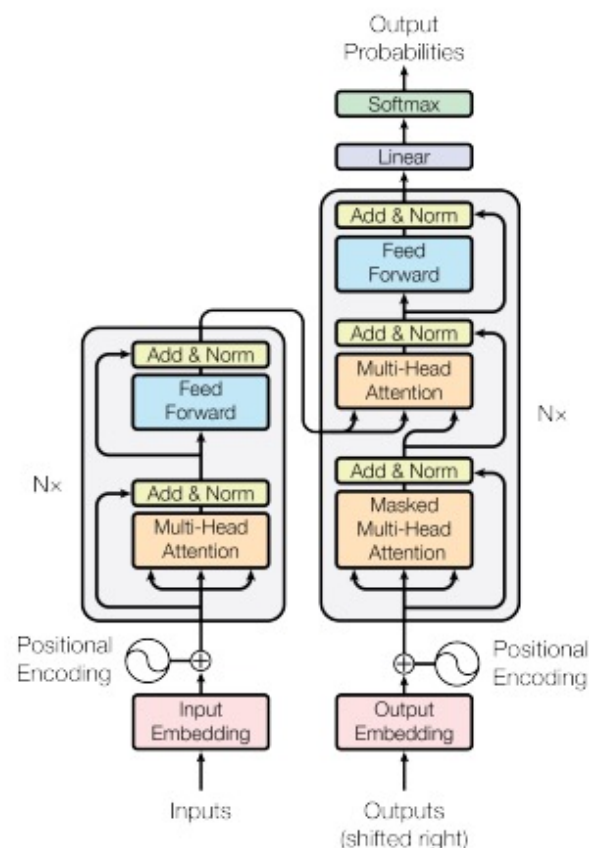


Figure 1: The Transformer - model architecture.

# Worksheet 9