



0. Setup

1. Transformer models

Transformers, why are they so damn cool?

Introduction

Natural Language Processing

Transformers, what can they do?

How do Transformers work?

Encoder models

Decoder models

Sequence-to-sequence models

Bias and limitations

Summary

End-of-chapter quiz

2. Using 🤖 Transformers

3. Fine-tuning a pretrained model

4. Sharing models and tokenizers



Pytorch

TensorFlow

Encoder models

Transformer models: Encoders



Encoder models use only the encoder of a Transformer model. At each stage, the attention layers can access all the words in the initial sentence. These models are often characterized as having “bi-directional” attention, and are often called *auto-encoding models*.

The pretraining of these models usually revolves around somehow corrupting a given sentence (for instance, by masking random words in it) and tasking the model with finding or reconstructing the initial sentence.

Encoder models are best suited for tasks requiring an understanding of the full sentence, such as sentence classification, named entity recognition (and more generally word classification), and extractive question answering.

Representatives of this family of models include:

ALBERT

BERT

DistilBERT

ELECTRA

RoBERTa

Next Section >