



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

Natural Language Processing

Before jumping into Transformer models, let's do a quick overview of what

natural language processing is and why we care about it.

What is NLP?

NLP is a field of linguistics and machine learning focused on understanding everything related to human language. The aim of NLP tasks is not only to understand single words individually, but to be able to understand the context of those words.

The following is a list of common NLP tasks, with some examples of each:

Classifying whole sentences: Getting the sentiment of a review, detecting if an email is spam, determining if a sentence is grammatically correct or whether two sentences are logically related or not

Classifying each word in a sentence: Identifying the grammatical components of a sentence (noun, verb, adjective), or the named entities (person, location, organization)

Generating text content: Completing a prompt with auto-generated text, filling in the blanks in a text with masked words

Extracting an answer from a text: Given a question and a context, extracting the answer to the question based on the information provided in the context

Generating a new sentence from an input text: Translating a text into another language, summarizing a text

NLP isn't limited to written text though. It also tackles complex challenges in speech recognition and computer vision, such as generating a transcript of an audio sample or a description of an image.

Why is it challenging?

Computers don't process information in the same way as humans. For example, when we read the sentence "I am hungry," we can easily understand its meaning. Similarly, given two sentences such as "I am hungry" and "I am sad," we're able to easily determine how similar they are. For machine learning (ML) models, such tasks are more difficult. The text needs to be processed in a way that enables the model to learn from it. And because language is complex, we need to think carefully about how this processing must be done. There has been a lot of research done on how to represent text, and we will look at some methods in the next chapter.

[Next Section >](#)