

LLM Course

LLM Course documentation

Natural Language Processing and Large Language Models

LLM Course

■ View all resources

Agents Course

Audio Course

Community Computer Vision Course

Deep RL Course

Diffusion Course

LLM Course

MCP Course

ML for 3D Course

ML for Games Course

Open-Source AI Cookbook

Robotics Course

a smol course

Search documentation

AR

BN

DE

EN

ES

FA

FR

GJ

HE

HI

ID

IT

JA

KO

MY

NE

PL

PT

RO

RU

RUM

TE

TH

TR

VI

ZH-CN

ZH-TW

Join the Hugging Face community

and get access to the augmented documentation experience

Collaborate on models, datasets and Spaces

Faster examples with accelerated inference

Switch between documentation themes

Sign Up

to get started

Copy page

Natural Language Processing and Large Language Models

Before jumping into Transformer models, let's do a quick overview of what natural language processing is, how large language models have transformed the field, 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.

The Rise of Large Language Models (LLMs)

In recent years, the field of NLP has been revolutionized by Large Language Models (LLMs). These models, which include architectures like GPT (Generative Pre-trained Transformer) and

Llama

, have transformed what's possible in language processing.

A large language model (LLM) is an AI model trained on massive amounts of text data that can understand and generate human-like text, recognize patterns in language, and perform a wide variety of language tasks without task-specific training. They represent a significant advancement in the field of natural language processing (NLP).

LLMs are characterized by:

Scale

: They contain millions, billions, or even hundreds of billions of parameters

General capabilities

: They can perform multiple tasks without task-specific training

In-context learning

: They can learn from examples provided in the prompt

Emergent abilities

: As these models grow in size, they demonstrate capabilities that weren't explicitly programmed or anticipated

The advent of LLMs has shifted the paradigm from building specialized models for specific NLP tasks to using a single, large model that can be prompted or fine-tuned to address a wide range of language tasks. This has made sophisticated language processing more accessible while also introducing new challenges in areas like efficiency, ethics, and deployment.

However, LLMs also have important limitations:

Hallucinations

: They can generate incorrect information confidently

Lack of true understanding

: They lack true understanding of the world and operate purely on statistical patterns

Bias

: They may reproduce biases present in their training data or inputs.

Context windows

: They have limited context windows (though this is improving)

Computational resources

: They require significant computational resources

Why is language processing 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.

Even with the advances in LLMs, many fundamental challenges remain. These include understanding ambiguity, cultural context, sarcasm, and humor. LLMs address these challenges through massive training on diverse datasets, but still often fall short of human-level understanding in many complex scenarios.

Update

on GitHub

←

Introduction

Transformers, what can they do?

→