AI 세션 - LLM

김민재 전효정 최동주

01 목표

논문 분석

LLM 프로젝트

라마 모델 파인튜닝하기

Attention Is All You Need

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Usz

The dominant sequence transduction models are based on con
encoder and decoder through an attention mechanism. We pro
convolutions entirely. Experiments on two machine translation t
model achieves 28.4 BLEU on the WMT 2014 English-to-Germa
French translation task, our model establishes a new single-mo
models from the literature. We show that the Transformer gene

Comments: 15 pages, 5 figures
Subjects: **Computation and Language (cs.CL)**; Machine Learning (cs.L(
Cite as: arXiv:1706.03762 **[cs.CL]**
(or arXiv:1706.03762v7 **[cs.CL]** for this version)
https://doi.org/10.48550/arXiv.1706.03762 ⓘ

Submission history

Attention Is All You Need

A Survey of Large Language Models

Wayne Xin Zhao, Kun Zhou, Junyi Li, Tianyi Tang, Xiaolei
Zhipeng Chen, Jinhao Jiang, Ruiyang Ren, Yifan Li, Xinyu

Language is essentially a complex, intricate system of human ex
grasping a language. As a major approach, language modeling h
models to neural language models. Recently, pre-trained langua
solving various NLP tasks. Since researchers have found that mo
larger size. Interestingly, when the parameter scale exceeds a ce
abilities that are not present in small-scale language models. To
PLMs of significant size. Recently, the research on LLMs has be
widespread attention from society. The technical evolution of LLN
algorithms. In this survey, we review the recent advances of LLN
namely pre-training, adaptation tuning, utilization, and capacity e

A Survey of Large Language Models

Making requests

You can paste the command below into your terminal to run your first API request. Make sure to replace `$OPENAI_API_KEY` with your secret API key. If you are using a legacy user key and you have multiple projects, you will also need to **specify the Project Id**. For improved security, we recommend transitioning to project based keys instead.
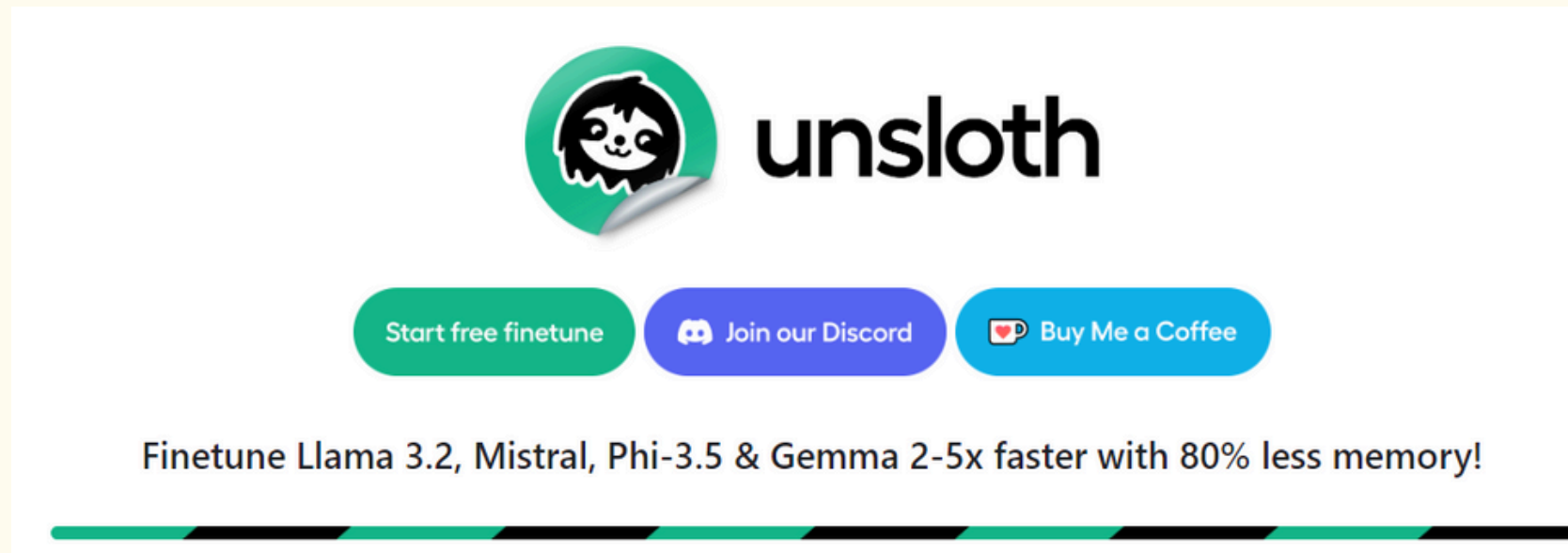
```
1  curl https://api.openai.com/v1/chat/completions \
2    -H "Content-Type: application/json" \
3    -H "Authorization: Bearer $OPENAI_API_KEY" \
4    -d '{
5      "model": "gpt-4o-mini",
6      "messages": [{"role": "user", "content": "Say this is a test!"}],
7      "temperature": 0.7
8    }'
```

This request queries the `gpt-4o-mini` model (which under the hood points to a `gpt-4o-mini` **model variant**) to complete the text starting with a prompt of "*Say this is a test*". You should get a response back that resembles the following:

LLM(GPT, LLAMA 등)을 사용하여 토이 프로젝트 진행

GPT API를 활용하여 기존의 있는 웹사이트를 발전시키기

허깅페이스, 코랩 등을 활용하여 직접 데이터셋을 만들고

파인튜닝하여 나만의 모델 만들어보기

새로운 방법론이 계속 등장하므로 다양한 방식으로 도전해보기

감사합니다