

# مودعیار

الضخم. الامهام متعددة اللغة فهم مدعوي نموذج بتقديم يقوم الامن سور هذا

متعددة أسئلة من يتكون الامور. من واسعة مجموعة عبر متتنوعة مهام أداء على الامن نموذج لقدرة شامل اختبار هو مدعوي ار والطب. القانون، التاريخ، الرياضيات، مثل متتنوعة مجالات تغطي الاريارات

البيانات: مجموعة روابط

البيانات  
المجموعات

```
import torch
from datasets import load_dataset
import requests
import json

# MMLU
subject = "abstract_algebra" #
dataset = load_dataset("cais/mmlu", subject, split="test")

#
def format_mmlu_prompt(example, few_shot_examples=5):
    prompt = "The following are multiple choice questions (with answers) about {}\n\n".format(subject.replace(" ", "_"))
    prompt += f"Question: {example['question']}\n"
    prompt += "Choices:\nA. {}\nB. {}\nC. {}\nD. {}\n".format(*example['choices'])
    prompt += f"Answer: {example['answer']}\n\n"

    #
    prompt += f"Question: {example['question']}\n"
    prompt += "Choices:\nA. {}\nB. {}\nC. {}\nD. {}\n".format(*example['choices'])
    prompt += "Answer:"

    return prompt

#
correct = 0
total = 0
```

```

for example in dataset:
    prompt = format_mmlu_prompt(example)

    #           llama-server
    url = "http://localhost:8080/v1/chat/completions"
    headers = {"Content-Type": "application/json"}
    data = {
        "messages": [{"role": "user", "content": prompt}],
        "max_tokens": 5,
        "temperature": 0,
    }

    response = requests.post(url, headers=headers, data=json.dumps(data))

    if response.status_code == 200:
        output_text = response.json()["choices"][0]["message"]["content"]
        predicted_answer = output_text.strip()[0] if len(output_text.strip()) > 0 else ""
    else:
        predicted_answer = ""
        print(f"Error: {response.status_code} - {response.text}")

    #

    if predicted_answer.upper() == example["answer"]:
        correct += 1
    total += 1

    #

accuracy = correct / total
print(f"Subject: {subject}")
print(f"Accuracy: {accuracy:.2%} ({correct}/{total})")

```