

# معار

الضخم. المهام متعددة اللغات فهم معار على لغوي نموذج بتقييم يقوم المنشور هذا

متعددة أسئلة من يتكون الموضوعات. من واسعة مجموعة عبر متنوعة مهام أداء على النموذج لقدرة شامل اختبار هو معار والطلب. القانون، التاريخ، الرياضيات، مثل متنوعة مجالات تغطي الاختبارات

البيانات: مجموعة روابط

```

#
#
```

```
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(" ", "_"))

    #
    few_shot_dataset = load_dataset("cais/mmlu", subject, split="validation")
    for i in range(few_shot_examples):
        ex = few_shot_dataset[i]
        prompt += f"Question: {ex['question']}\\n"
        prompt += "Choices:\\nA. {}\\nB. {}\\nC. {}\\nD. {}\\n".format(*ex['choices'])
        prompt += f"Answer: {ex['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}")

```