?????????????? ????? python

Запрос - ответ в существующую модель

```
import ollama
import requests

def chat_with_deepseek(prompt, model="deepseek-r1:7b"):
    response = ollama.chat(
        model=model,
        messages=[{"role": "user", "content": prompt}]
    )
    return response["message"]["content"]

def chat_with_deepseek_api(prompt, model="deepseek-r1:7b"):
    url = "http://localhost:11434/api/chat"
    data = {
        "model": model,
        "messages": [{"role": "user", "content": prompt}],
        "stream": False  # Отключить потоковый вывод
    }
    response_api = requests.post(url, json=data)
    return response_api.json()["message"]["content"]

# Пример использования
user_input = "Объясни, что такое ООП простыми словами."
#response = chat_with_deepseek(user_input)
response = chat_with_deepseek_api(user_input)
print("Ответ DeepSeek:", response)
```

Источник RAG

Дообучение модели на собственных данных

Дополнительные пакеты

```
python -m pip install ollama llama-index transformers torch sentence-transformers llama-index-
llms-ollama
```

Здесь должен был быть код, успешный результат (хотя бы результатик), но... Все уперлось в токенизацию. Напрямую 100 страничный файл оказался бессмысленным, узлов после 30 минутной обработки было создано 0. Причем различные варианты не помогли - простое предоставление файла в свободном форматировании оказалось бессмысленным занятием. Нужно погружаться как минимум в теорию Chunk'ов. Хотя скорее всего потребуется еще много чего.

Revision #4
Created 5 April 2025 13:09:44 by Admin
Updated 19 July 2025 07:50:46 by Admin