å¡¾é?·AI

Description

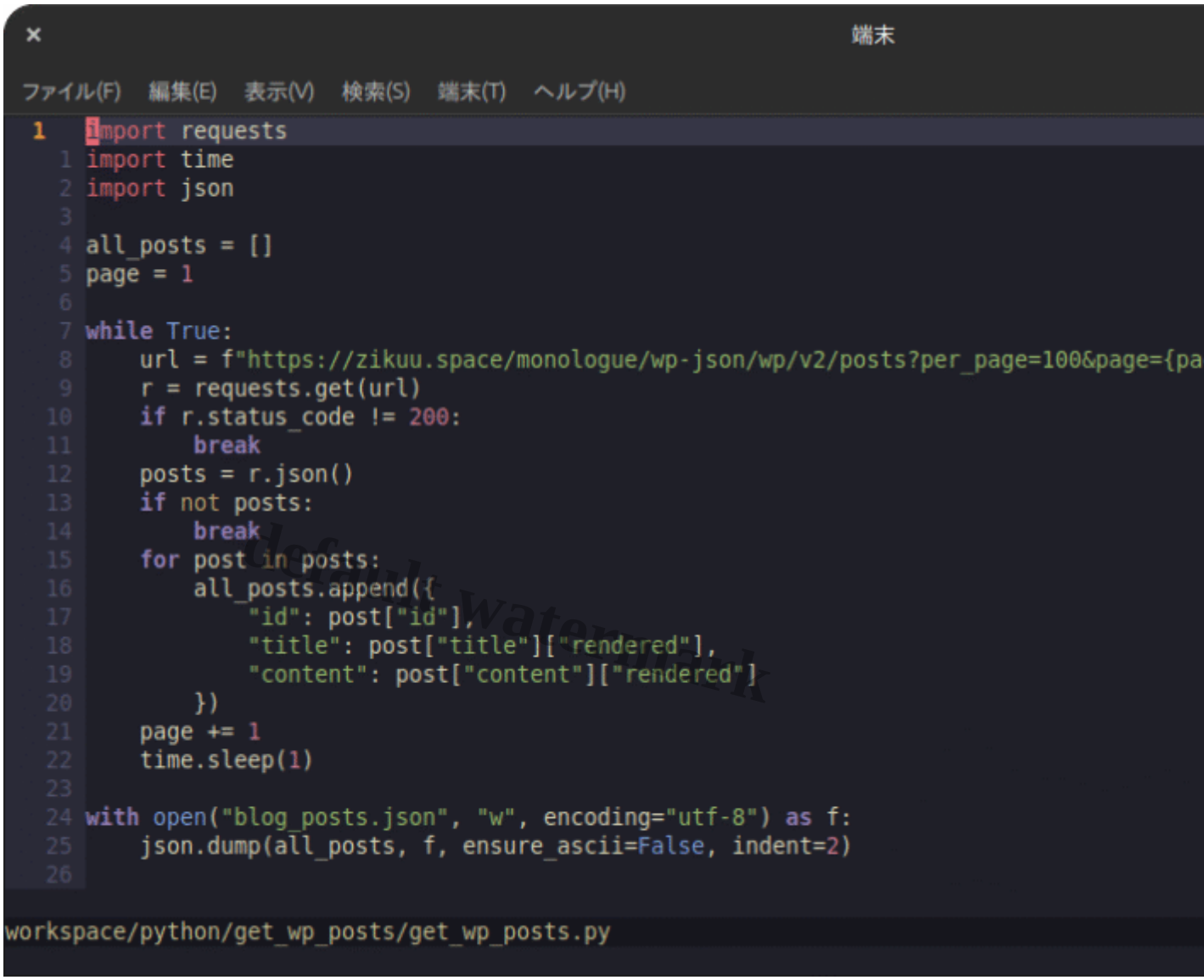
default watermark

è?ªå??ã•®å·¥æ?¿ã•®ã??ã?ã?°ã??ã•?ã•®ã??ã?ã?°ã??å•«ã?•ã•¦ã?•ç§•ã•®ã?¢ã??ã•¥ã••ã??ã•®æ´å•²ã•¯ã»ã•¼å®?å ¨ã••è¨?é?²ã•?ã??ã•¦ã•?ã•¾ã•?ã??ã•?ã??ã??ã•®è¨?é?²ã•«ã•¯ã?•ç§•ã•ªã??ã•§ã•¯ã•®è¦?ç?¹ã??æ?•è??ã??å•«ã•¾ã??ã•¦ã•?ã•¾ã•?ã??ã•?ã•?ã??ã?•ã•?ã??ã??ã??AIã•«å¦ç¿?ã•?ã•?ã?•ç§•ã•®ä»£ã?•ã??ã•«æ??ç« ã??ç??æ?•ã•?ã•¦ä¼?è©±ã•?ã•§ã••ã??ã??ã•ªAIã??ã?•å¡¾ã•®ä¸ã•§ç¨¼å?•ã•?ã•?ã•?ã•¨è??ã•?ã•¦ã•?ã•¾ã•?ã??ã•?ã??ã•¯ã?•ç§•ã•®èª?ã??å•£ã??è??ã•?æ?¹ã?•ã••ã•?ã•¦å®?é??ã•«è¡?ã•£ã•¦ã••ã•?ã•?ã•¨ã??ç¶?æ?¿ã•?ã??ä¸?ã•¤ã•®æ??æ®µã•§ã•?ã??ã•¨å•?æ??ã•«ã?•AIã•®å®?è·µç??ãªæ´»ç?¨ã??è©¦ã•¿ã??å®?é¨?ã•§ã??ã•?ã??ã•¾ã•?ã??

ã•¾ã•?ã?•ã•?ã••ã•«å®?ç•¾ã•§ã••ã??ã?•ã•?ã•§ã•¯ã?ã??ã•¾ã•?ã??ã•?ã?•ç§•ã•«ã•§ã••ã??ç¨?åº¦ã•®ã•?ã•¨ãªã??ã?•AIã??ã?ã??ã??ã??ã•«ä»»ã•?ã•¦ã•?ã•¾ã•?ã•®ã??æ?ªã••ãªã•?ã•¨æ?•ã•£ã•¦ã•?ã•¾ã•?ã??ã•?ã•¤ã•¦ã??è?ªå??ã•®ä»?äº?ã??ç?¡ã••ã•?ã?•ã•¨ã•?ã•?æ?•è??å®?é¨?ç??ãªè¨?äº?ã??æ?¸ã•?ã•?ã•?ã•¨ã??ã?ã??ã•¾ã•?ã•?ã?•è?ªå??ã•®ä»?äº?ã••å?ºå?·ã•?ã•?ã?•ä»?ã•®æ??æ®µã•§å®?ç•¾ã•§ã••ã?ªã??ã••ã•¡ã??ã••ã»»ã•?ã?•è?ªå??ã•¯æ?°ã?ãªæ??æ?¦ã•¸ã•¨é?²ã??ã•®ã•?è?¯ã?ã•¨è??ã•?ã•¦ã•?ã•¾ã•?ã??

ã••ã•?ã•§è©¦ã•?ã••ã??ã?ã?°è¨?äº?ã??ä¸?æ?¬ã??ã?¦ã?³ã?ã?¼ã??ã•?ã??ã?¹ã?¯ã?ªã??ã??ã??æ?¸ã?ã?¦ã•¿ã•¾ã•?ã•?ã??

ã•?ã•®ã?¹ã?¯ã?ªã??ã??ã§ã•?ã•®ã??ã?ã?°ã•®å ¨è¨?äº?ã??ã??ã?¦ã?³ã?ã?¼ã??ã•§ã••ã•¾ã•?ã??æ??è¿?ã•¯ã?ªã•?ã•?ã•?ã•?ã??ã?ã?°ã?©ã? ã•?æ?¸ã••ã••ã??ã•?ã••ã•ªã•£ã•¦æ??ã??é?£ã•?ã•§ã•?ã??



ã??ã?¦ã?³ã?ã?¼ã??ã•?ã•?è¨?äº?ã•®ä¸?é?¨ã??

ã•?ã??ã??QLoRAã??ã?¡ã?¤ã?³ã?•ã?¥ã?¼ã??ã?³ã?°ç?¨ã•®ã??ã?¼ã?¿ã?»ã??ã??ï¼?ã?¤ã?³ã?¹ã??ã?©ã?¯ã?·ã?§ã?³ã?»ã??ã?¼ã?¿ã?»ã??ã??ï¼?ã«å¤?æ•?ã•?ã??ã?¹ã?¯ã?ªã??ã??ã¯ã?•ã?ã•?ã•?æ??ã•?ã•§ã•?ã??

端末

ファイル(F) 編集(E) 表示(V) 検索(S) 端末(T) ヘルプ(H)

```python
from bs4 import BeautifulSoup
import json

with open("blog_posts.json", encoding="utf-8") as f:
    posts = json.load(f)

dataset = []
for post in posts:
    title = post["title"]
    content_html = post["content"]
    content_text = BeautifulSoup(content_html, "html.parser").get_text()
    dataset.append({
        "instruction": f"次のテーマで文章を書いてください：「{title}」",
        "output": content_text.strip()
    })

with open("qlora_dataset.json", "w", encoding="utf-8") as f:
    json.dump(dataset, f, ensure_ascii=False, indent=2)
```

def ... watermark

workspace/python/get_wp_posts/create_dataset.py

ã??ã?¼ã?¿ã?»ã??ã??ã•®ä¸?é?¨ã•§ã•?ã??

ã•?ã•®ã??ã?¼ã?¿ã?»ã??ã??ã??ä½¿ã•£ã•¦LLama3ã??Gemma3ã•ªã•©ã•®LLMã??ã??ã?¡ã?¤ã?•ã?¥ã?¼ã??ã?³ã?°ã•?ã??å¦ç¿?ã?³ã?¼ã??ã??ã••ã•?è¦?å?´ã•?ã•?ã•«æ?¸ã•?ã•¾ã•?ã??

å?•é¡?ã•¯ã?ã?¼ã??ã?¦ã?§ã?¢ã??å¡¾ã•®PCã•§ä¸?ç?ªå¼·ã•?ã•®ã•¯RTX A4000 (VRAM 16G)æ•è¼?ã•®PCã•§ã•?ã•?ã?•ã•?ã??ã• ã•¨ã•?ã•¾ã??å¤§ã••ã•ªLLMã•«å¦ç¿?ã•?ã•?ã??ã??ã•¾ã•?ã??ã??è©¦ã•?ã•«å°•ã•?ã•ªã?¢ã??ã?«ã•§ã•¨ã•¯æ?ã•£ã•¦ã•?ã•¾ã•?ã•?ã?•çµ•æ??ã•«ã•?ã•£ã•?ã??ã•?ã••ã•?ã•§æ°?ã•?é?²ã•¿ã•¾ã•?ã??ã??ã•¾ã•?ã?•ã••ã??ã•ªã•«é• ã••ã•ªã•?å°?æ•¥ã?•å®?ä¾¡ã•«ã??ã•£ã•¨å¼·ã•?PCã??å ¥æ??ã•§ã••ã??ã??ã•?ã•«ã•ªã??ã•§ã•?ã??ã•?ã•?ã?•ã•?ã•?ã•?ã•?ã•?ã•¨ã•®ã??ã??æ?¹ã??ç?¥ã•£ã•¦ã•?ã??ã• ã•?ã•§ã??ä¾¡å?¤ã•?ã•?ã??ã•¾ã•?ã??

Category

1. æ?¥ã? ã•®æ´»å??
2. ç?¬ã??è¨?

Tags

1. AI
2. LLM
3. QLoRA
4. ã??ã?¼ã?¿ã?»ã??ã??
5. ã??ã?¡ã?¤ã?³ã?•ã?¥ã?¼ã??ã?³ã?°
6. ä¸?é??å??å¸?
7. å¤§è¦•æ¨¡è¨?èª?ã?¢ã??ã?«
8. å±±æ¢¨ç??

Date Created
2025å¹´6æ??26æ?¥
Author
kazuo-tsubaki

default watermark