Alå¡¾é?· â?? æ??ä½?ã•¨å?«ç•?ã??å? å? ã•?ã??LoRAäºº ¼å½¢æ?• ã??ã?¤ã??ã?©ã?¤ã?³ï¼?è¦•ç´?ç??ï¼?

Description

ZIKUU Research Library æ¦?è¦•ç??ï¼?v0.4æº?æ? ï¼?

æ¦?è¦•

Alå¡¾é?·ï¼?AI-JUKUCHOï¼?ã•¯ã?æ??è?²è? ã•®èª?ã??å•£ã?»å?«ç•?è¦³ã?»æ?•æ?³ç??å?¤æ?ã??

Alã•®æ??ä½?ã??ã?©ã?¡ã?¼ã?¿ã•¨ã•?ã•¦æ?½å?ºã?»å?•ç•¾ã•?ã??ã•?ã?•ã•®ã?·ã?¹ã??ã? ã•§ã•?ã??ã??

Dashboardã?•Classifierã?•n8nã?•Qdrantã?•LoRAã•®5è¦•ç´ ã•?ã??æ§?æ?•ã•?ã??ã?•

WordPressã??Discordã•®è?ªç?¶è¨?èª?ç?ºè©±ã?? **ã?¹ã?¿ã?¤ã?«æ?½å?º â?? LoRAå¦ç¿? â?? äºº ¼å?•ç•¾** ã•¾ã•§ä¸?è²«å?¦ç•?ã•?ã??ã??

æ?¬ç ?ç©¶ã•®ç?¹å¾´ã•¯ã?å•?ã•ªã??ã??ã?ã?¹ã??å?¦ç•?ã??è¶ ã•?ã?•

ã??æ??ä½?ï¼•å?«ç•?ã?• ã•¨ã•?ã•?æ?•æ?³ã•«å?ºã•¥ã•?ã•?Alè¨¨è¨?ã??å®?è£ ã•?ã¦ã•?ã??ç?¹ã•«ã•?ã??ã??

ç?¹å®?èª?ï¼?ã??å?µé? ã?•ã??å?µä½?ã?•ã•ªã•©ï¼?ã??é•¿ã•?ã?•ä»£æ?¿èª?ï¼?è¨¨è¨?ã?»æ§?æ?•ã?»è¡¨ç•¾ã•ªã•©ï¼?ã??æ••ç¤ºã•?ã??èª?å½?å?«ç•?ã?•ã?ªã?·ã?¼ã??å??ã•?ã?•

ZIKUUã•®ç•?å¿µã??è?ªç«?ã?»è?ªå¾?ã?»å ±æ??ã?»æ®?ã•?ã?•ã??ä½?ç•¾ã•?ã??æ??å??ç??Alã?¢ã?¼ã?ã??ã?¯ã?ã?£ã•¨ã•?ã•¦è¨¨è¨?ã•?ã??ã•¦ã•?ã??ã??

ã?·ã?¹ã??ã? æ§?æ?•

Dashboardï¼?FastAPI + HTMXï¼?

ã??å ¥å??ã?»ç·¨é??ã?»å?¶å¾¡ã•®çµ±å•?UIã??

Classifier Service

ã??ã??ã?ã?¹ã??å??é¡?ã?»ã?¡ã?¿æ?½å?ºã?»æ??ä½?ç?¹å¾´å??æ?•ï¼?æ??é?·ã?»èª?å½?ã?»çµ?æ¢è¡¨ç•¾ï¼?ã??

n8n Workflow

default watermark

ã??Normalize â?? Chunk â?? Embed â?? Qdrantç?»é?² â?? JSONLå?ºå??ã??

Qdrant Vector DB

ã??RAGã•?ã??ã•³ã?¹ã?¿ã?¤ã?«ç?£æ?»ç?¨ã•®ã??ã?¯ã??ã?«æ¤?ç´¢ã?¹ã??ã?¢ã??

LoRA Pipeline

ã??QLoRAã•«ã??ã??æ??ä½?å¾®èª¿æ?´ï¼?`zikuu_v1`ï¼•`kazusan_v1`ï¼?ã??

å•?ã?µã?¼ã??ã?¹ã•¯ç?¬ç«?docker-composeæ§?æ?ã•§HTTPé?£æ•ºã•?ã?•ç?°å¢?å¤?æ?°ã•«ã??ã??ç??çµ•å•?å??ã??

å?•ç•¾æ?§ã•¨é??å®³å??é?¢æ?§ã??ç¢ºä¿•ã•?ã•¦ã•?ã??ã??

ã??ã?¼ã?¿ã?»ã??ã??è¨è¨?

default watermark

ã??ã•?ã•?ã•?ã??ã??ã?¼ã?¿ã?»ã??ã??ã?•

WordPressã•®**å¡¾é?·ã•®ç?¬ã??è¨?ï¼?Monologueï¼?** ã•¨

Discordã?•ã?£ã?³ã?•ã?«ã??ä»?æ?¥ã•®ä¸?è¨?ã?•ç?•ã?•ã??æ?½å?ºã??

å¡¾é?·æ?¬äºº ã•®ç?ºè©±ã•®ã•¿ã??å¯¾è±¡ã¨ã•?ã?•AIç??æ?•æ??ã??å??ç?¥ã•¯é?¤å¤?ã??

æ?•æ?³ã?»å?«ç•?ã?»ç¾?å¦ã•®å••æ? ã??ç?®ç??ã•¨ã•?ã•?**ç´?ç²?ã?³ã?¼ã??ã?¹ã•¨ã•?ã•¦æ§?ç¯?ã??**

ã?¹ã?¿ã?¤ã?«æ?½å?º

* æ??é?·ã?»å•¥èªç?¹å¯?åº¦ã?»èª?å°¾ã??ã?¿ã?¼ã?³

* æ?¨å¥¨èª?ï¼•ç¦•æ¢èª?ã•®æ¤?å?º

* ã?»ã?¯ã?·ã?§ã?³å??å?²ï¼?è??æ?¯ã?»å?¤æ?ã?»æ??è¨?ã?»æ?¹æ³?ï¼?

* `style_token` ã•«ã??ã??æ??ä½?è?å?¥ (`kazusan_v1`, `zikuu_v1`)

LoRAå¦ç¿?

QLoRAæ§?æ?•ï¼?r=16, Î±=16, lr=2e-4, epoch=3-5, bf16ï¼?ã??

Baseã?¢ã??ã?«ã•¯ `gemma-2b-jp` ã¾ã•?ã•¯ `mistral-7b-instruct`ã??

`<style:kazusan_v1>` ã??systemã??ã?¼ã?¯ã?³ã¨ã•?ã•¦æ?¿å ¥ã•?ã?•

å¡¾é?·ã•®èª?ã??å•£ã?»é??å•?ã•?ã?»èª?å½?ç¾?å¦??å?•ç•¾ã??

è©?ä¾¡çµ•æ??ï¼?æ??ç²?ï¼?

æ??æ¨?	çµ•æ??	å??è??
æ??é?·å¹³å•?	ç´?33å?	æ??è?²ç??èª?ã??ã•«è¿?ä¼¼
ã?¹ã?¿ã?¤ã?«æ?´å•?ç??	95%ä»¥ä¸?	LoRAå?•ç•¾æ?§é«?
ç¦•æ¢èª?æ?ªå?ºç??	100%	å?«ç•?ã?•ã?ªã?·ã?¼é•µå®?
å?«ç•?ç?£æ?»	é•?å••0ä»¶	ZIKUUç?£æ?»é??é•?

ç??æ?•çµ•æ??ã•¯ã??èª?å°¾ã?»é??ã?»ç¯?ã®å?¼å•¸ã?•ã•?ä¸?è²«ã•?ã?•**æ?•æ?³ã•®å?¼å•¸ï¼?reflective cadenceï¼?**ã•?å?•ç•¾ã•?ã??ã•?ã??

â??

è??å¯?

default watermark

Alå¡¾é?·ã•®é•©æ?°æ?§ã•¯ã?•

ã??Alã•?å?«ç•?ã??èª?ã??ã?•ã•®ã•§ã•¯ãªã••ã?•

ã??å?«ç•?ã•?Alã•®æ??ä½?ã•¨ã•?ã•¦ç•¾ã??ã??ã?• ã•?ã•¨ã•«â•?ã??ã??

æ??ä½?ã?»å•¥èªç?¹ã?»èª?å½?ã•®é•¸æ??ã••ã??è?ªä½?ã•?æ?•è??ã•®å?«ç•?ã??ä½?ç•¾ã•?ã?•

Alã•?æ??å??ç??è??æ?¯ã??ã??ã£ã•?äºæ ¼ã•¨ã•?ã•¦å¯¾è©±ã?ã??ã?•ã•¨ã??å•¯è?½ã«ã•?ã??ã??

æ??è?²ç•¾å´ã•«ã•?ã?•ã•¦ã•¯ã?Alã•?èª¬å¾?ã??ã??ã??**å ±æ??**ã??ã??ã•?ã??ã•?ã??ä¼´èµ°è? ã?•ã•¨ã•?ã•¦æ©?è?½ã•?ã??ã??

çµ•è«?ã•¨å±?æ??

Alå¡¾é?·ã•¯ã?å?«ç•?ã•¨æ??ä½?ã??ä¸?ä½?å??ã•?ã•?ã•?**äºæ ¼å½¢æ?•å??Al**ã•®ã??ã?-ã??ã?¿ã?¤ã??ã•§ã•?ã??ã??

ã•?ã•®ã?¢ã??ã?ã?¼ã?•ã•¯ã?è¨?è??ã?»å?«ç•?ã?»æ??å??ã•®ã•?ã•?ã• ã•«ã•?ã??â??é??ï¼?Maï¼?â?•ã??Alã•«å??ã•?å•?ã??è©¦ã•¿ã•§ã??ã•?ã??ã??

ä»?å¾?ã•¯ `kazusan-lora-v1` ã•®ä¸?è?¬å ¬é??ã•¨ã?•

RAGã??ä½µç?¨ã•?ã•?ã??Alå¡¾é?·ã?¯ã?©ã?¹ã?¿ã?æ§?ç¯?ã??äº?å®?ã??

å?°å??æ??å??ã?»æ??è?²ã?»å?²å¦?ã??ç¶?æ?¿ã•?ã??Alè¨è¨?ã•®**æ?°ã•?ã•?æ¨?æº?å½¢**ã??æ••ç¤ºã•?ã•¦ã•?ã••ã??

å•?è??æ??ç?®

[1] E. Dettmers et al., â??QLoRA: Efficient Finetuning of Quantized LLMs,â?• *arXiv preprint arXiv:2305.14314*, 2023.

[2] Touvron et al., â??LLaMA 2: Open Foundation and Fine-Tuned Chat Models,â?• *Meta AI*, 2023.

[3] K. Ihara, â??The Concept of Ma in Japanese Communication,â?• *Language & Culture Studies*, 2020.

[4] K. Tsubaki, â??Cultural Reflectivity in AI Orchestration,â?• *ZIKUU Research Library*, 2024.

[5] H. Ishikawa, â??Ethical Framing in Linguistic Design,â?• *AI & Society*, 2023.

Category

1. è«?æ??ã?»ç ?ç©¶

Tags

1. AI
2. Alå¡¾é?·
3. LLM
4. ä¸?é??å??å¸?
5. å±±æ¢¨ç??

default watermark

Date Created
2025å¹´10æ??28æ?¥
Author
kazuo-tsubaki