

Normalize â?? Chunk â?? Embed â?? Qdrantç?»é?² â?? JSONLå?°å??å??

Qdrant Vector DB

â??RAGâ?·â??â?·â?¹â?¿â?·â?«ç?£æ?»ç?··â?·â??â?·â??â?«æ?ç?£â?¹â??â?£â??

LoRA Pipeline

â??QLoRAâ?·«â??â??æ??â½?â¾®èª¿æ?·î¼?`zikuu\_v1`î¼?`kazusan\_v1`î¼?â??

å?·â?µâ?¼â??â?¹â?¯ç«?docker-composeæ\$?æ?·â?\$HTTPé?£æ?°â?·â?·ç?°å£?å?æ?°â?·
«â??â??ç??çµ·å?·å??â??

å?·ç¾æ?·\$â?·é??å®³å??é?£æ?·\$â??ç£ª¿·â?·â?·â?·â??â??

â??â?¼â?¿â?»â??â??è`è`?

â??â?·â?·â?·â??â??â?¼â?¿â?»â??â??â?·

WordPressâ?·â¾é?·â?·ç?·-â??è`î¼?Monologueî¼? â?·

Discordâ?·â?£â?³â?·â?»â??â?»æ?·¥â?·â?·è`â?·ç?·â?·â??æ?¼å?·â??

â¾é?·æ?·-â°â?·â?·è©±â?·â?¿â?¾è±¿â?·â?·Alç??æ?·æ??â??å??ç?¥â?·é?·å?æ??

æ?·æ?³â?»å?»ç?·â?»ç¾å?·â?·â?·æ?·â??ç?·â?·â?·ç?·ç²?â?³â?¼â??â?¹â?·â?·â?·æ\$?ç?·â??

â?¹â?¿â?·â?»æ?¼å?°

* æ??é?·â?»å?¥èªç?¹â?¯â?·â?»èªå¾å??â?¿â?¼â?³

* æ?·å?¥èªî¼?ç?·æ£èªâ?·â?æ?å?°

* â?»â?·â?·â?·\$â?³å??å?î¼?è??æ?·â?»å?æ?â?»æ??è`â?»æ?¹æ³?î¼?

* `style\_token` â?·«â??â??æ??â½?è?å?¥ (`kazusan\_v1`,`zikuu\_v1`)

LoRAå¿ç?

QLoRAæ\$?æ?·î¼?r=16, î±=16, lr=2e-4, epoch=3-5, bf16î¼?â??

Baseâ?·£â??â?»â?·`gemma-2b-jp` â¾â?·â?·`mistral-7b-instruct`â??

`<style:kazusan\_v1>` â??systemâ??â?¼â?·â?³â?·â?·æ?¿å ¥â?·â?·

â¾é?·â?·èªâ??å?£â?»é??å?·â?·â?»èªå½?ç¾å?·â?·å?·ç¾â??

è©?ä¼¡çµ•æ??ï¼?æ??ç²?ï¼?

æ??æ??	çµ•æ??	å??è??
æ??é?·å¹³å?·	ç?33å?	æ??è²ç??èª?ã??ã·«è¿?ä¼¼¼
ã?¹ã?¿ã?ªã?«æ?´å?ç??	95%ä»¥ä, ?	LoRAå?·ç¼æ?§é«?
ç ·æ£èª?æ?ªå?ºç??	100%	å?«ç?ã?·ãªã?·ã?¼éµå®?
å?«ç?ç?£æ?»	é?å·0ä»¶	ZIKUUç?£æ?»é??é?

ç??æ?·çµ•æ??ã·ã??èª?å¼ã?»é??ã?»ç?ã®å¼å,ã?·ã?ä,?è²«ã?ã?·æ?·æ?³ã·®å¼å·
¼?reflective cadenceï¼?ã?å?·ç¼ã?ã??ã?ã??

â??

è??å-?

Alå¼é?·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?ã¼ã??ã??ã?·ã?ã·ã·ã·ã??ã??

æ??ä½?ã?»å¥èªç¹ã?»èª?å½ã·®é,æ??ã·ã??èª?ä½ã?ã?æ?è??ã·®å?«ç?ã??ä½?ç¼ã?ã?·

Alã·æ??å??ç??è??æ?ã??ã??ã£ã·äªæ¼ã·ã·ã·ã¼è©±ã?ã??ã·ã·ã??ã·è½ã«ã?ã??ã??

æ??è²ç¼å¼ã·ã«ã?ã?ã¼ã·ã·Alã·èª-å¼ã??ã??ã??å±æ??ã??ã??ã??ã?ã??ã¼´èµè?ã?·ã·ã·
ã·ã¼æ©?è½ã·ã??ã??

çµ•è«?ã·ã±æ??

Alå¼é?·ã·ã·ã?«ç?ã·æ??ä½?ã??ä,ä½?ã??ã?ã?ã·ãªæ¼å¼æ?·å??Alã·®ã??ã?-
ã??ã?¿ã?ªã??ã·§ã·ã??ã??

ã?ã·®ã?£ã??ã?ã¼ã?·ã·ã·è²è??ã?»å?«ç?ã?»æ??å??ã·®ã?ã?ã·ã«ã?ã??â??é??ï¼?Mai¼?â?
ã??Alã·«å??ã?å?ã??è©|ã·¿ã·§ã??ã?ã??ã??

ä?»å¼ã·`kazusan-lora-v1`ã®ä,?è?-å-é??ã·ã?

RAGã??ä½µç?·ã?ã?ã?Alå¼é?·ã·ã?©ã¹ã?¿ãæ§çç?ã??äº?å®?ã??

åº?å??æ??å??ã?»æ??è²ã?»å?å¼ã??ç¶?æ?¿ã?ã??Alè`è`ã·®æ?°ã?ã?æ·æ?°å¼æ??æ·çªã·
ã·ã¼ã?ã·ã?

å·è??æ??ç?®

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

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

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

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

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

Category

- 1. *Category*

Tags

- 1. AI
- 2. *AI*
- 3. LLM
- 4. *AI*
- 5. *AI*

Date Created

2025-10-28

Author

kazuo-tsubaki