

ã?¢ã??ã•¥ã•ã??å¼¼ã•§ã??ã??ã•?ã•?ã•?æ?•ã•£ã•|ã•?ã??Alã•®ç ?ç©¶¶

Description

ç ?ç©¶¶ã•?æ?ã•ã•?ã°?ã•?ã•?ã•?è£?è£?ã•ã•®ã•§ã•?ã¼ã•?æ?ã•?ã•?æ?¹ã•?è?¬ã•?ã•£ã•?ã•?ã•?ã•?ã•?ã•?
¼ã•?ã•?ã•?

[ã?¢ã??ã•¥ã•ã??å¼¼ã•?ã•?ã¼ã•?ã•?ã•?æ??ç?¹ã•§ã•?ç«?æ?¼ã•?è?ã•?ã•?ã•?ã•?ã•?ã•?ã•?ã•?ã•?ã•?
ã•?](#)

ã•?§è|æ?jè?è?ã•?ã•?ã•?ã•?«i¼?LLMi¼?ã•?é?¢ã•?ã•?ã•?ã•?ã•?ã•?çµ?ã•?

RAGi¼?Retrieval Augmented Generationi¼?ã•?é?ã•?ã•?i¼?Quantizationi¼?ã•?ã•?©ã•?
®æ??æ³?ã•?ã¼ã•?£ã•?ã•?ã•?ã•?ã•?ã•?ã•?ã•?ã•?ã•?ã•?ã•?ã•?ã•?ã•?ã•?ã•?ã•?

- ç§ã•®ã•?ã•?ã•?è?ã°?ã•?ã•?ã•?REST APIã•?ã¼ã•?£ã•?ã•?ç«?ã•?ã•?ã¼ã•?ã•?
|ã•?ã•?ã¼ã•?ã•?ã•?ã¼ã•?¹ã•?ã•?ã•?
- ã•?ã•?ã¼ã•?ã•?ã•?ã¼ã•?¹ã•?ã•?ã•?ç«?ã•?ã•?ã•?ã•?ã•?ã¼ã•?ã•?ã•?ã¼ã•?¹ã•?ã•?ã•?ã•?ã•?
ã•?ã•?ã¼ã•?¹ã•?ã•?ã•?
- RAGã•?ã¼ã•?£ã•?ã•?ã•?ã•?è?ã°?ã•?ã•?ã•?ã•?ã•?ã•?ã•?ã•?ã•?ã•?ã•?ã•?ã•?ã•?ã•?
ã•?
- LLMã•?ã•?ã•?GPUã•?VRAMã•?§ã•?ã•?ã•?ã•?ã•?ã•?ã•?ã•?ã•?ã•?ã•?ã•?ã•?ã•?ã•?
ã•?
- ã?è|ã•?
¼ã•?ã•?è?ã•?ã•?
- é«?é??ã•?ã•?ã•?ã•?ã•?ã•?ã•?ã•?ã•?ã•?ã•?ã•?ã•?ã•?ã•?

1ã•?2ã•?cronã•?ã•?©ã•?§ã•?æ??ç?ã•?ã•?ã•?ã•?ã•?ã•?ã•?ã•?ã•?ã•?ã•?ã•?ã•?ã•?ã•?
®ã•?ã•?ã¼ã•?ã•?ã•?ã¼ã•?£ã•?ã•?ã•?ã•?ã•?ã•?ã•?ã•?ã•?ã•?ã•?ã•?ã•?ã•?ã•?

ã•?ã•?ã•?ã•?ã•?ç?ç?ã•?
ã•?

ç?»ã•?è?è?ã•?ã•?ç?©ã¼?æµ?ç?¥ã•?é?¢ã•?ã•?ã•?ã•?ã•?ã•?çµ?ã•?

ã•?ã•?ã•?Alã•®ã•?é??ã•?§ã•?
ã•?
ã•?
¬é?ã•?ã•?ã•?ã•?ã•?ã•?ã•?ã•?ã•?ã•?ã•?ã•?ã•?ã•?ã•?

Transformerã•?Diffusionã•?ã•?ã•?ã•?ã•?ã•?ã•?ã•?ã•?ã•?ã•?ã•?ã•?ã•?

è«?æ??ã•?ã•?ã•?ã•?ã•?ã•?ã•?ã•?ã•?ã•?ã•?ã•?ã•?ã•?ã•?ã•?ã•?ã•?
«ã•?¹ã•?
[ã?¢ã??ã•¥ã•ã??å¼¼ã•?ã•?ã¼ã•?ã•?ã•?æ??ç?¹ã•§ã•?ç«?æ?¼ã•?è?ã•?ã•?ã•?ã•?ã•?ã•?ã•?ã•?ã•?
ã•?](#)
§æ§?ã•?ã•?ã•?ã•?ã•?ã•?ã•?ã•?ã•?ã•?ã•?ã•?ã•?ã•?ã•?

