

ç ?ç©¶ã·æ?,ã·ñ·ă°?ã·?ă°şèç?è£?ã·ªã·®ã·şă?/?ă¼·ã·æ?,ã·?ã·?æ?¹ã·?è?¬ã·?ã·?ã·?ã·?ã·?ã·?ã·?
¾ã·?ã·?ã·?

ǎꞤꞥè|·æ̣jè̀?è^a?ã?ɕã??ã?«ĩ¼?LLMĩ¼?ã•é?ɕã?ã??ǎ?ã??çµ?ạ̃.

1. ç§•ã•®ã??ã?ã°è``ä°?ã•?ã??REST APIã??ã½¿ã•£ã•|æ??ç« ¤??ã°?ã¾?ã•?ã•
|ã??ã°¼ã°¿ã??ã°¼ã°¹ã°?ã•?ã°?

2. ã??ã°¼ã°¿ã??ã°¼ã°¹ã°®æ??ç« ¤??ã°?ã°?ã°?ã°¿ã°¿ã°¼ã°?ã°¼ã°¿ã°¿ã°¼ã°¹ã°«ã°?æ•?ã°?ã°?

3. RAGã°?ã½¿ã•£ã•|ã°?ã°?ã°è``ä°?ã°?ã°?ã°?ã°³ã°?ã°?ã°«è¿½¿ã°ã°?ã°|æ??æ°¿ç°?æ°?ã°?ã°
?ã°?

4. LLMã°¹ã°?ã°GPUã°®VRAMã°§ã°?ã°?ã°?ã°?ã°«é°?ã°?ã°?ã°?ã°?ã°?ã°?ã°?ã°?ã°?ã°?ã°?ã°?ã°?ã°?ã°?ã°?

5. å¿ è|•ã°«å¿?ã°?ã°|ã°?ã°|ã°?ã°³ã°?ã°?ã°¼ã°?ã°³ã°?ã°?ã°?è¿ã°?ç°?æ°?æ°?ã°®è¿ç°
¾ã°?èª¿æ°?ã°?ã°?

6. é«?é°?ã°?ã°?ã°?ã°?ã°?ã°«CTranslate2ã°?ã½¿ã°?

ã•?ã??ã??ã•«ç¿?ç??ã•?ã•ªã•?ã??ç??æ?•Alã??å®?æ¥å??ã•«é•©ç?¨ã•?ã??é•ç??ã??æ?£ã•?ã•!ã•?ã••
ã??

ç?»å?•è¹è?ã??ç?©ä½?æ¤?ç?¥ã«é?¢ã•?ã??å?•ã??çµ?ã¿

ã?·ã??ã?,Alã•®ã??é??ã•§ã•
 ?ã??ã?·ã?³ã?°ã?«ã??ã?¼ã??ã?³ã?³ã??ã?¥ã?¼ã?¿ã?¼ã??ã??ã?¤ã?³ã?³ã??ç?·ã?·ã•
 ?ã?·ã?¹ã??ã? é??ç?°ã•®ç?ç©¶ã•§ã?·ã??é•©ç?·"ã?é??ã?·"ã?¤ã?ã?«æ,jã??ã•®ã•§ã|ç¿ã??ã?¼ã??ã•
 ¯éã??ã?ªã•ç?·æ?·ã•§ã•ã??ã?§ã•?ã??ã?ã??

Transformerã??Diffusionã?¢ã??ã?«ã•ã•©ã??è?ªä½?ã?ã??

[illegible]

