

AI??????"??"

??

1.
 1.
 2.
2.
 1. $Y = A X_1 + B X_2 + C X_3 + \dots$
3.
 1.
 2. /

????

8

fp32 int8

=> kv cache

=>

=> AI

?????????

1.
2.
 1. => =>
 2.
 3.

Revision #2

Created 2025-01-11 09:44:05 UTC by Colin

Updated 2026-04-29 07:34:10 UTC by Colin