

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 #1

Created 11 January 2025 09:44:05 by Colin

Updated 11 January 2025 09:44:05 by Colin