

LLMs from Scratch: Assignment 1A

计科31 刘思远

2025.10.12

Acknowledgements I have discussed with my classmate Chenghan Yang, who helped me a lot in the implementation of flash attention.

Task 4

Problem 4.1 (a)ReduceSumConfig

(b)It calculates $\lceil \frac{a}{b} \rceil$.

(c)(2, 512, 64)

(1, 1024, 256)

(2, 512, 64)

Problem 4.2 S0: 2, S1: 512, B1: 64