

GPT-3 MoE 32x350M Training Step CUDA Kernel Summary: 8x4 A100 GPUs

