

GPT-3 MoE 8x350M Training Step CUDA Kernel Summary: 1x8 V100 GPUs

