

GPT-3 MoE 32x1.3B All-to-All Straggler Effect: 8x4 A100 GPUs

