

GPT-3 MoE 350M All-to-All Group Size (EP) vs Iteration Time: 1x4 A100 80GB

