

Packed-4 20M · fixed LR 0.008 · 131,072 tokens per batch
16 configurations · 48 runs · mean $\pm$ sample SD (three seeds) · lower is better

