

Packed-4 20M · 131,072 tokens per batch · microbatch 32 per model
 $\beta_1 = 0.9$ · 31 configurations · 93 runs · mean $\pm$ sample SD (three seeds)

