

Synchronous 20M · 131,072 tokens per batch · microbatch 128
72 runs · error bars: mean $\pm$ sample SD over three seeds · lower is better

