## throughput Wizardcoder-15b w b c o Throughput (Higher is Better) 6 5 20 40 60 80 $B_{atch}$ Size100 <sub>120</sub>