

Throughput vs TPOT for Different Arrival Rates
LLM: LLAMA-3.1-70B-Instruct
Batch Sizes: 4, 8

