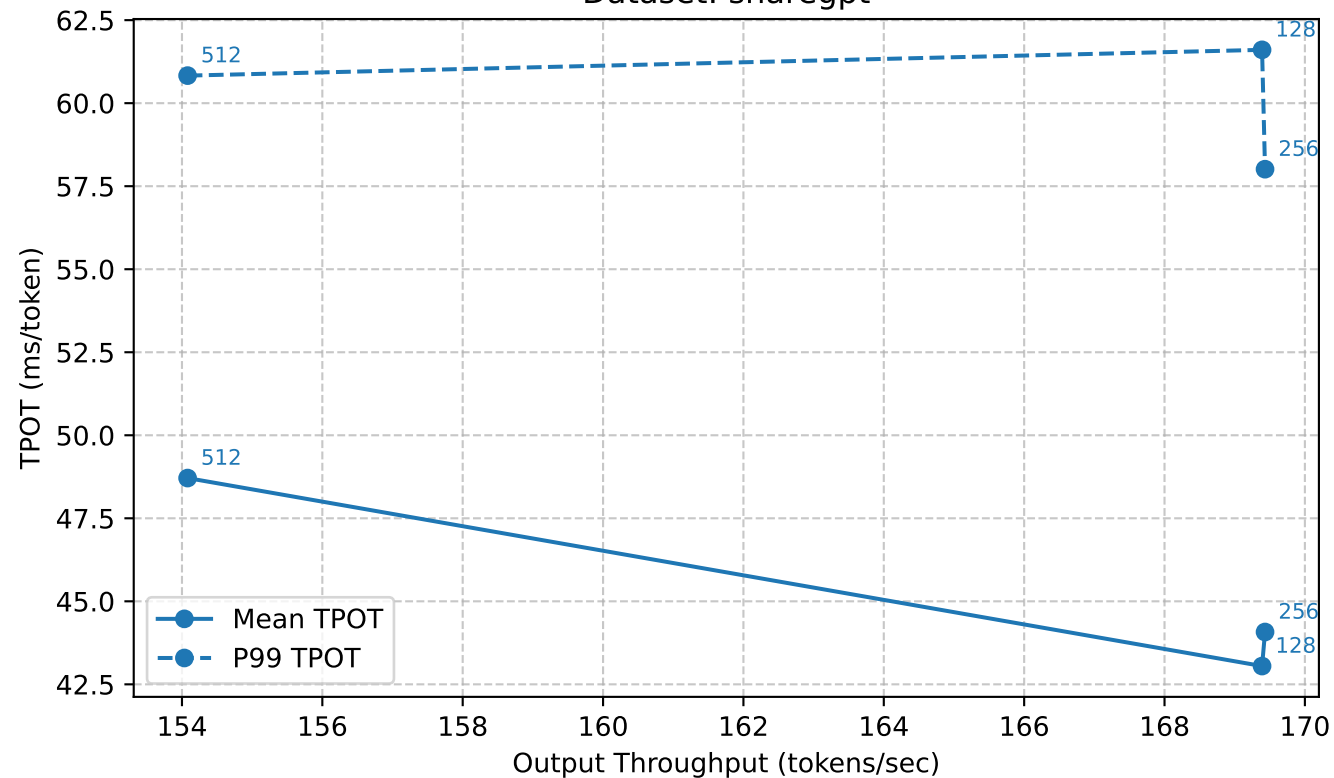


Throughput-Latency characterization
01-ai_yi-1.5-34b (TP=2)
Batch Size: 8
Dataset: sharegpt



Throughput-Latency characterization
01-ai_yi-1.5-34b (TP=2)
Batch Size: 8
Dataset: wildchat

