Normalized Latency on ShareGPT Dataset (Small Scale)