

TTFT vs Arrival Rate for Different Models and Batch Sizes
LLM: LLAMA-3.1-70B-Instruct

