

Average Number of Accepted Tokens per Step
LLM: LLAMA-3.1-70B-Instruct
Batch Sizes: 4, 8

