

Llama-3.1-8B Generation Time: KV Cache vs No KV Cache  
(Long Prompt)

