LLM Input Token Count vs. Latency (s), $R^2 = -0.00389$ 20.0 17.5 15.0 Latency (s) 0.01 7.5 5.0 2.5 400 500 100 200 300 LLM Input Token Count