

Total Inference Time vs Input Token Size

