Computational Cost: Inference Time vs Context Length