Performance for Model: 1B, Seq Len: 1024 on RTX5090