

RTX5090 Performance Report
Model: 8B, Sequence Length: 4096

