Mistral 3.1 24B Throughput tok/s Throughput tok/s AMD MI250X AMD MI300X **NVIDIA A100 NVIDIA GH200 NVIDIA H100**