gpt-oss-120B MXFP4 MoE - throughput vs config