

# Qwen MoE: Memory Usage vs Batch Size

## Expert Caching Memory Footprint

