

TTFT on One GPU with Batch Size 16 & Input Size 1024

