

# Inference Latency vs. Network Bandwidth (ALBERT-large)

