

Roofline Model: AB-Fused (K25) on A10G

(Analytical AI assumes dense A+B+C traffic; actual AI lower due to sparse overhead)

