

Experiment Decision Framework

Strong Loss
($p < 0.01$, -10%)

STOP

Guardrail Fail
(Latency +50%)

ROLLBACK

Null
($p = 0.3$, +1%)

STOP

Marginal Win
($p = 0.04$, +3%)

ITERATE

Strong Win
($p < 0.01$, +15%)

SHIP

Always consider: Statistical significance + Practical significance + Guardrails