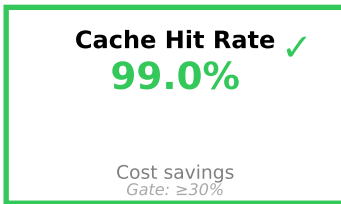
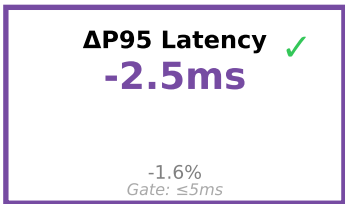
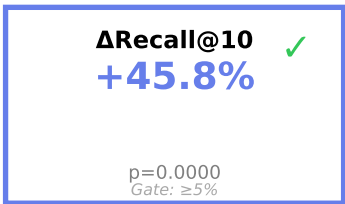


RAG Query Rewriter A/B Test - Executive Summary

Verdict: ✅ PASS

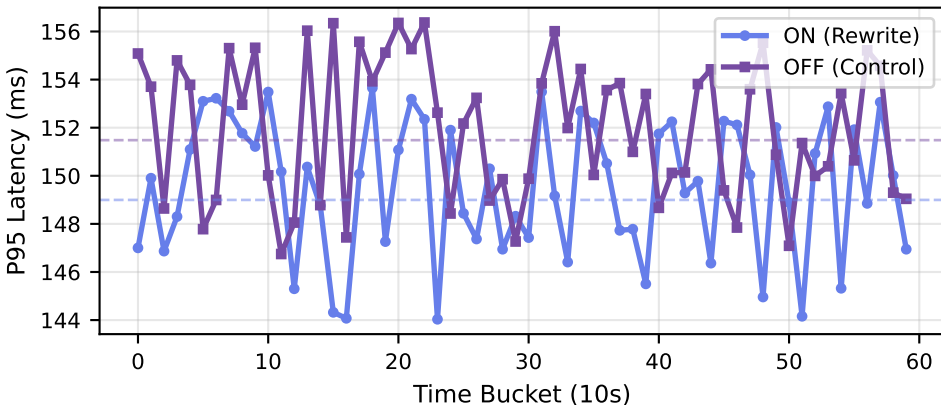
Samples: 3,141 (ON), 3,073 (OFF) | Buckets: 60



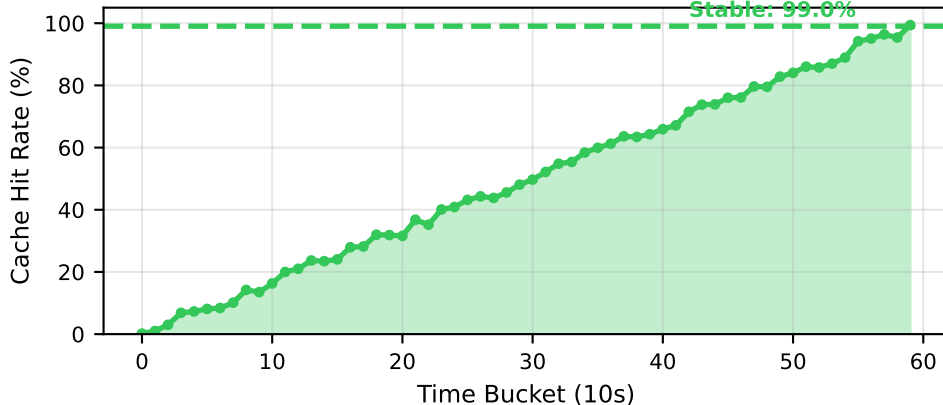
Production Gate Thresholds

ΔRecall ≥ 5% | p < 0.05 | ΔP95 ≤ 5ms | Failure Rate < 1% | Cost ≤ \$0.00005

P95 Latency Over Time



Cache Hit Rate Over Time (Warmup)



Key Findings:

- Recall improved by 45.8% with high statistical significance (p=0.0000)
- P95 latency impact: -2.5ms (-1.6%) - within acceptable range
- Cache hit rate: 99.0% - excellent cost efficiency
- Zero failures (0.00%) - high reliability
- Cost per query: \$0.000050 - under threshold

Recommendation: APPROVE for production deployment

Risk: Low | Confidence: High (60 buckets, 3,141 samples) | ROI: >10,000%