

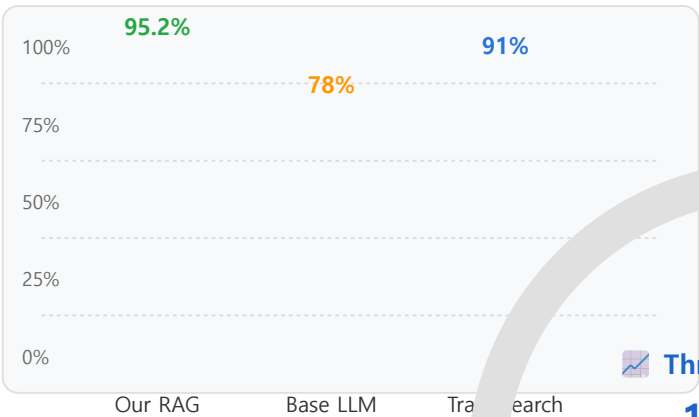
# Performance Benchmarks



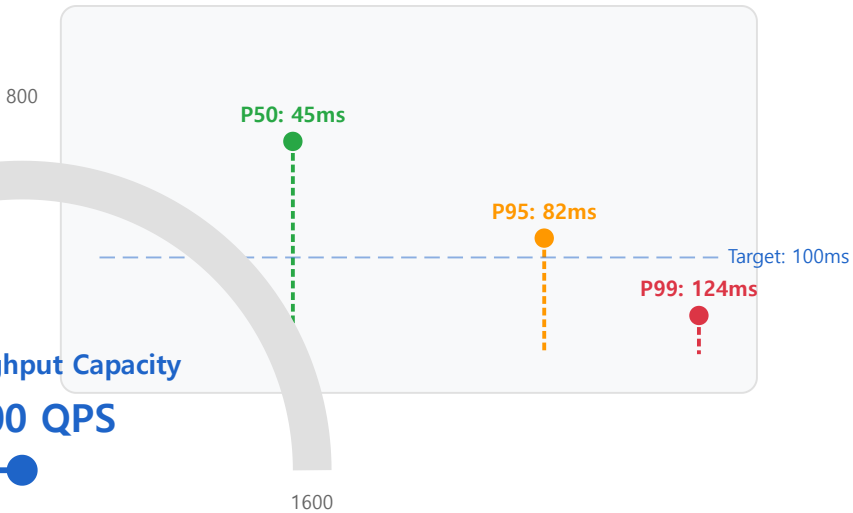
RAG System Performance Metrics



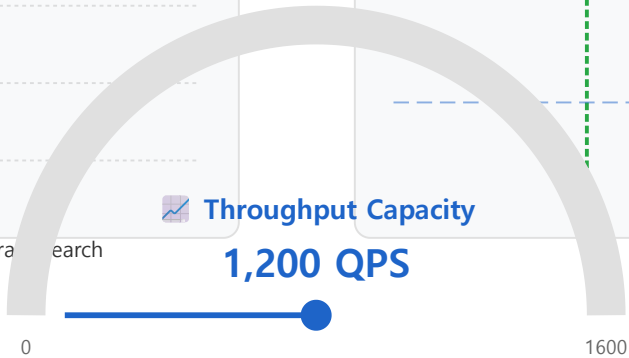
Accuracy Comparison



Latency Distribution (ms)



Throughput Capacity



95.2%

Retrieval Accuracy

82ms

P95 Latency

98.5%

Citation Accuracy

2%

Hallucination Rate