Performance Benchmarks

95.2%

Retrieval Accuracy

Baseline: 87%

82ms

P95 Latency

Target: <100ms

1,200

Queries/sec

Peak load

98.5%

Citation Accuracy

Manual verification

System Comparison

Factual Accuracy	**95%**	78%	91%
Latency (P50)	**45ms**	2000ms	120ms
Source Citations	**Yes**	No	Yes
Up-to-date Info	**Real-time**	Training cutoff	Index lag