

Scenario

Early prototype or internal-only tool

Customer-facing chatbot or assistant

Regulated industry (finance, healthcare, legal)

Cost-sensitive high-volume deployment

Autonomous agents executing actions

RAG-powered knowledge management

Recommended approach

AI observability only

Both

Both, with semantic emphasis

AI observability with selective semantic sampling

Both, with real-time semantic guardrails

Both, with retrieval quality focus

Rationale

Focus on operational stability. Semantic evaluation can wait until outputs reach external users.

Output quality directly affects customer experience and trust. Semantic evaluation catches errors that customers would notice.

Compliance requires demonstrable output accuracy and reasoning traceability, not just system uptime.

Full semantic evaluation on every request adds cost. Sample-based semantic checks balance quality assurance with budget.

Agents that act on outputs (processing transactions, modifying records, sending communications) require semantic validation before execution, not after.

Retrieval misalignment is a semantic problem that performance metrics alone cannot detect.