Checklist

Questions to Ask a Vendor to Determine Whether Their AI is “Enterprise-Grade”

Introduction

“Agentic AI” is quickly becoming table stakes. What’s rarer—and harder—is enterprise-grade agentic AI: systems that don’t just reason and act, but do so safely, audibly, repeatably, and at scale inside real enterprises.

Enterprise-grade is not a UX flourish. It’s an architectural and operational commitment. The gap between a demo that works and a system that ships is measured in operational reality. For agentic AI, "enterprise grade" is a specific set of capabilities, methods, and mindsets that separate production systems from prototypes.

Here is a list of 10 questions to ask AI technology vendors to determine if their solution is enterprise-grade. These questions are designed to move beyond high-level features and into the practical architectural and operational realities required for production.

Have you read our full Evaluation Guide on this topic? **You can access the full guide here.**

Data & Integration

1. Does your solution support bi-directional data integration (read and write) across our specific ecosystem, including legacy on-premise systems and cloud storage?

    - ***Why it matters:*** Enterprise-grade systems must do more than just read data; they must be able to "write back" by updating records and triggering workflows in your existing systems of record (like CRMs or ERPs) to produce actual outcomes.

2. How does your system capture and utilize the context and reasoning behind data rather than just the system of record (i.e. the final outcome)?

    - ***Why it matters:*** Agents trained only on final outcomes (like an invoice) miss the critical reasoning found in email threads and approval chains that generated those outcomes.

Knowledge & Context

3. Does your technology use a knowledge graph to map relationships and compound value over time?
    - ***Why it matters:*** Unlike a simple vector database, a knowledge graph captures domain-specific entity relationships and temporal context, ensuring the system gets smarter (and becomes more futureproof) as your business context grows.

Security & Governance

4. How does your platform ensure that agent permissions and data-level access controls strictly mirror our existing organizational hierarchy and role-based permissions?
    - ***Why it matters:*** Agents should never be able to see or act upon data that the human user they are representing is not authorized to access.

5. What specific "pre-action" guardrails are in place to prevent violations of regulations like GDPR, HIPAA, or SOC2 before an agent executes a task?
    - ***Why it matters:*** In an enterprise environment, governance requirements must be enforced proactively rather than discovered after a violation has occurred.

Transparency & Accountability

6. Can you demonstrate the audit trail for an agent's decision, including the reasoning trace, rejected alternatives, and the specific tools used?
    - ***Why it matters:*** "The agent made a mistake" is not an acceptable report. You need a searchable, complete history for compliance, debugging, and future training.

Evaluation & Reliability

7. What "batteries-included" evaluations do you provide for factual accuracy against internal knowledge bases, and how do you support custom evaluations for our domain-specific jargon?
    - ***Why it matters:*** Generic benchmarks don't predict performance on your specific workflows. You need continuous monitoring that triggers alerts when performance degrades.

8. How does your architecture achieve bounded risk regarding hallucinations and non-deterministic behavior?
    - ***Why it matters:*** While zero hallucinations are impossible, an enterprise system must use architectural defense-in-depth (like RAG grounding and citation verification) to catch errors before they cause damage.

Operations & Human Oversight

9. How does your model routing logic optimize for the tradeoff between cost, speed, and reasoning complexity for different types of queries?
    - ***Why it matters:*** Orchestration should automatically route simple tasks to cheaper models and escalate complex reasoning to frontier models while staying within your budget.

10. What specific handoff protocols and review interfaces are available to ensure humans remain in the loop for high-stakes decisions?
    - ***Why it matters:*** "Augmented AI" means the agent works for the human. You need clear protocols for when a human should override an agent and how that feedback improves the system.

Make sure you read our full Evaluation Guide on this topic for a broader discussion of each of these areas and considerations. You can **access the full guide here.**