CUBCLOUD AI · THOUGHT LEADERSHIP · 2026

AGENTIC AI EXPLAINER

What Is an AI Agent — And Why Should Your Business Care?

By Joshua Rhines, Founder & CEO, CubCloud AI

There is a version of this post that starts with "AI agents are the next big thing" and then lists ten bullet points. This is not that version.

AI agents are not hype. They are a specific architectural shift in how software works — one with real consequences for how your business operates, who controls your workflows, and whether the AI you invest in compounds your advantage or someone else's. Before you make a decision about them, you deserve an honest explanation of what they actually are.

$7.6B	40%	40%
AI agents market in 2025 (Grand View Research)	of enterprise apps will embed agents by end of 2026 (Gartner)	of agentic AI projects will be canceled by 2027 (Gartner)

** The market and the failure rate are both real. The organizations that succeed will be the ones that understand both.*

The Difference That Actually Matters

Most businesses have already encountered AI in the form of a chatbot, a writing assistant, or a recommendation engine. These are useful tools. You give them input; they give you output. You decide what to do with it. The interaction ends there.

An AI agent is categorically different. Not in degree — in kind.

> **An AI agent is a software component that has the agency to act on behalf of a user or system to perform tasks —**
>
> orchestrating complex workflows, coordinating activities among multiple agents, applying logic to thorny problems, and evaluating answers to queries. (McKinsey)

The operative word is *act*. A chatbot tells you a flight is available. An agent books it, sends the confirmation to your calendar, notifies your team, and flags the expense to your finance system — without being asked to do each step individually. MIT Sloan researchers describe it precisely: agents can execute multi-step plans, use external tools, and interact with digital environments to function as powerful components within larger workflows.

This is not a semantic distinction. It is the difference between a tool that informs decisions and a system that executes them.

How an AI Agent Actually Works

Under the hood, an agent combines four capabilities that previous AI systems kept separate:

Component	What It Does
Perception	Observes its environment — reads emails, monitors dashboards, ingests documents, watches for triggers. It knows what is happening.
Reasoning & Planning	Breaks a goal into steps, sequences them logically, and decides how to proceed. The LLM is not just generating text — it is thinking through a problem.
Memory	Retains context across steps and sessions. It knows what it has already done, what failed, and what the user said last week. Without memory, an agent starts from zero every time.
Action	Uses tools to do things: calling APIs, writing to databases, sending messages, executing code, triggering other systems. This is where the work actually gets done.

Strip any one of these out and you no longer have an agent. You have a chatbot with ambitions.

The infrastructure connecting agents to tools has also matured significantly. Anthropic's Model Context Protocol (MCP) — now the de facto standard for agent-to-tool connectivity, adopted by OpenAI, Google, Microsoft, and governed under the Linux Foundation — has dramatically reduced the engineering overhead of deploying agents in real enterprise environments. As of early 2026, the MCP ecosystem includes over 10,000 active public servers and 97 million monthly SDK downloads.

The Market Is Moving Fast — and So Is the Failure Rate

The AI agents market reached approximately $7.6–7.8 billion in 2025 and is projected to exceed $10.9 billion in 2026. Gartner forecasts that 40% of enterprise applications will embed task-specific AI agents by end of 2026,

up from less than 5% in 2025. That is an eight-fold increase in enterprise deployment in a single year.

But here is the number that does not make it into most AI vendor decks:

> **Gartner, June 2025: Over 40% of agentic AI projects will be canceled by end of 2027 —**
>
> due to escalating costs, unclear business value, or inadequate risk controls.
>
> "Most agentic AI projects right now are hype-driven experiments, often misapplied." — Anushree Verma, Senior Director Analyst, Gartner

Gartner also estimates that of the thousands of vendors claiming to offer agentic AI, only around 130 are building genuine agent capabilities. The rest are rebranding chatbots and rule-based automation — a phenomenon Gartner calls "agent washing."

The organizations that will succeed with agents are not the ones moving fastest. They are the ones moving deliberately.

What Agents Are Actually Good At — and Where They Break

Where Agents Excel	Where Agents Fail
Repetitive but complex multi-step workflows with consistent logic — claims processing, contract review, compliance monitoring	Poorly structured or siloed data where the agent cannot reliably read its environment
Data-intensive tasks requiring cross-system synthesis — financial reconciliation, supply chain analysis, clinical documentation	Ambiguous success criteria — vague goals produce vague actions, often expensively
Time-sensitive, high-volume workflows where speed and consistency matter — ServiceNow documented a 52% reduction in complex case resolution time	No human oversight checkpoints — autonomous action without governance is a liability, not a feature
Well-defined processes with clear inputs, outputs, and measurable success conditions	No observability — if you cannot see what the agent is doing and why, you do not have a production system

MIT Sloan researchers found that in a 2025 clinical AI agent deployment at a major cancer center, 80% of the work was consumed by data engineering, stakeholder alignment, governance, and workflow integration — not the AI itself. The technology is not the bottleneck. Organizational readiness is.

The Question Nobody Is Asking Loudly Enough

When an AI agent acts — when it sends an email on your behalf, updates a customer record, approves a transaction, or routes a patient case — it is making a decision. And the system making that decision is running on infrastructure that belongs to someone.

If that infrastructure is a third-party cloud platform, your operational decisions are being executed by systems you do not own, governed by terms of service you did not negotiate, running on hardware in a jurisdiction you did not choose. For most general-purpose workflows, that is a reasonable tradeoff. For regulated industries, proprietary processes, and any workflow touching sensitive data, it is not.

> **A chatbot that gives a bad answer is a nuisance.**
>
> **An agent that takes a bad action is an incident.**
>
> The governance requirements are fundamentally different when AI is executing rather than advising.

The organizations building agent infrastructure on hardware they own and control — with full observability, auditability, and the ability to halt, inspect, and correct any workflow — are building something qualitatively different from those deploying agents through a SaaS subscription and hoping the terms of service are sufficient protection.

What to Do If You Are Evaluating Agents Now

Principle	What It Means in Practice
Start with one workflow	Identify a specific process that is high-volume, well-defined, and consuming disproportionate human time. Prove value there before expanding.
Demand observability from day one	If you cannot see exactly what your agent is doing and why, you do not have a production system. Audit trails are not a feature — they are a prerequisite.
Keep humans in the loop for high-stakes actions	The goal is not to remove human judgment. It is to focus it where it matters by automating the steps where it does not.
Own your data pipeline	An agent is only as good as the data it operates on. Clean, accessible, well-governed internal data outperforms any model difference.
Ask where the agent runs	This is not a technical question. It is a strategic one. The answer should match your risk tolerance, regulatory environment, and competitive posture.

The Bottom Line

AI agents are the most significant shift in enterprise software since the cloud. They are also the most misunderstood, the most overhyped, and — for the organizations that get them right — the most consequential tool available for building durable operational advantage.

The market will separate into two groups. Organizations that deployed agents thoughtfully — on the right workflows, with proper governance, on infrastructure they control — will have compounding advantages that are genuinely difficult to replicate. Organizations that chased the hype, deployed agents without governance, or

outsourced the entire stack to a vendor they don't truly understand will be in the 40% Gartner is predicting.

The technology is real. The opportunity is real. The gap between those two outcomes is not a technology gap. It is an infrastructure and governance decision that most organizations are making right now, often without realizing it.

> **Sovereignty is not a political stance. It is an architectural decision.**
> And nowhere in the AI stack is that decision more consequential than in agentic AI —
> where the system is not advising you. It is acting for you.

About the Author

Joshua Rhines is the Founder and CEO of CubCloud AI, a sovereign AI infrastructure company headquartered in Missoula, Montana. CubCloud designs and operates private AI infrastructure — including agentic AI systems — for regulated industries and regional organizations across the Mountain West. CubCloud AI is the founding member of the MontanAI Alliance, a 501(c)(6) dedicated to building Montana's sovereign AI ecosystem.

Key Sources

- Gartner — "Gartner Predicts Over 40% of Agentic AI Projects Will Be Canceled by End of 2027" (June 2025)
- Gartner — Enterprise Application AI Agent Integration Forecast (2026)
- McKinsey — "What Is an AI Agent and How Will They Impact the World?" (2025)
- MIT Sloan Management Review — "Agentic AI, Explained" (2026)
- MIT Sloan / Boston Consulting Group — Spring 2025 AI Adoption Survey
- ServiceNow — AI Agent Deployment Case Study: 52% reduction in complex case resolution (2025)
- Anthropic — Model Context Protocol: modelcontextprotocol.io (2024–2026)
- Grand View Research / Gartner / IDC — AI Agents Market Size Forecast 2025–2030