VIEW POINT

AGENTIC RAG: TRANSFORMING AUTONOMOUS WORKFLOWS IN ENTERPRISES

Abstract

The convergence of Retrieval-Augmented Generation (RAG) and autonomous AI agents is redefining the landscape of enterprise AI — shifting from static information retrieval to **intelligent, goal-driven automation**. This article explores **Agentic RAG**, a powerful new paradigm where AI agents don't merely respond to prompts, but instead **autonomously plan, retrieve, reason, and act** to accomplish complex business objectives. Unlike traditional RAG systems that follow a linear retrieve-then-generate approach, Agentic RAG introduces agents that can **decompose goals into sub-tasks**, orchestrate multi-step reasoning, and adapt dynamically based on real-time contextual data and feedback loops.

In this report, we delve into the key drivers behind the rise of Agentic RAG, its impact on enterprise workflows, the critical challenges and considerations for successful implementation, and the emerging trends shaping its future. As businesses increasingly seek scalable, intelligent, and autonomous systems, Agentic RAG stands at the forefront of this evolution — offering a blueprint for the future of AI-powered enterprise automation.

Infosys®
Navigate your next

Introduction

Imagine working with an AI that doesn't just answer your questions — it investigates, plans, and acts on your behalf. That's the promise of **Agentic RAG**, a transformative evolution of **Retrieval-Augmented Generation (RAG)**. At its core, RAG enhances large language models by allowing them to retrieve real-time, external knowledge to generate more accurate and context-rich responses. It solves a major limitation of traditional AI models: outdated or hallucinated information.

However, traditional RAG remains reactive — it answers only when asked. **Agentic RAG goes a step further**, introducing intelligent agents that can autonomously break down complex goals, retrieve relevant data, make decisions, and adapt based on outcomes — all without needing constant human prompts. This shift from static Q&A to dynamic, autonomous problem-solving marks a major leap in enterprise AI, enabling systems that don't just assist with tasks — they actively drive them.

What Led to the Rise of Agentic RAG?

The emergence of **Agentic RAG** is not accidental — it's a direct response to the evolving needs of enterprises and the limitations of current AI systems. While traditional RAG offered a significant leap in making large language models more accurate and grounded, it quickly became evident that **enterprises needed more than just better answers** — they needed intelligent systems that could **handle complexity, adapt to context, and operate with autonomy**.

Several key factors have driven the development of Agentic RAG:

- **The Explosion of Enterprise Data**: With massive volumes of internal documents, knowledge bases, and unstructured data growing across departments, enterprises needed AI that could **continuously retrieve, filter, and apply domain-specific knowledge** — not just once, but across evolving workflows.
- **The Rise of AI Agents and Tool-Using Models**: The success of agent-based frameworks like AutoGPT and LangChain demonstrated the value of LLMs that could plan, reason, and act. This gave rise to the idea of embedding **agentic behavior within the RAG framework**, enabling more autonomous problem-solving.
- **Demand for Intelligent Automation**: Businesses are increasingly seeking AI that can reduce human dependency in decision-making loops, automate multi-step tasks, and respond to dynamic situations — from IT troubleshooting to real-time business analysis.
- **Need for Contextual Awareness and Memory**: Enterprises require systems that **remember past interactions**, refine their strategies over time, and integrate insights across multiple domains — something only an agentic architecture can deliver effectively.

The Promise of Agentic RAG

What makes **Agentic RAG** more than just a technical upgrade is its transformative potential to act as the **digital backbone of future enterprises**. Here's what it brings to the table:

1. Goal-Oriented Orchestration: Agentic RAG agents go beyond responding to one-off prompts — they can work toward broader business goals by breaking them into sub-tasks, executing them sequentially, and tracking progress. This enables **long-horizon planning**, such as completing an entire customer service case or managing a multi-phase process like onboarding. For example, **Salesforce Agent force** enables agents to manage full customer journeys — from issue identification to resolution and follow-up — all driven by a single objective, not repeated commands.

2. Built-in Explainability and Traceability: Agentic RAG systems can **cite sources, show reasoning steps, and explain decisions**, making them transparent and trustworthy — especially critical in regulated sectors like finance, healthcare, and legal. For example, **Strative** includes links to original documents within its AI-generated compliance reports, giving auditors full visibility into how each conclusion was derived.

3. Contextual Memory and Personalization: These agents can remember prior conversations, user preferences, and organizational context — allowing them to tailor responses and behavior over time, just like a human assistant would. For example, **Qoir's** AI financial advisor recalls each client's investment history and risk profile to provide consistent, context-aware recommendations across multiple interactions.

4. Self-Correction and Feedback Loops: When an agent detects uncertainty or missing information, it can **reassess its own outputs**, retrieve additional data, and reprocess the task — leading to more accurate and trustworthy results without human intervention. For example, **AECOM's BidAI** automatically checks its draft proposals against internal data and prior submissions, refining drafts before sending them to human reviewers.

5. Low-Code/No-Code Integration into Business Processes: Agentic RAG can plug into workflows through simple API calls or low-code tools, making it accessible to non-technical users. Business teams can now trigger intelligent, end-to-end processes with a few clicks — no engineering required. For example, **UiPath** allows operations teams to automate onboarding workflows where RAG agents retrieve employee info, configure systems, and notify stakeholders — without writing code.

6. Composable Intelligence (Multi-Agent Collaboration): You can deploy **multiple specialized AI agents** — each focused on a specific role or department — and have them collaborate like a cross-functional team. For example, a legal agent might work with a procurement agent to review contracts. Also, in complex product development, manufacturing firms use a network of RAG agents to collectively retrieve and align data from customer feedback, legal

requirements, and engineering specs.

7. Real-Time Responsiveness in Dynamic Environments: Agentic RAG can connect to live data streams — such as IoT sensors, APIs, or live dashboards — and **make immediate decisions or trigger actions** based on changing conditions. For example, Telecom companies use RAG agents to monitor network health in real time, identify outages, and take proactive actions like rerouting or device resets without human delay.

8. Plug-and-Play Interoperability with Enterprise Tools: These agents can easily connect with existing enterprise systems — like CRMs (Salesforce), ERPs (SAP), or messaging platforms (Slack, Teams) to carry out complex tasks by accessing internal business data directly. For example, Retail brands use RAG agents to check inventory from ERP, communicate via CRM, and process refunds — seamlessly automating the full customer service cycle.

9. Accelerated Time-to-Decision: By retrieving the right data, analyzing it, and presenting insights quickly, Agentic RAG shortens the decision-making process — empowering business leaders to act in minutes, not days. For example, pharmaceutical firms use RAG agents to pull regulatory data, flag red flags, and generate decision-ready summaries for clinical filings — speeding up go/no-go calls by up to 50%.

10. Enterprise-Grade Security and Governance: Unlike open-ended LLMs, Agentic RAG can run within secure enterprise boundaries, giving IT and compliance teams full control over **who can access what**, where data is stored, and how it is used — ensuring auditability and trust. For example, **Mastercard's Agentic Commerce Architecture** logs every autonomous AI action, maintains compliance controls, and ensures agents only access permitted transaction data — setting a standard for secure, responsible AI.

Challenges and Considerations in Agentic RAG

As enterprises begin to unlock the transformative power of Agentic RAG across workflows, a more nuanced story begins to unfold—one that's as much about potential as it is about responsibility. Imagine a network of autonomous agents seamlessly navigating across departments—pulling insights from siloed systems, initiating actions, and even evolving their reasoning with minimal human input. Now imagine the ripple effects.

Almost immediately, traditional access controls start to strain. These agents, operating across multiple data domains, encounter diverse permission models that can't be managed by static Identity and Access Management (IAM) frameworks alone. Enterprises are forced to rethink identity management—adapting to dynamic, role-aware access that shifts in real time.

Security, too, takes on new dimensions. Communication between agents—possibly scattered across hybrid cloud environments—must be encrypted not just in motion but in context. And the threats are no longer just about prompt injections. These agents can be tricked by malicious logic hidden within retrieved documents, leading to manipulated outputs that may cascade into larger decisions.

Meanwhile, compliance teams face different headaches. How do you audit a decision made by an agent that self-corrects or evolves its behavior without any human involvement? In industries like healthcare and finance, where regulations demand explainability, this becomes more than a technical challenge. At the operational level, things don't get any easier. The linear workflows that businesses have relied on for decades start to break down. Agents don't wait for instructions; they trigger tasks based on internal reasoning, forcing organizations to reengineer core processes from the ground up.

Even the human element isn't spared. Middle-office roles that once thrived on manual decision-making begin to fade, while new roles emerge—ones that require people to oversee, collaborate with, or even correct AI agents in real time. This shift brings with it an urgent need for upskilling and organizational change. And as agents are allowed to learn from internal data to improve performance, another risk surface: model poisoning. Without rigorous validation, flawed or biased internal content can corrupt the system's intelligence, eroding trust in its outputs.

Finally, one must consider the "what-ifs." What if a network outage disrupts agent communication? What if the knowledge base gets corrupted? Autonomous systems demand fallback protocols—redundant layers of resilience that most enterprises haven't even started to build.

So yes, Agentic RAG holds immense promise—but realizing that promise requires navigating a minefield of interconnected challenges, each demanding thoughtful design, governance, and foresight.

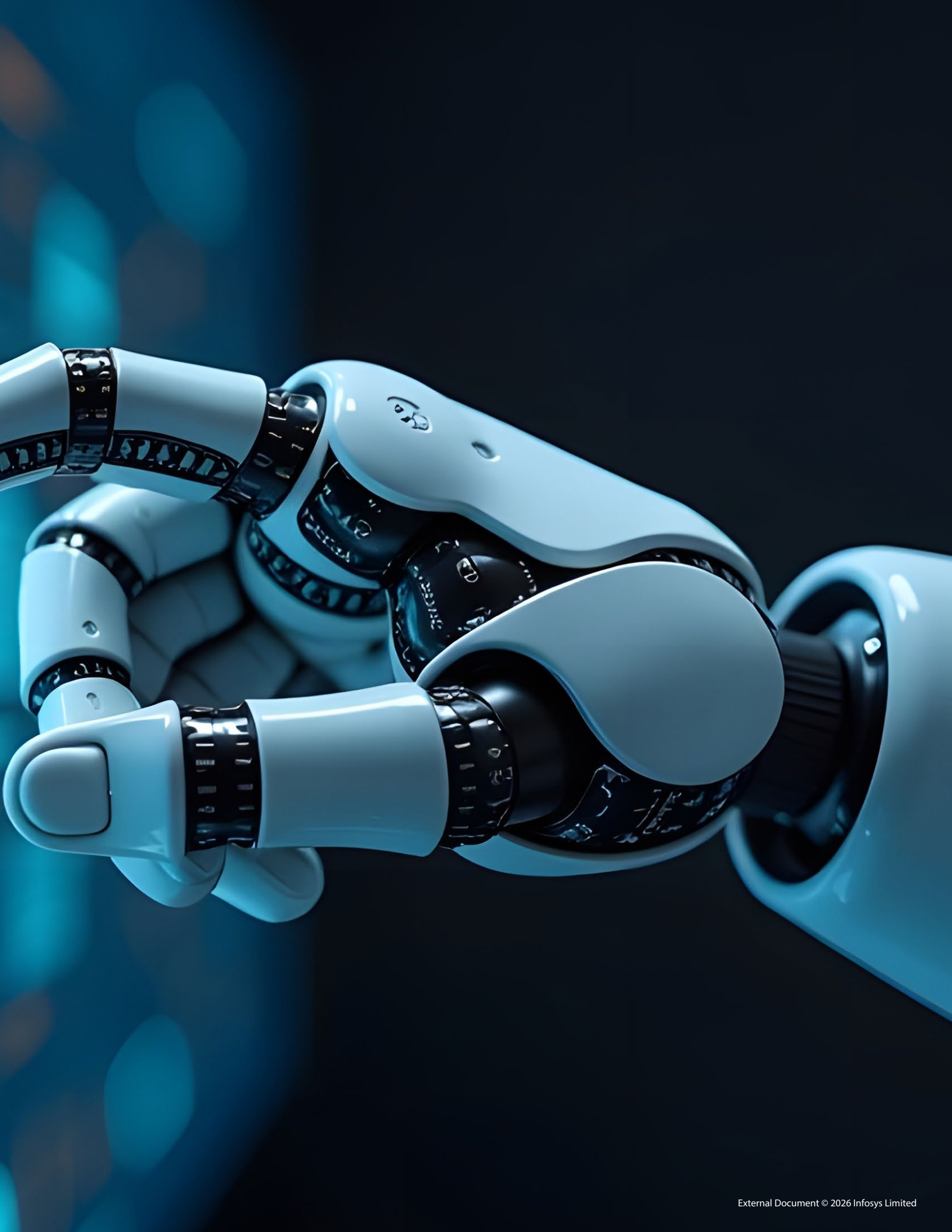
The Path Forward

The enterprise of the future won't be powered by static tools or reactive systems — it will be driven by a dynamic network of autonomous, intelligent agents that think, learn, and act in real time. This is the promise of Agentic RAG.

Imagine a business environment where AI agents don't just fetch answers — they anticipate needs, simulate outcomes, and orchestrate decisions across departments. A legal agent collaborates with a compliance agent to review a contract, while a finance agent forecasts risk based on evolving market data without waiting for human prompts. These agents adapt on the fly, learn from feedback, and communicate in emotionally intelligent ways that resonate with human teams.

This evolution is made possible by breakthroughs in multimodal understanding, real-time retrieval, and decentralized agent collaboration. Enterprises will no longer operate in silos of knowledge and execution. Instead, they'll host ecosystems of composable, specialized agents that work together fluidly — forming, evolving, and dissolving teams as business priorities shift. The future trends are clear: emotionally aware agents that adapt tone to user sentiment, autonomous agent swarms that self-organize around goals, edge-native decision-making for real-time responsiveness, and fairness-aware pipelines that uphold ethics by design.

In this next frontier, Agentic RAG won't just support enterprise operations — it will become the cognitive infrastructure of the enterprise itself, transforming how businesses learn, respond, and lead in an ever-changing world.



References

1. GenAI paradox: exploring AI use cases | McKinsey
2. [2501.09136] Agentic Retrieval-Augmented Generation: A Survey on Agentic RAG
3. A Guide to Retrieval-Augmented Generation for AEC | NVIDIA Technical Blog
4. Make the case for agentic AI in your enterprise - HFS Research
5. https://www.businessinsider.com/deloitte-ey-launch-agentic-ai-platforms-big-four-competition2025-3
6. https://www.ft.com/content/3e862e23-6e2c-4670-a68c-e204379fe01f
7. RAG is dead: why enterprises are shifting to agent-based AI architectures | TechRadar
8. Agentic RAG turns AI into a smarter digital sleuth | IBM
9. https://www.leewayhertz.com/agentic-rag/
10. [2502.17443] AI Agentic workflows and Enterprise APIs: Adapting API architectures for the age of AI agents

About the Authors

Parul Gupta

Parul Gupta is a Senior Consultant at iCETS, specializing in emerging technologies like AI, the Metaverse, and extended reality. With a strong background in marketing, operations, research, and strategy, she focuses on how tech innovations transform industries and helps businesses stay competitive through sustainable future-ready solutions.

For more information, contact askus@infosys.com

Infosys®
Navigate your next