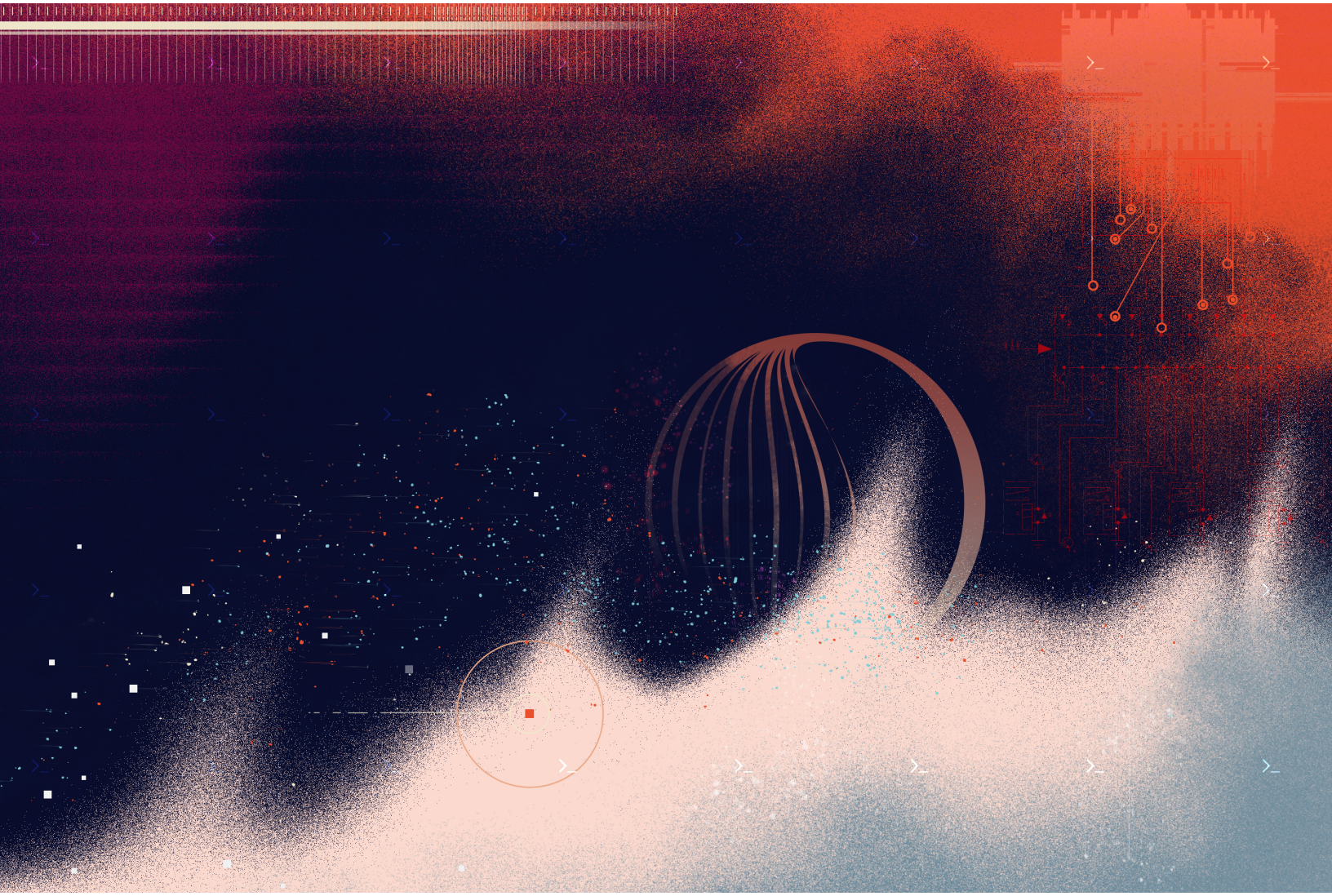
signal65



EXECUTIVE SUMMARY

Dell PowerEdge XE7745 RAG Evolution

Building Trust with Agentic RAG

AUTHOR

Brian Martin
AI and Data Center Lead | Signal65

IN PARTNERSHIP WITH

DELL Technologies

NOVEMBER 2025

The Enterprise AI Trust Challenge

As enterprises deploy large language models across critical business functions, the most pressing barrier to production adoption is not performance—it's trust. Without verifiable grounding in enterprise data, LLMs risk generating confident but inaccurate responses that can misinform decisions, breach compliance, or damage brand integrity. This is particularly critical in regulated domains like finance, healthcare, and government where every model output must be explainable, auditable, and traceable to authoritative sources.

Retrieval-Augmented Generation (RAG) emerged as the foundation for trustworthy AI at scale, connecting generative models to proprietary knowledge bases. However, traditional RAG approaches face significant limitations when dealing with complex, multi-faceted queries that require reasoning across multiple documents or understanding entity relationships.

Graph RAG recovered **15-20% more** supporting context

Agentic RAG **reduced hallucinations by ~40%**

Linear scale-out performance with Dell/ Broadcom networking

The Evolution from RAG to Agentic RAG

Traditional RAG combines embeddings, retrieval, and generation in a single pass—simple and effective for direct questions but limited to single-hop reasoning. **Graph RAG** extends this by representing knowledge as an interconnected graph of entities and relationships, enabling cross-document reasoning and recovering 15-20% more supporting context than flat retrieval. **Agentic RAG** represents the most sophisticated evolution: retrieval with autonomous reasoning. It introduces planning-and-reflection loops where an AI agent decides what to retrieve, how to verify it, and when to iterate, reducing hallucinations by approximately 40% through self-grading and adaptive querying.

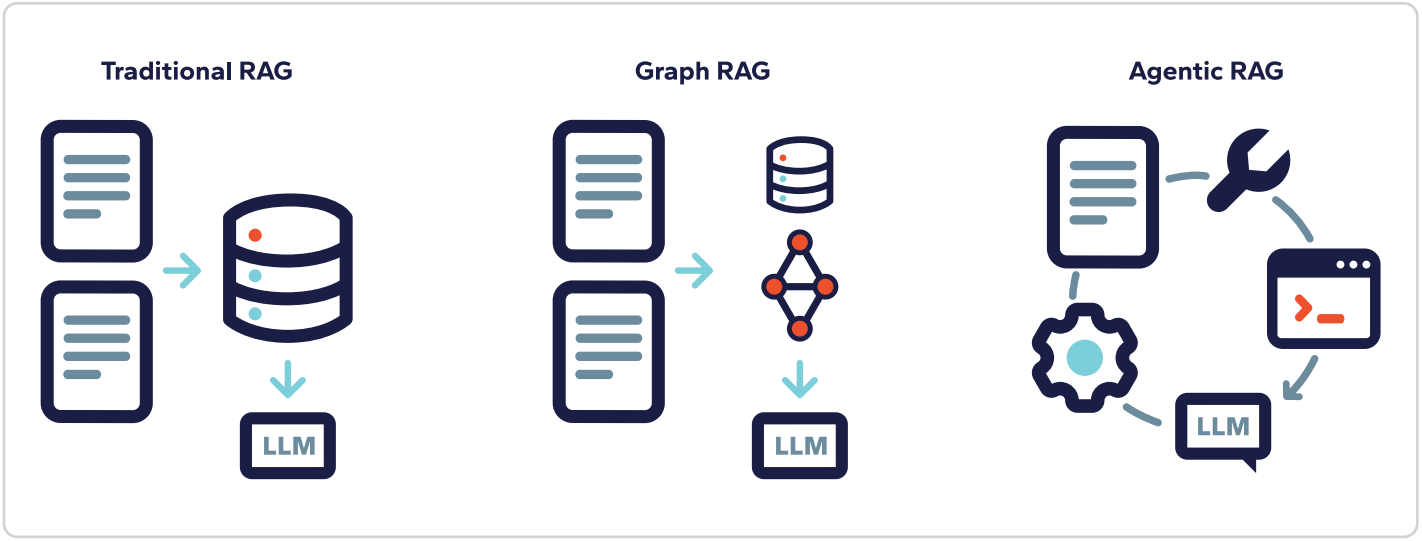


Figure 1: *RAG Models*

Performance Results and Business Impact

Signal65 tested F1 (precision and recall) and Accuracy (low hallucination) with Llama-3.3-70B-Instruct on HotPotQA and NQ-Open datasets. The results on Dell PowerEdge XE7745 infrastructure demonstrates measurable improvements across the RAG evolution:

- **Accuracy Gains:** Graph RAG achieved 89% faithfulness (vs. 79% baseline), while Agentic RAG reached 95% faithfulness—approaching human-level consistency
- **Hallucination Reduction:** Agentic RAG reduced factual errors by ~40% through iterative verification and self-grading mechanisms
- **Infrastructure Efficiency:** Linear scale-out performance with Broadcom BCM57608 400 GbE networking eliminated tail-latency spikes during complex agentic loops
- **Deployment Flexibility:** Support for 8 NVIDIA L40S, H100, H200, or RTX Pro 6000 GPUs enables organizations to scale from proof-of-concept to enterprise-wide deployment

Deployment Patterns and Use Cases

The choice of RAG architecture should align with specific business requirements:

Traditional RAG

- **Customer Support:** High-throughput FAQ resolution and documentation retrieval
- **Content Generation:** Simple summarization and report generation tasks
- **Requirements:** Sub-second response times, 80% accuracy acceptable

Graph RAG

- **Research & Analytics:** Cross-document reasoning and relationship discovery
- **Risk Assessment:** Analyzing entity relationships and market dependencies
- **Requirements:** Multi-hop reasoning, 90% accuracy needed

Agentic RAG

- **Compliance & Audit:** Traceable reasoning chains with full audit trails
- **Root Cause Analysis:** Multi-step investigation with iterative verification
- **Clinical Decision Support:** Evidence-based recommendations with source verification
- **Requirements:** 95%+ accuracy, full reasoning transparency, can tolerate 3-4 second latency

Dell PowerEdge XE7745: Purpose-Built for AI

The Dell PowerEdge XE7745 delivers enterprise-grade AI infrastructure specifically optimized for evolving RAG workloads:

Component	Specification
Processors	Dual AMD EPYC 9555 (64 cores/128 threads each)
Memory	Up to 6 TB DDR5 RAM
GPU Support	8x NVIDIA L40S, H100, H200, or RTX Pro 6000 Blackwell
Networking	8x Broadcom BCM57608 400 GbE controllers
Interface	PCIe Gen 5.0

This combination of high memory capacity, compute density, and network efficiency enables organizations to seamlessly evolve from basic document retrieval to advanced knowledge-reasoning workloads without infrastructure limitations.

Conclusion

For business leaders navigating AI transformation, the evolution from RAG to Agentic RAG represents more than technological advancement; it represents a critical shift in how organizations access, synthesize, and act upon enterprise knowledge. Organizations that invest in purpose-built AI infrastructure today will be best positioned to harness trustworthy AI systems that amplify human expertise while maintaining the transparency and reliability essential for enterprise deployment.

Important Information About this Report

CONTRIBUTORS

Brian Martin
AI and Data Center Lead | Signal65

PUBLISHER

Ryan Shrout
President and GM | Signal65

INQUIRIES

Contact us if you would like to discuss this report and Signal65 will respond promptly.

CITATIONS

This paper can be cited by accredited press and analysts, but must be cited in-context, displaying author's name, author's title, and "Signal65." Non-press and non-analysts must receive prior written permission by Signal65 for any citations.

LICENSING

DISCLOSURES

Signal65 provides research, analysis, advising, and consulting to many high-tech companies, including those mentioned in this paper. No employees at the firm hold any equity positions with any companies cited in this document.

ABOUT SIGNAL65

Signal65 is a leading research organization specializing in enterprise AI infrastructure optimization and deployment strategies. Our lab focuses on evaluating and optimizing AI hardware and software solutions for real-world enterprise applications, with particular expertise in large language models, retrieval-augmented generation systems, and distributed AI architectures.

For more information, visit signal65.com
or contact research@signal65.com

IN PARTNERSHIP WITH

ABOUT DELL TECHNOLOGIES

Dell Technologies helps organizations and individuals build their digital future and transform how they work, live and play. The company provides customers with the industry's broadest and most innovative technology and services portfolio for the data era, including cutting-edge AI infrastructure solutions like the PowerEdge XE series.

For more information, visit www.dell.com

CONTACT INFORMATION
Signal65 I signal65.com