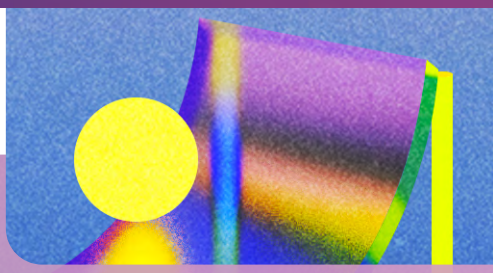
Embedded generative AI for smarter search and threat response

Many public sector organizations are under pressure to scale generative artificial intelligence (AI) solutions quickly to improve operational efficiency.

One of the fastest ways to achieve meaningful impact is by embedding generative AI directly into existing systems. Embedded generative AI can optimize the tasks your teams are already doing, support better decision-making, and save time for staff and constituents without requiring new workflows or retraining.

Elastic Cloud on Amazon Web Services (AWS) transforms how organizations access and act on information by embedding generative AI capabilities into the tools teams already use. Elastic enhances existing search and security functions with intelligent automation that delivers productivity results in days, not months.



What makes Elastic different?

Retire outdated keyword search

Legacy search functions often return pages of irrelevant results, wasting valuable time. Elastic replaces that frustration with a powerful vector database and conversational, context-aware search, so teams can ask questions like, "What did we learn from our cybersecurity incident last quarter?" and receive precise answers with full data source citations.

Revive forgotten knowledge

"Some legacy government formats are where knowledge has gone to die, and AI can get it back" says Dave Erickson, public sector distinguished architect at Elastic. Elastic's embedded generative AI search breathes new life into outdated repositories, allowing organizations to extract insights from data long considered unreachable.

Get instant visibility into high-volume data

Elastic enables generative AI-assisted search across a wide range of content—scanned PDFs, audio logs, decades of reports, and even Zoom meeting transcripts—without restructuring, duplicating, centralizing, or relinquishing control over any of your data.

Accelerate threat response and reduce alert fatigue

Elastic's AI Security Assistant helps organizations triage and investigate thousands of alerts, surfacing the ones that matter most. Analysts can use natural language to investigate threats and automatically open tickets, block access, or trigger remediation steps, all within the systems they already use.

Is Elastic right for my organization?

What human time are you trying to save?
If your staff spends substantial time searching for information or answering repetitive questions, Elastic delivers immediate productivity gains.

What's the most valuable data source that your teams struggle to access?
Elastic excels at searching across large, frequently updated datasets spread across multiple systems.

Do you need flexibility in AI models and deployment?
Elastic supports any LLM and works across cloud, hybrid, or air-gapped environments.

How quickly do you need to show AI value?
Elastic embeds into existing workflows with minimal disruption, delivering measurable improvements in days, not months.

Real-world impact: Generative AI-powered search transforms student financial aid process

Georgia State University, one of the largest public universities in the U.S., with over 50,000 students, used Elastic's embedded generative AI-powered search technology to tackle a critical challenge: helping students navigate the complex financial aid application process.

Using Elasticsearch, the university built a proof-of-concept generative AI tool that delivers personalized financial aid guidance. The system ingests data from multiple sources like student records, financial aid resources, deadlines, and eligibility requirements. Then, the tool uses semantic search and retrieval-augmented generation (RAG) to provide students with personalized answers about what financial aid resources are most relevant for them and the deadlines to submit for those resources—insights that would be nearly impossible to generate using traditional deterministic rules.

"Students who need assistance but can't find funding may be less likely to complete their college degrees," explains Jaroslav Klc, director of strategic initiatives and development for IT at Georgia State University. By embedding Elastic's generative AI search capabilities into their financial aid processes, Georgia State students can find and utilize critical aid opportunities, helping to improve retention and graduation rates across the university's diverse student body.

"You don't need to build a complex AI workflow. Just focus on increasing self-service access to the data your people need to do their jobs."

– Dave Erickson, Public Sector Distinguished Architect, Elastic

Elastic and AWS integration

Award-winning expertise
As an AWS Partner of the Year winner multiple years running, Elastic brings deep understanding of public sector requirements and proven experience supporting mission-critical public sector workloads.

Amazon Bedrock integration
Native integration with Amazon Bedrock provides access to leading LLMs, with built-in guardrails for responsible AI deployment. This combination delivers enterprise-grade AI capabilities with government-required security controls.

FedRAMP-enabled environments
AWS provides FedRAMP-authorized environments that support government customers at all security levels. Elastic runs natively in AWS GovCloud with full compliance for sensitive workloads.

Strategic advice for rapid generative AI adoption

Elastic experts share key considerations for embedding AI securely into existing workflows to accelerate generative AI adoption and begin seeing productivity results in days, not months:

Start with search for fast wins

Most organizations already operate search functions that frustrate users. Adding generative AI through retrieval augmented generation (RAG)—when AI searches your existing data to answer questions in natural language—offers immediate improvements with minimal risk. Focus on large, constantly changing datasets that require experienced staff to spend time answering repetitive questions. These deliver the highest impact while building organizational confidence in generative AI capabilities.

Embrace crawl-walk-run implementations

Begin with limited datasets, observe the results, then expand incrementally. This approach lets you prove value and verify governance controls work as expected before scaling to additional data sources.

Keep humans in the loop with full citations

Make all decisions based on citations from source material, not just large language model (LLM) output. All Elastic AI responses include links to original documents, enabling accountability and transparency while maintaining human oversight of critical decisions.

aws | elastic

Explore how Elastic helps public sector organizations embed generative AI into existing workflows for faster adoption and measurable productivity gains.

Learn more →