

The Path to Scalable, Privacy-First Generative AI

Across government agencies at all levels, generative artificial intelligence (AI) promises to advance mission outcomes, streamline constituent interactions and boost worker productivity. As made clear by the recent White House Executive Order on safe, secure, and trustworthy AI, however, there are perils on the path.

In this report, we will explore:

- → The challenges government faces in delivering context, confidentiality and compliance
- → How a solution that combines search technologies and cloud infrastructure can ensure relevant outputs from conversational AI, safeguard private data and meet compliance requirements
- → How such an approach will help elevate constituent experiences, deliver on agency mission goals, empower efficiency among government workers and provide strong security around generative AI applications

govloop elastic Google Cloud

CHALLENGE:
Providing Context, Confidentiality, Compliance

As technology leaders consider generative AI integration, government agencies typically face several impediments:

- **Agency modernization efforts are enhanced with AI-based search**, but people and organizations may use the same word to mean different things, making it difficult to pinpoint the content relevant to the query. To deliver accurate, meaningful outputs — through a constituent-facing chatbot, for example, or in support of a draft report — the AI needs a contextual understanding that often is lacking because it doesn't have access to an agency's private data.
- **It's difficult to ensure confidentiality of private data.** Generative AI needs to train on actual user data. With conventional approaches, feeding private information into the model risks exposing that data to prying eyes. For example, the AI may need to reach out to internal services or databases for data in order to answer a question. Those connections can be misconfigured, leading to vulnerabilities.
- **Agencies struggle to monitor AI applications for compliance in the cloud.** AI demands the hyperscale capacity of cloud infrastructure to process masses of data. When the AI processes interface with databases that perhaps are not cloud-certified, agencies risk falling short of their cloud compliance obligations.

"The right cloud provider is going to give you all of the tools needed to measure and remediate vulnerability within your application stacks"

— Charles Elliott, Google Cloud Solutions Engineer

SOLUTION:
Combining the Power of Search and Cloud

To deliver enhanced experiences using generative AI, government agencies need secure, world-class solutions that provide accurate, relevant and practical responses. A combination of search capabilities and a cloud service provider will help ensure accuracy, scale, relevance and security through:

- **Enhanced customer experience through conversational AI:** Semantic search enables the machine learning model to derive meaning from a user's query, beyond just keyword matching. In mental health, for example, it enables the contextual understanding to recognize "sad" (user's language) as another word for "depressed" (clinical language).
- **Leverage future-ready cloud technology with data privacy and security:** The right commercial cloud provider will have tools that identify the data being used in a generative AI model or application. It will flag anomalous behavior and leverage predictive analytics to ensure the confidentiality of private information.
- **Ability to monitor apps and infrastructure to meet compliance requirements:** With search-supported generative AI running in the cloud, agencies can be assured of the ability to meet their cloud compliance obligations. The cloud provider will offer tools to measure the cloud environment against various standards and regulations in order to ensure ongoing regulatory compliance, along with reporting to help guide remediation if needed.

Best Practices in Generative AI



Use Search to Unlock Insights

Vector search leverages machine learning to capture the unstructured data, including text and images, and transform it into a numeric representation. "It is the current state of the art in providing semantic search," MacKirdy said.

With hybrid search — a combination of traditional and vector search methods — the AI is better able to generate relevant, factual answers to users' questions. "Suppose the city or state legislature produces new legislation," MacKirdy said. In a constituent-service application, "hybrid search could boost the results of the new legislation, to make sure that individual is getting the most up-to-date and accurate information."



Obtain Relevant Responses with RAG

Retrieval-augmented generation (RAG) combines a retrieval model with a generation model, a technique that supplements text generation with information from private or proprietary data sources. This approach delivers grounded generative AI, a conversational experience in which the answers that the AI returns are derived from the right set of facts.

"Retrieval, for example, means utilizing secured, private data based on the role-based access and permissions of the user. Augmentation, then, is the pairing of that data with the user's question," MacKirdy said. "Together they support a large language model's work of building a response that is grounded in reality."



Prevent Cloud Misconfiguration with CSPM

In developing and deploying AI applications, you can identify and remediate risks using a cloud security posture management (CSPM) strategy. CSPM tools help detect and prevent misconfigurations and other threats that can lead to data breaches and compliance lapses.

"It addresses threat detection and analysis, giving you the ability to access very detailed levels of logging or threat data, so that you can identify anomaly behaviors," Elliott said.



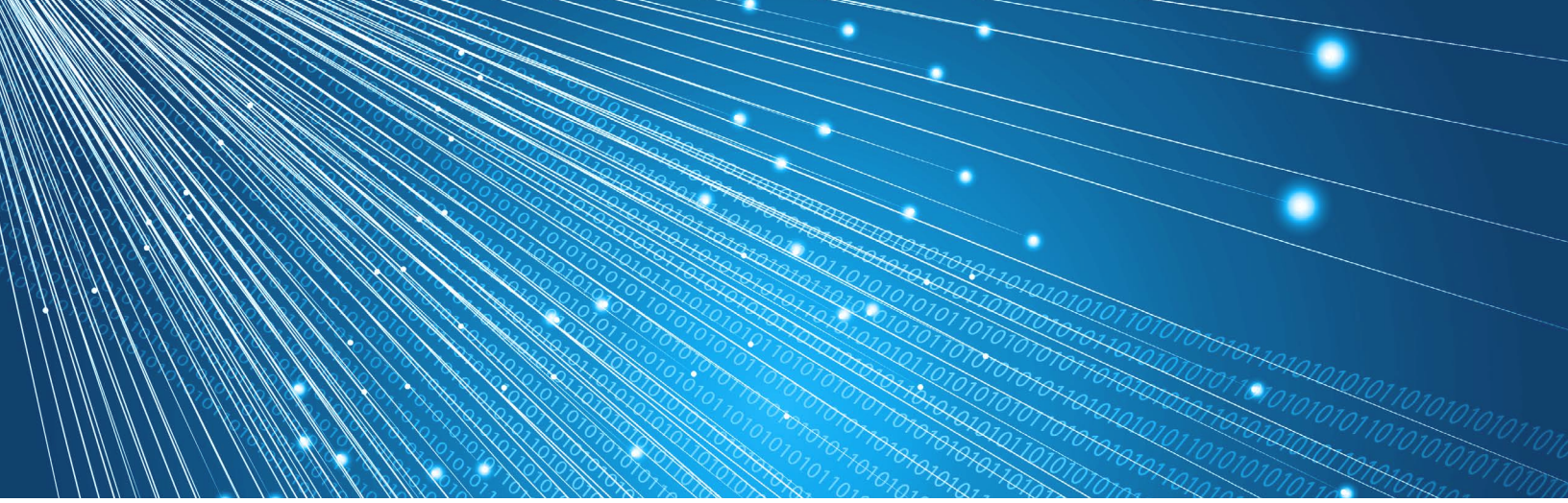
Continuous Evaluation of AI is Critical

Once the AI application is up and running, it's important to evaluate the end products, not just once but repeatedly over time.

"Create evaluating question-and-answer pairs for your use case. And make sure that that is evaluated on some frequency, to ensure it is not presenting any anomalous behaviors," Elliott said.

"If you feed AI training data that includes personally identifiable information, that data is potentially out there for everyone in the public domain."

— Sean MacKirdy, Area Vice President, SLED, Elastic



Use Cases for Government

Three government use cases help demonstrate the power of combining search and cloud in support of generative AI.

- **Improving the Constituent Experience**
  Suppose a New York constituent wants a business license for a general store that's going to sell beer and wine. They might not know they need a liquor license for that. Generative AI can help, working across silos to suggest the additional application, thus streamlining that constituent experience.

- **Improving the Employee Experience**
  Generative AI can make government employees more efficient and productive. HR teams, for example, can use AI for more effective onboarding, generating training materials that are specific to a given role. Employees themselves can get faster, more comprehensive answers to day-to-day questions as they use AI to query agency data.

- **Strengthening Security and/or Compliance**
  An AI assistant can significantly enhance an agency's cybersecurity posture. Junior analysts, for example, can use it to chat with a knowledge base, to ask questions when they see an indicator of compromise. Generative AI can effectively upskill that analyst in real time.

How Elastic and Google Cloud Can Help

Elasticsearch, the search analytics platform, harnesses search powered AI so that organizations can find and take action on all of their data, at scale, and in real time. Elastic is available in a Software as a Service model directly on the Google Cloud marketplace, with seamless integration with Google's Vertex suite of AI products, a partnership that offers robust security in cloud-scale architecture.

To learn more, visit https://www.elastic.co/enterprise-search/generative-ai

govloop elastic Google Cloud