CLOUD STORAGE SECURITY | aws

Securing Generative AI Models

Mitigating Risks and Protecting Your Business

Introduction

Generative artificial intelligence (AI) has exploded in recent years. The market for generative AI (GenAI) is projected to be valued at more than $36 billion by the end of 2024, and it's expected to exceed $350 billion by 2030. Only one year after its 2022 launch, OpenAI's ChatGPT reported 100 million weekly users.

Whether it's a foundation model, machine learning model, or a chatbot like ChatGPT or Copilot, GenAI is transforming modern life for businesses and individuals. Organizations are leveraging large language models (LLMs) and foundation models to enhance the customer and employee experience by building everything from chatbots and virtual assistants to completely self-service contact centers, analysis tools, data synthesis applications, and more. GenAI is also providing innovative ways to communicate and generate creative content such as images and social media posts that are personalized to target audiences.

Yet, along with technological advancements, GenAI is introducing the potential for cybersecurity threats across new attack surfaces. As enterprises adopt GenAI, one of the biggest challenges they face is understanding security for data models and preventing unauthorized disclosure of sensitive data. For example, what if GenAI chat output inadvertently reveals personally identifiable information (PII)? What would happen if the data used to train the model contains malicious payloads?

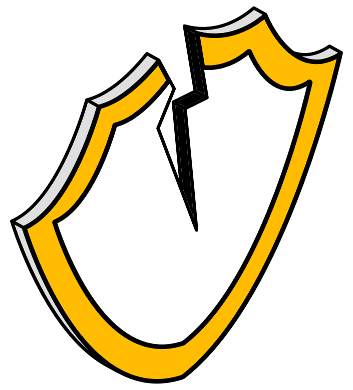
The Problem with Unscanned Data

The data used for an LLM-based AI application could compromise the system or expose sensitive information. If training data contains malicious code or sensitive information, it can create enormous risks for organizations when fed into GenAI technologies. Some recent examples highlight these risks:

- One in five UK companies has had potentially sensitive corporate data exposed via employee use of GenAI.
- A global electronics manufacturer banned its employees from using GenAI after discovering uploads of sensitive source code, possibly disclosing intellectual property to unauthorized users.
- ChatGPT experienced a data breach that exposed details of subscribers and their prompts to other users.
- The U.S. Space Force temporarily banned the use of AI tools because of data aggregation risks.
- 100 malicious models were uploaded to a widely used machine learning (ML) platform, potentially enabling attackers to inject malicious code onto user machines.

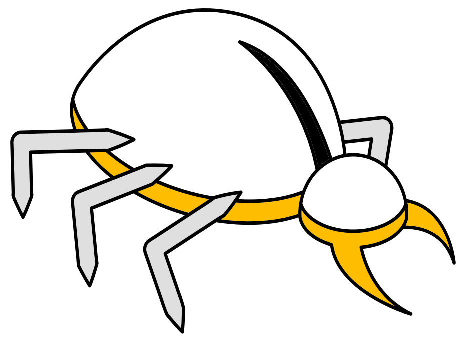
External data in AI models pose significant threats to businesses and their downstream data users. Data or model poisoning attacks, which pollute a machine learning model's training data, can impact a model's ability to output correct predictions. Additionally, "zero-day attacks"—the malicious events that exploit a vulnerability unknown to software developers—are major threats because many businesses lack the defenses to stop them. This kind of threat is enhanced by GenAI where malicious code hides within model data and activates against a weakness unknown to any party.

These examples underscore the importance of scanning data to ensure it is clean and that sensitive information is appropriately managed. Otherwise, unscanned data, whether it's ingested via models or shared via outputs, poses great risk for businesses. Specific challenges include:



Understanding data integrity

Infected models can create altered outputs and biased narratives that harm the business and mislead GenAI users and applications.

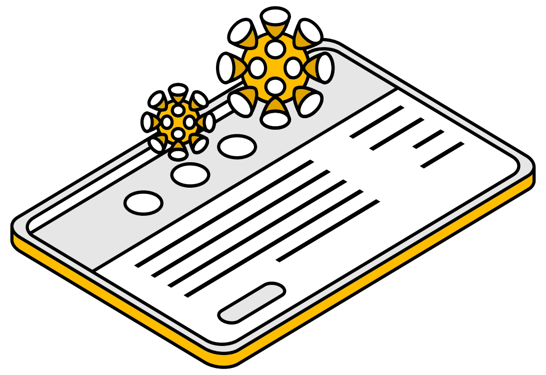


Malware intrusion via malicious third-party input data

AI models may be under threat from indirect attacks. For example, a malicious prompt could be injected into a model through a third-party API call or content received through retrieval-augmented generation (RAG).

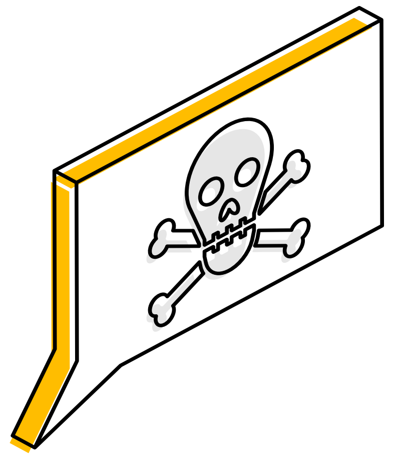
GenAI model output

Whether intentional or not, AI models can learn from the data around them and push content to unauthorized users, both internal and external. Model outputs can also be malicious, which creates significant risk for businesses.



Downstream data

AI models pose a major risk to downstream data use if the AI inputs, such as external data, are unscanned and unchecked. The inputs could contain malicious code, and models might leak data beyond the enterprise.

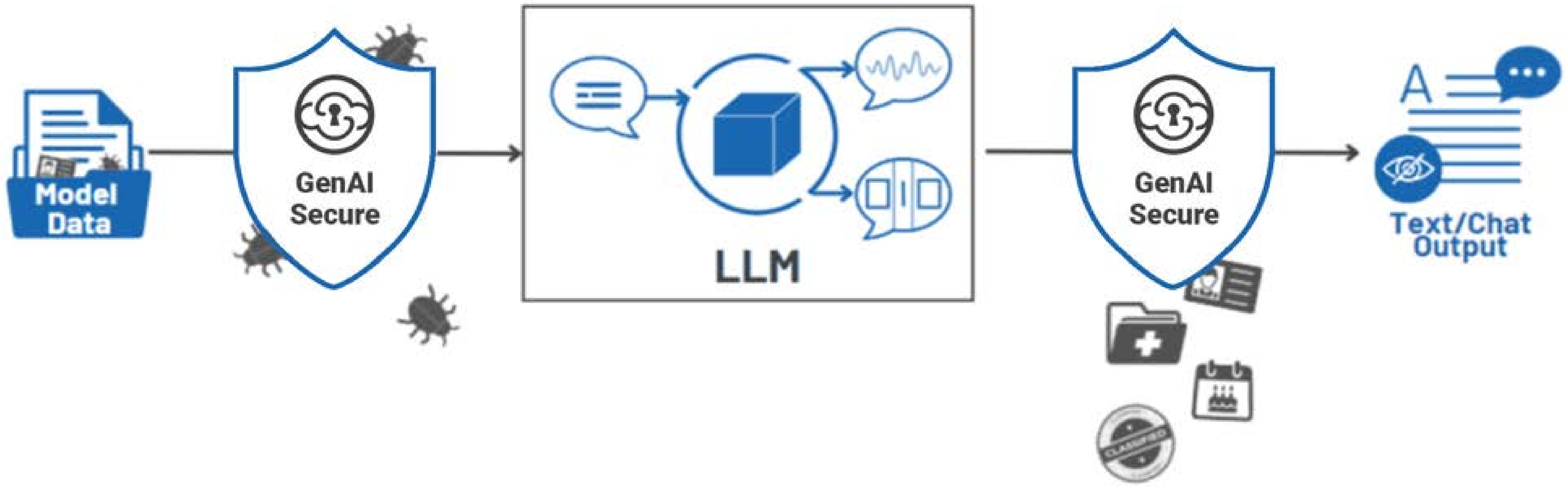


Sensitive information security

Chat or text outputs can unintentionally pass on confidential information to those who shouldn't be able to view it, both inside and outside the organization, if sensitive data isn't appropriately managed.

Overcoming the Challenge with Cloud Storage Security

To effectively protect applications integrated with GenAI, organizations need to scan all model data for malicious code as well as sensitive information beginning with training AI/ML models and ending with delivered output in text or chat served downstream. GenAI Secure by Cloud Storage Security (CSS) helps ensure you can safely use GenAI to improve processes and functions without jeopardizing the business and possibly exposing sensitive data.



GenAI large language model inputs can contain malicious code while data outputs can pass on sensitive information when they shouldn't. GenAI Secure by CSS addresses these challenges. Detect and quarantine suspicious files. Filter out sensitive data. Protect custom foundation models and GenAI applications.

CSS is a global company dedicated to solving the security and compliance challenges around data stored in the cloud. CSS is an AWS Public Sector Partner with an AWS Qualified Software Offering, AWS Security Competency, and an AWS Authority to Operate designation. Some quick facts:

>700
customers

>10 PB
of data located and scanned

>50,000
pieces of malware found

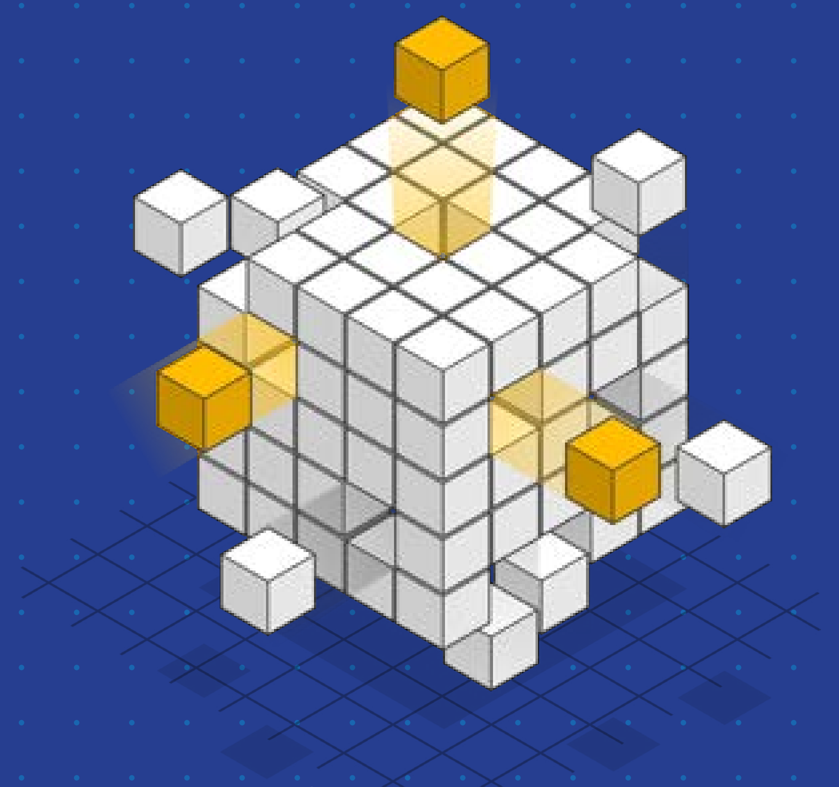
>50%
of customers find malicious code

GenAI Secure: Protecting Against Mounting AI Risks

GenAI Secure helps you protect both GenAI application model data and output, guarding against risks such as data poisoning and data loss, without disrupting workflows. Using GenAI Secure, you can protect against ransomware and safeguard sensitive data faster with improved visibility, control, and operational efficiency.

Financial Company Thwarts Payload

A financial services company is using third party model data to train its GenAI application, which is being built for internal financial data analysis. As a safeguard, the company integrated GenAI Secure into their input pipeline to scan the model data before it is used by their GenAI application. In doing so, the company uploads the data to an Amazon S3 staging bucket and GenAI Secure automatically initiates a malware scan. When the scan completes, malicious data is moved to a S3 quarantine bucket while clean data is moved to a S3 custom model bucket, which is used for training the FMs. This application of GenAI Secure quarantined numerous suspicious files, thereby preventing the transmission of potentially malicious data.



Built for and Powered by AWS to Secure AI Data

An AWS Partner since 2020, CSS designs solutions that integrate seamlessly with AWS services. GenAI Secure is built using AWS serverless architecture and runs within your AWS account, so data never leaves your environment.

GenAI Secure Features:

Easy in-tenant deployment

The solution installs within your AWS environment. Deployment via AWS CloudFormation template or HashiCorp Terraform module make it easy to get up and running.

Ability to scan over 300 different file types

Multiple detection engines are available, so you can scan more than 300 different file types of any size using one of three scan models: event-based, retro, or API. To scan data for sensitive information, leverage CSS to write your own custom rules.

AWS integrations

GenAI Secure integrates with services such as AWS Security Hub, AWS CloudTrail Lake, and Amazon EventBridge as well as your own security information and event management (SIEM) for consolidated notifications and reporting. Additional integrations include AWS Transfer Family for ingesting data and Amazon Bedrock to ensure GenAI data integrity.

Automated data security

GenAI Secure can protect GenAI models and output regardless of the application that's using it as long as cloud storage services such as Amazon S3, Amazon EBS, Amazon EFS, and Amazon FSx, are being used. Scan for both malicious code and sensitive data using GenAI Secure.

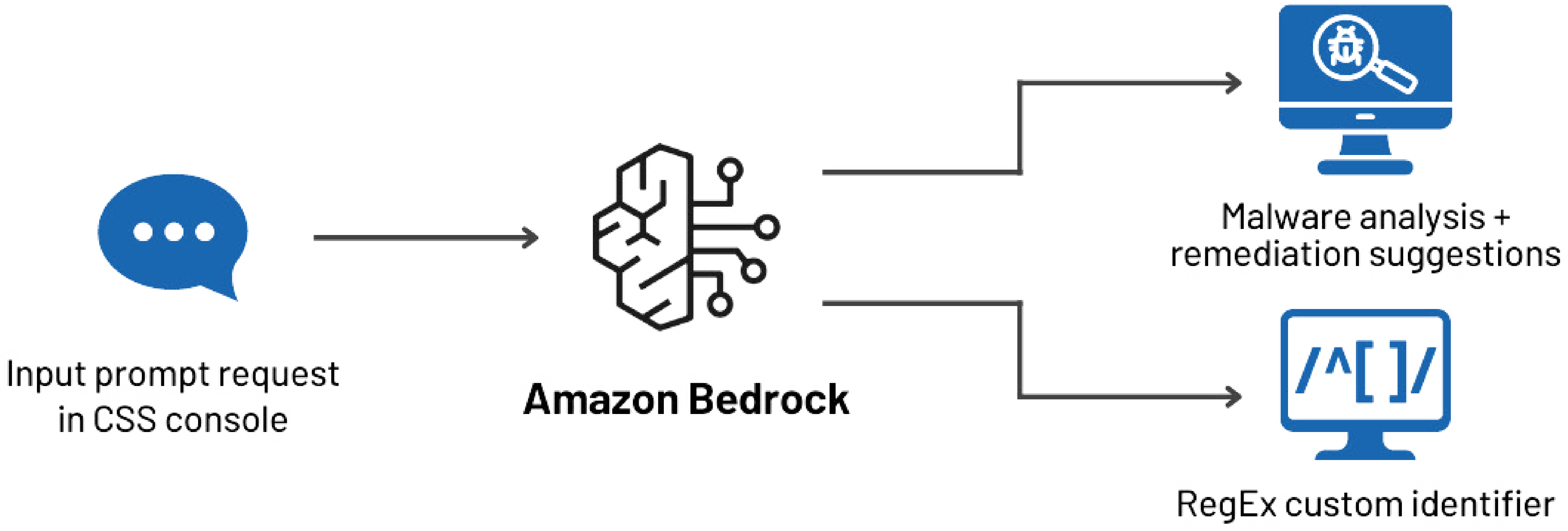
Built-in compliance

GenAI Secure is built for AWS using AWS Services in Scope of AWS assurance programs, which include control mappings that help customers create compliant AWS environments (e.g., HIPAA, FedRAMP, PCI). Additionally, GenAI Secure is available in AWS GovCloud.

Integration with Amazon Bedrock

GenAI Secure integrates with Amazon Bedrock, a fully managed service that offers a choice of high-performing foundation models through a single API, along with capabilities to help build GenAI applications. By integrating with Amazon Bedrock, GenAI Secure enables you to:

- Analyze data that's been flagged and quarantined as suspicious. Take action through forensic analysis and malware remediation
- Protect chat and text output with custom regular expression policies that are written for you simply by entering a prompt.
- Automatically protect PII like social security numbers and credit card numbers from exposure.


Input prompt request in CSS console
Amazon Bedrock
Malware analysis + remediation suggestions
/^[ ]/
RegEx custom identifier

Blocking malware, keeping sensitive data protected from exposure and writing policies to prevent data loss ensures your data remains safe. GenAI Secure's Amazon Bedrock integration amplifies the capability of your security team allowing your organization to comfortably leverage AI models and enhancements like Amazon Bedrock to move business forward.

Healthcare Company Stays HIPAA Compliant

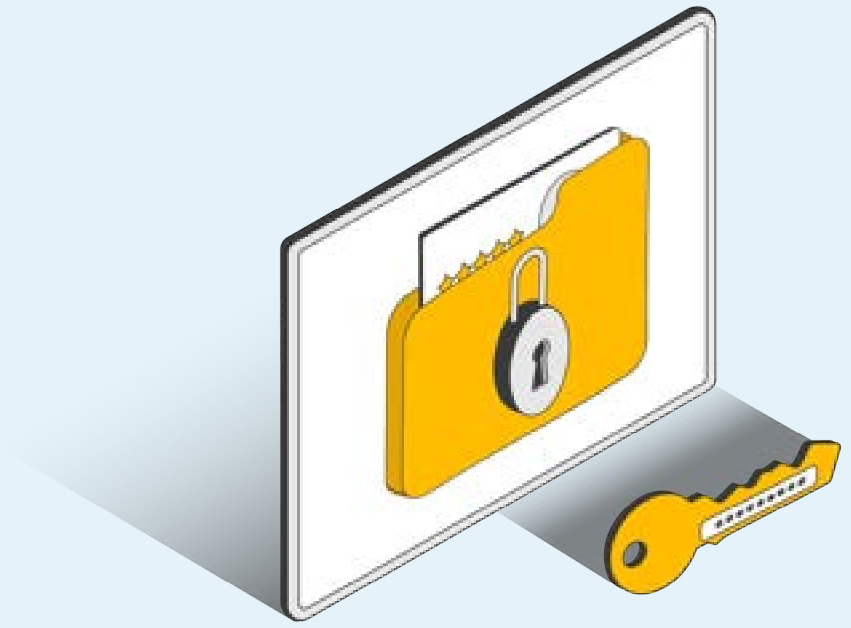
To ensure compliance with HIPAA, a healthcare company needs to prevent its AI chat bot from inadvertently sharing individually identifiable health information. So, they implemented GenAI Secure to scan the chat output. To do so, the chat output is sent to an Amazon S3 output bucket before it's delivered to the end user. GenAI Secure automatically kicks off a Lambda function to make HTTP POST calls to an Application Load Balancer, which distributes data across multiple API endpoints that automatically scale for sensitive data scanning. When the scan is complete, the API endpoint uploads non-sensitive data to an S3 application bucket, which is used to feed the downstream GenAI chatbot. This prevented a data breach and ensured the protection of sensitive customer data.

Protect Model Data Inputs and Application Outputs

CSS offers a variety of real-time, scheduled and on demand scanning models that help you ensure your data is clean no matter when it comes into the system. This means you can filter infected files and sensitive data out before they are used or delivered. Proactively scanning data and outputs protects downstream users and future models that may be trained on the same data or built from the same models.

Protecting GenAI model data inputs

For securing model data, malware scanning is best applied before the data is used to create or enhance models. This ensures that malicious code cannot disrupt services or alter the functions and processes of the AI model. Additionally, you can restrict the use of sensitive data within your models by removing that information before the data is fed to the application.



Protecting GenAI application outputs

If sensitive data is used to train your model, but you need to prevent sharing confidential information via the chat/text output produced by your GenAI application, GenAI Secure helps you classify the data before it is delivered to the end user so that you can restrict sharing sensitive data. This ensures that both internal and external users are only served data that meets their authorized clearance.

GenAI Secure Benefits

Completing a security risk assessment in 24 hours

A CSS customer completed a baseline scan of more than 120 million existing objects in a few hours. They were able to scan all objects within 24 hours.

Stopping malware

After scanning more than 300 million objects, a CSS customer discovered over 1,400 malicious files capable of disrupting business operations.

Eliminating dozens of hours of monthly maintenance

A customer used CSS automation to eliminate dozens of hours of maintenance time while still meeting real-time scan requirements.

98 percent reduction of resources

CSS shortened a customer's security review process from two days to one hour by automatically scanning terabytes of data across multiple AWS accounts and streamlining reporting in a single, centralized dashboard.

50 percent less expensive

By using CSS's custom pricing and AWS Fargate container-based solution, a customer's total cost of ownership was 50 percent lower than other solutions.

Protecting privacy

Customers are ensuring the right data is served to the right users.

Improving security

Customers implementing a private deployment meet compliance and security requirements because the data doesn't leave their environment to be scanned, which eliminates additional entry points for malware.

Reduced false negatives

With a single antivirus engine, research suggests you can typically decrease false negatives to 0.6 percent. Using a multi-engine approach in which the virus detection capabilities are independent of one another, you can bring the number down further—combining two engines reduces false negatives to 0.0036 percent; four engines to 0.000000002196. percent. GenAI Secure offers multiple engines that can be used together for increased efficacy.

Rapid scalability

You can scale faster by using multiple agents across a growing data pool. Because GenAI Secure uses AWS serverless architecture, spinning up multiple engines is simple and fast.

Malware and data loss prevention capabilities within a single stack

This means you can protect both custom models and GenAI model output.

Remediation recommendations

GenAI Secure quickly finds and flags data and provides best practices for how to respond, whether that means quarantining or deleting the data.

Conclusion

With the proliferation of GenAI technologies comes the potential for new cybersecurity threats that are creating significant risks for organizations today. The Cloud Storage Security platform, built on AWS, empowers organizations with cost-effective, automated solutions that help prevent breaches, manage sensitive data, and meet compliance requirements.

GenAI Secure by Cloud Storage Security, leverages years of experience in stopping malware and protecting data to help you protect both GenAI model data and output. Easily guard against risks such as data poisoning, data loss, and sensitive data exposure automatically so you can focus on letting secured GenAI applications transform your business.

Secure your GenAI model and application data

Don’t let GenAI become a security risk for your organization. Secure your models today!

Contact Us to Request a Demo

Reach out to our team today to find out just how easy it is to proactively secure your GenAI data and safeguard your organization against a critical security breach.

CLOUD STORAGE SECURITY | aws