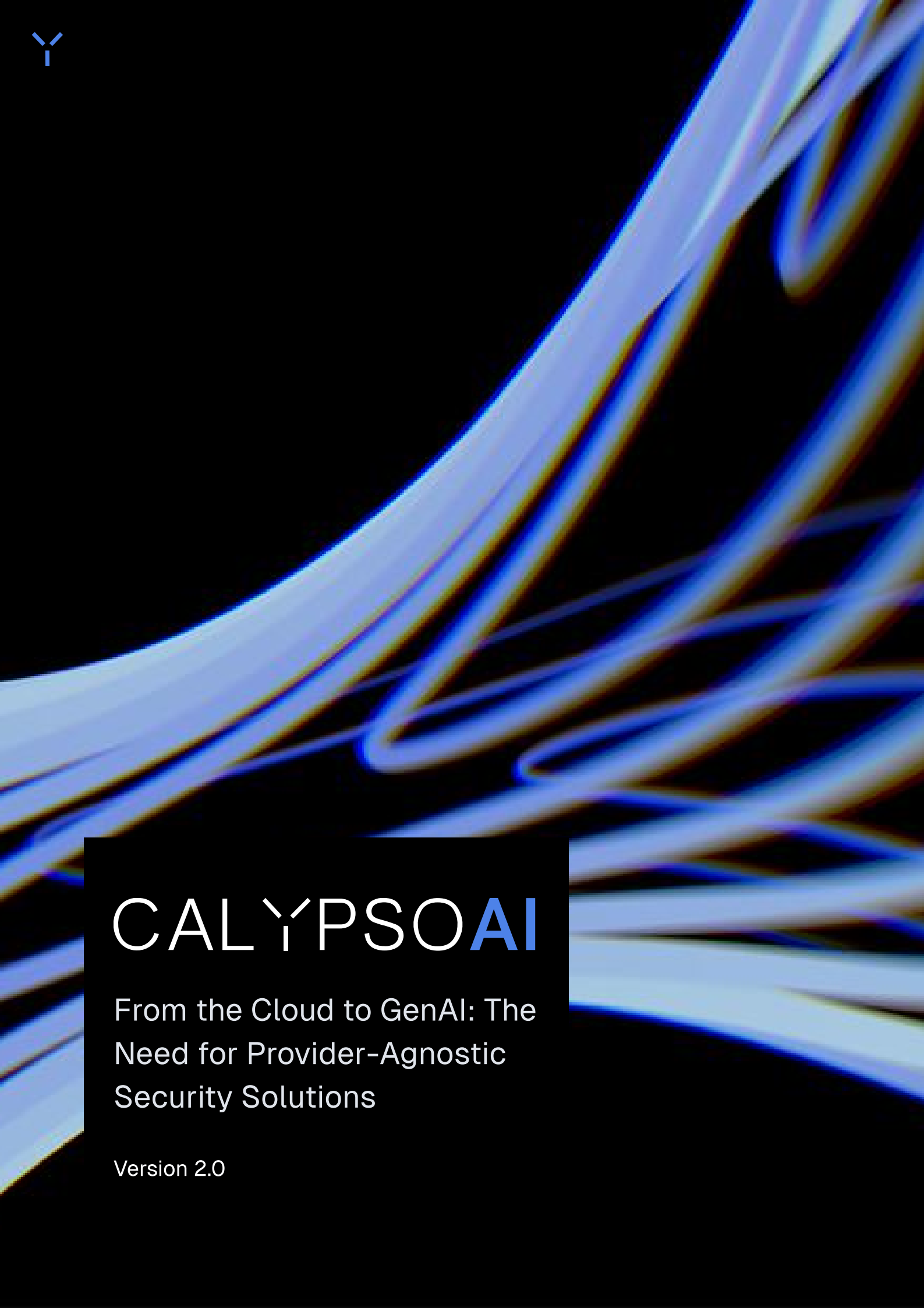


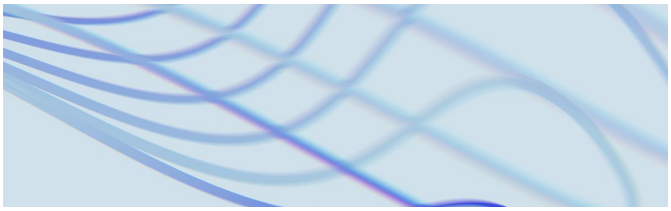
CALYPSOAI

From the Cloud to GenAI: The Need for Provider-Agnostic Security Solutions

Version 2.0

In just over one year, the transformative power of generative artificial intelligence (GenAI) models has rivaled that of the light bulb or combustion engine. AI has become part of the fabric of enterprise, with 69% of surveyed organizations reporting at least one AI project in production and 64% of businesses believing AI will substantially increase productivity. Large language models (LLMs), a subset of AI designed to comprehend and generate human-like text, have been instrumental in this widespread adoption. Based on a transformer model architecture, LLMs can understand context, answer questions, and classify and summarize text, improving customer interactions, data analysis, and decision-making.

For all the benefits that GenAI has brought to the business landscape, it has introduced an equally impressive set of challenges, specifically myriad new attack vectors and just as many bad actors willing to exploit them. Organizations' rapid adoption and reliance on GenAI models has vastly expanded the attack surface. This has been compounded by the large and rapidly growing number of models in use across the AI ecosystem and the number of providers offering models, none of which allow visibility into their systems—or vulnerabilities.



64% of businesses believe AI will substantially increase productivity
–Forbes

If this situation sounds familiar, it ought to. Fifteen years ago, the advent of the cloud ushered in a new computing era. While cloud service providers' early capabilities included only simple web hosting and data storage, hyperscale cloud providers like Amazon Web Services (AWS), Microsoft Azure, and Google Cloud Platform (GCP) quickly began offering more complex solutions, including Software as a Service (SaaS), Platform as a Service (PaaS), and Infrastructure as a Service (IaaS) to organizations of every size and configuration.

This paradigm shift from businesses exclusively running their own infrastructure to one in which they moved their data to third-party cloud providers gave rise to the need for provider-agnostic solutions that could provide security and visibility. And the cloud security industry came into being.

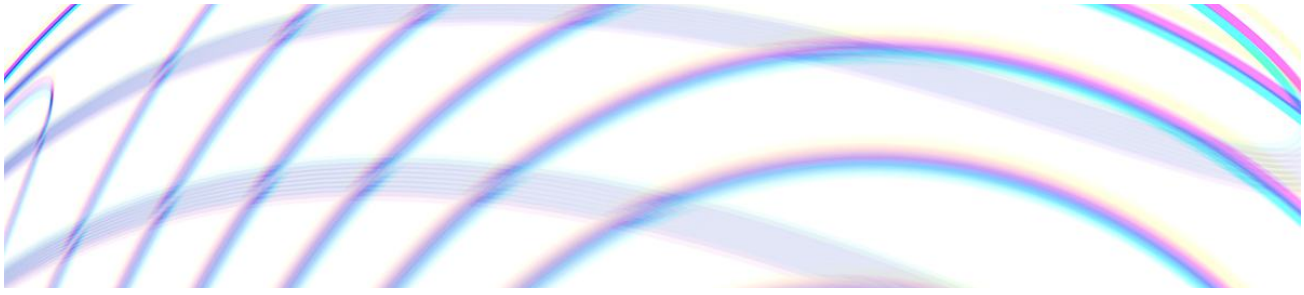
Security Redux: From the Cloud to GenAI

The 2010s saw a shift in infrastructure as enterprises adopted the cloud for computing, storage, and content delivery via SaaS, PaaS, and IaaS solutions rather than managing and maintaining all their data in their own data centers. Doing so allowed them to take advantage of the scalability, flexibility, and cost-efficiency cloud service providers offer, as well as the ability to run in multiple regions, which lowered latency.

However, as organizations began their cloud migration journey, they quickly sought increased control and realized that depending on one cloud provider like AWS or GCP wasn't enough. By 2020, 93% of cloud users employed a multi-provider strategy, which made fulfilling their part of the Shared Responsibility Model, which divides security responsibilities between providers and customers, increasingly difficult.

This framework arose when customers sought guidance regarding cloud security measures and their responsibilities in the face of various regulatory requirements and increasingly complex cloud service offerings. The Shared Responsibility Model delineates which aspects of security cloud providers handle and which parts customers are responsible for, enabling enterprises to better protect their data and manage security incidents. Not only did organizations like the Cloud Security Alliance (CSA) introduce frameworks that divided security responsibilities, but cloud providers like AWS, Microsoft Azure, and GCP also formalized the Shared Responsibility Model.

This shift toward multiple providers introduced security, data governance, and management challenges, spanning access control issues to data ownership challenges. It also meant organizations needed broad, provider-agnostic security solutions capable of providing visibility into resource utilization, user requests, and traffic, as well as managing the complexity associated with hybrid-cloud and multi-cloud strategies. However, even if an organization uses only one cloud provider, it still needs a robust cloud security solution.



Reasons Cloud Security is Necessary	
The Shared Responsibility Model	According to the Shared Responsibility Model, the cloud provider is responsible for securing the database, software, edge locations, and hardware. The customer must safeguard their data, platform, applications, access, and network, firewall, and operating system configurations within the cloud environment.
Fine-Grained Controls	Access controls enable enterprises to specify who can access which resources, effectively detect and respond to security threats, and uphold their part of the Shared Responsibility Model. For example, an organization can use its cloud security solution to monitor user activities in real time, gain comprehensive visibility across their cloud environments, view detailed logs, track configuration changes, and take advantage of advanced AI-driven threat detection and automated compliance assessments.
Multi-Cloud, Hybrid Cloud, Private Cloud, and Diverse Geographies	Many organizations use a combination of multiple cloud providers, private clouds, on-premise solutions, and multi-region strategies, but maintaining adequate visibility and control in such diverse environments can be challenging without a cloud security solution. The Shared Responsibility Model becomes more complex when a third-party software provider, such as a SaaS provider, is present. • The cloud-hosting provider, for example AWS, has a relationship with the client company that includes the shared security obligations. • The SaaS provider has its own relationship with AWS that involves resiliency and security obligations on both sides. • The client has a relationship with the SaaS provider with its own set of mutual obligations. This relationship grows in complexity as the number of third-party providers increases within this relationship entity.

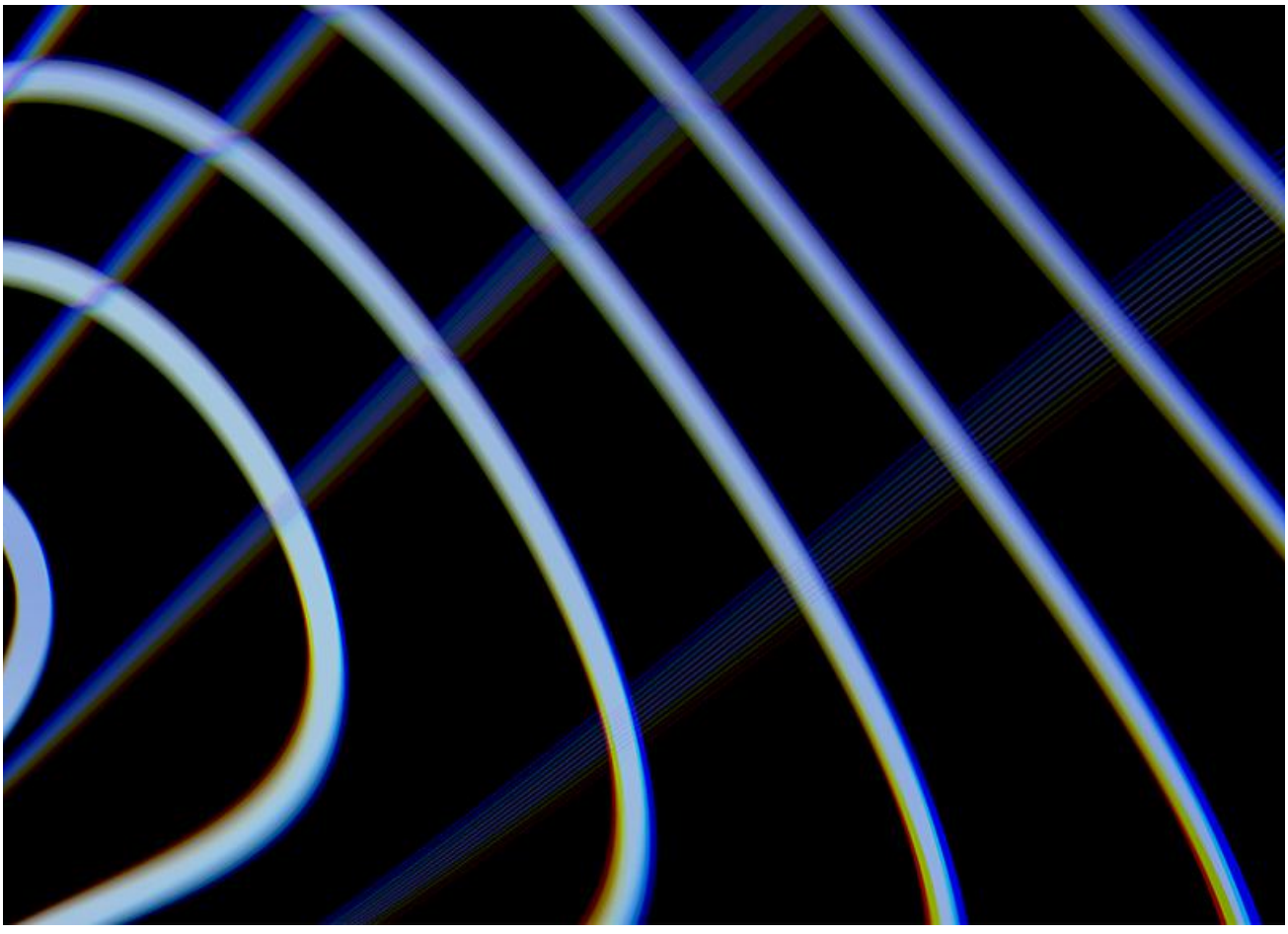
The GenAI security industry mirrors this evolution. Just as firms embraced multiple cloud solutions, they have adopted multiple and multimodal models. Companies are not just integrating apps with these models for internal use, such as Search, they are incorporating models into external apps to provide customer support chatbots, sentiment analysis, and language translations, and are fine-tuning external models or developing internal models. As firms shared security responsibility with cloud providers, they must share responsibility when using solutions from model providers.

GenAI Security: A Must in Modern Times

As enterprises are rapidly scaling AI, they are also:

1. Building their own internal models for information retrieval and retrieval-augmented generation (RAG), an AI framework designed to retrieve factual information from external knowledge bases.
2. Integrating apps from SaaS providers, from Salesforce to Grammarly, that have also incorporated LLMs into their solutions.
3. Relying on external base models, such as LLaMA 2, ChatGPT, and Gemini.
4. Developing or deploying agentic AI models that bring autonomy and the ability to adapt and evolve to make decisions and execute complex and dynamic tasks.

Organizations need a comprehensive, provider-agnostic solution to establish uniform security standards across multiple models and gain observability and control beyond what individual model providers can offer. Without one, establishing protective measures, safeguarding data, and complying with regulations becomes nearly impossible.



Indispensable Features for Effective GenAI Security

Since the Shared Responsibility Model also applies to GenAI, enterprises require a tool with the following capabilities:

The Ability to Work with Internal and External Models

Enterprises experiencing increased dependence on multiple, siloed internal and external models, as well as multimodal and other types of models, will incur fragmented, inconsistent security controls without an independent, model-agnostic trust layer as part of their security framework. Such a solution provides centralized model management and visibility, ensuring the same security protocols apply to every model. This unified security strategy results in fewer vulnerabilities, optimized resource allocation, and improved adherence to policies and regulations.

Key Controls for LLMs

Observability

To have sufficient visibility across the GenAI security infrastructure, make informed decisions regarding models that warrant continued investment versus which to discontinue, and maintain control of model usage, companies need a security solution that offers insights into which groups or individuals use each model, when, how often, and for what purpose.

Policy-Based Access Controls (PBAC)

In traditional data management systems, PBAC allow organizations to set granular access permissions, for instance, specifying which users can access specific documents in a Google Drive directory.

However, in environments deploying one or more GenAI models, the concept of PBAC faces a challenge. A model fine-tuned on an organization's entire data repository doesn't differentiate between sensitive and non-sensitive data in the same way that file permissions do. If a large or indiscriminate group has access to such a model, the company runs the risk that unauthorized employees could have visibility into sensitive information, such as payroll records.

To prevent such exposure, companies can develop and deploy several models, each trained on data meant for specific roles or departments. For instance, a general-use model might be trained on routine business data for widespread employee access, while another, containing sensitive financial information, is accessible only to the Finance team. This permissioning strategy improves security and ensures adherence to privacy regulations and corporate protocols, enabling the safe use of models without compromising data integrity.

- **Customization Options**

Each organization is unique and deserves a solution that can adapt to its needs. For example, customizable scanners can be adjusted to accommodate specific data types, compliance mandates, and security protocols, and narrowly targeted, bespoke scanners can be created to further expand capabilities aligning the security tool with an organization's security and management requirements.

- **Human Verification of Model-Generated Content**

AI-powered tools can generate inaccurate or out-of-date information and outright hallucinations, leading to misinformation and harmful content or data dissemination. Humans should verify all model output for accuracy and relevancy before using it in organizational documentation.

- **Auditability**

Auditability, the ability to track and review activities and changes, is vital for model and enterprise security because it preserves user and administrator activity. Security solutions capable of providing detailed team and individual data offer essential information regarding resource allocation, content, security threats, compliance issues, and model performance.

- **Scanning Capabilities**

Organizations require robust data scanning and filtering to maintain control, prevent the dissemination of harmful or sensitive content, ensure adherence to regulatory standards and internal acceptable use policies, safeguard privacy, and cultivate user trust. Scanning allows administrators to monitor prompt inputs and generated responses, while filtering content based on sentiment, toxicity, and the presence of banned terms. This allows administrators to manage the information flow to create a secure and compliant environment.

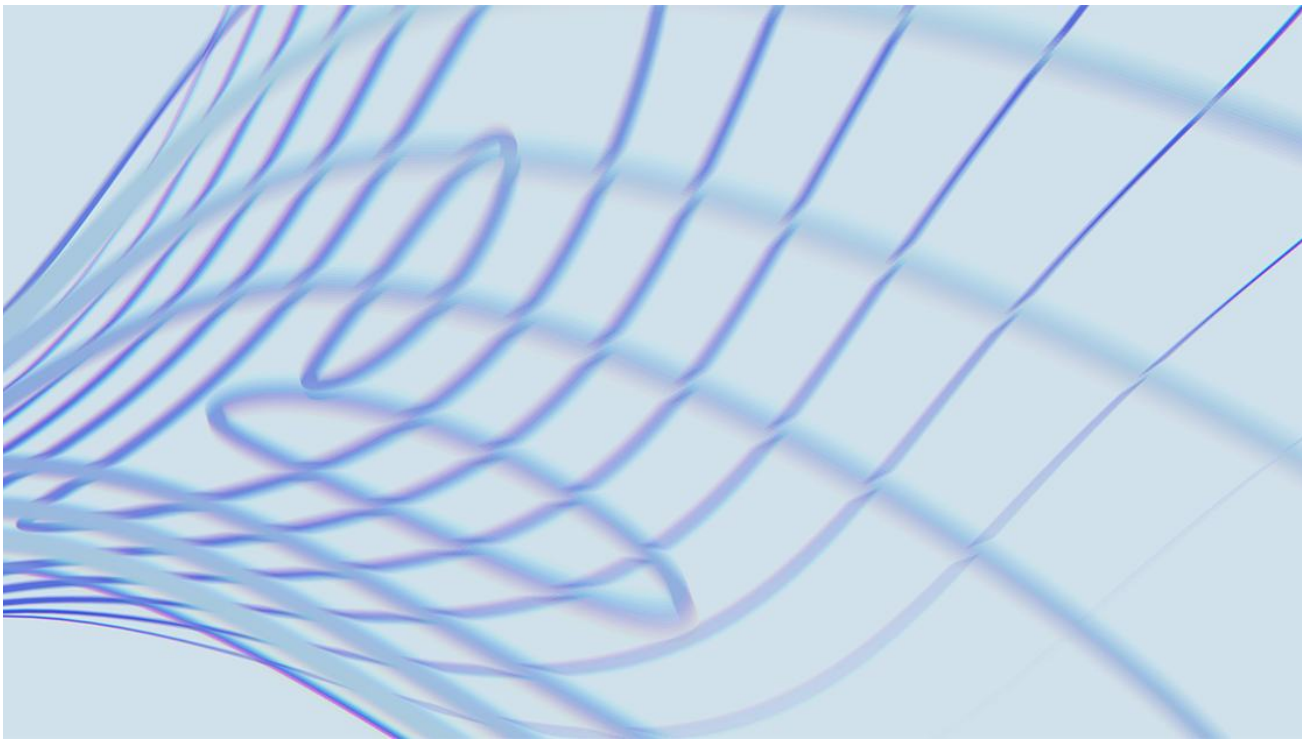
Regulatory Guidance

As organizations deploy GenAI models, they must consider existing and impending regulations, such as the EU AI Act, including AI-specific standards and directives, such as the Biden administration Executive Order.

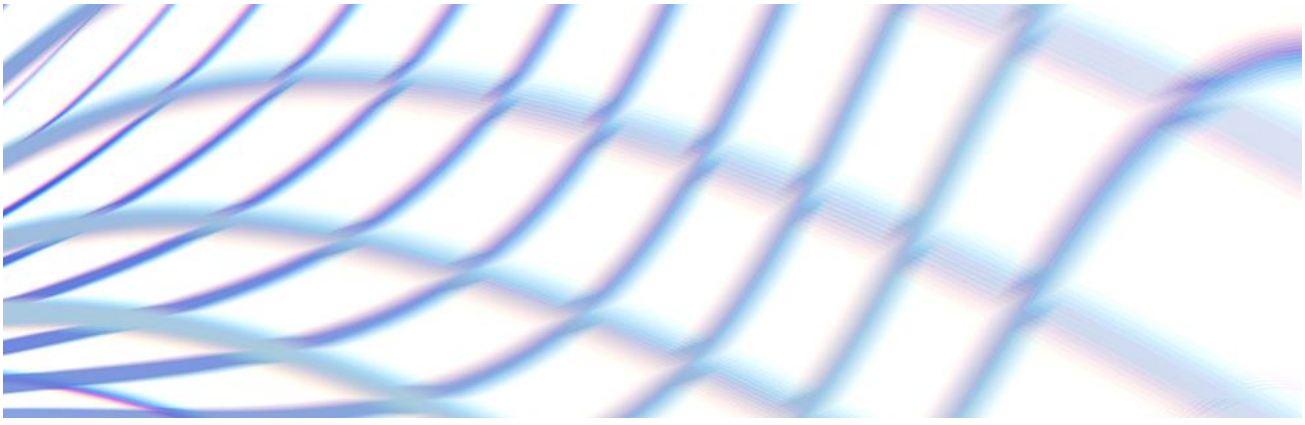
30th October, 2023 White House issues Executive Order on Safe, Secure, and Trustworthy Artificial Intelligence

Regulations affecting model usage often require diverse deployment strategies to comply with privacy and data protection standards, which can lead to an increase in the number of models an organization must manage. For instance, data residency requirements might necessitate hosting models across multiple locations to conform to local data sovereignty laws, which, in turn, expands the overall enterprise attack surface by adding more potential points of vulnerability.

This evolving regulatory environment is leading organizations to adopt more sophisticated security solutions. By proactively enhancing security measures, businesses can ensure and maintain compliance and future-proof themselves against new or updated regulations, while protecting their operations against emerging threats in an increasingly AI-integrated world.



Regulations and Recommendations	
Health Insurance Portability and Accountability Act (HIPAA)	This U.S. law requires organizations to protect sensitive health information from being shared without patients' knowledge or consent. However, it's not just the underlying data that must be stored in a HIPAA-compliant way. Organizations should also ensure models that include such information are secured and only provide data access to the appropriate personnel.
General Data Protection Regulation (GDPR)	This EU data protection and privacy regulation was created to safeguard individuals' data and give them more control. Organizations that monitor EU individuals' behavior, offer goods and services to EU residents, or have EU branches that process personal data must conduct information audits, use appropriate measures to protect data, and make it easy for customers to request their personal information on file, update inaccurate information, or delete it and prevent future data processing using it. This regulation also affects how data can be transferred from the EU, which includes data stored in AI systems.
Artificial Intelligence Act (EU AI Act)	The EU Artificial Intelligence Act is the world's first comprehensive AI legal framework, establishing standards for AI development and governance to prevent harmful outcomes. When the EU Artificial Intelligence Act goes into effect, cybersecurity will need to be a core component of all models.
Risk Management Framework Standards (RMF)	Since AI-powered systems process vast amounts of data, it is recommended that they be configured, managed, and monitored in alignment with the U.S. National Institute for Standards and Technology (NIST) framework's security and privacy guidelines. Doing so ensures that model usage doesn't inadvertently introduce vulnerabilities or expose sensitive information, preserving the system's integrity and users' trust.
Executive Order on Safe, Secure, and Trustworthy Artificial Intelligence	This Biden administration executive order identified NIST as the entity that will establish guidelines and best practices for "develop[ing] standards, tools, and tests to help ensure that AI systems are safe, secure, and trustworthy," further emphasizing the importance of robust cybersecurity solutions for GenAI models.



CalypsoAI

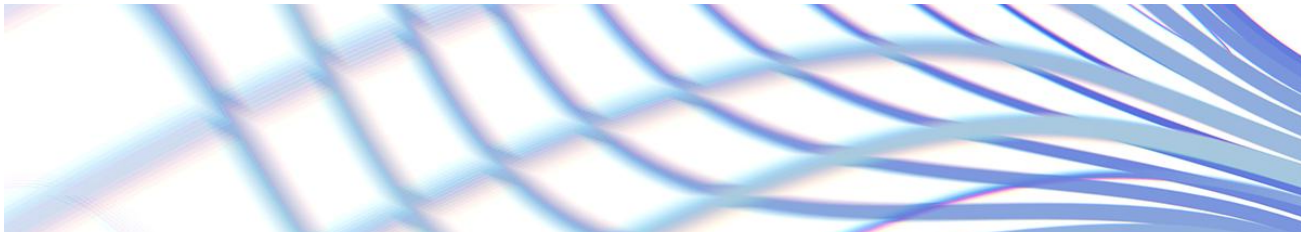
As enterprises continue to adopt and operationalize GenAI at an increasingly rapid pace, the need for security solutions that enable safe, efficient, effective, and responsible model use through features like security controls, observability, auditability, filtering, customization, PBAC, and human verification becomes vital.

CalypsoAI's API-accessible, SaaS-driven, Security and Enablement Platform is the only solution available today that offers the scope and scale of protections, innovations, and productivity enhancements organizations need.

CalypsoAI's holistic and model-agnostic approach to GenAI security provides the guardrails, visibility, and controls necessary for managing and protecting models and their data. The platform offers:

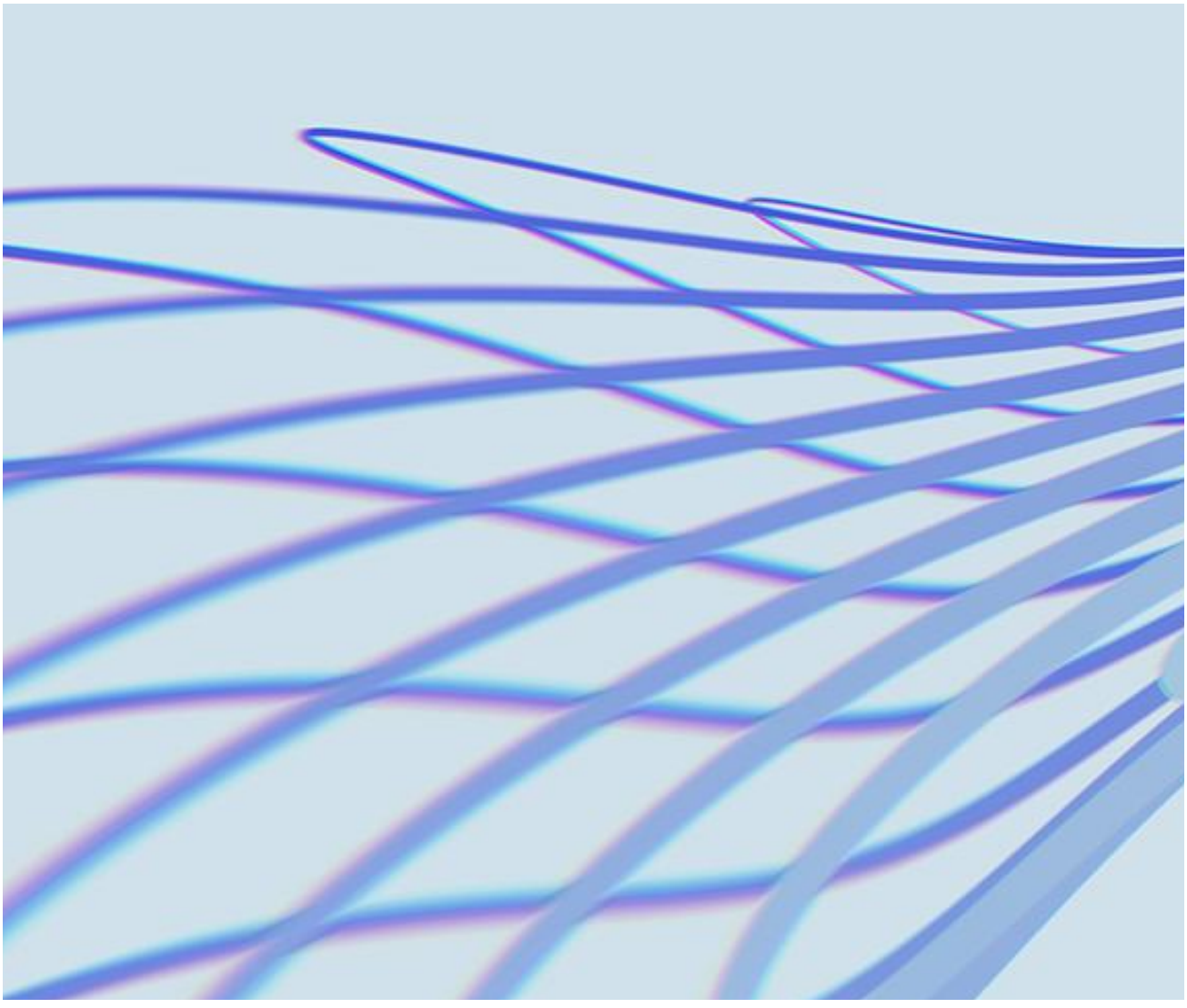
Feature	Description
Auditability and Monitoring	CalypsoAI records each user and administrator interaction with each model, providing full transparency and accountability. A detailed, interactive dashboard allows administrators to review usage analytics at the organizational, group, and individual levels. The platform retains a complete history of every prompt, whether blocked, redacted, or sent, and every response. The Prompt History includes results for each scanner applied. Retention options include indefinite retention, no retention, and setting a manual or automated purge cadence.
Integration with Workplace Chatbots	This first-of-its-kind model-agnostic bot seamlessly integrates with Slack and Microsoft Teams, optimizing AI functionalities while ensuring personal or corporate privacy. This tool supports external, proprietary, and fine-tuned models, which provides flexibility and compatibility across different environments with security features that go beyond basic encryption. It incorporates robust auditing and RAG assessments to ensure all operations adhere to strict security parameters. Every interaction within the communication platform is monitored and assessed to prevent potential security breaches and ensure sensitive data remains protected.
Customizable Scanners	Administrators can set thresholds and sensitivity for more than one dozen customizable scanners, enhancing security and responsible LLM usage. Settings can be applied globally, by group, or by LLM.

Feature	Description
LLM-Powered Proprietary Guardrails	Administrators can create and deploy detailed, bespoke LLM-powered AI guardrails (LPGs) with the content and thresholds tailored to the unique requirements of their enterprise.
Granular Access Control	Administrators can create numerous groups and define the access each has to specific models by configuring the PBAC through a user-friendly interface. This ensures only authorized personnel will have access to identified models and data.
Model Agnostic	The asynchronous multi-model and multimodal software enables businesses to diversify their solutions and address their needs efficiently by deploying any type of GenAI model across their enterprise without dealing with vendor lock-in issues.
Robust Security and Compliance	CalypsoAI's centralized security control plane provides customizable audit functionality and parameters via the broadest set of customizable scanners on the market, ensuring companies enhance their regulatory compliance and data protection capability.
360-Degree Observability & Visibility	The ability for oversight of every AI model on the system drives efficiency and transparency across the enterprise. Firms can optimize operations, enable informed decision-making, enhance security, cultivate trust, and maximize the benefits of GenAI while minimizing risks.
Effortless API Integration	Streamlined API integration allows seamless connectivity, enhancing efficiency and productivity, and accelerating implementation. Adoption requires entering only a few lines of code.
Simplified Model Configuration	Configuring and integrating self-hosted or external models, from ChatGPT to Gemini and others, and internal, fine-tuned, and RAG models is fast and straightforward.
Seamless Integration with Directory Services	CalypsoAI is compatible with Active Directory, Google Accounts, Lightweight Directory Access Protocol (LDAP), and OAuth.
Real-Time Generative AI Testing and Deployment Optimization	Our platform enables model evaluation and refinement before deployment, enhancing AI application reliability and accuracy, optimizing resource utilization, accelerating development cycles, and nurturing innovation by facilitating a test bed for rapidly iterating and improving model performance.



In an era in which GenAI plays an increasingly prominent role in every function and business unit across the enterprise, the path to integrating GenAI into daily business operations is fraught with security challenges, but they are surmountable. CalypsoAI's security and enablement solution is vital for safeguarding data and ensuring responsible AI usage, and ensuring businesses can safely and securely leverage the power of GenAI to transform their operations. This tool provides a trust layer atop all models deployed across an organization to provide visibility and security well beyond what individual model providers offer.

With CalypsoAI, enterprises have observability across their stack and the ability to compare models' usage and performance, enabling them to make data-driven decisions regarding everything from resource allocation to compliance improvements. Our platform offers the safeguards, monitoring, fine-grained controls, and customization features needed to ensure compliance, understand and regulate user activity, and protect enterprises from security threats.



Learn more about CalypsoAI today!

CalypsoAI is the leader in AI Security and Enablement. As a trusted partner and global leader in the AI Security domain, CalypsoAI empowers enterprises and governments to leverage the immense potential of GenAI solutions and LLMs responsibly and securely. CalypsoAI strives to shape a future in which technology and security coalesce to transform how businesses operate and contribute to a better world. For more information about our API-accessible, SaaS-driven Security and Enablement platform, please visit CalypsoAI.

CALYPSOAI