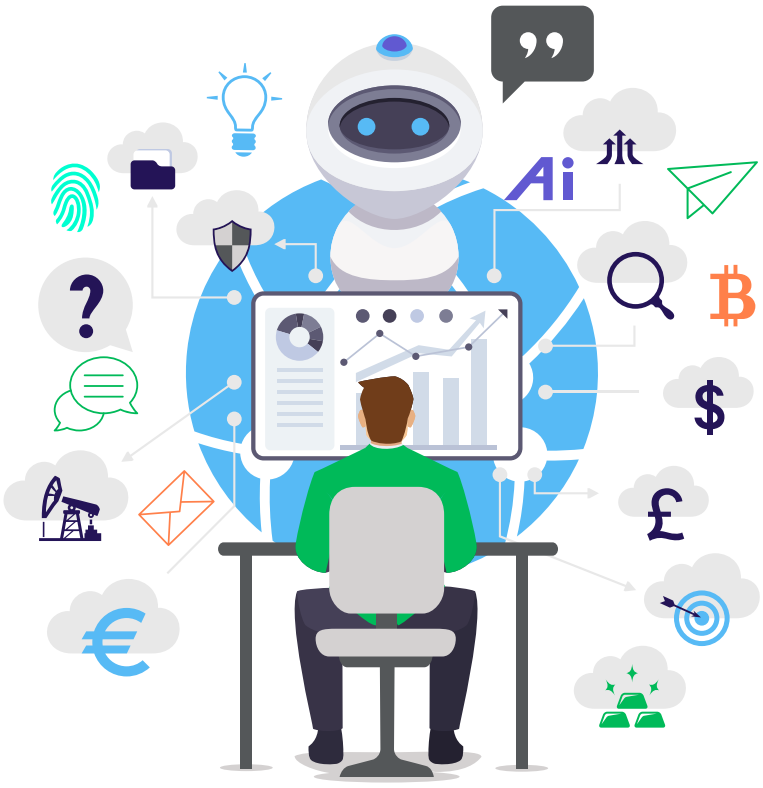
Delinea

Securing the Use of Generative AI Technologies

WHITEPAPER

Securing the Use of Generative AI Technologies

As Generative AI (GenAI) technologies continue to evolve, their potential to transform industries is undeniable. However, this rapid innovation comes with significant security, privacy, and compliance risks. Organizations must adopt a comprehensive approach that integrates governance, technology, and adaptive security measures to ensure the safe, responsible, and ethical use of GenAI. This whitepaper outlines Delinea's strategic framework for securing GenAI, empowering organizations to harness its potential while mitigating risk.



Establishing a strong AI governance framework

Effective GenAI security begins with robust governance. A well-structured governance framework ensures that AI initiatives align with organizational values, comply with evolving regulations, and proactively address ethical considerations.

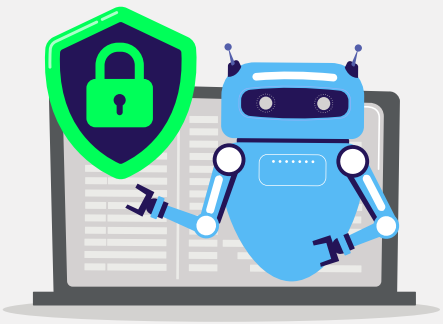
Key elements:

- **AI Governance Committee:** Create a cross-functional team to oversee AI projects, manage GenAI tool usage, and ensure compliance with global standards (e.g., EU AI Act, OECD AI Principles, GDPR). This committee should include representatives from security, legal, compliance, IT, and business units to provide diverse perspectives and holistic oversight.

- **Compliance & Ethics:** Define ethical use expectations to guide responsible AI deployment. Conduct regular impact assessments to evaluate organizational, economic, and societal impacts, while identifying potential risks such as algorithmic bias, unfair outcomes, or harmful misuse. Set clear guidelines that include:
  - Establish clear ethical standards that align with international AI principles, such as those from the OECD and UNESCO.
  - Ensure AI decision-making processes are understandable and explainable.
  - Mitigate biases in AI models, particularly in high-stakes applications like hiring, finance, and healthcare.
  - Hold AI developers and organizations responsible for AI system outcomes.
  - Protect data used by AI models and maintain compliance with privacy regulations.
  - Conduct proactive AI impact assessments to identify and mitigate potential risks throughout the AI lifecycle, from development to deployment.

- **Training & Awareness:** Establish ongoing training programs that educate users on responsible AI use, potential risks, and security best practices. Provide resources and simulations to help teams recognize and respond to AI-related threats. Focus on:
  - Developers receive practical guidance on designing AI systems that prioritize safety and fairness.
  - End users learn to critically evaluate and interpret AI-generated outputs to ensure accuracy and compliance.
  - Leaders learn how to identify risks, enforce rules and governance policies, and ensure that regulatory adherence.

- **Adaptive Governance:** Regularly update governance models to reflect evolving AI capabilities and regulatory changes. Use feedback loops and audits to refine policies and ensure governance remains agile and effective continuously.



Implementing strong technology controls

Technology controls are essential for protecting GenAI tools from both internal and external threats. By implementing proactive defenses and monitoring systems, organizations can secure their AI environments. This becomes increasingly critical as GenAI plays a larger role in business operations and decision-making processes.

Key elements:

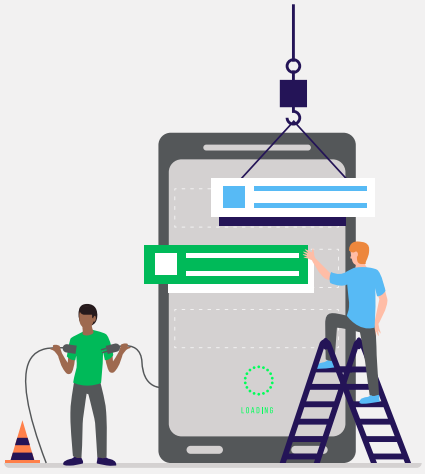
- **Secure Deployment:** The first step in safeguarding GenAI systems is to ensure secure deployment. This involves establishing logging mechanisms to track user interactions, AI-generated responses, and internal processes. Detailed logs provide an audit trail, making it possible to trace activities back when needed and quickly identify any irregularities, such as unauthorized access or unexpected behavior. Tools like Identity Threat Detection and Response (ITDR) can be deployed to monitor potential threats in real-time, quickly detecting attempts to tamper with models or breach systems.
- **Model and Data Integrity:** To protect the core of GenAI, it's critical to ensure the integrity of both AI models and the data they rely on. Strict access controls should be enforced for training datasets and model repositories to prevent unauthorized access. Regular audits of data pipelines can detect anomalies or unauthorized changes, and cryptographic hashing can be used to verify data integrity, preventing model poisoning or data manipulation.
- **Supply Chain Security:** Third-party libraries, models, and external services can introduce vulnerabilities into a GenAI system. Organizations should conduct regular assessments to vet the security of these external resources. Additionally, continuous patching and updating of GenAI tools will help mitigate risks, ensuring that no vulnerabilities are introduced via integrations or updates. Monitoring for unexpected behavior after updates is also essential.
- **Shadow AI Prevention:** As GenAI tools proliferate, it's important to control which tools are authorized for use within the organization. Endpoint and browser controls can restrict access to approved AI systems, minimizing the risk of unsanctioned "shadow AI." Continuous discovery mechanisms can also identify rogue AI instances and bring them under governance, ensuring they are properly monitored and controlled.

Delinea

Strengthening data and access controls

Securing data access and usage is essential to preventing breaches and ensuring that GenAI tools operate within organizational policies.

Key elements:



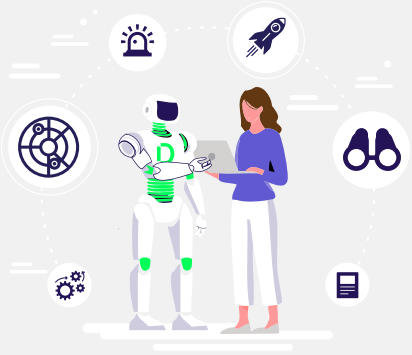
- **Identity Security:** A granular identity security model is essential for controlling both human and machine identities interacting with GenAI systems. Just-in-time access provisioning, multi-factor authentication (MFA), and the principle of least privilege can all help minimize exposure, ensuring that only authorized individuals and systems can interact with sensitive AI assets.
- **Data Protection:** Protecting training data is critical, especially when it involves personally identifiable information (PII) or other sensitive data. Before ingesting data into GenAI systems, organizations should strip out sensitive information and implement data classification and tagging to control access to various data sets. Encryption should also be employed to protect data both at rest and in transit, safeguarding it from unauthorized access or exposure.
- **Input Data Monitoring:** Real-time monitoring of input data is crucial to ensure compliance with privacy, ethical, and bias standards. Automated tools can flag non-compliant or anomalous data inputs, triggering alerts for further investigation. Additionally, requiring human-in-the-loop reviews for sensitive or high-risk use cases ensures that the AI operates within ethical and legal guidelines.
- **Output Data Security:** Protecting AI-generated outputs is equally important. Automated filters should be deployed to block or flag sensitive, harmful, or policy-violating content. Data outputs should be encrypted, tagged, and logged to maintain an auditable trail, helping organizations track the usage and impact of AI-generated content throughout its lifecycle.

Essential tools and capabilities for securing AI

To effectively secure AI environments, organizations must implement a comprehensive security stack that integrates identity management, privileged access control, and real-time monitoring.

Key security capabilities:

- **Privileged Access Management (PAM):** Restricts and monitors privileged access to AI models, datasets, and development environments, ensuring that only authorized users can modify or interact with sensitive AI assets.
- **Zero Trust AI Security:** Applies continuous authentication, real-time access validation, and just-in-time privilege elevation to mitigate insider and external threats.
- **Identity Threat Detection and Response (ITDR):** Detects and responds to identity-based threats targeting AI systems, providing real-time insights into anomalous access patterns and unauthorized activities.
- **Adaptive Policy Management:** Enables dynamic security policy adjustments based on AI-driven insights, user behavior, and evolving organizational requirements.
- **Comprehensive Compliance Monitoring:** Automates compliance reporting and provides audit trails to ensure adherence to industry regulations and governance policies.
- **AI Model Integrity Protection:** Protects AI models from adversarial attacks or unauthorized modifications, ensuring the integrity of both the model and its training data.
- **Secure DevOps for AI (AI-SecDevOps):** Integrates security practices into the AI development and deployment lifecycle, ensuring that security is built into AI models and datasets from the ground up.



By leveraging these tools and capabilities, organizations can strengthen their security posture, minimize AI-related risks, and ensure that AI-driven innovations remain secure and compliant.

Building security maturity: where to start and how to evolve

Securing GenAI is an ongoing journey. Organizations can build their security posture progressively by following a structured approach that balances immediate priorities with long-term maturity.

Phase 1: Foundation

- Form an AI governance committee and establish initial policies.
- Identify and catalog all GenAI tools in use.
- Implement baseline access controls, PAM, and logging mechanisms.
- Provide foundational AI security training to key teams.

Phase 2: Strengthening controls

- Expand access management to enforce least privilege and role-based access.
- Deploy advanced threat detection tools (e.g., ITDR) for GenAI environments.
- Enhance supply chain security through continuous vendor assessments.
- Automate compliance monitoring and data protection measures.

Phase 3: Full maturity

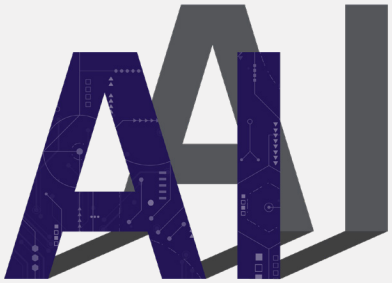
- Adopt continuous adaptive risk assessment processes.
- Leverage AI-driven analytics for anomaly detection and real-time response.
- Regularly refine governance policies to align with industry standards.
- Foster a security-first culture with ongoing training and incident simulations.

By following this phased approach, organizations can evolve their security capabilities, staying resilient against emerging GenAI threats while maintaining agility to seize AI-driven opportunities.

Common pitfalls to watch for

While securing GenAI offers immense benefits, organizations must remain aware of potential pitfalls that can undermine security efforts. Proactively addressing these challenges ensures a more resilient AI security strategy.

Key pitfalls:



- **Underestimating Shadow AI:** Even with strong policies, unsanctioned AI tools can creep into an organization. Failure to detect and govern these tools can introduce significant risk.
- **Over-Reliance on AI Outputs:** Trusting AI outputs without human validation can lead to inaccurate or biased decisions. Establish a human-in-the-loop process for critical workflows.
- **Incomplete Access Controls:** Neglecting machine-to-machine identities or failing to apply least-privilege principles increases the attack surface for adversaries.
- **Slow Adaptation to Evolving Threats:** The GenAI threat landscape evolves rapidly. Rigid security models that don't continuously adapt leave organizations vulnerable to new attack vectors.
- **Ignoring Supply Chain Complexity:** Overlooking the security of third-party AI components, libraries, or services can introduce hidden vulnerabilities into an otherwise secure system.

Conclusion

The adoption of Generative AI presents both opportunity and risk. By integrating governance, technology, and adaptive security measures, organizations can confidently navigate the complexities of AI adoption. Delinea's approach to securing GenAI empowers organizations to innovate without compromising security, trust, or compliance. As AI continues to evolve, staying vigilant and proactive will be key to harnessing its full potential while safeguarding the enterprise.

With Delinea's identity-first approach, organizations can build a secure AI ecosystem that not only drives innovation but ensures responsible and ethical AI use at scale.

Learn more about Delinea's AI security solutions at Delinea.com.

Delinea

Securing identities at every interaction

Delinea is a pioneer in securing identities through centralized authorization, making organizations more secure by seamlessly governing their interactions across the modern enterprise. Delinea allows organizations to apply context and intelligence throughout the identity lifecycle across cloud and traditional infrastructure, data, and SaaS applications to eliminate identity-related threats. With intelligent authorization, Delinea provides the only platform that enables you to discover all identities, assign appropriate access levels, detect irregularities, and immediately respond to identity threats in real-time. Delinea accelerates your teams' adoption by deploying in weeks, not months, and makes them more productive by requiring 90% fewer resources to manage than the nearest competitor. With a guaranteed 99.99% uptime, the Delinea Platform is the most reliable identity security solution available. Learn more about Delinea on **delinea.com**, **LinkedIn**, **X**, and **YouTube**.