delviom
INSIGHT | ANALYTICS | SECURITY

Shadow AI

The Rise of Generative AI and Its Risks

White Paper

44790 Maynard Square, Suite 280 | Ashburn, VA 20147 | Email: Info@delviom.com

Table of Contents

Restrictions

This material is for general informational purposes only and should not be considered legal, financial, or professional advice. While efforts have been made to ensure accuracy and relevance, no representation or warranty is given as to the completeness or applicability of the information for your particular situation. Government agencies, commercial organizations, and other stakeholders should consult qualified advisors for guidance specific to your organization or circumstances.

Shadow AI: The Rise of Generative AI and Its Risks

Timothy Mayers Jr., CISSP, CEH, CCSK, Delviom - Chief Cyber Solutions Architect

Introduction

Shadow AI is defined as the unauthorized use of artificial intelligence tools within organizations, poses significant risks to data security, compliance, and operational integrity. In 2024, generative AI (GenAI) applications saw explosive adoption across industries, with 94% of organizations integrating them into their workflows, according to the Netskope Cloud and Threat Report.[1] While this surge reflects the transformative potential of GenAI, it also exposes organizations to new and evolving risks, particularly around data leakage, compliance violations, and shadow AI usage. According to IBM's 2025 findings,[2] **breaches involving Shadow AI now cost enterprises an average of $4.63 million**; 16% higher than typical incidents; while 97% of affected organizations still lack fundamental AI access controls. Many organizations remain in the early stages of implementing robust governance and security controls, leaving gaps that malicious actors or careless insiders can exploit. According to the National Institute of Standards and Technology (NIST), mitigating these risks requires structured governance and technical controls. NIST's COSAiS (Control Overlays for Securing AI Systems) initiative[3] builds upon SP 800-53 and related publications to offer implementation-focused overlays tailored to various AI use cases, including generative models and multi-agent systems. These overlays emphasize protecting the confidentiality, integrity, and availability of AI components such as training data and model configurations. NIST encourages collaborative development of these controls through open forums, aiming to ensure that AI systems—authorized or otherwise—are secured against misuse and vulnerabilities.

[1] https://www.netskope.com/resources/cloud-and-threat-reports/cloud-and-threat-report-2025
[2] https://venturebeat.com/security/ibm-shadow-ai-breaches-cost-670k-more-97-of-firms-lack-controls/
[3] https://csrc.nist.gov/projects/cosais

Data Loss Prevention (DLP) Adoption and Gaps

DLP tools can help detect and block unauthorized AI usage by monitoring data flows for signs of sensitive information being accessed or transmitted to external AI platforms. To mitigate the risks of sensitive data exposure, Netskope reported that 45% of organizations have adopted DLP tools to monitor and control the flow of information into GenAI apps. However, adoption varies significantly by industry. For example, the telecommunications sector leads with 64% DLP adoption, while other sectors lag behind. This uneven implementation highlights the need for industry-specific strategies and broader awareness of how GenAI can inadvertently become a conduit for data exfiltration. By enforcing policies that restrict access to confidential datasets and flag anomalous behavior, DLP solutions reduce the risk of employees using generative AI tools without proper oversight. Integrating DLP with identity and access management systems ensures that only approved users and applications can interact with AI models, helping organizations maintain compliance and prevent Shadow AI incidents.

Empowering Users Through Real-Time Coaching

Beyond technical controls, Netskope reported that 34% of organizations are leveraging real-time interactive user coaching to guide employees in making informed decisions when interacting with GenAI tools. This approach is proving effective: 73% of users who receive warnings about potential policy violations choose not to proceed. These statistics underscore the value of combining automation with human-centric interventions to foster a culture of responsible AI usage.

Regular cybersecurity training is essential for maintaining a strong security posture in today's rapidly evolving threat landscape. As cyber threats become more sophisticated, often leveraging GenAI, organizations must ensure their workforce is equipped to recognize and respond effectively. Studies have shown that continuous training can lead to measurable improvements in security performance. For instance, according to a survey conducted by Keepnet Labs in 2025,[4] one survey **reported a 70% decrease** in phishing incidents after implementing a regular training program. This kind of training not only enhances technical skills but also fosters a proactive security culture, where employees are more vigilant and confident in identifying threats. Simulation-based exercises, workshops, and self-guided learning opportunities help reinforce best practices and prepare teams for real-world scenarios. Ultimately, consistent training empowers employees to serve as the first line of defense, reducing the likelihood of breaches and improving incident response times across the organization.

[4] https://keepnetlabs.com/blog/security-awareness-training-statistics

Blocking Unapproved GenAI Apps

Blocking unapproved GenAI applications is becoming a critical priority for organizations aiming to safeguard data privacy, intellectual property, and operational integrity. To further reduce exposure to Shadow AI, Netskope reported that 73% of organizations have implemented app-blocking mechanisms, with an average of 2.4 generative AI apps blocked per organization annually. Notably, the top quartile of organizations blocking GenAI apps have more than doubled their blocked app count, from 6.3 to 14.6, in just one year. This trend reflects growing awareness of the proliferation of unvetted GenAI tools and the need for proactive app governance. These apps, often cloud-based and rapidly evolving, can pose significant risks if employees use them without oversight. Unapproved GenAI tools may inadvertently expose sensitive information through prompts or outputs, especially when integrated with internal systems or used to process proprietary data. As a result, companies are increasingly implementing stricter access controls, network monitoring, and endpoint protection to prevent unauthorized usage.

The challenge lies not only in identifying and blocking these apps but also in educating employees about the risks associated with them. Many GenAI tools offer powerful productivity enhancements, which can tempt users to bypass official channels. To counter this, organizations are adopting comprehensive governance frameworks that include clear policies, approved AI tool lists, and regular training sessions. These efforts help foster a culture of responsible AI use, ensuring that innovation does not come at the cost of security or compliance.

From a technical standpoint, blocking unapproved GenAI apps often involves a combination of firewall rules, DNS filtering, and application whitelisting. Advanced solutions may also leverage AI-driven threat detection to identify anomalous usage patterns or unauthorized data transfers. However, **the most effective strategy combines technical controls with proactive communication and collaboration between IT, security teams, and business units.** In addition to blocking apps, many organizations are investing in real-time monitoring and behavioral analytics to detect unauthorized AI usage before it escalates into a security incident. These efforts are often paired with employee education campaigns and stricter data access policies to reinforce responsible AI usage. As the enterprise AI landscape evolves, combining technical controls with cultural and procedural safeguards is becoming essential to mitigate the risks posed by Shadow AI and maintain regulatory compliance. By aligning security with business goals, organizations can safely harness the benefits of GenAI while minimizing its risks.

Recommendations for Strengthening AI Security Posture

The Netskope Cloud and Threat Report[5] recommends a multi-layered approach to AI security, emphasizing that traditional defenses are no longer sufficient in the face of increasingly sophisticated threats. With phishing attacks now often powered by generative AI, organizations must move beyond basic user education and adopt advanced security measures that can adapt to evolving tactics. Investing in modern data protection solutions, such as behavioral analytics, AI-driven threat detection, and contextual access controls, is essential to counter increasingly convincing phishing attempts that target users across email, social media, messaging platforms, and search engines.

Moreover, the report highlights the importance of integrating AI-specific threat intelligence into existing security operations to identify and neutralize risks in real time. Organizations are encouraged to implement continuous monitoring of AI tool usage, enforce strict data governance policies, and establish clear protocols for evaluating and approving new AI applications. While addressing many Zero Trust requirements, secure internet browsers play a critical role in defending organizations against the risks posed by shadow AI. These browsers are designed with built-in security features such as data isolation, sandboxing, real-time threat detection, and policy enforcement, which help prevent unauthorized access to sensitive information and block unapproved GenAI applications from being used within corporate environments. By routing traffic through secure gateways and enforcing strict access controls, secure browsers can limit exposure to phishing attacks, malicious scripts, and data exfiltration attempts, especially those facilitated by AI-powered tools. Additionally, they enable organizations to monitor and log user interactions with web-based GenAI apps, providing visibility into potential misuse and supporting compliance with internal policies. When integrated with broader data protection strategies like DLP and user coaching, secure browsers become a powerful frontline defense against shadow AI threats.

Managing Insider Risk and Shadow IT

Insider threats in the context of AI are becoming increasingly complex, as employees and contractors gain access to powerful generative tools that can be misused intentionally or inadvertently, to exfiltrate sensitive data, manipulate models, or bypass security protocols. Unlike traditional insider risks, AI-enabled threats can involve subtle prompt engineering, unauthorized model training, or the use of Shadow AI tools that operate outside sanctioned environments. Common behaviors include uploading proprietary files to personal cloud accounts, using personal backups, or transferring data when leaving the company. To address this, organizations must limit access to apps based on business relevance, establish app approval workflows, and deploy continuous monitoring to detect misuse or compromise in real time. Organizations must also expand their insider threat programs to include AI-specific monitoring, enforce strict access controls on model inputs and outputs, and implement behavioral analytics to detect anomalous usage patterns that may indicate misuse or data leakage. Proactive governance and cross-functional collaboration between security, compliance, and data science teams are essential to mitigate these emerging risks.

[5] https://www.netskope.com/netskope-threat-labs/cloud-threat-report

Preparing for the Future of Generative AI in the Workplace

As GenAI becomes increasingly embedded in daily workflows, its adoption is expected to accelerate throughout 2025 and beyond, transforming how organizations operate across departments. To prepare for this shift, enterprises must implement robust governance frameworks that include controls to restrict the use of unauthorized GenAI applications and define acceptable use cases aligned with business objectives and compliance requirements. Enforcing data movement policies such as preventing sensitive information from being input into external AI tools, is critical to maintaining data integrity and minimizing exposure to Shadow AI risks.

Equally important is the ongoing education of employees through real-time coaching and contextual guidance, helping users understand both the capabilities and limitations of GenAI tools. Organizations should also invest in AI-aware security solutions, such as DLP, behavioral analytics, and model usage monitoring, to detect and respond to misuse proactively. Leveraging leading industry guidance is essential. The SANS Institute complements NIST's approach with its Critical AI Security Guidelines,[6] a community-driven framework designed to secure AI deployments across enterprise environments. These guidelines advocate for a multi-layered security strategy that includes defenses against model poisoning, prompt injection, and adversarial attacks. SANS also stresses the importance of adaptive governance frameworks that evolve with AI technologies, ensuring alignment with compliance standards like OWASP's AI Security and Privacy Guide. Their GitHub-hosted repository invites contributions from cybersecurity professionals to maintain the document as a living resource, reflecting the dynamic nature of AI threats. By balancing innovation with security, organizations can harness the productivity benefits of GenAI while safeguarding intellectual property, customer data, and regulatory compliance. This forward-looking approach ensures that GenAI becomes a strategic asset rather than a source of risk.

[6] https://www.sans.org/mlp/critical-ai-security-guidelines

What steps should organizations take?

To effectively mitigate the risks associated with shadow AI, organizations must adopt a proactive, multi-layered strategy rooted in industry best practices. This begins with ***educating employees*** about the hidden dangers of unauthorized AI use, such as data privacy breaches, security vulnerabilities, and compliance failures, while promoting the responsible use of approved tools. Next, organizations should ***establish clear AI usage policies*** that define ethical boundaries, data handling protocols, and approval processes. To enforce these standards, ***monitoring and auditing mechanisms*** must be deployed to detect and address unauthorized AI activity. Simultaneously, organizations should provide ***secure, vetted AI tools tailored to business needs***, supported by a streamlined request and approval system. Finally, ***fostering a culture of transparency*** is essential to encourage new ideas and open dialogue about AI usage while creating a safe, non-punitive environment where employees feel empowered to share their needs and concerns. Combined, these practices form a resilient framework for responsible AI adoption in the workplace.

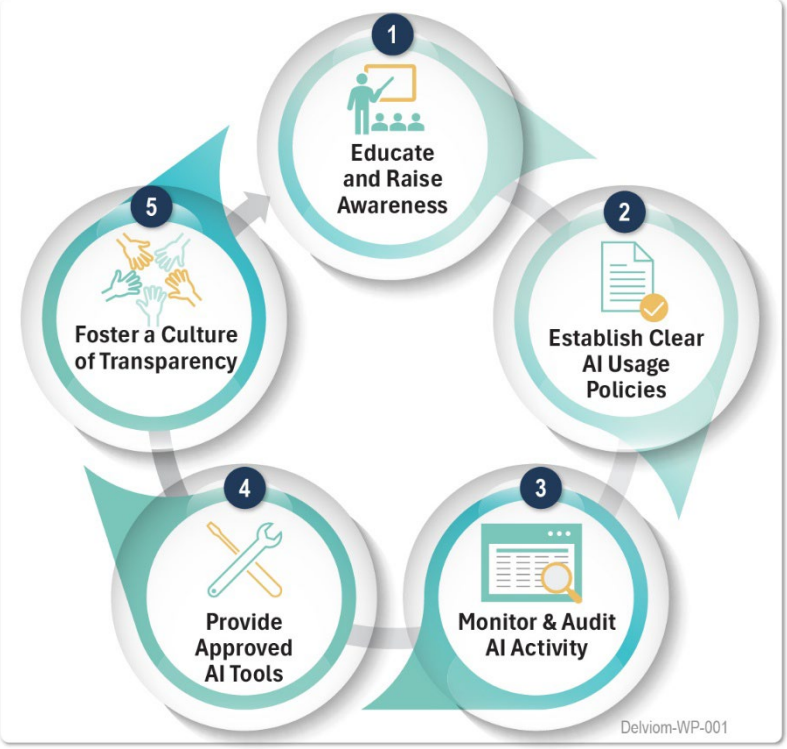

Figure 1: Five Steps to Safe AI Usage

How Delviom Supports AI Security Transformation

Delviom provides a full range of cybersecurity services for both federal and commercial clients, and holds multiple certifications including ISO 27001, CMMI Level 3, and DOD CMMC Level 2, reflecting our commitment to high-quality and secure service delivery. We offer comprehensive support to organizations navigating the complexities of AI security. By leveraging frameworks such as the NIST AI Risk Management Framework, OWASP Top 10 for LLMs, MITRE ATLAS, and MIT AI Repository, Delviom helps clients assess, model, and secure their AI pipelines. Our offerings include AI Governance, Penetration Testing, Red Teaming, Guardrails, Security Operations Center Support, and AI Security Posture Management, all tailored to deliver measurable outcomes. We collaborate closely with stakeholders to identify high-impact use cases and design full-stack solutions that align with business goals and regulatory requirements. By combining technical safeguards with strategic oversight, enterprises can better defend against the misuse of generative AI and reduce their exposure to Shadow AI-related breaches.

Contact Us

Web: Delviom.com
Email: info@delviom.com
Phone: +1 (703) 953 2535