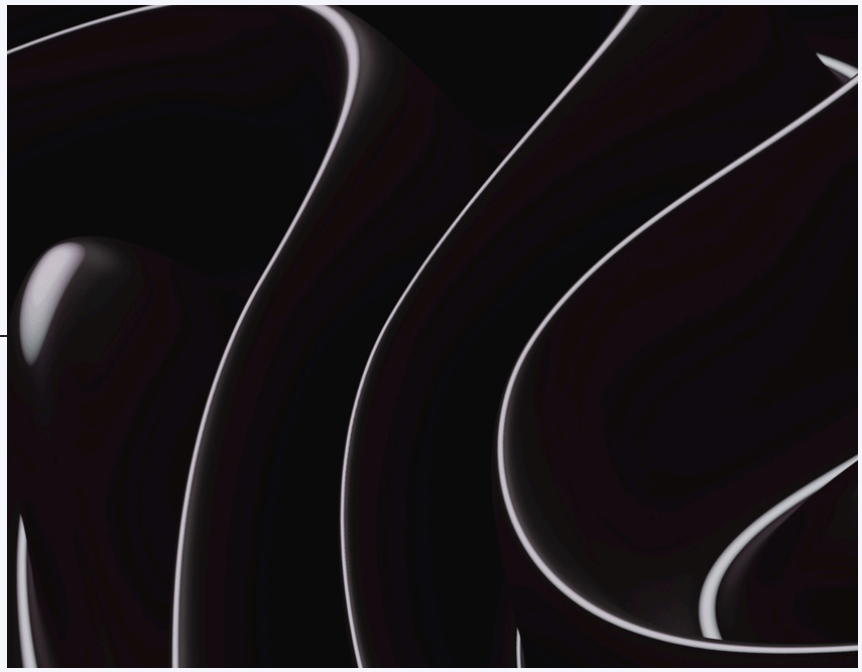

secnap

Navigating the Risks of GenAI: **A Comprehensive Guide to Safeguarding Your Organization**

White Paper

Prepared for Business Leaders, IT/Security Professionals, and Compliance Officers.

CONTENTS

secnap

EXECUTIVE SUMMARY

Generative AI (GenAI), which creates content such as text, images, code, or audio, is transforming operations across various sectors, including finance, healthcare, legal, retail, government, education, and manufacturing. From automating customer service to analyzing medical data, GenAI drives efficiency and innovation. However, its adoption introduces significant cybersecurity, privacy, and ethical risks, amplified by cybercriminals, internal missteps, and vendor vulnerabilities. This white paper equips CISOs, CIOs, CTOs, CFOs, CEOs, and legal counsel with strategies to mitigate these risks, leveraging standards such as NIST CSF 2.0, ISO 27001, and ISO 42001, as well as actionable steps like Privacy Impact Assessments (PIAs) and vendor oversight. By adopting robust frameworks and governance, organizations can harness GenAI's potential while safeguarding data, ensuring compliance, and maintaining trust.

INTRODUCTION

Imagine a healthcare provider's GenAI-powered chatbot, designed to streamline patient inquiries, is exploited by hackers to extract sensitive medical records, triggering a $7 million fine and eroding patient trust. This scenario, increasingly plausible in 2025, underscores the dual nature of Generative AI (GenAI) -- a powerful tool for innovation and a vector for risk. GenAI enables applications like automated financial forecasting, patient data analysis in healthcare, legal document drafting, inventory optimization in retail, policy analysis in government, personalized learning in education, and predictive maintenance in manufacturing. Yet, its data-hungry nature and complexity introduce cybersecurity, privacy, and ethical challenges. This white paper provides a roadmap for professionals to navigate these risks, drawing on real-world incidents and proven standards to ensure the secure adoption of GenAI across high-stakes sectors.

RISKS OF GEN AI IN CYBERCRIMINAL HANDS

Cybercriminals are leveraging GenAI to launch sophisticated attacks, exploiting its ability to mimic human behavior and scale malicious activities. The urgency to counter these threats is critical, as they target sensitive data and erode trust in industries like finance, healthcare, and legal services.

Hyper-Realistic Phishing Campaigns:

GenAI crafts convincing phishing emails and SMS messages tailored to deceive employees or customers. In 2023, hackers used GenAI to create targeted SMS phishing campaigns against Activision, breaching an HR employee's credentials to access employee email addresses, phone numbers, and salaries.[1]

Deepfake Impersonation Scams:

GenAI-powered deepfakes mimic trusted individuals, such as CFOs or legal counsel, to trick victims into sharing sensitive data or funds. In healthcare, deepfakes impersonating executives have surged, with fewer than 50% of C-suite leaders believing their organizations are well-positioned to thwart such attacks.[2]

Adaptive Malware and Ransomware:

GenAI enhances malware to evade detection and prioritizes critical files for ransomware encryption. In January 2023, Yum! Brands suffered an AI-driven ransomware attack, forcing the closure of 300 UK branches and exposing employee data.[3]

Data Poisoning and Model Evasion:

Cybercriminals manipulate GenAI training data or inputs to compromise model integrity, leading to flawed outputs. In finance, 47% of institutions reported successful prompt injection attacks exposing data due to compromised datasets.[4]

These incidents underscore the rapidly evolving threat landscape, with 93% of security leaders anticipating that AI-driven attacks are becoming a daily occurrence.[5] Organizations must act swiftly to protect sensitive data.

RISKS OF GENAI WITHIN YOUR ORGANIZATION

Integrating GenAI into operations, whether for financial forecasting, patient data analysis, or legal research, introduces internal risks. Its data-hungry nature can expose organizations to privacy breaches and operational disruptions if not managed carefully.

Data Exposure from Overreach:

GenAI systems require vast datasets, initiating the risk of unintended exposure of sensitive information. In 2023, Samsung banned GenAI tools after employees inadvertently leaked confidential code via ChatGPT, exposing proprietary data.[6]

Privacy Compliance Challenges:

GenAI's broad data processing capabilities risk violating privacy laws if left unchecked. Regulations like the EU's GDPR and California's CCPA mandate data minimization, but GenAI often processes more data than necessary, creating potential compliance issues. In healthcare, the Royal Free NHS Trust was fined for sharing 1.6 million patient records with Google's DeepMind without consent, highlighting transparency failures.[7]

Operational Overreliance:

Overdependence on GenAI without human oversight can disrupt operations. In 2023, a Stanford professor's AI-generated citations were dismissed in court due to inaccuracies, undermining a legal case on Minnesota's deep-fake election law.[8]

Reputational Risks:

GenAI errors or biases can significantly undermine trust. For instance, if a financial institution's AI chatbot delivers inaccurate investment information, it could compromise client confidence and harm the organization's brand reputation.

These risks necessitate robust internal controls to balance GenAI's benefits with its potential for harm.

RISKS FROM GENAI IN YOUR SUPPLY CHAIN

Third-party vendors providing GenAI tools introduce risks, particularly in finance, health care, legal, retail, government, education, and manufacturing, where data sensitivity is paramount. Vendor-related breaches can have a ripple effect throughout your organization.

Vendor Data Breaches:	Vendors handling GenAI systems may lack adequate security, exposing your data. In 2024, American Express reported a breach via a third-party merchant processor, compromising customer names and card details.[9]
Lack of Transparency:	GenAI vendors often fail to disclose their training data sources or access policies, which makes regulatory compliance more challenging. In healthcare, where compliance with HIPAA is crucial, more than 50% of organizations still fail to monitor AI usage across systems that handle sensitive patient data.[10]
Ethical Missteps:	Vendors that rely on biased training data risk generating harmful and discriminatory outputs. In 2023, an AI-powered legal tool faced criticism for issuing biased sentencing recommendations in U.S. courts, raising serious concerns about fairness and eroding the public's trust in the legal system.[11]
Financial and Regulatory Fallout:	Vendor breaches can be costly. Financial institutions face average breach costs exceeding $6 million, with regulatory penalties often surpassing this figure.[12]

Rigorous vendor oversight is essential to protect data and reputation.

GENAI LIFECYCLE MANAGEMENT

Managing GenAI across its lifecycle — development, design, deployment, and decommissioning — ensures security and compliance.

Development:

Embed security by design principles by using threat modeling to anticipate vulnerabilities such as data poisoning. Action: Implement secure coding practices for GenAI models.

Design:

Conduct adversarial testing to validate model integrity. Action: Use red-teaming to simulate attacks, as recommended by NIST.[13]

Deployment:

Monitor systems for anomalous behavior. In 2023, OpenAI's internal forums were compromised due to inadequate monitoring, exposing sensitive details about its AI technology.[14]

Decommissioning:

Ensure secure data deletion to eliminate residual risks. Action: Follow ISO 27002's data disposal controls (Control 8.3.2).[15]

REGULATORY LANDSCAPE FOR GENAI

Compliance with regulations is critical for GenAI deployment in regulated sectors.

EU AI Act:

Classifies high-risk AI systems, requiring transparency and risk mitigation. Action: Document GenAI data usage and implement risk plans.[16]

GDPR:

Mandates data protection by design (Article 25). Action: Limit GenAI training to non-personal data where possible.[17]

HIPAA:

Requires safeguards for healthcare data. Action: Ensure vendors encrypt patient data processed by GenAI.[18]

CCPA:

Grants consumer data rights. Action: Provide opt-out mechanisms for GenAI data processing.[19]

ETHICAL AND REPUTATIONAL CONSIDERATIONS

GenAI's potential for bias and lack of transparency can lead to ethical and reputational risks.

Bias in Outputs:

Biased training data can produce unfair outcomes, as seen in the 2023 legal tool incident.[20]

Transparency:

A lack of clarity in GenAI processes erodes trust.
Action: Publish transparency reports on GenAI usage.

Reputational Impact:

Errors in customer-facing GenAI tools, such as retail chatbots, can quickly erode brand credibility. Action: Introduce human-in-the-loop oversight to spot-check outputs and monitor high-impact decisions, ensuring accuracy and maintaining customer trust.

STANDARDS AND FRAMEWORKS TO MITIGATE GENAI RISKS

Established standards provide practical solutions to manage GenAI risks, ensuring security, privacy, and ethical alignment.

NIST Cybersecurity Framework 2.0:	Emphasizes proactive risk management. The "Govern" function establishes GenAI-specific policies, while "Protect" secures data pipelines. Action: Conduct risk assessments (ID.RA-1) to identify vulnerabilities.[21]
ISO 27001 and 27002:	Provide structured information security. Annex A.8.2.1 (ISO 27001) ensures secure asset management, while Control 5.15 (ISO 27002) limits access. Action: Restrict GenAI access to authorized personnel.[22]
ISO 42001:	Ensures AI governance across the lifecycle. Action: Establish an AI stewardship program.[23]
CIS Controls:	Control 6 (Access Control Management) restricts access, and Control 13 (Network Monitoring) detects intrusions. Action: Audit GenAI configurations using CIS benchmarks.[24]
NIST AI RMF:	Maps GenAI use cases and mitigates risks in categories such as robustness and privacy. Action: Conduct red-teaming for adversarial attacks.[25]
PIAs for GenAI:	Implement Privacy Impact Assessments to evaluate data flows and ensure compliance with HIPAA and GDPR. Action: Map GenAI inputs/outputs for minimization.[26]

VENDOR MANAGEMENT AND SUPPLY CHAIN CHECKLIST

potential weak links in your organization's security and compliance posture. In sectors such as finance, healthcare, legal, retail, government, education, and manufacturing, vendor-related breaches can expose sensitive data, disrupt operations, and trigger regulatory penalties. For example, a healthcare vendor's unsecure GenAI platform could leak patient records, violating HIPAA, while a legal firm's AI tool with biased outputs could erode client trust. To mitigate these risks, use the following comprehensive checklist to ensure GenAI vendors align with cybersecurity, privacy, and ethical standards, fostering resilience across your supply chain.

Cybersecurity Compliance

Require vendors to demonstrate compliance with robust standards like ISO 27001 or NIST Cybersecurity Framework (CSF) 2.0 through annual third-party audits. These standards mandate rigorous security controls, such as access management and incident response, to protect GenAI systems from breaches. For instance, a financial institution partnering with a GenAI vendor for customer analytics should verify that the vendor's data pipelines are audited against ISO 27001's Annex A.8 (Information Security) controls to prevent unauthorized access. Action: Request audit reports or certifications annually and include compliance verification in vendor contracts to ensure ongoing adherence.[27]

Standard	Key Features and Actions
NIST CSF 2.0	Govern function for policies, Protect for data security. Conduct risk assessments (ID.RA- 1).
ISO 27001 27002	Risk assessment (A.8.2.1), access controls (5.15). Restrict GenAI access.
ISO 42001	AI governance across the lifecycle. Establish a stewardship program.
CIS Controls	Access control (6), network monitoring (13). Audit configurations.
NIST AI RMF	Map use cases to mitigate robustness and privacy risks. Conduct red-teaming.
PIAs	Evaluate data flows for HIPAA/GDPR compliance. Map inputs/outputs.

Data Protection Measures

Ensure vendors implement strong encryption (e.g., AES-256) and access controls (e.g., role-based access, multi-factor authentication) for all GenAI data processing, aligning with GDPR's data protection by design (Article 25) and HIPAA's Security Rule. In healthcare, a vendor's failure to encrypt patient data processed by a GenAI diagnostic tool could lead to fines exceeding $1 million. Action: Mandate encryption for data at rest and in transit, conduct periodic access control reviews, and require vendors to comply with sector-specific regulations like the FTC's Safeguards Rule for financial or legal data.[28]

Transparency in Data Usage

Demand full disclosure of GenAI training data sources, model access policies, and data handling practices to ensure compliance and ethical integrity. Lack of transparency can lead to violations, as seen when a retail vendor's opaque AI model used customer data without consent, breaching the CCPA. Action: Require vendors to provide detailed data provenance reports and access logs, and establish contractual clauses mandating transparency to facilitate Privacy Impact Assessments (PIAs) and regulatory audits.[29]

Incident Response Plan

Confirm vendors have a documented incident response plan with a 72-hour breach notification commitment, as required by GDPR (Article 33). A government agency using a GenAI vendor for policy analysis could face delays in breach response if the vendor lacks a clear protocol, risking public trust. Action: Review vendor incident response plans annually, ensure they include rapid notification procedures, and test their effectiveness through tabletop exercises to prepare for real-world breaches.[30]

Ethical AI Practices

Insist vendors use bias-free training data and conduct regular audits to prevent discriminatory or harmful outputs, aligning with ISO 42001's ethical AI governance principles. In 2023, an AI tool used by a legal firm produced biased sentencing recommendations, sparking client backlash and reputational damage. Action: Require vendors to perform quarterly bias audits, document mitigation strategies, and provide evidence of ethical data sourcing to ensure fairness across applications.[31]

Contractual Safeguards

Incorporate Service Level Agreements (SLAs) mandating compliance with ISO 42001 and NIST AI Risk Management Framework (RMF), with penalties for non-compliance. For example, a manufacturing firm using a GenAI vendor for predictive maintenance should enforce SLAs to ensure secure data handling. Action: Draft SLAs specifying compliance with AI-specific standards, include penalty clauses for breaches, and require vendors to align with your organization's risk management framework to ensure accountability.[32]

Subvendor Oversight

Ensure vendors extend security and compliance requirements to their subcontractors, as mandated by the FTC's Safeguards Rule for third-party service providers in financial or legal contexts. A breach in an education vendor's subcontractor could expose student data, violating privacy laws. Action: Require vendors to document subcontractor compliance with security standards, conduct joint audits, and include subcontractor oversight clauses in contracts to maintain a secure supply chain.[33]

Regular Risk Assessments

Mandate vendors to conduct annual GenAI-specific risk assessments to identify vulnerabilities, such as prompt injection or data poisoning, aligning with NIST CSF 2.0's Identify function (ID.RA-1). For instance, a government vendor's unassessed GenAI model could be exploited, compromising sensitive policy data. Action: Require vendors to submit annual risk assessment reports, integrate findings into your organization's risk management strategy, and use red-teaming to simulate adversarial attacks on vendor systems.[34]

By implementing this checklist, organizations can strengthen their GenAI supply chain, ensuring vendors uphold the highest standards of security, privacy, and ethics, and minimizing risks across critical sectors.

CONCLUSION: A CALL TO ACTION FOR SECURE GENAI ADOPTION

GenAI holds immense potential for finance, healthcare, legal, retail, government, education, and manufacturing; however, its risks, including cybercriminal exploitation, internal exposures, and vendor vulnerabilities, require urgent attention. The 2023 OpenAI breach, where hackers accessed internal AI forums, underscores the stakes.[35] Robust frameworks can trans- form risks into opportunities.

Adopt Standards:	Implement NIST CSF 2.0, ISO 27001, ISO 42001, CIS Controls, and NIST AI RMF to secure systems.[36]
Conduct PIAs:	Map GenAI data flows to mitigate privacy risks, using AIAIC case studies.[37]
Engage Experts:	Hire consultants or build AI governance teams to align with GDPR, HIPAA, and the EU AI Act.[38]
Prioritize Risk Management:	Establish an AI stewardship program with continuous monitoring and red-teaming.
Leverage Resources:	Use CSA's AI Security framework, ISACA's governance guides, IAPP's privacy resources, CISA's playbooks, ENISA's guidelines, and NIST's AI Safety Institute.[39]

Start today by reviewing GenAI use cases, adopting these standards, and building a secure future.

GLOSSARY

- **GenAI:** Generative AI, technologies that create content like text, images, code, or audio.
- **PIA:** Privacy Impact Assessment, a process to evaluate data privacy risks.
- **ISO 27001:** International standard for information security management systems.
- **NIST AI RMF:** NIST AI Risk Management Framework, a guide for managing AI risks.

NOTES

1. Source: Hackers Allegedly Steal Activision Games and Employee Data, TechCrunch, February 21, 2023, https://techcrunch.com/2023/02/21/hackers-allegedly-steal-activision-games-and-employee-data/
2. Source: C-Suite Leaders Brace for Rise in Deepfake Financial Fraud, Poll Reveals, Deloitte, 2023, https://deloitte.wsj.com/cio/c-suite-leaders-brace-for-rise-in-deepfake-financial-fraud-poll-reveal
3. Source: Yum! Brands Says Nearly 300 Restaurants in UK Impacted Due to Cyber Attack, Reuters, January 19, 2023, https://www.reuters.com/business/retail-consumer/yum-brands-says-nearly-300-restauran
4. Source: Charting the Course of AI: Quantifying Risk for Financial Institutions, FS-ISAC, 2023, https://www.fsisac.com/knowledge/ai-risk
5. Source: 93% of Security Leaders Anticipate Daily AI Attacks by 2025, Security Magazine, 2023, https://www.securitymagazine.com/articles/100613-93-of-security-leaders-anticipate-daily-ai-attacks
6. Source: Samsung Bans ChatGPT and Other Generative AI Use by Staff After Leak, Bloomberg, May 2, 2023, https://www.bloomberg.com/news/articles/2023-05-02/samsung-bans-chatgpt-and-other-generative-
7. Source: UK Hospital-DeepMind Trial Failed to Comply with UK Data Protection Law, Inside Privacy, 2017, https://www.insideprivacy.com/health-privacy/uk-hospital-deepmind-trial-failed-to-comply-
8. Source: The Credibility Crisis: AI-Generated Citations in Expert Testimony, VMG Health, 2023, https://vmghealth.com/thought-leadership/published-article/the-credibility-crisis-ai-generated-cita
9. Source: American Express Notifies Customers of Data Breach at Third Party, PYMNTS, 2024, https://www.pymnts.com/news/security-and-risk/2024/american-express-notifies-customers-of-data-brea
10. Source: HIMSS Survey Report: Beware Insider Threats, Cybersecurity Involving AI, AI in Health- care, 2023, https://aiin.healthcare/topics/artificial-intelligence/himss-survey-report-beware-insider-
11. Source: Machine Bias: Risk Assessments in Criminal Sentencing, ProPublica, 2023, https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing
12. Source: Report: Average Data Breach Cost for Financial Sector Tops $6M, American Bankers Association, August 2024, https://bankingjournal.aba.com/2024/08/report-average-data-breach-cost-for-financia
13. Source: NIST AI Risk Management Framework, NIST, 2023, https://www.nist.gov/itl/ai-risk-management-fr
14. Source: OpenAI's Internal AI Details Stolen in 2023 Breach, Reuters, July 5, 2024, https://www.reuters.com/technology/cybersecurity/openais-internal-ai-details-stolen-2023-breach-nyt-reports-2024-07-05/

15. Source: ISO/IEC 27002:2022, ISO, 2022, https://www.iso.org/standard/75652.html
16. Source: EU Artificial Intelligence Act, European Commission, 2024, https://artificialintelligenceact.eu/the-act/
17. Source: General Data Protection Regulation, European Union, 2016, https://gdpr.eu
18. Source: HIPAA Security Rule, U.S. Department of Health and Human Services, 2023, https://www.hhs.gov/hipaa/for-professionals/security/index.html
19. Source: California Consumer Privacy Act, State of California, 2020, https://oag.ca.gov/privacy/ccpa
20. Source: Machine Bias: Risk Assessments in Criminal Sentencing, ProPublica, 2023, https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing
21. Source: NIST Cybersecurity Framework 2.0, NIST, 2024, https://www.nist.gov/cyberframework
22. Source: ISO/IEC 27001:2022, ISO, 2022, https://www.iso.org/standard/27001
23. Source: ISO/IEC 42001:2023, ISO, 2023, https://www.iso.org/standard/42001
24. Source: CIS Controls v8, CIS, 2023, https://www.cisecurity.org/controls
25. Source: NIST AI Risk Management Framework, NIST, 2023, https://www.nist.gov/itl/ai-risk-management-framework
26. Source: AI and Algorithmic Incident and Controversy Repository, AIAIC, 2023, https://www.aiaaic.org/
27. Source: ISO/IEC 27001:2022, ISO, 2022, https://www.iso.org/standard/27001
28. Source: General Data Protection Regulation, European Union, 2016, https://gdpr.eu}; HIPAA Security Rule, U.S. Department of Health and Human Services, 2023, https://www.hhs.gov/hipaa/for-professionals/security/index.html
29. Source: California Consumer Privacy Act, State of California, 2020, https://oag.ca.gov/privacy/ccpa
30. Source: General Data Protection Regulation, European Union, 2016, https://gdpr.eu
31. Source: Machine Bias: Risk Assessments in Criminal Sentencing, ProPublica, 2023, https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing
32. Source: ISO/IEC 42001:2023, ISO, 2023, https://www.iso.org/standard/42001
33. Source: Complying with the Safeguards Rule, FTC, 2023, https://www.ftc.gov/business-guidance/resources/ftc-safeguards-rule-what-your-business-needs-know
34. Source: NIST Cybersecurity Framework 2.0, NIST, 2024, https://www.nist.gov/cyberframework
35. Source: OpenAI's Internal AI Details Stolen in 2023 Breach, Reuters, July 5, 2024, https://www.reuters.com/technology/cybersecurity/openais-internal-ai-details-stolen-2023-breach-nyt-reports-2024-07-05/
36. Source: NIST, https://www.nist.gov; ISO, https://www.iso.org; CIS, https://www.cisecurity. org
37. Source: AIAIC, https://www.aiaaic.org/
38. Source: EU AI Act, European Commission, 2024; GDPR, European Union, 2016; HIPAA, U.S. HHS.
39. Source: CSA, https://cloudsecurityalliance.org/research/working-groups/ai-security; ISACA; https://www.isaca.org/resources/insights-and-expertise/artificial-intelligence; IAPP ; https://iapp.org/resources/topics/artificial-intelligence; CISA, https://www.cisa.gov/news-events/alerts/2025/05/22/new-best-practices-guide-securing-ai-data-released; ENISA, https://www.enisa.europa.eu/topics/artificial-intelligence-and-next-gen-technologies ; NIST AI Safety Institute, https://www.nist.gov/aisi