Rajah &Tann Cybersecurity

GenAI Cyber Risk

Adopt GenAI with a peace of mind with countermeasures and advanced solutions designed to prevent and mitigate GenAI cyber risks.

In today's rapidly evolving digital landscape, cyber threats associated with Generative AI pose significant risks. From deepfakes to AI-enabled phishing, these sophisticated tactics can lead to severe financial loss, identity theft, and compromised security.

What is GenAI?

Generative AI (GenAI) is a type of artificial intelligence that uses machine learning algorithms to produce, copy or rework content in various formats, including text, images, audio, code and more.

Cybersecurity Concerns of GenAI

Data Exposure
Risk of leaking sensitive info

Incorrect Outputs
AI-generated misinformation

Malicious Outputs
AI exposing vulnerabilities

Unsafe Output
Generation of inappropriate content

Shadow Usage
Unauthorised use of GenAI by users

Insecure Access
Weak identity and access management

Threats	Risks	Impact
Deepfakes and GenAI-Enabled Phishing	**People:** More convincing impersnation with deepfakes	Business email compromise (BEC), loss of personally identifiable information (PII), financial loss
	Process: Threat actors leveraging GenAI to create realistic content and personas for fraud and deception	Financial loss, identify theft, data exposure
	Technology: New account creation fraud, bypassing existing authentication with fake identities, spread of fake news with botnets	Fraud risk, security compromise, money laundering schemes, evasion of security protocol, market manipulation
Maware Generation and Enhancement	**People:** Malware sent through phishing links	Monetary loss, loss of PII
	Process: Inability to detect new malware	Delayed threat detection and containment of malware
	Technology: Polymorphic AI malware evading detection, outdated enterprise systems unable to detect malware	Bypass of security measures leading to loss of sensitive information, finacial loss and repuatation damage

INFO@RTCYBER.COM
WWW.RTCYBER.COM

Threats	Risks	Impact
Data Leakage from GenAI Deployment	**People:** Intentional or unintentional data leaks by employees to public GenAI models	Loss of customer data, PII, FI secrets Regulatory consequences and repuattional damage
	Process: Vulnerabilities or security weaknesses in in-house developed GenAI models, risks of supply chain attack arising from use of third party or open-source GenAI models	Data leakage leading to loss of sensitive information Backdoors and in-built vunlerabilities
	Technology: Inability to detect unusual user inputs, bypass of GenAI model guardrails	Loss of sensitive information, data leak of PII, reputational damage
GenAI Model and Output Manipulation	**People:** Insider threats, lack of access limitation to foundational model and training data	Unauthorised data access and loss of data integrity
	Process: Lack of proper access control to GenAI model, improper data governance for data used to train GenAI models, lack of contigency measures for GenAI solutions	Unauthorised data access, poisioning of foundation model data, impact fo business operations due to disruptions to GenAI solutions
	Technology: Inability to monitor model performance, model drift unexpected behaviours, inability to detect unusual model outputs	Incorrect information provided to users, reputational damage, regulatory consequences

GenAI Cyber Security Solutions

Pre-Adoption

- Threat Modelling and Risk Assessment
- Readiness Review (NIST AI RMF / ISO/IEC 42001 / SG Model AI Governance Framework for Generative AI)
- Safety Review of GenAI (using Adversial Testing and/or AI Verify Moonshot / MLCommons AI Safety Benchmark)
- Development of AI Incident Response Playbooks
- Employee Awareness Training

Post-Adoption

- Configuration Audit & Compliance Review (NIST AI RMF / ISO/IEC 42001 / SG Model AI Governance Framework for Generative AI)
- Adversial Testing (Based on MITRE ATLAS & OWASP Top 10 for LLM
- Regular Safety Review of inhouse GenAI (AI Verify Moonshot / MLCommons AI Safety benchmark)
- Incident Response Simulation Exercises with AI Scenarios

Readiness Review

NIST AI Risk Management Framework
Assess the readiness in accordance to NIST AI RMF which covers the 6 Govern areas, 5 Map areas, 4 Measure areas and 4 Manage areas.

ISO/IEC 42001
Assess the readiness in accordance to ISO/IEC 42001 AI Management System Management Clauses (1 to 10) and Annex A (A1 to A10).

SG Model AI Governance Framework for Generative AI
Assess readiness according to Singapore's AI Verify Foundation's Model AI Governance Framework, covering Accountability, Data, Development, Security, Testing, Safety, and AI for Public Good.

Safety Review

Adversarial Testing
Assess if popular GenAI tools expose client vulnerabilities or disclose sensitive information.

AI Verify Moonshot
Using AI Verify Moonshot, perform safety review of AI model to provide an overview of how safe the AI model is.

MLCommons AI Safety Benchmark
Using MLCommons AI Safety Benchmark, perform safety review of AI model to provide an overview of how safe the AI model is.

Digital Forensics and Incident Response

Pre-Incident Logging Configuration Review
Review if adequate logging of GenAI services is enabled and retained for sufficient amount of time.

Deepfake Detection
Help our clients to identify if videos, images or audio files, submitted by our clients, are deepfakes.

Post-Incident Forensics Investigation
Analyse GenAI logs and endpoint logs to identify cause(s) of cyber incidents, reconstruct the chronology of events and propose recommendations.

Rajah & Tann Cybersecurity

INFO@RTCYBER.COM

WWW.RTCYBER.COM