सत्यमेव जयते

Blueprint for Reducing Exposure and Defending against AI-Assisted Vulnerabilities Exploitation in Digital Infrastructure

Issued by:

Indian Computer Emergency Response Team (CERT-In)

Table of Contents

1. Executive Summary

The rapid advancement and accessibility of Artificial Intelligence (AI), including generative AI, large language models (LLMs), autonomous agents, and AI-enabled automation platforms, are significantly transforming the cybersecurity landscape. Threat actors are increasingly leveraging AI capabilities to accelerate reconnaissance, automate vulnerability discovery, generate highly targeted phishing campaigns, develop adaptive malware, and enhance the scale and speed of cyber-attacks.

AI-assisted cyber exploitation reduces the time required for adversaries to identify, weaponize, and exploit vulnerabilities, exposed services, weak identities, insecure APIs, and misconfigured systems. As organisations become increasingly dependent on interconnected digital infrastructure, cloud ecosystems, software supply chains, operational technologies, and AI-enabled platforms, the potential impact of AI-enabled cyber threats continues to increase across sectors.

In view of the evolving threat landscape, this document titled *"Blueprint for Reducing Exposure and Defending against AI-Assisted Vulnerabilities Exploitation in Digital Infrastructure"* has been developed by CERT-In to support organisations in strengthening resilience against AI-enabled cyber threats.

The blueprint provides a structured and implementation-oriented framework covering:

i. Governance and accountability mechanisms.

ii. Exposure reduction strategies.

iii. Technical defensive controls.

iv. AI-aware security operations.

v. Vulnerability and exposure management.

vi. Supply-chain security.

vii. Incident response and cyber resilience.

viii. Continuous security validation.

ix. Workforce preparedness and operational readiness.

Organisations are encouraged to implement the recommendations contained in this blueprint in a risk-informed manner based on operational criticality, technology dependencies, and threat conditions. Given the rapidly evolving nature of AI-assisted cyber threats, organisations should continuously reassess exposure, validate security controls, strengthen resilience capabilities, and enhance operational preparedness through ongoing audits, monitoring, testing, and coordinated cybersecurity governance.

2. Background and Context

The global cybersecurity landscape is undergoing a significant transformation driven by the rapid advancement, accessibility, and operationalisation of Artificial Intelligence (AI) technologies. While these technologies offer significant benefits for innovation and operational efficiency, they are also being leveraged by malicious actors to enhance offensive cyber operations.

AI-assisted cyber exploitation is increasingly characterised by:

i. Rapid reconnaissance and attack surface mapping.

ii. Automated vulnerability discovery and exploit development.

iii. Highly personalised phishing and social engineering campaigns.

iv. AI-generated malware and malicious scripting.

v. Deepfake-enabled impersonation and fraud.

vi. Automated attack orchestration.

vii. Adaptive evasion techniques.

The emergence of autonomous and agentic AI systems further increases the potential for semi-autonomous or fully automated cyber operations capable of accelerating multiple stages of the cyber kill chain, including reconnaissance, exploitation, privilege escalation, lateral movement, and data exfiltration.

At the same time, organisations across sectors are becoming increasingly dependent on cloud-native infrastructure, APIs, interconnected digital services, software supply chains, operational technologies, and AI-enabled platforms. This growing digital interconnectivity has expanded the organisational attack surface and amplified systemic cyber risk. Vulnerabilities affecting a single software component, cloud service, AI integration, supplier, or exposed interface may potentially create cascading impact across interconnected environments.

As exploitation timelines reduce, the duration for which vulnerabilities, exposed services, weak credentials, insecure APIs, and misconfigured systems remain discoverable and exploitable becomes a critical cybersecurity risk factor.

Organisations can no longer rely solely on periodic assessments or reactive security approaches. Instead, cybersecurity programs should increasingly prioritise:

i. Continuous exposure management.

ii. Rapid remediation and containment.

iii. Continuous monitoring and threat detection.

iv. Strong identity and access governance.

v. Segmentation and least-privilege architecture.

vi. Threat-informed defence.

vii. Operational resilience.

viii. Continuous Audits and validation of security controls.

This blueprint has been developed to assist entities in reducing exposure to AI-assisted cyber exploitation through structured governance, defensive controls, continuous monitoring, operational preparedness, resilience enhancement, and adaptive cybersecurity practices aligned with evolving threat conditions.

3. Objectives of the Blueprint

This blueprint aims to provide organisations with a structured and implementation-oriented framework for reducing exposure to AI-assisted cyber exploitation and strengthening cyber resilience against evolving AI-enabled threats.

The objectives of this blueprint are to:

i. Improve organisational understanding of AI-assisted cyber threats, including AI-enabled reconnaissance, phishing, malware generation, exploitation, impersonation, and automated attack techniques.

ii. Reduce exploitable exposure across internet-facing assets, identities, APIs, cloud environments, AI systems, and third-party dependencies.

iii. Strengthen cybersecurity governance, accountability, risk management, and executive oversight mechanisms.

iv. Enhance technical and operational security controls across identity, endpoint, network, application, cloud, AI, and operational technology environments.

v. Strengthen AI-aware security operations, threat monitoring, detection engineering, threat hunting, and incident response capabilities.

vi. Promote continuous vulnerability and exposure management, including rapid remediation, attack surface reduction, and validation of security controls.

vii. Strengthen supply-chain and third-party security assurance, including software, AI models, cloud, and dependency risk management.

viii. Promote continuous security validation through audits & assessments, adversarial testing, red teaming, and independent assurance mechanisms.

ix. Encourage timely threat intelligence sharing, coordinated response, and cybersecurity collaboration among relevant stakeholders and authorities, including CERT-In, where applicable.

4. Threat Landscape: AI-Assisted Cyber Exploitation

Key threat areas associated with AI-assisted cyber exploitation include:

4.1 AI-Enabled Reconnaissance and Vulnerability Exploitation:

Threat actors may use AI systems to automate:

i. Attack surface discovery.

ii. OSINT aggregation.

iii. Identification of exposed services and APIs.

iv. Vulnerability analysis.

v. Exploit adaptation and chaining.

vi. Malicious code generation.

vii. Automated exploitation workflows.

AI-assisted capabilities significantly accelerate attack preparation and exploitation timelines.

4.2 AI-Driven Phishing, Impersonation, and Social Engineering

AI technologies are increasingly being used to generate highly convincing phishing content, impersonation attempts, synthetic identities, and deepfake-enabled fraud.

Threat scenarios may include:

i. Spear phishing campaigns.

ii. Executive impersonation.

iii. Deepfake voice and video fraud.

iv. Business email compromise.

v. Credential theft campaigns.

vi. AI-generated social engineering at scale.

Such attacks may bypass traditional awareness-based detection mechanisms due to their realism, contextual accuracy, and personalization.

4.3 AI-Generated Malware and Automated Attack Operations

AI-assisted offensive tooling may enable:

- End-to-end Cyber Kill Chain (CKC) execution.
- Malware modification and obfuscation.
- Adaptive payload generation.
- Automated scripting.
- Evasion of static detection controls.
- Semi-autonomous attack execution.
- Lowering of technical entry barriers, enabling even non-expert / semi-skilled / untrained threat actors to launch sophisticated cyber-attacks at scale.

The emergence of agentic AI systems further increases the possibility within highly compressed time frame of automated multi-stage cyber operations involving reconnaissance, exploitation, persistence, lateral movement, and data exfiltration.

4.4 Adversarial Threats Against AI Systems

Organisations deploying AI-enabled systems may themselves become targets of adversarial attacks against AI models, inference systems, retrieval mechanisms, and AI-integrated workflows.

Potential risks include:

i. Prompt injection.

ii. Model manipulation.

iii. Training data poisoning.

iv. Insecure AI integrations.

v. AI model theft.

vi. Sensitive data leakage.

vii. Compromise of AI orchestration pipelines.

4.5 Risks to Vital Infrastructure and Digital Ecosystems

Vital sectors including government, finance, telecommunications, Digital Public Infrastructure, healthcare, energy, transportation, manufacturing, and digital services may face elevated exposure due to increasing dependence on interconnected digital infrastructure, cloud ecosystems, operational technologies, and AI-enabled systems.

AI-assisted cyber exploitation targeting such environments may result in:

i. Operational disruption.

ii. Compromise of sensitive information.

iii. Financial fraud.

iv. Disruption of critical services.

v. Broader national security implications.

4.6 Evolving Nature of AI-Assisted Threats

AI-assisted cyber threats are expected to continue evolving rapidly in terms of automation, scalability, adaptability, and operational sophistication. Organisations should recognise that:

- Exploitation timelines are reducing significantly.
- Attacks will become increasingly autonomous.
- Traditional static security approaches would become insufficient.
- Continuous monitoring, rapid remediation, adaptive defence using AI, and resilience-focused cybersecurity practices are increasingly necessary.

The evolving nature of AI-assisted cyber exploitation necessitates continuous threat assessment, proactive exposure reduction, operational preparedness, and ongoing enhancement of cybersecurity capabilities using AI across organisations and sectors.

5. Core Defensive Principles

Organisations should adopt AI enabled adaptive, intelligence-driven, continuously validated, and resilience-oriented cybersecurity practices to reduce exposure to AI-assisted cyber threats. Traditional perimeter-centric and periodic compliance-driven security approaches are required but may not be sufficient against rapidly evolving AI-enabled adversarial activity.

The following defensive principles should guide organisational cybersecurity strategy, architecture, implementation, and assurance practices:

S. No	Principle	Objective	Indicative Measures
1.	Assume Breach	Prepare for rapid detection, containment, and recovery from compromise scenarios.	Continuous monitoring, segmentation, telemetry, rapid incident response mechanisms, breach simulations
2.	Zero Trust Security	Enforce continuous verification and least-privilege access.	Multi-Factor Authentication (MFA), Privileged Access Management (PAM), micro segmentation, conditional access, session monitoring
3.	Defence-in-Depth	Implement layered controls across infrastructure, applications, identities, cloud, and AI systems.	Endpoint protection, Data Loss Prevention (DLP), secure configurations, backup and recovery, integrated monitoring
4.	Continuous Exposure Management	Continuously identify and reduce exploitable exposure.	Attack surface monitoring, vulnerability scanning, cloud posture assessment, remediation validation

S. No	Principle	Objective	Indicative Measures
5.	Secure-by-Design and Secure-by-Default	Embed security into systems, applications, and AI workflows from inception.	Threat modelling, secure coding, Continuous Integration (CI) / Continuous Delivery (CD) security testing, hardened configurations
6.	Threat-Informed Defence	Align defensive measures with evolving adversarial tactics and threat intelligence.	Threat intelligence integration, threat hunting, detection engineering, red/purple teaming
7.	Resilience-Centric Security	Maintain operational continuity during cyber incidents and disruption scenarios.	Business continuity, disaster recovery, immutable backups, crisis communication
8.	Security Automation with Human Oversight	Leverage automation while maintaining accountability for high-impact decisions.	SOAR workflows, automated triage, human approval for critical actions, audit trails
9.	Data-Centric Security	Protect sensitive and operationally critical data throughout its lifecycle.	Data classification, encryption, DLP, access governance, secure retention
10.	Supply-Chain Trust and Verifiability	Reduce risks arising from third-party software, AI models, and dependencies.	Vendor assessments, SBOM/xBOM, provenance validation, third-party governance

S. No	Principle	Objective	Indicative Measures
11.	Continuous Validation, Audits and Assurance	Continuously test security effectiveness against evolving threats.	Vulnerability assessments, penetration testing, adversarial simulations, independent audits
12.	Proportional and Risk-Based Implementation	Prioritise controls based on operational criticality and threat exposure.	Enhanced protection for critical systems, privileged identities, cloud management planes, OT environments

These principles should form the foundation for governance, technical controls, security operations, incident response, resilience planning, and assurance activities across organisational environments.

6. Cybersecurity Governance and Operational Readiness

Effective defence against AI-assisted cyber threats requires strong governance, defined accountability, continuous risk assessment, and organisational preparedness.

The following governance and organisational measures should be implemented:

S. No	Governance Area	Objective	Indicative Measures
1.	Leadership Oversight and Accountability	Ensure executive visibility and accountability for cybersecurity and AI-related risks.	Governance structure, executive review, risk ownership, cross-functional coordination, resource allocation
2.	Cybersecurity Governance Framework	Establish formal governance mechanisms aligned with organisational and regulatory requirements.	Security policies, risk management, incident governance, escalation mechanisms, assurance processes
3.	AI Governance and Responsible AI Usage	Govern organisational use of AI systems and reduce risks arising from insecure AI adoption.	AI usage policies, AI risk assessments, shadow AI monitoring, governance of AI integrations
4.	Risk Management and Exposure Assessment	Continuously assess cyber risk exposure across digital environments and dependencies.	Risk assessments, asset inventories using xBOM, exposure reviews, third-party risk assessment, impact analysis

S. No	Governance Area	Objective	Indicative Measures
5.	Policy and Standards Management	Establish and maintain cybersecurity policies and operational standards.	IAM policies, cloud security standards, secure development practices, AI usage controls, incident reporting procedures
6.	Organisational Roles and Responsibilities	Define accountability and operational ownership for cybersecurity functions.	System ownership, escalation responsibilities, third-party oversight, operational coordination
7.	Workforce Awareness and Capacity Building	Improve preparedness against AI-enabled phishing, impersonation, and social engineering threats.	Security awareness programs, deepfake awareness, role-based training, incident reporting awareness
8.	Third-Party and Supply-Chain Governance	Manage risks arising from vendors, cloud services, software dependencies, and AI providers.	xBOMS, Vendor assessments, contractual controls, dependency visibility, supplier reassessment
9.	Security Metrics and Governance Reporting	Monitor organisational security posture and operational readiness.	Vulnerability metrics, incident response timelines, exposure trends, monitoring coverage, governance reviews
10.	Incident Governance and	Establish governance for crisis management	Escalation procedures, executive communication, evidence

S. No	Governance Area	Objective	Indicative Measures
	Escalation Readiness	and cyber incident coordination.	preservation, regulatory coordination
11.	Security Assurance and Continuous Assessment	Validate effectiveness of governance, controls, and operational readiness.	Continuous cybersecurity audits and assessments from CERT-In empanelled Auditing Organisations in alignment with the *Comprehensive Cyber Security Audit Policy Guidelines* and other relevant guidelines issued by CERT-In (https://www.cert-in.org.in/PDF/Comprehensive_Cyber_Security_Audit_Policy_Guidelines.pdf) from time to time, control validation, risk reviews, compliance assessments.
12.	Continuous Improvement and Adaptive Readiness	Continuously refine governance and preparedness based on evolving threats and operational learnings.	Threat intelligence integration, periodic review, operational assessments, resilience enhancement

Cybersecurity governance and organisational readiness should be continuously reviewed and updated to address evolving AI-assisted cyber threats, changing technology environments, and organisational risk exposure.

7. Technical Defensive Controls

Organisations should implement layered, risk-based, and continuously validated technical controls to reduce exposure to AI-assisted cyber threats. Controls should prioritise protection of internet-facing systems, critical business applications, identities, cloud environments, APIs, sensitive data, AI-enabled systems, and operational infrastructure.

The following technical control areas should be implemented based on organisational risk exposure and operational criticality:

S.No.	Control Area	Objective	Indicative Controls
1.	Asset Visibility and Attack Surface Management	Maintain continuous visibility into exposed assets and reduce unmanaged exposure.	Asset inventory, attack surface monitoring, shadow IT/AI detection, cloud exposure monitoring, dependency tracking, adherence to **Technical Guidelines on | SBOM | QBOM & CBOM Version 2.0 issued by CERT-In.** (https://www.cert-in.org.in/PDF/TechnicalGuidelines-on-SBOM,QBOM&CBOM,AIBOM_and_HBOM_ver2.0.pdf)
2.	Identity and Access Security	Strengthen authentication, privileged access governance, and identity protection.	MFA, PAM, least privilege, adaptive authentication, service account governance, access reviews

S.No.	Control Area	Objective	Indicative Controls
3.	Endpoint and Server Security	Protect endpoints and workloads against malware, exploitation, and unauthorised activity.	EDR/XDR, system hardening, patching, application control, behavioural monitoring
4.	Network Security and Segmentation	Restrict unauthorised access and limit lateral movement.	Segmentation, microsegmentation, IDS/IPS, secure remote access, DNS protection
5.	Email, Messaging, and Collaboration Security	Reduce exposure to phishing, impersonation, and communication-based attacks.	Anti-phishing controls, SPF/DKIM/DMARC, executive verification procedures, collaboration monitoring
6.	Application and API Security	Secure applications and APIs against exploitation and abuse.	Secure SDLC, SAST/DAST/SCA, API security testing, secrets management, penetration testing, **Adherence to Guidelines for Secure Application Design, Development, Implementation & Operations issued by CERT-In.** (https://www.cert-in.org.in/PDF/Application_Security_Guidelines.pdf)
7.	Cloud Security	Protect cloud infrastructure and workloads from	Secure cloud configuration, identity and access controls,

S.No.	Control Area	Objective	Indicative Controls
		misconfiguration and identity compromise.	data protection mechanisms, and continuous monitoring of cloud environments
8.	Data Protection and Information Security	Protect sensitive and operationally critical data from unauthorised access and leakage.	Data classification, encryption, DLP, access governance, secure retention, AI data usage controls
9.	AI System Security	Secure AI models, inference systems, orchestration pipelines, and AI integrations.	Prompt injection protection, AI access controls, model monitoring, AI logging, adversarial testing
10.	Security Logging, Monitoring, and Telemetry	Maintain visibility for detection, investigation, and response activities.	Centralised logging, SIEM integration, telemetry correlation, anomaly detection, alert prioritisation
11.	Backup, Recovery, and Resilience Controls	Ensure recoverability and operational continuity during cyber incidents.	Immutable backups, restoration testing, recovery objectives, isolated backup infrastructure
12.	OT and Critical Infrastructure Security	Reduce Exposure and Protect operational technology and cyber-physical environments.	IT/OT segregation, defence-in-depth architecture, industrial monitoring, remote access restrictions, network isolation and air-gapped

S.No.	Control Area	Objective	Indicative Controls
			environments supported by data diodes/unidirectional gateways where appropriate, vendor and third-party access governance, supply-chain risk management, dependency visibility, secure backup and recovery mechanisms, and continuous monitoring of OT environments
13.	Secure Configuration and Hardening	Reduce exploitable exposure arising from insecure configurations.	Hardened baselines, configuration audits, drift monitoring, secure administrative controls
14.	Continuous Validation of Technical Controls	Continuously validate effectiveness of deployed security controls.	Vulnerability assessments, internal and external AI-assisted vulnerability assessments, penetration testing, adversarial simulations, exposure validation, recovery validation, continuous security control testing

Organisations should strengthen software, AI-model, and digital supply-chain visibility through adoption of Software Bill of Materials (SBOM), AI Bill of Materials (AIBOM), Quantum Bill of Materials (QBOM), Cryptographic Bill of Materials (CBOM), and related xBOM mechanisms. Such mechanisms help improve transparency,

component visibility, dependency tracking, provenance validation, vulnerability impact assessment, rapid exposure identification, and coordinated remediation across interconnected software, cloud, AI, and third-party ecosystems. Adoption of xBOM frameworks also supports improved supply-chain assurance, operational resilience, and defence against AI-assisted exploitation targeting vulnerable or compromised dependencies.

Technical controls should be continuously reviewed and updated based on evolving threat intelligence, organisational exposure, technology adoption, operational dependencies, and emerging AI-assisted attack techniques.

8. AI-Aware Security Operations and Continuous Monitoring

Organisations should strengthen security operations and monitoring capabilities to detect, analyse, and respond to AI-assisted cyber threats. Traditional static and signature-based approaches would be insufficient against rapidly evolving AI-enabled attack techniques involving automation, behavioural evasion, impersonation, and large-scale exploitation.

Security operations should support continuous visibility, intelligence-driven detection, rapid response, proactive threat hunting, and coordinated incident analysis across enterprise, cloud, AI, identity, application, and operational technology environments.

S. No.	Security Operations Area	Objective	Indicative Measures
1.	Security Operations Modernisation	Strengthen SOC capabilities using AI (Agentic SOC) for continuous monitoring and rapid response.	Centralised monitoring, telemetry correlation, alert prioritisation, visibility into critical and internet-facing systems
2.	Threat Intelligence Integration	Improve detection and response using	IOC correlation, monitoring of AI-assisted attack trends, threat

S. No.	Security Operations Area	Objective	Indicative Measures
		actionable threat intelligence.	intelligence integration, sector-specific threat tracking
3.	Detection Engineering	Develop adaptive and behaviour-based detection mechanisms.	Behavioural analytics, anomaly detection, detection rule tuning, monitoring of privilege escalation and suspicious activity
4.	Behavioural Analytics and Anomaly Detection	Identify abnormal activity that may evade traditional controls.	User behaviour monitoring, cloud anomaly detection, suspicious API activity, operational anomaly analysis
5.	Continuous Security Monitoring	Maintain visibility across enterprise, cloud, AI, and operational environments.	Endpoint monitoring, identity monitoring, network telemetry, cloud logging, API monitoring, AI activity logging
6.	Threat Hunting	Proactively identify malicious or stealthy activity within organisational environments.	Threat hunting exercises, investigation of anomalous behaviour, suspicious identity activity, cloud and AI misuse detection

S. No.	Security Operations Area	Objective	Indicative Measures
7.	AI-Assisted Defensive Operations	Improve operational efficiency through controlled use of AI-assisted security capabilities.	Automated triage, threat correlation, alert enrichment, investigation support, detection optimisation
8.	Monitoring of AI Systems and AI Usage	Maintain visibility into AI systems, integrations, and operational behaviour.	Monitoring AI access, prompt activity, inference behaviour, AI API usage, prompt injection detection
9.	Incident Correlation and Contextual Analysis	Correlate events across environments to identify coordinated attacks.	Cross-domain event correlation, attack path analysis, contextual enrichment, identity and cloud activity correlation
10.	Security Automation and Orchestration	Improve response speed and operational efficiency through controlled automation.	Automated workflows, IOC ingestion, endpoint isolation, response orchestration, approval-based automation
11.	Deepfake and Impersonation Detection Readiness	Strengthen readiness against AI-enabled impersonation and fraud.	Verification procedures, deepfake awareness, executive impersonation monitoring, escalation mechanisms including reporting to appropriate law enforcement

S. No.	Security Operations Area	Objective	Indicative Measures
			agencies and relevant authorities.
12.	SOC Workforce Readiness and Skills Development	Enhance operational capability of security personnel against evolving threats.	Detection engineering training, AI threat awareness, cloud security monitoring, incident investigation training
13.	Continuous Improvement and Operational Adaptation	Continuously refine monitoring and response capabilities based on evolving threats.	Detection review, telemetry gap assessment, incident learnings, operational tuning, threat-informed updates

Security operations and monitoring capabilities should be continuously reviewed and enhanced based on evolving AI-assisted attack techniques, operational observations, emerging threat intelligence, and organisational risk exposure.

9. Vulnerability and Patch Management

Organisations should adopt continuous, risk-based vulnerability and patch management practices to reduce exploitable exposure arising from vulnerabilities, misconfigurations, insecure APIs, exposed services, weak identities, cloud exposure, and third-party dependencies.

Vulnerability and exposure management should prioritise rapid identification, remediation, validation, and continuous reduction of attack surface across enterprise, cloud, AI, application, operational technology, and supply-chain environments. Organisations should also maintain continuous situational awareness regarding newly disclosed vulnerabilities, emerging threat intelligence, exploitation trends, adversarial techniques, and security advisories to ensure timely risk assessment, prioritisation, and proactive defensive response across their digital ecosystem.

S. No.	Control Area	Objective	Indicative Measures
1.	Continuous Vulnerability Management	Continuously identify and remediate vulnerabilities across organisational environments.	Vulnerability scanning, internet-facing assessments, cloud and API assessments, remediation validation
2.	Risk-Based Prioritisation	Prioritise remediation based on exploitability, exposure, and operational impact.	Known Exploited Vulnerabilities (KEV) prioritisation, Exploit Prediction Scoring System (EPSS) based exploit likelihood assessment, exploitability analysis, business criticality assessment, threat intelligence integration
3.	Patch and Remediation Management	Ensure timely and controlled remediation of	Patch management workflows, emergency remediation, exception handling, remediation tracking

S. No.	Control Area	Objective	Indicative Measures
		identified vulnerabilities.	
4.	Validation of Remediation Effectiveness	Validate that remediation actions effectively remove exploitable exposure.	Rescanning, penetration testing, configuration validation.

Indicative Risk-Based Remediation Timelines

S. No.	Finding Type	Indicative Remediation Expectation
1.	Known exploited vulnerability affecting internet-facing and crown-jewel systems	Immediate containment; patch, mitigate, or remove exposure within 12 hours where feasible.
2.	Critical externally exposed vulnerability	Patch, mitigate, or remove exposure within 1 day.
3.	Known exploited vulnerability affecting internal systems	Patch or mitigate within 1 day unless compensating controls are implemented and documented.
4.	Critical internal vulnerability affecting high-value systems	Patch or mitigate within 3 days.
5.	High-severity vulnerability	Patch or mitigate within 5 days based on risk prioritisation.
6.	No patch available	Implement temporary mitigation measures such as isolation, access restriction, WAF/API protection, enhanced monitoring, or feature

S. No.	Finding Type	Indicative Remediation Expectation
		disablement until remediation becomes available.

10. Incident Response and Cyber Resilience

Organisations should establish incident response and cyber resilience capabilities to rapidly detect, contain, investigate, respond to, and recover from cyber incidents.

S. No.	**Incident Response Area**	**Objective**	**Indicative Measures**
1.	Incident Response Governance	Establish structured governance and escalation mechanisms during cyber incidents.	Incident classification, escalation procedures, executive communication, evidence preservation
2.	AI-Aware Incident Preparedness	Prepare for AI-assisted attack scenarios and adversarial AI risks.	Deployment of mechanisms for detection, alerting, and response to AI-enabled phishing attacks, along with deepfake response procedures and cloud/AI incident handling
3.	Detection, Triage, and Containment	Rapidly identify, prioritise, and contain cyber incidents.	Telemetry correlation, alert validation, endpoint isolation, credential revocation, network containment

S. No.	Incident Response Area	Objective	Indicative Measures
4.	Investigation and Analysis	Assess scope, impact, and progression of incidents.	Log analysis, malware analysis, forensic review
5.	Backup, Recovery, and Operational Resilience	Restore operations securely following cyber incidents.	Immutable backups, restoration testing, recovery validation, business continuity planning
6.	Communication and Coordination	Enable coordinated response across stakeholders and authorities.	Internal communication, regulatory coordination, stakeholder notification, incident reporting
7.	Post-Incident Review and Continuous Improvement	Strengthen resilience through lessons learned and operational review.	Root cause analysis, control validation, response improvement, resilience testing

Organisations should conduct incident response exercises, cyber resilience testing, backup restoration validation, adversarial simulations, and table-top exercises to assess operational readiness against evolving AI-assisted cyber threats.

Organisations are encouraged to participate in technical exercises, cyber drills, simulations, and table-top exercises conducted by CERT-In from time to time for strengthening cyber resilience, incident coordination, and response preparedness.

Entities should ensure timely reporting of cyber incidents to CERT-In in accordance with the directions issued by CERT-In from time to time, including reporting of cyber incidents within 6 hours.

11. Security Validation, Assurance, and Continuous Testing

Organisations should establish continuous and risk-based security validation mechanisms to assess the effectiveness of cybersecurity controls, monitoring capabilities, incident response readiness, and operational resilience against evolving AI-assisted cyber threats.

S. No.	Validation Area	Objective	Indicative Measures
1.	Continuous Security Assurance	Continuously assess effectiveness of implemented controls and operational readiness.	Exposure assessment, monitoring validation, privileged access review, cloud and AI security review
2.	Vulnerability Assessment and Penetration Testing	Identify exploitable weaknesses across enterprise environments.	Internal/external VAPT, API testing, cloud assessment, segmentation validation, remediation verification
3.	Red Teaming and Adversarial Simulation	Evaluate resilience against sophisticated attack scenarios and adversarial techniques.	Multi-stage attack simulation, phishing simulation, cloud compromise testing, lateral movement assessment
4.	AI System Security Testing	Assess AI systems against adversarial manipulation and misuse.	Prompt injection testing, AI API assessment, model integrity review, AI workflow validation
5.	Supply-Chain and Third-Party Assurance	Validate security posture of vendors, dependencies, and externally integrated services.	Vendor assessments, dependency review, software provenance assessment, third-party access review

S. No.	Validation Area	Objective	Indicative Measures
6.	Security Operations and Monitoring Validation	Validate effectiveness of monitoring, detection, and response capabilities.	Detection testing, SIEM validation, alert quality review, telemetry assessment
7.	Business Continuity and Recovery Testing	Validate continuity and recovery capability during cyber incidents.	Backup restoration testing, disaster recovery exercises, operational continuity drills
8.	Tabletop Exercises and Crisis Simulations	Assess organisational coordination and incident response readiness.	Tabletop exercises, ransomware simulations, crisis communication testing
9.	Independent Audits and Assurance Reviews	Conduct independent evaluation of cybersecurity posture and control effectiveness.	Cybersecurity audits, resilience assessments, architecture reviews, sector-specific assessments

Organisations should conduct Red Teaming & cybersecurity audits, security assessments, adversarial simulations, and resilience validation exercises to assess effectiveness of implemented controls and operational preparedness. Where applicable, such assessments may be conducted through CERT-In empanelled Information Security Auditing Organisations in alignment with the *Comprehensive Cyber Security Audit Policy Guidelines* (https://www.cert-in.org.in/PDF/Comprehensive_Cyber_Security_Audit_Policy_Guidelines.pdf) **and other relevant guidelines issued by CERT-In from time to time.**

12. Secure Adoption and Governance of Artificial Intelligence System

Organisations are increasingly adopting Artificial Intelligence (AI) technologies across operational workflows, analytics platforms, automation environments, software development, cybersecurity operations, and decision-support systems. While AI adoption offers significant operational and efficiency benefits, it also introduces evolving cybersecurity, governance, privacy, operational, and supply-chain risks. The growing use of publicly accessible AI platforms, large language models (LLMs), autonomous agents, AI APIs, and AI-assisted automation tools may create unmanaged exposure if appropriate governance, security, monitoring, and validation mechanisms are not established. Organisations should therefore adopt a structured, risk-aware, and security-centric approach toward enterprise AI adoption and deployment.

Accordingly, the following areas should be considered for strengthening secure AI governance, operational resilience, risk management, and protection of AI-enabled organisational environments:

S.No.	Area	Objective	Key Organisational Measures
1.	**AI Governance and Oversight**	Establish governance and accountability mechanisms for secure AI adoption and operation.	•Define AI usage policies and accountability structures • Establish approval and review mechanisms for AI integrations • Define monitoring, audit, security, and privacy obligations
2.	**Inventory and Visibility of AI Systems**	Maintain visibility into AI systems, integrations, APIs, and third-party AI dependencies.	• Maintain AI asset inventories • Monitor AI APIs, plugins, and automation workflows • Identify shadow or unauthorised AI usage

S.No.	Area	Objective	Key Organisational Measures
3.	**Risk Assessment for AI Deployments**	Assess cybersecurity, operational, privacy, and compliance risks prior to AI deployment.	• Evaluate data exposure and integration risks • Assess third-party dependency and resilience risks
4.	**Secure Use of Public AI Platforms**	Prevent unauthorised disclosure of organisational or regulated information through public AI services.	• Restrict upload of sensitive information • Define acceptable usage policies • Conduct awareness and approval-based governance
5.	**AI System Security Controls**	Protect AI systems and associated workflows through layered security controls.	• Implement access control and authentication • Secure APIs and orchestration pipelines • Monitor AI usage and maintain audit logs
6.	**Prompt Injection and Input Manipulation Risks**	Mitigate risks arising from prompt manipulation and unsafe AI interactions.	• Validate and sanitise external inputs • Restrict AI permissions and execution capability • Conduct adversarial testing and behavioural monitoring
7.	**Data Protection and Privacy in AI Systems**	Ensure secure and compliant handling of data processed by AI systems.	• Classify and protect sensitive data • Define retention and deletion policies • Monitor AI-related data

S.No.	Area	Objective	Key Organisational Measures
			movement and third-party handling
8.	**Third-Party AI Provider and Supply-Chain Risk**	Manage risks arising from external AI models, APIs, platforms, and cloud services.	• Assess provider security posture • Review contractual and data handling obligations • Maintain contingency mechanisms for critical dependencies
9.	**AI System Monitoring and Logging**	Maintain operational visibility into AI activities and anomalous behaviour.	• Monitor model access and API usage • Correlate AI telemetry with enterprise security monitoring
10.	**Human Oversight and Decision Governance**	Ensure human validation and accountability for AI-assisted decisions.	• Validate AI-generated outputs • Restrict fully autonomous critical actions • Maintain auditability and approval mechanisms
11.	**AI in Software Development and DevSecOps**	Reduce risks arising from AI-assisted software development workflows.	• Review AI-generated code and dependencies • Conduct Static Application Security Testing (SAST), Dynamic Application Security Testing (DAST), and dependency analysis • Secure AI-assisted CI/CD workflows

S.No.	Area	Objective	Key Organisational Measures
12.	**AI Model Integrity and Trustworthiness**	Maintain integrity and trustworthiness of AI models and training pipelines.	• Validate model provenance and integrity • Protect against unauthorised modification • Maintain version control and inference validation
13.	**AI Usage Awareness and Workforce Preparedness**	Build organisational awareness regarding AI-related security and operational risks.	• Conduct awareness on phishing, deepfakes, and data exposure • Train personnel on secure AI usage • Establish reporting mechanisms for suspicious activity
14.	**Governance of Autonomous and Agentic AI Systems**	Secure deployment of autonomous AI systems operating with limited human intervention.	• Define operational boundaries and permissions • Maintain continuous monitoring and audit logging • Establish override and emergency shutdown mechanisms
15.	**Continuous Review and Adaptive Governance**	Ensure continuous improvement of AI governance and security posture.	• Periodically reassess AI risks and architecture • Conduct adversarial testing and intelligence review • Update governance policies and controls regularly

S.No.	Area	Objective	Key Organisational Measures
16.	**AI Security Assessment and Assurance**	Continuously assess AI systems for cybersecurity, operational, and trustworthiness risks.	• Conduct assessments for classical software and infrastructure vulnerabilities • Evaluate AI/ML-specific vulnerabilities and adversarial risks • Assess behavioural integrity, bias, fairness, robustness, explainability, and unsafe outputs • Perform continuous validation, red teaming, and assurance testing of AI systems

13. Recommended Implementation Roadmap

Organisations should adopt a phased, risk-based, and operationally practical approach for implementation of the recommendations contained in this blueprint. Implementation priorities should focus on reduction of exploitable exposure, strengthening of foundational controls, enhancement of monitoring capabilities, and improvement of operational resilience against AI-assisted cyber threats.

Phase	Timeline	Key Focus Areas	Indicative Activities
Phase I Immediate Risk Reduction	0–7 days	Foundational governance, exposure reduction,	Establish cybersecurity governance and accountability structures; identify critical assets and internet-facing systems;

Phase	Timeline	Key Focus Areas	Indicative Activities
		identity security, monitoring readiness	implement MFA for critical access; conduct vulnerability assessments; patch critical and known exploited vulnerabilities; reduce unnecessary exposure; establish incident reporting and escalation procedures; enable security logging and baseline monitoring; initiate workforce awareness for AI-assisted phishing and deepfake threats
Phase II Operational Strengthening	8-30 days	Continuous monitoring, exposure management, AI security governance, resilience enhancement	Strengthen SOC and monitoring capabilities; integrate endpoint, cloud, identity, and network telemetry; establish continuous vulnerability and attack surface management; implement behaviour-based detection and threat hunting; establish AI governance and AI system inventory; conduct cloud and API security assessments; strengthen third-party and supply-chain assurance; conduct tabletop exercises, ransomware simulations, and backup restoration testing
Phase III Advanced	31–60 days	Adversarial validation,	Conduct red team exercises and adversarial simulations; implement

Phase	Timeline	Key Focus Areas	Indicative Activities
Resilience and Adaptive Security		automation-assisted defence, advanced resilience capabilities	continuous control validation; enhance security automation and orchestration; adopt AI-assisted defensive operations where appropriate; strengthen operational resilience and continuity planning; conduct adversarial AI testing; validate model integrity and AI orchestration security; continuously reassess organisational exposure and resilience posture

14. Conclusion

The rapid advancement and accessibility of Artificial Intelligence (AI) technologies are significantly transforming the cybersecurity landscape and enabling increasingly sophisticated, scalable, and automated cyber threats. AI-assisted cyber exploitation is accelerating reconnaissance, phishing, impersonation, malware generation, exploit development, and large-scale attack operations across interconnected digital ecosystems.

In this evolving threat environment, organisations should adopt adaptive, intelligence-driven, continuously validated, and resilience-oriented cybersecurity practices rather than relying solely on static controls or periodic compliance-driven assessments.

This blueprint has been developed to support organisations and regulated entities in reducing exposure to AI-assisted cyber threats through strengthened governance, technical controls, security operations, vulnerability management, incident response, continuous validation, and operational resilience.

Organisations should implement the recommendations contained in this blueprint in a risk-informed manner based on operational criticality, threat exposure, and organisational maturity. Continuous monitoring, rapid remediation, adaptive defence, and coordinated cybersecurity preparedness are essential for strengthening resilience against evolving AI-assisted cyber threats and enhancing trust in India's digital ecosystem

Contact Information

CERT-In AI Cyber Defence Center

Email: aicompliance@cert-in.org.in

Incident Reporting: incident@cert-in.org.in

Contact No.: +91-11-22902657