REGULATORY COMPLIANCE WHITE PAPER

DefenGPT AI Security Platform

by Defenix Technology

Achieving Full Compliance with SEBI Circular

HO/13/19/12(1)2026-ITD-1_CIMGI/10873/2026 | Dated 05.05.2026

Advisory on Emerging Advanced AI Tools for Vulnerability Detection (like Mythos)

THE SEBI MANDATE

SEBI Circular 10873/2026 requires Regulated Entities to actively govern AI-driven threat vectors — not just react to them.

THE COMPLIANCE GAP

Traditional SOC, Firewall, DLP, and Cloud Security tools were not designed to govern AI model behaviour and prompt-level risks.

THE ONLY SOLUTION

DefenGPT AI Security Platform is purpose-built to close every gap — deployable on-prem, private cloud, or air-gapped.

10	100%	<30	0
SEBI Advisory requirements directly addressed	AI interaction coverage — prompt in, response out	minutes to deploy out of the box	data egress in air-gapped deployment

On-Premises | Private Cloud | Air-Gapped & Isolated | Public Cloud SaaS

Full Visibility • Total Control • Unified Management • Instant Deployment

SECTION 01

Executive Summary

On 5 May 2026, the Securities and Exchange Board of India (SEBI) issued Circular HO/13/19/12(1)2026-ITD-1_CIMGI/10873/2026, issuing a formal advisory to all Regulated Entities (REs) — including stock exchanges, depositories, mutual funds, brokers, custodians, investment advisors, and every other class of SEBI-regulated institution — on the risks posed by advanced AI-driven vulnerability detection tools such as Claude Mythos. This circular is not a future obligation. It is effective immediately and carries the weight of SEBI's enforcement authority under Section 11(1) of the SEBI Act, 1992.

The circular establishes ten specific technical and governance requirements covering AI risk assessment, API security, SOC monitoring, patch management, change control, asset inventory, system hardening, third-party vendor oversight, and long-term AI governance planning. It also constituted a dedicated task force — cyber-suraksha.ai — to coordinate threat intelligence sharing and vulnerability response across the securities market ecosystem.

This white paper makes the case that no existing cybersecurity product category — not next-generation firewalls, not SIEM/SOC platforms, not Data Loss Prevention (DLP) tools, and not Cloud Security Posture Management (CSPM) solutions which are architecturally capable of meeting the full scope of the SEBI circular's AI-specific requirements. Only DefenGPT AI Security Platform by Defenix Technology, a purpose-built AI governance and security layer, addresses every requirement in the circular — and does so in a deployment model that meets even the most demanding data sovereignty requirements: fully on-premises, private cloud, or air-gapped and completely isolated.

The Core Thesis

Traditional cybersecurity tools protect the network perimeter and data in transit. DefenGPT protects the AI interaction layer — the prompts, the models, the outputs, and the decisions — which is precisely the attack surface SEBI has mandated regulated entities to govern. These are not overlapping capabilities. They are complementary, and in the context of this SEBI circular, DefenGPT is the missing and essential layer.

SECTION 02

Understanding the SEBI Circular — What Is Required

SEBI Circular 10873/2026 was triggered by the recognition that AI-driven tools — specifically referencing Claude Mythos as an illustrative example — are now capable of identifying and exploiting vulnerabilities in financial market infrastructure at machine speed and scale. The circular acknowledges that the interconnectedness of market participants means a single compromised entity can trigger a cascading failure across the ecosystem.

The advisory in Annexure-A sets out ten distinct requirements. Each is examined below in the context of what standard tools can and cannot deliver.

1	**Immediate patching of all OS & applications; virtual patching as interim measure**	✗ Patch management tools exist but do not govern AI model interactions or AI-accelerated exploit discovery.	✓ DefenGPT virtual-patches AI interaction risks in real time —

			blocking AI-assisted exploitation before a system patch is available.
2	**Conduct VA using conventional AND AI-based tools; continuous security audits per SEBI CSCRF**	✗ VAPT tools scan for known CVEs but cannot assess the unique attack surface created by AI prompt interfaces.	✓ DefenGPT continuously monitors all AI interactions for anomalous patterns, prompt injections, and data exfiltration attempts that VAPT tools miss.
3	**Third-party vendors to assess AI-led vulnerability detection risks; implement VAPT, patching, hardening**	✗ Vendor risk management platforms track contractual compliance but cannot technically govern how vendors' AI tools behave.	✓ DefenGPT intercepts and inspects all AI traffic from third-party vendor tools, classifying risk and enforcing policy in real time.
4	**Change management: full documentation, impact analysis, rigorous testing, secure deployment**	✗ ITSM tools document changes but have no mechanism to assess AI-driven risk introduced by model updates or prompt changes.	✓ DefenGPT logs all changes to AI model configurations, prompt templates, and access policies with immutable audit trails.
5	**API Security: inventory, auth/authz, rate limiting, whitelist-only connections**	✗ API gateways enforce authentication but are blind to the semantic content of AI API calls and cannot detect malicious intent embedded in prompts.	✓ DefenGPT governs AI API traffic at the semantic layer: inspects prompt content, classifies intent, enforces whitelist policies, and blocks injection attacks.
6	**SOC Monitoring: 24x7 monitoring, SOAR integration, M-SOC onboarding, AI-driven attack awareness**	✗ SIEM/SOC tools detect network and endpoint anomalies. They generate no alerts for AI-specific threats like prompt injection, model manipulation, or LLM data leakage.	✓ DefenGPT feeds AI-specific threat signals — classified by intent, sensitivity, and risk score — directly into SIEM and M-SOC platforms via native integration.
7	**Risk Assessment: CSCRF-mandated periodic risk assessment including AI model capabilities as risk scenario**	✗ Generic risk assessment frameworks have no AI-specific risk models. GRC tools cannot assess LLM-specific threats without custom extensions.	✓ DefenGPT provides continuous AI risk scoring per user, per model, per department, with evidence packages mapped to SEBI CSCRF requirements.
8	**System hardening: secure config, disable unnecessary services, enforce least privilege & Zero Trust**	✗ Endpoint hardening tools enforce OS-level configurations but cannot apply least-privilege or Zero Trust principles to AI model access.	✓ DefenGPT enforces role-based, least-privilege access to all AI models. Users, agents, and applications receive only the AI access their role permits.
9	**Periodically update Asset Inventory and Software Bill of Materials for all critical applications**	✗ SBOM tools track software components but have no mechanism to inventory AI models, prompt libraries, fine-tuned weights, or AI API integrations.	✓ DefenGPT maintains a live AI asset inventory: all models in use, all API integrations, all prompt templates — updated in real time.
10	**Long-term AI governance plan; AI-augmented SOC; AI-accelerated threat recalibration; continuous VA using AI**	✗ No existing tool category provides a comprehensive AI governance roadmap or the operational infrastructure to govern AI at enterprise scale.	✓ DefenGPT is the enterprise AI governance platform: it provides the operational layer, the compliance evidence, and the strategic framework for long-term AI security.

SECTION 03

Why Standard Cybersecurity Tools Cannot Meet This Challenge

The SEBI circular explicitly references AI-driven vulnerability tools like Claude Mythos as a new and distinct threat category. This matters because it signals that SEBI has recognized that AI represents a qualitatively different risk — not merely a faster version of existing threats, but a fundamentally new attack surface that existing tools were not designed to protect.

To understand why, it is necessary to examine precisely what each standard tool category does and does not do.

Security Tool	What It Protects (Genuine Capability)	What It Cannot Do (AI-Specific Gap)
Next-Gen Firewall (NGFW)	✓ Network perimeter; inspects packets, blocks known-malicious IPs, enforces network segmentation	✗ Cannot inspect the semantic content of AI prompts. An NGFW sees 'HTTPS traffic to api.anthropic.com' it does not see that the prompt contains confidential client data or a prompt injection payload.
SIEM / SOC Platform	✓ Aggregates logs from network, endpoint, cloud; detects anomalies against known patterns; triggers alerts	✗ Has no AI-specific log sources, no AI threat taxonomy, and no ability to parse or classify AI prompt/response content. AI-driven attacks leave no traditional IOCs.
Data Loss Prevention (DLP)	✓ Scans files, emails, and web uploads for known sensitive data patterns (regex, fingerprinting)	✗ DLP cannot understand context. It may block 'account number' in an email but cannot detect that a prompt is asking an AI to 'summarize all client portfolio data' which is semantically the same exfiltration.
Cloud Security Posture Mgmt (CSPM)	✓ Audits cloud configuration against benchmarks; identifies misconfigured storage, IAM policies, exposed APIs	✗ CSPM is concerned with infrastructure configuration, not AI behaviour. It cannot assess how an LLM is being used, what data it processes, or whether its outputs are safe.
Endpoint Detection & Response (EDR)	✓ Monitors endpoint processes, file system, and memory for malware, exploits, and suspicious behavior	✗ AI model exploitation does not touch the endpoint. Prompt injection, data exfiltration via AI, and model manipulation happen entirely in the AI API layer — invisible to EDR.
API Gateway / WAF	✓ Enforces authentication on APIs; blocks OWASP Top 10 web attacks; rate-limits traffic	✗ API gateways enforce access control but are blind to the semantic content of valid, authenticated requests. A user with API access can still exfiltrate data through AI — and the gateway sees nothing wrong.
Vulnerability Assessment / VAPT	✓ Scans infrastructure for known CVEs; tests applications for web vulnerabilities; produces patch lists	✗ VAPT has no concept of AI prompt surfaces, fine-tuned model risks, or agentic AI behaviour. AI-driven vulnerability tools like Mythos operate at a layer VAPT cannot even access.
Identity & Access Management (IAM)	✓ Manages user authentication, role assignments, and access to systems	✗ IAM can confirm 'this user is allowed to use this AI tool' but cannot govern what the user does with the AI tool — the content, intent, or data sensitivity of their interactions.

The Fundamental Gap

All the above tools operate at the infrastructure, network, or data-at-rest layer. The SEBI circular is asking regulated entities to govern the AI interaction layer — the prompts, the model outputs, the agentic decisions, and the data flows within AI sessions. This layer did not exist until recently, and no traditional cybersecurity tool was designed to protect it. DefenGPT was built specifically and exclusively for this purpose.

SECTION 04

The Claude Mythos Threat — What SEBI Is Really Warning About

SEBI's explicit reference to Claude Mythos (the advanced AI capability within Anthropic's Claude platform) as a named threat illustrates a specific concern: that large language model-powered tools can now automate the discovery, analysis, and exploitation of vulnerabilities in financial market infrastructure in ways that were previously only possible for highly skilled human threat actors working over extended time periods.

01

AI-Accelerated Vulnerability Discovery

Mythos and similar LLMs can ingest API documentation, source code fragments, network diagrams, and configuration files, then rapidly identify exploitable weaknesses. What previously took a penetration tester weeks can now be accomplished in hours. SEBI is warning that regulated entities must assume adversaries have access to these same tools.

02

Prompt Injection via Third-Party Integrations

If a regulated entity's systems interact with AI models (for customer service, document analysis, trade research, compliance screening), adversaries can embed malicious instructions within documents, emails, or data feeds that manipulate the AI's behaviour — a technique known as indirect prompt injection. No firewall, DLP, or SIEM system has any awareness of this attack vector.

03

Data Exfiltration Through AI Reformatting

An adversary with access to an AI tool that can access internal data can instruct it to 'summarise' or 'translate' sensitive documents into an output format that bypasses DLP rules. The data is technically not 'copied' — it has been semantically reformatted by an AI — making traditional DLP fingerprinting ineffective.

04

Autonomous Agentic Actions Within Financial Systems

Advanced AI agents can be given access to APIs, databases, and trading systems, then instructed to perform multi-step tasks autonomously. If an agent is manipulated or misconfigured, it may execute transactions, exfiltrate records, or modify configurations without any human review — and generate no logs that traditional SIEM tools would flag as anomalous.

05

Shadow AI and Ungoverned Private LLM Deployments

Staff across regulated entities are using AI tools — both approved and unapproved — to process client data, draft communications, analyze trading patterns, and conduct research. SEBI's mandate implicitly requires that all such usage be governed. Without a purpose-built AI governance layer, there is no mechanism to enforce this requirement.

SECTION 05

DefenGPT AI Security Platform — Capabilities Aligned to SEBI

DefenGPT AI Security Platform by Defenix Technology is purpose-built to govern the AI interaction layer across all deployment environments. It addresses every SEBI circular requirement through four integrated capability pillars, each directly mapping to the advisory's specific mandates.

01 Data Classification & Sensitivity Engine

Governing what data flows into and out of AI systems — SEBI Requirements 2, 6, 7

- Real-time semantic scanning of all prompts and AI-generated outputs across 150+ data categories including PII, PHI, PCI, and securities market-specific data (trade data, MNPI, client portfolios, regulatory filings)
- Sensitivity scoring engine assigns risk levels — Public / Internal / Confidential / Restricted / Top Secret — to every AI interaction before processing
- Contextual classification understands financial domain semantics: recognises trade ticket data, ISIN codes, client account references, and regulatory submission formats as sensitive even when not in standard DLP dictionaries
- Supports custom data taxonomy aligned to SEBI CSCRF sensitivity categories and individual RE data classification policies
- Works across all input types: typed prompts, uploaded documents, spreadsheets, images, and structured API payloads
- Directly addresses SEBI Circular requirements for continuous risk assessment (Req. 7) and SOC-level monitoring of AI data flows (Req. 6)

02 User Intent Analysis Engine

Detecting malicious or risky use of AI before harm occurs — SEBI Requirements 2, 6, 7, 10

- Behavioral AI engine analyses the intent behind every AI interaction — distinguishing legitimate research from data exfiltration attempts, insider threat behaviour, and prompt injection attacks
- Dedicated detection models for SEBI-relevant threat vectors: MNPI leakage, client data harvesting, trade signal extraction, and regulatory circumvention via AI
- Prompt injection detection identifies both direct injection (malicious instructions in user prompts) and indirect injection (malicious instructions embedded in documents fed to AI agents)
- User risk scoring builds a continuous behavioral profile per employee, per department, per AI model — enabling early identification of insider threats and compromised accounts
- Alert severity scoring prioritizes genuine threats over noise, directly supporting SEBI's requirement (Req. 6) for SOC teams to investigate all alerts — including low-priority ones
- Integrates with Active Directory, Okta, and LDAP to correlate intent with role, department, and least privilege access rights

03 Guardrails Policy Engine

Enforcing SEBI-mandated controls automatically and in real time — SEBI Requirements 1, 3, 4, 5, 8

- Centralized policy console with 80+ pre-built policy templates including SEBI CSCRF, SEBI Circular 10873/2026 requirements, ISO 27001, and IRDAI cyber guidelines
- Granular policy actions: Allow / Warn / Redact / Block / Quarantine — applied per data class, intent category, user role, AI model, or API endpoint
- API security governance: enforces whitelist-only AI API connections (Req. 5d), rate limiting (Req. 5c), and least-privilege access control (Req. 5b) at the semantic layer — not just the network layer
- Virtual patching for AI risks: when a new AI-specific vulnerability is identified, DefenGPT deploys a policy-based mitigation in under 60 seconds — before a system patch is available (Req. 1)

- Change management logging: every policy change, model update, or prompt template modification is recorded with full documentation and impact analysis (Req. 4)
- Zero Trust enforcement for AI: every AI interaction is verified against identity, role, data sensitivity, and intent before being permitted to proceed (Req. 8)

04 Auditing, Reporting & AI Asset Inventory

Providing the evidence SEBI auditors, IT committees, and cyber-suraksha.ai require — SEBI Requirements 4, 7, 9, 10

- Immutable audit log captures every AI prompt, response, policy decision, classification result, and user identity with cryptographic tamper detection — satisfying SEBI's continuous audit requirements
- AI Asset Inventory: maintains a live, auto-discovered inventory of all AI models in use, all API integrations, all prompt libraries, and all fine-tuned model deployments — directly addressing Req. 9 (SBOM for AI)
- Pre-built SEBI CSCRF compliance evidence packages: one-click generation of all documentation required for SEBI periodic risk assessment, third-party vendor reviews, and IT committee reporting (Reqs. 7, 10)
- Executive and board-level dashboards: real-time AI risk posture score, violation trends, department-level AI usage heatmaps, and SEBI regulatory readiness indicators
- cyber-suraksha.ai integration: DefenGPT can be configured to automatically report qualifying AI security incidents and vulnerability intelligence to the SEBI task force at project-cyber-suraksha.ai@sebi.gov.in (Req. 6 / Circular Section C)
- M-SOC integration: native connectors for NSE/BSE Market SOC, Splunk, Microsoft Sentinel, and IBM QRadar — feeding AI threat signals into the centralized monitoring platform (Req. 6c)

SECTION 06

Deployment Flexibility — Designed for India's Regulated Entities

SEBI regulated entities span a spectrum of infrastructure maturity — from tier-1 exchanges and custodians with sophisticated multi-cloud environments to regional brokers and investment advisors with on-premises IT. DefenGPT AI Security Platform is engineered to serve all of them, with three certified deployment models that maintain identical functional capability regardless of where the platform runs.

AIR-GAP

On-Premises / Air-Gapped

✓ Zero external data egress — guaranteed
✓ Runs on RE's own hardware
✓ Fully isolated network deployment
✓ No internet connectivity required
✓ Ideal for exchanges, depositories

PRIVATE

Private Cloud

✓ Dedicated single-tenant VPC
✓ RE-controlled encryption keys
✓ Regional data residency in India
✓ FIPS 140-2 encryption in transit & rest
✓ BSE / NSE co-location option
✓ MSSP-hosted model available
✓ Custom SLAs and support tiers

HYBRID

Hybrid / SaaS

✓ Governs both cloud & on-prem AI
✓ Unified policy across all environments
✓ 24-hour rapid deployment
✓ SOC 2 Type II & ISO 27001 certified
✓ 99.99% availability SLA

✓ SEBI CSCRF data residency compliant
✓ Deploy in under 30 minutes

✓ Auto-scaling for peak market hours
✓ Free 30-day compliance trial

Critical for Regulated Entities: Air-Gapped Deployment

For stock exchanges, depositories, clearing corporations, and custodians operating under SEBI's most stringent data classification requirements, DefenGPT's air-gapped on-premises deployment provides absolute data sovereignty. No prompt, no classified output, and no audit log ever traverse a network boundary outside the RE's own controlled infrastructure. This is the only deployment model that can categorically satisfy a SEBI examiner's question: Can you guarantee that AI-processed data has never left your approved perimeter?' DefenGPT can.

Out-of-the-Box. Immediate Value. No Professional Services Required.

DefenGPT ships with everything a SEBI-regulated entity needs to begin governing AI on day one:

- Pre-configured SEBI CSCRF policy templates — activate in one click
- Built-in SEBI Circular 10873/2026 compliance checklist with evidence mapping
- Automated M-SOC and cyber-suraksha.ai incident reporting connectors
- AI Asset Discovery: scans your environment and builds the AI inventory automatically
- Role-based dashboards for CISO, DPO, IT Committee, and Board
- Native integration with NSE/BSE M-SOC, Splunk, Sentinel, QRadar, and all major IAM platforms

S E C T I O N 0 7

DefenGPT vs. Conventional Cybersecurity — The Definitive Comparison

The question is sometimes asked: 'Can we configure our existing tools to meet the SEBI AI circular?' The answer is no — not because the tools are inadequate for their original purpose, but because the SEBI circular has created a new compliance domain that existing tools were never designed for. The table below is definitive.

SEBI Circular Requirement	NGFW	SIEM/SOC	DLP	CSPM	EDR	API GW	VAPT	IAM	DefenGPT
AI Vulnerability Risk Assessment (Req.2)	✗	✗	✗	✗	✗	✗	○	✗	✓ FULL
AI Prompt Injection Detection (Req.2/6)	✗	✗	✗	✗	✗	✗	✗	✗	✓ FULL
API Semantic Governance (Req.5)	✗	✗	✗	✗	✗	○	✗	✗	✓ FULL
AI-Specific SOC Alerting (Req.6)	✗	○	✗	✗	✗	✗	✗	✗	✓ FULL
AI Risk Assessment Evidence (Req.7)	✗	○	✗	○	✗	✗	✗	✗	✓ FULL
AI Asset Inventory / SBOM (Req.9)	✗	✗	✗	○	✗	✗	✗	✗	✓ FULL
AI Guardrails Policy Engine (Req.8)	○	✗	○	✗	✗	○	✗	○	✓ FULL
Immutable AI Audit Log (Req.4)	✗	○	○	✗	✗	✗	✗	✗	✓ FULL

SEBI Circular Requirement	NGFW	SIEM/SOC	DLP	CSPM	EDR	API GW	VAPT	IAM	DefenGPT
Air-Gapped Deployment (Req.8/CSCRF)	✓	○	○	✗	✓	○	○	✓	✓ FULL
SEBI CSCRF Compliance Reports (Req.7/10)	✗	○	✗	○	✗	✗	✗	✗	✓ FULL

Legend: ✓ Full coverage ○ Partial / requires significant customisation ✗ Not applicable — tool category cannot address this requirement

SECTION 08

Use Cases — DefenGPT in SEBI-Regulated Environments

Stock Exchange / Depository

Mythos-class tools are used by red teams to probe exchange APIs for weakness. DefenGPT intercepts all AI-generated traffic targeting exchange APIs, classifies the intent as reconnaissance, generates an M-SOC alert with full session replay, and blocks the request while logging evidence for SEBI audit. Deployment: air-gapped on-premises within exchange perimeter. Time to deploy: 28 minutes.

Mutual Fund / AMC

Fund managers use AI to draft investment memos and summarize research. Without governance, NAV data, unreleased fund performance, and portfolio holdings flow into public AI. DefenGPT classifies all such data as Restricted, blocks external transmission, routes interactions to a private LLM instance, and generates weekly SEBI CSCRF compliance evidence.

Stockbroker / Trading Firm

Dealers use AI-assisted tools for order management and client communication. An adversary embeds a prompt injection in a client email instructing the AI to forward account details. DefenGPT's intent engine detects the indirect injection, blocks the AI action, alerts the broker's SOC team, and files an automated incident report to cyber-suraksha.ai.

Investment Advisor / Research Analyst

A research analyst uses an AI tool to process client financial statements for portfolio analysis. DefenGPT detects PII, PAN data, and income tax records in the upload, applies a SEBI-aligned data handling policy (redact before processing), logs the interaction for SEBI audit, and routes the anonymised document to the approved AI model.

SECTION 09

Conclusion — Compliance Is Not Optional

SEBI Circular 10873/2026 is unambiguous. Regulated entities are required to govern the risks posed by advanced AI tools — including those used offensively by adversaries and those used internally by their own staff. The circular is issued under Section 11(1) of the SEBI Act and carries full enforcement weight. Non-compliance exposes regulated entities to regulatory action, market suspension, and reputational damage in the world's fastest-growing capital market.

The technology gap is equally unambiguous. Traditional cybersecurity tools — firewalls, SIEM, DLP, CSPM, EDR, API gateways, VAPT, and IAM platforms — protect important parts of the enterprise security architecture, but none of them can address the AI interaction layer that SEBI's circular is specifically concerned with. They do not inspect prompts. They do not classify AI-processed data. They do not detect prompt injection or AI-assisted reconnaissance. They do not maintain an AI asset inventory. They do not generate SEBI CSCRF AI compliance evidence.

DefenGPT AI Security Platform by Defenix Technology is the only solution that does all of this — and does it out of the box, deployable in under 30 minutes, in any environment from a public cloud SaaS instance to a completely air-gapped, network-isolated on-premises installation inside the most sensitive financial infrastructure in India.

Ready to achieve full SEBI Circular compliance?

Defenix Technology offers three immediate pathways for SEBI-regulated entities:

01 — SEBI Compliance Gap Assessment: a structured review of your current AI security posture against Circular 10873/2026 requirements, delivered within 5 business days

02 — Proof of Concept Deployment: a fully functional DefenGPT instance in your environment — on-prem, private cloud, or air-gapped — operational within 30 minutes

03 — Compliance Evidence Package: a ready-to-present documentation set for SEBI examiners and your IT committee, generated automatically from your DefenGPT deployment

enterprise@defenix.com | www.defenix.com | SEBI Compliance Helpline: +91-22-DEFENIX