



AI-Driven Cyber Attacks and India's Digital Security Challenge

Mr. Ravindra Nishad (Research Scholar) Department of Defence & Strategic Studies,

Prof. Rajendra Singh (Rajju Bhaiyya) University, Prayagraj

Dr. Divya Dwivedi (Assistant Professor) Department of Defence & Strategic Studies,

Prof. Rajendra Singh (Rajju Bhaiyya) University, Prayagraj

DOI : <https://doi.org/10.5281/zenodo.23036594>

ARTICLE DETAILS

Research Paper

Received: 14-07-2026

Accepted: 13-08-2026

Published: 25-09-2026

Keywords:

Artificial Intelligence; Cyber Attacks; India Cybersecurity; CERT-In; AIIMS Ransomware; Digital India; AI Malware; Deepfakes; Phishing; Critical Infrastructure Protection; NCIIPC; Defence Cyber Agency; Digital Personal Data Protection Act; UPI Security; Aadhaar; Cyber Threats India; National Cyber Policy India; AI-

ABSTRACT

India's rapid and ambitious digitalization spanning a national biometric identity system covering 1.3 billion people a unified payments platform processing over 12 billion transactions per month a growing e-governance ecosystem has created one of the world's largest and fastest-expanding digital attack surfaces. Against this backdrop, artificial intelligence has emerged as both a transformative tool for cybersecurity defence and the most dangerous new capability now available to malicious actors targeting Indian infrastructure, institutions citizens. This article examines the intersection of AI-driven cyber threats and India's specific digital security context. It analyses five primary AI-enabled attack categories AI-powered phishing, deepfake-based social engineering, adaptive malware, automated vulnerability exploitation AI-controlled botnets with reference to documented incidents affecting Indian targets including the 2022 AIIMS New Delhi ransomware attack, the 2020 intrusion attempts against India's power grid attributed to Chinese state-linked actors, the 2022 Air India data breach and the documented targeting of India's banking and financial sector by AI-assisted fraud operations. The article situates these threats within India's governance context, examining the institutional landscape CERT-In,

Disclaimer: The opinions, interpretations, conclusions, recommendations, analyses, and statements expressed in this research article are solely those of the author(s) and do not reflect the views, policies, or positions of the Publisher, Chief Editor, Editors, Editorial Board, Reviewers, or their affiliated institutions. The author(s) are solely responsible for ensuring the originality, authenticity, and integrity of the manuscript and for ensuring that it is free from plagiarism, AI-generated content, copyright infringement, data falsification, and any other form of academic misconduct.



Enabled Fraud; State-Sponsored Cyber Attacks; Cyber Resilience India

NCIIPC, the Defence Cyber Agency, MeitY and the regulatory framework, including the Digital Personal Data Protection Act 2023 and CERT-In's 2022 mandatory incident reporting directions. It argues that India's unique combination of scale, digital ambition, governance complexity and adversarial exposure makes the AI cyber threat both more acute and more structurally complex to address than in smaller more institutionally consolidated states. The article concludes with recommendations grounded in India's specific institutional and operational context.

Introduction

On the morning of 23 November 2022, staff at the All India Institute of Medical Sciences (AIIMS) in New Delhi arrived to find the hospital's computer systems completely non-functional. The National Informatics Centre's servers which managed patient records, diagnostic data, billing administrative systems across India's most prestigious public hospital, had been encrypted by ransomware. For approximately fifteen days, AIIMS staff conducted operations using pen and paper manually recording patient data, manually tracking test results, manually coordinating care for a hospital that manages over 3,500 beds and tens of thousands of outpatient consultations daily. The attack disrupted care for some of India's most critically ill patients and exposed sensitive health records for potentially millions of individuals. The AIIMS attack was not exceptional. It was representative. Over the preceding two years, India had experienced an unprecedented surge in cyber incidents: ransomware attacks on state government systems in Maharashtra and Odisha, sustained intrusion campaigns against India's banking sector, a data breach at Air India exposing the personal and passport information of 4.5 million passengers most gravely documented attempts by Chinese state-linked threat actors to access the operational technology networks controlling India's power grid in Mumbai and other cities. In October 2020 a significant power outage in Mumbai coincided with what Recorded Future subsequently identified as an extensive Chinese intrusion campaign targeting Indian power infrastructure. These incidents occurred against a backdrop of extraordinary digital expansion. India's Digital India initiative has, over the past decade, created one of the most ambitious national digital ecosystems in the world. The Aadhaar biometric identity system now covers over 1.3 billion Indians. The Unified Payments Interface processes more digital transactions per month than any comparable system globally. The CoWIN platform managed COVID-19 vaccine registration for hundreds of millions of people. Government services from tax filing to land records to public distribution have moved online, bringing



genuine benefits in access, efficiency transparency and simultaneously creating a digital attack surface of extraordinary scale and complexity.

Into this environment, artificial intelligence has arrived as a threat multiplier. AI does not create new categories of cyber-attack so much as it removes the constraints that previously limited how sophisticated, personalized, scalable adaptive those attacks could be. For India specifically, this matters in ways that differ from smaller, more institutionally consolidated states. India's size means its attack surface is enormous. Its governance complexity multiple ministries, agencies regulators with overlapping cybersecurity mandates means coordination during incidents is harder. Its uneven digital literacy across a population of 1.4 billion means the human targets of AI-enhanced social engineering attacks include many who have limited capacity to recognize manipulation and its adversarial exposure facing sophisticated state-sponsored cyber operations from both China and Pakistan, as well as a highly active criminal ecosystem means the threat environment is among the most complex faced by any country. This article examines what AI-driven cyber threats mean for India specifically. Section 2 defines AI-driven cyber-attacks and the threat they represent. Section 3 analyses the five primary AI-enabled attack categories through the lens of India's documented experience. Section 4 explains why these threats are qualitatively different from conventional attacks why India's specific context amplifies certain risks. Section 5 surveys documented incidents affecting Indian targets. Section 6 examines India's institutional and regulatory framework for cyber defence. Section 7 analyses the defensive application of AI within India's governance context. Section 8 considers the trajectory of the threat over the coming decade. Section 9 presents recommendations grounded in India's specific operational and institutional context. The analysis draws on CERT-In annual reports, published research from the Observer Research Foundation, Carnegie Endowment Recorded Future, India's legislative framework including the Digital Personal Data Protection Act 2023 and the Information Technology Act 2000, government policy documents including the draft National Cybersecurity Strategy international cybersecurity research relevant to India's threat context. What Are AI-Driven Cyber Attacks?

AI-driven cyber-attacks are malicious operations in which attackers deploy machine learning models, generative AI systems, deep learning architectures AI-enabled automation to enhance the effectiveness, speed, scale adaptability of their operations. The term encompasses a wide range of techniques unified by a common characteristic: the use of artificial intelligence to remove human effort as the binding constraint on attack quality and scale. To understand what this means in India's context, consider the contrast between a conventional phishing attack targeting an Indian banking customer and an AI-enabled version. The conventional attack requires someone to write a generic message in Hindi or English, purchase or rent a list of targets, send it and hope that enough recipients are deceived. An AI-enabled version can ingest a target's



transaction history (if available through prior breach data) their publicly visible social media in their preferred regional language their likely banking relationships based on their location and demographic profile generate a highly personalized message in fluent Hindi, Bengali, Tamil or whichever language the target uses referencing plausible specifics about their account their recent transactions a government scheme they are likely enrolled in. India's linguistic diversity far from being a protective barrier becomes an attack surface that AI can exploit at scale.

Conventional Cyber Attacks	AI-Driven Cyber Attacks
Generic content, easy to spot	Personalized, context-aware, linguistically adapted
Requires skilled operators	Accessible to low-skill actors via AI tools
Slow limited by human bandwidth	Scales to millions of simultaneous targets
Static fixed attack pattern	Adaptive learns from defences in real time
One language typically	Multi-lingual, targeting India's 22 official languages
Known signatures detectable	Polymorphic evades pattern-based detection

Three structural characteristics of AI-driven attacks are particularly relevant to India's situation. First, the collapse of the skill barrier: tools specifically designed for offensive use Worm GPT, Fraud GPT and similar platforms documented in criminal markets in 2023 mean that sophisticated attack capabilities are no longer limited to technically expert actors. India faces a large and active domestic cybercrime ecosystem alongside state-sponsored threats; AI tools that lower the entry barrier for all of these actors represent a significant threat amplification. Second, adaptability: AI attack systems can observe defensive responses and modify their behaviour accordingly. For an Indian context in which CERT-In's incident reporting mandate and enforcement capacity are both relatively recent developments, this means attackers can learn what triggers detection faster than India's institutional response mechanisms can adapt. Third scale relative to institutional capacity: India's cybersecurity institutional infrastructure capable as it is becoming is managing a very large and rapidly growing digital ecosystem with resources that are not yet commensurate with that scale. AI-enabled attacks that generate vastly more incidents than conventional attacks will strain those resources further.



AI-Powered Phishing: Targeting India's Digital Citizens

Phishing deceptive communications designed to extract credentials, authorize fraudulent transactions enable system access is the most prevalent initial attack vector in India's cyber threat landscape. CERT-In's 2022 annual report identified phishing as the leading attack category by volume, accounting for a significant proportion of the 1.4 million incidents reported that year. That volume kept climbing sharply in subsequent years: CERT-In recorded roughly 2.04 million incidents in 2024 and 2.94 million in 2025, an increase of about 85 per cent over 2023. India's combination of a large and rapidly growing online population, expanding digital financial services and variable digital literacy creates conditions particularly favorable for phishing operations. AI has dramatically increased the sophistication and success rate of phishing targeting Indian users. Language, which was previously a partial barrier poorly translated Hindi or regional language phishing was relatively easy to identify has ceased to be protective. Large language models can generate grammatically correct, tonally appropriate messages in Hindi, Bengali, Tamil, Telugu, Marathi and other major Indian languages, drawing on scraped data to personalize content to individual targets.

The specific vulnerabilities in India's context are significant. UPI-linked phishing exploits the widespread use of the Unified Payments Interface by generating fake payment requests, fake bank alerts fraudulent merchant links that are visually and linguistically indistinguishable from genuine communications. Aadhaar-themed phishing messages claiming action is required to "update" or "verify" Aadhaar details exploits the universal reach of the biometric identity system and the genuine anxiety many citizens feel about maintaining their digital identity documents. Government scheme impersonation messages claiming to offer PM Kisan, PM Awas Yojana other benefit scheme payments target a population that actively engages with government digital platforms and has strong incentive to respond to such communications. The institutional response has been active but faces structural limitations. CERT-In's April 2022 mandatory reporting directions requiring organizations to report incidents within six hours were designed to improve visibility into the threat landscape. However, the six-hour window, while creating better data, does not by itself prevent attacks. And the volume of phishing incidents targeting individual citizens which typically fall below the enterprise reporting threshold remains largely invisible to formal institutional monitoring.

Deepfake Fraud: A Growing Threat to Indian Finance and Governance

AI-generated synthetic media voice cloning, video deepfakes, realistic document fabrication has emerged as a documented threat in the Indian context, particularly in financial fraud and political disinformation. The technology that enables deep-fake generation is commercially available, computationally accessible and increasingly capable of producing output that defeats human perception. In the Indian financial sector, voice



cloning fraud has been documented in multiple cases involving impersonation of bank officials, company executives and government representatives. Attackers clone a target's voice from publicly available audio social media videos, YouTube interviews, WhatsApp voice messages and use the synthetic voice to instruct employees or family members to authorize transactions. Several such cases have been documented in Indian courts and media in 2023 and 2024, though systematic data on total losses remains incomplete due to underreporting. The political dimension is equally concerning. India conducts the world's largest democratic elections, involving over 900 million eligible voters, extensive use of digital platforms for campaign communication and a multilingual media environment. AI-generated deepfake videos of political figures have already been documented in India's electoral context. In the run-up to the 2024 Lok Sabha elections, the Election Commission of India issued specific advisories about AI-generated content and deepfakes and several state-level electoral authorities documented complaints about fabricated video content attributed to political candidates. The deepfake challenge is particularly acute in India because the country's linguistic and cultural diversity means detection tools trained primarily on English-language content may perform poorly on content in Indian languages and regional dialects. Building robust detection capability that works across Hindi, regional languages and the diverse acoustic profiles of Indian speech patterns represents a significant technical challenge that has not yet been adequately addressed. The scale of the problem became clearer through 2025: independent surveys found that nearly half of Indian adults had been a victim of, or knew someone who had been a victim of, an AI voice-cloning or deepfake scam, roughly double the global average, and reported voice-cloning fraud complaints rose sharply through late 2025 and into 2026.

Adaptive Malware and Ransomware: Hitting India's Critical Institutions

The AIIMS New Delhi ransomware attack of November 2022 is the most publicly visible instance of a documented pattern: sophisticated malware attacks against India's critical institutions including hospitals government agencies public sector undertakings financial infrastructure. The attack encrypted five servers managing patient data for potentially 30 million patients and disrupted hospital operations for approximately fifteen days. The National Investigation Agency and CERT-In both investigated the incident; attribution remained officially inconclusive, though multiple cybersecurity researchers identified indicators consistent with Chinese or Chinese-linked threat actors. AI has changed what ransomware and adaptive malware can do in two important ways relevant to India's context. First, AI-enabled reconnaissance allows attackers to map target environments more thoroughly before deploying payloads identifying the most critical systems understanding backup architectures timing attacks for maximum impact. In the AIIMS case attack was effective precisely because it targeted the systems most central to hospital operations, suggesting significant prior reconnaissance. Second AI-enabled evasion means that malware can identify and avoid triggering



detection in environments with specific security tools deployed making it harder to detect before it deploys its payload. Beyond AIIMS documented ransomware and malware attacks against Indian targets in 2022–2024 include an attack on the Oil India Limited operational technology network in Assam (April 2022) multiple attacks on Maharashtra and other state government administrative systems an intrusion at Indian Council of Medical Research (ICMR) in late 2022 the compromise of systems at SpiceJet airlines. Each of these incidents followed a broadly similar pattern: extended dwell time during reconnaissance followed by deployment of a payload specifically configured for the target environment, a pattern consistent with AI-assisted operational planning.

Automated Vulnerability Exploitation: The Speed Problem for Indian IT

India's IT sector is one of its most significant strategic assets the country is home to the world's largest concentration of software services firms, managing critical applications and infrastructure for clients across the globe. This creates a distinctive vulnerability: an attack that successfully exploits a vulnerability in a major Indian IT services provider has the potential to cascade into the systems of that provider's clients across multiple countries and sectors. AI-assisted vulnerability discovery and exploitation compresses the window between vulnerability disclosure and active exploitation. For Indian organizations many of which operate on patch deployment cycles significantly longer than the global best-practice standard of 72 hours for critical vulnerabilities this compression is a direct threat. A vulnerability disclosed on a Monday that would previously have taken weeks for attackers to weaponize can now be exploited within hours.

The supply chain dimension is particularly significant for India. Indian IT companies Infosys, TCS, Wipro, HCL Technologies hundreds of smaller firms manage software development, IT operations digital infrastructure for clients including major financial institutions, healthcare providers government agencies in multiple countries. A successful supply chain compromise of a major Indian IT provider would have implications far beyond India's borders, making Indian IT infrastructure a high-value target for sophisticated adversaries.

AI-Controlled Botnets: Targeting India's Financial Infrastructure

India's rapid adoption of digital financial services has created a large, accessible high-value target for botnet-enabled attacks. The UPI ecosystem, with its hundreds of millions of active users and billions of monthly transactions, is a natural target for credential stuffing automated testing of username/password combinations from breach datasets to identify accounts that can be accessed and drained. AI-controlled botnets that can mimic legitimate human transaction patterns varying timing, transaction amounts



geographic patterns to avoid velocity-based fraud detection are documented tools of financial cybercrime operating against Indian banking targets. The RBI's payment fraud data shows consistent growth in digital payment fraud volumes even as security measures have improved, a pattern consistent with adversaries deploying adaptive, AI-managed attack infrastructure that adjusts to defensive countermeasures. DDoS attacks using AI-managed botnets have also been directed at Indian government infrastructure, particularly during periods of geopolitical tension. Following the 2020 Galwan Valley clash between Indian and Chinese forces, a significant surge in DDoS attacks against Indian government websites was documented by CERT-In and multiple private security firms. AI-managed botnets conducting these attacks demonstrated traffic patterns specifically designed to defeat India's deployed DDoS mitigation infrastructure .

Why AI Attacks Are Qualitatively Different India's Specific context

The qualitative shift that AI represents in the cyber threat globally. But several features of India's specific context certain aspects of that shift in ways that deserve explicit recognition. With 700 million internet users and growing, a unified national digital identity system, dominant digital payments platform an expanding e-governance ecosystem, India's digital attack surface is among the largest in the world. AI attack tools that can scale to millions of simultaneous targets far beyond what human-led operations could manage are therefore particularly consequential in India's context. The sheer volume of potential targets means that even attack campaigns with low per-target success rates can yield enormous absolute numbers of compromised accounts, fraudulent transactions breached datasets.

India's 22 officially recognized languages and hundreds of regional dialects have historically provided a partial barrier against phishing and social engineering poorly localized attacks were relatively recognizable. AI-powered language generation eliminates this barrier. Systems trained on multilingual datasets can generate grammatically correct, culturally appropriate phishing content in Hindi, Bengali, Tamil, Telugu, Marathi, Gujarati any other major Indian language. This linguistic capability has not yet been widely deployed against Indian targets at scale but it represents a near-term capability that will be for which Indian awareness training and detection systems are not currently prepared.

India's cybersecurity institutional architecture distributes responsibility across multiple agencies: CERT-In under MeitY for civilian incident response, NCIIPC under NTRO for critical infrastructure, the Defence Cyber Agency for military cyber operations, the RBI for financial sector cyber risk, TRAI for telecommunications, SEBI for capital markets state-level agencies for their respective domains. In routine operations, this distribution functions adequately. In the face of AI-driven attacks that operate faster than



human response cycles completing an attack chain within minutes the coordination costs imposed by this distributed architecture can mean the difference between containment and catastrophe.

India's digital expansion has brought hundreds of millions of first-generation internet users online. Many are accessing digital financial services, government platforms and social media for the first time, with limited prior experience of the deceptive techniques used in social engineering. AI-powered phishing and deepfake attacks that are difficult to detect even for experienced users are significantly more effective against a population that includes many who have never encountered such manipulation before. This is not a criticism of India's digital citizens it is a structural reality of rapid, large-scale digitalization that policy must account for.

India faces documented, sustained cyber operations from sophisticated state-linked adversaries primarily China and Pakistan. Recorded Future's analysis of Red Echo (attributed to China) documented extensive targeting of India's power grid, port infrastructure financial sector. Pakistani state-linked threat groups have consistently targeted Indian defence, government strategic sectors. These adversaries have both the resources to deploy AI-enhanced attack capabilities and strong strategic incentives to do so. India is not a peripheral target in the global AI cyber threat landscape it is a primary one.

Documented Incidents: India's Cyber Threat Record

India's most consequential documented cyber-attack on critical healthcare infrastructure. Ransomware encrypted five servers managing patient records for an estimated 30 million patients at India's premier public hospital. The attack disrupted clinical operations for approximately fifteen days, forcing reversion to manual paper-based processes exposed sensitive health data. CERT-In and the NIA both investigated; attribution was not officially confirmed. The incident prompted significant debate about cybersecurity preparedness in India's public health infrastructure and led to CERT-In issuing specific advisories for healthcare sector organizations .

In October 2020 a significant power outage in Mumbai disrupted the Bombay Stock Exchange and other critical systems. Recorded Future subsequently published analysis identifying extensive intrusion activity by a Chinese state-linked threat group (designated RedEcho) against India's power grid infrastructure, including load dispatch centers in multiple states. The analysis identified Trojan zed applications and malware specifically designed to target industrial control systems. The incident raised serious questions about China's pre-positioning of cyber capabilities in India's critical infrastructure capabilities that could be activated during a geopolitical crisis .



A breach of Air India's passenger services system vendor (SITA) exposed the personal data of 4.5 million passengers globally, including a significant proportion of Indian citizens. The compromised data included names, dates of birth, contact information, ticket details passport and credit card information. The breach demonstrated the supply chain vulnerability of Indian aviation infrastructure and the risks posed by third-party data processors operating with inadequate security standards .

Ransomware deployed against operational technology networks at Oil India Limited's facility in Duliajan, Assam, demanded a ransom of approximately US\$7.5 million. The attack targeted industrial control systems rather than IT infrastructure a significant escalation in attack sophistication consistent with AI-assisted reconnaissance and payload configuration. Oil India's operational systems were disrupted for several days the incident highlighted the vulnerability of India's energy sector operational technology to sophisticated cyber-attacks .

Indian cybersecurity researchers and the Indian Cyber Crime Coordination Centre (I4C) under the Ministry of Home Affairs documented the penetration of AI-enabled fraud tools including Worm GPT-style platforms into Indian criminal networks in 2023. These tools were identified as being used to generate Hindi-language phishing content, automate UPI fraud schemes and produce fraudulent KYC documentation. The commodification of AI attack tools in India's domestic cybercrime ecosystem represents a significant acceleration of the threat to individual Indian citizens .

A breach of the Indian Council of Medical Research's COVID-19 testing database was reported in November 2022, with data on approximately 81.5 million Indian citizens allegedly offered for sale on a dark web forum. The compromised data included names, ages, addresses Aadhaar numbers. If confirmed, it would represent one of the largest personal data breaches in Indian history and illustrated the risk posed by large, centralized health databases with inadequate security controls .

India's Cybersecurity Institutional and Regulatory Framework

India's institutional architecture for cybersecurity has developed substantially over the past decade but it continues to face significant challenges of coordination, capacity and statutory clarity that directly affect the country's ability to respond to AI-driven threats.

The Indian Computer Emergency Response Team (CERT-In), operating under the Ministry of Electronics and Information Technology (MeitY), functions as India's primary civilian cybersecurity response agency. CERT-In issues advisories, coordinates incident response, publishes threat intelligence and since April 2022 has operated a mandatory six-hour incident reporting requirement applicable to a wide range of



organizations. The six-hour reporting window, one of the shortest mandatory reporting timelines anywhere in the world has generated both improved incident visibility and significant compliance concerns from the private sector regarding operational feasibility.

The National Critical Information Infrastructure Protection Centre (NCIIPC), operating under the National Technical Research Organization (NTRO) is responsible for the protection of India's critical information infrastructure power, banking, finance, telecommunications, transport government. NCIIPC issues sector-specific guidelines, conducts vulnerability assessments and coordinates protective measures for designated critical sectors. Its mandate covers exactly the sectors most targeted by advanced AI-driven attacks, yet its resource base and institutional visibility remain limited relative to the scale of the challenge.

Established in 2019, the Defence Cyber Agency (DCA) operates under the Integrated Defence Staff and is responsible for military cyber operations. The DCA represents India's most explicit acknowledgement of offensive and defensive cyber capability in the military domain, though its mandate, capabilities and relationship with civilian cyber agencies remain largely opaque in public documentation. The separation between the DCA and civilian agencies CERT-In and NCIIPC means that India's military and civilian cyber responses operate on largely separate tracks, a structural feature that can complicate integrated responses to attacks that span both domains.

India's Digital Personal Data Protection (DPDP) Act, passed in August 2023, establishes a comprehensive framework for the protection of personal data a foundational element of India's digital security architecture. The Act establishes the Data Protection Board of India, creates obligations for data fiduciaries (organizations collecting personal data) sets out citizens' rights regarding their personal information. From a cybersecurity perspective, the Act's requirements for data breach notification and data minimization are relevant to reducing the consequences of successful attacks. However, implementation rules have not yet been fully notified the Act's interaction with CERT-In's existing reporting requirements creates some regulatory ambiguity. Much of that ambiguity was addressed when the Digital Personal Data Protection Rules, 2025 were notified in November 2025, operationalizing the Act with a phased implementation timetable, a two-stage breach-notification requirement and the formal constitution of the Data Protection Board of India.

The Gap: No Unified Statutory Framework

India's cybersecurity governance remains distributed across the Information Technology Act 2000 (and its amendments), CERT-In's directions, the DPDP Act 2023, sector-specific regulations from RBI, SEBI,



IRDAI TRAI the National Cyber Security Policy 2013 which has not been formally updated despite a draft National Cybersecurity Strategy having been circulated in 2020. This fragmentation creates real challenges: during a major incident, jurisdictional questions can slow response, accountability is diffuse long-term strategic planning is complicated by the absence of a comprehensive statutory framework that consolidates and clarifies responsibilities.

How India Is Fighting Back and Where the Gaps Remain

The same AI capabilities that enable more sophisticated attacks are being applied to defensive purposes in India, though deployment is uneven across sectors and the institutional infrastructure for AI-based defence lags behind the threat.

CERT-In has expanded its threat intelligence capabilities significantly since 2020, issuing more frequent advisories, establishing sectoral Computer Security Incident Response Teams (CSIRTs) developing partnerships with international CERTs and private sector threat intelligence providers. The mandatory reporting requirement, despite its implementation challenges, has improved CERT-In's visibility into the incident landscape and enabled more systematic pattern identification as a prerequisite for AI-assisted threat intelligence analysis.

The Reserve Bank of India has been the most proactive Indian regulatory authority in deploying AI for defensive purposes. RBI's fraud detection frameworks for UPI and card transactions increasingly incorporate machine learning models for real-time anomaly detection identifying transactions that deviate from a user's established behavioral patterns and flagging them for additional verification. NPCI, the operator of the UPI system, has deployed AI-based fraud scoring across the transaction processing pipeline. These measures have had measurable impact on fraud prevention, though the volume and sophistication of AI-enabled fraud attacks continues to grow faster than detection capabilities.

Major Indian IT firms TCS, Infosys, Wipro, HCL have deployed AI-based security operations centre capabilities for their enterprise clients, incorporating behavioral detection, automated threat intelligence AI-assisted incident response. Indian cybersecurity firms including Sequaretek, TAC Security Quick Heal have developed AI-enhanced products specifically adapted to the Indian threat landscape and Indian regulatory requirements. This private sector capability, however, is concentrated in larger enterprises; India's enormous base of small and medium enterprises many of which have limited IT security investment remains largely unprotected by AI-based defensive capabilities.

The persistent gaps



Several structural gaps in India's defensive AI capability deserve explicit identification. First the public–private trust deficit: effective AI-based defence requires continuous, real-time sharing of threat intelligence between government agencies and private sector organizations. India's institutional framework for such sharing modelled on information sharing mechanisms in the US and UK remains underdeveloped, with mandatory reporting substituting for the voluntary, relationship-based intelligence sharing that enables genuinely effective collective defence. Second the skills gap: India faces a documented shortage of cybersecurity professionals with AI security expertise. NASSCOM has estimated a shortfall of approximately 1 million cybersecurity professionals in India by 2025. Third the SME gap: the vast majority of India's economic activity occurs in small and medium enterprises that have neither the resources nor the expertise to deploy AI-based defences or even to implement basic cyber hygiene measures consistently.

The Decade Ahead: India's AI Cyber Threat Trajectory

India's cyber threat environment over the next decade will be shaped by three intersecting trends: the continued expansion of its digital attack surface, the increasing sophistication of AI-enabled attacks directed at Indian targets the gradual maturation of India's institutional and technical defenses.

Expansion of the attack surface

India's Digital India initiative continues to expand. The Ayushman Bharat Digital Mission is building a national health records system that will eventually cover all Indians. The National Digital Education Architecture is digitizing educational records and assessments. The expansion of smart city infrastructure is connecting physical traffic management, utilities and public safety to networked control systems. Each expansion creates new attack surfaces. AI-enabled attacks that can scan and map these surfaces autonomously will identify exploitable vulnerabilities faster than manual security assessments can keep up.

State-sponsored AI cyber operations against India

China's demonstrated capability and willingness to conduct sustained cyber operations against Indian infrastructure documented across the power sector, government systems research institutions represent the most serious near-term AI cyber threat to India. As Chinese state-linked threat actors integrate AI capabilities into their operational toolset, the sophistication and efficiency of these operations will increase. The 2020 power grid intrusion demonstrated China's interest in pre-positioning capabilities in Indian critical infrastructure; AI will make such pre-positioning easier to achieve and harder to detect. Pakistan-linked threat actors, while generally less sophisticated, are also adapting AI tools for operations against Indian targets.



AI-enabled election interference

India's 2024 general elections demonstrated the growing relevance of AI-generated content in the electoral context. As AI-generated audio and video quality continues to improve and the cost of production continues to fall, the risk of large-scale AI-enabled electoral disinformation in India's 2029 elections and in state assembly elections in the intervening period is significant. India's Election Commission has begun addressing this challenge, but the technical and regulatory frameworks for doing so at scale remain underdeveloped.

The regulatory trajectory

India's regulatory response to AI-enabled cyber threats is accelerating. The DPDP Act 2023 establishes a foundational data protection framework. The forthcoming National Cybersecurity Strategy, if formally adopted, will provide updated strategic guidance. CERT-In has indicated its intention to develop AI-specific guidelines. The challenge will be to ensure that regulatory development keeps pace with threat evolution a challenge made harder by the speed of AI capability development and the global nature of the threat environment.

Recommendations for India's Cybersecurity Ecosystem

The following recommendations are grounded in India's specific institutional context, threat environment operational constraints. They reflect the consensus of Indian and international cybersecurity research on the measures most likely to improve India's resilience against AI-driven threats.

For Government and Regulators

- Enact a comprehensive National Cybersecurity Act that consolidates and clarifies the mandates of CERT-In, NCIIPC, the DCA sector regulators, providing a unified statutory framework with clear authority, accountability escalation protocols for major incidents.
- Formally adopt and resource the National Cybersecurity Strategy. The 2020 draft represents a credible framework; its continued non-adoption leaves India without an authoritative strategic document against which policy can be evaluated and resources allocated.
- Develop AI-specific cybersecurity guidelines for critical sectors. CERT-In and NCIIPC should issue sector-specific guidance on AI threat detection, AI system security AI-enabled incident response in consultation with private sector stakeholders.



- Build real-time public-private threat intelligence sharing mechanisms. Mandatory reporting provides incident data retrospectively. India needs mechanisms modelled on ISACs and the UK's NCSC approach for sharing actionable threat intelligence in real time, with appropriate liability protections to incentivize private sector participation.
- Fund dedicated AI cybersecurity research capacity. DRDO, IITs IISc should receive dedicated funding for AI cybersecurity research both offensive threat research to understand adversary capabilities and defensive AI development relevant to India's specific threat context.

For Critical Infrastructure Operators

- Compress patch deployment timelines for critical vulnerabilities. India's average 30-day patch lag for critical vulnerabilities relative to the hours-to-days exploitation window AI attack tools create is one of the most significant operational vulnerabilities in India's critical infrastructure. Closing this gap should be a priority investment.
- Implement secondary verification protocols for all high-value transactions and access decisions. The deepfake fraud and voice cloning cases documented in India are preventable through procedural controls that do not require sophisticated technology simply requiring confirmation through an independent, pre-established channel for any significant financial or access decision.
- Conduct AI-specific red team exercises. Conventional penetration testing simulates historical attack patterns. Organizations should commission exercises that specifically deploy AI-generated phishing, simulated deepfake scenarios AI-assisted intrusion techniques against their actual defensive infrastructure.

For India's Digital Citizens

- Verify unexpected requests through independent channels before acting. Whether the request comes by phone, video call, email WhatsApp if it involves money, credentials sensitive personal information, call back on a number you already have, not one provided in the message.
- Be aware that AI can mimic voices and faces convincingly. If a family member or colleague makes an unusual or urgent financial request by phone or video, verify through a different channel before complying. The technology to fake these interactions is now commercially available.



- Use strong, unique passwords and enable two-factor authentication for all financial and government digital services. AI-assisted password cracking makes password reuse significantly more dangerous than it was even three years ago.

Conclusion

The AIIMS ransomware attack of November 2022 lasted fifteen days. For a quarter of those days, staff at India's most prestigious public hospital conducted clinical operations on paper. Patients with cancer, cardiac conditions and other critical illnesses experienced delayed diagnoses, disrupted treatment compromised care. No lives were officially attributed to the attack but the conditions for such a consequence were present. This is what AI-driven cyber-attacks mean for India. They are not abstract threats or distant hypotheticals. They are operational realities that have already disrupted hospitals, threatened power grids, compromised millions of citizens' personal data tested the resilience of financial infrastructure serving hundreds of millions of people. India's exposure to these threats is heightened by specific features of its situation: the extraordinary scale and ambition of its digital transformation, the governance complexity of managing cybersecurity across a federal democracy of 1.4 billion people, the uneven digital literacy of a population that includes hundreds of millions of first-generation internet users the sustained, sophisticated cyber operations directed at India by state-linked adversaries with both the capability and the motive to deploy AI-enhanced tools. None of this means India is losing the cybersecurity contest. CERT-In has become substantially more active and capable. The DPDP Act provides a foundation for data governance that can reduce breach consequences. RBI's AI-based fraud detection has demonstrated measurable effectiveness. The private sector cybersecurity ecosystem is growing. India has real strengths to build but the pace of AI capability development on the offensive side is accelerating faster than defensive adaptations. The structural gaps in institutional coordination, in public-private trust in cybersecurity workforce depth, in SME protection are real and consequential. Closing them will require not just technical investment but institutional reform, regulatory clarity a genuine national commitment to treating cybersecurity as a strategic priority commensurate with the scale of India's digital ambitions. The trajectory through 2025 and into 2026 shows both halves of this picture at once: recorded incident volumes and deepfake-enabled fraud continued to rise sharply even as the DPDP Rules, 2025 and an expanded CERT-In advisory programmer extended the institutional response.

India is building a digital future at a speed and scale that has no historical precedent. The security of that future the resilience of the systems that millions of Indians are coming to depend on for health, finance, identity governance depends on whether India can match that ambition with equally serious investment in



defending what it is building. The AI-driven threat makes that investment more urgent and not less achievable. The capabilities and the policy frameworks to address it exist. What is required is the political will and institutional discipline to deploy them.

References:

¹ CERT-In. (2022). *Annual Report 2021–2022*. Government of India, Ministry of Electronics and Information Technology; PTI. (2022, November 23). AIIMS Delhi server down due to cyberattack, manual work resumes. *Press Trust of India*.

¹ Recorded Future. (2021, February). China-Linked Group RedEcho Targets the Indian Power Sector Amid Heightened Border Tensions. *Insikt Group, Recorded Future*.

¹ Figures for 2023–2025 are drawn from CERT-In data disclosed to the Lok Sabha on 12 August 2026 and reported in ‘India Cyber Incidents Hit 29 Lakh in 2025, Up 85% from 2023,’ *TechObserver*, 2026; and Press Information Bureau. (2026, January 23). *CERT-In: India’s Frontline Defender against Cyber Threats*. Government of India.

¹ Election Commission of India. (2024). *Advisory Use of Artificial Intelligence and Deepfakes in Election Campaigns*. ECI.

¹ Observer Research Foundation. (2026, January 15). *Deepfakes and Financial Cybercrime: India’s Multi-Layered Response*. ORF Expert Speak.

¹ CERT-In. (2022). *Annual Report 2021–2022*. Government of India, Ministry of Electronics and Information Technology; India Today. (2022, November 25). AIIMS cyberattack: Chinese hackers suspected, NIA probe underway. *India Today*.

¹ CERT-In. (2020). *Annual Report 2019–2020*. Government of India, Ministry of Electronics and Information Technology.

¹ CERT-In. (2022). *Annual Report 2021–2022*. Government of India, Ministry of Electronics and Information Technology.

¹ Recorded Future. (2021, February). China-Linked Group RedEcho Targets the Indian Power Sector Amid Heightened Border Tensions. *Insikt Group, Recorded Future*.



¹ Air India. (2021). *Air India Data Security Incident: Passenger Data Breach Notification*. Air India Limited.

¹ PTI. (2022, April 13). Oil India Limited cyber attack: Operations disrupted at Duliajan facility. *Press Trust of India*.

¹ Indian Cyber Crime Coordination Centre (I4C). (2023). *Annual Cybercrime Report 2022–23*. Ministry of Home Affairs, Government of India.

¹ The Hindu. (2022, November 30). ICMR data breach: 81 crore Indians' data allegedly leaked on dark web. *The Hindu*.

¹ Ministry of Electronics and Information Technology. (2025, November 14). *Digital Personal Data Protection Rules, 2025*. Gazette Notification G.S.R. 846(E), Government of India.

¹ See CERT-In (2026) and Observer Research Foundation (2026), above, for the underlying 2025–2026 incident and fraud data, and Ministry of Electronics and Information Technology (2025) for the DPDP Rules.