IS AI AN ALLY OR ADVERSARY

A COMPREHENSIVE REPORT ON AI BASED SECURITY THREATS

2026

Cogenz Cybertech

Cyber Resilience · Business Excellence

EXECUTIVE SUMMARY

Artificial Intelligence (AI) has emerged as both a critical enabler and a growing threat in India's cybersecurity landscape. While AI-driven tools strengthen detection and response, they also empower attackers with scalable phishing, deepfakes, and social engineering campaigns. India is already the second-most targeted country globally, with deepfake-related crimes surging 4x in the last year alone.

Vulnerable groups particularly senior citizens and first-time digital users face disproportionate risks as cybercriminals exploit AI to create hyper-personalized scams. Existing laws and awareness mechanisms have not kept pace with this rapidly evolving threat environment.

This report examines the dual-nature of AI in cybersecurity, with a specific lens on India's vulnerable segments of the society, attack trends, and legal-preparedness gaps. It also outlines key recommendations to future-proof the nation's cybersecurity architecture. With perspectives from leading experts and practitioners from the government and industry, the report analyses the imperatives of technology and policy with a focus on a key disadvantaged and often vulnerable segments of the society.

The report also highlights India's preparedness gaps, and the urgent need for coordinated action. It recommends:

- Policy reforms under the Digital India Act with AI-specific safeguards.
- Citizen-centric programs tailored for senior citizens and rural populations.
- AI-enabled technology infrastructure in central and state cyber cells to combat this menace.
- Public–private collaboration to build indigenous and innovative defence technologies and tools.

For India to secure its digital future, a phased roadmap combining immediate safeguards, medium-term systemic strengthening, and long-term leadership is essential. Investing now in AI-driven defences will yield 8-10x returns by preventing massive economic and social losses.

FOREWORD

Dr. Shalini Rajneesh. IAS
Chief Secretary,Government of Karnataka.

Artificial Intelligence (AI) is revolutionizing industries across the globe and cybercrime is no exception. In India, where over a billion mobile users are expected to be digitally connected by the end of 2026, the emergence of AI-enabled cyber threats has created a formidable new challenge. From voice cloning and deepfake scams to intelligent phishing engines, cybercriminals are now leveraging AI to automate, personalize, and scale their attacks with alarming precision.

India's rapid digital adoption, especially with the use of Aadhaar, UPI in payments, Digital Health Mission, and the increasing use of e-governance in most of citizen services delivery, has brought both innovation and risk in equal measure. Traditional cybercrimes such as phishing and identity fraud have now evolved into AI-generated scam messages in regional languages, deepfake extortion videos, and voice-cloned emergency calls. The convergence of cheap compute power, public generative AI tools, and rising data digitization is fuelling a new wave of cybercrime faster, adaptive, and harder to trace.

In an increasingly digitized India, where even pension disbursement and hospital appointments are just a click away, a new demographic has entered the cybercrime crosshairs: Senior Citizens. With more than 138 million Indians aged 60 and above, this group is experiencing a surge in AI-enabled scams emotionally manipulative, multilingual, and alarmingly convincing. While AI is touted as a technological marvel, its darker side is being weaponized to target one of the most vulnerable sections of society. This report explores both sides of AI in cybersecurity how it is being misused, how it can be harnessed for defence, and how organizations and governments can steer it toward resilience through smart policies and sectoral strategies.

The Prime Minister recently in a public message, cautioned against 'digital arrests', and urged senior citizens to remain vigilant against fraudulent calls. It is crucial for all citizens to be aware of their fundamental rights especially digital. The recently unveiled Karnataka Cybersecurity Policy 2024 also identifies the senior citizens as a key disadvantaged segment that needs protectional intervention. While State-level enforcement aided with institutions like CERT-In and I4C are actively working to enhance AI-specific skills and resources with tools to detect AI threats, the race to bridge the ever-widening gap needs specific focus for commitment of resources and policy level interventions.

Karnataka and Bengaluru, being globally recognised for its IT industry and Startup ecosystem, is at the forefront of this technological revolution with not just innovative ideas by the private sector but also backed ably by the state government with many initiatives to strengthen the ecosystem. While technology has always been one step ahead of policy and regulation, it will need close collaboration between all players like the government, industry, startups and civil society to come together to keep this ever-evolving technology in check to harness its power for the welfare of its citizens.

Cybercrime powered by artificial intelligence is not just a technical issue, it's a social one. As AI tools become more accessible, so do their malicious uses. Protecting our elders from these invisible threats is not only a policy priority but a moral imperative. Because in the digital age, protecting the vulnerable isn't just good cybersecurity, it's a societal duty.

TABLE OF CONTENTS

AI AND CYBERSECURITY IN KARNATAKA: A ROADMAP FOR POLICY, REGULATION, AND PUBLIC TRUST

A Governance Perspective

PREPARED BY:

Dr. Shreevyas H M

Project Director- AI Cell @ Center for e-Governance, Government of Karnataka | Doctorate in AI | President Awardee I Ex-Research Scientist at Defence Institute of Advanced Technology (DIAT), DU, DRDO

AI and Cybersecurity in Karnataka: A Roadmap for Policy, Regulation, and Public Trust

Current Scenario: AI is Fuelling a Cybercrime Epidemic

India is rapidly moving online. By the end of 2026, the country will have one billion mobile users, and critical services like banking and healthcare are now digital.1 This shift brings convenience, but it has also created a new danger. Artificial Intelligence (AI) has made it incredibly easy for criminals to launch sophisticated cyberattacks.

Cybercrime is no longer for expert hackers. Widely available AI tools let criminals with few technical skills create scams that are automated, personalized, and massive in scale.1 These tools can write perfect phishing emails that get past spam filters, build fake websites, and create realistic social engineering traps.

The impact is clear, In Karnataka, AI was used in nearly 83% of all phishing email scams.These new attacks are different. They don't just trick you with logic; they trick your senses. By perfectly copying the voice of a family member or the face of a trusted official, AI scams create an emotional response that bypasses critical thinking.

Common AI-Powered Attacks

The scamsters are using AI to create new types of scams designed to manipulate people. Here are the most common methods:

Digital Arrest and Impersonation Scams:

Scammers pose as police, CBI, or other government officials.They call victims on video, often wearing uniforms, and accuse them of serious crimes like money laundering. They use personal details found in data breaches to make their claims sound real. The victim is placed under "digital arrest," where they are watched on a video call 24/7 and told not to contact anyone. The goal is to create intense fear and confusion, forcing the victim to transfer their savings to a "secure" government account for "verification." The money is never returned.

- An 83-year-old woman in Mumbai was held in "digital arrest" for a month and lost ₹7.7 crore.
- A 78-year-old man in Bengaluru lost his life savings of ₹83 lakh over two months.
- Two seniors in Mysuru, aged 81 and 73, lost a combined ₹1.92 crore in under two weeks.

Voice Cloning and Deepfake Scams:

This is one of the most dangerous threats. AI can create a perfect copy of someone's voice from just a few seconds of audio, often taken from social media. Criminals use this in "grandparent scams." A senior gets a panicked call from what sounds exactly like their grandchild, who claims to be in an accident or jail and needs money immediately. The emotional shock bypasses rational thought.

- A senior in Delhi lost ₹50,000 after hearing a cloned voice of his young relative begging for help.
- A 68-year-old businessman in Mumbai lost ₹80,000 after hearing his son's cloned voice asking for bail money.
-

The technology now includes deepfake videos. Criminals use fake videos of famous business leaders like N.R. Narayana Murthy and Mukesh Ambani to promote fake investment schemes.In Bengaluru, residents lost ₹95 lakh to these scams.This tactic is especially powerful in a developing economy where these figures are highly trusted.

Intelligent Phishing and Investment Fraud

AI has made old-fashioned phishing scams much more effective. AI can write personalized emails that look completely real, tricking people into clicking malicious links. These links often lead to fake investment websites. Scammers build trust by paying out a small return on an initial investment, which convinces the victim to invest a much larger amount before the criminals disappear.

Scam Type	AI Technology Used	Key Psychological Triggers
Digital Arrest	Social Engineering, Deepfake Video	Fear, Authority, Urgency, Isolation
Deepfake Investment	AI Voice Cloning.	Panic, Love, Shock.
Voice Cloning	AI Deepfake Video Generation	Greed, Trust in Public Figures.
Intelligent Phishing	Generative AI.	Credibility, Urgency.

The Financial Impact: A National Crisis

In 2024, Indians lost ₹22,812 crore ($2.78 billion) to cybercrime. This is nearly three times the amount lost in 2023 and ten times the amount from 2022.

Moreover, Cybercrime cases in Bengaluru have jumped up by 77% between 2022 and 2023. In rural Karnataka, cases nearly doubled in the same period, and for the first time, crimes were reported in tribal areas. This shows the threat is spreading everywhere.

Globally, India is a top target. In 2024, India was the second most-targeted country for crypto-related cyberattacks, after the United States.2 The huge number of new internet users, many with low digital literacy, makes the country a perfect target for criminals. This is not just a crime problem; it is a threat to our national economic security.

Why Seniors Are a Prime Target

Cybercriminals target senior citizens for specific reasons. Their strategy exploits a mix of technology gaps, psychology, and generational differences.

The "Digital Native" Gap

India has over 13 crore senior citizens, most of whom are "digital non-natives". They did not grow up with the internet and are naturally more trusting. In the analog world, a voice on the phone was real, and a person in uniform was respected. Criminals exploit this trusting nature, which is a major weakness online.

The numbers show the problem clearly:

Many seniors hesitate to report fraud, often due to embarrassment and concern about how their families might perceive it. The transition towards digital government services, such as online pension payments, creates a significant challenge, forcing seniors to engage with technologies they don't fully comprehend. The complex steps involved, like filling out online forms and remembering personal information for OTPs, can leave them feeling helpless. Essentially, we've provided technology to our elders without ensuring they have the necessary skills and support for safe and effective use.

Scammers Exploit Fear, Greed, and Trust of Elderly

Criminals are experts in psychological manipulation. Scams targeting seniors rely on three main emotional triggers: ignorance of technology, fear of authority, and the promise of greed or easy money.AI allows criminals to create urgent, personalized messages that amplify these feelings.

Social isolation makes things worse. Many victims are seniors who live alone, with their children living far away. When a scammer pretending to be a CBI officer calls, there is no one nearby to ask for a second opinion.

This isn't about education level. Highly educated people, including retired government officials and professionals, are also falling for these scams. The problem is that this generation was raised to respect authority. Criminals are now using that respect against them. Awareness campaigns need to go beyond technical tips. They must teach a new kind of digital scepticism and encourage seniors to verify everything, even when it seems to come from a trusted source.

The harm from these crimes is not just financial. Victims suffer from severe anxiety, depression, and a loss of trust. This also leads them to stop using online services altogether, cutting them off from essential support systems and defeating the goal of digital inclusion.

There is also a high cost to society. Government resources are spent on fraud prevention and investigation instead of other public needs. In a cruel twist, some victims are scammed a second time. After losing their savings, they search online for help and are targeted by fake legal aid services. One victim in Bengaluru who lost ₹1.5 crore was tricked into paying another ₹12.5 lakh to a fake company that promised to recover his money13. This cycle of fraud creates a climate of fear that threatens India's digital future.

The Governance Gap & Measures

The rise of AI-driven cybercrime shows that India's laws and institutions are not prepared. Our current system is reactive instead of pro-active and fragmented, leaving us vulnerable to these new, sophisticated threats. For instance, the India’s main cyber law, the Information Technology Act, 2000, was written for a different era. It covers basic cybercrimes like identity theft and hacking, but it has no specific rules for AI-generated threats like deepfakes or voice clones.

Prosecutors have to use broad, outdated laws to charge criminals for these very specific crimes. This makes it hard to get convictions and does little to stop criminals, who know how to exploit these legal grey areas.14 The law focuses on punishing the crime after it happens. We need a new approach that regulates the AI tools themselves, putting responsibility on the companies that create and distribute them.

Although, the government has tried to fix these gaps with new rules like the IT Rules, 2021, the Digital Personal Data Protection (DPDP) Act, 2023, and the Bharatiya Nyaya Sanhita (BNS), 2023.24

While these are good steps, they create a confusing patchwork of regulations instead of a clear, unified strategy. A major grey area still remains; there is no clear rule for who is liable when an AI system causes harm. The focus is on the illegal content (like a deepfake video), not the AI tools used to create it. This allows the developers of powerful AI models to avoid responsibility.

Four Key Roadblocks to a Strong Defence

Beyond weak laws, India faces several systemic problems that prevent an effective defence against AI cyber threats.

01

The Cost Barrier: Criminals can access powerful AI tools for free or at a low cost. But the AI-powered cybersecurity tools needed to defend against them are very expensive.A recent report by IBM found that 73% of Indian organizations have little to no AI-driven security, leaving them vulnerable

02

The Data Problem: Good defensive AI needs good data. India's public sector does not have a strong data governance framework to provide it.Initiatives like the National Data Governance Framework Policy are in progress but not yet fully effective.Worse, major data breaches, like those affecting Aadhaar, give criminals the very data they need to train their AI for targeted attacks.

03

Poor Coordination: India has several cybersecurity agencies, including CERT-In and the I4C. However, they do not share threat information effectively with each other, state police, or private companies like banks. Criminals operate globally, but our response is slowed down by bureaucracy.

04

The lack of local talent and the high cost of foreign technology create a "cyber-sovereignty risk." If India cannot build its own defences, it will become dependent on other countries for its national security. The "Make in India" initiative must be extended to digital defence

To fight AI-driven cybercrime, India needs a clear, multi-stakeholder action plan. We can learn from other countries and use AI's positive potential to build a secure digital future.

International Landscape on AI regulations

Several nations are also grappling with AI regulation. Their approaches offer valuable lessons.

United Kingdom: The UK uses a flexible, "pro-innovation" approach. Instead of one big AI law, it lets existing regulators in sectors like finance and healthcare apply key principles (e.g., Safety, Fairness, Accountability).This uses existing expertise and avoids rigid rules. The UK also uses AI fraud detection tool to fight fraud in the public sector.

United States: The US has no single federal AI law. States have taken the lead, passing laws against specific harms like election deepfakes.

Singapore: Singapore focuses on strong public-private partnerships. It created the AI Verify Foundation, where the government works with companies like Google and IBM to develop tools and standards for responsible AI.The approach is voluntary and collaborative, aiming to build a trusted ecosystem.

The key lesson is that the best models are flexible and collaborative, not rigid and top-down. India needs a hybrid approach: pass laws against clear dangers like deepfake fraud, but also work with industry to create standards that can adapt as technology changes.

Country	Legislative Approach	Key Principles	Governance Body	Key Takeaway for India
India	Amending existing laws (IT Act, DPDP Act).	Content moderation, Data protection	MeitY, CERT-In, I4C, State Police.	Current laws are not enough; a proactive framework is needed.
United Kingdom	Principles-based, sector-specific guidance.	Safety, Transparency, Fairness, Accountability.	Existing Sectoral Regulators.	Leverage domain expertise and stay flexible.
United States	Fragmented; State-level laws on specific harms.	Anti-fraud, Consumer protection.	Federal Trade Commission (FTC).	Strong federal-state coordination is important.
Singapore	Voluntary frameworks; Public-private partnerships.	Building a trusted ecosystem, Shared responsibility.	IMDA, AI Verify Foundation.	Public-private collaboration is powerful.

AI Isn't Just a Threat; It's an Opportunity

While the risks are serious, AI also has huge potential for good. The goal of regulation should be to manage the harm without killing innovation. India is already using AI to solve major challenges.

Healthcare: AI is helping to diagnose diseases like tuberculosis and breast cancer faster and more accurately. It is also powering telemedicine platforms that bring healthcare to remote villages.

Agriculture: AI is helping farmers increase crop yields and use fewer pesticides. In Telangana, the "Saagu Baagu" project used AI tools to help chili farmers double their income per acre.22
These examples show that AI can be a powerful tool for national development. We need a balanced approach that encourages this positive potential.\

Three Strategic Priorities:

A national strategy to fight AI-driven cybercrime should focus on three areas:

1. Update Our Regulations:

Create a Risk-Based AI Safety Framework: Classify AI systems by risk across sectors, High-risk AI (in finance, healthcare, etc.) should face strict rules like mandatory audits and pre-deployment testing.

Mandate "Safety by Design": Require AI developers to build safety features into their products, such as identity verification for voice cloning services and digital watermarks to identify synthetic media.

2. Build Government Capacity:

Launch a National Mission for AI in Government: Aggressively train at least 500,000 government employees from police to policymakers in AI literacy and cyber defence over the next five years.

Establish Public Sector AI "Sandboxes": Create controlled environments where government agencies can test new AI security tools from startups before deploying them widely.

3. Drive Public-Private-Civic Collaboration:

Develop a Joint Threat Intelligence Platform: Create a platform where banks, tech companies, and government agencies are required to share real-time data on new scams and fraudulent accounts.

Launch a National Digital Literacy Mission ("Digital Suraksha Abhiyan"): Start a massive, multi-lingual awareness campaign, especially for senior citizens, through trusted channels like banks and community centers. The campaign should teach simple, practical advice, like using a family "safe word" to verify urgent requests for money.

Incentivize "Cybersecurity for India": Use grants and tax breaks to support a domestic cybersecurity industry that builds affordable, AI-powered defence tools designed for India's needs.

The Way Forward: Karnataka Can Lead the Nation

Karnataka is in a unique position to lead India's fight against AI-powered cybercrime. With its world-class tech ecosystem and proactive government, the state can be a national laboratory for developing and testing the solutions India needs.

Karnataka Has the Right Ingredients: Tech, Talent, and Capital

The city of Bengaluru, recognized as the "Silicon Valley of India," provides an unparalleled ecosystem for cybersecurity advancement. The city anchors R&D centers for global giants like Google and Microsoft, headquarters Indian IT leaders such as Infosys and Wipro, and fosters a dynamic startup culture with over 45 unicorns. This industrial might is supported by a talent pool of over one million technology professionals, including the nation's largest concentration of AI/ML experts. The presence of premier academic institutions like the Indian Institute of Science (IISc) and the Indian Institute of Management (IIM) further solidifies the state's readiness to lead in the cybersecurity domain.The combination of industry, academia, and talent makes Karnataka the perfect place to create the next generation of cybersecurity defences for the nation.

Karnataka's Government Is Already Taking Action

The state government has shown strong political will and a forward-thinking approach to cybersecurity by launching the cyber security policy in 2024 to combat the digital threats. The strategic initiative aimed at fortifying the state's digital infrastructure and creating a secure and resilient cyberspace for its citizens, businesses, and government entities. The policy, backed by a financial outlay of ₹103.87 crore over five years, focuses on a multi-pronged approach encompassing awareness, skill development, innovation, and public-private partnerships.

It focuses on five pillars: awareness, skill-building, innovation, industry promotion, and collaboration. It includes practical steps like paying stipends to cybersecurity interns and giving grants to startups.25 The policy has a public part for building the ecosystem and a confidential part for securing government systems, a smart model for other states to follow. This forward-looking policy positions Karnataka as a proactive leader in addressing the challenges of the digital age and underscores its commitment to fostering a safe and secure online environment for all.

Karnataka is the first state to create a unified Cyber Command Unit led by a DGP-rank officer. This brings all 45 of the state's cyber police stations under a single command, fixing the problem of fragmented enforcement. The state's IT Minister, has pushed for balancing regulation with innovation, announced a plan to train 500,000 people in new technologies, and is launching a state-level AI Mission.

The state is thoughtfully debating how to regulate AI, considering a hybrid approach of a dedicated "AI Mission" for innovation and embedding AI governance rules into its broader IT policy. This pragmatic strategy is a good model for the rest of India.

Recommendation: Make Karnataka a National Pilot

Karnataka is the ideal "policy sandbox" for India. New laws, public-private partnerships, and awareness campaigns can be piloted in Karnataka's diverse environment.
The collaborative way Karnataka drafted its cybersecurity policy involving multiple government departments and academic partners like IISc is exactly what is needed at the national level. By providing federal resources and a formal process to scale successful programs from Karnataka to the rest of the country, a state-level success can become a national shield.

Author's Sources

1 https://cxotoday.com/press-release/a-new-report-reveals-how-ai-is-empowering-cybercrime-and-what-india-must-do-about-it/
2 https://www.newindianexpress.com/states/karnataka/2025/Jun/26/ai-driving-force-behind-828-per-cent-of-phishing-emails-in-karnataka
3 https://timesofindia.indiatimes.com/city/mumbai/83-year-old-south-mumbai-woman-kept-under-digital-arrest-for-a-month-loses-rs-7-7-crore-cyber-police-register-cheating-and-extortion-case/articleshow/123311368.cms
4 https://timesofindia.indiatimes.com/city/mumbai/83-year-old-south-mumbai-woman-kept-under-digital-arrest-for-a-month-loses-rs-7-7-crore-cyber-police-register-cheating-and-extortion-case/articleshow/123311368.cms
5 https://www.thehindu.com/news/national/karnataka/gang-of-four-digitally-arrest-senior-citizen-in-bengaluru-rob-83-lakh/article69846419.ece
6 https://timesofindia.indiatimes.com/city/mysuru/two-mys-women-both-sr-citizens-lose-nearly-rs-2-cr-in-online-fraud/articleshow/123263846.cms
7 https://www.ndtv.com/ai/ai-scams-surge-voice-cloning-and-deepfake-threats-sweep-india-6759260
8 https://timesofindia.indiatimes.com/city/mumbai/mumbai-businessman-falls-victim-to-ai-voice-cloning-loses-rs-80000/articleshow/109225530.cms
9 https://nquiringminds.com/cybernews/bengaluru-residents-lose-rs-95-lakh-to-deepfake-investment-scams/
10 https://timesofindia.indiatimes.com/technology/tech-news/bengaluru-residents-duped-of-rs-95-lakh-by-deepfake-videos-of-narayana-murthy-and-mukesh-ambani/articleshow/114955868.cms
11 https://cxotoday.com/press-release/a-new-report-reveals-how-ai-is-empowering-cybercrime-and-what-india-must-do-about-it/
12 https://www.storyboard18.com/digital/45-of-indian-seniors-struggle-to-identify-online-fraud-cybercrimes-on-the-rise-58331.htm
13 https://www.thehindu.com/news/cities/bangalore/techie-held-for-running-fake-online-legal-aid-service-for-cybercrime-victims-in-bengaluru/article69892624.ece
14 https://www.vintagelegalvl.com/post/deepfake-technology-and-it-s-legal-regulation-in-india-a-doctrinal-and-comparative-study
15 https://ardorcomm-media.com/majority-of-indian-firms-lack-ai-governance-amid-soaring-data-breach-costs-ibm-report/
16 https://dig.watch/resource/indias-national-data-governance-framework-policy-draft
17 https://www.gov.uk/government/publications/ai-regulation-a-pro-innovation-approach/white-paper
18 https://www.gov.uk/government/news/criminals-should-be-aware-says-minister-as-government-upgrades-ai-fraud-detection-tool
19 https://www.aoshearman.com/en/insights/ao-shearman-on-tech/zooming-in-on-ai-tackling-deepfakes-around-the-world
20 https://www.pdpc.gov.sg/help-and-resources/2020/01/model-ai-governance-framework
21 https://indiaai.gov.in/article/ai-for-social-good-the-role-of-rigorous-impact-evaluations-in-maximizing-ai-s-potential
22 https://farmonaut.com/asia/ai-in-agriculture-india-7-ways-transforming-farming
23 https://www.gov.uk/government/publications/ai-regulation-a-pro-innovation-approach/white-paper
24 https://www.deccanherald.com/india/karnataka/ai-disruptions-will-be-brief-as-reskilling-drives-new-jobs-karnataka-it-minister-priyank-kharge-3663223
25https://www.cyberpeace.org/resources/blogs/karnataka-government-launched-the-cyber-security-policy-2024
26 https://www.deccanherald.com/india/karnataka/ai-disruptions-will-be-brief-as-reskilling-drives-new-jobs-karnataka-it-minister-priyank-kharge-3663223

AN INDUSTRY PERSPECTIVE

PREPARED BY:

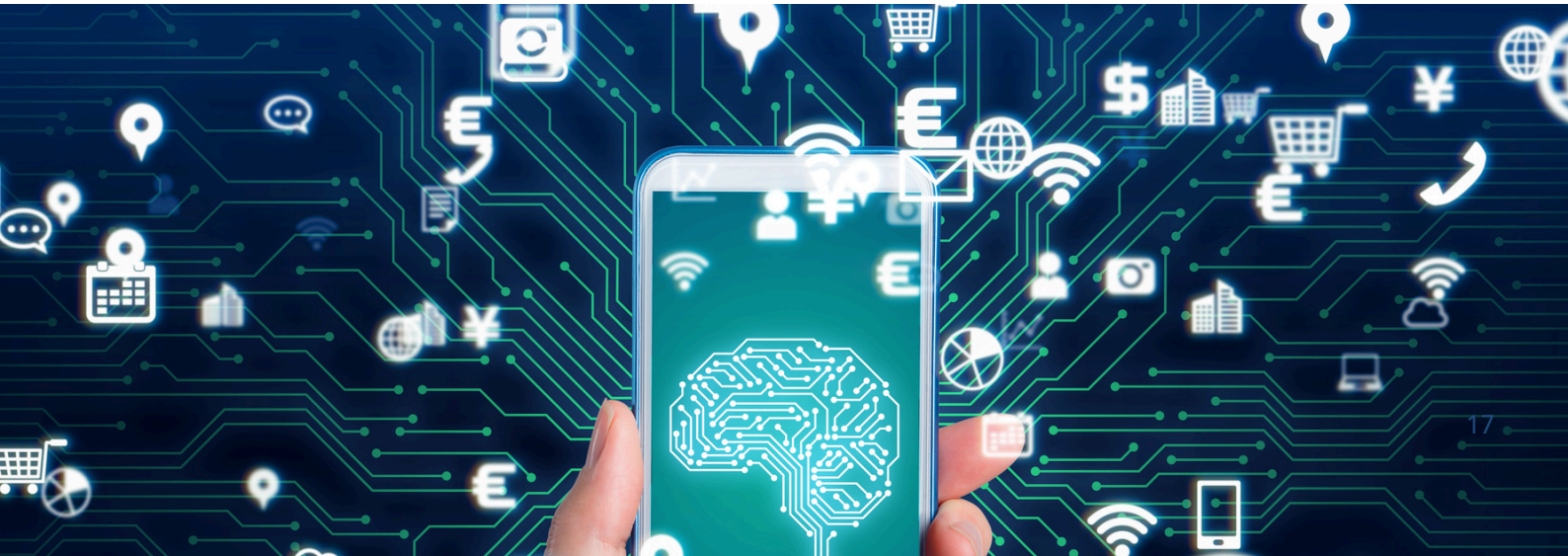
MR. SANAT RAO

Board Member & ex-CEO, Infosys Finacle I Digital Anthropologist & AI Ethicist I Fellow - Royal Anthropological Institute I Co-founder & MD - WithinTheBox.ai I Alumnus of Indian Institute of Management Bangalore

Introduction: A New Kind of Danger

"Help me, Appa!" Those words echoed over the phone to a retired banker in Bangalore, in what sounded exactly like his son's panicked voice. And of course, his first instinct was to panic. In reality, it was an AI-generated voice clone crafted by scammers to demand a ransom, something that is not uncommon in different parts of the world. Thankfully, this person had the presence of mind, and he double-checked and discovered his son was safe. But that is not typically the case with many others who get hoodwinked into acting exactly as the scammers would have them. How often have we heard stories of someone receiving an email from a family member or a close friend asking for help. Such incidents underline a growing threat that of cybercriminals leveraging Artificial Intelligence (AI) to mimic loved ones' voices and personalities. The logic is simple – they prey on trust and fear on the one hand, and a certain amount of ignorance on the other to scam their victim.

India's rapid digitalization has set the stage for this new breed of cybercrime. The country is poised to have 1 billion smartphone users by 2026, even as public services like pension disbursement and hospital appointments are all moving online. As India's digital footprint expands, so does the "target market" for bad actors exploiting technology. AI, often hailed as a revolutionary force for good, is now being increasingly weaponized by criminals to automate, personalize, and scale their attacks with alarming precision. The result is a wave of highly convincing and seemingly genuine scams. The techniques vary, from cloned voices and deepfake videos to intelligent phishing emails. There is also a new demographic that is the target of such scams, namely senior citizens. India has roughly 138 million people aged 60 and above. This is generation that was born well before the internet but it's a group that is coming online in increasingly large numbers. Many of them are internet-literate and know the right from the wrong. But a large majority are now firmly in the crosshairs of AI-powered cybercrime. These scams are emotionally manipulative, sometimes multilingual (thanks to AI's translation and voice capabilities), and tailored to exploit the trusting nature and vulnerability of older people. In an increasingly digitized India, senior citizens, many of whom are digital non-natives, are facing a surge in AI-enabled scams that would have been unimaginable just a few years ago.



Why Are Senior Citizens at the Epicentre of Such Scams?

There are no easy answers to this question. Suffice to say that is the convergence of several factors including demographic, behavioural, and systemic which places India's senior citizens at the highest risk:

Digital Literacy Gap: The country maybe the 2nd largest market globally for smartphones and its usage in the country is only growing. And yet, despite rising smartphone penetration, it is only a small share of about 30% of Indian seniors who report they feel comfortable with digital banking or online communication and more than half of the elderly fear making errors while interacting with digital means (HelpAge India, 2025).

High Trust Quotient: Senior citizens were not born in the computer age. Their exposure to emails, computers and smartphones came very late in life. They tend to trust phone calls and official-looking emails rather easily. Having grown up in a largely non-digitised era, their instinct is to believe what they hear or see, making them far more vulnerable to deepfakes, cloned voices and the like.

Loneliness and Social Isolation: With the increasing migration of their younger generations to cities or abroad, many elderly Indians live alone. And when one of the partners passes away, the elder is even more isolated and lonely. This factor of loneliness has increased substantially after Covid. Given this state of mind, if anything, they are even more vulnerable to scamsters who exploit the emotional void with fake emergencies, false friendliness and other means.

Access to Resources: The senior citizens often have lump-sum savings from years of working, many of them receive decent pensions, gratuities, and returns from investments made either in real estate or in mutual funds or sometimes even in the stock market. Other than possibly medical expenses, their other expenses are minimal at that stage in life. One successful scam can yield substantial sums thereby making them prime high-value targets.

Low Reporting Rates: Finally, there is yet another unfortunate reason. When they become victims to such scams, many of the senior citizens possibly feel ashamed of being duped, or they fear any judgment on crime reported, or many times unfortunately they simply don't know what to do and where to report scams. Therefore it is not surprising that a large majority of financial frauds against senior citizens actually go unreported.

This pattern is not dissimilar to many other countries. Research reports highlight that older adults are more likely to be victims of impersonation fraud and less likely to recover funds due to lower tech savviness and slower institutional response. India has an opportunity to get ahead of this curve.

The Real Cost of Inaction

Failing to address AI-powered fraud against seniors could lead to wide-ranging repercussions:

Financial Ruin: For many seniors, especially those without extended family support, a large financial fraud can be devastating, indeed life changing. With limited earning potential, they risk long-term dependency in the absence of which they lead to a state of financial ruin.

Erosion of Trust in Digital Platforms: Repeated exposure to fraud risks making digital systems appear unsafe. There is a fear factor and quickly a trust-deficit sets in. At a time when the country is making huge strides in digitalisation of a host of public services, the erosion of trust can easily set back national goals such as Digital India, JAM trinity (Jan Dhan-Aadhaar-Mobile) and other initiatives.

Emotional and Psychological Harm: Besides shame and loss of face, many victims also often suffer great trauma, severe depression, and deep anxiety. Several cases have been reported of victims of such scams who lost their entire savings to a deepfake scam were hospitalised due to severe shock.

Intergenerational Distrust: It takes just one bad experience to immeasurably cause damage. The scars from such incidents are deep and long-lasting. As seniors become wary of any digital interaction, even legitimate communications from banks or family members may be met with scepticism, affecting family dynamics and care systems.

Wider Societal Costs: Increased dependence on welfare, growing healthcare burdens due to stress-related conditions, and a general sense of fear among the elderly can reduce overall societal resilience.

Harnessing AI for Protection: Flip the Script

While AI is being used to perpetrate fraud, it can also serve as a powerful tool for defence. There is a good side to the AI too. Strategic applications include:

Behavioural Anomaly Detection: AI systems can flag suspicious banking transactions based on a user's historical behaviour patterns. These systems learn normal spending habits, transaction timings, and recipient patterns, automatically triggering alerts when elderly users suddenly transfer large amounts to unfamiliar accounts or exhibit uncharacteristic digital behaviour that could indicate coercion.

Deepfake Detection Tools: Emerging startups, including some India-based companies, are developing real-time detection for synthetic media. These tools analyse micro-expressions, voice spectral patterns, and video compression artifacts that are difficult for current AI generation tools to perfectly replicate, providing a technological arms race advantage for defenders.

Caller Verification Algorithms: Telecom providers can integrate real-time caller verification APIs to detect spoofed numbers or voice modulation. Advanced systems can cross-reference caller location data, voice pattern analysis, and historical communication patterns to assign confidence scores to incoming calls, warning users when authenticity is questionable.

Fraud Alert Systems: AI can be embedded into banking apps to prompt extra verification when unusual behaviour is detected (e.g., "Are you sure this is your grandson? Click here to confirm their identity via OTP"). These context-aware systems can detect emotional urgency in transaction descriptions and automatically suggest verification steps like calling known family numbers or requiring multi-person approval for large transfers.

Proactive Scam Baiting: Virgin Media O2 in the UK launched an innovative initiative to keep the scamsters at bay. They have introduced 'Daisy' (dAIsy), a custom-made human-like chatbot that answers calls in real-time, keeping fraudsters on the phone for as long as possible in a bid to annoy them. They have automated the practise of 'scambaiting' which involves people posing as potential victims to squander scammers' time and resources, then publicly expose their nefarious activities etc. Inspired by the UK's Virgin Media O2 "Daisy AI" initiative, India could develop local-language bots that keep scammers occupied and flagged. These AI systems can engage fraudsters in lengthy conversations in multiple languages, wasting their time while simultaneously gathering intelligence on scam tactics and potentially identifying criminal networks through call pattern analysis

These solutions must be complemented by human support systems and legal safeguards.

Empowering the Elderly Through Awareness

Technology alone cannot solve the problem. Equipping the elderly with knowledge and digital confidence is crucial. Suggested initiatives include:

Senior-Centric Digital Literacy Campaigns: Partner with NGOs to run awareness drives at pension offices, religious gatherings, and community centres. These sessions should use relatable scenarios and hands-on demonstrations, allowing seniors to practice identifying scams in a safe environment with immediate feedback from trained facilitators.

Local Language Training Modules: Develop simple, visual training content in Kannada, Tamil, Hindi, and other languages covering common scams, how to verify calls, and what to avoid online. Content should include audio examples of real vs. AI-generated voices and step-by-step visual guides that seniors can refer to when unsure about suspicious communications.

Intergenerational Mentoring: Promote family-driven support where grandchildren act as "digital buddies" for elders. This could include scheduled weekly check-ins about digital activities and establishing family code words that can be used to verify authentic emergency calls from relatives.

Media Campaigns: Launch emotive storytelling ads on TV and radio demonstrating real scam cases and protective measures. These campaigns should feature actual senior citizens sharing their experiences and recovery stories, making the content more relatable and reducing the stigma around being targeted by scammers.

Community-Based Hotlines: A state-wide senior digital helpdesk could offer advice and support in regional languages. Karnataka could pilot this. The helpline should operate 24/7 with trained counsellors who can walk seniors through suspicious scenarios in real-time and provide immediate guidance on whether to proceed with or report suspicious communications.

Policy Recommendations for Karnataka and Beyond

The Government of Karnataka can take the lead nationally by implementing pioneering policies:

India's Global Moment: From Risk to Leadership

India, with its world-leading IT services industry and vibrant startup ecosystem, is well-positioned to pioneer frameworks for AI and digital safety. Karnataka, home to India's tech capital Bengaluru, could:

- Launch a national blueprint on senior digital safety.
- Partner with UK institutions such as NCSC, Ofcom to adopt and adapt global best practices.

Facilitate academic-industry partnerships to develop AI ethics curricula and fraud-detection technologies.

The UK's approach to elder protection through tools like "Friends Against Scams" and the Office of the Public Guardian's awareness drives offer replicable models. India can build upon these with scale and multilingual reach.

Conclusion: A Call to Protect Those Who Once Protected Us

The image of an elderly parent or grandparent being duped by a faceless scammer is a distressing one. Yet, it is an increasingly frequent occurrence. AI-powered cybercrime targeting senior citizens is not a distant, futuristic threat; it is already here, and it is growing. However, as we have explored, this is a challenge that can be met. It requires us to shine a spotlight on the issue – exactly what this white paper aims to do – and then to follow through with collaborative action.

This is not a problem isolated to one country or one city. It is a larger problem, though its gravity may vary. The scams might have local flavors such as a deepfake scam in Bengaluru, a tech support fraud in London, a voice clone con in New York, but they all trade on human vulnerabilities and technological opportunism. The Government of Karnataka, by focusing on this issue, has a chance to become a leader in India's response to AI-enabled cybercrime. Karnataka, being a technology hub, is uniquely positioned to foster the kind of industry-government collaboration needed to develop cutting-edge defences. The state has already shown the initiative through several steps. The next step could be convening that expert group or task force to formulate concrete policy recommendations.

Protecting our senior citizens from online threats is not just about one demographic; it's about upholding the promise of a digital society for all. If the internet and AI become tools that only the young or the "in the know" can use safely, we will have failed a significant portion of our populace. Conversely, if we succeed in creating a secure digital environment for the elderly, it will likely be safer for everyone else as well. Think of it like wheelchair ramps on sidewalks, initially installed for those who need them, but ultimately making the path easier for all pedestrians.

In conclusion, alongside the growing threat of cybercrime targeted at the elderly, there is also an opportunity. By combining technological innovation, widespread education, and strong policy measures, we can ensure that AI's story in India is not about how it empowered criminals to exploit the vulnerable, but about how it was ultimately harnessed to protect and uplift the more vulnerable sections of society.

PART C

A TECHNOLOGY PERSPECTIVE

PREPARED BY:

MR.ANIL KUMAR NS

Co-Founder - Cogenz Cybertech I Indian Administrative Fellowship @ Government of Karnataka I Ex-UIDAI-GoI, Infosys, IBM and Cisco I Economic Times Awardee I Alumnus of Indian Institute of Management Bangalore.

Global Landscape of AI-Powered Cyber Threats

Globally, the AI cybersecurity threat landscape has intensified dramatically. 87% of organizations worldwide have faced AI-powered cyber-attacks in the last year, with 74% of IT security professionals reporting significant impact from AI-powered threats. AI has revolutionized phishing attacks globally with close to 82.6% of phishing emails now using AI technology with a 202% increase in phishing over the previous year alone. 78% of people open AI-generated phishing emails, with 21% clicking on malicious content. Credential phishing attacks increased by 703% compared to 2024.

Notable AI-Powered Attack Cases

- Several high-profile cases demonstrate the sophistication of AI-driven attacks:
- Midnight Blizzard used AI to launch phishing attacks via Microsoft Teams in 2023.
- Gmail AI hack early in 2024 used AI-generated emails and deepfake audio to bypass two-factor authentication.
- $25 million deepfake fraud against UK engineering firm Arup in February 2025, using AI-generated video of the company's CEO.

AI-Driven Cyber Threats in India

India is no exception to this trend. Aon's 2025 Asia-Pacific Cyber Risk Report notes a sharp rise in AI-fuelled attacks in India and the South Asian region. Cyber incidents in APAC jumped 29% year-over-year, and India saw especially targeted compromises. AI-generated deepfakes are directly blamed for a 53% jump in social engineering incidents, and fraud claims in India increased 233% in 2024. Attackers in India are increasingly unleashing advanced payloads malicious files are planted in enterprise networks to gain remote access, stolen data is used for ransomware extortion, and cloud credentials are hijacked to pivot within systems.

Experts attribute these trends to the rapid AI adoption in India. As one Aon analyst observes, "Data poisoning attacks can compromise the integrity of critical AI systems, and deepfake technology is now being used to craft convincing malicious content - making social engineering attacks more sophisticated than ever". In short, AI makes phishing and identity fraud harder to detect and trace. For example, voice-clone scams (already observed worldwide) have also emerged here, with fraudsters mimicking business executives to authorize fake payments. The Government of India recognizes this risk.

Scale of Impact

India recorded Rs 22,812 crore ($2.78B) in losses from digital fraud in 2024; AI-driven schemes were cited as a major driver by a report released in Bengaluru, noting AI was used in 80% of phishing campaigns.

Cybercrime complaints reached 1.91 million in 2024, up from 1.55 million in 2023, nearly 10x higher than 2019 levels, highlighting steep growth in AI-accelerated threats and fraud.

Karnataka data illustrates the acceleration: Rs 2,396 crore lost in 2024,Rs 562 crore in 2023.Rs 113 crore in 2022, with authorities noting that a majority of incidents leveraged AI tools (e.g., phishing, cloning sites, deepfakes).

Financial and Operational Costs

The global cost of data breaches has reached $4.88 million on average over the past year, representing a 10% increase and an all-time high. Cybercrime is projected to cost a staggering $13.82 trillion to the global economy by 2032.

Key Drivers of AI-Enabled Cybercrime

- Accessibility of generative AI tools for content creation (emails, voices, videos) enabling highly personalized social engineering at low cost.
- Automation and scale: AI accelerates reconnaissance, credential attacks, phishing campaigns, and malware mutation, expanding attack surfaces and reducing defenders' reaction windows.
- Identity-centric weaknesses: Reliance on weak identity verification in workflows, especially approvals via voice/video or email, raises risk against deepfake impersonation.
- Cloud and hybrid ecosystems: Misconfigurations and sprawling identities in multi-cloud and remote/hybrid work environments increase exposure to AI-driven probing and exploitation.

Current Gaps in India's Cyber AI Preparedness

- Reliance on manual monitoring over AI-driven automation.
- Absence of national datasets for training AI detection models.
- Fragmented R&D ecosystem with insufficient coordination between government, academia, and startups.
- Limited focus on AI robustness and assurance testing, leaving critical systems exposed to adversarial AI.

Industry-Specific Playbook

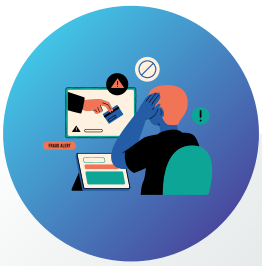
1.Banking, Insurance and Financial Services

Banks are being hammered by AI-assisted social engineering and "business email compromise" (BEC) that now includes voice/video deepfakes of senior executives. Attackers use language models to write fluent emails, tune persuasion styles, and automate follow-ups; they also clone voices or faces to authorize urgent transfers on live calls. The Hong Kong/Arup case, where a fake "CFO" on a video call duped staff into sending ~US$25M, is the canonical example of how convincing these scams have become

Commonly used tactics

Deepfake video meetings to approve urgent payments.

AI-generated email/write-ups that mimic internal tone and formatting



Prompt-injection tricks against AI copilots used by finance teams

Some key defences you can adopt:

- Out-of-band verification for any new/changed beneficiary or urgent transfer (mandatory dual human checks)
- Staff drills on deepfake red flags (latency, micro-sync issues, refusal to turn camera, and "we changed process just for today" claims
- Lock down AI assistants integrated with email/drive: enable output filtering, content provenance checks, and restrict tool/agent actions
- Align with RBI's new recommendations for safe AI adoption in finance (governance, auditability, guardrails).

2.Healthcare & Life Sciences

Healthcare staff face AI-sharpened data theft risks. This is critical since patient data and research IP are valuable and clinical operations are time-critical. Recent Indian cases show scammers using staged video calls and forged documents to coerce payments. In healthcare, AI already assists in radiology and pathology, allowing faster diagnosis and reducing the load on doctors. For child nutrition programs, simple image recognition tools could flag malnutrition risks in Anganwadi centres. Telemedicine and assistive robots can extend care to senior citizens, especially in remote areas.

Commonly used tactics

AI-tailored spear-phishing that references real clinical trials, procurement SKUs, or vendor tickets.

Model manipulation against hospital copilots (indirect prompt injection in documents, emails, calendars used by staff AIs)

Some key defences you can adopt:

- Strict "two-channel" verification for any external payment demands or PHI access approvals.
- Disable risky auto-actions in AI agents tied to EHR, email, or cloud storage; force explicit user confirmation.
- Run NCSC/OWASP-aligned hardening for LLM apps (prompt-injection defences, output handling, supply-chain checks).

3.Government & Public Sector

As government agencies pilot AI assistants for citizen services and internal workflows, adversaries try to plant hidden instructions in PDFs, websites, or calendar invites so that the assistant performs unintended actions (data exfiltration, spoofed alerts, or risky tool invocations). Recent research demonstrations against Gemini show how a poisoned invite or document title can trigger actions across integrated tools.

Commonly used tactics

Indirect prompt-injection via open data portals, RFP docs, and citizen emails.

Data-poisoning and insecure output handling, listed among OWASP's LLM Top 10 risks.

Some key defences you can adopt:

- Treat AI like any other critical system: follow NCSC's secure ML deployment principles and OWASP LLM Top 10.
- Restrict which tools an assistant can invoke; require human approvals for sensitive actions (payments, data exports).

4.Utility and Transport Industry

Utilities and transport industries increasingly use predictive AI across OT/IT in their operations. Attacks shift to poisoning data sources, manipulating maintenance notes, or using agent chains to reach into control-adjacent systems via enterprise tools (e.g., email/drive/calendar) rather than hammering OT directly. Guidance now stresses lifecycle security for AI and agent integrations.

Commonly used tactics

Training-data and context poisoning to skew models or drive bad recommendations.

Social engineering via deepfake calls to plant urgent “work orders” or override approvals

Some key defences you can adopt:

- Build an AI risk inventory (what models, what data, what tools they can touch).
- Enforce context provenance (only ingest trusted, signed data sources into AI workflows).

5.Education & Research

Universities are rolling out AI copilots for admin and student support. Attackers seed hidden prompts in docs, LMS pages, or shared drives to elicit credential collection pages or exfiltrate class lists. Security bodies continue to warn that prompt injection is the most common LLM attack class and must be assumed in the threat model.

Some key defences you can adopt:

- Disable "auto-open links/auto-run tools" from AI outputs; require staff clicks and approvals.
- Apply OWASP LLM Top 10 controls; scan shared course materials for hidden instructions.

Attack pattern	What it looks like	First control to apply
Deepfake CFO/CEO	Urgent video/voice request for funds	Call-back to a known number; dual approvals
"Summarize this email" trap	Hidden text makes AI add fake warnings/links	Disable auto-actions; output filtering; user confirmation
Poisoned calendar/doc title	AI triggers tool actions from metadata	Sanitize metadata; limit tool scope; require approvals
LLM data poisoning	Model goes "off" after new data	Signed data sources; retrain hygiene; change control
Insecure output handling	AI prints links/commands that look official	Never trust links blindly; client-side URL/mail filters

Essential Policy & Regulatory Frameworks

India is advancing AI governance with the AI Safety Institute, RBI's FREE-AI framework, and a tolerant supervisory stance to balance innovation with oversight. Policies also stress zero-trust, AI-aware defences and AI-driven cybersecurity for stronger resilience.

RBI's FREE-AI Framework for Finance

The RBI's panel recently introduced the "FREE-AI" framework for Responsible and Ethical Enablement of Artificial Intelligence to guide the financial sector in balancing innovation with risk mitigation through data protection, transparency, and accountability

India AI Safety Institute

In early 2025, India launched the India AI Safety Institute under the IndiaAI Mission. It supports safe and ethical AI deployment in collaboration with academia, industry, and international partners. It focuses on inclusive, responsible AI grounded in India's unique social context.

Tolerant Supervision of AI Use

The RBI also recommends a "tolerant supervisory stance" toward initial AI errors, provided institutions demonstrate strong safety protocols. This approach encourages experimentation while maintaining oversight.

Zero Trust & AI-Aware Defences

The RBI has urged institutions to adopt zero-trust architecture and AI-informed defences, warning against systemic risks from vendor lock-in. This is part of a broader push to build resilience in banking.

Cybersecurity Architecture

AI would be instrumental in automating threat detection, remediation, and mitigation, as part of the outlined strategies for securing government ICT infrastructure. The policy stresses continuous review and adoption of latest threat management practices and technology tools, implicitly including AI-based advancements to enhance cyber defence maturity.

Emerging AI Technologies in Cybersecurity

1.AI-driven Threat Intelligence Platforms

What it is: Uses machine learning (ML) to aggregate, correlate, and analyze threat feeds, darknet chatter, malware signatures, and IoCs (Indicators of Compromise).

Why it matters: Helps organizations predict and prevent attacks by identifying patterns of emerging threats (e.g., ransomware gangs, phishing kits).

Example: Recorded Future, ThreatConnect in India, CERT-In is exploring AI-assisted cyber threat intelligence.

2. Deep Learning for Malware Detection

What it is: Deep neural networks analyze behavior, API calls, and system changes to detect zero-day and polymorphic malware.

Why it matters: Outperforms traditional signature-based antivirus by catching previously unseen malware.

Example: Microsoft Defender's AI-based cloud protection.

3. User and Entity Behavior Analytics (UEBA)

What it is: AI models baseline "normal" user activity (logins, keystrokes, network use) and flags anomalies.

Why it matters: Detects insider threats and account takeovers that bypass firewalls and other basic security measures.

Example: Splunk UEBA, Exabeam, Banks in India use UEBA to catch unusual financial transaction patterns.

4. AI-powered SOAR (Security Orchestration, Automation, and Response)

What it is: SOAR platforms leverage AI to automate detection, triage, and incident response workflows.

Why it matters: By reducing manual work, Security Operations Centers (SOCs) can cut response times from hours to minutes.

Examples: Palo Alto Cortex XSOAR, N-Able Adlumin XDR/MDR

5. Generative AI for Red Teaming & Penetration Testing

What it is: Large Language Models (LLMs) simulate phishing campaigns, malicious code snippets, and adversarial tactics to test organizational defenses.

Why it matters: Security teams can think like attackers, identify weak points, and strengthen defenses proactively.

Examples: OpenAI's cybersecurity initiatives, DARPA's AI Cyber Challenge (AIxCC).

6. AI in Deception Technology

What it is: Advanced honeypots, honeyfiles, and decoys powered by ML lure attackers into controlled fake environments.

Why it matters: These tools give defenders early warnings while wasting attackers' time and resources.

Examples: Attivo Networks (now part of SentinelOne); pilot projects in Indian defense cybersecurity.

7. AI for IoT/OT Security

What it is: Machine learning models monitor Industrial Control Systems (ICS) and Internet of Things (IoT) devices for abnormal commands or traffic.

Why it matters: Safeguards critical infrastructure such as power grids, hospitals, and transportation systems.

Examples: Darktrace AI in smart factories; AI-driven SCADA monitoring pilots in Indian utilities.

8. Explainable AI (XAI) in Security

What it is: AI models that not only predict but also explain why an alert was triggered.
Why it matters: Builds trust with CISOs and auditors, and helps meet regulatory compliance (important in India's DPDP Act context).

Example: DARPA's XAI initiative, IBM's explainability modules.

Top 5 recommendations for policy makers, industry bodies and technologists.

AI-Specific Cybersecurity legislation.

1. Provisioning for amendments under the upcoming Digital India Act to address AI-driven threats, deepfakes, and phishing scams.
2. Need to implement standards for Mandatory Watermarking & Provenance
3. Creation of digital signatures and watermarking for AI-generated audio, video, and text to enable traceability.

Cyber Awareness Mission

1. There is a need to launch a Karnataka-led program with banks, telecoms, and NGOs to train senior citizens in recognizing AI-enabled scams.
2. Organization of awareness camps in rural areas for the vulnerable segments.
3. Creating a repository of cases on deepfake or AI-related fraud, integrated with CERT-In and state cyber police.

AI Cybersecurity Innovation Hubs

1. Bengaluru's startup ecosystem can be leveraged to build real-time detection tools for deepfakes and AI scams in Indian languages.
2. Invite citizen participation to explore ways to strengthen the existing CERT-In's AI Capabilities
3. Explore adoption of latest technologies to strengthen CERT-In with advanced AI monitoring, early warning systems, and partnerships with global threat intelligence networks.

Cross-Sector Cybersecurity Collaboration

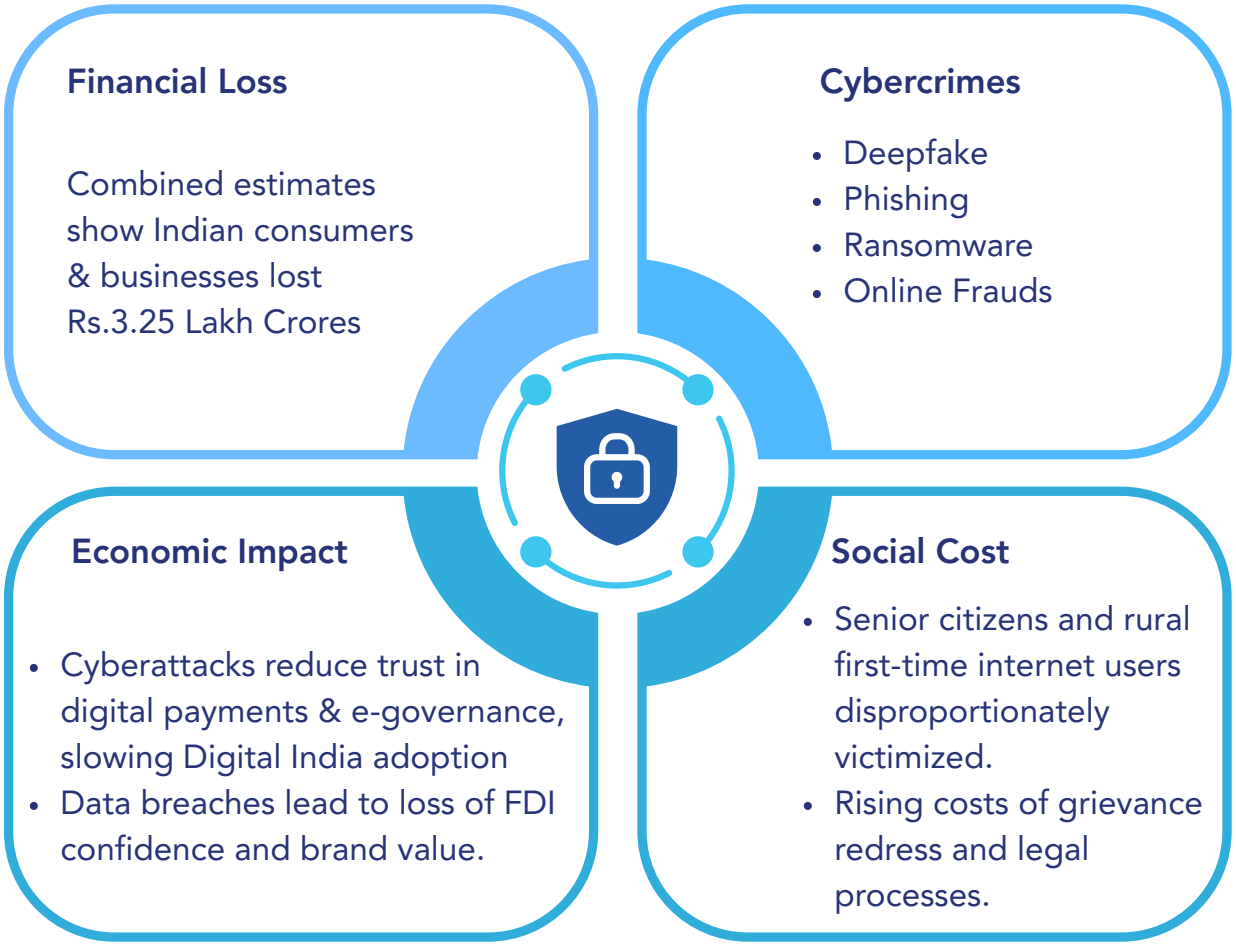
1. Conduct regular simulations involving government, banking, telecom, and civil society to test resilience against AI-driven threats.
2. Foster Public–Private Partnerships (PPP) for AI Security
3. Incentivize Indian startups and global tech firms to co-develop AI defense tools under Make in India and Digital India programs.

National AI-Cybersecurity Taskforce

1. Set up a central taskforce to coordinate legal, technical, and policy responses, ensuring agility in addressing evolving threats.
2. Opportunity to position Karnataka as a Pilot State
3. Karnataka model can be positioned as the national model for AI in cybersecurity by piloting awareness campaigns, detection tools, and regulatory sandboxes.

B. Indicative Cost–Benefit Framing: Cybercrime vs. Defensive Investment

1. Estimated Cost of Cybercrime (India, 2024 estimates)



2.Estimated Cost of Defensive Investment (Proposed)

- Awareness & Training (Senior Citizens, Citizens, Businesses):
  - Nationwide awareness campaigns + community workshops ≈ ₹500 crore annually.
- AI Fraud Helplines & Cyber Police Integration:
  - Setting up & operating state-level helplines ≈ ₹1,000 crore nationwide annually.
- CERT-In AI Upgrade & Innovation Hub:
  - AI anomaly detection + Bengaluru innovation hub ≈ ₹2,500 crore over 3 years.
- Regulatory & Legal Reforms (Digital India Act + AI watermarking):
  - Drafting, compliance enforcement ≈ ₹250 crore.
- Total Investment (first 3 years): ≈ ₹4,250 crore (~$500 million).

3. Cost–Benefit Comparison

Without Investment

Losses projected to exceed ₹5 lakh crore ($60 billion) annually by 2027

With Investment

A ₹4,250 crore investment in cyber defense delivers ₹32,500 crore in annual savings, with fraud reduction projected at 20 to 30% over time

Cybersecurity spending should be framed as economic protection, not just IT cost. By Investing ₹1 today in AI-driven defense prevents ₹8–10 in losses tomorrow. Defensive investment safeguards Digital India, fintech trust, and citizen welfare, especially vulnerable groups like senior citizens.

C.Critical Technology Domains and Strategic Needs

Generative AI & Deepfakes:

- Risk: Fraudulent impersonation, misinformation, financial scams.
- Response: To establish a national deepfake detection infrastructure, create a mandate watermarking of synthetic content, and integrate detection into citizen services.

AI-Powered Threat Hunting:

- Risk: There is a high risk of hidden attacks on financial institutions, telecom infrastructure, and citizen-facing digital platforms.
- Response: Need to embed AI-driven threat detection into state cyber cells for real-time monitoring and proactive defense.

Agentic AI (Autonomous Defenses)

- Risk: If uncontrolled, agentic AI could be weaponized to create adaptive, autonomous malware.
- Response: Investments in ethical and controlled deployment frameworks for agentic AI in cybersecurity.

Adversarial AI & Model Manipulation:

- Risk: Of attackers poisoning training datasets or corrupting detection models.
- Response: Set-up AI Assurance Labs in India to certify robustness of AI models against adversarial manipulation.

4.-Phased Roadmap for AI & Cybersecurity in India

In the Short-Term (0–18 months)
Objective: To Build immediate safeguards against rising AI-enabled cyber threats.

- Legislation: Insert AI-specific provisions into the upcoming Digital India Act to cover deepfakes, AI phishing, and data misuse.
- Senior Citizen Cyber Awareness Mission: Pilot in Karnataka and expand nationally in partnerships with banks, telecoms, NGOs etc. for outreach.
- Cyber fraud prevention Helplines: Establish 24/7 state-level helplines, integrated with CERT-In, police cyber cells, and grievance redress platforms.
- Mandatory Watermarking Pilot: Require AI provenance/watermarking on government-issued digital content.
- CERT-In AI Upgrade Phase 1: Deploy AI anomaly detection tools to track phishing and fake websites.

Medium-Term (18 months–3 years)
Objective: To Build resilient systems and institutional frameworks.

- National AI-Cybersecurity Taskforce: Create a multi-stakeholder body (govt, startups, academia, civil society) for strategy oversight.
- Cross-Sector Cybersecurity Drills: Conduct annual national drills simulating AI-driven attacks across telecom, BFSI, and e-governance.
- AI Cybersecurity Innovation Hub (Bengaluru): Support startups building AI deepfake detection and multilingual scam prevention tools.
- Expansion of AI Fraud Helplines: Scale to all states with integration into digital public infrastructure (e.g., DigiLocker, Aadhaar services).
- CERT-In AI Upgrade Phase 2: Expand to behavioral analytics for social engineering, voice spoofing, and GAN-generated fakes.

Long-Term (3–5 years)
Objective: To Position India as a leader in AI-driven cybersecurity defense.

- Legislation 2.0: Enforce mandatory watermarking and provenance standards for all AI-generated media across public and private sectors.
- Karnataka as National Pilot Model: Make Karnataka a regulatory sandbox for AI security solutions, then scale nationwide.
- Public–Private Partnerships at Scale: Co-develop indigenous AI defense platforms under Digital India and Make in India.
- Global Leadership Role: Position India in international AI governance forums (e.g., G20, UN, QUAD cyber initiatives).
- Next-Gen CERT-In: Transition CERT-In into a global AI cyber defense center, with predictive capabilities and exportable Indian technologies

AI tools and technologies used by cyber criminals

Cybercriminals are now actively using a range of specific AI tools and techniques in recent cyberattacks. Here is a breakdown of the major categories and example tools:

1. Malicious Generative AI & Large Language Models (LLMs)

WormGPT: An illicit chatbot built on GPT-J (an open-source equivalent to GPT-3) designed for writing convincing phishing emails and Business Email Compromise (BEC) attacks. It is available on dark web marketplaces and bypasses protections found in legitimate AI models.

FraudGPT: Purpose-built for cybercriminals, this tool promises features for generating malicious code, scam websites, undetectable malware, phishing pages, and more. It is subscription-based and designed to provide "no boundaries," making it easy for scammers to tailor attacks at scale.

DarkBard: Another malicious LLM mentioned in underground forums, focused on automated social engineering and phishing.

Custom Jailbroken Variants of ChatGPT and Similar Models:Attackers frequently jailbreak existing popular models (like ChatGPT, GPT-4, LLaMA, or Mixtral) to remove guardrails and enable malicious outputs, including personalized phishing or malware code.

2.Advanced Deepfake Generators

- Voice and Video Deepfake Apps: Tools that synthesize a specific spokesperson's voice or mimic their appearance in real-time video calls, used for high-value BEC scams or "virtual kidnapping" and "digital arrest" frauds.

- AI-Powered Media Synthesis: Attackers now stitch together realistic audio, images, and videos scraping from social media to impersonate close relatives or trusted colleagues, frequently leveraging advanced open-source and commercial models.

3.Malware Generation and Evasion Bots

AI-Enabled Malware Engines: Cybercriminals use models trained on malware repositories to write and mutate malicious code, nurture polymorphic (shape-shifting) malware, and create new ransomware variants that adapt to evade detection and signature-based antivirus.

AI Code Bots: Some use open-source LLMs (fine-tuned on malware datasets) to generate reverse shells, logic bombs, or backdoors that change with every deployment.ies used by cyber criminals.

4.Conversational AI for Scams

Chatbots & Virtual Agents: Fraudsters deploy AI chatbots to automate phishing and scam interactions, making social engineering scalable and convincing by mimicking human conversation and emotional responses in real-time

AI Tool/Technique	Attack Purpose	Example/Details
WormGPT/FraudGPT	Phishing, malware, BEC	Custom LLMs for dark web
Deepfake Generators	Social engineering, BEC, fraud	Executive impersonation
Jailbroken LLMs	Phishing, malware	Unrestricted ChatGPT, GPT-4
AI OSINT & Profiling	Automated reconnaissance	Mapping targets from social/OSINT
AI-enabling Malware Bots	Polymorphic malware	Adaptive code for evasion
AI Chatbots	Conversational scam automation	“Human-like” scam chat

5.Automated reconnaissance and profiling

Cybercriminals leverage AI for automated reconnaissance and profiling by using machine learning and generative AI tools to rapidly collect, analyze, and exploit vast amounts of publicly available and internal data. Here's how these tactics work in practice:

5.1 **Automated Data Collection**

AI-powered OSINT Tools: Cybercriminals deploy AI models to crawl and aggregate data from platforms like LinkedIn, Facebook, Twitter, company websites, GitHub, and exposed databases. These tools automate what used to be manual, combing through millions of profiles, posts, employee lists, and credential leaks in seconds.

Vulnerability Mapping: AI scours hacker forums, dark web leaks, and asset search engines (like Shodan or Censys) to identify targets with vulnerable applications, outdated software, low security hygiene, or exposed infrastructure.

5.2 **Target Profiling and Prioritization**

Machine learning models analyse OSINT data to score and rank targets (individuals or organizations) based on their job roles, access privileges, online behaviour, and relationships. For instance, C-level executives, finance staff, or system admins get flagged as "high-value" for phishing or BEC (business email compromise) attacks.

Behavioral Analysis: AI tools can monitor how organizations and people interact online detecting habits, routines, language style, travel plans, corporate events, and more which enables the generation of highly personalized phishing messages or social engineering lures.

5.3 **Automated Attack Blueprint Development**

Social Graph Building: AI systems model internal and external networks—mapping who reports to whom, frequent collaborators, and trusted partners—so cybercriminals can impersonate or exploit trusted relationships for success in BEC or insider threats.

Attack Simulation: Some tools simulate phishing campaigns or intrusion methods virtually, learning the fastest paths to compromise by analyzing target defences and employee behavior using AI-powered red teaming.

5.4 **Scaling and Speed**

Unprecedented Scale: AI allows cybercriminals to simultaneously profile thousands or millions of potential victims, rather than focusing on just a handful, massively expanding their attack surface and probability of success.

Real-Time Adaptation: As victims change passwords, move departments, or change organizations, AI-enabled bots autonomously update target profiles and adapt attack paths without manual intervention

6.Adversarial AI methods used by criminals

Adversarial Examples and Input Manipulation

Attackers craft inputs designed to fool AI-based detection systems (like antivirus, intrusion detection, or spam filters) by subtly altering malicious payloads so they appear benign to AI classifiers. This includes adding noise or perturbations that confuse AI models without disrupting the malware's function.

Polymorphic and Metamorphic Malware Generation

AI tools automatically generate malware variants that continuously change their code structure, signatures, and behavior to bypass signature-based and heuristic detection. Polymorphic malware morphs its code every time it propagates, while metamorphic malware rewrites its code to evade sandboxing and AI-evaluated static analysis.

Model Poisoning and Data Manipulation

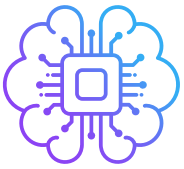
Attackers intentionally feed poisoned data into AI models training or operational data streams, causing those AI models to misclassify malicious activity as safe. This weakens detection over time and leads to blind spots in cybersecurity systems.

Evasion of Behavioural Detection

AI-crafted payloads mimic legitimate user or software behaviours, making it harder for anomaly detection systems to distinguish between normal and malicious activities. Attackers may time attacks or throttle actions to remain under statistical thresholds monitored by AI.

AI-Powered Automation to Test Defences

Attackers use AI bots to continuously probe networks with diverse attack vectors generated and adapted in real-time, learning which attempts succeed and modifying methods on-the-fly to evade AI-driven defences.



Role of Agentic AI in Cyber Security

1. Autonomous Threat Hunting

What it does: Instead of waiting for analysts to query logs, agentic AI actively scans across endpoints, networks, and cloud environments for suspicious activity.

Why it matters: Detects advanced persistent threats (APTs) and stealthy attackers faster than human SOC teams.

Example: An AI agent could autonomously correlate login anomalies, phishing attempts, and malware signatures, then flag a likely coordinated attack.

2. Adaptive Incident Response

What it does: Agentic AI doesn't just raise an alert it can take real-time actions (e.g., isolate infected endpoints, block IPs, revoke credentials).

Why it matters: Reduces mean-time-to-response (MTTR) from hours/days to seconds/minutes.

Example: A compromised IoT camera in a smart city grid could be automatically quarantined by an AI agent.

3. Dynamic Red Teaming & Penetration Testing

What it does: Agentic AI can simulate adaptive attackers probing defenses, changing strategies, even generating phishing emails or exploit payloads.

Why it matters: Provides a more realistic stress test of cyber defenses compared to static scripts.

Example: DARPA's AI Cyber Challenge (AIxCC) is funding research into autonomous "AI hackers."

4. Continuous Policy Enforcement

What it does: AI agents monitor compliance with cybersecurity policies (e.g., access controls, data protection rules) across all systems.

Why it matters: Prevents accidental insider leaks and enforces zero trust architectures without constant manual oversight.

Example: An agent can automatically revoke excessive user privileges or flag data transfers outside approved geographies.

5. Deception & Counter-Intelligence

What it does: Deploys AI agents as active defenders in deception environments (honeypots, honeyfiles, fake credentials).

Why it matters: Confuses attackers, collects intelligence on their tools, and buys time for defenders.

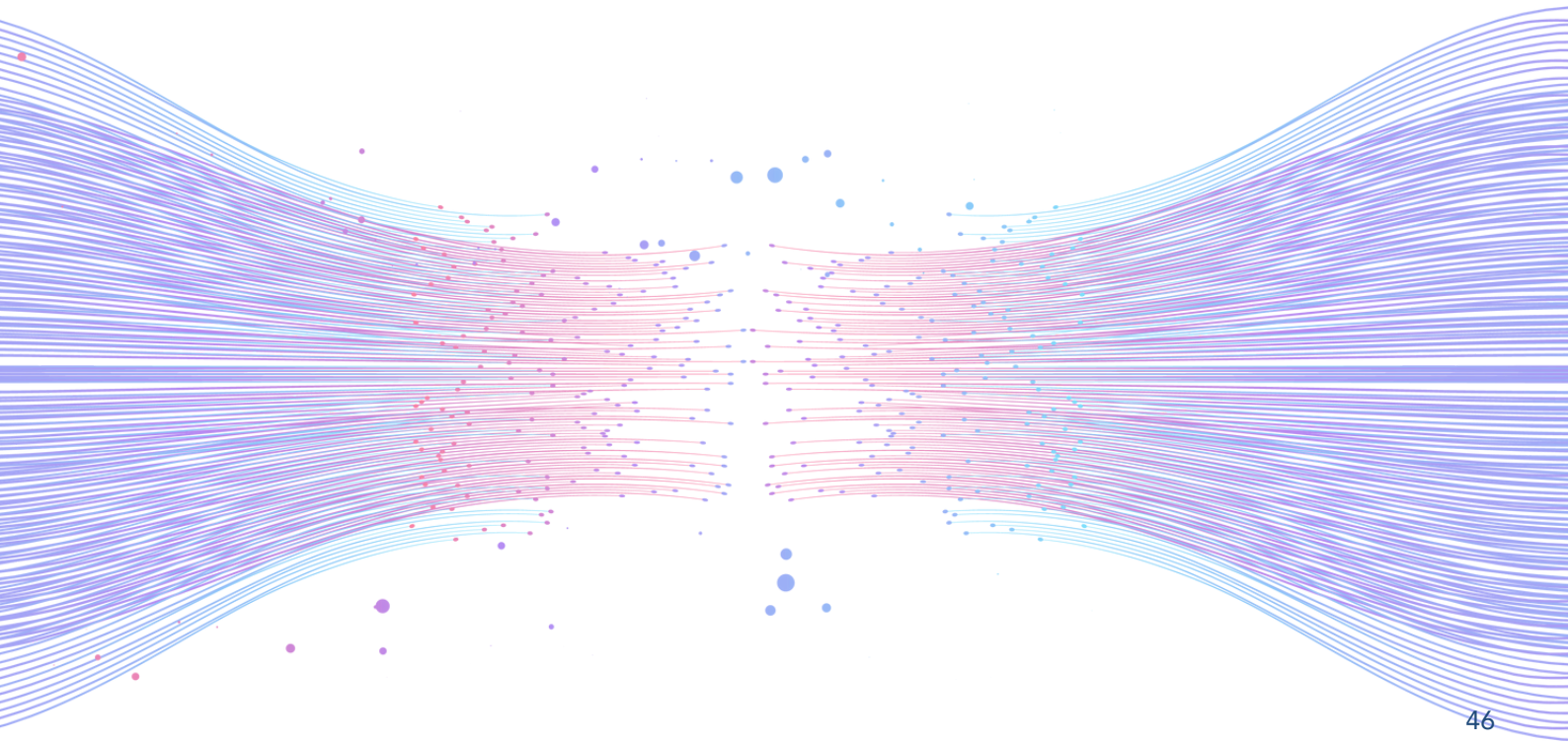
Example: An AI agent could "chat back" to adversaries inside a honeypot, learning their techniques.

6. AI vs. AI Cyber Battles

What it does: As attackers also adopt AI, agentic AI will be used for autonomous cyber defense duels (attack AI vs. defense AI).

Why it matters: Human-only response will be too slow; agentic AI provides machine-speed defense.

Example: Detecting and countering AI-generated phishing campaigns in real time.



Conclusion & Call to Action

AI's role in cybersecurity will only grow deeper. Criminals will continue leveraging AI for faster, harder-to-detect attacks, while defenders must harness AI to stay ahead. The way forward is multi-fold and urgent. Organizations that serve customers and citizens must integrate AI into their security operations using machine learning for anomaly detection, automated incident response, and predictive risk analysis while maintaining strong governance and human oversight. Industry and Government should collaborate to share threat intelligence and AI safeguards.

Policymakers must likewise act decisively. India's regulators have set important benchmarks with laws against deepfakes (IT Act, DPDP, etc.) and new initiatives like the IndiaAI Safety Institute. Expanding these efforts for example by mandating transparency standards for AI-generated content or funding cybersecurity AI research will help national resilience. On the global stage, India should engage with international AI norms (e.g. EU's AI Act, UN AI governance talks) to ensure cross-border security.

Karnataka, as one of India's leading technology and startup hubs, is uniquely placed to confront this challenge. The state must frame AI not as an abstract concept, but as a set of applied technologies with both opportunities and risks that demand foresight and responsibility. AI holds immense potential to improve governance. Predictive analytics can help identify irregularities in transactional data involving citizens' welfare distribution and also safeguard data. Natural Language Processing (NLP) can support multilingual digital platforms that bring citizen services to rural areas in Kannada and other regional languages. For Karnataka, the lesson is clear. While drawing from global examples, the state must craft an AI framework suited to India's democratic and federal context, balancing innovation with rights and accountability.

Above all, public-private collaboration is key. Technology companies, financial institutions, critical infrastructure operators and law enforcement must cooperate to implement frameworks to build secure AI tools. Civil society and educational institutions should raise awareness and training at all levels. The stakes are high, if unchecked, AI-driven cyber-attacks could undermine public trust and economic stability; but if guided properly, AI can become a powerful ally in creating a safer digital world. The time to act is now by integrating AI responsibly, investing in defences, and enforcing ethical standards, we can realize AI's potential as a tool for security and innovation, rather than let it be exploited as a weapon.

Cogenz Cybertech

At Cogenz , we make cybersecurity simple, effective, and accessible for businesses of all sizes. Our goal - To protect what matters most-your data, your operations, and your peace of mind. Cogenz Cybertech is a cybersecurity solutions & services provider based in Bengaluru, India dedicated to helping businesses protect their digital assets and stay ahead of threats. We specialize in delivering comprehensive security strategies that empower organizations to safeguard their infrastructure, ensure compliance, and mitigate risks effectively.

With a strong focus on innovation, reliability, and expertise, we work closely with businesses of all sizes to implement tailored security solutions that address modern cybersecurity challenges. Our team of professionals is committed to providing top-tier protection, strategic insights, and proactive support to help organizations build a secure and resilient future.

At Cogenz, we believe that cybersecurity is more than just defense it's about enabling businesses to operate with confidence in an increasingly digital world.

Visit us: www.cogenz.in

Sources Cited

- Times of India coverage of the GIREM–Tekion "State of AI-Powered Cybercrime: Threat & Mitigation Report 2025," national metrics and phishing with AI (80%); India's complaint volumes; "digital arrest" losses.
- Times of India Bengaluru data on Karnataka losses and prevalence of AI-enabled tactics (phishing, site cloning, deepfakes).
- Fortinet–IDC survey findings on 72% Indian firms targeted; readiness and detection gaps; tactic mix (credential stuffing, deepfakes, AI phishing, polymorphic malware).
- World Economic Forum Global Cybersecurity Outlook 2025 and industry statistics context on AI-enhanced tactics and rising sophistication/frequency.
- Case study detailing Arup/Hong Kong deepfake BEC via multi-person video call leading to ~US$25.6M loss and methodology.

References:

- Eviden (Atos Group), "Is AI our greatest ally or our worst enemy in the fight against cyber threats?" (June 12, 2024)eviden.com.
- Economic Times (BFSI), "RBI's FREE-AI Framework: Pioneering Responsible AI in Finance" (Aug. 14, 2025)bfsi.economictimes.indiatimes.combfsi.economictimes.indiatimes.com.
- Press Information Bureau (Govt. of India), "India well-equipped to tackle evolving online harms and cyber crimes; Government to Parliament" (Aug. 8, 2025)pib.gov.inpib.gov.in.
- Economic Times (CFO), "AI-driven cyber attacks surge in India and APAC, warns cyber risk report" (Aug. 6, 2025)cfo.economictimes.indiatimes.comcfo.economictimes.indiatimes.com.
- Brennan Center for Justice, "Gauging the AI Threat to Free and Fair Elections" (May 9, 2024)brennancenter.orgbrennancenter.org.
- IndiaAI (Digital India), "Call for Partnerships as part of the IndiaAI Safety Institute" (May 9, 2025)indiaai.gov.inindiaai.gov.in.
- CERT-In (2023). Annual Cyber Security Report 2023. Government of India, Ministry of Electronics and IT (MeitY).
- https://www.cert-in.org.in
- National Cyber Security Strategy 2020 (Draft). National Security Council Secretariat, Government of India.
- MeitY (2024). Digital India Act – Consultation Papers. Ministry of Electronics and IT.
- NITI Aayog (2021). Responsible AI for All: Approach Document. New Delhi.
- World Economic Forum (2023). Global Cybersecurity Outlook 2023.
- https://www.weforum.org/reports/global-cybersecurity-outlook-2023
- MITRE (2022). Adversarial Machine Learning Threat Matrix. MITRE Engenuity.
- Gartner (2023). Emerging Technologies: AI in Security Operations. Gartner Research Note.
- European Union (2024). EU AI Act – Regulatory Framework on Artificial Intelligence. European Commission.
- NIST (2023). AI Risk Management Framework (AI RMF 1.0). U.S. National Institute of Standards and Technology.
- McAfee Labs (2024). The Cybercrime Economy Report. McAfee Inc.
- Kaspersky (2024). The Rise of Deepfakes: Global and India Threat Landscape. Kaspersky Labs.
- Brookings Institution (2023). Generative AI and the Future of Cybersecurity. Washington, DC.
- PwC India (2023). Artificial Intelligence and Cybersecurity in India: Trends and Implications.
- Carnegie Endowment for International Peace (2022). AI and Cybersecurity: Emerging Global Challenges.
- ISACA (2023). State of Cybersecurity 2023 Report. ISACA Global.