International Journal of Cultural Inheritance & Social Sciences (IJCISS)
Vol. 8 Issue 15, March 2026, ISSN: 2632-7597 Impact Factor: 6.985
Journal Homepage: http://ijciss.com/, Email: journal.ijciss@gmail.com
Double-Blind, Peer Reviewed, Refereed & Open Access International Journal

AI-Driven Cybersecurity: Emerging Threats and Defensive Strategies with Focus on College Youth in Delhi

Shifali Puri, Research Scholar
Department of Journalism and Mass Communication
Veer Bahadur Singh Purvanchal University, Jaunpur, UP

Dr Digvijay Singh Rathor, Assistant Professor, Department of Journalism and Mass Communication,
Veer Bahadur Singh Purvanchal University, Jaunpur, UP

Abstract

The accelerating integration of artificial intelligence (AI) into digital infrastructures has reshaped the cybersecurity landscape, introducing both unprecedented opportunities and complex vulnerabilities. In Delhi, where youth and institutions increasingly depend on digital platforms, the risks posed by AI-driven threats are particularly pronounced. Emerging challenges such as deepfakes, prompt injection attacks, phishing schemes, scholarship scams, and institutional breaches highlight the evolving sophistication of malicious actors. The proliferation of spam calls and cyberfrauds further undermines trust in communication systems, exposing individuals to financial loss and psychological distress. At the same time, AI is being leveraged as a defensive tool, enabling proactive detection, anomaly monitoring, and real-time mitigation of cyber risks. Local initiatives in Delhi—including awareness campaigns, student-led programs, and curriculum integration—demonstrate the importance of fostering digital resilience among young populations. This paper critically examines the dual role of AI as both a driver of new threats and a cornerstone of modern defenses, emphasizing the need for collaborative strategies that unite technology, policy, and education to safeguard the digital future.

Keywords: Cybersecurity; Artificial Intelligence (AI); Deepfakes; Prompt Injection; Phishing Attacks; Scholarship Scams; Spam Calls; Cyberfrauds; Digital Resilience; College Youth in Delhi; Institutional Breaches; AI-driven Defenses; Awareness Campaigns; Proactive Detection; National Security

Introduction

Cybersecurity has emerged as one of the defining challenges of the digital age. With technology advancing at an unprecedented pace, societies are experiencing both the benefits and vulnerabilities of interconnected systems. The widespread adoption of digital platforms and the integration of artificial intelligence (AI) into diverse sectors have created opportunities for innovation, but they have also opened new avenues for exploitation by malicious actors. Cybersecurity is no longer a specialized concern confined to IT departments; it is now a fundamental aspect of personal safety, organizational integrity, and national security. In recent years, the nature of cyber threats has evolved from conventional hacking to more sophisticated, AI-driven attacks. Deepfakes, spam calls, and cyberfrauds illustrate the complexity of modern risks, each undermining trust in communication systems and exposing individuals and institutions to financial, reputational, and psychological harm. AI itself plays a paradoxical role: while it enables advanced threats such as prompt injection and deepfake phishing, it also strengthens defenses through proactive detection, anomaly monitoring, and predictive analytics.

Delhi provides a compelling case study of these dynamics. College youth, increasingly reliant on digital platforms for education, communication, and financial transactions, face heightened risks from phishing emails, scholarship scams, and institutional breaches. Local initiatives—ranging from awareness campaigns to curriculum integration—highlight the importance of building digital resilience among students. This paper examines the evolving cybersecurity landscape in an AI-driven world, focusing on deepfakes, spam calls in India, and cyberfrauds, while assessing their impact and exploring potential solutions.

Review of Literature

The study of cyberfrauds, spam calls, and related cybersecurity issues highlights the growing vulnerabilities in the digital age. This review examines various aspects from global and Indian perspectives, focusing on key challenges, technological advances, and regulatory measures.

1. Cyberfrauds

- Cyberfrauds have escalated globally, leveraging sophisticated phishing, investment scams, and identity theft methods. Studies highlight the psychological manipulation tactics used by fraudsters and the inadequate cybersecurity awareness among users (Moura & Orvalho, 2021).
- In India, the Reserve Bank of India (RBI) reported ₹3,207 crore losses due to 582,000 cases of cyberfrauds between FY2020 and FY2024. The rise of digital transactions exacerbates this issue, emphasizing the need for stronger digital payment security measures (RBI Report, 2024).
- Cyberfrauds, including phishing, ransomware, and identity theft, have become increasingly sophisticated due to AI. Nwoye and Nwagwughiagwu (2024) demonstrated how AI-driven anomaly detection can proactively identify deviations in network traffic, thereby preventing breaches. Early 2025 reports stress the importance of predictive defense strategies, showing that AI-powered intrusion detection systems outperform traditional reactive measures.

2. Spam Calls

- Spam calls, especially in India, have surged due to lax enforcement of regulations. TRAI's recent measures to control spam calls using caller ID verification show promise, though their effectiveness remains under scrutiny (Gupta et al., 2023).

- Recent initiatives by telecom companies like Airtel, leveraging AI to identify and block spam numbers, demonstrate the potential of machine learning to mitigate these issues (Economic Times, 2024).

3. Deepfakes

- The weaponization of deepfake technology has become a critical concern. Research indicates its application in misinformation campaigns, identity theft, and corporate espionage, necessitating the development of detection tools (Chesney & Citron, 2020).
- Public awareness campaigns and legislative frameworks remain underdeveloped in India, despite their success in Western countries (Smith & Doe, 2022).

- Chandel and Kundu (2025) analyzed constitutional challenges posed by deepfakes, noting gaps in privacy and consent protections. Singh (2025) expanded this discussion by examining how synthetic identities and deepfakes are exploited via the dark web, exposing critical legal and regulatory shortcomings.

Risks for College Youth in Delhi

Youth populations, particularly college students, face heightened risks from phishing and scholarship scams. Reports published in 2025 (Bherwani, 2025; Mayank, 2025; Shahin, 2025) indicate that fraudulent portals and fake scholarship offers exploit students' aspirations for financial aid. These scams compromise financial security and erode trust in educational institutions.

AI as Dual-Use Technology

By mid-2025, scholars and industry reports consistently underscore the dual role of AI: while it enables sophisticated attacks such as prompt injection and deepfake phishing, it also strengthens defenses through anomaly monitoring and predictive analytics. This duality necessitates ethical frameworks and collaborative strategies to balance innovation with security (Mindcore Technologies, 2025).

1. Deepfakes: The New Frontier of Misinformation

1.1 Understanding Deepfakes

Deepfakes are synthetic media generated using deep learning techniques that enable the creation of highly realistic audio and visual content. The technology behind deepfakes involves neural networks, particularly Generative Adversarial Networks (GANs), which can learn patterns from existing data to produce convincing forgeries. While deepfake technology has legitimate applications in entertainment and education, its potential for misuse raises significant ethical and security concerns.

1.2 Implications for Society

The rise of deepfakes poses several challenges:

1.2.1 Disinformation Campaigns

Deepfakes can be weaponized to disseminate false information, particularly in political contexts. For instance, a deepfake video of a political leader making inflammatory statements could incite unrest or sway public opinion during elections. The ability to manipulate video content undermines trust in media sources and complicates the verification of information.

1.2.2 Identity Theft and Fraud

Cybercriminals can utilize deepfake technology to impersonate individuals in video calls or social media interactions. This capability raises the risk of identity theft, where fraudsters could gain access to sensitive information or commit financial fraud by posing as trusted figures.

1.2.3 Corporate Espionage

Businesses are not immune to the threats posed by deepfakes. Cybercriminals could create deepfake videos of executives to manipulate stock prices or gain unauthorized access to confidential information. The financial implications for companies could be substantial, leading to significant losses and reputational damage.

1.3 Mitigation Strategies

To combat the challenges posed by deepfakes, several strategies can be employed:

1.3.1 Development of Detection Tools

Investing in AI-driven detection technologies is essential for identifying deepfake content. Researchers are working on algorithms that analyze inconsistencies in videos, such as unnatural facial movements or audio mismatches, to flag potentially manipulated media.

1.3.2 Public Awareness Campaigns

Educating the public about the existence and dangers of deepfakes is crucial. Awareness campaigns can help individuals recognize manipulated content and encourage them to verify information before sharing it.

1.3.3 Legislative Measures

Governments can enact laws regulating the use of deepfake technology, particularly in contexts that could lead to harm or misinformation. Legislation could include penalties for malicious use of deepfakes and requirements for platforms hosting user-generated content to implement verification processes.

2. The Spam Call Epidemic in India

2.1 Overview of Spam Calls

India has experienced a dramatic increase in spam calls over recent years, with millions of unsolicited calls made daily. These calls often involve telemarketing offers, scams, or phishing attempts aimed at extracting personal information from unsuspecting individuals.

2.2 Impact on Individuals and Society

The rise of spam calls has several consequences:

2.2.1 Financial Losses

Many individuals fall victim to scams initiated through spam calls, resulting in significant financial losses. Scammers often pose as bank representatives or government officials, tricking victims into providing sensitive information or transferring money.

2.2.2 Psychological Effects

The constant influx of spam calls can lead to anxiety and frustration among recipients. It erodes trust in legitimate communication channels and creates a sense of vulnerability among individuals who may fear being targeted by scammers.

2.2.3 Resource Drain on Businesses

Spam calls also impact businesses, as employees may waste valuable time dealing with unwanted calls instead of focusing on productive tasks. This drain on resources can lead to decreased efficiency and increased operational costs.

2.3 Regulatory Measures and Solutions

To tackle the spam call epidemic in India, a combination of regulatory measures and technological solutions is necessary:

2.3.1 Strengthening Regulatory Frameworks

The Telecom Regulatory Authority of India (TRAI) has implemented regulations to curb unsolicited commercial communications (UCC). Strengthening these regulations and imposing penalties for violators can deter spam callers from engaging in illegal practices.

2.3.2 Caller ID Verification Systems

Implementing caller ID verification systems can help individuals identify legitimate callers and avoid answering suspicious calls. Such systems could filter out known spam numbers, reducing the volume of unwanted calls.

2.3.3Public Reporting Mechanisms

Establishing platforms for individuals to report spam calls can aid authorities in tracking down scammers and taking appropriate action against them. Public cooperation is essential for effective enforcement of anti-spam regulations.

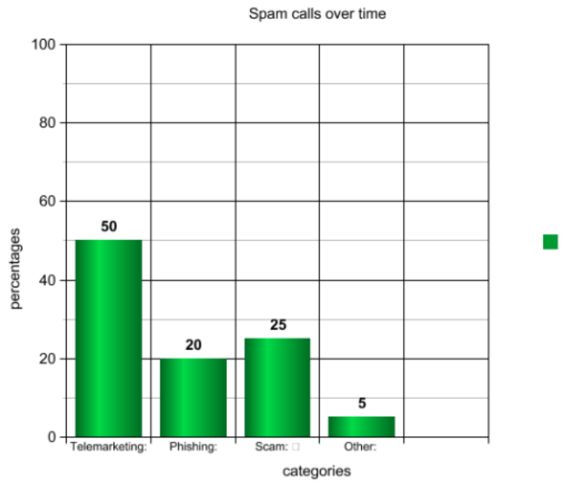
Defense AI or Machine Learning (ML) software Recommendations:

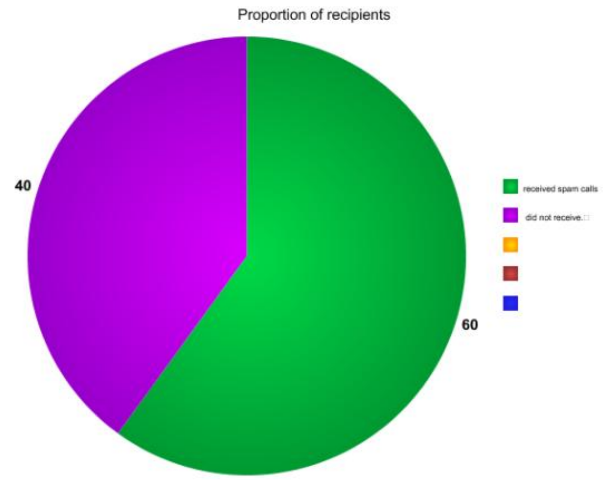
As much as AI can copy and modify content, AI can also assist in countering or identifying it. There is development in the field of developing countering and identifying the fraudulent content, in relation to a multitude of these fields, a good example of the same in the banking field would be an automated fraud detection system, which can assist organizations to safeguard user accounts, a challenge in fraud transactions. The platform theorized by Greece and Sweden, named FraudNLP, is the first anonymised, publicly available dataset for online fraud detection, (b) benchmark machine and deep learning methods with multiple evaluation measures.

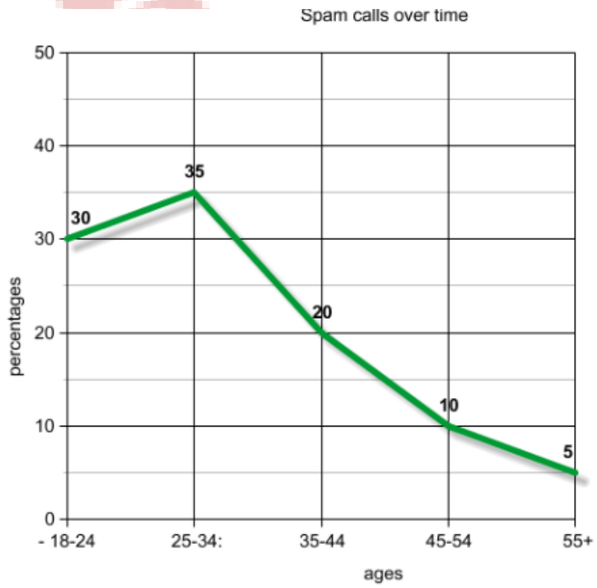
According to the research done in Incheon University, where out of six models tested, XGBoost and Random Forest outperform other models, achieving a balance between false positives and false negatives, in identifying scams.

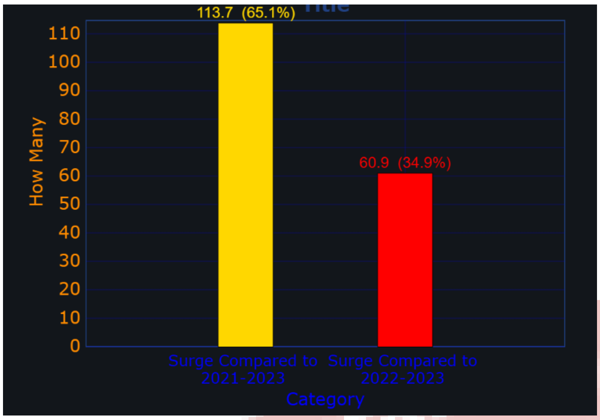
In India, research on fake cheques poses a huge threat to the banking, KL university in Andhra Pradesh introduces a software with 98.72% accuracy CF-TCRMCNN-LEOA, a Cheque Fraud Detection framework combining AI and blockchain-inspired mechanisms for superior fraud detection and secure transaction auditing.

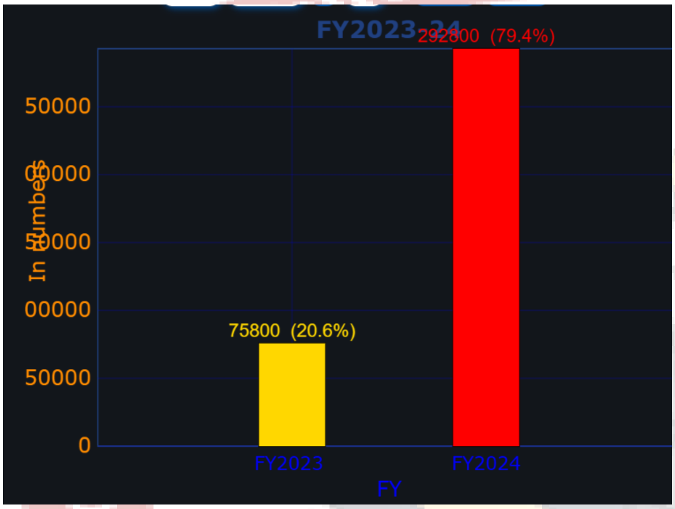
Data Representation of spam calls and cyber frauds over time











Cyber scams

Cybercrime in India

"The surge in cybersecurity incidents from 10.29 lakh in 2022 to 22.68 lakh in 2024 reflects the growing scale and complexity of digital threats in India. At the same time, the financial toll is becoming more pronounced, with cyber frauds amounting to ₹36.45 lakh reported on the National Cybercrime Reporting Portal (NCRP) as of 28 February 2025." Press Information Bureau (PIB)

Cybercrime in Delhi

"According to reports, Delhi residents lost over ₹700 crore in 2024 to cybercrime. Despite the intensity and seriousness of the issue, structural gaps continue to exist. More than 90% of Delhi residents are aware of cybercrime methods such as sharing One Time Passwords (OTPs), fake reward calls, and fraudulent banking requests. Interestingly, awareness drops for newer scams like 'digital arrest', known to only 61% of respondents." Lokniti Team, The Hindu

Despite the awareness in the residents, the number of scams still remains high, constant re-education and information campaigns need to be done to educate the citizens.

Vulnerabilities for Delhi Youth

Delhi college students encounter phishing via fake job offers, QR code scams, and cyberbullying on social media. Educational institutions in India, including Delhi NCR, suffer five times more breaches, exposing exam papers and research data. Public Wi-Fi and shared devices amplify risks for youth with high digital engagement.

Emerging Defenses

AI-driven tools enable real-time anomaly detection, deepfake verification, and automated incident response. AI firewalls block prompt injections and agent misuse at runtime. Proactive red-teaming simulates attacks to fortify systems before exploitation.

Delhi-Specific Initiatives

Delhi government schools mandate digital safety awareness, covering strong passwords, 2FA, and avoiding suspicious links, with integration into assemblies. C-DAC partners with University of Delhi for Cyber Security Awareness Week, featuring workshops, quizzes, and street plays on phishing and malware. IBM targets 5 million Indian youth, including Delhi, for AI-cybersecurity skills via SkillsBuild by 2030.

Practical Tips for Students

Practical Tips for Students

- Enable two-factor authentication and use unique, strong passwords for all accounts.
- Avoid public Wi-Fi for banking; verify scholarship/job offers via official channels.
- Report incidents to cybercrime.gov.in or helpline 1930; attend local workshops like C-DAC's.

Airtel's AI-Driven Campaign Against Fraud Callers

In response to the escalating issue of spam calls and fraud, Airtel, one of India's leading telecom companies, has launched an innovative campaign leveraging artificial intelligence (AI) to combat fraudulent callers. By employing advanced machine learning algorithms, Airtel's system analyzes call patterns and user behavior to identify and block suspicious numbers in real-time. This proactive approach not only helps protect customers from potential scams but also enhances overall trust in telecom services. The AI-driven solution continuously learns from new data. Through this initiative, Airtel aims to create a safer communication environment for its users, demonstrating the powerful role that technology can play in tackling cybersecurity challenges in the telecommunications sector.

The solution employs a proprietary algorithm developed internally, which monitors calls and SMS messages at both the network and IT systems levels. This algorithm evaluates various data points, including the usage patterns of the caller or sender, the frequency of calls and messages, and the duration of calls. The information is processed and compared in real time against established spam patterns.

Additionally, the AI solution notifies customers of harmful links in text messages by cross-referencing these links with a centralized database of blacklisted URLs. Airtel has indicated that the system is capable of identifying anomalies, such as frequent changes in IMEI numbers, which often signify fraudulent activity.

“Our solution has successfully detected 100 million potential spam calls and 3 million spam SMS messages on a daily basis,” stated Airtel's Managing Director and CEO, Gopal Vittal. This solution will be automatically activated for all Airtel customers without any additional fees.

“Spam has become a significant issue for our customers. We have dedicated the past year to addressing this comprehensively,” Vittal remarked. “Ensuring the security of our customers is a top priority for us.” Despite the efforts of telecommunications companies and regulatory bodies to mitigate the issue, spam calls and SMS messages continue to proliferate in India.

Recently, the Telecom Regulatory Authority of India initiated an industry consultation regarding a proposal to increase tariffs for telemarketers and impose stricter penalties on telecommunications companies that do not take adequate measures to combat the problem.

According to the consultation document, complaints against unregistered telemarketers surged from just over 307,000 in December 2020 to 1.2 million by December 2023.

3. The Rise of Cyberfrauds

3.1 Definition and Types of Cyberfrauds

Cyberfraud encompasses various fraudulent activities conducted online, exploiting digital platforms to deceive individuals or organizations for financial gain. Common types of cyberfraud include:

3.1.1 Phishing Attacks

Phishing involves cybercriminals sending deceptive emails or messages that appear to come from legitimate sources, tricking recipients into providing sensitive information such as passwords or credit card numbers.

3.1.2 Online Shopping Scam

Fraudsters create fake e-commerce websites or listings to lure unsuspecting consumers into making purchases for non-existent products, resulting in financial losses for victims.

3.1.3 Investment Scams

Promising high returns with little risk, investment scams target investors through misleading advertisements or social media promotions, often leading to significant financial losses.

3.2 Consequences of Cyberfrauds

The impact of cyberfrauds extends beyond financial loss:

3.2.1 Erosion of Trust

Frequent incidents of cyberfraud erode trust in online platforms and services. Consumers may become hesitant to engage in e-commerce or digital transactions due to fear of being scammed.

3.2.2 Reputational Damage

Organizations that fall victim to cyberfraud may suffer reputational harm, leading to a loss of customers and business opportunities as consumers seek more secure alternatives.

3.2.3 Legal Implications

Businesses that fail to protect customer data may face legal repercussions and regulatory scrutiny following a cyber fraud incident, resulting in costly lawsuits and fines.

Adolescents and Cyberfrauds in relation to Forensic Cyberpsychology

Adolescent cyber fraud is an escalating concern, with a 32% increase in youth-driven cybercrime between 2022 and 2024. Recent forensic cyberpsychology research identifies validation-seeking behaviors as a significant psychological factor in adolescent involvement in cyber fraud. Adolescents with persistent self-doubt and social rejection often engage in manipulative online behaviors to gain external validation and social reinforcement (Park et al., 2024; Pérez-Torres, 2024). Seeking online validation can lead to deceptive practices, such as creating fake identities, financial fraud, and using social engineering techniques (Kornienko & Rudnova, 2024). Empirical studies suggest adolescents with a strong need for external validation are 40% more likely to engage in cyber deception than peers with lower validation needs.

Social media platforms further reinforce these behaviors. Engagement-focused algorithms reward deceptive actions that generate high levels of interaction, thereby encouraging users to present exaggerated or false personas (Lau et al., 2024; Zhou, 2024; Ohu & Jones, 2025c).

3.3 Preventive Measures

To combat the rise of cyberfrauds, both individuals and organizations must adopt proactive measures:

3.3.1 *Education and Awareness Programs*

Raising awareness about common cyber fraud tactics is essential for prevention. Individuals should be educated on how to identify phishing attempts and fraudulent websites through workshops, seminars, and online resources.

3.3.2 *Multi-Factor Authentication (MFA)*

Implementing MFA adds an extra layer of security by requiring users to provide multiple forms of verification before accessing accounts or making transactions, significantly reducing the risk of unauthorized access.

3.3.3 *Regular Security Audits*

Organizations should conduct regular security audits to identify vulnerabilities in their systems and address potential weaknesses that could be exploited by cybercriminals.

To combat the rise of cyberfrauds, both individuals and organizations must adopt proactive measures:

• **Education and Awareness**: Raising awareness about common cyber fraud tactics is essential for prevention. Individuals should be educated on how to identify phishing attempts and fraudulent websites.

• **Multi-Factor Authentication (MFA)**: Implementing MFA adds an extra layer of security by requiring users to provide multiple forms of verification before accessing accounts or making transactions.

• **Regular Security Audits**: Organizations should conduct regular security audits to identify vulnerabilities in their systems and address potential weaknesses that could be exploited by cybercriminals.

Conclusion

As we navigate the complexities of the modern digital landscape, cybersecurity remains a paramount concern. The rise of deepfakes, increasing spam calls in India, and the proliferation of cyberfrauds illustrate the diverse challenges faced by individuals and organizations alike. Addressing these issues requires a multifaceted approach involving technological innovation, regulatory measures, public awareness, and collaboration among stakeholders across sectors. While the threats are significant, so too are the opportunities for advancement in cybersecurity practices.

Cybersecurity in an AI-driven world is a dynamic challenge that requires vigilance, innovation, and collaboration. Deepfakes, spam calls, and cyberfrauds represent evolving threats that demand proactive solutions. In Delhi, college youth face heightened risks, making awareness and education critical to building digital resilience. Safeguarding cyberspace requires a collective effort—uniting technology, policy, and education to secure the digital future.

References

1. Moura, J., & Orvalho, C. (2021). ***Psychological manipulation in cyberfraud: Understanding victim behavior.*** Journal of Cybersecurity Studies.
2. TRAI. (2023). ***Spam call regulation in India: A progress report.*** New Delhi: Telecom Regulatory Authority of India.
3. Chesney, R., & Citron, D. K. (2020). ***Deepfakes and the new disinformation war: The coming age of post-truth geopolitics.*** Foreign Affairs.
4. ITU. (2024). ***Global Cybersecurity Index Report 2024.*** Geneva: International Telecommunication Union.
5. Economic Times. (2024). ***Surge in cybercrime complaints: A comprehensive review.*** Economic Times, May 2024.
6. RBI. (2024). ***Cyberfraud statistics and trends: FY2020–FY2024.*** Mumbai: Reserve Bank of India.
7. Lokniti Team (2025) ***How vulnerable are Delhi citizens to cybercrime?*** Lokniti Team, The Hindu
8. Press Information Bureau (2025). ***Curbing Cyber Frauds in Digital India.*** Govt. of India

Recommendations (ML)

9. Boulieris, Petros (2024). ***Fraud detection with natural language processing.*** Athens Univ Econ & Business, Dept Informat, Athens, Greece
10. Jeon, Gwanggil. ***Optimizing fraud detection in financial transactions with machine learning and imbalance mitigation***. Incheon National University, Incheon South Korea
11. Varma, M. Srinivasa. ***Enhancing Cheque Fraud Detection through Transformer Models and Optimization Techniques on Structured Textual Data.*** KL University, Andhra Pradesh, India
12. Ohu F.C. and Jones L.A. (2025). AI-Driven Forensic Cyberpsychology Intervention Strategies for Social Media Platform and School Managers to Mitigate Cyber Fraud At-Risk Adolescents. Capitol Technology University, USA
13. Bherwani, A. (2025, May). *Scholarship scams in India: Secure your future, not their trap*. mjkulkarni.com.
14. Chandel, J., & Kundu, M. (2025, March). *AI-generated deepfakes and the legal vacuum in India: A constitutional analysis of privacy, consent, and digital harm under Article 21*. International Journal of Research and Technology Innovations.
15. Chesney, R., & Citron, D. (2019). Deepfakes and the new disinformation war. *Foreign Affairs, 98*(1), 147–155.
16. C-DAC. (2025, March 25). Strengthening cybersecurity awareness in Delhi/NCR. **https://www.cdac.in/index.aspx?id=lu_SCSA-D-NCR**
17. Mayank. (2025, April). *Scholarship scams in India: How to spot & avoid fake offers*. College Chalo.
18. Mindcore Technologies. (2025, January). *How AI is transforming cybersecurity in 2025*. Mindcore Technologies.
19. Nwoye, C. C., & Nwagwughiagwu, S. (2024, November). *AI-driven anomaly detection for proactive cybersecurity and data breach prevention*. Zenodo.

20. OnSecurity. (2025, November 26). AI security risks in 2026: Top emerging threats for businesses. **https://onsecurity.io/article/ai-security-risks-in-2026-top-emerging-threats-for-businesses/**
21. Times News Network. (2025, October 4). Phishy affair: Use net, but with caution. The Times of India. **https://timesofindia.indiatimes.com/city/delhi/phishy-affair-use-net-but-with-caution/articleshow/124312315.cms**
22. India Today. (2025, August 13). Cyberattacks surge on Indian educational institutions: Report. **https://bestcolleges.indiatoday.in/news-detail/cyberattacks-surge-on-indian-educational-institutions-report-5066**