

Artificial Intelligence and Cybersecurity in Modern Information Systems

S. V. Raghavan

Research Scholar, Department of Information Systems, Indian Institute of Technology Madras
(IIT Madras), Chennai, India

Abstract

Modern information systems store, process and transmit the data on which businesses, governments and citizens depend, and they are exposed to an ever-growing number of cyber attacks. Artificial intelligence (AI) has become important in this area in two ways. It strengthens defence by detecting intrusions, malware and fraud faster than manual methods, and it creates new risks because attackers can use the same technology and because AI systems themselves can be manipulated or used without proper control. This paper examines the relationship between artificial intelligence and cybersecurity in modern information systems on the basis of secondary data from the Indian Computer Emergency Response Team (CERT-In), the annual Cost of a Data Breach reports of IBM Security and the Ponemon Institute, and published research. Three comparison tables are prepared: year-wise growth of cyber security incidents in India from 2019 to 2024, a comparison of global breach cost, breach lifecycle and the savings from security AI and automation in 2023, 2024 and 2025, and a comparison of breach outcomes between organisations that use AI extensively for defence and those exposed to unsanctioned or shadow AI. The analysis shows that the number of incidents tracked in India increased more than five times in five years, that organisations using AI and automation extensively saved between 1.8 and 2.2 million US dollars per breach and identified and contained breaches 80 to 108 days faster, and that shadow AI increases the cost and the exposure of sensitive data. The paper concludes that AI is a necessary but not a sufficient element of cyber defence, and that it must be combined with governance, secure design, skilled people and human oversight.

Keywords: Artificial Intelligence, Cybersecurity, Information Systems, Machine Learning, Intrusion Detection, Data Breach, Adversarial Attacks, Shadow AI

Introduction

An information system is an organised combination of hardware, software, data, networks, people and procedures that collects, stores, processes and distributes information for decision making and operations. Modern information systems are no longer confined to a single office. They run on cloud platforms, mobile devices, web applications and connected sensors, and they exchange data with partners and customers continuously. The security of such systems is traditionally described by the three objectives of confidentiality, integrity and availability, which mean that information should be accessible only to authorised persons, should remain accurate and unaltered, and should be available when needed (Whitman and Mattord 8; Stallings and Brown 22). Anderson explained that security failures arise not only from technical flaws but also from weak incentives, poor design and human error, so security has to be treated as a system-level problem (Anderson 3).

The threat landscape has widened with the spread of digital services. Attackers use phishing, ransomware, malware, credential theft, denial of service and exploitation of software vulnerabilities,

National Conference on Emerging Trends in Arts, Humanities and Multidisciplinary Research (NCETAHMR)

Organized by- Navyug College of Education, Hisar (16 Aug 2026)

In IJESH ISSN: 2250-3115

and they increasingly operate as organised and profit-seeking groups. In India, the national agency CERT-In tracks and responds to such incidents and has reported a sharp rise in the number of incidents in recent years (Ministry of Electronics and Information Technology, Lok Sabha Reply 2). The cost to organisations is also large, since the annual global study of IBM Security and the Ponemon Institute reported an average cost of a data breach of 4.88 million US dollars in 2024 (IBM Security and Ponemon Institute, 2024 Report 5).

Artificial intelligence in cybersecurity mostly means machine learning, which is the use of algorithms that learn patterns from data, and its subfield deep learning. Buczak and Guven surveyed data mining and machine learning methods for intrusion detection and described how misuse detection, which recognises known attack signatures, and anomaly detection, which recognises deviation from normal behaviour, can both be supported by learning algorithms (Buczak and Guven 1154). Shaukat and co-authors reviewed the use of machine learning in the last decade for malware detection, intrusion detection, spam and phishing detection and found that supervised and unsupervised techniques are widely used (Shaukat et al. 222311). Ferrag and co-authors compared deep learning approaches for intrusion detection on modern datasets and showed that deep models can achieve high detection accuracy (Ferrag et al. 5). Nguyen and Reddi reviewed deep reinforcement learning in cybersecurity and described its use in autonomous defence, in which a software agent learns the best response through interaction with the environment (Nguyen and Reddi 3780).

In practice, AI is used in security operations centres to correlate alerts, to prioritise incidents, to detect abnormal user behaviour, to filter malicious email, to analyse malware and to automate the first steps of response. Sarker, Furhad and Nowrozy described this as AI-driven cybersecurity, in which security intelligence is modelled from data and used for prediction and automated action (Sarker et al. 3). The advantage is speed and scale, because the volume of logs and alerts in a modern organisation is too large for human analysts alone.

The same technology creates risks. Goodfellow, Shlens and Szegedy showed that small and deliberately designed changes in the input can make a machine learning model give a wrong output with high confidence, which are known as adversarial examples (Goodfellow et al. 2). Papernot and co-authors demonstrated that an attacker with limited knowledge of a deep neural network can craft such inputs (Papernot et al. 373). Biggio and Roli reviewed ten years of research on adversarial machine learning and described evasion attacks at the time of use and poisoning attacks on the training data (Biggio and Roli 318). A security tool based on machine learning can therefore itself become a target.

AI also assists the attacker. Brundage and co-authors warned that the malicious use of artificial intelligence could expand existing threats, introduce new threats and change the character of attacks, for example through automated phishing and faster discovery of vulnerabilities (Brundage et al. 3). Kaloudi and Li surveyed AI-based cyber threats and found examples across malware, social engineering and attacks on systems (Kaloudi and Li 3). Taddeo, McCutcheon and Floridi described trusting artificial intelligence in cybersecurity as a double-edged sword, because the benefits depend on reliability, transparency and human control (Taddeo et al. 557). A newer risk is shadow AI, which

National Conference on Emerging Trends in Arts, Humanities and Multidisciplinary Research (NCETAHMR)

Organized by- Navyug College of Education, Hisar (16 Aug 2026)

In IJESH ISSN: 2250-3115

refers to employees using unsanctioned AI tools and applications with company data, outside the control of the security team (IBM Security and Ponemon Institute, 2025 Report 20).

Objectives

1. To explain the role of artificial intelligence in securing modern information systems.
2. To analyse the growth of cybersecurity incidents in India.
3. To compare the cost of data breaches and the benefits of AI-based cybersecurity over recent years.
4. To compare the outcomes of AI-enabled defence with those of unsanctioned AI use.

Review of Literature

The literature is reviewed under four heads: machine learning for detection, limits and adversarial vulnerability, malicious use of AI, and economics and governance.

Buczak and Guven provided a detailed survey of machine learning and data mining methods for intrusion detection and noted that the choice of method depends on the data and on the type of attack (Buczak and Guven 1160). Shaukat et al. surveyed machine learning techniques in cybersecurity across malware, intrusion and phishing detection (Shaukat et al. 222315). Dasgupta, Akhtar and Sen presented a comprehensive survey of machine learning in cybersecurity and discussed applications and open challenges (Dasgupta et al. 58). Ferrag et al. compared deep learning models for intrusion detection and found that performance varies across datasets and models (Ferrag et al. 10). Nguyen and Reddi covered reinforcement learning methods that allow adaptive defence (Nguyen and Reddi 3782).

Sommer and Paxson argued that machine learning for network intrusion detection faces difficulties which are different from those in other fields, namely the high cost of errors, the wide variety of normal traffic, the lack of good training data and the difficulty of explaining results, and that these explain why anomaly detection is less common in operational networks than in research (Sommer and Paxson 306). Apruzzese and colleagues examined the role of machine learning in cybersecurity and highlighted the gap between research results and real-world deployment (Apruzzese et al. 3). Goodfellow et al., Papernot et al. and Biggio and Roli established that learning-based models can be fooled or poisoned (Goodfellow et al. 2; Papernot et al. 373; Biggio and Roli 318).

Brundage et al. and Kaloudi and Li documented how AI can increase the scale, speed and success of attacks (Brundage et al. 3; Kaloudi and Li 3). Taddeo et al. concluded that the use of AI in cyber defence must be accompanied by safeguards because it can be exploited (Taddeo et al. 557).

Gordon and Loeb developed an economic model of investment in information security and showed that the optimal spending on protecting an information set should not exceed about thirty-seven per cent of the expected loss from a breach, which means that the most vulnerable assets are not always the ones that deserve the highest spending (Gordon and Loeb 439). The National Institute of Standards and Technology published a Cybersecurity Framework with five core functions, namely identify, protect, detect, respond and recover (National Institute of Standards and Technology, Cybersecurity Framework 6), and an AI Risk Management Framework with four functions, namely govern, map, measure and manage (National Institute of Standards and Technology, AI Risk

National Conference on Emerging Trends in Arts, Humanities and Multidisciplinary Research (NCETAHMR)

Organized by- Navyug College of Education, Hisar (16 Aug 2026)

In IJESH ISSN: 2250-3115

Management Framework 20). The annual breach reports provide empirical evidence on cost and on the effect of AI-based defence (IBM Security and Ponemon Institute, 2023 Report 5).

The literature shows that AI is useful for detection but not free from weaknesses. There are fewer studies which put the national incident data and the cost data together with the effect of AI on defence, and this paper tries to do so using published secondary data.

Research Methodology

The study is descriptive and analytical and uses only secondary data. Data on the number of cyber security incidents observed in India from 2019 to 2024 were taken from the replies of the Ministry of Electronics and Information Technology in the Lok Sabha, which are based on the information tracked by CERT-In. Data on global breach cost, breach lifecycle and the effect of security AI and automation for the years 2023, 2024 and 2025 were taken from the annual Cost of a Data Breach reports published by IBM Security and conducted by the Ponemon Institute. The conceptual background was collected from books, journals and standards listed in the references. Year-on-year growth, an index with 2019 as the base year, percentage change and the share of savings in the average cost were calculated for comparison. All figures are rounded and the latest values should be checked in the original reports before use.

The limitations of the study are as follows. CERT-In figures count incidents that are reported to or detected by the agency, and they are influenced by the extent of monitoring and the reporting rules, so they do not give the true number of attacks. The IBM and Ponemon studies are based on a sample of organisations that have suffered a breach and are not a representation of all organisations, and the reported savings from AI show association and not strict causation. The studies of different years may differ in sample and method. The paper does not use primary data from organisations.

Data Analysis and Discussion

Table 1 presents the number of cyber security incidents tracked by CERT-In from 2019 to 2024, the growth in each year and an index in which 2019 is taken as 100.

Table 1: Year-wise Cyber Security Incidents Tracked by CERT-In in India (2019 to 2024)

Year	Cyber Security Incidents	Growth over Previous Year (%)	Index (2019 = 100)
2019	3,94,499	-	100
2020	11,58,208	193.6	293.6
2021	14,02,809	21.1	355.6
2022	13,91,457	-0.8	352.7
2023	15,92,917	14.5	403.8
2024	20,41,360	28.2	517.5

Source: Ministry of Electronics and Information Technology, replies in the Lok Sabha based on CERT-In data; growth and index calculated by the author.

The number of incidents tracked in India increased from about 3.9 lakh in 2019 to more than 20 lakh in 2024, which is more than five times the original level (Ministry of Electronics and Information

National Conference on Emerging Trends in Arts, Humanities and Multidisciplinary Research (NCETAHMR)

Organized by- Navyug College of Education, Hisar (16 Aug 2026)

In IJESH ISSN: 2250-3115

Technology, Lok Sabha Reply 2). The largest jump, nearly 194 per cent, took place in 2020, when the pandemic forced organisations and individuals to shift quickly to remote work, online learning, digital payment and cloud services, and this enlarged the attack surface. The growth moderated in 2021 and the number was almost the same in 2022, even though CERT-In issued directions in April 2022 which made reporting of incidents mandatory (Ministry of Electronics and Information Technology, Lok Sabha Reply 3). The rise again in 2023 and the high growth of 28.2 per cent in 2024 shows that the threat is still increasing. It should be noted that these numbers reflect both the real number of attacks and the capacity to detect and report them. CERT-In operates automated threat exchange and monitoring systems, so a part of the growth may come from better detection. The data are therefore a good indicator of the scale of the problem, but not an exact count of attacks. The volume of more than twenty lakh incidents in a single year supports the argument of Buczak and Guven and of Shaukat et al. that manual analysis alone cannot keep up and that learning-based automation is needed (Buczak and Guven 1154; Shaukat et al. 222311).

The cost of a breach and the time taken to identify and contain it are two important measures of the impact of a cyber attack. Table 2 compares these measures for 2023, 2024 and 2025, together with the savings reported by organisations that used security AI and automation extensively.

Table 2: Comparison of Global Data Breach Cost, Breach Lifecycle and Savings from Extensive Use of Security AI and Automation

Report Year	Global Average Breach Cost (US\$ million)	Change over Previous Year (%)	Breach Lifecycle (days)	Saving with Extensive Security AI and Automation (US\$ million)	Saving as Share of Average Cost (%)	Lifecycle Shorter with Extensive AI (days)
2023	4.45	-	277	1.8	40.4	108
2024	4.88	9.7	258	2.2	45.1	98
2025	4.44	-9	241	1.9	42.8	80

Source: IBM Security and Ponemon Institute, Cost of a Data Breach Reports 2023, 2024 and 2025; change and share calculated by the author. Savings are rounded as reported.

The global average cost of a breach rose to a record level of 4.88 million dollars in 2024 and then fell by about nine per cent to 4.44 million dollars in 2025, which was the first decline in five years (IBM Security and Ponemon Institute, 2025 Report 5). During the same period the breach lifecycle improved steadily from 277 days to 241 days, a reduction of 36 days. The IBM reports attribute part of this improvement to faster detection, with greater use of security AI and automation and with more breaches being detected internally (IBM Security and Ponemon Institute, 2024 Report 5; IBM Security and Ponemon Institute, 2025 Report 5). In each of the three years, the organisations that used AI and automation extensively had much lower costs, with savings of 1.8, 2.2 and 1.9 million dollars, which is equal to about 40 to 45 per cent of the average breach cost. Their breach lifecycle was also shorter by 108, 98 and 80 days. The size of the advantage in days has narrowed over the three years,

National Conference on Emerging Trends in Arts, Humanities and Multidisciplinary Research (NCETAHMR)

Organized by- Navyug College of Education, Hisar (16 Aug 2026)

In IJESH ISSN: 2250-3115

which may indicate that AI-based tools are becoming common and that organisations without them have also improved. This result supports the view that AI-based detection and response improve the speed of defence, and the view of Gordon and Loeb that security investment should be judged by the expected loss that it prevents (Gordon and Loeb 439). Since the study includes only organisations that suffered a breach, and since organisations that adopt AI may also be better managed in other ways, the savings should be read as an association and not as a guaranteed effect.

AI as Defence versus AI as Risk: Comparison of Breach Outcomes in 2025

The 2025 report provides a direct comparison between the benefit of AI used in defence and the cost of AI used without control. Table 3 compares breach outcomes for these groups.

Table 3: Comparison of Breach Outcomes for AI-enabled Defence and Shadow AI (2025)

Indicator	Group A	Group B	Difference (B compared with A)
Average breach cost (US\$ million): global average (A) versus breaches involving shadow AI (B)	4.44	4.63	+0.19 (+4.3%)
Average breach cost (US\$ million): organisations with no AI or automation (A) versus extensive AI and automation (B)	5.52	3.62	-1.90 (-34.4%)
Customer personal data compromised (% of breaches): global average (A) versus shadow AI incidents (B)	53	65	+12 percentage points
Intellectual property compromised (% of breaches): global average (A) versus shadow AI incidents (B)	33	40	+7 percentage points

Source: IBM Security and Ponemon Institute, Cost of a Data Breach Report 2025; differences calculated by the author.

Table 3 shows a contrast. When AI is used with proper control in security operations, the average breach cost is about 34 per cent lower than for organisations that use no AI or automation. When AI is adopted without control, the cost is higher and more sensitive data are exposed. The report found that one in five organisations had suffered a breach due to shadow AI, that only 37 per cent had policies to manage AI or detect shadow AI, and that organisations with a high level of shadow AI had on average 670,000 dollars higher breach costs (IBM Security and Ponemon Institute, 2025 Report 20). Personal data were compromised in 65 per cent of the shadow AI incidents against 53 per cent in all breaches, and intellectual property in 40 per cent against 33 per cent. The report also noted that 13 per cent of organisations had breaches of AI models or applications, and that nearly all of these lacked proper access controls for AI (IBM Security and Ponemon Institute, 2025 Report 20). These findings are consistent with the argument of Taddeo et al. that trust in AI for cybersecurity has to be earned through reliability and oversight (Taddeo et al. 557), and with the risk management approach of the NIST AI Risk Management Framework, which begins with governance of AI systems (National Institute of Standards and Technology, AI Risk Management Framework 20). The comparison

National Conference on Emerging Trends in Arts, Humanities and Multidisciplinary Research (NCETAHMR)

Organized by- Navyug College of Education, Hisar (16 Aug 2026)

In IJESH ISSN: 2250-3115

indicates that the benefit of AI depends not on the technology alone but on the control framework in which it is deployed.

Major Findings

The study offers the following findings. First, the number of cyber security incidents tracked in India increased from about 3.9 lakh in 2019 to more than 20 lakh in 2024, which is more than five times, with the sharpest rise in the pandemic year of 2020. Second, the global average cost of a data breach rose to 4.88 million dollars in 2024 and fell to 4.44 million dollars in 2025, and the average breach lifecycle improved from 277 to 241 days between 2023 and 2025. Third, organisations that used security AI and automation extensively saved between 1.8 and 2.2 million dollars per breach, about 40 to 45 per cent of the average cost, and detected and contained breaches 80 to 108 days faster. Fourth, unsanctioned or shadow AI raised the breach cost and exposed more personal data and intellectual property, and most organisations had no policy to manage it. Fifth, the literature shows that AI-based defences can themselves be attacked through adversarial inputs and data poisoning, and that attackers can also use AI, so AI cannot replace governance and human oversight.

Suggestions

Organisations should deploy AI and automation in the core tasks of detection, alert triage, phishing filtering and incident response, and should measure the result through indicators such as the time to detect, the time to contain and the cost per incident. AI models used for security should be tested against adversarial inputs, protected from data poisoning and monitored for performance drift (Biggio and Roli 318).

A governance policy for AI should be adopted, with an approved list of AI tools, access controls, data classification rules and employee training to prevent shadow AI (National Institute of Standards and Technology, AI Risk Management Framework 20). Security investment should be proportionate to the expected loss and should cover people and processes as well as tools (Gordon and Loeb 439). Human analysts must remain in the loop for important decisions, and the explainability of alerts should be improved so that analysts can trust and verify them (Sommer and Paxson 306). The national agencies should continue to expand incident reporting, publish more detailed and timely data, and support small and medium organisations with affordable tools and training. Future research should use organisation-level data from India to measure the effect of AI-based defence on detection time and loss.

Conclusion

Artificial intelligence has become a central element of cybersecurity in modern information systems. The data analysed in this paper show a very large and growing number of incidents in India and a high cost of breaches worldwide, and they show that organisations that use AI and automation extensively detect and contain breaches faster and at a much lower cost.

At the same time, the literature and the recent breach data show that AI introduces new risks, which include adversarial manipulation, malicious use by attackers and unsanctioned use within organisations. The security of information systems in the age of AI therefore depends on using AI wisely: as a tool within a framework of governance, secure design, skilled people and human

National Conference on Emerging Trends in Arts, Humanities and Multidisciplinary Research (NCETAHMR)

Organized by- Navyug College of Education, Hisar (16 Aug 2026)

In IJESH ISSN: 2250-3115

judgement. Organisations and regulators that combine technological adoption with responsible control will be better able to protect data, maintain trust and sustain the digital economy.

References

1. Anderson, R. (2020). *Security Engineering: A Guide to Building Dependable Distributed Systems* (3rd ed.). Wiley.
2. Apruzzese, G., Laskov, P., Montes de Oca, E., Mallouli, W., Brdalo Rapa, L., Grammatopoulos, A. V., & Di Franco, F. (2023). The role of machine learning in cybersecurity. *Digital Threats: Research and Practice*, 4(1), 1-38.
3. Biggio, B., & Roli, F. (2018). Wild patterns: Ten years after the rise of adversarial machine learning. *Pattern Recognition*, 84, 317-331.
4. Brundage, M., Avin, S., Clark, J., et al. (2018). *The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation*. arXiv:1802.07228.
5. Buczak, A. L., & Guven, E. (2016). A survey of data mining and machine learning methods for cyber security intrusion detection. *IEEE Communications Surveys & Tutorials*, 18(2), 1153-1176.
6. Dasgupta, D., Akhtar, Z., & Sen, S. (2022). Machine learning in cybersecurity: A comprehensive survey. *Journal of Defense Modeling and Simulation*, 19(1), 57-106.
7. Ferrag, M. A., Maglaras, L., Moschoyiannis, S., & Janicke, H. (2020). Deep learning for cyber security intrusion detection: Approaches, datasets, and comparative study. *Journal of Information Security and Applications*, 50, 102419.
8. Goodfellow, I. J., Shlens, J., & Szegedy, C. (2015). Explaining and harnessing adversarial examples. *International Conference on Learning Representations (ICLR)*.
9. Gordon, L. A., & Loeb, M. P. (2002). The economics of information security investment. *ACM Transactions on Information and System Security*, 5(4), 438-457.
10. IBM Security & Ponemon Institute (2023). *Cost of a Data Breach Report 2023*. IBM Corporation.
11. IBM Security & Ponemon Institute (2024). *Cost of a Data Breach Report 2024*. IBM Corporation.
12. IBM Security & Ponemon Institute (2025). *Cost of a Data Breach Report 2025*. IBM Corporation.
13. Kaloudi, N., & Li, J. (2020). The AI-based cyber threat landscape: A survey. *ACM Computing Surveys*, 53(1), Article 20.
14. Ministry of Electronics and Information Technology (2022). *Reply in Lok Sabha on Cyber Security Incidents* (answered 19 July 2022). Parliament of India, New Delhi.
15. Ministry of Electronics and Information Technology (2024). *Reply in Lok Sabha on Cyber Security Incidents and Mandatory Reporting Directions of CERT-In* (answered 7 February 2024). Parliament of India, New Delhi.
16. Ministry of Electronics and Information Technology (2025). *Reply in Lok Sabha Unstarred Question on Cybersecurity Incidents (data for 2020 to 2024)*. Parliament of India, New Delhi.

**National Conference on Emerging Trends in Arts, Humanities and
Multidisciplinary Research (NCETAHMR)**

Organized by- Navyug College of Education, Hisar (16 Aug 2026)

In IJESH ISSN: 2250-3115

17. National Institute of Standards and Technology (2018). *Framework for Improving Critical Infrastructure Cybersecurity, Version 1.1*. NIST, Gaithersburg.
18. National Institute of Standards and Technology (2023). *Artificial Intelligence Risk Management Framework (AI RMF 1.0)*, NIST AI 100-1. NIST, Gaithersburg.
19. Nguyen, T. T., & Reddi, V. J. (2023). Deep reinforcement learning for cyber security. *IEEE Transactions on Neural Networks and Learning Systems*, 34(8), 3779-3795.
20. Papernot, N., McDaniel, P., Jha, S., Fredrikson, M., Celik, Z. B., & Swami, A. (2016). The limitations of deep learning in adversarial settings. In *IEEE European Symposium on Security and Privacy* (pp. 372-387).
21. Sarker, I. H., Furhad, M. H., & Nowrozy, R. (2021). AI-driven cybersecurity: An overview, security intelligence modeling and research directions. *SN Computer Science*, 2, Article 173.
22. Shaukat, K., Luo, S., Varadharajan, V., Hameed, I. A., & Xu, M. (2020). A survey on machine learning techniques for cyber security in the last decade. *IEEE Access*, 8, 222310-222354.
23. Sommer, R., & Paxson, V. (2010). Outside the closed world: On using machine learning for network intrusion detection. In *IEEE Symposium on Security and Privacy* (pp. 305-316).
24. Stallings, W., & Brown, L. (2018). *Computer Security: Principles and Practice* (4th ed.). Pearson.
25. Taddeo, M., McCutcheon, T., & Floridi, L. (2019). Trusting artificial intelligence in cybersecurity is a double-edged sword. *Nature Machine Intelligence*, 1, 557-560.