ARTICLE

Integrating Artificial Intelligence into Ethical Hacking: A Framework for Automated Threat Mitigation Defensive Security

Aranya Nath* **& Gautami Chakravarty+**

Abstract

Digital ecosystems are growing rapidly as more people turn to artificial intelligence in their everyday lives. As a result, high-tech cyber threats have emerged, fundamentally reshaping the vision of cybersecurity. Ethical hacking has become a necessary tool for combating cybercrime, and the conventional approach used by ethical hackers is often insufficient. Consequently, AI-based automation and real-time threat prevention are revolutionising ethical hacking. This article provides a clear indication of the development of ethical hacking in the age of artificial intelligence and, more specifically, automated threat mitigation. It also examines how AI technologies have changed defensive cybersecurity practices and assesses the ethical and legal issues that autonomous security systems present to the contemporary cyber-legal landscape. The research approach used is a doctrinal and analytical design, which implies a detailed analysis of the interdisciplinary literature on AI, cybersecurity and ethical hacking, along with a review of the current international standards of cybersecurity and new AI regulation systems. The practical uses of AI-driven automated threat mitigation are addressed through selected case studies and industry practices. Moreover, the analysis of AI reveals that it positively affects ethical hacking, as it enables complete monitoring, threat recognition, and automatic responses. Nonetheless, it also uncovers the most vital problems associated with accountability, clarity, information bias, and susceptibility to adversarial attacks in AI. Automated mitigation systems are effective, but they need organised human monitoring to avoid ethical and legal non-compliance. Therefore, hybrid solutions for ethical hacking, which involve automation and human decision-making with AI, should be developed. On the legal front, the study illuminates the equilibrium of AI-powered cybersecurity in data security legislation, the development of AI security laws and regulations, and in international security frameworks. Ethical governance frameworks must be put in place to achieve responsible use of autonomous security technologies. The article concludes that AI-based automated threat mitigation may be used to enhance cybersecurity resilience, but its responsible adoption would involve hybrid designs that implement automation in ethical governance and legal compliance.

Keywords –Ethical Hacking, Artificial Intelligence, Automated Threat Mitigation, Cybersecurity, AI Ethics, Cyber Law

I. INTRODUCTION

People and organisations are still exposed to sophisticated cyber-attacks in the modern technological environment, which keeps being propagated in many internet-based environments. Such harmful

* Research Assistant, DPIIT IPR Chair, HNLU, Raipur
+ LLM (IPR) NLUJAA, Assam

attacks are aimed at disrupting business, stealing computing resources, and interfering with confidential information, which may often cause great financial losses. Malware, ransomware, phishing, and advanced persistent threats (APTs) are the most common types of such threats.[31] Conventional security controls, such as firewalls and antivirus software, are increasingly ineffective against such dynamic foes, thus compelling the implementation of AI-based threat-detection tools that can significantly shorten the response time.[32]

Although manual penetration testing can be used to identify known vulnerabilities, it is often not comprehensive or slow enough to respond to constantly evolving strategies used by cybercriminals. As a result, ethical hacking is taking a central stage in enhancing cybersecurity by identifying vulnerabilities within authorised boundaries. Through simulation of real attacks, ethical hackers unveil vulnerabilities in the systems and networks, hence allowing organisations to deal with the possible threats before they are exploited. They specialise in creating new ways of copying the mindset and strategies of the bad actors, thereby predicting and preventing security attacks. Preemptive approaches have led to financial reward for ethical hackers who can detect and report vulnerabilities, which is often done under bug-bounty programmes.[33]

The concept of Artificial Intelligence was introduced in 1956, and then it developed into actual solutions that are applied in various fields.[34] Machine learning was first introduced to the field of cybersecurity during the 1990s with the introduction of anomaly detection systems (ADS) and intrusion detection systems (IDS).[35] The methods that are based on signature detection and predefined rules fail to keep up with the new and emerging threats. Artificial Intelligence, conversely, is used to improve the vulnerability assessment by identifying risks and autonomously acting on incidents through machine learning, deep learning, and natural language processing using real-time data processing, pattern recognition, and anomaly detection.

Significant benefits are:

- Precision and Breadth: AI systems can ingest and analyse data from a large range of sources in real time, meaning that possible threats are noticed early and treated accordingly.
- Improved Accuracy: AI-powered solutions minimise false positives by using existing data, which gets progressively better over a period, which reflects in increased threat detection certainty.

Problem Statement

Even though a lot of technological advancement has been seen in the last few years, there is still a huge discontinuity between the classical ethical hacking methodologies and the empirical realities that are characteristic of AI-infused attack surfaces. Classical approaches have a mainly reactive

[31] 'What is a Cyber Attack? How They Work and How to Stop Them', Search Security. Accessed: Mar. 28, 2025. [Online]. Available: https://www.techtarget.com/searchsecurity/definition/cyber-attack

[32] F. Asif, F. Sohail, Z. H. Butt, F. Nasir, and N. Asgar, 'Ethical Hacking and its role in Cybersecurity', Aug. 28, 2024, arXiv: arXiv:2408.16033. doi: 10.48550/arXiv.2408.16033.

[33] A Survey of Bug Bounty Programs in Strengthening Cybersecurity and Privacy in the Blockchain Industry'. Accessed: Mar. 28, 2025. [Online]. Available: https://www.mdpi.com/2813-5288/2/3/10

[34] M. Z. Alom et al., 'The History Began from AlexNet: A Comprehensive Survey on Deep Learning Approaches', Sep. 12, 2018, arXiv: arXiv:1803.01164. doi: 10.48550/arXiv.1803.01164.

[35] A. D. Joseph, P. Laskov, F. Roli, J. D. Tygar, and B. Nelson, 'Machine Learning Methods for Computer Security (Dagstuhl Perspectives Workshop 12371)', Dagstuhl Manifestos, vol. 3, no. 1, pp. 1–30, 2013, doi: 10.4230/DagMan.3.1.1.

posture, are time-consuming and rely on pre-determined signatures or known vulnerabilities. Therefore, these techniques have limited capabilities to contend with adaptive enemies who use AI to refine attack plans, bypass defenses, and use zero-day attacks. Moreover, the implementation of AI-based defensive systems brings in new problems, such as the opacities in the algorithms, responsibility of automated decision making, and vulnerability to adversarial machine-learning attacks. Lack of clear ethical and legal frameworks of autonomous threat mitigation gives rise to anxiety about how it can be abused, how it can cause unintended harm, and how it may be non-compliant with the law. The given paradox of technological ability and governance highlights the importance of a strict analysis of AI-powered ethical hacking.

II. FUNDAMENTALS OF ETHICAL HACKING

Ethical hacking, also known as white-hat hacking or penetration testing, is a controlled attempt to assess computer systems, networks, and applications to expose weaknesses and, as a result, install improved cybersecurity measures. Ethical hackers adopt similar techniques to those used by harmful agents, though within the scope of the law, and to enhance security and not compromise strengths. With the current upward trend in the frequency of cyberattacks, ethical hacking has become an urgent exercise, which allows organisations to address security vulnerabilities before attacks can be made on them. The field is a key factor in not only ensuring that security standards are met, but also in preventing data breaches and optimising risk-management models. Many organisations in different sectors now rely on ethical hacking to secure sensitive information and maintain the integrity of their online systems and networks.[36] Ethical hackers apply various methodologies to help them locate vulnerabilities in the security systems in a systematic manner. Penetration testing, in which the assessor tries to simulate real attacks with the aim of measuring system defence strength, is a widely used method. Social engineering, based on human psychology, may force humans to share personal information by way of phishing or pretexting. Network scanning requires building diagrams of the underlying architecture to identify areas of weakness, like open ports of entry points to the network by potential intruders.

The systematic discovery of security vulnerabilities is the major activity of testing web applications, which can be described as threats of vulnerabilities, including SQL injection and cross-site scripting (XSS) threats. Security researchers can map the possible ingress points in a simulation of a penetration attempt in a real wireless network environment. Combining these investigation methods, it is possible to perform a comprehensive evaluation of the security architecture of an enterprise, which preconditions the creation of more effective defensive positions.

Audits on security will be very effective when carried out by ethical hackers who install specialised equipment. Some of the important tools include Nmap, which is used to scan the network comprehensively; Metasploit, designed to enable a detailed examination of packets; and Burp Suite, which is used to identify vulnerabilities in web applications. By using such tools, ethical hackers will be able to automate the testing process and identify gaps with a greater level of accuracy. The methodical approach, including the initial data gathering on the target system, followed by scanning, breaching it to achieve and maintain access, and finally drafting a formal report, gives a strict

[36] 'Scarfone et al. - 2008 - Technical guide to information security testing an.pdf'. Accessed: Mar. 28, 2025. [Online]. Available: https://nvlpubs.nist.gov/nistpubs/legacy/sp/nistspecialpublication800-115.pdf

structure to further documentation and recovery. In the ethical hacking terminology, the first stage of this well-organised approach is usually known as reconnaissance. In this phase, the researcher gathers exhaustive data on the target system, which is a collection that guides all other stages of the assessment. After reconnaissance, the process advances to a series of scans to target access credentials and the sustenance of said access, to the ultimate report, which leads to finalization of findings before final documentation. After reconnaissance, the process continues through specific scanning, the procurement of access, access continuation, and the ultimate reporting before the actual documentation of results occurs.

The ethical ethos of hacking practice is comprised of security models and guidelines that are built by recognised organisations like NIST and the Open Web Application Security Project (OWASP)[37] Although it is undeniably true that ethical hacking plays an extremely central role in the larger cybersecurity context, it is also important that such practices follow a carefully laid-out legal and moral framework. The second reason why corporations that need to reduce their legal exposure should ensure that they obtain express permission from the concerned stakeholders before they embark on the security testing process. The behaviour of these exercises is regulated by legislative tools, most prominently the Computer Fraud and Abuse Act (CFAA) in the United States and the General Data Protection Regulation (GDPR) in the European Union, and ethical hackers should stay vigilant to prevent disruption of organisational operations and privacy of customers, that is, these tests should not interfere with normal business activities. Professional ethical hacking and legal compliance are maintained through compliance with the principles that are established in the Hacker Code of Ethics[38] such as the importance of honesty, secrecy, and responsible disclosure.

III. ARTIFICIAL INTELLIGENCE IN CYBERSECURITY

Cybersecurity is a domain that has been radically transformed by artificial intelligence (AI), primarily by augmenting the anomaly detection process, automating the incident response process, and enhancing the threat recognition process. Early in the 21st century, machine learning (ML) algorithms were initially implemented in intrusion detection systems and antivirus platforms, thus creating a base on which more advanced AI programs can be built in the same sphere. Modern AI-based systems use deep-learning systems and natural-language-processing (NLP) frameworks to perform real-time threat detection, automatically identifying malicious behaviour by querying large amounts of data to identify harmful patterns. Moreover, AI-enhanced ethical hacking uses ML and deep-learning models to enhance threat identification, automate penetration-testing processes, and analyse text in search of signs of social-engineering efforts. AI also forms the core of cyber-threat intelligence, as it automates the data processing and matches the dissimilar threat indicators, thus streamlining the cumbersome processes that would be manual in the past. Despite these numerous benefits, such as increased detection effectiveness and reduced human error, AI in cybersecurity is

[37] Open-Source Intelligence Testing Using the OWASP Version 4 Framework at the Information Gathering Stage (Case Study: X Company)'. Accessed: Mar. 28, 2025. [Online]. Available: https://www.researchgate.net/publication/342662106_Open_Source_Intelligence_Testing_Using_the_OWASP_Version_4_Framework_at_the_Information_Gathering_Stage_Case_Study_X_Company

[38] 'Cybercrime and the Law: Primer on the Computer Fraud and Abuse Act and Related Statutes | Congress.gov | Library of Congress'. Accessed: Mar. 28, 2025. [Online]. Available: https://www.congress.gov/crs-product/R47557

facing significant challenges, such as adversarial examples of AI solutions, inherent biases to produce false positives and false negatives, and significant resource consumption in terms of deployment and model training. Current academic research is still questioning the limits of AI and how it can be applied in strengthening ethical hacking and threat defence procedures.

Threat Landscape and Core Machine Learning Techniques

Artificial intelligence and machine learning are used in cyberattacks to transform the security environment, and are becoming more complex and multidimensional. Such advanced automated attacks do not tend to be overcome by traditional defence mechanisms. Consequently, AI-based ethical hacking has become an inevitable aspect of simulated attacks, vulnerability discovery, and mitigation, as well as software that simulates the behaviour of an adversary.

Several factors characterize the threat environment:

The Emergence of AI-Powered Attacks: There is empirical evidence showing that modern cyber threats use AI in various areas of operation, such as automated reconnaissance and exploit generation.[39] Malware powered by AI can dynamically change its signatures to avoid detection, and reinforcement learning models can change attack strategies in response to changing conditions.[40]Also, AI-enhanced phishing attacks' improvisation of persuasive messages, which circumvent standard filtering mechanisms, is based on natural-language generation.[41]

1. Attack Scaling and Supply Chain Exploitation: Portal incidents like the SolarWinds intrusion can be used to describe how attackers can utilise vendor relationships as a method to hack into thousands of downstream systems without detection. The automation encourages the implementation of massive attacks with the least possible human control.[42]
2. Hacking Attacks: One of these attacks is exploited by cybercriminals by using vulnerabilities in system configurations and algorithms of AI, so that they can interfere with data confidentiality and network integrity. Artificial intelligence also promotes the creation of intelligent malware that is undetectable by detection tools.
3. Effect of Data Breach: The prevalence of data breaches teaches of the high organisational susceptibility, which leads to the massive consequences of cyber-adversarial attacks.
4. Phishing Scams: Phishing attacks are advancing because criminals use AI technologies to blackmail people into revealing sensitive data. AI may be used to analyse email data and web behaviour to enhance phishing identification, hence the need to have advanced, AI-based protective measures to overcome such sophisticated attacks.

---

[39] 'Cyber Shadows: Neutralizing Security Threats with AI and Targeted Policy Measures'. Accessed: Oct. 21, 2025. [Online]. Available: https://arxiv.org/html/2501.09025v2

[40] D. Abbadi, 'Cyber threats in the age of artificial intelligence: Exploiting advanced technologies and strengthening cybersecurity', International Journal of Science and Research Archive, vol. 13, pp. 2576–2588, Oct. 2024, doi: 10.30574/ijsra.2024.13.1.1961.

[41] 'What is Polymorphic Malware? Examples & Challenges'. Accessed: Oct. 21, 2025. [Online]. Available: https://www.sentinelone.com/cybersecurity-101/threat-intelligence/what-is-polymorphic-malware/

[42] 'Understanding Supply Chain Attack Tactics With Case Study - Brandefense'. Accessed: Oct. 21, 2025. [Online]. Available: https://brandefense.io/blog/understanding-supply-chain-attack-tactics-with-case-study/

IV. Automated Threat Mitigation Frameworks

The pace at which enterprises are adopting this new technology has increased faster as more and more businesses are expected to embrace the Internet of Things (IoT), with an expected proliferation of 45 billion IoT devices by 2021. The IoT wave has spawned a huge amount of data due to the improvement of remote device management and the introduction of more intelligent processes. However, the available legacy protocols like Telnet and Secure Shell that use password authentication have vulnerabilities that, through brute force attacks, can be compromised. To prevent these risks, it is impossible not to include artificial intelligence (AI) in the threat detection systems. In contrast to traditional signature-based solutions, AI uses machine learning and deep-learning methods in the recognition of real-time trends of malicious actions and zero-day vulnerabilities before their official acknowledgement.[43] Predictive analytics also enhances security by using previous data to predict possible cyber threats. Models that have been trained using a mixture of data types enable the differentiation between regular and suspicious behaviours, whereas the deep-learning anomaly detection can effectively detect small changes that suggest an imminent attack.[44] The AI solutions also provide real-time monitoring of large volumes of endpoints, which in turn increases organisational response capacity compared to that of human analysts. It is also the technology that is utilized to complement already existing systems, such as intrusion detection systems (IDS) and security information and event management (SIEM), to become more effective[45]. Despite the serious advances in AI-based threat detection, it has been facing issues, including the necessity to use highly polished training samples, algorithmic bias, and the creation of AI-driven adversarial attacks. Nevertheless, AI is a key aspect in developing proactive and scalable defence strategies with these challenges.

Cyber Threats and Vulnerability Patterns in Digital Systems

The volume and intricacy of digital systems are growing, and the sophistication of cyber threats and vulnerability patterns is keeping pace with the same effect. The intensive digital transformation that is being experienced in the banking, healthcare, and governmental sectors in India has increased the vulnerability of the country to a wide range of cyberattacks. Ransomware, phishing, data breaches, and advanced persistent threats against critical infrastructure-all these have their roots in the easy availability of vulnerable IoT devices, cloud computing, and mobile banking systems. Vulnerabilities, outdated software, and weak security setups are under constant attack. Since adversaries always seek to exploit weaknesses before detection, Indian organisations need AI-powered detection techniques for threat detection.[46]

---

[43] 'Internet of Things is a revolutionary approach for future technology enhancement: a review | Journal of Big Data | Full Text'. Accessed: Nov. 05, 2024. [Online]. Available: https://journalofbigdata.springeropen.com/articles/10.1186/s40537-019-0268-2

[44] I. Ullah, D. Adhikari, X. Su, F. Palmieri, C. Wu, and C. Choi, 'Integration of data science with the intelligent IoT (IIoT): current challenges and future perspectives', Digital Communications and Networks, Mar. 2024, doi: 10.1016/j.dcan.2024.02.007.

[45] O. Ogundairo and P. Broklyn, 'Predictive Analytics for Cyber Threat Intelligence', Journal of Cyber Security, Aug. 2024.

[46] 'Cyber Threats to Critical Infrastructure: Assessing Vulnerabilities Across Key Sectors'. Accessed: Nov. 05, 2024. [Online].

AI Algorithms for Predictive Threat Analysis

Artificial Intelligence(AI) has become central to modern cybersecurity, being an automated method of identifying, analysing, and eliminating dangers by deploying machine learning algorithms and advanced processing methods. Using AI to examine large volumes of data will make it easier to identify possible threats at their initial stage, thus allowing security teams to identify the weaknesses in their systems. Additionally, the automation of incident response or response management is impossible without AI systems, as cyber threats are of uncountable types and forms that evolve as fraudsters learn to use different methods. The threats include generative AI exploits, phishing campaigns, polymorphic malware, and zero-day vulnerabilities, each of which is a challenging adversary to AI-based protective measures. Moreover, AI-enhanced threat detection improves cybersecurity resilience to new challenges presented by mobile devices, the Internet of Things (IoT), and cloud deployments; it does so by taking into consideration the increased frequency and complexity of threats, such as ransomware. Threat analysis based on artificial intelligence algorithms to predict threats enables cybersecurity systems to respond to them with early interventions based on trend analysis using historical data. Predictive analytics is also utilised in the Indian scenario to create anti-fraud systems that help banks and payment providers to prevent fraud. These systems use unsupervised and supervised machine learning methods to analyse massive data sets to discover patterns that could be dangerous and would otherwise remain undetected by human analysts. In turn, predictive analytics and AI integration are one of the key aspects of a proactive cybersecurity solution for Indian organisations that seek to prevent cyber threats.[47]

Machine Learning and Deep Learning for Anomaly and Intrusion Detection

Recently, machine learning and its use in detecting anomalies in IoT systems have received significant interest as an anomaly detection tool to track suspicious behaviour. Four distinct learning strategies, supervised, unsupervised, semi-supervised, and reinforcement learning, make up the method. When data is labelled, supervised learning can find outliers; when data is unlabeled, autonomous learning can find them by comparing data structures. Machine learning (ML) has several uses, including the capacity to automate processes, optimize resources, and detect threats in real-time. It also supports decisions in predictive maintenance and makes systems scalable. Anomaly detection is an important issue in many different industries, and several machine learning algorithms provide the best answers. Using both labelled and unlabeled IoT data entering, the main operation then runs a system to check if the data is normal or anomalous. Research on connected device assaults and anomaly detection using intrusion detection systems (IDSs) is presented in the literature to show how effective anomaly detection using machine learning is. Attacks against the Internet of Things necessitate erratic system behaviours, such as the transfer of harmful payloads, troublesome

Available:https://www.researchgate.net/publication/378078435_Cyber_Threats_to_Critical_Infrastructure_Assessing_Vulnerabilities_Across_Key_Sectors

[47] 'Riding the AI Waves: An Analysis of Artificial Intelligence's Evolving Role in Combating Cyber Threats'. Accessed: Nov. 05, 2024. [Online]. Available: https://www.researchgate.net/publication/378068138_Riding_the_AI_Waves_An_Analysis_of_Artificial_Intelligence's_Evolving_Role_in_Combating_Cyber_Threats

behavioral patterns, and unforeseen network streaming. Using ML frameworks, researchers have conducted a thorough examination of many anomaly detection areas.[48]

Anomaly Detection

Various classification approaches, such as "K-Nearest Neighbours, Multilayer Perceptron, Decision Trees, Linear Discriminant Analysis, Logistic Regression, and Naïve Bayes," are utilized in the research to examine anomalies in the Internet of Things (IoT). When it comes to non-time-series data, Decision Trees and linear discriminant analysis achieve an accuracy of 80%. However, when it comes to time-series data, such as trends, Neural Networks with memory gates produce superior results.[49]

Research has examined multiple ML models to solve IoT infrastructure cybersecurity issues through the detection of assaults and irregularities by analysing open-source Kaggle dataset samples.[50] Random Forest, together with "Artificial Neural Network (ANN), produced superior results to Decision Trees during testing and training phases with 99.4% accuracy, because Support Vector Machine and Logistic Regression provided underwhelming outcomes.

The proposed research provided a new approach towards detecting attacks and deviations in connected devices through a feature-transformation-based classifier for calculating missing data values.[51] To assess the DS2OS dataset, our research utilized Decision Tree (DT), Naïve Bayes (NB), Support Vector Machine (SVM), and Random Forest (RF). With the use of cutting-edge imputation technology, the proposed method reduced the dataset's capacity by using a feature modification strategy to fill in missing values. This article presents a novel supervised and unsupervised approach to detecting anomalies in industrial sensor IoT networks, and it explores the current limits of ML-based anomaly detection systems. The study assessed the efficiency of the Random Forest and Decision Tree models with regard to minimising false positive results in the context of privacy concerns. The intrinsic sensitivity of data makes it difficult to apply the results of the studies. In a study, traffic pattern anomalies were analysed by researchers to assess the risks that IoT networks pose to privacy and security. Researchers identified shallow learning methods, such as Naïve Bayes and Support Vector Machine, trained on the UNSW-NB15 and DAD datasets. Industries such as telecommunications and e-commerce can utilise ML-DL in real time for attack detection. Deep learning requires layered neural networks to handle complex attack detection; therefore, ML is not sufficient on its own. Indian telecommunications companies have built ML- and DL-based platforms for detecting malicious activities to improve network security. AI-driven insider threat detection is essential through behavioural analyses, as it allows for the continual monitoring of user activities to identify deviations that reflect potential malfeasance. This approach greatly enhances the

[48] A. Chatterjee and B. S. Ahmed, 'IoT anomaly detection methods and applications: A survey', Internet of Things, vol. 19, p. 100568, Aug. 2022, doi: 10.1016/j.iot.2022.100568.

[49] S. Brady, D. Magoni, J. Murphy, H. Assem, and A. O. O. Portillo-Dominguez, 'Analysis of Machine Learning Techniques for Anomaly Detection in the Internet of Things', in 5th IEEE Latin American Conference on Computational Intelligence, Gudalajara, Mexico: IEEE, Nov. 2018, pp. 1–6. doi: 10.1109/LA-CCI.2018.8625228.

[50] 'Attack and anomaly detection in IoT sensors in IoT sites using machine learning approaches - ScienceDirect'. Accessed: Nov. 05, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2542660519300241

[51] 'A machine learning approach for imputation and anomaly detection in IoT environment - Vangipuram - 2020 - Expert Systems - Wiley Online Library'. Accessed: Nov. 05, 2024. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1111/exsy.12556

cybersecurity posture of sectors such as administration and healthcare in India and helps in the mitigation of both external and internal threats.[52]

Multi-Source Telemetry Fusion

The telemetry fusion of many sources, where endpoint events, network traffic, syslogs, and cloud activities come in and interact into a single model, is the foundation of AI-based detection.[53]As an example, endpoint telemetry, such as process creation events and registry changes, network telemetry, such as packet traffic metrics and DNS anomaly logs, and identity telemetry, such as user login events and instances of privilege elevation, are all combined to form a complete behavioural graph.[54] These combined datasets are then input into advanced graph neural networks (GNNs) and Bayesian inference models to identify lateral movement or abuse of privileges in enterprise networks.[55]

Case Studies: Applications of AI in Real-World Threat Detection in the Context of India

The high pace of digitalization in India demands more effective detection of the threat, and artificial intelligence (AI) is the key to cybersecurity. The use of AI by financial institutions such as the ICICI Bank and the State Bank of India to identify and curb fraud is also effective because the incidents of scams are on the rise, following the rise in internet banking and online payments. The AI machine learning algorithms are used to monitor transactions in real-time, ensuring that anomalous behaviour is identified more effectively than the conventional solution would otherwise have been the case[56]. Telecommunication companies such as Bharti Airtel and Reliance Jio apply AI to secure their networks against cyber-attacks and use deep learning to detect malware, so their networks are secure and reliable in terms of customer data and network stability in a competitive environment.[57]

The use of AI is essential to defend critical infrastructure, as well as the case of the National Critical Information Infrastructure Protection Centre (NCIIPC), which is increasing its response to cyber threats by monitoring trends that may suggest state-backed attacks. The actions should be proactive in the context of increased geopolitical tensions on national security. Healthcare is one of the fields that has incorporated AI in data security, and organizations such as Apollo Hospitals and Max Healthcare are employing automated technologies to make sure that patient records are not accessed by unauthorized persons and to track unauthorized access patterns (after the emergence of

---

[52] L. Vigoya, D. Fernandez, V. Carneiro, and F. J. Nóvoa, 'IoT Dataset Validation Using Machine Learning Techniques for Traffic Anomaly Detection', Electronics, vol. 10, no. 22, Art. no. 22, Jan. 2021, doi: 10.3390/electronics10222857.

[53] '(PDF) Multimodal Anomaly Detection: Combining Data from Different Sources'. Accessed: Oct. 21, 2025. [Online]. Available: https://www.researchgate.net/publication/391459064_Multimodal_Anomaly_Detection_Combining_Data_from_Different_Sources

[54] D. Preuveneers and W. Joosen, 'Privacy-preserving correlation of cross-organizational cyber threat intelligence with private graph intersections', Comput. Secur., vol. 135, no. C, Dec. 2023, doi: 10.1016/j.cose.2023.103505.

[55] O. Koucham, S. Mocanu, G. Hiet, J.-M. Thiriet, and F. Majorczyk, 'Cross-domain alert correlation methodology for industrial control systems', Computers & Security, vol. 118, p. 102723, Jul. 2022, doi: 10.1016/j.cose.2022.102723.

[56] R. Achary and C. Shelke, Fraud Detection in Banking Transactions Using Machine Learning. 2023, p. 226. doi: 10.1109/IITCEE57236.2023.10091067.

[57] T. H. Bureau, 'AI-powered spam detection system deployed by Airtel', The Hindu, Oct. 08, 2024. Accessed: Nov. 05, 2024. [Online]. Available:https://www.thehindu.com/news/cities/Hyderabad/ai-powered-spam-detection-system-deployed-by-airtel/article68729750.ece

telemedicine during the COVID-19 pandemic). The AI anomaly detection can protect against illegal sharing of sensitive health data[58]

Also, the Indian government is working on AI in the cybersecurity sphere within its Digital India programs. Monitoring tools, based on AI, evaluate government websites and other networks within the government sector on the vulnerability and malicious activities. The Cyber Emergency Response Team of India (CERT-In) uses AI-upgraded threat intelligence tools to detect potential cyber threats to prevent the threat of attacks on the population networks (both domestic and external).

V. DEFENSIVE AUTOMATION & SOAR INTEGRATION

Very complicated hybrid networks are limited by the situations of high stress and the necessity to make a quick decision. SOAR software is deployed by organisations and is based on machine learning and artificial intelligence that is designed to assist in automating routine business operations, ranking issues intuitively, and orchestrating cross-domain actions by sending real-time alerts[59].

Such SOAR playbooks are capable of assisting security groups in the incident detection, prioritisation, and investigation up to resolution. Other than that, contextual enrichment and predictive prioritisation are notable advancements in artificial intelligence. Isolation (isolating endpoints that have the affected code embedded and blocking malicious IP addresses) as an applied technique is imperative and is enhanced by rollback and validation that ensure the integrity of the service. Finally, human monitoring is needed along with approval gates and escalation processes for making sure responsibility is in place with corporate and regulatory requirements.[60]

Task Category	**Automated by SOAR**	**Human Analyst Role**
Alert ingestion & correlation	Log parsing, IOC matching, anomaly scoring	Define correlation logic and validate anomalies
Initial triage	Severity classification, contextual enrichment	Review uncertain classifications
Containment	Endpoint isolation, IP blocking	Approve high-impact containment actions
Remediation	Patch deployment, credential reset	Validate patch plan and monitor outcomes

[58] 'The potential impact of AI on the Indian healthcare industry - Express Healthcare'. Accessed: Nov. 05, 2024. [Online]. Available: https://www.expresshealthcare.in/news/the-potential-impact-of-ai-on-the-indian-healthcare-industry/439611/

[59] R. Kaur, D. Gabrijelčič, and T. Klobučar, 'Artificial intelligence for cybersecurity: Literature review and future research directions', Information Fusion, vol. 97, p. 101804, Sep. 2023, doi: 10.1016/j.inffus.2023.101804.

[60] C. Alejandro, R. Gómez, E. Boyle, Y. Adebayo, and A. Adeyemo, 'Designing Intelligent Security Orchestration, Automation, and Response (SOAR) Systems with AI', pp. 2422–2208, Oct. 2025.

Task Category	Automated by SOAR	Human Analyst Role
Learning feedback	Retrain ML models from labelled incidents	Curate and label ambiguous or new threat samples

VI. AI-ENHANCED AUTOMATED PENETRATION TESTING

Digital enterprises in modern contexts face operational challenges, significantly associated with cybersecurity vulnerabilities resulting from accelerated technological developments.[61] Penetration testing, more commonly known as pen-testing, is one of the most important tools to identify such weaknesses; however, challenges remain regarding the standardisation of processes and the financial control of costs.[62] Artificial intelligence and machine learning techniques are now being used to help speed up risk assessment workflows. The pen testing methodology has several stages of planning, identifying vulnerabilities and post testing remediation, all of which require significant expertise and time investment to execute well.[63]AI-enhanced penetration testing tools follow a quintuple-phase implementation methodology of needs assessment, data collection and preprocessing, model development, tool amalgamation, execution, and deployment. The strict data preparation procedure guarantees the delivery of high-fidelity data sets, which are required to train the AI models and ensures that the continuous improvement through feedback will take place to guarantee the continued efficacy. The paradigm of pen-testing life cycle has shifted towards the proactive paradigm and takes advantage of artificial intelligence to create precision and speed, but needs constant monitoring as well as flexibility to the evolving threat landscape.

VII. AI IN CYBERSECURITY: ETHICAL AND LEGAL CONSEQUENCES

The legal and ethical issues being brought about by artificial intelligence integration in ethical hacking are a significant challenge. The lack of policies regarding transparency and reliability of the AI tools is exacerbated by the fact that the old methods of defense cannot be successfully used to fight the constantly growing and increasingly complex cyber threats. Marginalized groups are disproportionately affected by biases inherent in AI, which is a result of poor data and training algorithms. As an example, most facial-recognition systems are less successful with women and people of colour. Privacy issues, since AI is sensitive to vast volumes of data to perform any threat detection, there exists a conflict between the requirement of security and the requirement of privacy. The overuse of AI endangers human expertise, and this may lead to wrongful judgment when practicing cybersecurity. Accordingly, ethical hackers should enhance their work with the help of AI and maintain transparency, human control, and compliance with ethical guidelines. Legally, the Information Technology Act of 2008 is not very effective at dealing with AI-specific issues in the ethics of hacking because it does not give enough recognition to the ethics of hacking activity. Illegal

[61] D. Sharma Shria Verma, 'Automated Penetration Testing using Large Language Models', IJSR, vol. 13, no. 4, pp. 1826–1831, Apr. 2024, doi: 10.21275/SR24427043741.
[62] 'Cybersecurity | Penetration Testing in Today's Digital Age and Its Importance'. Accessed: Nov. 05, 2024. [Online]. Available: https://www.penntech-it.com/2024/09/23/the-importance-of-penetration-testing-in-todays-digital-age/
[63] 'AI-Based Application Penetration Testing: Tools, Types and Process'. Accessed: Mar. 27, 2025. [Online]. Available: https://qualysec.com/ai-penetration-testing/

AI-based unauthorized penetration testing is punishable by law, and many court cases have proven that. Besides, regulations concerning the handling of personal data are stringent in the Digital Personal Data Protection Act of 2023, which the ethical hacking tools must meet, such as the user consent and security tools. Issues of ethics are important in AI-driven cybersecurity, because only in this way can the implementation of the DPDP Act and other privacy laws be avoided, and therefore, the necessity of a set of ethical requirements regarding AI-driven cybersecurity at the state and even international level.

VIII. LEGAL PRECEDENTS RELATED TO ETHICAL HACKING IN INDIA

Shreya Singhal vs. In the case of the Union of India, the Supreme Court struck down Section 66A of the Information Technology Act of 2000, according to which offensive speech on the Internet was criminalized, as it was unconstitutional according to Article 19(1)(a) of the Indian Constitution. This historical ruling showed the need to have limits on regulation in cybersecurity and ethical hacking, highlighting the need to balance between lawful and unlawful security measures[64]. Since AI is still underdeveloped in the field of cybersecurity, the legal frameworks are changing to distinguish between ethical and malicious hacking. Legal differences depend on factors like authorization, purpose, adherence to laws of data protection and accountability. Organizational-permitted ethical hacking is done to improve security, whereas malicious hacking takes advantage of the vulnerabilities and uses these to gain power. International standards such as the GDPR of the EU and the NIST AI Risk Management Framework in the US provide ethical standards regarding the use of AI in cybersecurity. To overcome the ethical hacking challenges associated with AI, the DPDP Act and the IT Act in India need to be amended to accommodate privacy, liability, and authorization to create a strong regulatory background of ethical AI hacking and secure cyber assets.

IX. CONCLUSION

The very idea of artificial intelligence is changing how businesses tackle the problem of cyber threats, as well as engage in ethical hacking, and offer faster and more convenient responses in the form of machine learning and automation. AI significantly decreases how long it takes to resolve issues because it makes it easier to detect vulnerabilities and process the incidents much faster through Security Orchestration, Automation, and Response (SOAR) tools. Furthermore, predictive analytics allow for preventing threats before they happen. However, the implementation of these tools is associated with several new duties: the process of automation needs to be consistent with the current legal and ethical standards, and human control should be maintained to provide responsibility and make strategic decisions. Even though AI enhances the effectiveness of ethical hackers, transforming subjective judgments into data-based and fast judgments, it can only act as an addition, but not a replacement for human expertise. The collaboration between human analysts and AI, therefore, creates a greater security posture, which is necessary because the threat environment is

[64] 'Section 66A, Information Technology Act, 2000, Shreya Singhal Case'. Accessed: Aug. 04, 2024. [Online]. Available: https://www.livelaw.in/law-firms/law-firm-articles-/section-66a-information-technology-act-2000-shreya-singhal-case-183816

increasingly becoming more complex. The human-AI cooperation today is a reality, and automation can lead to stronger defence, whereas the ethical principles can regulate the operational activities.