International Journal of Basic and Applied Sciences, 14 (7) (2025) 139-148

International Journal of Basic and Applied Sciences

Website: www.sciencepubco.com/index.php/IJBAS
https://doi.org/10.14419/a3kpw407
Research paper

AI Revolutionizing Cybersecurity: An Overview

Bharathi S [1], Alexandros Konios [2], Nandhakumar Manikandasamy [3], Sudha V K [4], Chairman M [5 *], Senbagam B [5]

[1] Associate Professor, Department of Electronics and Communication Engineering, Dr. Mahalingam College of Engineering and Technology, Pollachi, India
[2] Assistant Professor, Department of Computer Science, Nottingham Trent University, U.K
[3] PG scholar, Department of Computer Science, Nottingham Trent University, U.K
[4] Professor, Department of Electronics and Communication Engineering, Dr. Mahalingam College of Engineering and Technology, Pollachi. India
[5] Assistant Professor, Department of Electronics and Communication Engineering, Dr. Mahalingam College of Engineering and Technology, Pollachi, India
**Corresponding author E-mail: bharathi_mani@yahoo.com*

Received: August 10, 2025, Accepted: September 26, 2025, Published: November 5, 2025

Abstract

Data, or digital information, is one of the main engines that drive human society. With ever-evolving digital footprints and human needs, the demand for data protection also increases significantly. Cybersecurity cuts across every nook and corner of the digital world, from physical environments to clouds. With multiple attack vectors evolving day by day, the need for more robust enterprise infrastructure and Threat intelligence also evolves. Artificial Intelligence plays a pivotal role, offering advanced simulation and reasoning capabilities for both offensive and defensive cybersecurity strategies. This work examines the application, ability, advantages, and disadvantages of various cyber strategies. With some limitations around the corner, the recent advancements made it clear that with proper utilization of AI, the human effort involved can be reduced with autonomous threat intelligence and attack capabilities. It also highlights the fact that adversarial training on models makes them more cyber-aware against the effectiveness of the model.

Keywords: *Artificial Intelligence, Attack vectors, Cyber-aware, Cyber strategies, Digital information*

Motivation and scope

Motivation

As cyber threats grow in complexity and frequency, traditional security measures struggle to keep pace with evolving attack vectors. The increasing reliance on digital infrastructure across industries necessitates more intelligent and adaptive security solutions. With features like automatic response, real-time threat detection, and predictive analytics, artificial intelligence (AI) has become a disruptive force in cybersecurity. The motivation behind this study is to explore how AI-driven cybersecurity can enhance defense mechanisms, reduce human intervention, and improve the resilience of enterprise security frameworks.

Scope

This work delves into the integration of AI in cybersecurity, examining its applications, advantages, and limitations. It covers AI-powered threat intelligence, adversarial training, and autonomous security mechanisms. Additionally, the study draws attention to issues including algorithmic prejudice, adversarial assaults, and ethical dilemmas, emphasizing the need for responsible AI deployment. By analyzing recent advancements, this research aims to provide insights into how AI can enhance cyber resilience while addressing its potential risks and limitations.

Structure of the article

The article begins with an Introduction, highlighting the growing importance of AI in addressing evolving cybersecurity threats. It then covers the Evolution of cybersecurity, from basic protections to advanced AI-driven techniques. The Framework section details how the Application section presents actual use cases and how AI may be used in cybersecurity procedures. It discusses various Cyber-Attacks and how AI mitigates these risks, followed by an exploration of the Challenges in AI implementation, including data quality and ethical concerns. The article then addresses Advanced Cyber-Resilience, emphasizing AI's role in strengthening defenses. The Future Work section

suggests areas for further research, and the Conclusion summarizes the importance of AI in cybersecurity while acknowledging the need for proper frameworks.

1. Introduction

Security can be broadly classified into two types: defensive and offensive. AI plays a major role in both defensive and offensive strategies, with capabilities ranging from threat intelligence to attack simulation. AI can be a game-changer in cybersecurity with its reasoning and learning capabilities substantiating its significance in developing automated tools for offensive methodologies. One fine example would be integrating OpenAI capabilities for the reconnaissance phase of ethical hacking, also making the model learn the derived outcomes from the system. Also, care must be taken in open-sourcing the trained models, as it could potentially alleviate the heinous crimes and trigger information warfare with AI capabilities. Human information, also called DNA, is the biological genome information, and it decodes the human's behavior. However, it is the right time we call the publicly available digital footprints the 'Human information' as it enables potential attackers to digitally attack a person and demand ransom rather than physical means. Defensive strategies range from the robust enterprise network to the adhering secure coding practices in an organization. Since it can potentially eliminate many avenues for vulnerabilities, offensive strategies focus on simulating attack vectors, crafting malformed payloads, and analyzing application code to develop malware. They also involve training models on specific use cases and using a feedback loop of corrective measures to become more proactive in targeting systems.

This is potentially highlighting the need for cyber-resilience across the enterprise customers, countries, and we as humans. Data can be exploited in all stages, data exploitation during storage, retrieval, and transmission. The type of data exploited results in adverse effects, including paying huge ransoms and loss of privacy. Digital privacy is one of the most sought-after needs of the human era in recent days. With malware capable of learning system information from the host, it can reproduce its current state based on the transfer learning capabilities with changes to its build. This doesn't limit itself to malware, however, with more automation capabilities and reinforcement learning techniques, we are moving towards the world of AI with enhanced cyber-awareness.

One of the prominent leverages in AI involves conditional generative adversarial networks (CGAN) and developing attack obfuscation strategies [16]. There are also learning capabilities developed across competing defensive and offensive AI strategies with the self-reinforcement technique. Even though AI promises stronger security, its rapid adoption can make systems more vulnerable because attackers can use AI to bypass defensive mechanisms [18]. Figure 1 illustrates the Trend showing the increasing interest of Google search involving 'cybersecurity' and 'AI' [15]. We shall examine the various applications of AI across diverse environments in the field of cybersecurity.

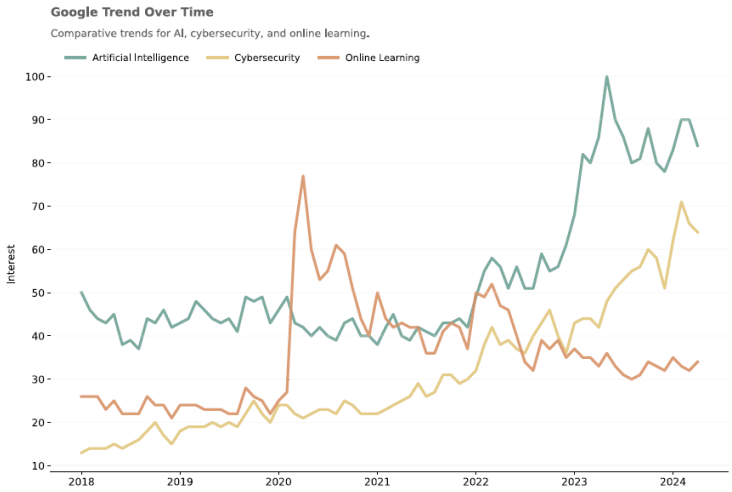

Fig1: Trend showing gaining interest of Google search involving 'cybersecurity' and 'AI' [15]

1.1 Intelligent cyber-evolution

Beginning with simple safeguards like encryption and passwords, which at first worked well but quickly proved insufficient as cyber threats got more complex, cybersecurity has undergone a dynamic evolution. As malware, viruses, and worms became more prevalent, security measures grew to incorporate intrusion detection systems, firewalls, and antivirus software [10]. Cyberattacks had become a major worldwide issue by the 2000s, affecting people, businesses, and even vital infrastructure. Rapid threat detection and behavior-based authentication have been made possible by AI and machine learning, which have transformed cybersecurity in recent years. However, as attackers use AI to create adaptive cyber threats, these technologies also present new concerns. In the future, post-quantum cryptography will be developed because of the significant threat posed by the emergence of quantum computing to the existing encryption techniques. Innovative and robust solutions will be essential as cybersecurity develops further to combat the constantly shifting threat landscape [4, 8, 9].

1.2 Framework of AI in cybersecurity

AI technologies come in many different forms, such as natural language processing, machine learning, deep learning, etc. It becomes important to set up a framework and scope for AI to work under the curtain to potentially regulate the threats caused by AI to humans. It is also significant that we need to formally define a framework that legislates the use of technology for what to explore its maximum potential of. Figure 2 shows the Framework for AI in cybersecurity [4].

Deep learning has become a powerful tool in cybersecurity, enabling the detection of intricate traffic anomalies and sophisticated phishing attempts that often evade traditional filters [5]. Beyond threat identification, AI-driven automation enhances cybersecurity by minimizing human intervention, reducing errors, and accelerating response times. Autonomous AI systems can apply security fixes, stopping criminal

activity and isolating infected networks, allowing security teams to focus on strategic initiatives. Once a threat is detected, AI executes predefined actions such as quarantining files or disabling user accounts, ensuring rapid containment and mitigating damage [25].
Furthermore, AI-driven behavioral analysis continuously monitors user and system activity to detect anomalies that may indicate emerging threats. Unlike traditional signature-based methods, AI identifies zero-day vulnerabilities and insider attacks by recognizing deviations from normal behavior, providing a proactive and adaptive security framework [4]. Even though AI has improved cybersecurity, integrating it presents several difficulties. Its susceptibility to adversarial assaults, in which manipulated inputs may deceive AI models and compromise security protocols, is a significant worry. For example, data poisoning taints training datasets to reduce a model's precision in identifying risks. Research on creating strong defenses against these kinds of assaults is still ongoing. Concerns about privacy and ethics also surface, especially in relation to algorithmic bias, transparency, and data protection [6]. For instance, AI-driven monitoring presents serious privacy concerns, underscoring the need for ethical frameworks that strike a balance between security, individual liberties, and public confidence [7].

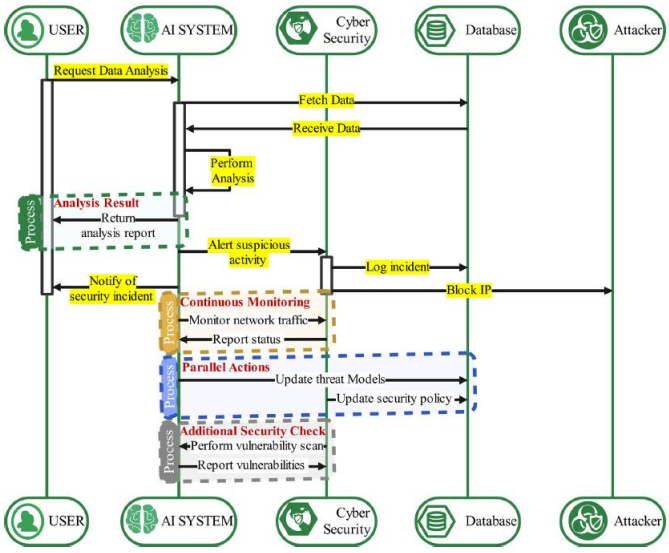

Fig 2: Framework for AI in cybersecurity [4]

2. Applications for AI in Cybersecurity:

Because AI tackles the increasing complexity and scope of contemporary cyber threats, its application in cybersecurity is essential. While AI can handle enormous volumes of data in real time, identify abnormalities, and react to threats more quickly than humans, traditional security systems frequently find it difficult to keep up with the sophistication and speed of assaults. Let's examine a few AI applications.

2.1 Reconnaissance

It is the first stage in ethical hacking that involves information gathering of the target. With the development of transformer models, the freely available reasoning search engines can be effectively integrated to make information gathering for a particular host feasible and easy. With more advancements in the search engines, it becomes easier to mitigate the hidden truths on the web [36,37].

2.2 Cybersafe smart cities

The cornerstone of smart city infrastructure is made up of information and communication technologies (ICTs), which enable services to be supplied seamlessly in all aspects of urban life. One of the primary benefits of AI-based security systems is their ability to manage the vast and diverse volumes of data that enter smart cities. Traditional security methods usually need help to keep up with the fast expansion of data and the increasing complexity of IoT networks. However, AI-powered systems can easily and swiftly collect and evaluate this data in real time, finding any risks and abnormalities that would otherwise go unnoticed. Machine learning algorithms can swiftly detect abnormalities that suggest a security breach by establishing common behavioral patterns through previous data training. [1,17]

2.3 Secure manufacturing systems

AI can be diligently used in securing manufacturing systems. These systems generate a lot of data that can be analyzed on a regular basis to be proactive in detecting threats and anomalies. Organization-wide implementation and maintenance, adverse training on models would be a great cause for the systems to respond to any unknown threats and attacks [2].

2.4 LLM in cybersecurity

Future cybersecurity frameworks that incorporate LLMs can take advantage of their capabilities to create stronger and more advanced defenses against changing cyber threats. Figure 3 shows Parameter minimization upon traditional fine-tuning [3]. This paper's strategic orientation seeks to steer future investigations and implementations, stressing the value of creativity and adaptability in preserving digital infrastructures.

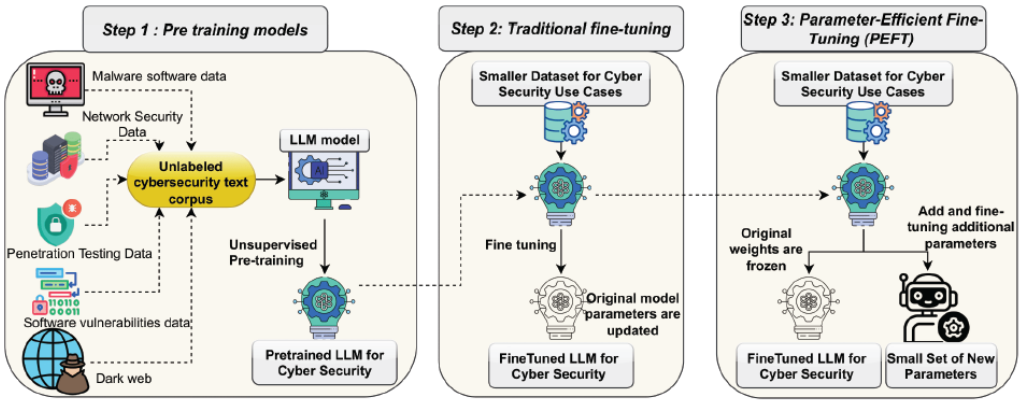

Fig 3: Parameter minimization upon traditional fine-tuning [3]

Figure 3 [3] describes a thorough three-stage procedure for training a sizable language model with a focus on cybersecurity. The first phase is unsupervised pre-training on a large corpus of cybersecurity literature, which covers a wide range of data such as malware, network security, and dark web stuff. The model is then subjected to traditional fine-tuning using a more concentrated, smaller dataset to improve its performance for certain cybersecurity tasks. To prepare the LLM to successfully meet the issues, the Parameter-Efficient Fine-Tuning (PEFT) approach entails freezing the initial model weights and fine-tuning a limited number of additional parameters. This improves the model's flexibility and efficiency while decreasing the likelihood of overfitting.

A practical advantage of PEFT in cybersecurity is its ability to quickly adapt large models to emerging threats without retraining the entire network. For instance, LoRA-based fine-tuning has been shown to achieve comparable accuracy to full fine-tuning while reducing GPU memory requirements by more than 60% [41]. In phishing detection tasks, PEFT enables models to incorporate small, recent datasets of malicious emails and generate timely defenses against evolving social engineering tactics. Similarly, in malware analysis, adapter-based PEFT methods allow security systems to recognize new obfuscation strategies with minimal computational overhead. Compared to alternative approaches such as prompt tuning or feature-based transfer learning, PEFT provides a balance of efficiency, robustness, and adaptability, making it a particularly valuable strategy for future cybersecurity frameworks.

2.5 Threat Intelligence and Critical Infrastructure

The following sectors are some of the major critical infrastructures that mandate threat intelligence. Threat refers to the automatic detection of threats and the successful mitigation of the same. These infrastructures, if disrupted, could cause a major loss to the human fraternity. Involving AI in safeguarding these infrastructures along with the existing cyber practices, would help enhance the process. It includes widespread usage of AI for monitoring, automation, recovery, and logging. Feedback-based learning could also aid in faster learning and response.

Figure 4 explores the possibility of an SIEM (security information and event management) architecture involving these dashboards that could be built to explore the nature of the incidents and enhance the faster learning rate of the model [22].

Gauhar et al. implemented behavior detection modules for early-stage detection of Advanced Persistent Threats (APTs) and enhanced the existing dataset with new alerts for malicious activities. To minimize false positive APT alarms, they developed a sophisticated machine learning-based prediction model, achieving an accuracy rate of 99.6% while significantly reducing the False Positive Rate (FPR). Additionally, they automated the incident response process, reducing manual intervention during post-alert decision-making by leveraging machine learning techniques for more efficient and accurate responses.

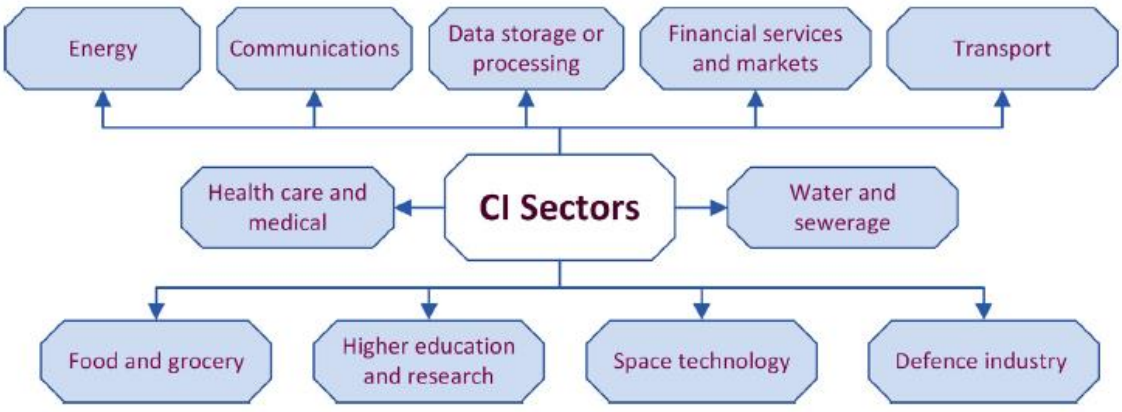

Fig 4: critical infrastructure sectors that AI could come in [21,41]

Explainable AI can also be trained to detect spam threats. It could stop a huge cyber-kill chain at the starting stage, as it could notify the victim timely manner [23]. Although several cybersecurity mitigation strategies, such as Artificial Intelligence (AI), the Internet of Things (IoT), blockchain, and quantum computing [24,26], have the potential to greatly improve the security of agricultural technology, especially those in agriculture and critical infrastructure, their application is fraught with difficulties. By protecting data, streamlining supply chains, these cutting-edge solutions aid in reducing cyberthreats. They do, however, carry certain hazards, such as the possibility of employee task overload, which might result in stress at work and unfavorable opinions about cybersecurity best practices. Figure 5 shows Security and event management [22]. The workforce may find it difficult to handle the heightened complexity and security requirements, which might

impede the adoption of new technologies and diminish their efficacy. This also explains the employee adoption in critical infrastructures [27,29]

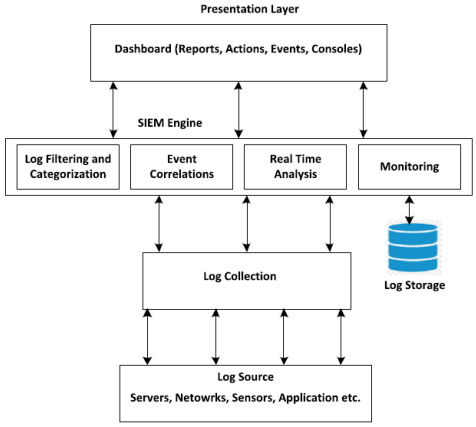

Fig 5: Security and event management [22]

2.6 Malware detection

Malware has a significant and varied effect on people, organizations, and society. Malicious software is frequently used in cybercriminal assaults with a variety of goals, such as stealing confidential information for financial gain, blocking access to computer systems, or putting ads and buy offers into websites [34]. While AI models, such as those used for malware detection, frequently excel in classification and prediction tasks, their inner workings are opaque. Particularly in high-stakes industries like healthcare, banking, and cybersecurity, this opacity raises questions about justice and dependability. Ensuring explainability, the capacity to be comprehended and trusted by humans, is just as vital as preserving good performance. Table 1 gives classification results for each dataset and model [35]. This is addressed by Explainable AI (XAI), which offers techniques and resources to make AI predictions and suggestions more understandable. XAI assists stakeholders in evaluating reliability, spotting biases or mistakes, and making defensible decisions based on AI insights by providing concise, understandable, and practical explanations. The studies also conclude the lack of framework evaluation metrics against XAI.
Different XAI approaches vary in their applicability to cybersecurity tasks. For example, feature attribution methods such as LIME and SHAP [34] provide local interpretability by highlighting input features that drive predictions, but they are computationally expensive and prone to instability in adversarial settings. In contrast, intrinsic XAI models such as decision trees or attention-based networks [35] offer more transparent structures but often sacrifice predictive accuracy when compared to deep learning methods. A critical limitation noted across studies is the absence of standard evaluation metrics to assess the reliability of explanations [36,37], making cross-model comparison difficult. This highlights the need for future research to balance interpretability, robustness, and performance, especially in high-stakes cybersecurity applications such as malware detection and intrusion response.

Table 1: Classification outcomes for each dataset and model [35]

Dataset	Model	Accuracy	Precision	Recall	F1
Mal-API-2019	LSTM	47.53	49.62	47.82	48.47
	BiLSTM	52.88	54.89	53.99	54.31
	GRU	47.69	49.10	48.62	48.27
	Attention	52.18	53.92	53.32	53.40
	MultiHeadAttention	47.69	49.10	48.62	48.27
API Call Sequences	LSTM	99.43	95.69	91.81	9k3.79
	BiLSTM	98.91	93.86	81.86	86.85
	GRU	98.91	91.62	84.30	87.49
	Attention	99.32	94.59	91.08	92.71
	MultiHeadAttention	98.99	92.87	84.92	88.38
Alibaba	LSTM	83.69	66.60	67.54	67.05
	BiLSTM	82.90	67.21	62.88	63.55
	GRU	84.09	72.02	65.20	65.81
	Attention	85.71	69.51	67.78	68.32
	MultiHeadAttention	84.72	67.23	66.85	66.47

Table 1 presents the classification outcomes for different malware detection models. Notably, the BiLSTM model on API Call Sequences achieved 98.91% accuracy, significantly outperforming its performance on the Mal-API-2019 dataset (52.88%). This discrepancy can be attributed to differences in feature representation: sequential API calls preserve temporal dependencies that BiLSTM architectures exploit effectively, whereas Mal-API-2019 offers a more heterogeneous feature space, limiting sequential learning benefits. In contrast, classical machine learning methods such as Random Forests perform better on structured, tabular datasets but struggle to capture long-range dependencies in behavioral logs. These results highlight that model performance is highly dataset-dependent, emphasizing the importance of selecting architectures aligned with the characteristics of the input data.

3. Cyber Attacks and AI

Cyber-attacks can be broadly classified into these three types. Figure 6 shows cyber attack types[12].

3.1 Network and Infrastructure Attacks

Network and infrastructure attacks are designed to disrupt the availability and integrity of services, often through overloading systems or intercepting communications. Distributed Denial of Service (DDoS) attacks attempt to overwhelm a network with traffic, leaving it inaccessible to normal users. Man-in-the-Middle (MitM) attacks are conducted by intercepting communications between two parties to steal or modify critical information. AI helps mitigate these threats by detecting anomalies in real-time. Machine learning monitors traffic blocks suspicious activity during DDoS attacks and identifies irregularities in encrypted data to prevent MitM attacks. AI also detects abnormal botnet patterns, reducing their spread and impact [12].

3.2 Data Breach and Unauthorized Access Attacks

Data breaches and unauthorized access attacks aim to compromise sensitive data or systems for malicious purposes. Phishing attacks are one of the most common ways attackers trick individuals into revealing sensitive information like login credentials and financial details. SQL injection exploits vulnerabilities in a website's database system, allowing attackers to inject malicious commands into queries and retrieve or manipulate confidential data. AI can analyze email content, detect phishing indicators, and flag suspicious messages. For SQL injection, machine learning algorithms can analyze query patterns and identify malicious input. In cases like APTs and BECs, AI-driven behavioral analysis can detect deviations from normal user behavior, signaling the presence of an attacker. These systems can identify insider threats or compromised accounts, minimizing the risk of data breaches.

3.3 Malicious Software and Exploits

Malicious software like ransomware, zero-day exploits, and crypto-jacking pose significant threats. Ransomware encrypts files for ransom, zero-day exploits target unpatched vulnerabilities, and crypto-jacking hijacks computing power for cryptocurrency mining. AI detects ransomware by identifying unusual encryption patterns, predicts zero-day exploits by analyzing software behavior, and spots crypto-jacking by monitoring abnormal CPU/GPU usage. Early detection allows quick responses, minimizing damage from these exploits.

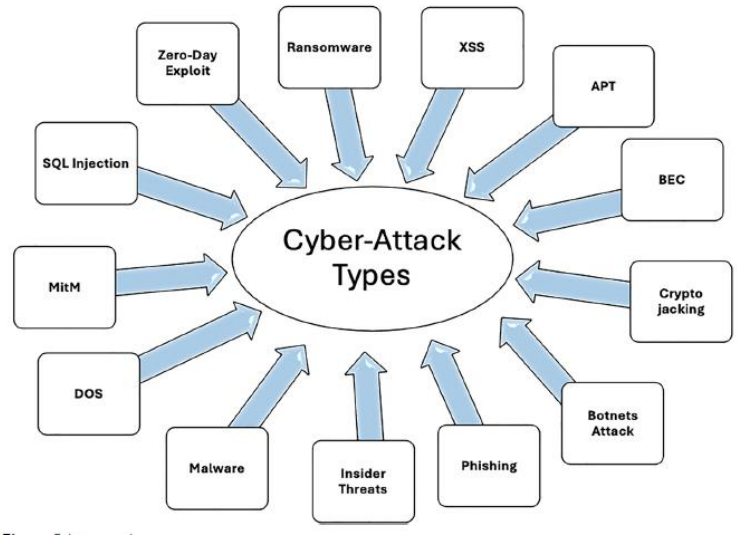

Fig. 6: Cyber attack types [12]

As shown in Figure 6, cyber-attacks encompass a wide spectrum ranging from social engineering threats (e.g., phishing, business email compromise) to technical exploits (e.g., SQL injection, XSS, zero-day attacks). AI plays a critical role in tailoring defenses to each category. For instance, natural language processing (NLP) models are effective in detecting phishing or BEC attempts by analyzing linguistic and semantic patterns, while anomaly detection algorithms are suited to spotting DoS or botnet traffic in high-volume network streams. Similarly, machine learning–based intrusion detection systems can help uncover APT activity by correlating weak signals across logs, whereas cryptojacking detection often relies on identifying abnormal CPU or GPU utilization. This highlights the importance of aligning AI techniques with specific threat classes rather than adopting a one-size-fits-all approach.

4. Considerable Contributions of AI in Cybersecurity

There are possible applications where AI can be considered, and its impact could be irreplaceable.

4.1 BCDR drills

AI could be a game-changer in implementing Business Continuity and Disaster Recovery (BCDR) drills. It could help in simulating server attacks, misconfigurations, and then recovery of the systems in minimal time. Thus, it reduces the business downtime and speeds up the resilience test.

4.2 Predictive maintenance of enterprise infrastructure

It follows 4.1, where the diagnostics come into place with threat response. The systems regulate and learn for future threats impending from past alerts, create an on-demand defense mechanism by altering the configurations, or suggest the same

4.3 Self-sustaining continuum

One possible edge case would be to create a 'self-sustaining' continuum that follows both threat response and recovery model. It includes predicting, diagnosing, alerting, monitoring, recovery, healing, and learning of the cyber lifecycle in the enterprise. With current limitations, achieving data privacy would be a real challenge.

4.4 Template security and AI

AI can also be combined with steganographic techniques, image analysis, feature, and template security images with a widespread scope for enhancement. It includes using AI for dataset building, reinforcement learning, and performing template evasion techniques standards to assert the secure nature of the template.
A practical example is the use of steganography in malicious templates, where attackers embed hidden payloads inside invoices, resumes, or image-based forms. Such payloads often bypass traditional signature-based detection because they are concealed within pixel values or metadata. AI-driven image analysis, particularly CNN-based steganalysis, has been applied to detect subtle statistical irregularities introduced during embedding. For instance, models trained on large corpora of clean versus tampered templates can identify hidden anomalies with high precision, improving resilience against document-based malware. However, adversaries have also begun exploring adversarial steganography, where hidden data is crafted specifically to fool AI detectors. This creates ongoing challenges in terms of dataset diversity, computational cost, and the need for adaptive retraining pipelines. Integrating such AI-based steganalysis into template security frameworks, therefore, offers promising protection, but its effectiveness depends on continuous updates and balancing detection accuracy with scalability.

4.5 Cyber monitoring

Cyber monitoring involves invoking AI agents at various stages of the cyber-kill chain lifecycle to enhance threat detection and response capabilities. This process includes the customized development of AI agents tailored for each phase of the kill chain, from initial reconnaissance to exploitation, installation, command and control, and ultimately the execution of the attack. These agents are designed to continuously learn and adapt to evolving threats, improving their detection accuracy and response times over time. By analyzing patterns of attack and identifying potential vulnerabilities, AI agents enable proactive measures, mitigating the risks posed by cyberattacks. As they are integrated into the cybersecurity infrastructure, AI-driven cyber monitoring systems can autonomously detect anomalies, isolate threats, and implement security measures, reducing the need for manual intervention and minimizing the impact of potential breaches.

4.6 Digital Forensics

Analyzing large volumes of data, such as system logs, network traffic, and digital storage devices, is a laborious and challenging process in traditional forensic investigations. The time needed for forensic investigation may be significantly decreased by using AI techniques to automate the identification of abnormalities, spot suspect behavior patterns, and rank evidence according to its importance.

5. Challenges and Limitations of AI in Cybersecurity

AI-driven cybersecurity solutions face several key challenges that hinder their full potential. One of the primary obstacles is the quality and availability of training data. AI models require vast amounts of diverse, high-quality datasets to accurately detect and mitigate threats. However, incomplete, biased, or poor-quality data can lead to ineffective AI models, causing false positives or negatives. Furthermore, the rise of adversarial attacks poses a significant threat, where malicious actors manipulate AI models by feeding them deceptive data, potentially bypassing detection systems. AI models are also frequently seen as "black boxes," making their decision-making processes difficult to explain, reducing trust and transparency in crucial cybersecurity applications [31,32].
Another challenge stems from the continuous need for AI systems to be updated and monitored to keep pace with evolving cyber threats. As cybersecurity threats rapidly change, AI models require frequent fine-tuning to remain effective. This demands significant resources, including specialized knowledge, to adapt AI-driven systems to new attack vectors. The requirement for human oversight remains another limitation, as while AI can automate certain processes, complex or novel security issues still necessitate human intervention. Furthermore, ethical issues about the use of AI in cybersecurity are increasing [24, 26]. Privacy problems, the possible exploitation of AI systems, and the collection of enormous volumes of sensitive data for training create serious questions regarding data protection and user rights.
Moreover, the open-sourcing of AI-trained models presents both opportunities and risks. While open-source models allow for greater collaboration and development, they also pose a security risk by making these models vulnerable to exploitation by malicious actors. In the context of information warfare, the ability to manipulate AI models for deceptive purposes or to influence the security of entire networks becomes a critical concern. The future of AI in cybersecurity is also influenced by quantum computing, which may render current encryption methods obsolete, leading to a need for the development of new cryptographic systems. While AI holds tremendous promise in strengthening cybersecurity measures, these challenges must be addressed for it to reach its full potential in protecting critical digital infrastructure [11,13].
Recent studies propose several defense strategies against adversarial attacks. Adversarial training [11] improves robustness by exposing models to perturbed samples, but it is resource-intensive and may not generalize well to unseen attack types. Defensive distillation and gradient masking reduce model sensitivity to input perturbations, yet they can be bypassed by adaptive attacks [16]. Other approaches, such as certified robustness, offer formal guarantees but remain computationally prohibitive for large-scale cybersecurity systems. A key gap across these defenses is the trade-off between robustness and efficiency: stronger defenses often come at the cost of higher latency, which may hinder real-time intrusion detection. This suggests that a hybrid strategy combining lightweight adversarial training with adaptive monitoring could be more effective in operational contexts. Recent approaches propose privacy-preserving AI techniques such as differential privacy, which introduces statistical noise to protect sensitive training data, and federated learning, which enables collaborative model training without sharing raw datasets. These methods are particularly relevant in domains like smart cities, where large volumes of citizen data are processed for security monitoring. Moreover, addressing algorithmic bias is essential, as imbalanced training data may lead to under-detection of threats originating from less-represented attack vectors. Embedding fairness auditing tools, explainability (XAI), and human-in-the-loop oversight into cybersecurity pipelines could help mitigate these risks. By combining ethical frameworks with technical safeguards, AI systems can be both effective and trustworthy in critical applications.

6. Advanced Cyber-Resilience and Weaponization

Advanced cyber-resilience refers to the capacity of an organization's digital infrastructure to not only withstand and recover from cyberattacks but to continuously function in the face of evolving threats. cyber-resilience involves the use of automation in incident response, which reduces human intervention and response times and aids in effective damage control. Another critical feature is the concept of "Fail-Safe" systems and redundant infrastructures that automatically redirect traffic, reroute services, or isolate compromised systems to prevent lateral movement of threats [19,20]. The efficacy of benign AI algorithms is seriously threatened by the weaponization of AI, which also makes more sophisticated assault scenarios possible in both digital and physical realms. Prioritizing national and international stability and well-being over political goals is essential to addressing the rising worry of AI-driven cyberattacks. AI has two sides to its use in cybersecurity: although it may strengthen defenses, it can also be abused for malevolent ends. Policymakers, engineers, researchers, and other stakeholders must work together and have an open discussion to combat the weaponization of AI and its possible abuse. Putting money into media and digital literacy builds sustainable solutions and increases social resilience [30].

7. Future Scope

The convergence of AI, blockchain, quantum computing, cloud, and edge computing is poised to redefine the technological landscape. While this integration unlocks vast potential, it also introduces complex challenges that demand focused research. Moving beyond theoretical synergy, the future scope involves tackling specific, interdisciplinary problems to ensure these technologies are scalable, secure, and ethical.
Building on the foundational synergies, the following concrete research questions and challenges emerge as critical directions for future work:

- **AI & Quantum Computing for Cybersecurity:**

  - Research Question: Can hybrid quantum-classical machine learning models significantly outperform classical AI in detecting zero-day exploits and advanced persistent threats (APTs) in real-time? What quantum algorithms are most suited for analyzing large-scale network traffic for anomaly detection?
  - Challenge: The current limitations of Noisy Intermediate-Scale Quantum (NISQ) hardware require research into error-resistant algorithms and efficient quantum-classical hybrid frameworks for practical security applications.

- **Blockchain for AI Integrity and Trust:**

  - Research Question: What are the scalability and throughput limitations of using permissioned blockchains to immutably log AI model training data, parameters, and decisions for auditability in critical systems (e.g., autonomous vehicles, financial forecasting)?
  - Challenge: Developing lightweight consensus mechanisms that can handle the high-volume, high-frequency data generated by AI systems without creating prohibitive latency or computational overhead.

- **Federated Learning (FL) for Distributed Security:**

  - Research Question: How can federated learning architectures be optimized for resource-constrained IoT devices (e.g., in smart cities or industrial IoT) to enable collaborative threat detection without compromising data privacy or draining device batteries?
  - Challenge: Mitigating security risks inherent to FL, such as model poisoning attacks from malicious participants, and developing techniques to ensure robust aggregation of models from heterogeneous and non-IID (Independent and Identically Distributed) data sources.

- **Post-Quantum Cryptography (PQC) in Distributed Systems:**

  - Research Question: Which PQC algorithms offer the best balance of security, performance, and key size for securing machine-to-machine (M2M) communication in edge computing environments and for protecting distributed ledger consensus mechanisms against quantum attacks?
  - Challenge: The integration of PQC into existing blockchain protocols and cloud-native encryption systems requires significant architectural changes and performance benchmarking to avoid system degradation.

- **Ethical & Regulatory Frameworks for Autonomous AI:**

  - Research Question: How can Decentralized Autonomous Organizations (DAOs) or smart contracts be designed to provide transparent oversight and enforce ethical constraints on self-learning AI systems operating in critical domains?
  - Challenge: Creating technically enforceable regulatory frameworks (RegTech) that can dynamically adapt to the evolution of AI models while ensuring accountability and aligning with human values.

Proactively addressing these specific questions will be paramount to developing the resilient security measures, scalable architectures, and ethical guidelines necessary to responsibly harness these transformative technologies and build a secure and intelligent digital future.

8. Conclusion

Thus, this study has been concluded by highlighting the importance of AI in cybersecurity. In the current era, AI is increasingly capable of outperforming human capabilities in various cybersecurity tasks. From the development of AI agents that can autonomously detect and respond to threats to reinforcement learning for continuous improvement of defense mechanisms, AI enhances the ability to prevent and

mitigate cyberattacks. Additionally, AI aids in the timely response and analysis of vast amounts of data, offering insights through explainable AI systems that help users understand the findings and take informed actions. AI can simulate attack vectors, identify vulnerabilities, and mitigate payloads, thus strengthening the overall security posture of organizations. Despite its potential, the implementation of AI in cybersecurity remains in a relatively nascent stage, requiring further refinement and understanding. As AI continues to evolve, its integration into cybersecurity must be accompanied by the establishment of solid frameworks and ethical considerations. Key challenges include ensuring the robustness and accuracy of AI models, reducing biases, and addressing privacy concerns. The success of AI-driven solutions is partly determined by the quality of the data used for training and their capacity to react to new, emerging threats. Therefore, while AI offers tremendous promise, its successful implementation requires continuous research, innovation, and careful planning to maximize its impact while minimizing potential risks.

References

[1] Ali J, Singh SK, Jiang W, Alenezi AM, Islam M, Daradkeh YI, Mehmood A., A deep dive into cybersecurity solutions for AI-driven IoT-enabled smart cities in advanced communication networks, Computer Communications 229(2024) ,doi:10.1016/j.comcom.2024.108000

[2] Alqudhaibi A, Albarrak M, Jagtap S, Williams N, Salonitis K., Securing industry 4.0: Assessing cybersecurity challenges and proposing strategies for manufacturing management, Cyber Security and Applications. 3(2025), doi.org/10.1016/j.csa.2024.100067.

[3] Ferrag MA, Alwahedi F, Battah A, Cherif B, Mechri A, Tihanyi N, Bisztray T, Debbah M. Generative AI in Cybersecurity: A Comprehensive Review of LLM Applications and Vulnerabilities. Internet of Things and Cyber-Physical Systems, 5(2025) 1:46.

[4] Ali S, Wang J, Leung VC., AI-driven fusion with cybersecurity: Exploring current trends, advanced techniques, future directions, and policy implications for evolving paradigms-A comprehensive review, Information Fusion, 118(2025).

[5] Manoharan A, Sarker M., Revolutionizing Cybersecurity: Unleashing the Power of Artificial Intelligence and Machine Learning for Next-Generation Threat Detection,1(2023), doi. org/10.56726/IRJMETS32644..

[6] Nadella GS, Gonaygunta H., Enhancing Cybersecurity with Artificial Intelligence: Predictive Techniques and Challenges in the Age of IoT, International Journal of Science and Engineering Applications. 2024,13(04):30-33.

[7] Familoni BT., Cybersecurity challenges in the age of AI: theoretical approaches and practical solutions, Computer Science & IT Research Journal, 2024, 5(3), 703-724.

[8] Kumar S, Gupta U, Singh AK, Singh AK., Artificial intelligence: revolutionizing cyber security in the digital era, Journal of Computers, Mechanical and Management, 2023, 2(3), 31-42.

[9] Dunn Cavelty M, Wenger A., Cyber security meets security politics: Complex technology, fragmented politics, and networked science, Contemporary Security Policy, 202041(1), 5-32.

[10] Welukar JN, Bajoria GP., Artificial Intelligence in Cyber Security-A Review, International Journal of Scientific Research in Science and Technology, 488(2021): 488-91.

[11] Kaur R, Gabrijelcic D, Klobucar T. Artificial intelligence for cybersecurity: Literature review and future research directions. Information Fusion. 97(2023), doi.org/10.1016/j.inffus.2023.101804

[12] Salem AH, Azzam SM, Emam OE, Abohany AA., Advancing cybersecurity: a comprehensive review of AI-driven detection techniques, Journal of Big Data, 2024 11(105), doi.org/10.1186/s40537-024-00957-y

[13] Jada I, Mayayise TO., The impact of artificial intelligence on organisational cyber security: An outcome of a systematic literature review, Data and Information Management. 2024, 8(2): doi.org/10.1016/j.dim.2023.100063.

[14] Rampasek M, Mesarcik M, Andrasko J., Evolving cybersecurity of AI-featured digital products and services: Rise of standardisation and certification?, Computer Law & Security Review, 2025 56:106093.

[15] Parambil MMA, Rustamov J, Ahmed SG, Rustamov Z, Awad AI, Zaki N, Alnajjar F., Integrating AI-based and conventional cybersecurity measures into online higher education settings: Challenges, opportunities, and prospects.,Computers and Education: Artificial Intelligence, 7(2024): doi.org/10.1016/j.caeai.2024.100327

[16] Coppolino L, D'Antonio S, Mazzeo G, Uccello F, The good, the bad, and the algorithm: The impact of generative AI on cybersecurity, Neurocomputing, 623(2025): doi 10.1016/j.neucom.2025.129406

[17] Zeng H, Yunis M, Khalil A, Mirza N., Towards a conceptual framework for AI-driven anomaly detection in smart city IoT networks for enhanced cybersecurity, Journal of Innovation and Knowledge. 2024, 9(4):100601.

[18] Vasalou A, Benton L, Serta A, Gauthier A, Besevli C, Turner S, Gill R, Payler R, Roesch E, McAreavey K, Bauters K., Doing cybersecurity at home: a human-centred approach for mitigating attacks in AI-enabled home devices, Computers and Security,148(2025):104112.

[19] Sarker IH, Janicke H, Mohsin A, Gill A, Maglaras L., Explainable AI for cybersecurity automation, intelligence and trustworthiness in digital twin: Methods, taxonomy, challenges and prospects, ICT Express, 2024,10(4), 935-958.

[20] Muheidat F, Mallouh MA, Al-Saleh O, Al-Khasawneh O, Loai AT., Applying AI and Machine Learning to Enhance Automated Cybersecurity and Network Threat Identification, Procedia Computer Science, 251(2024):287-94.

[21] Sarker IH, Janicke H, Ferrag MA, Abuadbba A., Multi-aspect rule-based AI: Methods, taxonomy, challenges and directions toward automation, intelligence and transparent cybersecurity modeling for critical infrastructures, Internet of Things. 2024, 25(2), doi:10.1016/j.iot.2024.101110

[22] Ali G, Shah S, ElAffendi M., Enhancing cybersecurity incident response: AI-driven optimization for strengthened advanced persistent threat detection, Results in Engineering, 25(2025):104078.

[23] Filali A, Sallah A, Hajhouj M, Hessane A, Merras M. Towards Transparent Cybersecurity: The Role of Explainable AI in Mitigating Spam Threats. Procedia Computer Science. 236(2024), 394-401.

[24] Karki S, Hasan AM, Sanin C, Use of ML and AI in Cybersecurity-A Survey, Procedia Computer Science. 246(2024):1260-1270.

[25] Kumar R, Aljuhani A, Javeed D, Kumar P, Islam S, Islam AN, Digital twins-enabled zero touch network: A smart contract and explainable AI integrated cybersecurity framework, Future Generation Computer Systems, 156(2024):191-205.

[26] Khan K, Khurshid A, Cifuentes-Faura J., Is artificial intelligence a new battleground for cybersecurity?., Internet of Things, 28(2024):101428.

[27] Maraveas C, Rajarajan M, Arvanitis KD, Vatsanidou A., Cybersecurity threats and mitigation measures in agriculture 4.0 and 5.0., Smart Agricultural Technology. 9(2024), doi.org/10.1016/j.atech.2024.100616.

[28] Saleh AM., Blockchain for secure and decentralized artificial intelligence in cybersecurity: A comprehensive review, Blockchain: Research and Applications. 2024 doi:10.1016/j.bcra.2024.100193

[29] Campoverde-Molina M, Luján-Mora S., Cybersecurity in smart agriculture: A systematic literature review, Computers and Security, 150(2024): doi.org/10.1016/j.cose.2024.104284.

[30] Nobles C., The weaponization of artificial intelligence in cybersecurity: A systematic review, Procedia Computer Science. 239(2024):547-555.

[31] Chaudhuri A, Behera RK, Bala PK., Factors impacting cybersecurity transformation: An Industry 5.0 perspective, Computers and Security, 150 (2025):104267.

[32] Michael K, Vogel KM, Pitt J, Zafeirakopoulos M., Artificial intelligence in cybersecurity: A socio-technical framing, IEEE Transactions on Technology and Society. 2025,6(1):15-30.

[33] Alhamdi MJ, Lopez-Guede JM, AlQaryouti J, Rahebi J, Zulueta E, Fernandez-Gamiz U. AI-based Malware Detection in IoT Networks within Smart Cities: A Survey. Computer Communications. 253(2025):108055.

[34] Baghirov E., A comprehensive investigation into robust malware detection with explainable AI. Cyber Security and Applications. 3(2025): doi.org/10.1016/j.csa.2024.100072.

[35] Galli A, La Gatta V, Moscato V, Postiglione M, Sperlì G., Explainability in AI-based behavioral malware detection systems, Computers and Security.141(2024): doi.org/10.1016/j.cose.2024.103842.
[36] Raman R, Calyam P, Achuthan K., ChatGPT or Bard: Who is a better Certified Ethical Hacker?. Computers & Security, 140(2024): doi.org/10.1016/j.cose.2024.103804.
[37] He Y, Zamani E, Yevseyeva I, Luo C. Artificial intelligence–based ethical hacking for health information systems: simulation study. Journal of medical Internet research. 25(2023): doi: 10.2196/41748.
[38] Akhunzada A, Al-Shamayleh AS, Zeadally S, Almogren A, Abu-Shareha AA, Design and performance of an AI-enabled threat intelligence framework for IoT-enabled autonomous vehicles, Computers and Electrical Engineering, 119(2024):109609.
[39] Hassan A, Nizam-Uddin N, Quddus A, Hassan SR, Rehman AU, Bharany S., Navigating IoT Security: Insights into Architecture, Key Security Features, Attacks, Current Challenges and AI-Driven Solutions Shaping the Future of Connectivity, Computers, Materials and Continua. 2024 81(3): 3499-3559
[40] Cyber and infrastructure security centre, department of home affairs, 2023, Australian Government https://www.homeaffairs.gov.au/. (Accessed: 20 July 2024).
[41] Hu, E.J., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S. and Chen, W., LoRA: Low-rank adaptation of large language models, (2021), doi.org/10.48550/arxiv.2106.09685.