AI in Cybersecurity: A Literature review

Dr. Neha Yadav[1], Arooj Naimuddin Khan[2], Ayesha Naimuddin Khan[3], Diya Pratap[4]

[1] *Department of Artificial Intelligence & Machine Learning, Dr. Akhilesh Das Gupta Institute of Professional Studies*
[2] *Department of Artificial Intelligence & Machine Learning, Dr. Akhilesh Das Gupta Institute of Professional Studies*
[3] *Department of Artificial Intelligence & Machine Learning, Dr. Akhilesh Das Gupta Institute of Professional Studies*
[4] *Department of Artificial Intelligence & Machine Learning, Dr. Akhilesh Das Gupta Institute of Professional Studies*

---***---

Abstract - In cybersecurity, artificial intelligence is revolutionizing incident response, risk management, and threat identification in a progressively hostile cyber threat landscape. This research presents a thorough literature review on AI in cybersecurity, focusing on both aspects of the balance sheet. This paper discusses how AI-driven technologies like machine learning, deep learning, natural language processing, and expert systems improve security frameworks through predictive analytics, real-time threat intelligence, and anomaly detection. The research explores various uses of AI such as network protection, cloud safety, healthcare, and finance, highlighting how AI-driven solutions enhance the resilience of cybersecurity against attacks. Nonetheless, there are drawbacks too, primarily associated with algorithmic prejudices and aggressive attacks. This paper discusses AI cybersecurity tools like Cylance, Darktrace, and IBM Watson, analyzing their influences and effects on security operations. The study also explores recent advancements and enhancements in AI-driven cybersecurity, ethical concerns, and regulatory structures. To establish a safe digital space, this document highlights the importance of a unified strategy that integrates AI with human expertise, ethics, and regulatory adherence.

***Key Words*:** Artificial Intelligence, Machine Learning, Deep Learning, Expert Systems, Natural Language Processing, Threat Detection, AI-driven Threat Intelligence, Dynamic Threats, Cyber Attack, Cyber Security.

1. INTRODUCTION

In a period characterized by technological advancement and interconnected systems, reliance on cyberspace for individual, commercial, and governmental functions has achieved an unmatched degree. As cyber threats advance, conventional security measures find it challenging to keep up with increasingly sophisticated attack methods [9]. The area of cybersecurity has progressed notably in recent decades, propelled by the swift growth of internet and digital technologies [7]. Originally, cybersecurity emphasized fundamental protective strategies such as firewalls and antivirus programs, aimed at safeguarding against recognized threats. Nonetheless, as cyberattacks grow increasingly sophisticated and intricate, cybersecurity systems are required to evolve [7].

Artificial Intelligence (AI) has assisted numerous organizations in enhancing their security posture efficiently and minimizing breach risks [2]. Machine learning, especially deep learning within artificial intelligence, enhances the ability to analyze substantial data sets, identify patterns, and anticipate potential threats in real-time [7]. In contrast to conventional signature-based detection, AI-powered systems analyze data patterns to promptly identify anomalies and threats (Introduction 3). Machine learning algorithms can detect concealed patterns, zero-day vulnerabilities, and breach-associated actions by analyzing extensive data sets (3). AI models assess user and entity actions to establish a standard for typical behavior in behavioral analysis [3]. The swift technological progress of the digital era has led to new opportunities as well as difficult obstacles [3]. Safeguarding against cyber threats can solely be achieved through advanced technologies [8]. Integrating AI into cybersecurity involves not only protecting boundaries but also evaluating asset worth, recognizing threats, and ranking responses to establish a defense-in-depth strategy that guarantees service continuity [1]. AI enhances cyber resilience, but it can simultaneously be utilized for penetration [1]. Studies are ongoing to investigate the possibilities and dangers of AI in cybersecurity, highlighting the necessity for a balanced strategy [1]. Combining the power of AI with ethical standards, privacy issues, and potential algorithmic biases is crucial for creating a safe and fair digital landscape [9]. AI alone cannot resolve all cybersecurity issues; it needs to be integrated into a comprehensive security strategy that includes human expertise, effective policies, and ongoing monitoring [3]. Adopting AI's potential is essential for strategy and crucial for strengthening the digital protections that support our interconnected globe [9].

2. KEY AI TECHNOLOGIES IN CYBERSECURITY

2.1 Machine Learning (ML)

Machine learning models enable machines to offer decisions or predictions based on the mathematical representation of organized data [1]. Machine Learning (ML) is a crucial component of AI, where algorithms are created to learn from data. As these algorithms are subjected to more data over time, they enhance their capacity to make predictions or decisions, a process that resembles human learning [5]. It constitutes a subset of AI that allows systems to learn from data, recognize patterns, and make choices without being specifically programmed. Common methods include supervised learning, unsupervised learning, and reinforcement learning [7].

Supervised Learning: Supervised learning is a category of ML where the algorithm learns from labelled data, meaning it is provided with inputs along with corresponding outputs during training processes [12]. It aids in recognizing patterns and making predictions. Supervised learning is applied for malware detection and spam filtering tasks in cybersecurity. Unsupervised Learning: Identifies patterns without labelled data, useful for anomaly detection. Unsupervised learning can be utilized to recognize unusual network behaviour that may signify a cyberattack [9].Reinforcement Learning: In reinforcement learning, AI agents learn through interaction with their environment and receive feedback in the form of rewards or penalties. This method is increasingly being implemented in cybersecurity to develop AI systems capable of learning optimal reactions to dynamic and evolving threats [7].

2.2 Deep Learning

An area of machine learning that utilizes neural networks with several layers is typically referred to as deep learning. Deep reinforcement learning (DRL or deep RL) is a subset of machine learning and artificial intelligence in which intelligent robots can learn from their experiences similarly to how humans learn. It combines reinforcement learning (RL) techniques like Q-learning with deep learning concepts, including neural network learning [14]. Deep learning excels at handling vast amounts of data and is applied in tasks such as image and voice recognition. In the realm of cybersecurity, it can be used for advanced threat identification and evaluation [9]. Deep Learning (DL) is particularly adept at detecting complex and intricate attack patterns, owing to its capacity to handle large volumes of data and create high level models [7]. The fundamental architecture of deep learning includes an input layer, one or more hidden layers, and an output layer. [1] The hidden layers in artificial neural networks (ANN) can comprise numerous layers to train the network for adapting to complex patterns and delivering efficient analysis based on correlated data [1].

The prevalent structure in deep learning, Convolutional Neural Network (CNN), has gained significant acclaim for its effectiveness and achieved remarkable success through the idea of deep learning [1]. Initially created for image recognition, CNNs are now extensively utilized in cybersecurity to identify threats across different types of data. CNNs are particularly effective at handling grid-structured data, like patterns in network traffic or visual forms of malware code [7].

The constraints of the CNN were understood and addressed by the Recurrent Neural Network (RNN), which can retrieve the information lost in hidden layers through the principle of gradient descent, utilizing loops within the network [1]. RNNs are created to handle sequential data and are especially effective for forecasting threats in cybersecurity. These networks excel at evaluating time-series data, like logs of network activities, to identify attacks throughout time. RNNs can detect abnormal event sequences that could indicate the initial phases of a cyberattack, like numerous failed login attempts or dubious file access behaviors that occur before a ransomware attack [7].

2.3 Natural Language Processing

A sector of AI is dedicated to communication between computers and human languages. It allows machines to comprehend, analyze, and produce human language in a manner that is significant and contextually appropriate [7]. The vast amount of information available on social media and websites spurred the examination and interpretation of unstructured text data to uncover current trends in cyber threats and risks stemming from social media sentiments. NLP models seek to identify and separate threat reports and malware code insertions from text analysis. It assists in analyzing and retrieving pertinent details to support the recognition of incidents from security logs [1]. Recognizing phishing emails is a common issue to prevent the compromise of personal information, such as banking details, social security numbers, and user passwords. Phishing attacks pose a serious risk, and NLP is essential for their detection can be combined with these ML and DL models for effective classification of email contents to accurately detect phishing attacks. Drawing from the cognitive analysis conducted by ML models, NLP facilitated the creation of domain ontologies through a dual approach: symmetry stage and adjusted machine stage. Utilizing these ontologies, a prototype termed a cybersecurity analyzer was developed to yield a comprehensive architecture featuring critical components and an implementation initializer. NLP can further enhance cybersecurity operations by automatically handling and responding to text-based information like security reports, user questions, or threat intelligence updates. For example, chatbots or virtual assistants driven by NLP can automatically assess logs or user inquiries to detect possible threats and trigger predetermined responses. This can significantly lighten the burden on cybersecurity experts and accelerate the reaction to new threats [7]. Cyberattack Sensing and Information Extraction (CASIE) was introduced to improve the effectiveness of NLP in detecting cyber threats from text, as the system was designed to update the public about cyber incidents via knowledge-based graph visualization in various online articles**.**

2.4 Expert Systems

As a subset of AI, Expert Systems are computer-driven systems that replicate problem-solving and automated decision-making procedures by integrating expert human knowledge systems with rule-based inference engines [1]. An expert program is technology designed to find solutions to issues presented by either a client or a specific technology within a particular technological domain. This could be utilized in virtual environments [8]. The advisory functions of Expert System enhance the system's resilience and strength regarding cybersecurity. Rule-based IDS serve as a typical illustration of this idea, where established rules are applied in the detection mechanism to observe atypical activities within the network or system [1]. These guidelines or heuristics assist in detecting signs of dubious patterns in system logs for offline security services and in real-time traffic monitoring systems to notify about irregularities. Rules and decision-tree-based IDS, termed RDTIDS, was introduced to differentiate between malicious and benign network traffic by utilizing an integrated rule-based expert system and decision tree methodology [1]. A different all-inclusive rule-based IDS utilizing Java ExS Shell (JESS) offers guidelines for the Pattern-based Intrusion Detection Engine (PIDE) to evaluate user 4 interactions within the system [1].

Expert systems utilize different types of representations, with stabilizers interpretation being the most general one. Artificial intelligence could offer extra features for simulation [8]. Expert systems can be advantageous in cybersecurity management by offering effective oversight of the large volume of traffic [1]. The expert system is essential in cybersecurity, utilizing specialized knowledge and rule-based reasoning to improve security operations, threat identification, incident management, and compliance oversight [1].

3. METHODOLOGY

Artificial Intelligence (AI) is currently employed in, or is actively being investigated for, several aspects of cyber security solutions. For instance, Gmail utilizes AI to detect and block unwanted spam and phish emails. Gmail's Artificial Intelligence learned from the millions of existing users - each time users click on an email or mark it as not spam, they contribute to training the AI to identify spam in the future.

3.1 Enhanced threat detection

The initial step in protecting your organization's network is identifying threats. It would be perfect if you could quickly identify problems such as unreliable data. It will safeguard your network from lasting damage. Combining AI with cyber security is the most effective way to identify and address attacks instantly [2]. AI has the capability to learn and identify common behaviors of users and systems, which helps in detecting unusual or questionable activities that may indicate a possible security breach [9]. It rapidly identifies patterns, atypical behaviors, and emerging threats that might evade conventional methods. It detects intricate attacks sooner and can autonomously react by preventing harmful traffic and issuing alerts. This aids in saving time and reduces the risk of data breaches. A fraud detection In system utilizing Artificial Intelligence, which uses algorithms cantered on anticipated consumer behaviors to detect fraudulent transactions via MasterCard's Decision Intelligence. It analyzes the typical buying behaviors of the customer, the vendor, the transaction site, and various other intricate algorithms to assess whether a purchase is atypical [4]. Artificial intelligence can be employed to identify fraudulent activities in financial transactions, including money laundering or credit card fraud. AI algorithms can identify suspicious activities in transaction data by examining patterns and irregularities. This can assist in preventing financial losses and protecting individuals and organizations against fraud [6]. The capacity of AI in an environment filled with vast amounts of unstructured data enhances the precision of threat identification, minimizes the need for manual efforts, and illustrates the potential of AI in thwarting advanced cyber-attacks [1].

3.2 Anomaly detection

Anomaly detection is an essential component of cybersecurity aimed at recognizing departures from established normal behavior within systems. The procedure encompasses various essential methodologies and frameworks, mainly statistical, to efficiently recognize and reduce potential dangers. Anomaly detection recognizes departures from standard behavior, which is determined by organizational needs and resilience approaches. Multiple statistical models are utilized to detect patterns of attack vectors. These models aid in differentiating between typical and irregular patterns, frequently employing a blended methodology that integrates various techniques for improved precision. These instruments are essential for tackling cybersecurity issues across various industries [1]. The integration of AI improves anomaly detection abilities by facilitating predictive modeling and decreasing incident response durations [1]. AI methods can examine typical network activity patterns, user actions, and system operations. Divergence from these patterns may be recognized as possible risks [9].

3.3 Scalability

Artificial intelligence can be expanded to analyze vast quantities of data and operate in real-time, making it an exceptional resource for assessing large, intricate data sets and identifying potential risks. Scalability enables AI to function concurrently with additional AI systems. AI can be utilized to analyze historical data and reveal patterns that could help predict future cyberattacks. This kind of analysis is referred to as "predictive analytics." Predictive analytics utilizes past data to anticipate possible future risks. Through the application of machine learning to examine patterns, AI systems are able to predict cyberattacks prior to their happening. This proactive strategy allows organizations to implement preventive security measures instead of reactive ones, decreasing the chances of breaches [7]. AI enables companies to recognize potential risks and implement proactive measures to avert them by analyzing data from previous attacks. This can be achieved by analyzing information from earlier attacks [6].

3.4 Automation of Cybersecurity processes

Artificial intelligence can enhance numerous functions in cybersecurity, including performing vulnerability assessments, overseeing patch management, and responding to incidents. This ability can assist organizations in reducing the strain on their security staff while enhancing their overall security stance. Particular tasks that gain from AI encompass vulnerability evaluations, patch management, and response to incidents. Additionally, AI can reduce the likelihood of human mistakes, a major factor in cybersecurity incidents. Through the examination of information from various origins such as threat intelligence feeds, social media platforms, and dark web discussions, AI can swiftly detect possible threats. This kind of information is referred to as "real-time threat intelligence." Such insights can help organizations sidestep emerging threats and respond swiftly to new attack vectors, vital for preserving security.

3.5 High data analysis and capacity

Artificial Intelligence Can Efficiently Manage Massive Data Collections. The network of a company is presently undergoing a significant amount of activity. Generally, a mid-sized company experiences considerable traffic, signifying that a large volume of information is communicated daily between customers and the business. This data must be protected from harmful people and programs. Nonetheless, cybersecurity experts cannot examine every piece of data for possible threats. AI is considered the optimal choice for identifying threats that might disguise themselves as normal behaviors. Its automated features enable it to effectively examine vast amounts of data and traffic. Technologies powered by AI, like personal proxies, enable data transfer while recognizing and distinguishing potential dangers hidden within the disorder. Moreover, AI can analyze and examine large amounts of security data in real-time, uncovering essential insights and detecting patterns that may suggest potential threats. Incorporating AI in cybersecurity improves threat detection's efficiency and precision, allowing organizations to adjust to the changing nature of cyber threats. With the increasing complexity of cyber threats, artificial intelligence will continue to be essential in enhancing security and protecting digital assets.

3.6 Intelligent Password Management

Artificial intelligence can be employed to enable biometric authentication techniques like fingerprint identification, voice identification, and facial identification, making it harder for unauthorized persons to access. Additionally, AI can greatly enhance identity and access management (IAM) systems by examining user behavior and detecting anomalies that could suggest unauthorized access or compromised accounts. This ability enables organizations to identify and react to possible security threats instantly, thus enhancing their overall IAM security structure.

Furthermore, AI can examine password trends and detect weak passwords that are vulnerable to guessing or hacking, thus improving password security. It can likewise detect efforts to utilize compromised passwords and encourage users to alter them. An instance of this application is PassGAN, which employs a Generative Adversarial Network to autonomously understand the distribution of genuine passwords from authentic data breaches, removing the necessity for manual password examination.

3.7 Cost Effective

Artificial intelligence (AI) could serve as a budget-friendly answer for cybersecurity as it can automate various tasks that would otherwise require substantial resources if performed manually. For example, artificial intelligence can streamline the examination of malware, a task that is frequently tedious and costly when done manually. AI can continuously monitor networks and systems, assisting businesses in identifying risks in real-time and reacting accordingly. This form of oversight is referred to as "continuous monitoring." This can aid in reducing the duration between a cyberattack happening and the organization detecting the attack and responding to it [6].

3.8 Incident Response & Mitigation

Incident response and mitigation are essential elements of cybersecurity, centering on the measures implemented after a security event is identified to assess, address, and avert future weaknesses. The identification of adverse events is aided by different monitoring methods that recognize risks. When a threat is identified, it is essential to assess the circumstances to comprehend the type of vulnerability and develop an appropriate reaction. Artificial Intelligence (AI) significantly contributes to improving incident response abilities. AI models, including statistical decision trees and neural networks, are utilized to enhance decision-making processes in the event of incidents. They facilitate automated reactions to unusual occurrences, thus enhancing the speed and effectiveness of incident management [1]. AI can aid in quickly triaging security incidents by evaluating the seriousness and significance of alerts, enabling security teams to prioritize their response actions. AI is capable of examining extensive datasets to recreate the sequence of events in a security incident, which assists in investigations after the incident [9]. AI enhances endpoint security by continuously monitoring and evaluating endpoint devices for possible harmful activities. This proactive strategy aids in the prompt detection and elimination of vulnerabilities, minimizing the overall effect of threats throughout the network. AI technologies facilitate adaptive decision-making processes, enabling organizations to react flexibly to changing threats. This encompasses the deployment of smart decision support systems that use historical information and sensor data to aid in threat prevention and risk management. The incorporation of AI in incident response not only assists in prompt threat alleviation but also contributes to understanding previous incidents to enhance future defenses. With the rise of sophisticated technologies used by attackers, it is crucial for defenses to adapt accordingly [1].

4. AI-POWERED CYBERSECURITY SOLUTIONS

4.1 Darktrace

Darktrace is a cybersecurity firm utilizing AI that employs machine learning to identify and respond to cyber threats instantly [9]. The AI algorithms created by Darktrace analyse network traffic, user behaviour, and system activities to identify possible security vulnerabilities and irregularities [6]. When a possible danger is detected, the system can automatically initiate protective measures, like isolating compromised systems or preventing harmful traffic. This AI-driven autonomous response significantly diminishes the time required to counter attacks, enhancing the overall security stance of organizations [7]. Another feature of Darktrace's AI is the automation of incident response, enabling businesses to respond swiftly and efficiently to potential threats. One unique feature that distinguishes Darktrace from other network monitoring software is its ability to comprehend the anticipated behaviour of an organization's network and systems over time. As a result, Darktrace is able to identify any anomalies or deviations from the norm that might indicate the existence of a potential threat. Darktrace has demonstrated its efficacy in detecting and reacting to a range of cyber threats, such as insider threats, ransomware incidents, and zero-day vulnerabilities [6]. Darktrace's AI system continuously adapts to the typical behaviour of a network and can detect irregularities that suggest possible security threats. It subsequently reacts to these dangers independently, reducing risks as they occur. This forward-thinking strategy assists organizations in evading new threats and reducing the effects of cyberattacks [9]. As an illustration, in a mock ransomware incident, Darktrace effectively identified the attack at its initial phase and stopped the encryption of vital files, thereby averting data loss [7].

4.2 Cylance

Cylance represents a cutting-edge artificial intelligence solution for endpoint security, employing machine learning algorithms to detect and mitigate malware threats. The AI-driven algorithms of Cylance analyse the behaviour of files and applications on endpoints to identify potential malware risks. By leveraging artificial intelligence to discern patterns and behaviours linked to malware, Cylance can proactively identify and neutralize threats before they are activated. This methodology significantly bolsters cybersecurity by establishing a pre-emptive defence against both emerging and unidentified threats. Additionally, Cylance's AI streamlines threat management, allowing organizations to react swiftly and efficiently to potential dangers. A notable aspect of Cylance is its capacity to foresee and thwart both novel and recently identified malware variants. The artificial intelligence algorithms utilized by Cylance can detect potential malware threats by observing the behaviour of infections, regardless of whether the specific malware has been previously encountered. Cylance has proven its efficacy in identifying and blocking various malware attacks, including advanced persistent threats and zero-day exploits.

4.3 IBM Watson

IBM Watson for Cyber Security integrates artificial intelligence with cognitive technologies to evaluate security information and deliver insights to cybersecurity professionals. This solution employs natural language processing and machine learning techniques to analyse security data from diverse sources. By enabling security analysts to efficiently sift through vast quantities of data, IBM Watson for Cyber Security aids in the identification of potential threats. Through the examination of aggregated security information, it can recognize possible threats and recommend appropriate countermeasures. A key feature that sets IBM Watson for Cyber Security apart from similar offerings is its capability to handle extensive volumes of security-related data from various origins, such as security alerts, threat intelligence feeds, and historical records. Furthermore, it can respond to inquiries from security analysts in natural language, thereby providing them with enhanced information and insights regarding potential threats. Instances of cyberattacks, including phishing, ransomware, and advanced persistent threats (APTs), have been effectively detected and mitigated using IBM Watson for Cyber Security. The incorporation of AI not only expedites incident response times but also improves the overall effectiveness of cybersecurity operations.

5.CHALLENGES AND THREATS

5.1 False positives and negatives:

The initial technical challenge would involve the cybersecurity concerns related to AI technology itself. Given that perfection is unattainable in real-life scenarios. There is no way to prevent software that utilizes AI technology from producing unexpected faults, such as false positives [13]. AI systems may overlook threats or create false alarms, resulting in inefficient use of resources and security vulnerabilities. False negatives involve not recognizing genuine threats, while false positives consist of mistakenly identifying harmless actions as threats. Even with their abilities, AI systems can generate false positives (identifying non-threats) or false negatives (overlooking real threats). Elevated false positive rates can cause alert fatigue, whereas false negatives may lead to unnoticed security violations [9].

5.2 Complexity:

The complexities of artificial intelligence systems can make them hard to understand. As a result, security analysts might struggle to assess the choices made by AI systems and identify any errors or biases that exist within these frameworks. Moreover, incorporating AI into current cybersecurity systems can require substantial resources and time. Altering legacy systems to accommodate AI technologies frequently poses significant challenges. The intrinsic complexity of AI presents considerable hurdles when trying to integrate AI-powered cybersecurity tools into existing security structures. Organizations might encounter pushback or see a sluggish adoption rate of AI technologies because of integration challenges, potentially reducing the expected advantages. Moreover, scalability poses an issue; as the amount of security data increases, AI algorithms might find it difficult to meet the rising demands. This scenario may lead to an increased occurrence of false positives or false negatives, thus compromising the overall efficacy of the existing cybersecurity protocols.

5.3 Interoperability:

There is a chance that AI-driven cybersecurity solutions may not work with the security technologies that were developed earlier. This could lead to challenges when incorporating AI-driven solutions into existing security systems, potentially necessitating the investment of extra resources to address. Mistakes Made by People Artificial intelligence algorithms can be susceptible to errors resulting from human contributions. For example, if the data used to train the AI model is inaccurate or biased, it could result in errors in the AI's decision-making process [6].

5.4 Adversarial attacks:

AI systems, while effective at identifying cyber threats, are susceptible to adversarial attacks [7]. Accenture describes adversarial AI as a situation where it "leads machine learning algorithms to misinterpret inputs into the system and react in a manner advantageous to the attacker." Essentially, this happens when the neural networks in an AI application are deceived into incorrectly identifying or misrepresenting objects due to intentionally modified inputs [8]. These are intentional alterations of input data aimed at misleading machine learning models, resulting in inaccurate predictions or classifications [7]. This may result in misclassifications, false positives, or false negatives, undermining the effectiveness of cybersecurity protections [9]. In the absence of proper safeguards or precautions, Cyber Security implementations can be almost limitless. Luckily, cybersecurity researchers acknowledge the dangers linked to adversarial AI [8]. For instance, a hacker could make slight alterations to malware code or network traffic that go unnoticed by people but result in an AI system incorrectly categorizing a harmful file as harmless. This manipulation may make AI-based defences useless.

Various methods are being created to protect AI systems from adversarial attacks, such as adversarial training (where AI models are intentionally trained on altered data to identify and resist these attacks) and robust optimization (to reduce the sensitivity of AI models to adversarial changes). Moreover, ensemble learning—using several models together—can assist in safeguarding against the individual limitations undermining the entire system [7].

5.5 Bias and fairness:

AI models are only as effective as the data utilized for their training. If the training data has biases—whether associated with race, gender, location, or attack styles—these biases may be sustained and even intensified in AI systems. In cybersecurity, this might result in biased threat detection that unjustly focuses on particular users or neglects certain attack methods [7].AI models depend greatly on the quality and variety of training data. Partial or skewed data can lead to biased models and unfair results [9]. In the realm of cybersecurity, algorithmic bias may lead to increased false-positive occurrences for particular user groups, or certain threat types might be overlooked due to their underrepresentation in the training dataset [7]. Biases in AI systems may result in ignoring specific threats or incorrectly categorizing typical behaviour as harmful [9]. To address these concerns, researchers and professionals are creating methods to enhance fairness, including the use of varied and representative datasets for training, utilizing explainable AI (XAI) models that ensure transparency in decision-making, and implementing audit trails for the decision-making process. Ethical AI frameworks are being created to guarantee that AI-driven cybersecurity solutions are both accountable and transparent [7].

5.6 Data Privacy:

Cybersecurity systems powered by AI frequently depend on vast quantities of data to identify threats. Nevertheless, this data may contain sensitive details regarding individuals, including browsing behaviours, communication trends, or personal identifiers. The comprehensive surveillance needed by AI-based systems to identify anomalies raises major worries regarding privacy infringement [7]. The application of AI in cybersecurity brings up ethical dilemmas, including privacy concerns, data exploitation, and the risk of misuse. Ethical issues may restrict the range of AI deployment and could lead to regulatory difficulties [9]. The collection and examination of extensive datasets can unintentionally reveal personal details, resulting in privacy violations [1]. Although AI can improve threat detection and response, it should be applied in a manner that honours user privacy. A possible solution is to anonymize or encrypt confidential information prior to its processing by AI systems, guaranteeing that user privacy is preserved while permitting AI to detect patterns within the data. Gathering data to train AI models also brings up ethical concerns. Utilizing personal data without clear consent or transparent data usage guidelines may result in legal and public relations issues. Making certain that data collection adheres to privacy laws such as the General Data Protection Regulation (GDPR) is essential for sustaining trust [7]. Thus, it is essential to enforce strict data protection practices, including data anonymization, encryption, and safe data storage and access procedures. Additionally, AI models should be created to strike a balance between ensuring security and maintaining user privacy [1].

5.7 Ethical considerations:

The incorporation of artificial intelligence (AI) within cybersecurity presents numerous ethical issues, especially concerning the potential for AI systems to make choices that unfairly affect specific individuals or populations. Moreover, AI-based technologies raise considerable privacy concerns, particularly regarding data collection and analysis. To guarantee fairness and precision, it is crucial to tackle the possible biases existing in algorithms and the risks linked to misrepresenting dangers caused by the utilization of biased data sets. Organizations need to be accountable for assessing the ethical consequences of AI in cybersecurity and put in place measures to reduce any possible harm that might emerge.

5.8 Legal and regularities challenges:

As AI becomes more crucial in cybersecurity, current laws and regulations need to adapt to meet the distinct challenges that AI presents. Existing cybersecurity regulations frequently overlook the intricacy and lack of transparency of AI models, which complicates efforts to hold organizations responsible when AI systems do not safeguard against new dangers. Numerous current cybersecurity regulations emphasize the obligation of organizations to adopt appropriate security measures, yet they might not cover how AI systems ought to be incorporated into security protocols [7]. The application of AI in cybersecurity could lead to issues regarding legal and regulatory compliance [6]. For example, the General Data Protection Regulation (GDPR) requires that companies give individuals access to their personal data and clarify how that data is being used [6]. GDPR highlights data protection yet fails to explicitly tackle the transparency and accountability of AI systems deployed in cybersecurity [7]. Conversely, deploying AI algorithms could render it difficult to ensure this level of control and transparency. Reliance on External Suppliers Numerous AI-driven cybersecurity solutions are provided by external vendors. This establishes a reliance on these external suppliers. This may result in a detrimental reliance on these providers, which can be risky if the vendor experiences a data breach or ceases their service [6]. To guarantee the ethical use of AI in cybersecurity, fresh regulations are required. These could involve regulations aimed at AI transparency, required assessments of AI-based systems, and the establishment of accountability structures for AI choices. Moreover, there could be demands for worldwide pacts on AI standards in cybersecurity to guarantee adherence to best practices on a global scale [7].

6. FUTURE ASPECTS

Future malicious entities are already utilizing ChatGPT to generate malware, identify coding bugs, and bypass user access limitations, highlighting the impact of artificial intelligence on cybersecurity. Social engineers utilize artificial intelligence to create more believable and persuasive deepfakes and phishing techniques, while hackers employ AI-assisted password guessing and captcha bypassing to obtain sensitive information without authorization. The technology industry association predicts that spending in the United States will exceed 6 billion dollars, accounting for 0.35 percent of GDP, over the next three years. Gartner Inc. expects an 82 percent rise in global spending. Furthermore, it is expected that blockchain technology will create the highest potential net benefit in the United States, valued at approximately 407 billion $ [4]. Certainly, 85% of security professionals observing a rise in cyberattacks over the past year attribute the surge to harmful individuals. The implementation of generative AI for cost savings in businesses is around 12%. Nevertheless, due to these benefits, artificial intelligence, AI machine learning, predictive analytics, and natural language processing are being utilized to improve cybersecurity in distinct manners, such as rapidly addressing security incidents, identifying attack vectors, and highlighting concealed malfunctions. By the end of 2023, nearly half (48%) of its leaders plan to invest in AI-powered cybersecurity, with 82% intending to do so in the next two years. Most research studies indicate that the success rate for identifying cyber-attacks is between 86 and 98 percent. An artificial intelligence development firm called Darktrace claims that it has a 99 percent success rate and caters to thousands of clients worldwide [4]. Most security professionals (63%) believe that AI can enhance security measures, especially by improving threat detection and response abilities. Only 12% of security experts believe that AI will completely take over their roles. Many believe it will enhance their skill set (30%), bolster their general role (28%), or take over a large part of their responsibilities (24%), allowing them to pursue other activities. In contrast to their employees (11%), C-suite executives indicate a notably greater level of familiarity with AI technologies (52%). The year 2024 is expected to be a transformative one for the use of AI in the security sector. This year, over 50% of organizations (55%) plan to utilize generative AI solutions.

7. CONCLUSIONS

The input of AI has been assessed and evaluated across different domains and uses. However, the domain of cybersecurity is crucial and cannot withstand the shortcomings of AI techniques, as security serves as the sole defense against attackers targeting sensitive information and services. Consequently, understanding the limitations and strengths of AI is essential for future researchers and developers to implement necessary adjustments and modifications for informed decision-making. This paper is prepared to address the limitless possibilities of AI in the cybersecurity field while also confronting the threats that arise if the associated constraints are not considered during implementation. The evolution of AI is also showcased to help researchers understand the progression that led to current AI technology and to connect this evolution with the potential future advancements, as depicted by numerous researchers, experts, and individuals throughout history. The strength of AI can be maintained by the advancing cyber-attacks if AI is developed at a comparable speed to the attacks. Considering the possible integration of additional technologies in AI-driven security applications, attackers can be effectively protected or at least mitigated. Moreover, regarding shifts in other sectors as a factor, AI cybersecurity is transforming the labour market by pushing it toward more intellectually challenging positions, necessitating a blend of skills and fostering growth and flexibility in an evolving technological landscape.

ACKNOWLEDGEMENT

We would like to express our profound gratitude to Dr. Neha Yadav for her invaluable guidance, support, and encouragement throughout the development of this project, *AI in Cybersecurity : A Literature review*. Her insightful feedback, deep knowledge of the subject matter, and unwavering commitment to academic excellence played a pivotal role in shaping the direction and quality of this work. We are truly thankful for her patience, availability, and willingness to aid whenever we encountered challenges during the research and writing process.

We would also like to extend our sincere appreciation to the faculty and staff of the Artificial Intelligence and Machine Learning Department, Dr. Akhilesh Das Gupta Institute of Professional Studies, for their constant support, timely suggestions, and for providing access to relevant resources and a conducive environment for academic inquiry. Their commitment to fostering a learning-oriented atmosphere greatly contributed to the successful completion of this study.

Furthermore, we are grateful for the moral support and motivation provided by our peers, who, directly or indirectly, encouraged us to persevere and maintain focus throughout the course of this project.

Finally, we would like to thank all those who contributed in any way to the realization of this work. Their collective support and cooperation have been invaluable, and we remain deeply appreciative of their contribution.

REFERENCES

1. https://www.researchgate.net/publication/377235308_Artificial_Intelligence_in_Cyber_Security
2. https://www.researchgate.net/publication/364126309_Artificial_Intelligence_in_Cyber_Security
3. https://www.researchgate.net/publication/385881488_Artificial_intelligence_in_cybersecurity_Enhancing_threat_detection_and_prevention_mechanisms_through_machine_learning_and_data_analytics
4. https://ijsrst.com/paper/9159.pdf
5. https://www.intechopen.com/chapters/1185407
6. https://www.researchgate.net/publication/377720256_The_Two_Sides_of_AI_in_Cybersecurity_Opportunities_and_Challenges
7. https://www.researchgate.net/publication/387437835_Artificial_Intelligence_in_Cybersecurity_Identifying_and_Neutralizing_Emerging_Threats
8. https://www.researchgate.net/publication/353419449_Artificial_Intelligence_in_Cyber_Security
9. https://www.ijsrp.org/research-paper-1223.php?rp=P14413195
10. https://doi.org/10.1186/s40537-024-00957-y
11. https://ieeexplore.ieee.org/abstract/document/10491270
12. https://doi.org/10.1080/08839514.2024.2439609
13. https://doi.org/10.1109/DASC-PICom-CBDCom-CyberSciTech49142.2020.00064
14. https://doi.org/10.22541/au.166379475.54266021/v1