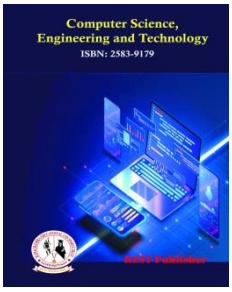
Computer Science, Engineering and Technology

Vol: 3(1), March 2025

REST Publisher; ISSN: 2583-9179 (Online)

Website: https://restpublisher.com/journals/cset/

DOI: https://doi.org/10.46632/cset/3/1/1



Review on Effects of Artificial Intelligence on Cyber-Security

P. C. Khanzode, *Aniruddha Pise, Prajwal Sinkar, Rathijeet Jawalkar, Subodh Yadao, Vaishnavi Bajaj

Sipna College of Engineering and Technology, Maharashtra, India.

Corresponding Author Email: aniruddhaspise03@gmail.com

1. INTRODUCTION

Cybersecurity encompasses a collection of technologies, processes, and practices designed to protect networks, software, and data from intrusion, damage, and unauthorized access. The rapid advancements in digital technology and infrastructure have led to a significant increase in cyberattacks, posing severe threats. Additionally, researchers observe the growing sophistication of nation-state-backed malicious campaigns and the evolving complexity of cyberattacks, which are continuously developing new and innovative ways to target even the most vigilant of targets. The frequency, scale, and impact of cyberattacks are increasing due to this evolution, making intelligence-driven cybersecurity essential to manage vast amounts of data and provide a dynamic defense against changing cyber threats. To identify, prevent, detect, respond to, and recover from cyberattacks to avoid future security incidents, organizations like the National Institute of Standards and Technology (NIST) are advocating for the adoption of more advanced and adaptive approaches. They promote real-time assessments, continuous monitoring, and data-driven analysis.

Artificial Intelligence (AI) is a promising tool that can offer analytics and intelligence to combat constantly evolving cyber threats by rapidly analyzing millions of events and covering a broad spectrum of cyber threats to predict and address issues before they arise. Consequently, AI is being increasingly integrated into cybersecurity systems and used for various purposes, such as automating security tasks or assisting human security teams in their work. The vast field of cybersecurity and the growing interest of researchers from both AI and cybersecurity domains have led to numerous studies addressing issues related to the identification, protection, discovery, response, and recovery from cyberattacks. The convergence of AI and cybersecurity knowledge represents a significant advancement in how organizations defend themselves against the onslaught of cyber threats. This combination has the potential to enhance human training and response capabilities while revolutionizing threat discovery.

Although AI is becoming increasingly prevalent in cybersecurity practices, its impact and limitations are complex and require careful study. AI's adaptive human training approach also holds great potential for creating a security-conscious workforce by tailoring training materials to meet unique learning needs and knowledge gaps. These factors demonstrate how AI can elevate cybersecurity practices and serve as a crucial safeguard against a dynamic landscape of online threats.

Moreover, the vulnerability of AI systems to malicious attacks highlights a dangerous trend in which the very tools of defence could be turned against the defenders. Data bias, particularly when AI systems unintentionally reinforce biased outcomes, raises justice and ethical concerns. To maximize AI's benefits and minimize its drawbacks as organizations navigate these complex relationships, it is essential to strike a balance between the technology's promise and its limitations.

In recent times, several studies on AI applications and cybersecurity have been published. However, to the best of our knowledge, there is not yet a comprehensive analysis that covers current research to explain the specifics of how AI methods are used and the cybersecurity operations they address. Therefore, to serve as a resource for future researchers and practitioners, our objective was to present a thorough analysis, an overview of AI use cases in cybersecurity, limitations, and a discussion of the research challenges associated with the adoption and use of AI for cybersecurity.

2. LITERATURE REVIEW

Selma Dilek, Hüseyin Çakır and Mustafa Aydın in their paper they proposed in this ultramodern period of I.T. sector associations and associations are having intrusion issues and not only the associations but also nontechnical individualities are not safe either from cybercriminals currently. They need a system having adaptive, literacy, dynamic features and protean enough and can descry high-position intrusions and can do real- time intelligent opinions (1). Ozlem Yavanoglu, Murat Aydos's study was to clarify and look at the most typically employed datasets. His paper was centered around the datasets employed in artificial intelligence strategies that are employed for studying and assaying system transitions and relating variations from the norm (1). Jian- Hua LI sum up being examination trials regarding fighting cybercrimes and attacks exercising A.I., probing the counterassaults from which A.I. itself may endure, dissect their attributes, and order the relating protective ways. At last, from the corridor of developing translated neural system and understanding a defended combined deep literacy, and unfold the current examination on the stylish way to fabricate a safe A.I. frame (1). Dr Pranav Patil proposed the idea of using artificial neural nets in cybersecurity systems that will drastically ameliorate the intrusion discovery systems and can help cybercriminals effectively. So, he said cybersecurity issues can be handled by exercising artificial intelligence (1). Arab Mohammed Shamiulla directed his disquisition on the job of artificial intelligence in cybersecurity and how A.I. not only can descry and help the cyberattacks and cybercrimes but can also prognosticate and distinguish them because of its literacy and prophetic capabilities and it can also store the information and keep on literacy and acclimatize through it (1). Shikha Goyal in her composition gave some nitty- gritty examination over" What's Cyber Crime and how it's expanding step by step", as to how cybercrimes are an unlawful demonstration where the P.C. is employed as an instrument or target or both. (1). Katharina Buchholz composed a composition in which she gave her disquisition on cybercrimes and assaults, and cybercrimes have in verity observed a huge proliferation in India during the most recent decade. Wrongdoing cases coming about because of cybercrimes expanded right around 30- crinkle since 2010, notwithstanding the way that a couple of assaults lead to a case (1). Nick Heath clarified the abecedarian ideas, highlights and uses of A.I. in his composition. And he added that artificial intelligence can fluently help humans and mortal sapience similar as in thinking, prognosticating, proposing, feting, controlling, an information storing, imagination and in understanding social and new abecedarian information (1). Martin Armstrong, a data intelligencer delved the determined worldwide total artificial intelligence profit and application in the middle of 2016- 2025. In his composition, he gave a statistical knowledge from firm Tractica gave us the rate of A.I. is being used worldwide by associations. (1). Ron Tolido, Greet V. D. Linden, Luis Delabarre, Anne- Laure Thieullent, Allan Frank, Luis Delabarre, Jerome Buvat, Jeff T, Sumit Cherian, Yashwardhan Khemka reviewed 850 elderly chiefs, scientists, I.T. sector employs worldwide. They bandied the current issues and cyber problems with seniors, CIOs and CISOs in different associations, and also led interviews with assiduity settlers and scholastics, looking at the current status and effect of A.I. in cybersecurity (1). Sarah Feldman overviewed and broke down the advantages of exercising A.I. alongside cybersecurity bettered the intrusion discovery fabrics. In her composition establishment Consumer Technology Association delved and reached to the conclusion that artificial intelligence is perfecting I.T. sector associations security and findings systems (1). Anagnostopoulos exfoliate lights on the uses of deep literacy out how to ameliorate the cybersecurity. And he said that with a deficit of information and indeed with high chops a trouble can arise but if we make the machine gather much further information and it learns all of it and keeps on passing it also the gathered data can be stored in the proper arrangement of information for machine literacy and can be used in prognosticating the attack and will affect in further robust and secure intrusion system (1). Pandey, M. bandied the application of expert systems in cybersecurity. He proposed the idea that experts' systems are one of the most effective artificial intelligence operations. And because of its features and capabilities, it can be used in making logical and right opinions (1). Ansari Q., Patki T., Patki B., and Kumar V. bandied the data mining strategies to plan to acquire the particular information and store it in one particular corresponding database in between several databases. And similar storing is n't an easy task and normal ways can't suitable to do it effectively though we can use data mining strategies to complete this task with proper distribution by exercising machine literacy strategies, expert systems, neural network strategies etc (1). Azzah Kabbas, Atheer Alharthi, Asmaa Munshi gave us the idea of artificial intelligence and its strategies. likewise, the artificial intelligence along its operations can help in stopping cybercrimes and assaults, therefore, perfecting the overall cybersecurity (1). Jan Pospisil, Senior Data Scientist, Siemens Cyber Defence Centre enlightened us concerning their association and advantages of exercising A.I. in cybersecurity for better security and guaranteed us about the 170- year old worldwide invention colonist, must watch out for the ever- developing scene of cyber-attacks and assaults. They shield their guests from the cyberattacks and assaults (1).

3. CYBERSECURITY

Cybersecurity is the field of protecting computer systems, networks, and data from online threats. These cyberattacks typically aim to disrupt normal business operations, gain, alter, or delete sensitive data, or demand ransom payments from victims. Cybersecurity tools, techniques, processes, and specialized methodologies are employed to prevent damage, unauthorized use or modification, or exploitation of information and communication systems and their contained data. The rapidly evolving technological landscape and the ever-changing nature of cyber threats further complicate the situation. In response to this dynamic challenge, AI-based cybersecurity solutions have emerged, assisting security teams in effectively mitigating threats and enhancing security.

To analyse the literature on using AI for cybersecurity, a widely recognized and unified taxonomy is necessary due to the diversity of AI and cybersecurity. Such a taxonomy will enable researchers and practitioners to better understand the specific techniques and services that leverage AI to improve cybersecurity performance. Therefore, the proposed framework employs a well-established cybersecurity framework, proposed by NIST, to understand the tasks required to identify, prevent, respond to, and counter cyberattacks.

The NIST cybersecurity framework outlines the fundamental principles for strengthening an organization's cybersecurity. The framework's four main components are functions, categories, subcategories, and informative references. The identified AI use cases were categorized using the first two components of the NIST framework, consisting of 23 tasks and five cybersecurity functions. The functions provide a comprehensive overview of the cybersecurity lifecycle process. The tasks under each function serve as an excellent starting point for exploring AI use cases to enhance cybersecurity. The primary purpose of selecting these two components is to offer a simple and accessible categorization system for grouping the existing AI for cybersecurity literature into relevant tasks.

As depicted in Figure 1, the proposed taxonomy includes a third level that aligns with the first two components by outlining AI-based use cases for each cybersecurity framework component. Cybersecurity is a broad field encompassing several academic disciplines. It comprises seven primary pillars:

a. Network Security
b. Cloud Security
c. Endpoint Security
d. Mobile Security
e. IoT Security
f. Application Security
g. Zero trust

4. ARTIFICIAL INTELLIGENCE

Artificial intelligence (AI) leverages computers and machines to mimic the problem-solving and decision-making capabilities of the human mind. AI has been a game-changer in cybersecurity, offering cutting-edge techniques for detecting and mitigating cyber threats, fundamentally altering the approach to cybersecurity. AI is becoming an increasingly essential tool in cybersecurity strategies for numerous businesses, and its adoption in this field is growing rapidly. According to a report published by Markets and Markets, the global AI in cybersecurity market is projected to increase at a compound annual growth rate (CAGR) of 23.3%, from $8.8 billion in 2020 to $38.2 billion by 2026.

Prior to the advent of AI, traditional cybersecurity heavily relied on manual-based detection methods. These systems operated by comparing incoming traffic to signatures of known malicious or suspicious threats in a database. When a match was found, the system would raise an alert and take appropriate action to block or isolate the threat. However, manual- based detection systems could produce numerous false positives, as legitimate traffic might inadvertently resemble the characteristics of a known threat. This resulted in security analysts spending a significant amount of time investigating false positives, potentially draining valuable resources.

Traditional cybersecurity also involved manual analysis, where security analysts would manually examine security alerts and logs, searching for trends or indicators that might suggest a security breach. AI-based cybersecurity solutions differ from conventional methods in several ways. As previously discussed, manual-based detection systems were limited to detecting known threats, which meant that new and unknown threats might go unnoticed. In contrast, AI-based solutions utilize machine learning algorithms capable of real-time threat detection and response for both known and unknown threats.

Machine learning algorithms are trained on vast amounts of data, including historical threat data and data from the network and endpoints, to identify patterns that are difficult for humans to discern. This enables AI-based systems to detect threats and take appropriate action in real-time, without requiring human intervention. For example, machine learning algorithms can analyse network traffic patterns to spot unusual activity that might indicate a cyberattack. They can also notify security staff or even initiate automatic actions to mitigate the threat.

Another distinguishing aspect of AI-based solutions is their ability to continuously learn and adapt. This sets them apart from conventional methods. The application of AI to cybersecurity signifies a significant shift in how businesses approach cybersecurity. AI disciplines include, but are not limited to, fuzzy logic, case-based reasoning, genetic algorithm, Bayesian optimization, evolutionary algorithm, planning graph, artificial neural network, deep learning, support vector machine, natural language processing, text mining, sentiment analysis, image processing, convolutional neural networks, object recognition, and speech processing. This allows businesses to better protect sensitive information and critical systems.

5. AI'S IMPACT ON CYBER OPERATIONS AND THREATS

This section examines how artificial intelligence (AI) will influence the effectiveness of cyber operations and the subsequent implications for the cyber threat landscape over the next two years. It does not delve into the potential cyber security vulnerabilities of AI tools themselves or the risks associated with integrating them into system architectures.

Assuming no significant breakthroughs in transformative AI during this period, the assessment remains subject to review. Such breakthroughs could significantly impact malware and zero- day exploit development, altering the cyber threat landscape.

The impact of AI on the cyber threat will be counterbalanced by its use to enhance cyber security resilience through detection and improved security by design. Further research is needed to understand the extent to which AI advancements in cyber security will mitigate threat impacts.

Key Findings:

- **Increased Volume and Severity of Cyberattacks:** AI will likely increase the volume and severity of cyberattacks in the next two years. However, the impact on the cyber threat will vary.
- **Evolution of Tactics, Techniques, and Procedures (TTPs):** The threat to 2025 stems from the evolution and refinement of current TTPs.
- **AI Adoption by Threat Actors:** All types of cyber threat actors, state and non-state, skilled and less skilled, are already employing AI to varying degrees.
- **Enhanced Reconnaissance and Social Engineering:** AI enhances capabilities in reconnaissance and social engineering, making these activities more effective, efficient, and harder to detect.
- **Sophisticated AI Applications:** More sophisticated AI applications in cyber operations are likely to be limited to threat actors with access to quality training data, significant expertise in both AI and cyber, and substantial resources. Advanced AI-based attacks are unlikely to become widespread before 2025.
- **Amplified Impact of Cyberattacks:** AI will almost certainly amplify the impact of cyberattacks as threat actors can analyse exfiltrated data more rapidly and effectively, using it to train AI models.
- **Increased Accessibility for Inexperienced Actors:** AI reduces the difficulty for inexperienced cybercriminals, hackers-for-hire, and hacktivists to conduct effective access and information gathering operations. This increased accessibility will likely contribute to the global ransomware threat over the next two years.
- **Widespread Availability of AI-Enabled Capabilities:** Looking toward 2025 and beyond, the widespread availability of AI-enabled capabilities in criminal and commercial markets will almost certainly make advanced capabilities accessible to both cybercriminals and state actors.

6. CHARACTERISTICS AND TYPES OF AI-BASED ATTACKS

Artificial Intelligence (AI) has become a pivotal asset in the IT strategies of organizations but has also emerged as a potent tool for cybercriminals. AI-driven cyberattacks leverage AI or machine learning (ML) algorithms to automate, enhance, or accelerate various stages of an attack. This includes identifying vulnerabilities, launching attacks across specific vectors, advancing attack methodologies, creating backdoors in systems, stealing or manipulating data, and disrupting system functions.

Like all AI systems, those used in AI-based attacks can learn and improve with time. This adaptability allows AI-driven attacks to evade detection or develop new patterns that security systems may struggle to recognize.

Key Features of AI-Powered Cyberattacks

AI-driven cyberattacks typically exhibit the following five characteristics:

- **Automation of Attacks:** Traditionally, most cyberattacks required direct human involvement. However, with the increasing availability of AI-based and generative AI tools, attackers can now automate both the research and execution of their attacks.
- **Efficient Data Collection:** The initial stage of any cyberattack involves reconnaissance, where attackers gather information on potential targets, vulnerabilities, and valuable assets. AI can expedite this process, allowing adversaries to shorten the time spent on research while increasing the accuracy and scope of their findings.
- **Personalization:** AI excels at data extraction, gathering information from public sources like social media platforms and corporate websites. In a cyberattack, this information can be used to craft highly personalized and timely messages that form the basis for phishing or other social engineering attacks.

- **Reinforcement Learning:** AI systems are designed to learn and adapt continuously. Just as AI tools refine their insights to benefit corporate users, they also enable attackers to improve their methods or evade detection by evolving their techniques over time.
- **Employee Targeting:** Similar to attack customization, AI can identify individuals within an organization who are considered high-value targets. These are people with access to confidential data or extensive system permissions, those who may lack technical proficiency, or individuals with close ties to other critical personnel.

7. TYPES OF AI-POWERED CYBERATTACKS

AI and machine learning (ML) have enabled various forms of cyberattacks. Some of these include:

➢ **Adversarial AI**

Adversarial AI or adversarial machine learning (ML) aims to disrupt or manipulate AI/ML systems by deceiving them. These attacks can occur at various stages of the machine learning model's development, from corrupting training data to introducing flaws or biases in ML models. Attackers may also craft misleading inputs to trick the system into generating incorrect outputs, and they can combine these tactics to amplify the effectiveness of their attacks.

Unlike traditional cyber threats, such as malware or phishing, adversarial AI specifically targets the decision-making processes of AI systems. This allows adversarial malware to bypass even well-trained machine learning models. As a result, adversarial AI/ML is becoming a major concern for modern security operations (SecOps) teams.

Main Concerns Associated with Adversarial AI/ML :

- **Growing Complexity of AI/ML Models:** As these models grow in complexity, they become increasingly attractive targets for attackers.
- **Prevalence of AI/ML:** AI and ML technologies are widely used across many industries, which increases the potential impact of a successful adversarial AI attack.
- **Enhanced Attacker Capabilities:** As AI/ML technologies continue to evolve, so too do the tools, skills, and strategies that attackers use to exploit them.
- **Defending Against Adversarial AI Attacks**

To effectively defend against adversarial AI attacks, security teams should be aware of the following techniques:

- **Data Poisoning:** Modifying training data in a way that causes the resulting model to behave incorrectly or make poor decisions.
- **Model Tampering:** Altering the parameters or structure of an ML model, compromising its ability to deliver accurate results.
- **Attack Transferability:** Utilizing successful attack techniques across multiple AI/ML systems by leveraging common tools and methods.

Deepfakes: A Growing Threat

The rapid advancements in artificial intelligence have ushered in significant breakthroughs but have also given rise to challenges, such as deepfakes. These AI-driven forgeries create realistic yet fabricated images, audio, or video. As AI becomes more sophisticated and widely available, deepfakes will become increasingly difficult to differentiate from real content, posing a growing threat to information authenticity.

Understanding Deepfakes: While deepfakes highlight the impressive capabilities of AI and machine learning, they also present a serious threat to the integrity of digital content. These manipulations can influence public perception, tarnish reputations, and even disrupt political processes. This section delves into how deepfakes are created and their broader societal implications.

The Technology Behind Deepfakes: Deepfakes are often produced using a type of AI called generative adversarial networks (GANs). A GAN consists of two competing neural networks: one, known as the generator, attempts to create realistic fake content, while the other, called the discriminator, evaluates whether the content is genuine or not. Through repeated iterations, the generator's output becomes increasingly convincing, making it difficult to distinguish between real and fabricated material.

GANs, a rapidly advancing technology, enable the creation of deepfake videos and audio that can deceive even the most discerning observers.

Categories of Deepfakes: Deepfakes generally fall into two main types: video and audio manipulations. Video deepfakes involve altering someone's appearance, such as their face or body, to make them look like another individual. These are frequently used in contexts like celebrity face swaps or spreading false political narratives. Audio deepfakes, on the other hand, replicate someone's voice, creating convincing audio clips of people saying things they never actually said. Both forms of deepfakes can be used for a variety of purposes, from innocent entertainment to more harmful activities like fraud and political manipulation.

Addressing the Malicious Use of Deepfakes: Fighting the harmful applications of deepfakes is a challenging and ongoing struggle, particularly in the field of cybersecurity. As AI technology advances rapidly, the methods used to create deepfakes also become more sophisticated, making them increasingly difficult to detect. This technological "arms race" against malicious AI emphasizes the complex nature of the issue. Cybersecurity systems, detection algorithms, and defense mechanisms must constantly evolve to keep up with the advancements being leveraged by bad actors.

In response to this growing threat, cybersecurity experts have developed a range of tools and techniques aimed at detecting and neutralizing deepfakes. These include:

- **AI-powered detectors** that identify irregularities in video and audio content.
- **Digital forensics methods** to assess the authenticity of media files.
- **Blockchain technology** to confirm the legitimacy of digital content using immutable records.
- **Identity protection solutions** to prevent individuals' digital personas from being exploited in deepfake attacks.

Despite these efforts, the fight against deepfakes remains complex, and a completely foolproof solution has yet to be found.

AI-Powered Malware: A New Era of Cyber Threats

An AI-generated malware dubbed BlackMamba successfully evaded cybersecurity technology, including enterprise-grade endpoint detection and response (EDR), in an experimental mission led by researchers at Hyas. BlackMamba employs a large language model (LLM)—a deep learning algorithm capable of summarizing and generating textual content—to create a polymorphic keylogger. This means that whenever BlackMamba runs, it mutates, enabling it to slip through predictive cybersecurity software.
Imagine this AI malware as a deadly disease that constantly mutates. It would be difficult to develop a permanent cure since the malware can change on the fly. BlackMamba can be introduced through an executable file, which contains instructions to alter a device's system. Cybercriminals could create malware similar to the experimental BlackMamba and disseminate it through seemingly innocent software programs.

Building BlackMamba: To illustrate the capabilities of AI-based malware, researchers at HYAS constructed a simple proof of concept that leverages a large language model to synthesize polymorphic keylogger functionality on the fly, dynamically editing benign code at runtime—without any command-and-control infrastructure to deliver or verify the malicious keylogger functionality.
To create this proof of concept, HYAS researchers combined seemingly disparate ideas. The first was to eliminate the command and control (C2) channel by using malware equipped with smart automation that could push back any attacker-specific data through a benign communication channel. The second was to utilize AI code-generative techniques that could synthesize new malware variations, altering the code to evade detection algorithms.

BlackMamba utilizes a benign executable that reaches out to a popular API (OpenAI) at runtime, receiving synthesized, malicious code necessary to steal an infected user's keystrokes. It then executes the dynamically generated code within the context of the benign application using Python's exec() feature, ensuring the malicious polymorphic element remains entirely in-memory. Whenever BlackMamba executes, it re- synthesizes its keylogging functionality, making the malicious component of this malware truly polymorphic. BlackMamba was tested against an industry-leading EDR, which remained anonymous, resulting in zero signals or detections.

Data Exfiltration via MS Teams: Once a device was compromised, a method was needed to exfiltrate data. MS Teams, like other communication and collaboration tools, can be exploited by malware authors as an exfiltration channel. In this context, an exfiltration channel refers to the method by which an attacker removes or extracts data from a compromised device and sends it to an external location, such as an attacker-controlled Teams channel through a webhook.
With its keylogging capability, BlackMamba can collect sensitive data, such as usernames, passwords, credit card numbers, and other personal or confidential information that a user types into their device. Once this data is captured, the malware uses an MS Teams webhook to send the collected information to the malicious Teams channel, where it can be analyzed, sold on the dark web, or used for other nefarious purposes.

The Delivery: Auto-py-to-exe

Auto-py-to-exe is an open-source Python package that enables developers to convert their Python scripts into standalone executable files that can be run on Windows, macOS, and Linux operating systems. While intended for legitimate use, this package can also be exploited by malware authors to package their Python-based malware into executable files that can be distributed and run on a target system without requiring Python installation.

When using auto-py-to-exe, the malware creator first writes their Python-based malware code and imports any necessary libraries or modules. They then utilize the auto-py-to-exe package to generate an executable file from their Python code.

Once the executable file is generated, the malware writer can distribute it. When a victim runs the executable file, the malware is executed on their device and can perform various malicious actions, including stealing sensitive information, altering system settings, or downloading additional malware—such as keylogging in our case.

Moving Forward: The Growing Threat

The threats posed by this new breed of malware are very real. By eliminating C2 communication and generating new, unique code at runtime, malware like BlackMamba is effectively undetectable by today's predictive protection solution.

The Delivery: Auto-py-to-exe

Auto-py-to-exe is an open-source Python package that allows developers to convert their Python scripts into standalone executable files that can run on Windows, macOS, and Linux operating systems. While intended for legitimate use, this package can also be exploited by malware authors to package their Python-based malware into executable files that can be distributed and run on a target system without requiring Python installation.

How Malware Authors Use Auto-py-to-exe

1. **Writing the Malware Code:** The malware creator first writes their Python-based malware code and imports any necessary libraries or modules.
2. **Generating the Executable:** They then utilize the auto-py-to-exe package to generate an executable file from their Python code.
3. **Distribution and Execution:** Once the executable file is generated, the malware writer can distribute it. When a victim runs the executable file, the malware is executed on their device and can perform various malicious actions, including stealing sensitive information, altering system settings, or downloading additional malware—such as keylogging in the BlackMamba case.

Moving Forward: The Growing Threat

The threats posed by this new breed of malware are very real. By eliminating command- and-control (C2) communication and generating new, unique code at runtime, malware like BlackMamba is effectively undetectable by today's predictive protection solutions.

Call to Action

Staying safe in this evolving threat landscape requires vigilance and proactive measures. Here are some steps you can take to protect yourself:

- **Be cautious of unsolicited attachments or downloads.** Only open files from trusted sources.
- **Use a reputable security software suite** that includes real-time protection and behavior- based detection capabilities.
- **Keep your software up to date.** Software updates often include security patches that address newly discovered vulnerabilities.
- **Be aware of social engineering tactics.** Phishing emails and social media scams can trick you into downloading malware.

By following these tips and staying informed about the latest cyber threats, you can help to keep yourself and your devices safe.

- **AI-Driven Social Engineering:**

AI-driven social engineering attacks leverage AI algorithms to assist in the research, creative concepting, or execution of a social engineering attack. A social engineering attack is any type of cyberattack that aims to manipulate human behavior to achieve a goal, such as sharing sensitive information, transferring money or ownership of high-value items, or granting access to a system, software, database, or device.

In an AI-driven social engineering attack, an algorithm may be used to:

- **Identify a target:** This includes both the overall corporate target and an individual within the organization who can serve as a gateway to the IT environment.

- **Develop a persona and online presence:** This persona is used to communicate with the attack target.
- **Create a plausible scenario:** Develop a situation that would generate interest and appear credible.
- **Craft personalized messages or multimedia assets:** This could include audio recordings or video footage to engage the target.

Steps of a Social Engineering Attack:
Social engineering attacks typically follow these basic steps:

1. **Research:** The attacker identifies victims and chooses a method of attack.
2. **Engage:** The attacker makes contact and begins the process of establishing trust, appealing to greed, helpfulness, or curiosity, and creating a sense of urgency.
3. **Attack:** The attack is launched, and the attacker collects the payload.
4. **The Getaway:** The attacker covers their tracks and concludes the attack.

Best Practices to Prevent Social Engineering Attacks:
Security awareness training is the best way to prevent falling victim to social engineering attacks. As part of security awareness programs, organizations should continually remind their employees of the following practices:

- **DON'T click on links sent by people you don't know.** Hover over them first; trust but verify!
- **Avoid opening attachments in emails from unknown senders.**
- **Do not provide your username, password, date of birth, Social Security number, financial information, or other personal information in response to an email or robocall.**
- **Check for misspellings or incorrect domains within a link** (for example, an address that should lead to a .gov ends in .com instead).
- **Before transferring money or information, confirm via voice or video call.**

8. CHALLENGES OF IMPLEMENTING AI IN CYBERSECURITY

Artificial Intelligence (AI) offers immense potential to enhance cybersecurity, but its effectiveness hinges on addressing several critical challenges. Let's delve into some key obstacles:

a. **Bias in AI:** AI systems, like humans, can be influenced by biases present in their training data. If this data is discriminatory, the AI may produce biased outcomes, impacting cybersecurity decision-making. Leading AI platforms are actively working to minimize bias in their systems through thoughtful machine learning training.
b. **Misinterpretation:** Even advanced AI systems can experience "AI hallucinations," misinterpreting information and making decisions based on incomplete or inaccurate data. This can lead to incorrect threat assessments, potentially leaving threats undetected or increasing false positives.
c. **Overreliance:** Excessive reliance on AI can create vulnerabilities as AI-driven errors may accumulate and impact cybersecurity systems. Organizations must be cautious about relying solely on AI for defense, as novel cyberattacks may exploit AI-managed protections.
d. **Skills Gap:** The shortage of skilled security professionals and IT specialists capable of effectively deploying and managing AI systems poses a significant challenge. Insufficient expertise can lead to poor implementation, misconfigurations, and inadequate protection against cyber threats.
e. **Privacy and Legal Concerns:** AI applications in cybersecurity often involve the processing and analysis of vast amounts of personally identifiable data, raising privacy concerns. Legal compliance is essential before deploying AI models to ensure adherence to privacy regulations.
f. **Data Manipulation:** AI models rely on historical data to learn and make decisions. However, this dependence can make them vulnerable to manipulation by malicious actors. Hackers may gain access to training data and introduce biases, compromising the efficiency and accuracy of AI models.
g. By addressing these challenges and overcoming them, organizations can harness the power of AI to enhance their cybersecurity posture and protect against emerging threats.

9. USE CASES OF ARTIFICIAL INTELLIGENCE IN CYBER SECURITY'S

a. **Use Case 1: Threat Detection and Prevention**

i. AI-Powered Threat Detection and Response

1. AI excels at threat detection by analyzing vast amounts of data from various sources and identifying anomalous patterns in user behavior that may indicate a cyberattack.
2. When potential threats are detected, AI-powered systems trigger real-time alerts and notifications to cybersecurity teams, enabling rapid and effective responses.
3. By automating incident response actions, such as isolating affected systems or blocking malicious activities, AI minimizes opportunities for attackers and limits the potential impact of security breaches.

ii. Malware and Phishing Detection

1. AI-based cybersecurity systems offer significantly enhanced efficacy in detecting malware and phishing attacks.
2. AI analyzes email content and context to distinguish between spam, phishing attempts, and legitimate messages.
3. Machine learning algorithms and advanced threat intelligence enable AI to evolve and adapt to new threats, recognizing the signs of sophisticated attacks like spear phishing.

iii. Endpoint Security

1. AI-driven endpoint protection adopts a dynamic approach, establishing baselines of normal endpoint behavior and detecting deviations in real time.
2. By continuously learning from network behavior, AI can identify potential threats, including zero-day attacks, without requiring signature updates.

iv. Encryption

1. Encryption algorithms like AES and SHA are designed to be highly resistant to cracking, making it difficult for AI or any attacker to predict their behavior.
2. While AI can accomplish remarkable feats, breaking strong encryption remains a significant hurdle.

v. Security Log Analysis

1. AI utilizes machine learning algorithms to analyze vast amounts of real-time log data, detecting patterns and anomalies to identify and respond to potential security breaches promptly.
2. AI excels at detecting potential insider threats through a comprehensive analysis of user activity across multiple systems and applications.

b. **Use Case 2: User Behavior Analytics**

- AI models leverage deep and machine learning techniques to continuously analyze network behavior and detect deviations from normal patterns.
- Over time, these models self-correct and adapt, enhancing their accuracy in identifying anomalies and potential threats.
- AI-driven behavioral analytics enhances threat-hunting processes by creating deployed application profiles and analyzing extensive user and device data.
- This enables organizations to identify and investigate suspicious activities that may indicate potential security breaches.

Amazon, through its AWS platform, provides a range of AI-driven security services that have transformed how businesses identify and prevent threats.

- AWS GuardDuty is a managed threat detection service that examines various data sources such as AWS CloudTrail logs, VPC Flow Logs, and DNS logs. It detects suspicious activity, including abnormal spikes in API calls, irregular network traffic patterns, and unauthorized access attempts.
- AWS Inspector offers continuous monitoring to identify security vulnerabilities within an organization's AWS infrastructure, enabling proactive threat management.
- AWS Macie is another key service that leverages machine learning to discover, classify, and protect sensitive data in AWS environments. It specializes in detecting critical information, such as personally identifiable information (PII), financial data, and intellectual property (IP).

c. Use Case 3: Vulnerability Assessment and Management

As cybercriminals increasingly use advanced techniques, organizations face challenges in handling emerging vulnerabilities. AI-powered tools like User and Entity Behavior Analytics (UEBA) monitor user, server, and device activities to detect anomalies,

including zero-day attacks. These systems offer proactive protection by identifying and mitigating undisclosed vulnerabilities before they are exploited.
Splunk's Enterprise Security platform utilizes machine learning to analyze large volumes of data from sources such as network logs, system events, and user activity. This AI-powered approach enables real-time detection of patterns and anomalies that could indicate vulnerabilities or malicious activities.

A significant advantage of Splunk's AI-driven Vulnerability Assessment and Management is its ability to intelligently prioritize threats. By applying AI algorithms to the data, the platform can accurately assess the severity and potential impact of vulnerabilities, allowing security teams to focus on the most critical risks first.

d. Use Case 4: Security Operations and Automation

AI streamlines security operations by automating the identification and resolution of threats, reducing response times, and minimizing the risk of human error in critical tasks. This automation allows cybersecurity teams to focus on strategic decision-making and improving defense mechanisms.

Security Operations and Automation in Plaid
Plaid leverages advanced machine learning to analyze multiple data points, such as customer names, addresses, and Social Security numbers. The AI system performs bank account verification quickly and accurately, minimizing errors and fraud risks. By automating these processes, Plaid enhances the onboarding experience for financial institutions, eliminating the need for manual intervention and paperwork while improving data security.

The software industry is witnessing a concerning surge in newly discovered vulnerabilities, with over 22,000 reported in 2022 — the highest number in over a decade. As cybersecurity professionals grapple with the challenge of keeping pace with constantly evolving threats, machine learning-based cybersecurity systems offer a promising solution.

Leading tech companies such as Google, IBM, and Microsoft are pioneering advanced AI systems designed to identify and mitigate threats. Google's Project Zero, for instance, has committed $10 billion over five years to bolster cybersecurity efforts. The initiative is dedicated to finding and fixing web vulnerabilities, helping to secure the internet from potential attacks.

This paper offers valuable insights into the intersection of cybersecurity and AI technologies, highlighting key research gaps that could drive future developments. Currently, AI's role in cybersecurity is not as impactful as it could be due to certain limitations. For instance, AI systems often produce false positives and false negatives, which can result in actual threats being overlooked or overshadowed by irrelevant alerts. This not only undermines user confidence but also reduces the overall effectiveness of security operations.

Adversarial attacks can exploit AI's vulnerabilities, potentially compromising systems meant to enhance security. Moreover, AI systems that overly depend on historical data may struggle to adapt to novel, sophisticated attack methods, which limits their ability to respond to emerging threats. While AI has great potential, it still requires human oversight because it struggles to grasp contextual nuances and may misinterpret user behaviors or intentions. To fully realize AI's benefits in cybersecurity, it is crucial to address these limitations and strike a balance between automated systems and human expertise.

10. CONCLUSION

The paragraph delves into the application of artificial intelligence (AI) in the realm of cybersecurity. AI's remarkable computational capacity empowers proactive identification of potential threats, and its customized recommendations foster a culture of cyber warfare. Despite these advancements, challenges such as inherent biases, adversarial vulnerabilities, and false positives can compromise its effectiveness and reliability. To fully harness AI's potential while mitigating its drawbacks, a judicious balance between its strengths and human capabilities is indispensable. This paper meticulously examines the evolution of AI in cybersecurity, encompassing diverse roles, solution categories, specific use cases, and AI methodological approaches.

The analysis findings reveal that while the volume of publications in this field has surged, practical implementation of AI-based cybersecurity solutions necessitates a heightened emphasis on collecting and presenting historical data pertaining to various cybersecurity functions. The classification of pivotal research, aimed at synthesizing the existing body of literature and comprehending AI's significance in cybersecurity, constitutes the primary contribution of this work. Moreover, the paper proposes future research avenues to address emerging challenges and optimize AI's effective application in the domain of cybersecurity.

REFERENCES

[1]. Islam, S., Hayat, M. A., & Hossain, F. (2023). ARTIFICIAL INTELLIGENCE FOR CYBERSECURITY: IMPACT, LIMITATIONS AND FUTURE RESEARCH DIRECTIONS. In © 2023 JETNR | (Vol. 1).
[2]. Noor, A., Nafis, T., Wazir, S., & Sarfraz, M. (2021). Impact Of Artificial Intelligence In Robust & Secure Cybersecurity Systems: A Review. SSRN Electronic Journal. https://doi.org/10.2139/ssm.3834207
[3]. https://zvelo.com/ai-powered-malware-holds-potential-for-extreme-consequences/
[4]. https://zvelo.com/ai-powered-malware-holds-potential-for-extreme-consequences/
[5]. https://zvelo.com/ai-powered-malware-holds-potential-for-extreme-consequences/
[6]. https://www.impactmybiz.com/blog/how-ai-generated-malware-is-changing- cybersecurity/
[7]. https://www.hyas.com/blackmamba-research-whitepaper
[8]. https://www.techmagic.co/blog/ai-in-cybersecurity/
[9]. https://www.ncsc.gov.uk/report/impact-of-ai-on-cyber-threat