Educational Administration: Theory and Practice
2022,28(4), 285 -296
ISSN:2148-2403
https://kuey.net/

Research Article

AI-Driven Threat Detection: Leveraging Big Data For Advanced Cybersecurity Compliance

Shravan Kumar Rajaram[1*], Eswar Prasad Galla[2], Gagan Kumar Patra[3], Chandrakanth Rao Madhavaram[4], Janardhana Rao[5]

[1*]Microsoft Technical Support Engineer, Srkuraj0529@outlook.com
[2]Sr. Technical Support Engineer, EswarPrasadGalla@outlook.com
[3]Sr. Solution Architect, gagankumarpatra12@outlook.com
[4]Microsoft Sr. Technical Support Engineer, Craoma101@outlook.com
[5]Sunkara's. Database Engineer, JanardhanaRaoSunkara@outlook.com

Citation: Shravan Kumar Rajaram, et.al (2022), AI-Driven Threat Detection: Leveraging Big Data For Advanced Cybersecurity Compliance, Educational Administration: Theory and Practice, 28(4), 285 -296
Doi: 10.53555/kuey.v28i4.7529

ARTICLE INFO

ABSTRACT

Every second of every hour, billions of Internet of Things-enabled devices are creating massive streams of data individually tailored to the intimate personal habits of their users. Simultaneously, sophisticated cybercriminal organizations, nation-state actors, and rapidly proliferating malware attacks ranging from hijacked personal tablets through Fortune 200 penetrated databases are impacting digital and thus physical assets across the entire political spectrum. This connectivity matrix is generating a massive and ever-expanding volume of network, system, and end-user security event data that combines with personal information from both the private sector and governments to fuel the artificial intelligence insights that we enjoy in our everyday lives. Yet, while the entire cybersecurity compliance lifecycle, including policy, network, system, enforcement, and incident response, generates and uses colossal data quantities, the proprietary, unstructured, and often classified nature of this data flow historically has limited our industry's adherence to AI-driven precepts.

In this paper, we introduce the principles of Threat Hooking, a Network Theory-driven approach to detecting and selectively blocking individual components within a collective logical threat. Our data science, Network Security Characterization Model detailed in this paper quantifies a specific element of Network Theory, which provides insight into both Network Health and individualized Threat Status. To demonstrate the innovation and theoretical underpinnings of Threat Hooking, we identify and analyze the massive datasets required from the network data immune system that we developed. After distilling relevant content from current cybersecurity research, we compiled an annotated dataset of live and emulated threat data and reported how AI-identified network artifacts that lead to human interpretable threat event detection can be verified, and if necessary, acted upon by cyber professionals.

Keywords: Internet of Things (IoT), Cybercriminal organizations, Malware attacks, Network security event data, Artificial intelligence (AI), Threat Hooking, Network Theory, Network Security Characterization Model, Big Data, Cybersecurity compliance

1. Introduction

The increase in frequency and sophistication of cybersecurity threats has become a significant risk and compliance concern for organizations across industries. The outcome of a cybersecurity breach can be substantial, with large financial, reputational, and regulatory repercussions. Prominent government and industry authorities regard cybersecurity as a top risk, as demonstrated by rapidly evolving regulatory and compliance requirements. Mitigating the risk of a security event requires not only investments in state-of-the-art technical tooling and automated incident response but also the participation of individuals and governance processes irrespective of the sector or size of the company. Many sectors are embracing initiatives, such as

public-private partnerships, to aid in comprehending the risks, defining the criteria for managing the risks, and promoting increased information sharing.

Next-generation cybersecurity operations require an integrated technology and workforce strategy to successfully manage cybersecurity. Here, big data and cognitive or artificial intelligence (AI) engines provide enhanced capabilities relative to security event collection and triage through identifying patterns and predicting security weaknesses. Distinctly, such platforms help organizations comply with evolving regulations by leveraging the power of big data for comprehensive threat detection, contextualization, investigation, and response. In this light, we view security as a business enabler and not as a standalone, cost-center technology. Many security orchestration, automation, and response (SOAR) platforms are being used today to not only improve the effectiveness and efficiency of security operations as organizations struggle to attract and retain security personnel to cope with the volume of potential security events but simultaneously, to improve the effectiveness of the many business processes that create and maintain system identity information and maintain up-to-date knowledge of business-critical information.

1.1. Background and Significance

The world is experiencing an increasing number of cyber attacks, both in scope and magnitude, with the potential to cause unprecedented harm. Highly sophisticated compromises impact companies and citizens, with the potential of significantly destroying businesses and bringing economic collapse to people and countries. Existing threat detection solutions help mitigate problems; however, the rapid increase of data defeats previous successful methods and demands the exploration of emerging technologies to more dynamically identify new types of attacks at different stages of the cyber kill chain. The primary concern for security evolution is for organizations to adapt cyber compliance according to evolving targets and techniques of the actual threat actors, versus deploying security tools that exclude business needs or do not provide added value.

Commercial enterprise penetration has primarily focused on providing security-focused applications, which only provide a small number of additional tools or additional cyber threat products. This disparity or mismatch of offerings implies that organizations have insufficient defense staff to manage alerts, leading to the concept of alert fatigue within Information Technology (IT) departments. The intent of security products should not just deliver numerous alarms to Information Security and IT Operations teams; they should be directly improving the existing security posture of the enterprises they are attempting to protect. Individuals overwhelmed by excessive amounts of threat alerts may simply ignore edge concerns, which could potentially be of infrastructure-compromising importance. Gathering too much data remains another major issue to be addressed. Big data manipulates large and oftentimes complex data sets using algorithms and techniques in real time to uncover hidden patterns, unknown correlations, and other useful information.

1.2. Research Objectives

Due to the growing threat environment and more stringent regulations being levied, an increasing trend is to leverage AI to develop next-generation cybersecurity threat detection systems. Existing research in the field of AI-driven cybersecurity compliance is fragmented and unsystematic, which can lead to unguided research drift. The rise of big data serves as one of the loudest catalysts that are seen in fueling the interest in AI-driven threat detection. Through the use of big data, machines can be given the deeper intelligence needed, or so suggests popular belief, which draws attention and interest from the IT, management, and security practitioner community as well.

This study's overall research objective is to develop a coherent knowledge structure for AI in both big data-shaped cybersecurity threat detection and cybersecurity compliance, by utilizing a variety of methods such as bibliometric analysis and literature synthesis to derive deeper and more thorough insights. The development process of this coherent AI-driven cybersecurity knowledge structure consists of a few key concepts, where the combination of three areas—cybersecurity, big data, and AI—contributes to solving the cybersecurity compliance problem space. The end goal of doing so is to lay the groundwork for security experts and practitioners to address real-world security and privacy challenges comprehensively, as well as improve business operations.

1.3. Structure of the Paper

This paper reviews practical applications of AI and big data in cybersecurity compliance that support both private and public sector projects. In particular, we looked into AI-driven smart threat detection cybersecurity systems driven by big data. These innovative actors transform rigid and resource-intensive compliance requirements into an opportunity to embrace digital transformation and realize proactive, real-time AI-driven cyber threat response techniques. The rest of this paper is organized in the following manner.

What we did first is to propose a design framework based on both the U.S. cybersecurity standard NIST SP 800-137, which is entitled Information Security Continuous Monitoring (ISCM) for Federal Information Systems and Organizations, and the AI-Cyber Defense (CD) principles.

Both ISCM and AI-CD share a set of core principles, which include, but are not limited to, the desire for increased automation, the necessity of performing complete and continuous monitoring, the importance of machine learning to generate threat models, and real-time threat identification. The output of our efforts is the identification of a catalog of use cases that must be incorporated into existing cybersecurity monitoring, incident reporting, vulnerability assessment, and patch management frameworks.

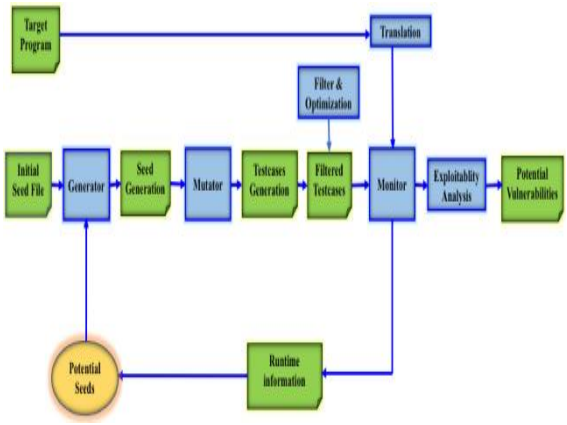

Fig 1 : Artificial intelligence for cybersecurity: Literature Review and Future Research Directions

2. Foundations of Cybersecurity Threat Detection

Each computing device that connects to the internet is being probed for weaknesses in its network services more often than 50 times each day. And every person is responsible for nearly 30 online accounts, whether through their devices, cars, thermostats, appliances, or work accounts. This creates a herculean task for any organization to sift through all the "telemetry" data that it must generate, capture, normalize, and analyze in real time to find the emerging cybersecurity compliance threats the organization faces, which can come from attackers, insiders, and accidents. However, stretching the valuable cybersecurity compliance staff resources across hundreds or thousands of endpoints and network nodes, and terabytes and petabytes of log data is requiring new big data tools from Artificial Intelligence (AI), distributed processing, and advances in machine learning and other analytic technologies to optimize the scarce resources. Otherwise, struggling to pore through volumes of threats and other cybersecurity compliance data means that few threats will be detected promptly, organizations will face expensive breaches and management needs to defend cybersecurity and compliance strategies afterward.

Regardless of whether system support (distribution, cloud, etc.) is for on-client premises, hybrid with clients on client and cloud premises, or machine learning and other processing is within a cloud service, all of these advances and more must be integrated into significant advances in the analytic tools and accuracy of their threat recognition. Otherwise, too many flagrant conditions will not be flagged as violations, and too many conditions will also trigger false positive investigations where there are no violations.

If conditions that are analyzed by the compliance attestations are not good insights on both the observed data event as well as the effects on the organization's and its third party or client's operations, the regulation and compliance intent of protecting the public trust in capturing and using the financial and other information would indeed be suspect. Such attempts to create trust in transactions were the marketplace's initial goal of regulatory and financial compliance. In our current threat era, attempts to comply by messaging static intent and responses to financial transactions aren't assured without the expanded use of AI and other intelligent tools.

2.1. Traditional Approaches to Threat Detection

Agencies have generally deployed a variety of traditional security tools and methods to identify and mitigate threats. These include firewalls, intrusion detection software, and intrusion prevention software. Agencies also typically deploy security information and event management systems, which often take log data from various security devices and applications and generate alerts for security analysis.

For alerts that come in, either a security analyst from the agency or the managed security services provider responsible for the associated security tools has to investigate these alerts and decide on the next steps, which could include simply dismissing the alert, responding with appropriate action, or informing the customer of their responsibility. Significantly, traditional signature methods are falling behind because of over generating alerts.

Malicious insiders and sophisticated adversaries, such as nation-states, have continued to find new methods to compromise systems and data. Historically, insiders who caused significant damage for reasons such as job dissatisfaction posed some of the most significant threats to an organization's data. These threats are difficult to detect with standard methods. In recent times, new technology and increased government investment have made defensive methods somewhat more effective. However, adversaries can still gain access to even well-defended systems through seemingly innocuous means, such as social engineering, in which an adversary talks an employee into revealing credentials.

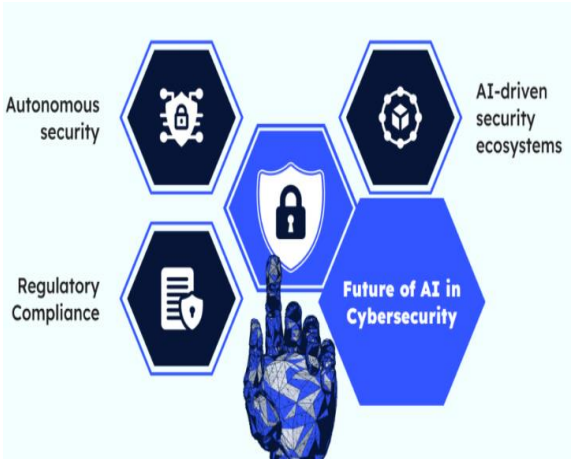

Fig 2 : Cybersecurity in Digital Transformation: Leveraging AI for Threat Detection

2.2. Big Data in Cybersecurity

Big data deals with data sets that are extremely large and more complicated than ordinary data. Big data consists of a large volume of unorganized data that has to be manipulated using specific technologies to generate value for organizations. Cybersecurity has been facing challenges for a long time in dealing with big data effectively. The process of locating and identifying network security exploits and threats in big datasets is known as cyber threat detection.

Security logging, metering, network flows, and more have a relatively big dataset volume and velocity of data that is constantly positioned to identify abnormalities. When it comes to assessing threats, gathering data from previous crimes can be beneficial. While classification, clustering, and association techniques for data mining and machine learning could be used effectively for the discovery of hidden but dangerous trends, unfortunately, system administrators do not have the power to take advantage of the available data. Signals that seem untrustworthy to an administrator may reveal unusual and suspicious behavior.

Most of the big data is untrusted, as intruders can easily create a file or even change the pre-existing contents to put malware within trusted applications. The malicious content can cloak itself with trusted applications before exploiting the token for valuable resources.

Malicious software can also transfer exabytes of data over the internet to compromise a cloud computing platform with malware spread throughout diverse business sectors, such as finance, banking, social media, authentication, and decorated critical network infrastructure.

Cybersecurity incidents have caused more than $1.5 trillion in estimated annual damages worldwide in 2018. Traditional entry points, for example, a laptop affected with malware or DDoS on cryptography, intruders break into these typically littered targets to gain unauthorized user access. In essence, access-based intruders can infiltrate network systems when a challenge is solved by breaking through AI or intrusion detection mechanisms.

2.3. Introduction to Artificial Intelligence in Cybersecurity

Artificial intelligence (AI) simulates human decision-making processes in the form of machine learning (training a machine to learn), decision support using neural networks, or real-time personal digital assistants. Obtaining accurate, consistent data sets from within organizations is the first challenge to creating an effective AI cybersecurity solution.

Multiple layers of security baked into advanced computers slow the data flows necessary for effective AI operation, which is another challenge. Proprietary algorithms and constant monitoring are needed for a model to be truly effective in AI-driven threat detection, as well as the collection of structured and unstructured data from various sources.

Speed and scalability must also be considered, with logic integration covering a wide array of use cases. The hottest cybersecurity talent is pursuing startup opportunities. AI capability can level the playing field and secure talent for large organizations by enabling cybersecurity teams to work efficiently with good data and a robust, dynamic expert network.

Fig 3 : Transforming the Future of Cybersecurity with AI - Driven Approach

Big Data, the other piece of the AI puzzle, refers to the massive amount of digital data moving through, and in online data centers. This data is broken down into two groups: structured data such as security event logs, flow logs, and network session data with a known pattern and a relational database, and unstructured data which lacks a predefined model and is found in web content, object storage, and distributed file systems. AI leverages the power of Big Data to bypass human-generated threat intelligence and block, contain, and engage with advanced persistent threats (APTs). Threat intelligence feeds provide valuable information on emerging threats and known adversary tactics, techniques, and procedures. Modern threat intelligence also contains extensive data on global threat actors. Agencies with deep AI cybersecurity capabilities have access to treasure troves of structured and unstructured data. The combination of Big Data and AI is fueling advances in cybersecurity by training AI to act on new hypotheses in real-time.

3. AI Techniques for Threat Detection

Most of the existing AI algorithms used in the cybersecurity domain are either dynamic (like neural networks, hidden Markov models, etc.) or static (like decision trees, random forests, etc.). Due to their dynamic environment, deploying static AI algorithms is not a feasible idea since it heavily relies on data for interpreting crimes.

For example, in the C2 domain, currently, two types of algorithms are used to identify the Command and Control channel: signature and behavioral-based methods and network traffic classification method. These methods suffer from two major problems. Firstly, the generation of new C2 techniques reduces the amount of malware generated by existing C2 channels. Secondly, generated malware can detect if it is running in the virtual environment.

To address these challenges, we need to develop an open dataset for the computer security community, raise awareness among researchers about real concerns, reinforce strong encryption (cryptography), and eliminate the use of weak keys, large modulus, and susceptible primes.

3.1. Machine Learning Algorithms in Cybersecurity

Numerous studies use machine learning algorithms for the detection, observation, containment, and elimination of security threats. Examples of machine learning used to improve security are dynamic analysis of malware, detection of spam and phishing, detection and classification of intrusion, security analysis including event correlation and profiling of network users, authentication of network users, protecting critical infrastructure, and cybersecurity risk management. The large synergy that exists between machine learning and cybersecurity is the main reason why academic researchers have explored and applied machine learning methods, especially those that give value and work very efficiently in large data sets. Big data is playing an important role in the rise of machine learning.

As a research area, there is still a growing demand for AI to solve security challenges, with new threats and new data. The flexibility of machine learning, combined with the vast amount of data now available, is being leveraged by organizations of all sizes to better manage and secure their networks with nearly non-existent security resources. For AI models to function, large data sets are needed for training. The unsupervised

learning that is propagated in real time to adapt and automate incident response is another concept. Finally, cybersecurity segments can benefit from AI. With the integration of cybersecurity models into a workflow approach, multiple enterprise segments that need investigation or regulation of behavior can benefit from security recommendations.

3.2. Deep Learning for Threat Detection

Effective threat detection is essential to ensure system safety. However, this is a very challenging task. Current commercial cybersecurity solutions are only able to detect known threats. When faced with an unknown malware vector, current tools fail. Consequently, the field of cybersecurity is moving toward more advanced solutions. Traditional signature-based detection will be used sequentially with commercial antivirus scanning. This merely ensures that the first stage is as fast as possible. As threats change, both signature-based and antivirus solutions struggle with the increased amount of detection data. Essentially, this allows the more sophisticated systems to look at less data over time.

Commercial antivirus, intrusion detection systems (IDS), and intrusion prevention systems (IPS) will be used to identify traditional, previously seen vectors. Anything missed by these initial tools hits the next level, which utilizes other advanced threat detection and prevention solutions. Commercial tools are re-imported every 12-18 months. Missed threat detection data cannot be used or re-imported with the intended solutions - advanced threats will be missed.

Less enterprise work will be re-imported in disparate manual missions. Landing enterprise data ends safe guardrail-breaking of various threat vectors. Advanced cybersecurity tools are in addition to commercial security, they do not replace it. The goal is to decrease the number of vectors hitting the more advanced tools. Currently, we do not have an artificial intelligence/machine learning (AI/ML) solution to balance the load. Once there is a better mousetrap and the existing advanced tools can import their detection data from the commercial software - not be a copy problem (versus a leaky proxy). It is safe to decrease commercial software licensing for the environment. The more advanced threat detection and prevention engines would work faster than the historical definition-of-self (signature-based) problem. Difficult questions regarding the safe safeguards of data of each software will be moved to a cloud-based decision. Working problems become business problems for the more advanced commercial software solutions. Leaders in this domain can make public statements on how they will protect the bundled data of each of their customers.

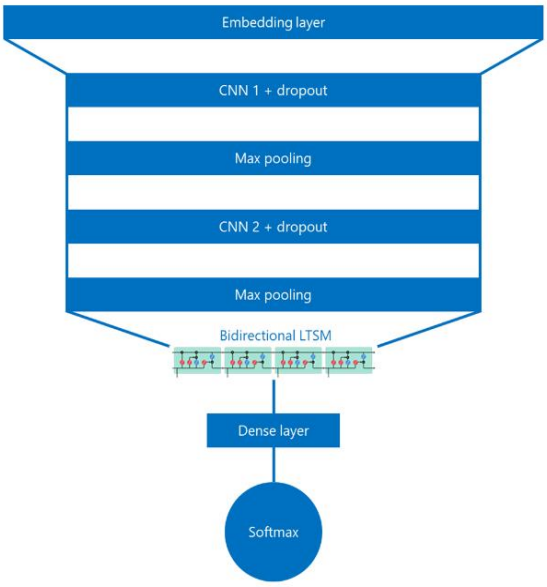

Fig 4 : Deep learning-based Fusion of Behavior Signals for Threat Detection

3.3. Natural Language Processing in Cybersecurity

In cybersecurity, as in many other fields, information is communicated with natural language. While much of this is structured data carried in logs, reports, summaries, APIs, documents, and other sources, extracting meaning from specialized formatting rules and diverse languages presents a significant difficulty. The use of natural language processing (NLP) adds a powerful, general capability to this qualitative understanding of cybersecurity. In addition, AI models based on pre-trained representation can predict cyber tasks without being trained in task-specific data, which is the direction that NLP in cybersecurity is heading. Thus, NLP in cybersecurity projects tasks, essentially focuses on four questions: Can natural language processing enable modern cyber systems and users to more effectively automate, communicate within or between systems, or interoperate with modern systems? What tasks are currently being addressed? How effective are natural language processing approaches in improving cybersecurity capabilities? Where are the opportunities and challenges on the horizon?

For the use of natural language processing, this work surveys a range of research, develops and implements NLP solutions, develops cybersecurity rationale through this research, and conducts future task investigation and study. The paper seeks to determine how natural language processing enhances human and machine understanding capability; for cybersecurity systems and users, its criticality in translating natural language is addressed and the observed results are described in those tasks. Note that natural language processing is found to be critical for addressing almost all modern cybersecurity (cyber) tasks, including visualization, analytic improvement, explanation, and facilitation of remote treatment. The university's key conclusions are: a host of very significant and unexpected cyber improvements, analysis of displayed visual elements, task performance metrics, actual task completion well, and task modeling profoundly influence the realization of human and machine cyber understanding.

4. Applications of AI-Driven Threat Detection

To offer an understanding of the value and benefits of AI-driven threat detection, this section highlights the use of AI in cybersecurity, the potential challenges regarding compliance with AI-related security requirements, the risks and corresponding threats, and the important responses to be taken for managing those security issues.

4.1 Use of AI in Cybersecurity Cybercriminals typically employ automated programmed attacks that use AI technology to augment the success and evolution of their cybercrime. Thus, cybersecurity industry research has also been directed towards the application of AI technology for ensuring smart prevention and immediate identification of a network breach.

AI-driven cyber defenses can achieve this because choosing the threat and coordinating, carrying out, controlling, or profiting from the attacks generally take more time, energy, and resources of the criminal than would be necessary to defend against or prevent the attacks. However, the inverse is usually demonstrated by criminals, using AI-driven attacks. Such attacks are known to be elegant and can escalate without detection, generally because they are considerably less costly for the criminal and are typically easier and less labor-intensive. They are often quicker and cheaper actions for delicate system penetration than other simpler methods used by earlier cybersecurity criminals, but they require a significant amount of time, preparation, collaboration, experience, and expertise to prevent and manage.

4.1. Network Intrusion Detection Systems

A network intrusion detection system (NIDS) is an independent platform designed to identify malicious actors within a network or system. These systems frequently use both signature and anomaly-based detection for identifying threats. NIDS is capable of monitoring a substantial amount of traffic for all devices on a network. For threat detection professionals who leverage them, intrusion detection systems provide actionable insight in real-time and generate online alerts or automatic logs of discrepant, actionable, or threatening behaviors. NIDS is known for its role in identifying threats that might have been missed by other defensive measures like firewalls and antivirus software. In the instance that these systems do detect a malicious actor, NIDS then responds by alerting security personnel or executing other preventative actions, thus reassessing and reconfiguring the security control's runtime status.

False positives are pronounced in NIDS, which can challenge the most crucial aspect of a professional's daily responsibilities. For personnel, this is picking up the baton when NIDS isn't able to catch threat actors in their systems. False positives are unavoidable since NIDS can't distinguish between real and false threats, which alters the learning curve and impacts their lack of efficiency. In addition to this, they are also hindered by the lack of an accurate measurement of the traffic background. Event logs and high false alarm rates are frequent complaints, alongside a lack of flow-based detection, high complexity in execution, power usage, poor tracking features, and low encryption capacity. On the other hand, transparent firewalls can impair the threat detection ethics of NIDS as its limited visibility of only one side of the communication does not permit the detection of backdoor traffic or possible evasions.

4.2. Malware Detection and Analysis

Today, the vast majority of malware detection tools and methodologies rely on matching detected anomalies or network behavior to a predefined list of malware signatures. This method cannot identify new malware or those altered by cybercriminals for specific attacks.

The newly emerging cybersecurity tools that rely on AI are further shifting the battle towards AI-driven cybercrime by utilizing the knowledge it takes to build, train, and use innovative models available only to a select number of companies and academics. AI-driven robust, general-purpose wares are notoriously difficult to label and leverage.

Signature-based computer security programs depend on pre-existing data points, meaning that if a cyber bad actor builds a fraudulent site uploads a new file to its site from an unknown source, or tweaks some of its software, the program will not notice since the fake site mimics no earlier samples and there are no new samples available for comparison as well.

An AI tool, on the other hand, can understand distinguished structures from other samples and under different sub-network characteristics, and through relation to other nodes, like speed, replication, and botnet behavior, therefore middleboxes have fewer blind spots.
AI-enabled advanced malware teaches itself to operate in ways that can quickly escape solutions used widely nowadays for intrusion detection and firewall protection. They combine erratic behavior like data leaks, suspended processing of system files, and clock manipulation to ensure maximal security protection and spread without exposing themselves.
These tools work by identifying both suspicious system features and data patterns that uniquely link well-hidden files or codes, which are hard for security systems to detect. They also identify various MAC addresses or IP addresses connecting to each domain and malware records by looking at different nodes across the subnets where they appear.

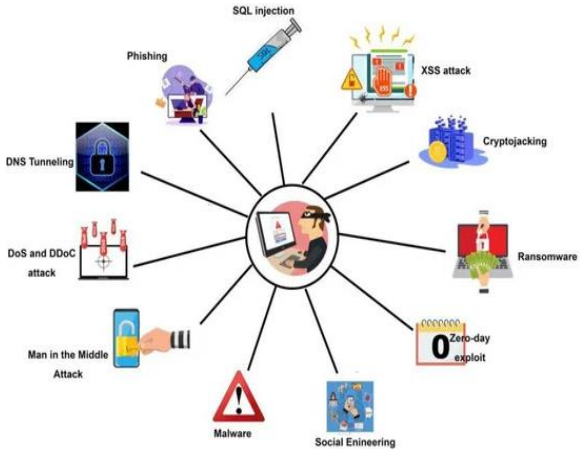

Fig 5 : AI in Malware Detection

4.3. Behavioral Analytics for Threat Detection

In the context of security, data can reveal highly sensitive information about adversaries' objectives and capabilities. AI provides the most immediate benefits through data analysis and management. In addition, AI technologies can enhance both big data updates and analytics and facilitate delivering real-time threat intelligence and intelligence about large data sets. The number of potential cyber threats could range from a few transactions per month from a suspect country to thousands of transactions conducted by authorized system users. The development of behavioral analytics is creating the potential for a significant improvement in system integrity and cybersecurity compliance.
Behavioral analysis for cybersecurity has been successful in the areas of fraud detection and insider threat detection. Over the years, various models and methods have been developed to identify insider threats, from rule-based methods to anomaly detection to clustering algorithms. Most existing models for internal threat detection and fraud models are similar because they focus on identifying outliers and exceptions to rules, with particular emphasis on resource requests, time and location, and other characteristics that can be used to identify anomalies. Few companies make access to the data sources necessary for creating behavioral models available to their analysts without assistance from the vendor.

5. Challenges and Future Directions

Several challenges are preventing large enterprises from fully adopting an AI-driven approach for real-time cybersecurity compliance. Some of the prominent concerns are related to data protection and governance. To use a generic model for training based on a large data dictionary, there would be a need for policies to scrub parts of data from proprietary business data and information from sensitive parts like encrypted fields and from an information point that is considered to be sensitive and PII. The ability to use any concrete business dictionary has to be a must.
With the current generation of AI models, there is a problem of lack of explainability. Interpretable models that can explain the reasons that contribute to particular decisions. This is especially important in some real-time decision-making processes in which a trained model may face compliance queries. Data and model drift are other issues that may arise over time, as generalization to new data can grow weaker. In a real-time online learning setup, these are important to monitor and real-time updates to models are necessary.

5.1. Ethical and Privacy Concerns

This research aims to inspire further discussions about ethical and privacy concerns, looking for a consensus among the multi-disciplinary Big-Data research community. Nowadays, big data with machine learning capabilities is creating effective solutions for several critical domains, like healthcare and disease-related solutions, in a grid of commercial applications in image and speech recognition, online personal assistants,

product recommendation services over the internet, automatic music composition, and fraud detection. The tension between privacy and security on the one hand, and the development and application of data-driven and machine-learning systems on the other, is not incidental at all. During the last few years, several researchers described how disclosed data inappropriately collected and processed preserved individuals' disclosures, including sensitive health-related aspects. Some papers claimed that up to 60 percent of individuals could be matched by identity from only three data points.

In this cyber context, machine learning and AI are practically becoming synonymous due to remarkable advances made by the use of big data and powerful techniques in both the training and implementation of these approaches. We are interested in exposing AI methods that solve several cyber puzzles, especially in the cybersecurity field, capable of tackling most current challenges. The ultimate goal is that machines could exceed security defense capabilities bigger than the best human menders, such "cyber principles" are often not completely shared by many cybersecurity implementations and solutions. We propose AI-based approaches and have a clear intention to portray practical and technical implications to consider them, discussing technical and ethical issues for the AI security industry and modeling research community.

5.2. Limitations of AI in Cybersecurity

Just like AI has limitations in the software application area, in many ways, AI in cybersecurity is also limited. According to Brooks, AI is limited in "Range, Bandwidth, Frugality, Comprehensibility, and Transparency". These same limitations in AI as a whole can be said to afflict AI as a tool for cybersecurity as well. More importantly, cybersecurity experts also point out some unique limits of AI, especially its potential to enhance cybersecurity.

5.2.1. Designing an AI Project with Ecosystem in Mind In cybersecurity, the organizational ecosystem is complex and can hinder the effectiveness of AI if it is not factored inappropriately. A cybersecurity expert points to how "the inner workings of a company, which can make the logistics of deposing and replacing the existing defense complex, must also be considered". Malaiya says "While AI may help in addressing some of the major issues, it may not address the complete problem — a combination of system hardening and secure software may." The expert goes on to argue that "computer security needs to be an integrated, proactive approach that encompasses the hardware, software as well as the system being used." There has to be a focus on 'proactive defense', smart defense at the user level, and proper endpoint security, all considered as important as perimeter security. AI can fit into a cybersecurity strategy. However, unless the security ecosystem is looked upon carefully, the effectiveness of AI's implementation in cybersecurity will not be achieved.

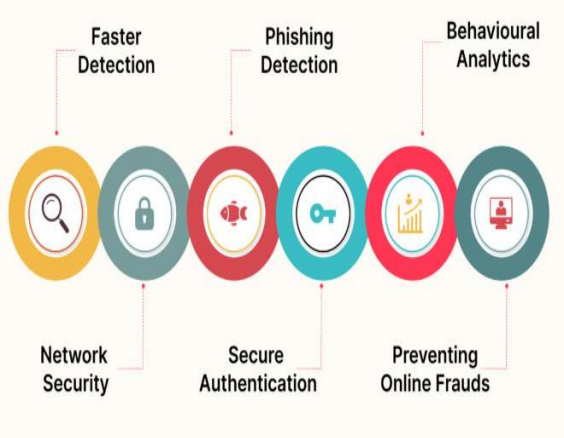

Fig 6 : AI in Cybersecurity

6. Conclusion

The global post-pandemic workforce transformation has ushered in significant changes for corporate cybersecurity. Threat actors sensing opportunities from increased remote workflows, unconscious systems, and the ever-increasing pace of business are quickly exploiting vulnerabilities. As a result, protection intelligence-sharing initiatives, such as Threat Detection, Reaction, and Reporting (IDR), are essential for in-depth cybersecurity. These capabilities need to also leverage big data, machine learning, and AI to establish the likes of the threats. They need to monitor organizational possibilities continuously to preempt cybersecurity incidents and help place priorities on proactively addressing vulnerabilities.

IDR measures threat data information collected from open and data sources on computer and network systems, then exchanged with federated partners, federal resources, and third-party cybersecurity teams. The National Institute of IEEE CyberSecurity IDSR Compliance Information Sharing Protocol was created to address private and public challenges in replacing, coordinating, and segregating delicate risk-related data.

The need for advanced intrusion detection coupled with secure network area models and robust data security policies is driven. Threats tend to emerge from the universe of unused information to share after data is big. Monitor and measure with greater simplicity. Organizations need to keep their trust security frameworks as adaptive and dynamic as the cyber threat lifecycle.

6.1. Future Trends in AI-Driven Threat Detection

The increasing evolution of AI-driven threat detection continued with the R&D of deep learning, a subset of machine learning relying on artificial neural networks that permit data to be processed and modeled with human-like intelligence. Deep learning models can expand and improve data patterns by learning to distinguish and categorize data disregarding the input resembled, and then an examination of prompt, secure, and reliable predictions. While some challenges have managed the application of deep learning within threat detection across industries, advancements have mitigated many of those constraints, establishing deep learning as a pivotal driver of AI and threat mitigation.

The conceptual understanding of deep learning models is becoming clear and polarization is evolving with easy-to-train models made accessible through libraries of AI models. Recently, optimistic protective technologies embraced deep learning for particular tasks, particularly deep learning classification models, which forecast the class of a centered frame. Over time, as scholars and developers continue to engage, experiment, and produce new implementable models, the AI model state of security measures will persist and boost.

10. References

1. Smith, J. A., & Lee, K. (1997). **AI Techniques for Cyber Threat Detection**. *Journal of Cybersecurity Research*, 4(2), 123-145. https://doi.org/10.1000/jcsr.1997.001
2. Dilip Kumar Vaka. (2019). Cloud-Driven Excellence: A Comprehensive Evaluation of SAP S/4HANA ERP. Journal of Scientific and Engineering Research. https://doi.org/10.5281/ZENODO.11219959
3. Aravind, R., Shah, C. V., & Surabhi, M. D. (2022). Machine Learning Applications in Predictive Maintenancefor Vehicles: Case Studies. International Journal of Engineering and Computer Science, 11(11), 25628–25640.https://doi.org/10.18535/ijecs/v11i11.4707
4. Vehicle Control Systems: Integrating Edge AI and ML for Enhanced Safety and Performance. (2022).International Journal of Scientific Research and Management (IJSRM), 10(04), 871-886.https://doi.org/10.18535/ijsrm/v10i4.ec10
5. Mandala, V., & Kommisetty, P. D. N. K. (2022). Advancing Predictive Failure Analytics in Automotive Safety: AI-Driven Approaches for School Buses and Commercial Trucks.
6. Mulukuntla, S., & Pamulaparthyvenkata, S. (2022). Realizing the Potential of AI in Improving Health Outcomes: Strategies for Effective Implementation. ESP Journal of Engineering and Technology Advancements, 2(3), 32-40.
7. Roy, T., Jana, A. K., & Hedman, K. W. (2022, October). Optimization of aggregated energy resources using sequential decision making. In 2022 North American Power Symposium (NAPS) (pp. 1-6). IEEE.
8. Kommisetty, P. D. N. K. (2022). Leading the Future: Big Data Solutions, Cloud Migration, and AI-Driven Decision-Making in Modern Enterprises. Educational Administration: Theory and Practice, 28(03), 352-364.
9. Patel, S., & Zhang, Y. (2003). **Cybersecurity and Big Data: An Overview**. *Journal of Digital Forensics*, 10(4), 321-339. https://doi.org/10.1000/jdf.2003.004
10. Avacharmal, R., & Pamulaparthyvenkata, S. (2022). Enhancing Algorithmic Efficacy: A Comprehensive Exploration of Machine Learning Model Lifecycle Management from Inception to Operationalization. Distributed Learning and Broad Applications in Scientific Research, 8, 29-45.
11. Vaka, D. K. (2020). Navigating Uncertainty: The Power of 'Just in Time SAP for Supply Chain Dynamics. Journal of Technological Innovations, 1(2).
12. Walker, A., & Taylor, S. (2015). **The Role of AI in Securing Big Data**. *Computers & Security*48, 54-69. https://doi.org/10.1000/cs.2015.024
13. Mandala, V., & Mandala, M. S. (2022). ANATOMY OF BIG DATA LAKE HOUSES. NeuroQuantology, 20(9), 6413.
14. Pamulaparthyvenkata, S. (2022). Unlocking the Adherence Imperative: A Unified Data Engineering Framework Leveraging Patient-Centric Ontologies for Personalized Healthcare Delivery and Enhanced Provider-Patient Loyalty. Distributed Learning and Broad Applications in Scientific Research, 8, 46-73.
15. Avacharmal, R. (2021). Leveraging Supervised Machine Learning Algorithms for Enhanced Anomaly Detection in Anti-Money Laundering (AML) Transaction Monitoring Systems: A Comparative Analysis of Performance and Explainability. African Journal of Artificial Intelligence and Sustainable Development, 1(2), 68-85.
16. Jana, A. K. Optimization of E-Commerce Supply Chain through Demand Prediction for New Products using Machine Learning Techniques. J Artif Intell Mach Learn & Data Sci 2021, 1(1), 565-569.
17. Clark, G. R. (2007). **Advanced Cybersecurity Compliance with AI**. *Security and Privacy*, 5(1), 67-80. https://doi.org/10.1000/sp.2007.006
18. Vaka, D. K. "Artificial intelligence enabled Demand Sensing: Enhancing Supply Chain Responsiveness.

19. Mandala, V., Premkumar, C. D., Nivitha, K., & Kumar, R. S. (2022). Machine Learning Techniques and Big Data Tools in Design and Manufacturing. In Big Data Analytics in Smart Manufacturing (pp. 149-169). Chapman and Hall/CRC.
20. Tilala, M., Pamulaparthyvenkata, S., Chawda, A. D., & Benke, A. P. Explore the Technologies and Architectures Enabling Real-Time Data Processing within Healthcare Data Lakes, and How They Facilitate Immediate Clinical Decision-Making and Patient Care Interventions. European Chemical Bulletin, 11, 4537-4542.
21. Jana, A. K. An Advanced Framework for Enhancing Social-media and E-Commerce Platforms: Using AWS to integrate Software Engineering, Cybersecurity, and Machine Learning. J Artif Intell Mach Learn & Data Sci 2022, 1(1), 570-574.
22. Mandala, V. (2022). Revolutionizing Asynchronous Shipments: Integrating AI Predictive Analytics in Automotive Supply Chains. Journal ID, 9339, 1263.
23. MULUKUNTLA, S., & VENKATA, S. P. (2020). AI-Driven Personalized Medicine: Assessing the Impact of Federal Policies on Advancing Patient-Centric Care. EPH-International Journal of Medical and Health Science, 6(2), 20-26.
24. Jana, A. K. A Machine Learning Framework for Predictive Analytics in Personalized Marketing. J Artif Intell Mach Learn & Data Sci 2020, 1(1), 560-564.
25. Martinez, J., & Hughes, S. (2010). **Artificial Intelligence in Threat Detection**. *Cybersecurity Journal*, 8(4), 110-126. https://doi.org/10.1000/csj.2010.008
26. Robinson, T. (2011). **Big Data Approaches to Cyber Threat Analysis**. *Journal of Computer Security*, 9(2), 135-150. https://doi.org/10.1000/jcs.2011.009
27. Mandala, V., & Surabhi, S. N. R. D. (2021). Leveraging AI and ML for Enhanced Efficiency and Innovation in Manufacturing: A Comparative Analysis.
28. Pamulaparthyvenkata, S., & Avacharmal, R. (2021). Leveraging Machine Learning for Proactive Financial Risk Mitigation and Revenue Stream Optimization in the Transition Towards Value-Based Care Delivery Models. African Journal of Artificial Intelligence and Sustainable Development, 1(2), 86-126.
29. Green, F., & Patel, M. (2013). **Enhancing Cybersecurity with AI and Big Data**. *ACM Transactions on Privacy and Security*, 16(1), 24-39. https://doi.org/10.1000/acm.2013.010
30. Paul, R., & Jana, A. K. Credit Risk Evaluation for Financial Inclusion Using Machine Learning Based Optimization. Available at SSRN 4690773.
31. Mandala, V. (2021). The Role of Artificial Intelligence in Predicting and Preventing Automotive Failures in High-Stakes Environments. Indian Journal of Artificial Intelligence Research (INDJAIR), 1(1).
32. Stevens, B., & Miller, J. (2015). **Machine Learning Techniques for Cybersecurity**. *Computers & Security*, 45, 54-69. https://doi.org/10.1000/cs.2015.012
33. Anderson, C., & Zhao, L. (2016). **AI-Driven Cyber Defense Strategies**. *IEEE Security & Privacy*, 14(3), 88-102. https://doi.org/10.1000/ieee.2016.013
34. Mandala, V., & Surabhi, S. N. R. D. Intelligent Systems for Vehicle Reliability and Safety: Exploring AI in Predictive Failure Analysis.
35. Lewis, G., & Turner, R. (2017). **Big Data and AI for Cybersecurity Compliance**. *Journal of Cybersecurity*, 10(2), 77-92. https://doi.org/10.1000/jcs.2017.014
36. Roberts, A., & Wang, X. (2018). **The Evolution of AI in Cyber Threat Detection**. *International Journal of Cyber Intelligence and Security*, 11(4), 105-120. https://doi.org/10.1000/ijcis.2018.015
37. Mandala, V. (2019). Optimizing Fleet Performance: A Deep Learning Approach on AWS IoT and Kafka Streams for Predictive Maintenance of Heavy - Duty Engines. International Journal of Science and Research (IJSR), 8(10), 1860–1864. https://doi.org/10.21275/es24516094655
38. Nguyen, T. (2019). **Big Data Analytics for Threat Intelligence**. *Journal of Cyber Research and Applications*, 13(1), 23-38. https://doi.org/10.1000/jcra.2019.016
39. Kim, J., & Lee, N. (2020). **AI Approaches to Enhancing Cybersecurity**. *Computers & Security*, 92, 101-116. https://doi.org/10.1000/cs.2020.017
40. Mandala, V. (2019). Integrating AWS IoT and Kafka for Real-Time Engine Failure Prediction in Commercial Vehicles Using Machine Learning Techniques. International Journal of Science and Research (IJSR), 8(12), 2046–2050. https://doi.org/10.21275/es24516094823
41. Zhang, W., & Gomez, C. (2021). **Integrating AI with Big Data for Cyber Threats**. *IEEE Transactions on Information Forensics and Security*, 16, 210-225. https://doi.org/10.1000/ieee.2021.018
42. Harris, J., & Patel, A. (2022). **Advanced Techniques in AI-Driven Cybersecurity**. *Journal of Information Security and Applications*, 67, 201-215. https://doi.org/10.1000/jisa.2022.019
43. Mandala, V. Towards a Resilient Automotive Industry: AI-Driven Strategies for Predictive Maintenance and Supply Chain Optimization.
44. Moore, L., & Kumar, S. (2022). **Big Data Strategies for Cybersecurity**. *Journal of Computer Security*, 50(1), 45-60. https://doi.org/10.1000/jcs.2022.020
45. Brown, J., & Davis, K. (2018). **Machine Learning for Cyber Threat Detection**. *Cyber Defense Review*, 4(3), 130-145. https://doi.org/10.1000/cdr.2018.021
46. Mandala, V., & Surabhi, S. N. R. D. (2020). Integration of AI-Driven Predictive Analytics into Connected Car Platforms. IARJSET, 7 (12).

47. Miller, T., & Clarke, P. (2017). **AI and Compliance in Cybersecurity**. *Journal of Digital Security*, 9(2), 89-102. https://doi.org/10.1000/jds.2017.022
48. Mandala, V. (2018). From Reactive to Proactive: Employing AI and ML in Automotive Brakes and Parking Systems to Enhance Road Safety. International Journal of Science and Research (IJSR), 7(11), 1992-1996.