JOURNAL OF NORTHEASTERN UNIVERSITY

Volume 25 Issue 04, 2022 ISSN: 1005-3026 https://dbdxxb.cn/ Original Research Paper

RISK AND ATTACKS AGAINST AI/ML, AS WELL AS THE CHIEF INFORMATION SECURITY OFFICER'S (CISO) POSITION ON IT

Amaresh Bose
Department of BCA, Saint Xavier College, Maharo, Jharkhand

Naghma Khatoon
Department of Computer Science, Usha Martin University, Ranchi, Jharkhand

Abstract:
Cyberattacks are today more pervasive, sophisticated, and technically advanced than ever. This circumstance calls for a quick response supported by artificial intelligence and capable of being predictive, cognitive, prescriptive and auto-reactive. Only security based on artificial intelligence will be able to identify previously unknown attack pattern, zero-day attacks, and former vulnerabilities. This is particularly crucial in light of the emergence of polymorphic malware, ransomware and persistent complex threats. Artificial intelligence offers a more diversified range of skills than machine learning. The phrase “machine learning” (ML) and "artificial intelligence" (AI) refer to the process of teaching computers to do human-like functions, such as learning and problem-solving (including hardware and software). The bulk of these AI systems are used in cybersecurity. The CISO is responsible for convincing manufacturers that their products must meet the criteria of the next generation of identity and access management systems and cyber security solutions, which will be enhanced by artificial intelligence.
Keywords: AI/ML, Cyberattacks, cybersecurity, CISO, Security, Network, System, Machine.

1. Introduction:

1.1.AI, ML, and DL for cybersecurity:

Before AI, ML, and DL for the many models of cyber security were even discussed in the field of information technology, essential security aspects were put in place using manual analytical and proactive reactions to information depends on hash-based identifiers or known past encounters made of earlier attacks, infections or oddities. This was done before AI, ML, and DL were discussed a lot as the future of cyber security worldwide. All the security measures for the network, host, server, device, application, and perimeter were based on this.

Today, cyberattacks are more common, sophisticated, and complicated than ever before. This calls for a real-time, artificially intelligent predictive, prescriptive, cognitive, and auto-reactive response. With the rise of ransomware, polymorphic malware, and sophisticated, persistent threats, only security based on artificial intelligence can find attack patterns, zero-day attacks, and other problems that haven't been seen before (APT).

Submitted: 08/11/2022 **Accepted**: 27/11/2022

Before we start, I want to point out that artificial intelligence has a broader range of skills than machine learning. "Artificial intelligence" (AI) is the process of teaching machines to do things like learn and solve problems like humans (including hardware and software). Most of this AI is used in cybersecurity through machine learning (ML) (and information technology in general). Even though ML is often thought of as its field, it is a subset of artificial intelligence (AI), where its main ideas come from. Machine learning (ML) is primarily interested in whether computers and other systems can be taught to act intelligently. In cybersecurity and IT, Despite the fact that these terms are often used simultaneously, there is a difference to be made between machine learning (ML) and artificial intelligence (AI). The phrase "AI/ML" is used throughout this paper, from the paper's title to its body, to show how these two ideas are similar and how they are different in small ways.[1-5]

1.2.In the Relatively Recent Past:

At first, only the following patterns were used to protect against cyberattacks.

The packet - filtering pattern was utilized by the first generation of stateless firewalls to offer network edge security. Firewalls execute firewall security actions at the OSI layer 3 and 4 integrated approach, taking the Source and Destination IP address and TCP and UDP Source and Destination Interfaces into consideration. The firewall and an early version of the forward proxy used similar data structures to quickly match domain names and IP addresses to make black-and-white list. This was done by setting up a firewall or a forward alternative.

The intrusion detection system (IDS) finds harmful activity on the network or either scanning for patterns (such as data - structures in network traffic) or by recognizing known dangerous instruction sequences used by malware. IDSs were not in a line because it would hurt their performance, and they could usually only spot certain types of attacks. Antivirus (AV) systems can protect endpoints and servers from malicious software using basic ideas like hash matching and digital signature matching. By comparing their hashes (often MD5 hashes) and digital signatures, executables were checked against a list of known harmful programs (binary patterns).[6]

1.3.The Present Situation:

The present condition of cybersecurity in the IT industry is a direct result of installing more security protection patterns and the early stages of supervise learn. Please remember that many RPA systems need more ingenious and can't use supervised machine learning. Even though RPA has helped automate routine tasks, this is still the case.

Here is a summary of some of the essential parts of the current situation and how they can be made better in the not-too-distant future with the help of AI-related skills.

1.3.1. ***Perimeter security-*** With the help of a web access firewall (WAF), superior proxies, an interruption prevention system (IPS), and a next-generation firewall (NGFW), the company has been able to do many things (for DDoS mitigation).

1.3.2. ***The next generation firewall (NGFW)-*** Using stateful multi-level inspection (SMLI) and deep packet inspection (DPI) covering OSI levels 2 through 7, you may compare each packet quickly to known data bits (such as malware) and other criteria to decide whether to block or permit data. This lets them choose whether to let data in or not. Next-generation firewalls (NGFWs) use a "stream-based detection" method to find threats. This method involves watching and analyzing real-time data moving across the network.

If malicious software or other threats aren't found by the Next-Generation Firewall (NGFW) in time, or if the Next-Generation Firewall (NGFW) doesn't post out a TCP rearrange pack in time to stop the traffic flow, the internal network could be broken into. The NGFW scans, called "stream-based scanning," may not catch malware that spreads slowly through fragmented packets. This is the case because of the scanning process.

Security companies have added more features to NGFW, like App-ID by Palo Alto Networks firewalls, which make it easier to figure out what a program is, no matter what port, protocol, encryption (SSH or SSL), or other ways it tries to hide. For example, several security companies have used Palo Alto Networks' App-ID firewalls. There are many different ways to classify the flow of network data, making it easy to find the correct application. These methods include application signatures, decoding the application protocol, and using heuristics. With this feature, the firewall can tell what software is running. There are so many ways that next-generation firefighting systems (NGFWs) could use supervised and unsupervised machine learning.

1.3.3. ***A web access firewall (WAF)-*** On HTTP servers that host web applications, rule-based logic, parsing, and signatures may avoid common security issues such as cross-site scripting (XSS), cross-site request forgery (CSRF), and SQL injection. These three security techniques are complementary to one another. Cross-site scripting (XSS), cross-site request forgery (CSRF), and SQL injection are the most prevalent kinds of these attacks.

You are able to use either controlled or uncontrolled machine learning algorithms inside the WAF. It is one of the first perimeter-based solutions that use both supervised and unsupervised machine learning techniques. The Web Application Firewall(WAF) was among the first instances of perimeter-based technologies.

1.3.4. ***Web Forward Proxies (WFP)-*** Customer safety is the most important thing. They offer a "quarantine" check for outbound web transfer, during whom they inspect all HTTP/HTTPS traffic among (local) client browsers and (remote) HTTP/HTTPS site (servers), labeling each URL so that harmful websites or page can be blocked-up. In contrast, good quality URLs can still be accessed based on the user's policies. Because WFP can classify URLs, IT security teams and administrators can set up more nuanced policies that block access to specific malicious information while letting users use the site. WFP can sort websites by URL, while NGFW can only sort them by domain. The main job of a reverse web proxy is to protect a server from unwanted outbound connections.

1.3.5. ***Next-generation Intrusion Prevention Systems (IPS)-*** Inline defenses can find and stop new or unknown attacks using signature-based detection and prevention methods based on anomalies (vulnerabilities and threats). This is done by setting up systems that can find and stop attacks before they happen.

When a firewall finds a malicious packet, it can either throw it away immediately, reset the connection (TCP), or stop all communication from the aberrant IP address. Plus, they are "context-aware," meaning they can take into account information about the network, the application, the user's identity, and their behavior when making decisions about finding and blocking anomalies. Many IPS implementations still use static allowlists but supervised machine learning algorithms are becoming more popular as a way to cut down on false positives. In the next version of IPS, you will be able to use together supervise and unsupervised machine learning method. It's also significant to note that some of these features are already done well.

1.3.6. ***Next generation antivirus (NGAV)-*** Using a mix of heuristic and behavior analysis, hash matching, and digital signature matching, malware on endpoints and servers can be found and stopped. Many vendors have also made ways to prevent malicious programs from running that don't use signatures. Supervised machine learning techniques are used to create signatures, heuristics, hashes, and other algorithms for scanning malicious software. Most screening processes still involve people so that relevant information is kept. Thanks to the work of manufacturers, AV systems that are ready for the future now have endpoint detection and response (EDR) capabilities. The main reason for this is skills in user behavior analysis (UBA), which are based on many different types of basic supervised ML algorithms.

1.3.7. ***The next-generation traffic scrubbing service-*** Using the Border Gateway Protocol (BGP), all of a protected entity's network traffic is sent to a centralized and

decentralized scrubbing facility somewhere in the world. So, it can defend against a wide range of distributed denial of service (DDoS) attacks, including volume-based ones (like UDP and ICMP floods), protocol-based ones (like SYN floods and pings of death), and application-layer ones (like HTTP GET/POST floods).

1.3.8. ***Security incident and event management (SIEM)-*** have grown to the point where it can store and link logs, events, and Syslog data from all security and fire alarms (network, perimeter, mobile, endpoint and cloud,). SIEM would be a great place to run supervise MLA and DL paradigms because it has many different training datasets.

1.3.9. ***Network access control (NAC)-*** Allows policy to be enforced using role-based constraints to prevent unauthorized access to (configured) wireless network resources and wired by endpoints, network devices, and computing assets until user and device-level authentication is complete. This is sometimes done to keep hackers from getting into parts of a network that aren't open to everyone. As a bonus, it can use 802.1X protocols, such as EAP-TLS and EAP-PEAP, to encrypt data transfers between wired and wireless networks.

1.3.10. ***Data loss prevention (DLP)-*** When the function is upgraded, it may protect against accidental copies or transfers of data in both a logical and a physical way. It is possible to offer these kinds of protections. Physically, it can't write to standard protocols like USB, SDHC, SDSC, SDXC, and UHS used by removable media. When standard protocols like SMTP/S, FTP/S, or HTTPS are used, sending data with regulated data components (such as PII, NPI, CTI, or ITAR) may be impossible. Machine learning guided by a human could work well in this situation.

1.3.11. ***SSL/TLS decryption-*** This ability is currently used for data loss prevention (DLP), forward/reverse proxy, and finding anomalies. A wide range of other network visibility technologies uses this ability to decrypt to find out more by looking for strange patterns in encrypted network data.

1.3.12. ***EDR (endpoint detection and response)-*** This capability may be used to find, investigate, and fix security problems and suspicious behavior on hosts and endpoints. We are excited about how this platform could make it easier to create supervised and semi-supervised machine learning algorithms.

1.3.13. ***Data protection and malware detection-*** Two approaches are widespread:

(a) data containerization for encryption, which is supported by companies like good quality (now Blackberry) and AirWatch, even though each of these companies has its way of displaying and protecting data;

(b) semi-supervised machine learning algorithms can be used by systems like those made by Zimperium and Lookout to scan for malware and find threats.[7-14]

1.4.A Cybersecurity Code of Ethics for Artificial Intelligence:

Artificial intelligence (AI), also called machine learning, is one of the most valuable tools a CISO can use. Artificial intelligence can't work without algorithms, which are basically sets of rules. Machine learning is how an AI system learns to improve by trying things out and making up its own algorithms. This artificial intelligence doesn't just respond to inputs humans taught it to recognize as triggers. It also responds to data and information that help it learn. It can now adapt to more situations because it is more flexible. To stop malicious activities, especially ones that a human security expert wouldn't be able to see with the naked eye, it's essential to have access to AI systems that can scan and analyze vast amounts of data at lightning speed. Still, even when people mean well, we've all seen science fiction stories where robots rebel against humans and their institutions. Even though CISOs haven't seen "Transformers" or been in the "South Park" situation where computers decide to kill humans because they are to blame for global warming, their arsenal of AI systems could be compared to a collection of double-edged swords. If nothing is done to stop it, algorithms made by artificial intelligence could wipe out whole industries. Without proper human oversight, AI algorithms that can't be controlled may destroy critical organizational processes. This could be fixed by adding a bit of common sense to a system driven only by logic. In particular, businesses often need to maintain legacy systems to do essential business tasks. But organizations need more resources to keep updating and maintaining these systems to make them more resistant to cyber-attacks. Because dangerous threats can change quickly and affect older systems, AI algorithms may mistakenly try to destroy them.

In the last ten years, the chances of being attacked online have increased considerably. There are many different kinds of cyberattacks and security flaws that can happen to any business that uses IT.

Threat actors who have right of entry to higher hacking tools, sophisticated malware attacks, persistent threats, insider threats from country governments, and other restricted and global criminal elements can disrupt a company's operations, profits, and services for customers. The Chief Information Security Officer (CISO) must have the foresight to predict future security problems and implement the incident response, necessary security controls and business permanence and disaster revival capabilities to get rid of, reduce, or transfer the risk that these threats pose. First, we need to talk about what "CISO's Next Frontier" means. Location, level of technical development, and time in history all play a role in where a person or group stands

about the boundary. Unrealized technologies like quantum computing, better encryption, and machine learning that don't need to be supervised may define the edge.

To put it another way, what makes it what it is are the people who go to the edge. The future of cyber technologies is also vital to the CISO's current responsibilities since it is impossible to know when technological advances will lead to cyber threats. Chief information security officers (CISOs) and their teams depend on continuously improving present state cyber innovations to protect their institutions as we navigate the dangerous new frontier of cyber threats, for which there are frequently no clear geographical or, at times, technological boundaries. This is due to the fact that cyber attacks may originate from any location and any sort of technology. Cyber threats can come from anywhere and can be made with any technology.

The success of "current" and "future" technology is critical to CISOs today, and rightly so (CISOs). This is because CISOs today have to use a wide range of technologies to deal with the growing number of threats that the internet poses. Ethically, CISOs and their cybersecurity teams must keep an eye on and manage how their AI algorithms work, choosing the best ones for the job. When designing AI systems to Cybersecurity teams must make difficult choices in order to safeguard the interests and well-being of the business and technology enterprises they serve. For example, they have to decide whether or not to use unassisted (or supervised) AI algorithms for cutting-edge cybersecurity defense and whether or not to use human-assisted (or run) algorithmic systems to stop a company's AI from destroying other non-malicious actor systems.

1.4.1. The Code:

If security teams want to follow ethical rules, they must build AI algorithms that can allow specific procedures and principles while protecting against the damage these behaviors could cause. Here is a basic set of rules that all AI programs should follow:

1.4.1.1. ***Provide for System Preservation-*** the primary goal of AI algorithms (or systems) is to keep a computer network or system from looking like it has been damaged.

1.4.1.2. ***Algorithms must be straightforward-*** to understand. Managers in charge of cybersecurity should be able to use both supervised and unsupervised AI algorithms (or systems) that make decisions in real-time. This includes the logic and reasoning used in processing, the shown intelligent behavior with the input from training data or other sources. Administrators must guarantee that the tools and systems utilized for cyber security monitoring and response include predictive, cognitive, prescriptive, and auto-reactive use cases. align with these explicit algorithms. Both synchronous and asynchronous situations need this.

1.4.1.3.***Make sure that algorithms can be held accountable***- Entities that use supervised or un - supervised artificial intelligence algorithms or networks for cyber security surveillance and reactive reaction must implement accountability and auditing procedures for the judgments made and actions (detective, protective, or otherwise) done by these algorithms. These standards should ensure that people are accountable for their choices and that algorithmic activity can be checked.

1.4.1.4.***Allow algorithmic identification to be used***- The artificial intelligence algorithms (or systems) must be able to identify themselves (in their current state) to the environment in which they live and work, and vice versa. These needs and skills must be taken into account.

1.4.1.5.***Make sure you can do evolutionary computing***- The AI system must be able to self-correct, self-optimize, and self-heal by removing unwanted algorithmic logic or implementations predictably. There should be checks and balances built into the system to ensure it never does anything wrong.

1.4.1.6.***Help Get Rid of Discrimination-*** Because the data sets used to train the algorithms are already skewed, it is essential to find, fix, and eliminate any biases that may appear. To stop tendencies from forming, it's necessary to use a wide range of up-to-date training datasets.

1.4.1.7.***Organize the Checking-*** It is very important to give ways to check the results or conclusions to make sure they are right. When an AI system needs to make a decision, it must be able to check on its own whether that decision is correct. The system must validate the output to meet moral, ethical, and legal standards.

1.4.1.8.***Make sure that everything is kept secret***- AI algorithms (or systems) can use any system data (including PII and NPI) to make decisions and calculations based on algorithms (including PII and NPI). This includes both structured and unstructured files, as well as files that have been encrypted.

1.4.1.9.***Due to the sensitive nature of the data it handles***- an AI system must have failsafe measures in place to protect user identities and their right to privacy.

1.4.1.10. ***Stick to the rules in the letter***- The artificial intelligence algorithms (or systems) must be made with the knowledge and ability to work within the corporate, municipal, state, or national regulatory requirements of the operational ecosystem(s). Some ecosystems may have different requirements.[15-17]

1.5.The Risks of AI and ML:

Even though AI and ML are increasingly used in cyber security, finding and avoiding their risks is essential.

1.5.1. ***Insufficient Level of Domain Awareness*** - As of now, machine learning algorithms have yet to learn about specific domains. This means they are working in an area of IT or security that they need to know more about. Cybersecurity is one of these fields. Because of this, the way decisions are made may need to be fixed, leading to false positives and negatives.

1.5.2. ***Deeply embedded Prejudice*** – Since they learn from training data, all supervised machine learning algorithms are prone to biases like sampling data, knowledge, and correlation biases. These biases can lead to mistakes in judgment, biases that are created by algorithms, and wrong assessments. Because of this, the data used to train an algorithm could change how it makes every decision.

1.5.3. ***Verification*** – Because it is hard to tell if the output or decisions made by machine learning algorithms, especially unsupervised ones, are correct, there is a chance that unreliable (or unfair) systems will be made.

1.5.4. ***Probability*** – AI and ML can determine how people will act in the future by looking at how they have behaved in the past (training data). Even though it is possible to make a very accurate prediction of what will happen in the future by looking at what has happened in the past and figuring out how the things that interact with each other work, there is no guarantee that the prediction will always be correct because there are always outliers in the data, rather than having to show a complicated mathematical proof, he just said, if (*f*) is a word or phrase that describes the connection between two things. (x, y), such that f(x) = E + y, where E is the mistake factor, then usually E→0 Even a tiny amount can have a significant effect if the information from the feedback loops is used to train people of E as *f* approaches *0,* it is still possible to prove mathematical oddities that might lead to erroneous results.

1.5.5. ***As a result of a statistically significant correlation*** – AI and ML use the statistical relationship between two or more variables to make predictions. But in other situations, like when complex systems or even people's actions are involved, a statistical correlation alone is not enough to prove cause and effect. This often leads to preemption, another word for wrong predictions that are hard to check.

1.6.The Attacks on AI/ML:

AI and machine learning can be broken like any other technology can. Because AI and ML systems are trusted, and their inner workings need to be better understood, these attacks may be hard to spot and could have terrible effects on the security of a system that relies on them.

1.6.1. ***Malicious Manipulation of Data*** - this is a method by which an attacker can trick an algorithm into coming to the wrong conclusions by changing its training data. An attacker may change geolocation data to hide the fact that malicious IP addresses come from a hostile country or a place where people hang out on the dark web. This is one way to get around a blocklist or other restrictions on access based on AI or ML.

1.6.2. ***Generative adversarial networks (GAN)-*** are usually made up of two AI systems that are rivals but work together to beat another AI by making data or functions that are typical of both systems. Two AI systems, a generator and a discriminator are at the heart of the approach. The plan's ultimate goal is to copy data or a function used by supervised machine learning methods (f). The generator makes a random sample of data, and then the discriminator tells the generator what it did wrong so it can get better at what it does. This process will be repeated until the generator creates content or a function that is the same as the source. Threat actors have been seen using GANs as adversarial entities to break passwords, avoid malware, and bogus facial detection to mess up supervise machine learning system used for biometric verification.

1.6.3. ***RPA Bot Manipulation***– supervised machine learning can trick second-generation RPA agents into making a bad call or going down the wrong branch of the decision tree. RPAs or bots with AI have been beaten or fooled before. If an RPA mediator (bot) used for high-frequency stock trade or to let people reset their passwords is broken into, it could lead to bad things.

1.7.AI/ML with Supervision:

Simply put, it is an algorithm that can see or evaluate pairs of input and output data to find or predict data that hasn't been seen or is missing (also called labeled data). The same set of tools can be used for behavior and anomaly detection. In the training set, which is made up of N samples, inputs and outputs are kept separate- (x1, y1), (x2, y2), . . .(xN, yN), where an unknown function produced each y. y = f(x), Using supervised AI, we can find a function h almost the same as f but not quite. x and y don't have to be whole numbers for their values to make sense here. Most of the time, machine learning is used to solve problems in IT and computer security, such as finding spam and viruses, using OCR, recognizing faces, and recognizing voices. It could "actively gather" the missing (label) data from various sources, allowing it to add tags to existing data sets that might be missing them. Even though it is still

in its early stages, Active Learning in cyber defense looks promising. One way to get this information is to ask for it directly.

Strategies for classification, regression, and ranking are all examples of supervised machine learning models that can be used. Most supervised machine-learning techniques used in cybersecurity fall into these two categories.

1.7.1. Characterization of AI/ML with Supervision:

These supervised machine learning algorithms use the ideas of statistics classification and sub-categorization to sort the information in a dataset. These models use standard identifiers like credit cards or social security numbers to put data into groups. Both binary classifications, where there are only two possible values, and multiclass, where there are several, are the two most essential classification systems. The decision tree model is also part of this group. Decision trees can forecast future values or distinct labels by categorizing current (response) variables. This is because they are built to classify data.

1.7.2. Regression:

This group of supervised ML models can look at and estimate the properties of a connection role (f) among input (x) and output (y) variables –

$$f(x) = y$$

If you know x and y, you can figure out f for a given data set. If you know the function f, you can figure out the output based on what you put in.

1.7.3. Ranking:

Putting two things in a collection in order of importance based on how they relate to each other is called "ranking." Among the most effective security datasets for supervised ML algorithms include actual network taps, logs, software configuration data, and purpose-built structured data. Some examples of structured data specific to a function are listed below.

1.8.AI/ML with Unsupervised:

An unsupervised machine learning (ML) algorithm can find or predict data that has yet to be seen. It does this by analyzing or evaluating unlabeled data (which may or may not include input and output data pairs) to build a function that may help map the input to the output. This type of machine learning is also called "unsupervised learning" (UL). This claim is easier to make than to prove since it is hard to objectively measure how well the (unsupervised) algorithm works without labeled data. In several ways, unsupervised machine learning is used in cyber security, such as finding anomalies, putting things into groups that help find threats, exploring and classifying data, etc. All of these programs can help you find problems and the solutions to those problems. The clustering and dimensionality approaches could help show unsupervised Learning better. These are not on topic and won't be talked about anymore.[18-20]

1.9.AI/ML with Semi-Supervised:

In quasi ML, an algorithm employs both labeled and unlabeled data to produce a function that translates input to output. So, the software can find or make predictions based on information that has been hidden. This method is often used when there are more unlabeled data points than labeled data points. Most of the time, network intrusion detection is the most common way that semi-supervised machine learning is used in IT security today. The best use of these machine learning strategies is when they take in Cybersecurity datasets comprised of real-time network taps, logs, software configuration data, and custom-built data structures.

1.10. AI/ML with Reinforcement:

During the testing phase of reinforcement machine learning, the results from the training phase are sent back into the system to help create a program that may assist in mapping input to output. Allowing the system to recognize or foresee data it hasn't seen yet. A process that will enable map inputs to outcomes is used to do this. To master this technique, you must first figure out when to move from the tuition phase to the testing stage. This is a decision that is heavily influenced by how your practice runs go. Reinforcement machine learning algorithms are the most effective way to learn from security datasets. Information like this could come from real-time network taps, configuration management data, logs and custom-built, well-organized datasets.

1.11. AI/ML with Bayes' Law:

Machine learning algorithms often use Bayes' Law to determine how likely an event is given the variables, circumstances, and other probabilities are already known to be linked to it. ML algorithms often use Bayes' Law. This is essential to Bayesian models, Bayesian probability, and Bayesian inference.

1.12. AI Applications in the Field of Cybersecurity:

Artificial intelligence has been a buzzword in cyber security for the past ten years. It was primarily used to describe schemes that mainly used robotic process automation (RPA) or similar technique to mechanize or add to event answer and threat uncovering use cases done manually. In cyber security, supervised machine learning has been used in some exciting ways in the last three to four years, such as for user and entity behavior analysis (UEBA), network security, threat detection. But there is still a lot of space in cyber security for semi-supervised and unsupervised ML paradigms.

1.12.1. Standard Conditions:

Before AI can be helpful in cybersecurity use cases, it needs to meet the following conditions:

a) You need to be able to do the next in REAL TIME (or as close up to real-time as is technically probable) with live network traffic that can be accessed through SPAN or TAP port or straight more than the wire when possible:

i. Find things that don't fit with what you know or what you've learned, like suspicious activity or strange behavior from people, systems, or devices;

ii. Find APTs and zero-day attacks by using supervised or unsupervised machine learning techniques as well as indicators of compromise (IOCs) that have already been found;

iii. Compare real-time record data by means of current safety measures logs in the safety measures incident and event management, system to find attack patterns and unusual behavior;

b) Make sure that cross-functional application, security logs and system can be correlated after an event (asynchronously) to look for unusual or malicious behavior.

i. Find out if there have been any cases of money laundering, fraud, unauthorized withdrawals, account takeovers, or other suspicious things.

ii. Tasks need to be run every day, week, month, or three months, so algorithms need to be changed (in batches).

iii. Massive data sets from application and security logs must be easy to analyze, train, and label.

iv. Data from production applications must be able to be analyzed across functions.

c) Taking a right unsupervised alerting or blocking actions as described below in response to identified threat patterns, anomalies, and user behavior, as it relates to the optimized cyber kill-chain [reconnaissance, breach (weaponization + delivery), infection (installation, privilege escalation, and code execution), and actions (data exfiltration)].

i. Improve and add to the machine learning methods that are already in place.

ii. Find and get rid of weaknesses, wrong settings, and too many access permissions;

iii. Being able to sort and fix data as it is being stored;

vi. AI should be able to figure out steganography and other ways of hiding information.

1.12.2. AI/ML active Models:

AI paradigms can be used in cyber security in many ways because there are many models and algorithms to choose from. In order of preference, the following are the most beneficial artificial intelligence and machine learning algorithms for use online:

1.12.2.1. Decision Tree:
This method uses branches to make graphical representations of the data it receives. This lets it predict a visual conclusion or flow for any options. Nodes, edges, branches, and leaf nodes are the main parts of a decision tree's graphical representation.

There are two distinct subcategories of decision trees:

- ***A classification tree-*** Most of the time, simplified decision trees give a yes/no or binary answer as their primary output.

- ***A regression tree-*** is a structure for making decisions that can lead to continuous values. Decision trees are used in cyber security as part of machine learning models. They help track the steps of cyberattacks. In the past, decision trees were used to find malware; some of these methods are still used today. Deep learning neural networks are becoming increasingly popular as the best way to implement many of these uses.

1.12.2.2. Naïve Bayes:
This method can create a predictive classifier that can estimate how likely an event will happen based on a set of highly independent variables, situations, or other probabilities. The idea behind this method is Bayes' Law. This strategy aims to make predictive classifiers that can determine how likely it is that something will happen based on a set of variables, conditions, and other probabilities.

$$\frac{P(A|B)}{} = \frac{P(B|A)P(A)}{P(B)}$$

Where:

• Events A and B are very different from each other;

• The probability is P (), but the conditional probability is P (|). This means that P(AB) is the chance that A will happen if B is accurate, and P(C) is the chance that C will happen if A doesn't. The first, P (), is an unconditional probability, while the second, P (|), is a conditional probability.

1.12.2.3. The K-Nearest Neighbors:
At the heart of this method for classifying data are multiclass classification and regression ideas.

1.12.2.4. The K-NN Classification:
In this collection of data, each person is put into a category. When the algorithm is done with the item, it will be placed in the class with the most similar items (based on a threshold of k).

In the simplest case (k=1), a thing only belongs to the category that its next-door neighbor belongs to. The number k is often a minimal, odd number. Pattern recognition, data mining, and detecting intrusions are where this method may be helpful in cybersecurity, and it can be improved with the help of this method.

1.12.2.5. The K-NN Regression:

In regression, as in classification, the input is the object's k closest neighbors, and the output is the average value of these neighbors. This non-parametric method could help make statistical estimates and find patterns in cybersecurity.

1.12.2.6. The Deep Learning:

Deep Learning is a machine learning technique that integrates the results of one or more (lower) level (usually supervised) ML algorithms with the outcomes of one or more (higher) level (typically unsupervised) ML algorithms as inputs or in combination. This enables a better degree of accuracy and a lower rate of false positives while performing cybersecurity tasks such as pattern recognition (including the detection of malware). DL can be used to do cybersecurity tasks like figuring out patterns. Deep Learning is becoming more critical in cyber security, especially for finding security holes, lousy software, and phishing and spam operations. It is used by image recognition programs to spot when a website has been changed. Speech recognition and NLP are two more ways this technology can be used (NLP).

1.12.2.7. The Restricted Boltzmann Machine (RBM):

Since RBM is a two-layer random-deterministic generative artificial neural network, it can learn the probability distribution (how often outcomes happen) across all input sets. Researchers from Cambridge University in the United Kingdom came up with RBM. There are two tiers, each with both open and closed nodes, and a rule that says a node in one level can only connect to one or more nodes in the other tier. Deep Belief Networks (DBNs) are becoming more popular, and RBMs are used to build them to learn without being watched.

1.12.2.8. The Neural Networks:

In artificial intelligence, neural networks based on the human brain have been used to improve unsupervised machine learning. The (artificial) neural network is made up of connected nodes. It is based on the way neurons in the human brain are connected. It does this by using classification techniques, predictive linkage, and error correction in order to improve its outcomes. Because neural networks are better at learning complicated temporal sequences, they are good at finding outliers in data sets made from streaming data that is close to real-time. This is because neural network learning has the potential to improve machine learning paradigms that don't have a teacher. During training, algorithms in this group are shown a lot of clean, non-threatening data. This helps them learn how systems usually work. Because it has been trained, it can tell the difference between standard and strange behavior. Putting this information into UEBA (user entity behavior analysis) scenarios would be beneficial.

1.12.2.8.1. Convolutional Neural Networks (CNN):
These methods are called feed-forward neural networks because the input can only move from one hidden layer to the next before it reaches the final output. Between the information and the work, there is often more than one level of abstraction. This method was used here because it helps find problems without the help of a person.

1.12.2.8.2. Recurrent Neural Networks (RNN):
These methods, similar to neural network models, can be used to look for patterns in large sets of random numbers or events. The following study uses the results of previous analyses on the same data sets.

1.12.2.8.3. Pseudo-Randomness Detection (PRD):
This method is part of the family of neural networks. It can be used to get better at putting random numbers and events together, and it can also be used in these areas. If they were used in SIEMs, business processes would generally be better.

1.12.2.8.4. Extreme Value Theory (EVT):
This method can determine how an object acts or build a pattern from a small amount (sparse) or extensive data. This feature is called "extreme data availability." This avoids the time-consuming data optimization process, making it possible to find patterns more quickly and accurately. This is important for cyber security uses that combine supervised and unsupervised machine learning.[21-22]

1.13. AI/ML Standards and Advice for Implementation:
This document highlights the cyber security use cases and specifications that have been mapped to an altered version of Lockheed Martin's cyber death chain. This is valuable knowledge for unsupervised and supervised machine learning algorithm implementations that leverage the aforementioned core approaches.

Even though the Lockheed Martin kill chain was chosen for this book, more improvements could be made by switching to the more well-known and complicated "MITRE ATT&CK" design and using the existing mapping between the two.

But this book can only cover some needs by giving solutions and implementation-level details. Those things might be added to a future version. These kinds of implementations are different from what this book is about. This is meant to be both a general guideline and a big-picture plan of action:

1.13.1. The Supervised Machine Learning — Guidance for Implementation:
Both old and new algorithms should be trained with real-world examples of each step of the cyber kill chain of bad behavior. Utilize training data from the application, systems, and

security logs for particular cyber kill chain use situations. This is called "supervised machine learning" (from the SIEM).

1.13.2. The Unsupervised Machine Learning - Guidance for Implementation:

Using supervised machine-learning techniques that have already been taught as inputs for more advanced unsupervised machine-learning algorithms makes it possible to build unsupervised machine-learning algorithms for specific cyber kill-chain mapping application situations. For example, you can use the Expectation Maximization (EM) method to find the parameters of a hidden Markov model (HMM). This led to the development of a paradigm called deep learning.

Here's one possible way to put the plan into action:

a) All methods for supervised machine learning should be put into subcategories within the current kill-chain mapping.

b) Check to see if the machine learning (ML) structures and techniques can meet the mathematical needs of the different subclasses. For example, a method for finding pseudo-randomness can be used to find IP and Port scanning, both types of survey.

c) Rigidly train each supervised algorithm with the publicly available labeled information for the necessary criteria and identified application circumstances that have already been written down.

d) If needed, make supervised variations of the strategy to meet the needs of each subclass's many use cases.

e) Put the results of every supervised algorithm trained for a subclass into an unsupervised AI/ML and DL algorithm that can find outliers in situations where the underlying supervised method was not trained. Since the supervised technique wasn't introduced in these subclasses, the unsupervised approach can find oddities. This is an excellent example of AI at its best.

f) The "Cyber kill-chain"-related detective techniques listed below need to be used in many areas of such as endpoint security, cyber security, network security, cloud security, application security, perimeter security, data security, monitoring, and incident response.

It's important to remember that the different tools that makeup today's corporate security stack use both commercial and open-source versions of algorithms. There are also a lot of different algorithms that can be used.

In the following, we'll talk about the requirements for artificial intelligence and machine learning for each link in the kill chain.

1.13.3. AI/ML Intelligence gathering:

During this stage, the threat actor monitors the target host or system. These activities, which could be compared to scans or scouts, are meant to find vulnerabilities, open ports or services, or other flaws that could be exploited remotely. Simply put, the threat actor is trying to figure out how to take advantage of the victim from a distance. For this kind of scanning, people often use Nmap and Shodan. Scanning is just one option for this stage. Other actions, like getting information from social networking sites like LinkedIn, Twitter, and Facebook, or publicly available sources, could also be added. Once the threat actor has enough information, they might try to steal login information from the target organization by sending phishing emails. At this point, the main job of machine learning algorithms is to find and report any reconnaissance actions (often external or internal) that threat actors take to find a weak system or service component. Firewalls, web application firewalls, and behavior analysis engines for endpoint security would all be much better if they worked well with EDR solutions.

1.14. Detection of various applications of AI and ML:

1.14.1. IPSweep:

A host can "IP Scan" many remote hosts at once with the help of IPSweep. In this method, the IP address that sends the request quickly sends several ICMP echo requests to a wide range of IP addresses (or hosts).

One type of vulnerability assessment that is done from an IP address outside of an organization's private network is an external IP scan. Similar scans can also be done from a remote network's internal IP address, called an "internal IP scan." Below, we'll discuss the details of the three different kinds of IPSweeps. All of these should be the goal of AI and ML implementations.

1.14.1.1.*The External IP Scans*- are done from far away, often using a virtual private network (VPN) to hide a temporary IP address and tools like Nmap and Shodan. Even though skilled attackers could hide behind a virtual private network (VPN) to do these scans, most IPs are immediately flagged as bad.

1.14.1.2.*The Internal IP Scans*- things that happen within the boundaries of a network. An attacker or insider threat may be able to hide their information exchange in plain sight by mixing it in with the regular east-west traffic of the network while trying to pretend to be a regular user or piece of code.

1.14.1.3.*Random (External) IP Scans-* are also done by an attacker looking in from the outside. Still, they tend to happen over a long period to trick Cybersecurity algorithms that search for certain scanning pattern in less time.

1.14.2. PortSweep:

PortSweep lets a host "scan" the ports of another host by "sweeping" over them. The source IP address will quickly send several TCP SYN messages to many target IP ports (also known as a host). There are two main kinds of PortSweeps, which we'll discuss in more detail below. All of these should be the goal of AI and ML implementations.

1.14.2.1.*The External Port Scans-* use tools like Nmap to learn about a target network while the attacker is elsewhere. They often carry out attacks remotely using a Virtual Private Network (VPN) to stay anonymous. Even though skilled attackers could hide behind a virtual private network (VPN) to do these scans, most IPs are immediately flagged as bad.

1.14.2.2.*The Internal Port Scans* - The user is still connected to the network during these scans. By acting like a real user or application, an attacker or insider threat may be able to hide this communication among the regular east-west traffic of the network.[23-24]

1.15. Additional Scans (incoming and outgoing) of Miscellaneous Items:

Not only can IP and Port scans be used for malicious recon, but there are many other scanning methods to watch out for. AI and ML systems should be able to pick up on some of them, as we'll see in the next section.

(a) Scanning with programs like Nmap, such as (Masscan, Solarwinds, Zip, etc.)

(b) SSH search of the whole target network's IPv4/v6 addresses.

(c) Unauthorized use of the famous H.323 protocol for video conferences, which companies like Cisco, Avaya, and Polycom use.

(d) It uses flaws in the Googlebot web crawler or other tools like it to commit fraud.

(e) Leakage of temporary file or folder names that use the tilde character (" ") in Microsoft Internet Information Services vulnerabilities or features.

(f) Bad things people do on a network Since the UPnP protocol doesn't require authentication, evil people could use it to spread malware.

(g) I am using a more powerful version of the Simple Service Discovery Protocol to scan (SSDP).

(h) Firewall scans like Firewalk can use a traceroute to look at the replies to IP packets. This information can then be used to find gateway ACL filters and map the network.

1.16. The CISO Position:

Chief Information Security Officers are in charge of a company's security architecture and make sure that the company's data, systems, and infrastructure are safe. It is the CISO's job to make sure that the AI code of ethics described below is followed as the company grows its use of AI in cyber security to get the benefits of automation and speed, to avoid any intentional or accidental abuse or bias, and to protect the safety or IT systems and users. The CISO must also ensure that the rules for ethically developing AI are followed. When companies that make security products add AI to their products and security suites, CISOs must work with their CTOs to ensure the code of ethics. This is possible if you work with the CTOs of the companies that make security products. The goal of this code of ethics is to give people a way to use AI in cyber security in a responsible manner. This is needed to protect the country's most crucial infrastructure, as well as the safety of the general public and the commonwealth, from the dangers that malicious AI poses.

No one can argue against the fact that machine learning and other types of AI are becoming more and more useful for a wide range of cybersecurity applications. It lets us find and stop cutting-edge attacks from well-trained attackers, some of whom may use tools with AI. This allows us to find and avoid threats from well-organized enemies. Attackers today need less time than ever to launch an attack and take advantage of a known or unknown vulnerability (zero-day), so the response must be automated and lightning-fast. Cyber threats can only be found and stopped with the help of AI-powered tools. With AI and ML-enhanced security solutions. Improvements to asynchronous algorithms made with AI and ML are beneficial in finding fraud inside financial service providers. The CISO's job is to convince manufacturers that their products must meet the needs of the next generation of cyber security solutions and identity and access management systems, which will be strengthened by artificial intelligence.[24-25]

Conclusion:

AI and machine learning are being used in cyber security to reap the benefits of automation and speed, prevent intentional or unintentional abuse or bias, and protect security or information technology systems and their users. In addition, the CISO is responsible for ensuring that the ethical criteria for developing AI are followed. When firms that develop security solutions use AI in their products and security suites, their CISOs must engage with their CTOs to guarantee compliance with the code of ethics. Collaboration with the Chief

Technology Officers of companies that create security solutions will make this possible. This code of ethics aims to give people the tools to employ AI for cyber security ethically. The nation's most essential infrastructure, as well as the safety of the general population and the commonwealth, must be protected against the threats presented by malicious artificial intelligence. As they transition from supervised artificial intelligence and machine learning to unsupervised AI and ML as the foundation of their security toolkits, chief information security officers face several dangers and challenges. Utilizing AI in this fashion poses various risks and problems, including the inability to validate the findings reached by an AI response system, the development of biases, and the production of false positives. AI and ML will be included in the tools of the future.

Reference:

1. Blackman, Reid. "A practical guide to building ethical AI." *Harvard Bus. Rev* 15 (2020).
2. Badhwar, Raj. "AI code of ethics for cybersecurity." *The CISO's next Frontier*. Springer, Cham, 2021. 41-44.
3. Ismail, K. "AI vs. Algorithms: What's the Difference." *CMS WiRE* (2018).
4. Carreira, Tulio. "Ethical Issues on AI-powered Social Media Apps."
5. Vousinas, Georgios L., et al. "Mapping the Road of the Ethical Dilemmas Behind Artificial Intelligence." *Journal of Politics and Ethics in New Technologies and AI* 1.1 (2022): e31238-e31238.
6. Russell, Stuart J. *Artificial intelligence a modern approach*. Pearson Education, Inc., 2010.
7. Schuld, Maria, Ilya Sinayskiy, and Francesco Petruccione. "An introduction to quantum machine learning." *Contemporary Physics* 56.2 (2015): 172-185.
8. Martin, Lockheed. "GAINING THE ADVANTAGE, Applying Cyber Kill Chain Methodology to Network Defense." *Lockheed Martin Corporation* (2015).
9. Koushik, C. S. N., et al. "Determination of Age, Gender, Dress Color and Type of a Person by Convolutional Neural Network (CNN)." *Proceedings of the International Conference on Innovative Computing & Communications (ICICC)*. 2020.
10. Badhwar, Raj. "The Case for AI/ML in Cybersecurity." *The CISO's Next Frontier*. Springer, Cham, 2021. 45-73.
11. Deng, Li, and Dong Yu. "Deep learning: methods and applications." *Foundations and trends® in signal processing* 7.3–4 (2014): 197-387.
12. Delplace, Antoine, Sheryl Hermoso, and Kristofer Anandita. "Cyber Attack Detection thanks to Machine Learning Algorithms." *arXiv preprint arXiv:2001.06309* (2020).
13. Deng, Li. "Dong Yu και others." *Deep learning: methods and applications. Foundations and Trends R in Signal Processing* 7.3-4 (2014): 197-387.
14. Badhwar, Raj. "The Case for AI/ML in Cybersecurity." *The CISO's Next Frontier*. Springer, Cham, 2021. 45-73.

15. Klimek, Thomas. "Generative adversarial networks: What are they and why we should be afraid." (2018).
16. Mahalakshmi, A., N. Swapna Goud, and G. Vishnu Murthy. "A survey on phishing and it's detection techniques based on support vector method (Svm) and software defined networking (sdn)." *International Journal of Engineering and Advanced Technology* 8.2 (2018): 498-503.
17. Viet, Hung Nguyen, et al. "Using deep learning model for network scanning detection." *Proceedings of the 4th International Conference on Frontiers of Educational Technologies*. 2018.
18. Koutroumbas, Konstantinos, and Sergios Theodoridis. *Pattern recognition*. Academic Press, 2008.
19. Stepanic, Pavle, Ilija V. Latinovic, and Zeljko Djurovic. "A new approach to detection of defects in rolling element bearings based on statistical pattern recognition." *The International Journal of Advanced Manufacturing Technology* 45.1 (2009): 91-100.
20. Bouboulis, Pantelis, and Sergios Theodoridis. "The complex Gaussian kernel LMS algorithm." *International Conference on Artificial Neural Networks*. Springer, Berlin, Heidelberg, 2010.
21. Kläs, Michael, and Lisa Jöckel. "A framework for building uncertainty wrappers for AI/ML-based data-driven components." *International Conference on Computer Safety, Reliability, and Security*. Springer, Cham, 2020.
22. Porambage, Pawani, et al. "Sec-EdgeAI: AI for edge security Vs security for edge AI." *The 1st 6G Wireless Summit,(Levi, Finland)* (2019).
23. Ouaissa, Mariyam, et al., eds. *Artificial Intelligence of Things in Smart Environments: Applications in Transportation and Logistics*. Walter de Gruyter GmbH & Co KG, 2022.
24. Traore, Issa, Isaac Woungang, and Sherif Saad, eds. *Artificial Intelligence for Cyber-Physical Systems Hardening*. Vol. 2. Springer Nature, 2022.
25. Alexandrou, Alex. *Cybercrime and Information Technology: The Computer Network Infrastructure and Computer Security, Cybersecurity Laws, Internet of Things (IoT), and Mobile Devices*. CRC Press, 2021.