Artificial Intelligence in Cybersecurity and Network security

By

Christoffer Sjöblom

Master's thesis
Computer Engineering
Faculty of Science and Engineering (FNT)
Åbo Akademi University

Supervisors: Annamari Soini & Sebastien Lafond
Advisors at If Insurance Company: Niclas Gers & Tommi Nurmi

Spring 2021

ABSTRACT

Cyber threats and network attacks are becoming more sophisticated each day. The number of daily attacks globally is rising, and the ways and techniques used to infiltrate company systems and personal devices are diverse. Behind the attacks are lone attackers, organized groups, and entire governments. The recourses for carrying out attacks are increasing, and today cyber attacks can have severe, widespread effects and consequences. These factors combine to create problems for security teams to keep up with the pace, and smarter solutions are required.

This thesis gives an overview of how artificial intelligence (AI) approaches, and sub-domains such as machine learning and deep learning, can be applied to cybersecurity issues. This thesis also aims to showcase existing AI technologies and how they improve cyber security. With AI security systems can become more predictive and proactive in detection and response, as well as save time in manual repetitive security tasks. In this thesis, employees at If insurance company who work with network monitoring and network security were interviewed. The purpose of the interviews was to find out what sort of problems they have encountered in their tasks. Additionally, possible AI-based solutions for the defined problems are represented in this thesis.

Table of Contents

List of Figures

List of Tables

ACKNOWLEDGMENTS

I would like to express my gratitude to my supervisors Annamari Soini and Sebastien Lafond for helping me in the making of this thesis. The effort they put in to review my thesis is much appreciated. I would also like to thank my advisors Niclas Gers and Tommi Nurmi at If insurance company for giving me the opportunity to conduct this thesis in collaboration with them. Finally, a special thanks to my partner and family for supporting me through the writing process and for making sure I finalize my studies.

1. Introduction

1.1 Overview

In today's work environments and everyday life computers are essential and unreplaceable. As our technologies evolve and increase, the demand for security of information grows. The collection of data is vast, and data is collected by corporations, military, financial, medical, and governmental applications all the time. This highlights the importance of cybersecurity. The term cybersecurity is a prominent one and cybersecurity is an issue we have to encounter and develop continuously. Cybersecurity, which also goes by the name computer security and information technology security, is the action of securing data and information on networks, mobile devices, computers, and other electronical devices or connections. The goals are to ensure the integrity, availability, and confidentiality of the data and to protect it from malicious attacks, unauthorized access, or other damage.

Cybersecurity today is mostly implemented with rule-based systems. A rule-based system is the way we give a system rules in order for it to know how to manipulate, store, and sort data. The rules in the system are implemented by the creator or the vendor and these rules can usually be changed via updates. What happens when an attack occurs is that the system checks for the ways it should react in the list of rules. If there are no implemented actions for a certain intrusion, the system has to be shut down and put on hold. After this the creators have to identify the problem and, either by patching or software updates, correct the system manually. The rate of new attacks and various forms of the same attack is huge and one major problem with rule-based systems is that they are not very suitable for change and adaptation. The exercise of manually fixing these ever-occurring attacks is time-consuming and requires many man hours, thereby decreasing efficiency and productivity.

In order to solve these issues, we would have to create a system that could adapt to the environment and have the ability to learn from its previous encounters and change the rules accordingly for future malicious attacks. The system would, in other words, have the capacity to patch itself and find its own ways of repairing the lacking security aspects. In addition to this, the system would be able to log its attack history and recreate itself as a more robust system from the new rules it has come up with by itself.

Utilizing *artificial intelligence (AI)* in cybersecurity is a way of tackling the problems mentioned above. Cybersecurity tools based on artificial intelligence are used more and more and have been developed rapidly during the last decade. The upscaling of these methods will help reduce the number and risk of breaches, as well as increase the efficiency of cybersecurity-related tasks.

1.2 Goal of this Thesis

This thesis aims to illustrate how artificial intelligence-based systems and algorithms can advance cybersecurity and network security. The field AI is vast and not every aspect is covered in this research. The number and types of attacks are also very diverse, so every type of attack and how to encounter it is not included. The main goal is to review practical existing systems and research how they can be applied to network security. By introducing more AI methods in computer security, intrusion and detection systems become better at detecting and altering deviating behavior and possible attacks. Lastly, this thesis aims to show what factors are stalling the further development of utilization of AI tools in cybersecurity and what the future of AI in this field looks like.

1.3 Methods

The research methods used in this thesis are both literature reviews and interviews. In order to gain knowledge about the field of AI and cybersecurity, a general literature review was done. Firstly, the literature review strives to give an overview of artificial intelligence and various types of cybersecurity. Secondly, the background contains theory about machine learning and deep learning, as well as how these methods are used in the detection of cyberattacks and intrusions.

The second method used in this thesis was freeform interviews with employees at If insurance company. The point of the interviews was to find problem areas in network monitoring and network security, and how AI could be utilized to solve these issues.

1.4 Thesis structure

This thesis starts by defining the terms artificial intelligence and cybersecurity and by presenting an overview of the threat landscape and how AI can improve cybersecurity. Chapter 3 describes how machine learning and deep learning approaches in cybersecurity work and how they can be used in attack and intrusion detection. Chapter 4 describes the methodology in greater detail. Chapter 5 presents the result from the interviews. Chapter 6 presents AI-based solutions for problems highlighted in the interviews. Chapter 7 offers a discussion of the results as well as what the future of AI in this field entails. Chapter 8 concludes the thesis.

2. Artificial Intelligence and Cybersecurity

2.1 Definition of Artificial Intelligence

The term artificial intelligence has of late become a veritable buzzword in almost every branch in our society. The wide usage of the word has thereby caused some ambiguity when it comes to the definition of the term AI. The list of definitions is quite wide, and the following definitions should give a pretty comprehensive understanding of the concept for the coming thesis.

- "*Artificial Intelligence is the study of computations that make it possible perceive, reason, and act*" [1].
- Artificial Intelligence is the science and engineering of creating computer systems and machines that mimic the concept of intelligence [2].
- Artificial Intelligence is the concept of making computers do things that would be considered intelligent, if done by humans [3].
- Artificial Intelligence as a field of study is to develop systems and computers to have the ability of human-like decision making. These abilities consist of self-correction, learning and adaptation to the environment. It is also a science of expanding the human capabilities in a sense, where powerful computational capabilities are utilized [4].
- Artificial Intelligence strives to find the connection of computations and cognition [5].

Artificial Intelligence as a term has been around for many decades and in the early days the simple definition was the capacity of machines mimicking human behavior and decision making. Today we can use AI-based tools to diagnose diseases, solve complex mathematical equations using only symbols, understand human speech and language, develop organic chemical compounds, and analyze electrical grids and circuits [6]. As we can see, the world of AI is almost limitless, being a growing body of engineering and computer power that will have a great impact on technical solutions and civilization itself.

2.1.1 History of Artificial Intelligence

The actual birth and origins of artificial intelligence can be diffucult to pinpoint but the first examples of the core principles of AI can be traced back to the 1940's. During this decade the writer Isaac Asimov wrote fiction about robots that could emulate human behavior and decision making [7]. Another, more practical example during the same decade, took place in England, where the English mathematician Alan Turing created *The Bombe.* This machine, also regarded as the first computer, was able to crack and decipher the German Enigma code during the Second World War. Turing later published an article describing his methods of testing computers and their intelligence that still stands as a springboard for AI as we know it today [7].

Between 1964 and 1966 the first computer program able to process natural language was created. This program called *Eliza* could emulate human language and simulate a dialog with a real human being. Primitive problem-solving computer programs were also invented during the same era and these could automatically solve some games such as Towers of Hanoi [7]. The successes in the field prompted the funding of artificial intelligence and research in it during the 1970's. However, the ever-increasing allocations of assets into AI met some pushback. Those against further development of AI reasoned that machines could never achieve the complexity of human behavior [7].

The greatest problem stalling AI progress in the beginning was the way the human behavior aspect was implemented. The initial approach was to create a stack-like hierarchy for decision making in the form of multiple if-then statements [7]. These rule-based systems, for example expert systems, do work well in a limited environment such as chess and other games, and in the 1990's a machine was able to defeat the World Champion of chess with ease. The next step for AI was to have the capacity to process external data, learn from it, and adapt to the changing environment that it works in. The previous sentence provides roughly the definition of artificial intelligence we use today, therefore the previous examples dating back to the 60's and 70's do not exactly represent AI as we know it today.

Since the beginning of the new millennium artificial intelligence has seen some great advances, and the inventions of neural networks and deep learning have further accelerated the evolution of AI programs. These two concepts will be explained later in this thesis as they are essential in the AI solutions presented in this study. In 2015, one cornerstone of AI technology was created when Google developed AlphaGo, which is a computer program designed to play the game Go. AlphaGo utilized the computational power of machine learning and neural networks to defeat the best Go players in the world and is still considered as one of the most sophisticated AI algorithms [7]. Inventions such as speech and face recognition, advertisement algorithms, and smart speakers are all very recent advances in the realm of AI, and the full potential of AI tools is yet to be explored.

2.2 Definition of Cybersecurity

In the following section, we will define the word *cybersecurity* and explain the various traits that characterize it. According to Cybersecurity & Infrastructure Security Agency (CISA) [8], "*Cybersecurity is the art of protecting networks, devices, and data from unauthorized access or criminal use and the practice of ensuring confidentiality, integrity, and availability of information*". In recent years, cybersecurity has become more prioritized in business since users and companies are more aware of the threats concerning their private data. In addition, with specific regulation efforts such as the General Data Protection Regulation (GDPR) in 2016 and the Data Protection Act (DPA) in 2018, there are now clear instructions and consequences to motivate organizations to set up a proper cybersecurity infrastructure and management.

Cybersecurity itself encompasses several types of security. This includes network, endpoint, application, data, cloud, and mobile security. Therefore, sufficient cybersecurity solutions have to take the whole system and infrastructure into account and not just focus on the most crucial components in the system [9]. As mentioned at the beginning of this section, the three main components of cybersecurity and information security in general are confidentiality, integrity, and availability. This set of attributes is commonly called the CIA Triad.

Confidentiality is the means of protecting data and information from unauthorized access. In other words, the data we want to protect must be ensured in some way, most commonly by encryption. This end goal aims to ascertain that only the right people have access to the information sent between computers and devices [10].

Integrity in the CIA Triad refers to the idea of preserving the data and keeping it accurate and consistent. In the case of changes to the data, this has to be done in an authorized way. Changes to data can occur due to careless usage of it as well as by faults in the security system itself. When integrity is implemented well, the data is kept unchanged during both storage and transmission phases [10].

Availability is the last component of the triad and sometimes overlooked. Availability is the practice of enabling the data to be accessed at the right place and the right time. The lack of availability is often caused by errors in the system. The ability to access information is crucial in some instances and hick-ups in this section can be very problematic and have serious consequences [10].

2.2.1 Types of Cybersecurity

As mentioned in the previous chapter, there are various of types of cybersecurity and each of them has its own field of problems to tackle and solve. In this section, we present a couple of them that are most relevant for this thesis.

Network security is, in simple terms, the practice of protecting a network from being accessed by unauthorized actors. The main point is to deflect malicious activity in the network and the connections between the devices [9]. The most commonly used network protocol today for transmitting traffic flow is Internet Protocol (IP). The network consists, of course, of many layers and all of the layers are vulnerable to attacks. The two most frequently occurring types of malicious activity on a network are network intrusion and distributed denial of service (DDoS) attacks [11]. Network intrusion can be exploited by introducing unwanted packages to the network whose purpose is to consume resources of the network, interfere with the functions of the resources, and gain information and knowledge about the system that can be used later for more severe attacks [11].

The main objective of application security is to secure its functionalities and to ensure they are implemented in a way that protects them from attacks. Software updates and testing are ways to enhance the security of the application [9]. In recent studies, the area in cybersecurity with most breaches has particularly been web applications. Testing, as mentioned, is the key ingredient in preserving application security and this should be repeated throughout the lifespan of the application. The three phases of an application are generally development, quality assurance, and production [12]. While most organizations perform diligent testing and checking for vulnerabilities during the first and second stage, the actual problems occur in the production stage. Continuous testing post-production is important to maintain integrity concerning the security aspects of applications [12].

As vast amounts of our data and information are stored in digital environments, such as clouds, the protection of these storage spaces is as important as protecting any physical storage units [9]. Most attacks on cloud platforms are very similar to the types of attack we can see in the above-mentioned security types. However, the implementation of security measures on cloud environments is more difficult due to their characteristics [13]. Cloud services rely on virtual machine solutions and these are surfaces for various attacks such as data breaches and DoS attacks. The protection and isolation of these platforms are crucial to ensure cloud security. Another type of cloud service attack is a service injection attack [13]. In this instance, the attacker inserts a malicious input to the system which alters the function of a command.

The last aspect of cybersecurity to describe is endpoint security. Endpoint security, in short, is the practice of securing the entry point of the network from being exploited or attacked. By end- and entry points we mean various devices such as laptops, mobile phones, printers, desktops, smart watches and any other “smart” device that could be accessed via a network [14]. Endpoint security is not a new concept but sometimes an overlooked one. Today, the average antivirus software might not be sufficient to address the complexity of malicious attacks encountered every day. Data is the most valuable asset in most markets and companies, which means there need to be tools to protect it. Since the number of entry points in a network can be several thousands, automatic detection systems and endpoint protections platforms (EPP) are the way in the future to keep up with the pace [14].

2.3 Current Cybercrime Situation

As the number of technological solutions in all aspects of life, services, machines, and companies grows, the number of cyberattacks and cyber criminality has also increased. In 2019, there were 15 billion records (data structures) exploited, which compared to 2010 with a total of exposed records of merely 103 million, is a huge increase [15]. Criminals and other attackers today can get around even sophisticated intrusion and detection systems and the sheer volume of exploitations that companies and private persons are exposed to is very high. In the future, 5G technology will enable more complex IoT networks and exponentially increase the number of possible devices to be attacked. Furthermore, the bandwidth of 5G technology enables more data to be transmitted [16]. An example of how cyber criminals take advantage of uncertain times comes from the recent past and the outbreak of COVID-19. According to Reed Smits's statistics [17], email and online scams have increased by 400%. Scare tactics work, and human emotions are difficult to rule out of the equation.

Currently, there are many challenges with cybersecurity, one of them being the geographically distant connections that expose a weak point in the chain at least in one place. Many security systems still work on manual detection labor and, as noted earlier, there is not enough time to tackle all attack attempts. In addition, computer systems still often use a reactive recovery method instead of a proactive one [18]. A reactive method is not sufficient in the long run and does not necessarily prevent the same problem from occurring again and again. The transparency of cybersecurity methods, as well as available research and data, help criminals to understand the systems and more easily exploit them. Lastly, hackers often use Virtual Private Networks (VPN) and proxy servers in order to change IP-addresses, which keeps them from being caught [18].

Estimations for global costs caused by cybercrime are showing some alarming trends as well. According to McAfee's report from 2020 [19], the global losses related to cybercrime are now over $1 trillion. In 2018, the number was around $600 billion, which shows a doubling in only two years, *Fig. 1*. The report also highlights the fact that two thirds of the companies participating in the study have encountered at least one cyber incident in 2019.

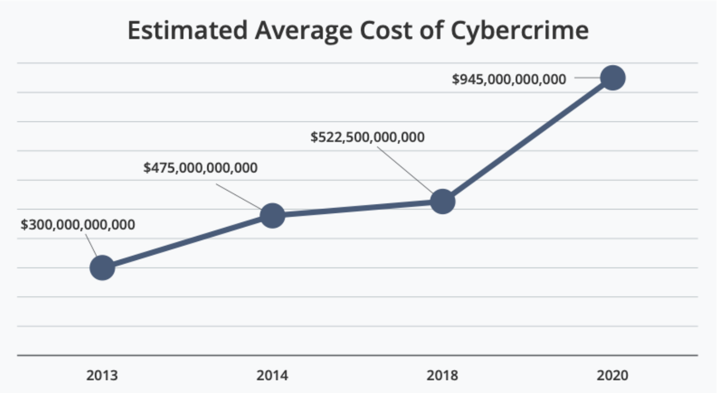

Figure 1: Cost estimation of cybercrime [19]

The total sum of the cost may seem inflated and excessive at first glance. However, the damage to monetary assets and intellectual property is not the only factor in this situation. A big part of the costs comes from the damage to performance in companies. There are many hidden costs in cybercrime, and they are often overlooked by companies [19]. Included in these hidden costs are system downtime, reduced efficiency, incident report costs, and brand/reputation damage. Downtime of systems after an incident shows that the cost for the company lies somewhere between $100 000 to $500 000. In worst case scenarios, the number rose up to approximately $750 000. Time spent on damage recovery during system downtime is by average nine hours [19]. The interruptions in systems are critical for companies and a waste of valuable company time. Additionally, the incident report costs are also adding to the expenses. Depending on the gravity of the incident, the costs vary. Often the reports can be done in-house but in bigger cases external actors must step in, which could be expensive. Today the average consumer is quite aware of the importance of cybersecurity and especially assurance of personal information. Therefore, companies have the responsibility to ensure this factor to the customer. This means that money spent on saving the company's reputation and trustworthiness is a contributor to costs related to cybercrime [19].

2.4 How Artificial Intelligence Improves Cybersecurity

As we can notice, there is a clear incentive to deploy artificial intelligence solutions in future to fight cyber criminality. In this section, we take a closer look at what the actual benefits of AI methods are and why they are worthwhile.

The computational and analytic power of AI tools is faster than human brain power. Compared to current methods, artificial intelligence can achieve a much higher detection speed. In addition to faster identification of threats, unknown attacks can be recognized faster, and proper response methods can be created without a previously implemented method [20]. Human errors are still a big contributor in cybersecurity issues. By implementing AI technology, the number of cases caused humans could be remarkably reduced. This certainly goes for small repetitive tasks that are conducted every day, but artificial intelligence can be exploited in decision making as well [20]. When making decisions, data and software can be tested with AI algorithms and therefore, unnoticed errors and hidden security hazards could be detected early on. The computing power of artificial intelligence may be greater than people's, but creative thinking and innovation are still up to humans. Therefore, AI tools should be implemented in tasks that are routine and repetitive. This frees up more time for security workers to focus on creative thinking and improving the process themselves [20].

Another area in security that can benefit from AI-based solutions is threat hunting [19]. The basic idea for threat hunting is traditionally to have the security software look for indicators of threats. The setback of this method is that the threats must be previously known in order to establish the right signatures for the threat in question. So, in the case of a totally new threat type the technique is insufficient. Signature-based techniques are, however, effective methods and can detect up to 90% of the cases [19]. A complete replacement of signature-based solutions with AI solutions would not be the way forward either since the results would yield a high number of false positives. A combination of these two methods is, however, a profitable approach.

As discussed in section 2.3, the increase in cyber threats requires improved security measures in order to keep up with the pace of the emerging attacks. *User and Entity Behavioral Analytics* (UEBA) is a security process that monitors so called normal system behavior and normal usage of endpoints in e.g. networks [21]. UEBA is a form of artificial intelligence in that it can detect anomalies in the system and react faster when these are occurring. UEBA uses machine learning and statistical analysis to determine if the behavior deviates from the normal patterns, also known as a baseline. In order to detect anomalies, the UEBA system analyzes packet information, files, and aggregated data such as logs and reports [21]. A specific attribute of a UEBA system is that it tracks users and entities in the system instead of monitoring individual devices or security events. An entity in this context refers to servers, routers, and applications. Features that these systems use in the anomaly detection algorithms include time stamps of accessed resources, duration between the first and last access, and the total sum of uploaded and downloaded bytes [83]. When generating an alert of an anomaly the system takes two score values into consideration. The severity score, which is determined by a security expert, is a value of how important an event is regarding business operations. For example, the security events regarding a CEO have a higher severity score than events concerning a server with few devices connected to it [83]. The second score is the confidence score. The value of this score represents how confident the system is about a detected anomaly. If the system is not certain about an event and whether it is an anomaly, the system does generate an alert but with a lower confidence score value [83].

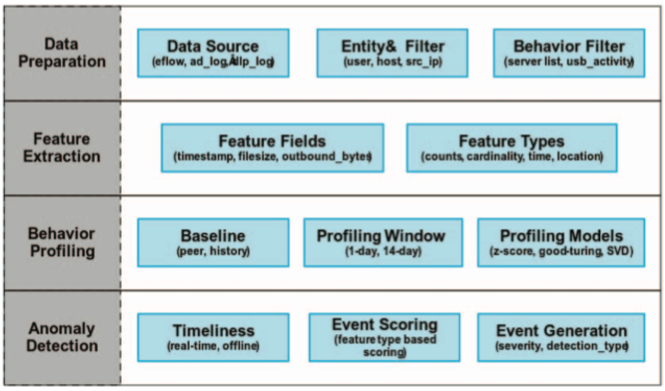
Data Preparation
Data Source (eflow, ad_log, dlp_log)
Entity& Filter (user, host, src_ip)
Behavior Filter (server list, usb_activity)
Feature Extraction
Feature Fields (timestamp, filesize, outbound_bytes)
Feature Types (counts, cardinality, time, location)
Behavior Profiling
Baseline (peer, history)
Profiling Window (1-day, 14-day)
Profiling Models (z-score, good-turing, SVD)
Anomaly Detection
Timeliness (real-time, offline)
Event Scoring (feature type based scoring)
Event Generation (severity, detection_type)

Figure 2: UEBA system architecture [83]

A typical architecture of a UEBA system is illustrated in *Fig 2.* The first section of the architecture is data presentation. In this phase, the relevant data is collected from the existing data sources, grouped into the identified entities, and prepared for the next stage. The next step is feature extraction where the obtained data from the various fields are categorized by each entity. The features in this step are then stored and computed for the coming profiling. In the third phase, called behavior profiling, the extracted features in every entity are grouped into configured baselines. Machine learning is then applied to create a behavioral profile for each specific entity [83]. The last stage in the architecture is anomaly detection. Here the features are compared to the generated behavioral profile and the system yields a confidence score [83].

UEBA systems provide greater insight into system behavior due to automated analytics tasks using machine learning. The detection of novel and diverse types of attacks is also made easier, since machine learning methods can analyze large amounts of data and find hidden patterns that security analysts cannot. Examples of such attacks are insider threats, breaches of protected data, and brute-force attacks [21].

2.5 Cybersecurity Threat Landscape

Cyber threats and cyber criminality are vast and complex subjects that keep on evolving. Threats are becoming more difficult to assess and understand. In this section, the current threat landscape is described. *European Union Agency for*

Cybersecurity (ENISA) listed the top 15 threats in cybersecurity for 2019 and 2020 in their annual threat landscape report [22]. These 15 where malware, web-based attacks, phishing, web application attacks, spam, DDoS, identity theft, data breach, insider threats, botnets, information leakage, ransomware, cyberespionage, cryptojacking, and physical manipulation, damage, theft and loss. A couple of these attack types, relevant for this thesis, will be discussed in section 2.6.

Out of these 15 threats phishing, information leakage, ransomware, insider threat, and identity theft saw an increase during the time period between 2019-2020. Contrarily, spam, DoS, and botnets saw a decrease during the same time interval. According to the report, there were two main factors that created the fluctuations and changes in the threat landscape. Firstly, the outbreak of the COVID-19 pandemic forced organizations, companies, and states globally to quickly adapt to a new working environment that then introduced its own new challenges.

For states, new technologies had to be adopted in order to coordinate national health services, distance learning, international response and responsibility to mitigate the spread of the virus and lockdown measures. Regimes also had to work and make important decisions only by telecommunication. For companies, the problem has been to create a safe and secure cyber environment when working from home. Remote working introduced a large new attack surface with computers and mobile devices connected to the internet as entry points. Remote access, unsecured internet connections, and cloud services are all examples of possible weak points, and the usage of these technologies has grown during the pandemic. The combination of tackling the new challenges and maintaining business productivity and efficiency has been a great obstacle to encounter. This has forced cyber security teams to work at full capacity and ability [22].

The second factor contributing to the shift in the threat landscape are advanced adversary capabilities. These refer to the lately growing trend where attackers use more complex and advanced methods to expose systems. With the outbreak of the pandemic, it was discovered that adversaries adapted to the new environment quickly [22]. The key characteristic of the threat trend for 2019-2020 was personalized attack vectors. An attack vector is a method or a technique used by an attacker that aims to illegally access a network or device. This realm includes advanced types of credential stealing, where the malicious attackers gain access to critical data often by e.g. targeted phishing attacks. Another common threat type shown in the ENISA report is credential stuffing. Credential stuffing is a tactic where the lists of usernames and passwords acquired from credential stealing are used to send thousands of login requests in order to gain even more unauthorized access [23]. New advanced techniques of malware obfuscation, which is the process of altering binary and textual data and thereby bypassing antivirus software, were trending. Furthermore, social engineering and penetration of mobile platforms are among the techniques frequently used during the pandemic.

The overall trends in cybersecurity show us a new perspective on how companies and organizations should prepare for the future [22]. The technological transformation in products, services, and utilities shows no sign of slowing down, leading to a bigger attack surface for an attacker to act on. As the use of social media platforms has increased, their ability to affect people's minds, influence political elections, and undermine democracy has grown. This means that the cybersecurity environment, such as social media applications, needs to be even more reliable and robust. Another trend is state organized attacks that are highly targeted and aim to access intellectual property and state secrets.

Financial gain is still the main goal of attacks. Attacks are more distributed, more widespread, and shorter lasting but have a wider impact. As mentioned before, and also brought up in the report [22], most of the incidents in cybersecurity still go unnoticed and the response time is too long to keep up with the number of attacks. Social engineering and human errors will be exploited more. Humans will be considered the weakest link in the system as cyber defense systems become more sophisticated. One of the main conclusions in the summary of the ENISA report is to further investigate artificial intelligence tools in the cyber threat intelligence community. Motivation for this is that these reduce the number of manual steps when tackling attacks and add value to machine learning functions [22].

2.6 Types of Threats

In this section we take look at some of the threats and attack types listed by the ENISA report [22]. For the purpose of this thesis, a few of these will be defined and explored deeper in the chapter.

2.6.1 Malware

Malware is an often-used method in cyber criminality. Malware is malicious software used to accomplish identity theft, cyber espionage, and disruptions in systems. Malware appears in the form of viruses, Trojan horses, and ransomware [24]. In contrast to a typical bug, malware is an attempt designed to cause harm. As seen in *Fig 3,* malware stands for the majority of the attack distribution of 2020.

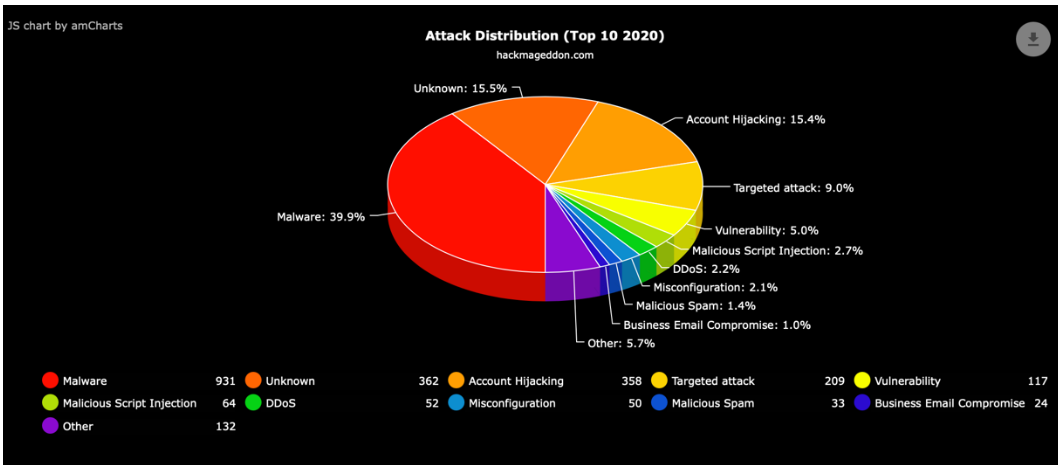

Figure 3: Attack distribution in 2020 [77]

The usage of malware has seen a shift from consumer targets to business targets. Further statistics show that 50% of malware attacks were designed to steal personal information and that 71% of businesses which were targeted saw malicious software spreading between employees [24]. A new alarming trend in malware is the concept of Malware-as-a-Service (MaaS). Contrary to Software-as-a-Service (SaaS), MaaS is a black-market business for criminals that sell and rent malware to other criminals [25]. This enables people without the technical skills to perform cyber attacks. The ease in how almost anybody can launch an attack via MaaS organizations has led to an increase in botnets [25]. A botnet is a network of devices with the goal of launching attacks controlled by one host server. The upscaling of botnets makes it challenging to identify the host in charge and responsible for the malicious activity.

2.6.2 Web-based attacks

Web-based services are increasing in the world and this offers an appealing opportunity to be exploited by malicious actors. Web-based attacks are a way to, for instance, inject malicious script of false URLs to users, and by that redirect the user to a desired webpage [26]. This method can also be used to have the victim download malicious files and to inject harmful content into webpages that are trusted but have compromised features. An overload of login requests with usernames and passwords, such as brute-force attacks, affects the key features of cybersecurity, such as availability of web sites as well as the confidentiality and integrity of the information, accessible in the web services.

A typically used attack technique based on this method is *Drive-By-Downloads*. In these situations, the end-user only has to click a compromised link or visit (drive-by) a webpage that contains malicious code that is then downloaded onto the device [27]. A second tactic, practiced in web-based attacks, is *Watering hole attacks.* Here the malicious actors observe user tendencies and look what web sites certain groups in organizations use and visit. The specific webpages are then injected with harmful scrips and false advertising that can potentially be clicked [26].

2.6.3 Phishing

The COVID-19 pandemic caused a massive increase in regard to phishing attacks. ENISA reports [28] that the number of phishing scams in one month rose with 667%. Phishing attacks are most often seen attached to emails. The email is meant to look trustworthy but contains malicious attachments and links. These emails try to tap into human emotions and cause human errors as a consequence. In order to achieve that, keywords such as “payment” are used to evoke rushed and ill-advised decisions from the victim [28]. The overall uncertainty and fear caused by the pandemic has given cybercriminals potential to exploit humans effectively. For instance, criminals send emails, claiming to be big reliable organizations such as World Health Organization or a national health organization, with harmful attachments. Remote working has also caused problems in the prevention of phishing scams. Microsoft programs are used in almost every company and organization. Phishing emails with fake error messages and fake update suggestions are typical examples of phishing attacks [28]. With usage of AI, mitigation is possible with pattern seeking and learning abilities. For instance, AI can recognize typical dialog patterns, interaction characteristics, and grammar and syntax anomalies, and it can scan images attached to emails to detect fraudulent links and login requests [29].

2.6.4 Web application attack

The utilization of the internet has led to more web applications and these play a vital role in companies' ability to provide services. Web applications rely heavily on databases that have the task of storing and transmitting the requested data to the user [32]. The typical attack methods used to threaten databases are SQL injections (SQLi) and cross-site scripting (XSS). SQLi attacks exploit vulnerabilities in web security by injecting malicious code to database queries, thus gaining access to data and altering it in a harmful way [33]. XSS attacks work with the same principle, but here the malicious code is planted in web applications and websites that can then redirect the end user to malicious websites [34].

2.6.5 Distributed denial of service (DDoS)

DDoS is the phenomenon where a user is not able to access certain data or resources in a system. To accomplish this, the attackers flood the host network or target with requests and traffic, resulting in the system crashing for not being able to respond [35]. DDoS attacks are not unfamiliar to cyber security experts, but the techniques used by malicious actors are becoming more advanced. Trends in DDoS attacks show that over half of the attacks last for less than 15 minutes and that multiple attack vectors are used at the same time [36].

2.6.6 Data breach

A data breach is a type of incident where data and/or parts of an information system is accessed without authorization [30]. Once having accessed the breached system, the attacker can misuse and destroy the data. Data breaches and human errors are closely linked, since many vulnerabilities and unintentional exposure of data are due to insufficient implementation and configuration of the system itself [30]. The sophistication of the methods cybercriminals use has led to the fact that many organizations and companies are not even aware of a data breach. In addition to the novelty of the crimes, systems can be lacking in visibility and classification that prevents them from detecting the attack happening. According to IBM's report [31], the average time it takes to identify and contain a breach is around 208 days and the average total cost is $3.86 million. These time and cost estimations are based on how long it takes to return back to normal until the breach is totally remediated, and all the data recovered. Increase in in-house and on-premises solutions for data storage and cloud infrastructures has presented new challenges in maintaining security policies, since the attack surface has grown [30]. The most used methods to achieve data breaches are email phishing and web application attacks. The third relevant threat type in this scenario is insider threats. This does not only refer to corrupt employees but generally the human errors made by internal workers [30].

3. Artificial Intelligence Algorithms in Cybersecurity

3.1 Machine Learning & Deep Learning

When talking about artificial intelligence we talk about a broad field with several underlying fields of technology. During the last decades, AI has benefited greatly from improvements in deep learning (DL) and machine learning (ML) technology. *Fig. 4* illustrates how AI, ML, and DL relate to one another.

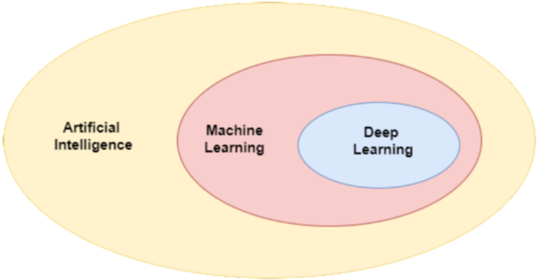

Figure 4: Relationship between AI, ML, and DL [86]

ML technology is a sub-domain of AI that enables computer systems to learn from previously gathered data. This data is then used by the ML application to learn and create its own solutions and methods, without explicit programming [37]. Specific machine learning tasks include classification, regression, and clustering. Classification is based on the idea of assigning each data point or item into a category. For instance, in image recognition the ML algorithm is fed pictures of two types of objects and its task is to categorize them into two groups. Classification with two categories is called binary classification, and a model with more than two is called multiclass classification. ML algorithms can also be utilized in prediction tasks. This is called regression, and here the algorithm aims to predict a value for an item. An example of regression is the prediction of stock values. The prediction is made possible by having the algorithm learn specific relations between data points provided to it. The accuracy of the prediction is determined by how far off the predicted value was to the true value [84]. Clustering methods are useful for large data sets. Clustering is the concept of dividing data into similar or homogenous sub-groups called clusters. The difference between clusters is not always clear, but the main point is to have data points more similar to one another in one cluster and those dissimilar in another cluster [84]. For efficient classification, regression, and clustering the dimensionality of the input data needs to be reduced. This can be done by recognizing specific features from the data. A feature is one or many attributes associated with a data point [84]. Features from a data set can be distinguished by feature selection or feature extraction. Feature extraction is the concept of translating the raw input data into features. The features have to be represented in such a format that they are suitable for the algorithm. Feature selection is the process of selecting a subset of the discovered features that are relevant for the task and discarding the rest [85]. Input data fed into an algorithm can be labeled or unlabeled. A label is a value or category assigned to the raw input data [84]. By labelling data, one provides context to the information.

ML applications have played a big role in cyber security-related task for a long time, but there are drawbacks to them. Even though ML methods can be powerful, they rely on previously known data and information [38]. In order to detect malware and malicious attempts using ML, the system has to be informed about the characteristics and features of the malware in advance. This, however, means that traits associated with the malware that are not pre-defined cannot be detected by the ML application [38]. As a solution to the problem, DL methods were introduced.

DL is a sub-domain of ML that uses artificial neural networks (ANN) when making decisions. An artificial neural network is a concept that mimics the human brain, which is a biological neural network. Individual cells in the brain, called neurons, are connected to each other and send signals to each other via synapses [87, 88]. ANNs use interconnected neurons also known as processing units. The input to a processing unit can either be from an external source or from the output of another unit. The output from the units can be sent to other units or sent back to itself [87]. The connection between units and how the information is transformed within the network depends on the weight. The weight value describes the strength of a connection between two units. The higher the weight value is, the greater influence it has. The weight value can be negative, in which case the input is considered unimportant [88]. By adjusting the weights, the ANN gives more accurate results.

Another important term in neural networks is transfer function. A transfer function is a mathematical function that translates the input data to output signals [90].

There are three types of layers in an ANN. These are the input layer, middle or hidden layer, and output layer. The most rudimentary form of an ANN is one where there is one of each layer. The input layer receives information from external sources and outputs its information to the hidden layer. The same procedure is done from the hidden layer to the final output layer. The term hidden refers to the fact that the units in the middle layer are not inputs or outputs [89]. An ANN with several hidden layers is called a multilayer perceptron or deep neural network [88]. A type of ANN is a feedforward neural network. In a feedforward network the information moves in one direction, from the input layer all the way out to the output layer. The output from the previous layer serves as the input for the next layer. A neural network where the flow is not only one-directional and where feedback loops to single units are possible is called a recurrent neural network [89]. An artificial neural network that has all its units interconnected all the way from the input layer to the output layer is called a fully connected network [84, 86].

Both machine learning and deep learning can be categorized based on how the training data is received and how the algorithm learns from the data. In the next section, the various categories will be explained.

Supervised learning works with labeled data that is then outputted as known results [86]. The term supervised refers to the concept of having humans label the input data and define the output as output categories. The learning algorithm then trains the model to make predictions of unseen items [84]. Examples of supervised learning are classification and regression. Unsupervised learning algorithms are given unlabeled data as input. The algorithm tries to learn patterns in the unlabeled data and structure it without explicit predefined rules [84,86]. A typical example of unsupervised learning is clustering. Competitive learning is a type of unsupervised learning. In this learning mechanism, the individual units compete for executing tasks in the learning process [91].

Semi-supervised learning is a combination of supervised and unsupervised learning methods. Semi-supervised learning algorithms receive both labeled and unlabeled data in order to make predictions of new data points that are fed to it [84]. This method is useful when feature extraction is difficult, or the labelling of data is too time consuming. Another common approach is reinforced learning. In this case, the algorithm learns by constantly interacting with its environment [86]. The idea is to reward certain action by the network by giving feedback on the action performed. The punishment and reward methods help the algorithm to find the most optimal way to perform a task [86].

In ML and DL, the core function is to generalize data. Generalization describes how well the model learns from the given data and how well it can apply the knowledge to new information [92]. A problem called overfitting is associated with generalization. Overfitting occurs when the model does not react accurately to new data after the training stage [92].

The following section will present the main practical differences between ML and DL methods [39]. The first distinction between ML and DL is data dependency. The performance of ML and DL algorithms varies depending on the amount of data available. DL methods do not work that well with small amounts of data. ML, in turn, can perform on smaller volumes of data, as long as the rules for its actions are established. Depending on the algorithm, there are also certain hardware dependencies that affect the outcome. A key element in DL algorithms are matrix operations. These operations require GPU power and as the volumes of data increase, high-performance GPUs are necessary for DL methods to work.

Another key difference between the two concepts is feature processing. This process encompasses feature extraction where complex data is divided and reduced into smaller manageable groups. In ML, the features and general data patterns must be determined by humans. Therefore, the performance of an ML algorithm depends on the initially determined extraction of features. DL utilizes unsupervised and semi-supervised learning more that ML methods, which makes the feature extraction more automated and better for obtaining high-level features from complex data. ML and DL often have different approaches to solving a problem.

ML methods work by breaking the initial problem down into smaller sub-problems. The final result is then obtained by solving the smaller sub-problems first. In DL, there is more of an end-to-end problem-solving approach, whereas many mid-layer steps are eliminated, hence optimizing performance.

There is also variation in execution time. ML and DL differ in time consumption regarding training and testing. In the training phase, DL methods are more time consuming than ML since DL algorithms have a larger number of parameters. On the contrary, when testing operations DL is faster when it comes to performance. The time it takes to process data in the testing phase with ML algorithms is proportional to the amount of data. Lastly, the way results are interpreted depends on what type of algorithm is used. The predetermined rules of the algorithm in an ML provide humans with better insight into how a certain problem was solved and what decisions the algorithm took. On the contrary, DL methods, and the results they generate, are more difficult to explain due to the nature of e.g. deep neural networks. The problem solving is, as mentioned earlier, characterized as an end-to-end solution and many of the algorithm's decisions are made in the hidden layers.

Typically, ML methods encompass 4 steps [39]. 1. Identify and choose the attributes and features for the basis of your prediction. 2. Select the machine learning algorithm for the task e.g. classification, regression. 3. Train the model by changing parameters and algorithms, and then choose the best performing one. 4. Use the selected model to predict the data. The same principle can be applied to deep learning methods.

In the coming section, we shall look at various approaches of ML and DL algorithms in the field of cybersecurity. We first take a look at more traditional machine learning techniques and in the second part analyze deep learning methods.

3.2 Machine learning approaches

ML techniques utilized in cybersecurity are a powerful tool. The ability to learn enables computer systems to enhance performance and create execution models from previous data.

3.2.1 Neutral network approach

The neural network (NN) approach is a technique used to predict user behavior in a system. The main advantage of an NN is that it can interpret unknown data that is either unspecific or imprecise and create a solution model by itself. In addition to the tolerance of imprecise data, NN algorithms work well when generalizing data. However, neural networks require initial training. In the realm of cyber security, examples of attacks, non-attacks, and the difference between them need to be set in the training phase for the system to work properly [40].

A way in which NN techniques can be used is self-organizing maps (SOM). SOM is a neural network that takes high-dimensional data, which is a dataset with many attributes, and transforms it to a simpler one- or two-dimensional grid. The concept was introduced by the Finnish professor Teuvo Kohonen and is also known by the name Kohonen Map [41]. The SOM technique is frequently used in tasks that require pattern recognition and data compression. SOM builds topological maps from the input data, revealing underlying structures. A topological map is a representation of a data set where only the necessary details are visualized. The SOM is similar to many clustering techniques. The nodes are placed in a two-dimensional grid in such a way that nodes from the input data that resemble each other are placed close to one another [41]. This is achieved with competitive learning where the vector weight of the node is compared to the input vector, and the node most similar to the input vector is placed closer to the input vector. This node is then the best matching unit (BMU) and the distance to the BMU determines the weight of the other nodes in the network [41].

Pachghare et al. [41] conducted a study into how self-organizing maps could be used in intrusion detection. The study used unlabeled packets from real networks for the training model. The results of the experiment showed that SOM is an effective method for intrusion detection systems as it could recognize intrusive behavior by itself. Since the training data needed to be labeled first for the experiment to work, the process was quite time consuming.

3.2.2 Support Vector Machines

Support vector machines (SVM) are often used in classification and regression related tasks. SVM is an unsupervised learning technique that is based on finding a way to separate the data into two classes with a so-called hyper plane [40]. This hyper plane is a line drawn between the data points in a two-dimensional graph. The datapoints closest to the hyper plane are called support vectors, and they determine the separation between the two classes. In case the placement of the hyper plane is not possible in 2-D, the datapoints have to be mapped out in higher dimensions, with a process called kernelling, see *Fig 5* [40].

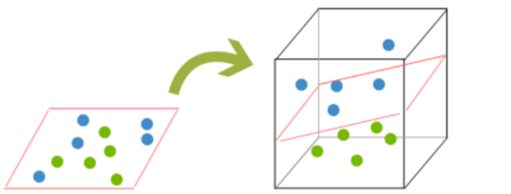

Figure 5: Finding hyper plane in 3-D [78]

One exceptional aspect with SVM is its ability to generalize data despite the number of features. Additionally, SVM has been shown to perform better than artificial neural networks (ANN) in both scalability and prediction accuracy [40]. A study by Chen et al. [42] compared SVMs and ANN tools in intrusion detection, and their findings show a couple of advantages with SVM. SVMs use the structural risk minimizing principle, which works by selecting the simplest classifier in order to avoid too specific boundaries. By applying this method new unseen observations are more easily classified. In other words, we minimize the risk of wrongly sorting the data. SVM techniques are also more suitable for global solutions and they require fewer parameters compared to NN where the number of hidden nodes, transfer functions, and hidden layers need to be determined [42].

3.2.3 Markov model

The Markov model is a probabilistic approach to map out states. A state is, for example, a set of values or a specific situation. Within this concept there are two core ideas: Markov chains and hidden Markov models.

A Markov chain consists of a set of states that are connected to each other with certain probabilistic transition values [40]. In other words, a Markov chain tells one the likelihood of where the system will transition from one state to another. In detection of threats, Markov chains can be used to establish the transition probabilities in the training stage so that they reflect normal system behavior. Hixon and Gruenbacher [43] show in their study that TCP/IP headers can be used to identify harmful packets by looking at their transitions in the system. If the transition does not match the thresholds established in the training stage, this can be marked as an anomaly or a potential threat.

The hidden Markov model (HMM) is an approach where the hidden state is predicted by observing the symbols that the hidden state emits [44]. For example, one does not know the weather, but based on the clothing of people one can make a pretty accurate guess of what sort of weather it is. HMM techniques for intrusion detection can be found in many studies. Mahoney and Chan presented a packet header anomaly detection (PHAD) system [45]. The study showed that by applying HMM methods the system could learn to obtain normally behaving profiles in different layers of a network.

3.2.4 Decision trees

A commonly used tool in ML for classification and prediction are decision trees. This method can be used to approximate discrete functions by classifying data with a set of rules [38]. The three components of a decision tree are nodes, arcs, and leaves [40]. The nodes are labeled with features that split the attributes of the node. The arc coming out of the node is labeled with an attribute that shows the next step. The final stage is a leaf that is labeled with a class or a category [40]. Decision tree methods can be applied to detect DoS and injection attacks. Vuong et al. [46] proposed decision trees in detection of attacks against robotic vehicles. In the study, features of DoS and injection attacks were used to build the decision tree. In addition to cyber features such as network traffic, physical features of the vehicle such as speed, power consumption, and vibration were taken into consideration. The result showed that attacks have different impacts on cyber features and physical features. By adding the physical features to the equation, the accuracy of the attack detection increased. Moon et al. [47] researched the usage of decision trees in intrusion detection to prevent advanced persistent threats (APT). An APT is a sophisticated type of cyberattack where the intruders aim to enter a network without detection, and establish a presence for a long period of time [48]. Moon et al. [47] looked at APT attacks that could change after having entered the network, and proposed decision trees to analyze the behavior of the system. Based on the system behavior the decision tree could determine if an intrusion was occurring. The detection accuracy of the proposed method was 84,7%, which in the case of APT detection is high.

3.2.5 Clustering technique

In clustering methods, the core idea is to sort the data into clusters by looking at the similarities and differences between two instances. To create the clusters, a number of fixed points that best represent the attributes of the cluster called representative points are selected and when a new instance is presented, the proximity to the corresponding representative points determines the cluster it belongs to [92, 40]. For anomaly detection, there are two main approaches. The first approach uses unlabeled data of both normal network traffic and attack traffic in the training stage. For this method to work, we must make the assumption that anomalous traffic constitutes only a small percentage of the entire data. As the data is clustered, we can deduce that the large clusters represent normal network activity and the smaller ones that are outliers represent attacks [40]. In the second approach, only normal traffic is given as input to the system in the training stage, thus creating a model for normal behavior.

A common clustering technique is k-nearest neighbor (k-NN). The k-NN method relies on labeled data in the training stage and the distance measurement between two data point (x, y) can be established with Euclidean distance, which is defined as:

$$d(x,y) = \sqrt{\sum_{k=1}^{n}(x_k - y_k)^2}$$

Liao & Vemuri [49] presented an algorithm based on the k-NN technique for intrusion detection. In the experiment, program behavior was classified as normal or intrusive. The results showed that they were able to effectively detect intrusive behavior and reduce calculation time for classification.

3.3 Deep learning approaches

In the following section, we present deep learning methods against cyber attacks. ML and DL have many similarities as mentioned earlier. However, the main difference between these two approaches is the feature selection. In DL, the feature selection is automated contrary to ML where the selection must be done manually. The purpose of DL approaches is to gain a deeper understanding of the input data and obtain data features that could not possibly be detected by humans [38].

3.3.1 Deep Belief Network approach

A deep belief network (DBN) is a generative model that can graphically represent all the possible values in a case [39]. A DBN is a set of stacked restricted Bolzmann machines (RBM). RBMs are a type of neural network that can create probabilistic distributions among the set of inputs, and they consist of a visible layer and multitude of hidden layers. Each layer contains many units, *see Fig 6.* The word restricted in the name refers to the idea that the units in a specific layer are not interconnected [98].

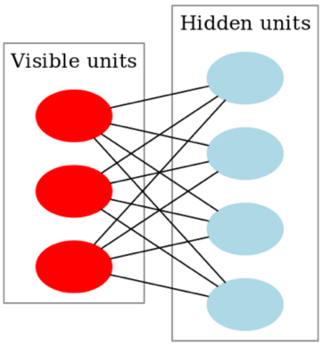

Figure 6: Restricted Boltzmann machine with 3 visible and 4 hidden units [79]

As mentioned, a DBN is a bigger neural network built out of many RBMs, which enables the DBN to efficiently train the dataset [39]. In practice, the hidden layer of the previous RBM works as the visible layer to the next set of RBMs and so on, until all the layers of the DBN are trained. A key characteristic of a DBN is that each RBM layer is trained to know the entire input and this enables the network to detect patterns in the data. The fine-tuning and labeling of the patterns from the training stage requires only a small set of labeled data, which is advantageous when it comes to performance [50].

Studies into how DBN can be used in detection of intrusion and malware have shown that it is a useful concept and can outperform many ML approaches. Ding et al. [51] researched how DBNs could be used in malware detection. The study investigates how DBN performance compares to decision trees, kNNs, and SVMs, which are all shallower neural networks. In the study, an unsupervised training approach was used for the DBN to find features. The features were then used in a feed-forward neural network. This means the signal flow is one-directional, input to output, and goes through every layer one at the time. The unsupervised pre-training approach decreased overfitting and made the training of networks with many hidden layers easier. In the conclusion Ding et al. also mention that DBNs can take unlabeled data and learn from it, which resulted in this technique performing better in classification tasks than decision trees, kNNs, and SVMs [51].

3.3.2 Recurrent neural networks

Recurrent neural networks (RNN) are types of artificial neural networks that use data sequences. Contrarily to a one-directional feed-forward network (FNN), an RNN allows the signal to travel in both directions by adding loops to the network, see *Fig 7.*

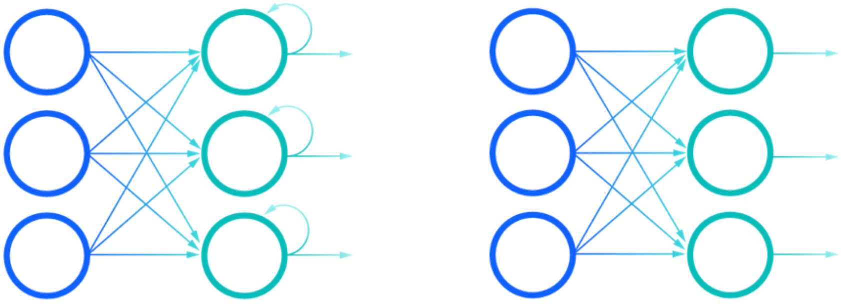

Figure 7: RNN on the left. FNN on the right [80]

By introducing the loops, RNN has a type of memory that allows it to remember the data from previous instances and therefore affect the current input and output [39]. RNN models are suitable for situations where the data patterns change over time and for situations where one wants to find out how multiple input variables influence the output.

Recurrent neural networks can be utilized in cybersecurity. Yin et al. [52] proposed an intrusion detection system based on RNNs. The goal of the study was to compare traditional ML methods in binary and multiclass classification against RNNs. The results show that RNNs give a higher accuracy in detecting intrusions compared to the traditional ML techniques and are therefore very suitable for the task.

3.3.3 Convolutional neural networks

Convolutional neural networks (CNN) have during the last decade excelled in the field of DL, and are a discussed topic in tasks such as image recognition and speech analysis. A CNN is a feed-forward neural network with four types of layers, *fig 8* [39,53]. The first layer is a *convolutional layer* that convolves the data with kernels, also known as filters. In the realm of image recognition, the first layer uses the filters to extract broad features of the image and sums up their weight as output to the coming layer [53]. The second layer, called ReLu (Rectified Linear Unit) or *activation layer*, is a linear function that outputs the previous input from the convolutional layer if it is positive, otherwise the output will be zero. The ReLu activation is widely used in DL methods to avoid slowdowns in the training of the model. Together, the convolutional layer and the activation layer create very complex and large patterns. To simplify the data a layer called the *pooling layer* is added. As a result of the pooling layer, the dimensionality of the data is reduced, which enables the CNN to focus on the desired patterns. In conjunction, these three layers discover a multitude of patterns without any connections. In order to allow the network to classify and understand the processed data, a *fully connected layer* is added at the end [39,53].

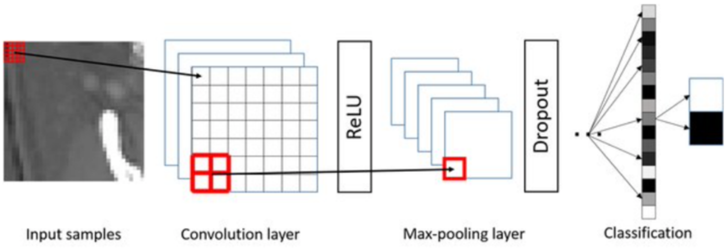

Figure 8: CNN diagram [81]

One main advantage with CNNs is their capability to discover and learn hierarchical features, which makes them a good candidate to apply in cyber security. Kolosnjaji et al. [54] implemented their network with the purpose of classifying detected malware into predefined classes. The constructed neural network consisted of convolutional and feed-forward layers. The features for the classification were obtained from portable executable file headers. The result shows that the addition of convolutional layers improved the classification accuracy and outperformed e.g. SVMs and simple feed-forward networks.

3.3.4 Automatic encoder approach

An automatic encoder, or autoencoder, is a neural network that uses unsupervised learning. In autoencoder networks, the dimensionality (number of units) of the output layer is the same as in the input layer. An autoencoder works so that it takes the input data and reduces the dimensionality of it. The middle layers, or hidden layers, have therefore a smaller number of units to work with. This allows the network to reduce noise and extract important features of the input. The output is a reconstruction from the reduced version in the middle layer [55]. The architecture of an automatic encoder typically consists of three components, as visualized in *Fig 9*.

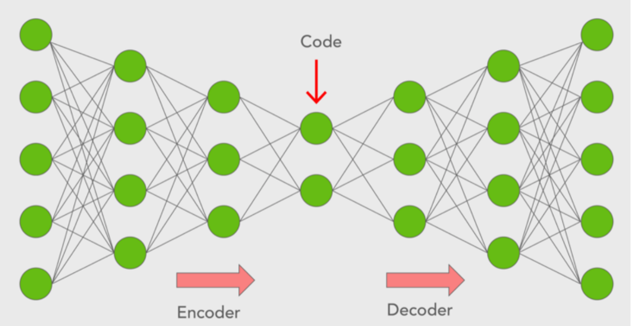

Figure 9: Automatic encoder diagram [82]

The first component in the network is called encoder. This fully connected feed-forward network takes the input and reduces its dimensionality into a latent space representation of the original input data. The code component receives the compressed version of the input. After this the data is fed into a similar type of structure as the encoder, called the decoder. This component does the opposite and reconstructs the data to represent the input, with the same number of units and dimensionality [55]. In order to understand inherent structures and features of more complex data sets, autoencoders can be stacked, creating a deeper network. The concept of autoencoders is used in CNNs, RBMs, and DBN.

Yu et al. [56] used a combination of a CNN and stacked autoencoders called dilated convolutional autoencoder (DCAE) for intrusion detection. The advantages of both methods are especially useful in extracting important features from data that is varied and unlabeled. The data sets in the study consisted of both normal network traffic as well as adversarial types such as botnets and web-based malware. The experiment was conducted with three different classifications tasks, each working on a mixture of normal data and malware traffic. The overall accuracy in the classification tasks showed good results and the accuracy for precision, recall, and test accuracy called the f-measure was 98,44%, 98,40% and 98,40 % [56].

4. Methodology

This thesis was conducted in collaboration with If insurance company and their network monitoring and security teams. The research is based on conducting interviews with various workers at If in order to find out the most common problems regarding network monitoring and network security. Ten participants took part in the inquiry. The goal was to investigate how AI solutions could be utilized to tackle the main problems and to see what the future of AI-based network security and monitoring looked like.

4.1 Research methodology

The method used in this thesis is a qualitative research. A qualitative research is based on the deeper emotions of individuals and their way of acting. The focus is the behavior of individuals and groups and their perspective of things. Qualitative research is often based on interviews, observations and personal views [93].

Small numbers of interviewees are chosen in a qualitative research. Even though the number of participants may be small, the information flow shall still be large, and the information collected shall be enough for the research question to be answered. The interviewees should be chosen wisely, and the fact that the candidates are appropriate for the study is crucial [94].

4.2 Interview structure

The interviews conducted in this study were done via video calls with Microsoft Teams. Ten employees from If insurance company were interviewed, *see Table 1*. The discussions in the interviews were free-form and the goal with the interviews was to pinpoint the problems of network monitoring and security tasks at If. The ten interviewees were chosen by my advisor at If. All the participants were either a part of, or in some way connected to, the network and security monitoring tasks at If.

The interviews were semistructured, which means that the discussions were directed in a specific way towards a specific topic and research question. The interviews were still built as free-form discussions, where the interviewees could describe the situations in their own words. The reason semistructured interviews were chosen as a method is that these are effective for pinpointing specific improvement ideas and possible issues, which form a critical part of this thesis [94].

4.2.1 Analyzing the interviews

In qualitative interviews there are two main approaches when it comes to interview transcription. The deductive approach in a qualitative research consists of a well-planned and predetermined base, with a clear connection between research and interviews. Inductive methods can be divided into two subgroups called thematic content analysis and narrative analysis. Thematic content analysis strives to find common themes and patterns in the data from the interviewees, without a predetermined framework. A narrative analysis is based on analyzing interviewees' responses and highlighting key observations [95].

For this thesis a thematic content analysis approach was selected. This was done because the main purpose of the interviews was to find common problems that the participants have encountered.

4.3 Conducting the interviews

Interviewee	Position
1	Head of Service Delivery
2	Service Desk Analyst
3	Service Delivery Manager
4	Head of Architecture and Platforms
5	Site Reliability Engineer
6	Information Technology Security Specialist
7	Junior Site Reliability Engineer
8	Site Reliability Engineer
9	Nordic Site Reliability Engineer Manager
10	Network Service Specialist

Table 1: Interviewees' job positions at If insurance Company

The overall framework of the interviews was the following. The interviews started by discussing the existing network and security monitoring tasks at If, and what sort of tasks the specific participant worked with daily. The participants were then asked what sort of problems they have encountered in their tasks and how the problems affect monitoring and security aspects. Improvement suggestions were discussed as well. Each interview lasted approximately 30-45 minutes. Notes about the problems and the improvement suggestions were taken and later transcribed. The majority of the participants were only interviewed once, but shorter and more frequent meetings were held with a couple of the interviewees. Notes were taken during all the interviews.

4.4 Limitations

One of the limitations in the study was that the interviews were not recorded. Notes were taken during the interviews, but since no recordings were made no direct quoting is possible.

None of the participants were specific AI experts, which made it difficult to discuss the topic of AI in general. In hindsight, this was not optimal since a more in-depth discussion and elaboration about AI tools and their implementation would have been very helpful.

Another limitation for the study was that I had no access to or training in If's network monitoring tools. The COVID-19 pandemic also prevented face to face discussions and on-premise work at the office.

5. Results

The issues, suggestions, and solution ideas from the conducted interviews are presented in this chapter. The participants are listed as numbers from 1-10 in order to keep the interviewees anonymous. We will first list the problems that the interviewees brought up, *see Table 2.* In the second table, *Table 3,* we will list the suggested improvement ideas that came up in the discussions. Lastly the interview results are summarized.

Interviewee	Problems brought up
1	Network data visibility
2	Lacking relevant information from network data Lack of real-time data
3	Alert accuracy and handling
4	Planned software updates require new hardware, which makes the updates unfeasible.
5	Too much sifting through of data (manual monitoring work)
6	Data visibility Lateral movement monitoring The monitoring systems could be more proactive
7	No information available about devices in the network
8	Lateral movement monitoring Lack of real-time visibility and alerts of network traffic anomalies and threats
9	The monitoring tasks and detection tasks are too reactive
10	Network data visibility

Table 2: Problems brought up in the interviews

Interviewee	Suggestions of solutions and points of interest
1	Requirements for AI-based network infrastructure
2	(No answer.)
3	Smart monitoring Clear dashboards Have good network monitoring tools
4	Management of network infrastructure Introduce clear responsibility allocation when developing network monitoring
5	Introduce skill certifications in the field of network monitoring and security Automation of parameter and feature selection
6	Take a deeper look into NDR, SIEM, and SOAR technologies
7	Real-time network data visibility Complete and clear network data
8	Endpoint protection and detection system improvements
9	Find information of what solutions for network monitoring and detection are available
10	More network monitoring tools

Table 3: Suggestions of solutions and points of interest

5.1 Analysis of interview discussions

In this section, a more detailed description of the discussion with each interviewee is provided. The content is based on the notes taken during the interviews. Each section contains themes that are listed in *Table 1* and *Table 2*.

5.1.1 Interviewee 1: Head of Service Delivery

The main problem encountered in network monitoring according to the interviewee was the insufficient visibility of network data. No clear solutions were brought up in this interview, but this person was interested in seeing what AI could bring to anomaly detection. Another topic discussed was what reforms and changes an existing network needs to go through in order to make the anomaly detection task more automated.

5.1.2 Interviewee 2: Service Desk Analyst

In this discussion, the key issue brought up by the interviewee was also related to data visibility. The person mentioned that the data about an anomaly that is analyzed sometimes lacks relevant information, and that a solution for this would be to gather more information about what caused it and what the root problem was. Furthermore, an important aspect of network monitoring mentioned by the interviewee is to have in-house development of the monitoring and security systems. The person claimed that this would make it easier to customize the infrastructure according to the organization's needs.

5.1.3 Interviewee 3: Service Delivery Manager

This person brought up that there could be improvements regarding alert handling and alert accuracy. In future, the interviewee would also like to see that incident managers would be more active in developing and improving the monitoring systems and that this would lead to more accurate alerts and more reliability in the system outputs. Solutions brought up by this person were to have smart monitoring and to introduce clear dashboards that the monitoring team could work with.

5.1.4 Interviewee 4: Head of Architecture and Platforms

In this interview, aspects regarding management were discussed. This person mentioned that it is important to have clear allocation and delegation of who is responsible for certain parts of the monitoring and security process. Another aspect that this person thought was important was to have in-house capabilities and that this approach should be encouraged in the future. An issue that was brought up was that some planned software updates would require massive hardware updates, and this was not viable or worthwhile from a cost benefit perspective. Another technical problem was that some software used in the company was difficult or impossible to interconnect to other parts of the system.

5.1.5 Interviewee 5: Site Reliability Engineer

Interviewee number 5 pointed out that some monitoring tasks and root cause investigations were still done manually. This person said that solutions for this would be to have more automated anomaly detection solutions. According to the interviewee, a specific suggestion that could eliminate some of the manual labor would be to have automated feature selection. Another wish for the future that the interviewee brought up was that employees would be trained more in tasks related to monitoring. The interviewee mentioned that skill certification would be a possible solution and that this would improve the overall monitoring process.

5.1.6 Interviewee 6: Information Technology Security Specialist

This interviewee also put forward the issue of data visibility. The specific type of network visibility discussed was lateral movement and how to detect any attacks and threats within the network. Lateral movement is a technique used by attackers to get deeper into the network once gaining access to it. This interviewee brought up many technologies such as network detection and response systems, and security operations automation and response technology. The interviewee said that these types of technologies were relevant for the company’s future plans regarding network security and monitoring. Skill certification was also on the list of possible improvements for the future.

5.1.7 Interviewee 7: Junior Site Reliability Engineer

Interviewee 7 had in his/her daily task noticed that there was not enough information about some devices connected to the internal network. Solutions that this person would like to see in future were more complete data about the endpoints and more real-time information about network traffic. The reason why the interviewee brought up these solutions would be that the improvements would make the interviewees work more efficiently.

5.1.8 Interviewee 8: Site Reliability Engineer

In this interview, the topic of data visibility was brough up again. According to this person the system currently used lacked real-time visibility and alerting of network anomalies and threats. Points of interest expressed by the interviewee included endpoint detection technologies and how they could improve the monitoring process. A second feature that would be useful in future, according to this person, was to investigate which technologies would improve the visibility of lateral movement and threats associated with it.

5.1.9 Interviewee 9: Nordic Site Reliability Engineer Manager

Interviewee number 9 mentioned that the current system was too reactive. This person was interested in knowing how AI could make a network monitoring system more proactive and how an AI-based system could provide more real-time data and information. The interviewee was also interested to know what AI-based solutions there are available on the market and how the company would benefit from these systems.

5.1.10 Interviewee 10: Network Service Specialist

The last interviewee also highlighted the network visibility issue. The interviewee said that a solution for this problem would be to have more monitoring tools and well selected tools that were specifically designed for a certain problem. The motivation for the suggested solution was that this would make the monitoring tasks more efficient and that this would make the monitoring easier.

5.2 Summary and analysis of the interview results

As can be seen in *Table 2,* there are common themes that are brought up by many of the interviewees. The visibility of network data was brought up by interviewees number 1, 6, 8 and 10. Closely related to the visibility issue is the data relevance and absence of relevant information of network traffic and devices within the network. These types of issues where brough up by interviewees number 2 and 7. Visibility regarding network monitoring is the gathering and analysis of data flowing inside the network as well as in and out of the network. The traffic flow within a network, such as server-to-server traffic is called east-west traffic. Issues regarding east-west or lateral traffic monitoring was brought up by interviewee number 6 and 8. Issues regarding visibility were brought up by over half of the interviewees (interviewee number 1, 2, 6, 7, 8, and 10). This problem was not only uttered by people from a specific team or department, which shows that the visibility problem has a widespread effect. This clearly indicates that solutions and action need to take place in order to tackle this in future. Good visibility of an enterprise network is vital in many ways. Network data visibility is a necessity in an environment where the number of cyber attacks is growing, as discussed in section 2.6.6.

The second most common problem that came up in the interviews dealt with real-time network data and information. The importance of real-time data is vital since it gives the monitoring and security teams a better chance to react to anomalies and threats faster. This problem was brought up by interviewees number 2 and 8. Since they have a different job description and different tasks this issue is also worth highlighting and acknowledging.

Issues regarding alert accuracy and alert handling are also high on the priority list. Alert accuracy is important and problems regarding this are a common phenomenon in the field of network monitoring. False positives and false negatives are a biproduct of more automated detection systems. This issue was brought up by one interviewee only (number 3). However, this does not mean the issue is irrelevant and alert accuracy is an important feature of any security monitoring system. UEBA can provide solutions to alert accuracy and confidence as mentioned in section 2.4.

Interviewee number 5 mentioned that there were still many manual tasks in the monitoring process such as feature selection. Implementation and utilization of DL can be applicable to this issue as described in section 3.3.

A couple of themes that are not directly solvable by AI were also put forward by the interviewees. A common theme that was mentioned in the interviews was the importance of finding solutions that are conducted in-house. This provides the company with more control over the system infrastructure and architecture. In the case of monitoring this enables the company to solve problems faster without waiting for vendors to give their insight or solutions. The management of the systems was mentioned by interviewee number 4. The further development of the monitoring and security system requires clear management and well defined goals in order to improve the system.

6. Presentation of solutions

A part of my job at If was to find and present AI-based solutions and technologies. This chapter strives to present solutions and present existing technological concepts for the problems brought up by the interviewees at If, described in Chapter 5. This chapter will first take a look at what needs to be taken into consideration when implementing an ML-based detection system and what design principles are necessary to accomplish this. Since one of the key aspects presented in the interviews was how to change the reactive measures to be more proactive, the planning of the system is important to highlight.

The next section is about intrusion detection systems (IDS) and intrusion prevention systems (IPD) and how they work in network monitoring and security. These are common monitoring systems that utilize AI. The concepts and technologies of IPD and IDS are shifting towards so called NDR systems. Therefore, the following section considers network detection and response systems and how they provide an AI-based approach to tackle many of the problems brought up in the discussions such as visibility, lateral movement, proactive response methods and alert handling, *see Table 2* and *Table 3*. This section also includes security orchestration, automation, and response technology and how these can further enhance security operations with AI.

6.1. Design principles for Machine Learning Anomaly Detection Systems

Analyzing data patterns in the network has become a necessity in business. A big part of the network monitoring process is anomaly detection. Anomaly detection is a task in data analytics that detects abnormal and irregular patterns in the data [57]. Deviations from normal network behavior and detection of them are important, since the detection of rare but crucial events enables the network administrator to act and react to potential attacks and hacking attempts. AI can be utilized in this process, as described in section 2.4.

Anomalies can be categorized in different ways, depending on the nature of the anomaly and on the context in which it is analyzed in [57]. A *point anomaly* is a particular data point, and/or instance that clearly deviates from the normal metric values of the dataset. *Contextual anomaly* is an anomaly due to the context and circumstances of the case. Certain times or seasons during the year, for instance, are expected to deviate from normal data behavior, but if this happens outside the expected timeframe it is considered a contextual anomaly. Lastly, a *collective anomaly* is a set of data instances that collectively deviate from the normal network behavior. In this example, a single abnormal value is not considered an anomaly [57].

Anomaly detection is still widely done with manual methods. One typical way in which networks are monitored is to have dashboards that are analyzed weekly. Network monitoring staff then look for spikes and dips in the traffic activity and investigate them if they are out of the ordinary. This approach is limited to the number of metrics set initially and is difficult to upscale [58]. While this approach can detect bigger anomalies, smaller deviations can easily go undetected. As many attacks today are more distributed and shorter lasting, the system may not even know what to look for. The inspection of dashboards, after the fact that the anomaly has occurred, is retro-active and delays reaction time for incidents significantly [58]. Another manually conducted anomaly detection technique used is setting thresholds for the established metrics. This method relies on generating alerts if the measurements of the metrics go outside the threshold. For this technique to work, one needs to set the thresholds right, which even in a smaller network is very complicated, as there could be thousands of metrics and parameters to review. In case the threshold is set too high or too low, the number of false positive alerts increases. Additionally, wrongly set threshold limits also cause anomalies to go undetected [58].

Some of the interviewees highlighted problems regarding real-time information and expressed future desire to introduce more automation in the anomaly detection process, *see Table 2* and *Table 3.* For an ML-based automated anomaly detection system to work, there are five key design aspects to consider in the planning process

[58]: (1) timeliness, (2) scale, (3) rate of change, (4) conciseness, and (5) definition of incident.

Timeliness tackles the question of how fast the anomaly needs to be detected. Real-time anomaly detection includes alerts of attacks, threats, and established harmful network behavior. Recognition of this aspect is important for companies, since this affects the selection of which type of ML algorithm should be implemented, and what task and purpose it is supposed to address.

Scale and the amount of data are also a big part of the detection system infrastructure and its implementation. The datasets in various monitoring tasks vary in size. Important here is to investigate if the system deals with large scale data or smaller sets of information [58]. Since ML and DL algorithms react differently to data size and whether the data is labeled or unlabeled, this affects the planning of the system.

Rate of change is the rate at which the data that is measured fluctuates. Depending on what metrics are being measured, the rate of change has implication for the ML or DL algorithm that is selected for the task. When network monitoring is conducted, the system may see changes frequently in the metrics, which means the algorithm needs to have adaptive features in order to work efficiently [58].

Conciseness is the concept of taking multiple metrics and parameters into account when detecting anomalies. This idea helps the users and the system to gain a holistic view and understanding of the root cause of the anomaly. Investigation into one metric does not often show the whole picture, i.e. updates of certain parts of the system may cause latency in another part of the system. There are three ways to approach the conciseness of an ML anomaly detection system. Univariate anomaly detection takes each metric in the system and creates a map of normal behavior. This method is easy to scale up, since it handles each instance individually. However, root cause analysis is not feasible with this method, and one single unexpected incident that affects many metrics at the same time yields a high number of anomalies. Multivariate anomaly detection takes multiple inputs and analyses them as a group. Here we group such metrics together as form a certain model for a possible incident. The drawback of this is that the anomaly output does not give any indication of which parameter yielded an anomaly. Additionally, the input signal types must be similar for the system to have the necessary calculation capabilities. A hybrid approach of these two methods takes a univariate single metric technique, but instead of grouping many of them in a black box it analyses how anomalies are linked together. This gives better results when conducting root cause analysis [58].

The final design principle to contemplate is *definition of incident.* Since a fully automated intrusion or anomaly detection system is not yet possible, defining what constitutes an incident is still required [58]. For an incident to be well defined all, or at least most of, the causes for anomalies must be presented. This works for a system with few parameters and a finite number of metrics which, in most cases, is not the reality. In this phase of the system planning, ML and DL principles of supervised, unsupervised, and semi-supervised learning need to be reviewed. In a situation where many metrics are well defined and the system intent is to output classification or regression, a supervised algorithm is applicable. These algorithms require labeled training data to establish a normal behavior baseline. Because of the training model for supervised learning-based anomaly detection, this method is not that well suited for the task of detecting novel intrusions and anomalies. Since all possible incidents are often impossible to predefine, unsupervised learning is a viable approach. With an unsupervised algorithm, the system learns what normal system behavior is over time and anomalies are detected whenever the analyzed data deviates from the trained model. The training dataset is unlabeled, which means the system can detect a new incident whether it is known or unknown. The success of an unsupervised learning algorithm is dependent on how accurately and well normal system behavior is defined. This puts great emphasis on the training stage, since this creates the threshold values and foundation for the algorithm to work on. These two methods are on opposite sides of the spectrum and a combination of them yields a well-performing technique. Semi-supervised learning accomplishes this. Here the algorithm uses a small number of labeled data for training to learn the input structure. In most cases, a mixture of both supervised and unsupervised techniques gives a better result in anomaly and intrusion detection.

As discussed in the previous section, the concept of normal or abnormal behavior is fundamental to the automatization of intrusion and anomaly. The question then becomes, how do we assess or confirm normal behavior? As a general framework for anomaly detection, *see Fig. 10,* the process is straightforward and easy to understand.

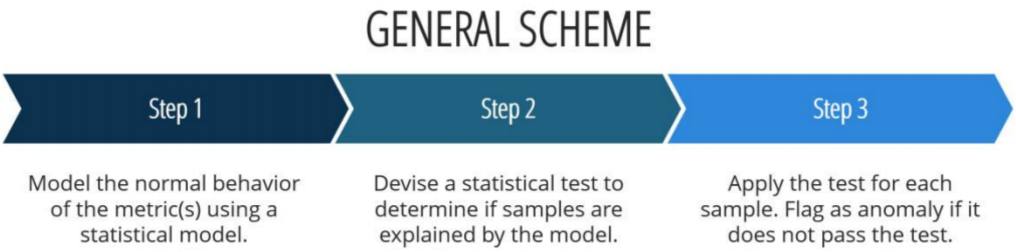
GENERAL SCHEME
Step 1
Step 2
Step 3
Model the normal behavior of the metric(s) using a statistical model.
Devise a statistical test to determine if samples are explained by the model.
Apply the test for each sample. Flag as anomaly if it does not pass the test.

Figure 10: Anomaly detection framework [59]

Input data in networks are often large scale and in addition to that, data can behave non-linearly and unexpectedly. Choosing a single technique when modeling normal behavior is not possible, and a combination of many different models is more reliable for accurate baseline creation [59]. Seasonality is an important factor to consider when establishing the baseline and not acknowledging this can lead to contextual anomalies, as mentioned earlier in this section. Secondly, as ML and DL algorithms take the input data and constantly update the new normal, anomalous datapoints should be assigned a lesser weight for them not to alter the baseline and therefore generate false results in the future [59].

A common issue in anomaly and intrusion detection is how the output is represented. When implementing a detection system, the significance of the anomalies needs to be represented in some way to see how much these deviate from the normal and how the system should address them. Generally, the anomalies can be labeled in two ways, by scoring and by binary measures [57]. In scoring based detection the detected anomaly is given a score based on how much it has deviated from the defined normal thresholds. The scores then rank the anomalies and further investigation can be conducted. The second method is binary in nature and simply classes an instance as either normal or anomalous [57]. The system should, in other words, have the capacity to evaluate anomaly and intrusion significance and react only when necessary. This decreases the number of unnecessary alerts and diminishes the manual labor network analysts need to do.

Related to the problem of having several alerts at the same time, is the matter of finding out how metrics are related to one another and how they could stem from the same root cause. Therefore, behavioral topology learning is an important factor when designing the automated system. One approach to find similarities in anomalies is to look at the abnormal instances and see if there is a causal link between them. One machine learning technique to find the causality is to apply clustering algorithms [60]. The cluster could then be represented with the two methods mentioned above, namely scoring and binary labelling. A prominent candidate for clustering large scale data is a latent Dirichlet allocation (LDA) algorithm. LDA is an ML technique used in topic modelling to discover underlying patterns in the data and how parts of the data are similar [61]. Most clustering algorithms classify a data point only as belonging to a specific class. The advantage of LDA is that it allows metrics to belong to two or more classes and thereby discovers similarities between them [60]. Huang et al. [61] proposed an intrusion detection method using an LDA approach. In this study all monitored events were assigned a vocabulary and a single event was regarded as a word in that vocabulary. A type of security incident is called a topic and each topic has a probability distribution connecting it to one or several words. Normal network operations are regarded as topics as well and a mixture of these topics is signed as a document. Data threats are then discovered by looking at the topic distribution. If a new set of events (document) has a high probabilistic distribution to security incident topics, this document is logged as a security breach and an alert is prompted.

The last design principle for automated anomaly detection systems is to consider the scalability of the system [60]. A solution for large-scale data to be more digestible is to split the data and metrics into smaller groups. Specific ML algorithms can then be assigned to each group, depending on what task is at hand. This decreases the time it takes for the algorithm to learn patterns and generate faster results.

6.2. Intrusion Detection System

One of the most prominent fields in cyber- and network security that has adopted ML algorithms is intrusion detection systems (IDS), this was presented thoroughly in section 3.2. In this section we discuss how an IDS works, and how IDS classification takes form.

An IDS is a security software that automatically alerts network administrators of malicious activity, system compromise, and security policy violations [62]. IDS's can either be a hardware solution or a software embedded in the firewall that monitors the network, identifying malicious network traffic. A report of malicious activity is often a combination of outputs from network resources, sent through a security information and event management system (SIEM). An IDS is one part of a system that incorporates firewalls and intrusion prevention systems (IPS).

There are, however, clear distinctions between these parts and how they work. An IPS is similar to an IDS but this system can terminate the connection of networks. IPS is an active inline system, often placed directly after the firewall, that prevents attacks from entering the system. An IDS, contrary to an IPS, is a passive system that monitors network packets and compares signature patterns to defined normal patterns, alerting the system of possible malicious activity [63].

Intrusion detection systems can be classified into two categories [62]. Network intrusion detection systems (NIDS) monitor network traffic and host-based intrusion detection systems (HIDS) monitor the operating system files. A NIDS works by placing sensors in many locations over the network, which then monitor the network traffic. HIDS is a host or device specific implementation that only monitors the traffic on the selected host or device [64]. Both NIDS and HIDS use two methods in order to detect malicious network traffic: signature and anomaly-based intrusion detection.

Signature based intrusion detection is based on detecting previously known threats. This method requires a list of pre-defined compromise indicators that it looks for when monitoring the traffic. Examples of indicators are byte sequences, file hashes, and email subject headings. As it monitors the packets, the system compares them with a database of threat indicators and flags the packets if they constitute suspicious network traffic [64]. The second method, called anomaly-based intrusion detection, utilizes machine learning techniques. Here the system uses ML algorithms to train the detection system to recognize normal behavior and accomplish a baseline. This method does not look for threats but instead compares network activity to the baseline. In other words, the system triggers an alert of any behavior that is not in alignment with the baseline. The benefit with this approach compared to signature-based techniques is that it can detect unknown threats [64].

IDS's and IPS's are widely used in most monitoring systems and infrastructures. There are, however, a couple of drawbacks that prevent certain aspects of the monitoring task. An IDS is primarily designed to monitor north-south network traffic. North-south traffic is a client-to-server form of traffic, referring to traffic flows into and out of the network or data center. In other words, an IDS can detect threats and anomalies that come from outside the corporation network. IDS's lack the capability to detect an attack happening inside the network [65]. The internal network traffic, such as server-to-server is called east-west and the lack of visibility into this is a problem. Furthermore, IDS methods rely on signature libraries which prevent them from detecting novel attack types. Automated analysis and investigations are not possible either and these require human interaction or collaboration with an IPS [65]. Like IDS products, intrusion prevention systems have a couple of disadvantages. Large scale data sets introduce a lot of unnecessary alerts and an IPS cannot distinguish useful signals from noise. IPS products use signature models to detect anomalies and threats, and constant updating of these is necessary for an IPS to work properly [66].

IDS and IPS methods are a necessary part of any security operation and cannot be replaced, but there are techniques that can tackle the problems mentioned before. A new wave in network security has emerged in the form of network detection and response systems (NDR). In the following sections we take a look at these systems and how they use machine learning to their advantage.

6.3. Network Detection & Response systems

A trending new field of study in network security and monitoring is network detection and response (NDR) systems and solutions. The lack of visibility in lateral movement, and detection of threats already existing within the network, cause unwanted data breaches and potentially severe consequences financially. These issues were also brought up in the interviews, *see Table 2*. Response time to data breaches is too slow as mentioned in section 2.5. Real-time event monitoring and a proactive manner of handling attacks are tasks for machine learning tools. In the interview with interviewee number 6 NDR technologies were brought up as possible solutions, *see Table 3.* Therefore, in this section we take a look at how NDR systems work and how they complement the already existing cyber security measures, and what role ML has in this.

6.4. Gartner's SOC Visibility Triad

The security operations center visibility triad was coined by research and advisory company Gartner Inc. in 2015. This concept consists of three parts, SIEM, EDR, and NDR, *see Fig 11,* and strives to shorten the response time for attacks, improve the visibility of network traffic, and detect east-west attack types [67].

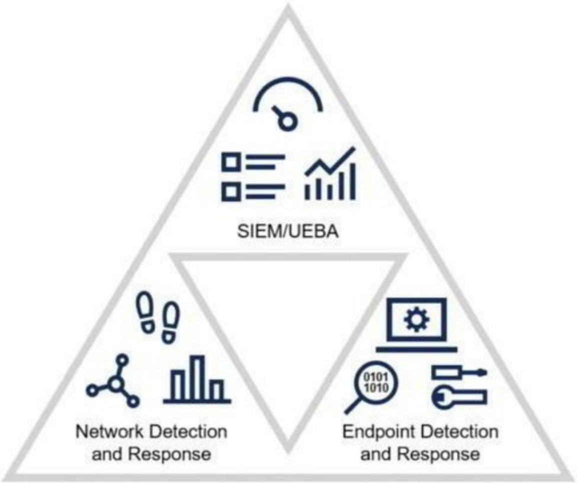
SIEM/UEBA
Network Detection and Response
Endpoint Detection and Response
0101
1010

Figure 11: SOC Visibility Triad [68]

6.4.1. SIEM

Security information and event management (SIEM) is a software tool to log and collect network data. This data can be gathered from applications, endpoints, cloud services and security devices and SIEM brings them together in a centralized format [68]. From this aggregated data, threats and events can be classified into categories and generate alerts based on predetermined rules. User and entity behavior (UEBA) technology is an AI approach and has been adopted into the SIEM methods. By using UEBA the task of establishing the baseline for the network is more automated [68].

6.4.2. EDR

The second component of the visibility triad is endpoint detection and response (EDR) technology. The task of this technology is to detect malicious data and traffic that happens on an endpoint, i.e. servers and laptops. As basic antivirus software works on the predefined rules for the system, EDR also uses behavioral analysis to detect malicious activity [68]. EDR is a way to augment SIEM and these two components have been for a long time the basis for security operations. While logs can be deleted by the attacker, leading to undetected threats using SIEM, endpoint systems can detect this activity and flag it as an anomaly [68]. EDR systems also allow security operations workers to isolate an infected endpoint from the network, preventing the attack from spreading laterally.

6.5. NDR

The latest addition to the triad is, as mentioned, an NDR system. NDR tools complement EDR and SIEM techniques by enabling detection without having rules or signatures that determine the system actions [69]. In addition to monitoring north-south traffic at network perimeters, NDR can track east-west communication with strategically placed sensors. NDR systems take advantage of AI methods to provide companies with better detection capabilities, and advanced NDR systems can detect novel unknown attacks. In addition, ML has enabled NDR systems to improve the confidence in determining risk levels of threats [97].

The complementary fashion of the triad is important to highlight. NDR does not simply replace SIEM and EDR systems but helps these technologies to function better and be more efficient. Endpoint detection systems are dependent on constant maintenance and management, which leads to worse visibility of the internal network. NDR uses behavioral analysis powered by ML and AI-based algorithms, which can detect attacks communicating and spreading between devices within the intranet. NDR real-time monitoring capabilities, and EDR's signature-based methods create a more robust security infrastructure [67]. NRD systems also use cloud-based ML techniques, which decreases the load on the system and offloads demanding modeling tasks. Additionally, machine learning methods used in NDR introduce the capability to automatically update detection models and signature libraries. The feature of isolating specific endpoints from the network provided by EDR systems can be improved with real-time network information provided by the NDR component [67].

SIEM systems are an efficient way of logging data and they are a very common approach in security operations among enterprises. For most cases, when the rules for malicious activity are well defined, SIEM can conduct early detection of threats. Data analyzed with SIEM systems cause many false alarms, which, on the other hand, prompts alert fatigue that is an undesired result when dealing with security systems [67]. As the amount of data is large in bigger corporations the logging of network traffic is turned off during certain times of the day or week. This introduces a window of opportunity for attackers to execute an attack and destroy or alter previous logs, which disrupts attack detection [67]. Log data also poses a problem when it comes to network activity and how this is reported. NDR systems use wire data, which is network packets showing the network communication and how it has moved within the network. Opposed to log data, wire data cannot be altered, and this provides the SIEM system with complete, trustworthy metadata to perform its tasks on [67].

In the next sections we shall present a couple of companies that provide NDR system solutions and how they take benefit from ML and AI techniques in their systems.

6.5.1. Darktrace

Darktrace is company that focuses on creating AI and ML driven technological security solutions. Darktrace has developed a patented ML algorithm security product called Enterprise Immune System. This system uses unsupervised learning methods to detect and defend against known and unknown malicious network traffic. This approach tackles the problem of having a rule-based system, as it can create models of normal and abnormal behavior by learning from the data [70]. An example case [71], showing the AI function in their product, demonstrates how it detects and responds to ransomware attacks in real time. In this case, the ransomware got into the network via corporate email, where a malicious Word document was opened by an employee. The employee's device started to connect to suspicious external domains and search for SMB shares. SMB stands for server message block which is a network protocol that allows devices within the same network to share files [96]. Furthermore, the ransomware started encrypting the SMB shares, which could impact the whole enterprise network and cause serious harm. This incident happened during off-hours, so no personnel was on premise to handle it. The Darktrace system detected this in nine seconds, and 24 seconds later the engine had blocked the encryption activities and neutralized the situation without any human interaction. The fast reaction time of the Darktrace system limited the damage to a small portion of the network. Ransomware is a problematic type of attack due to its unpredictable behavior in a network. With unsupervised learning algorithms the detection of these types of threats is more feasible and the Darktrace case shows it is possible.

6.5.2. Vectra

Another prominent vendor in the network detection and response field is Vectra AI Inc. Vectra AI is a company founded in 2012 which has recently done well in the cybersecurity market. Vectra AI develops NDR cybersecurity platforms that are driven by AI to detect attack behavior [72]. The solution proposed is called Vectra Cognito and like Darktrace's product, Vectra AI claims to have patented ML algorithm models that can learn system functions and behavior when detecting attacks. Another core feature of Vectra Cognito is to gather useful enriched metadata to get a clearer view of the system instead of using all the network data available. This function also reduces the noise in the data that SIEM and EDR systems work with.

6.6. SOAR technology

Security operations, automation, and response (SOAR) technology is a fairly new tool and implementation concept in cyber and network security. SOAR technology aims to tackle three key aspects of modern cybersecurity. These three functions are threat and vulnerability management (orchestration), automation of security operations (automation), and incident response (response) [75]. A SOAR system has the capability to take input from both internal and external sources and give a better visibility of network events and traffic flow. With the combination of SOAR and AI a system would be able to automatically respond to events instead of just recommending actions to the security operations team. The task of the orchestration functions is to combine the manual task and the automated steps of the information system and coordinate them [74]. Normally analysts have to switch between many different systems and cooperate with other team members when conducting the investigation and data collection concerning an incident or a threat. Orchestration improves the efficiency when remediating attacks [74,75]. The automation functions in SOAR technology use machine learning to automate parts of the investigation, which supports the orchestration process and reduces the time necessary to collect meaningful and relevant data [74].

7. Discussion

The advances in network and cybersecurity during the last decades have been significant and there are good examples of how AI has been adopted in this field. The findings reported in this thesis show that AI techniques are suitable for cybersecurity and network security. By applying AI into security and monitoring systems the efficiency of detecting anomalies and threats can be improved. This also decreases the amount of repetitive work security teams have to do on a daily basis. In order to adopt AI methods companies need to evaluate the current system and find what sort of problems they encounter. In this study there were clear themes that came up from multiple sources and most of these problems could be solved or mitigated with AI methods.

However, AI is a term that is used both as a marketing word and as a promising theory for a solution to all problems. There are certain artificial intelligence techniques that have shown promising results in security tasks, such as NDR, but these are mostly machine learning techniques that are not as complex as could be possible. For example, these systems are not intelligent in the sense that they would have the capability of being aware or have humanlike problem-solving knowledge. Most of the ML methods still depend on heavy human intervention and manual oversight. The supervised techniques are restricted to specific tasks in the security infrastructure and lack the self-learning ability that unsupervised learning methods provide. Unsupervised methods and deep learning approaches have met to a certain extent the expectations concerning a self-learning and decision-making machine and as mentioned, DL techniques have brought the field of AI to a whole new level.

Future research of AI in cybersecurity should further invest resources in experimenting with unsupervised techniques. Important to remember is that AI is still a tool for security teams and that AI in no way can replace the human deduction skill or problem-solving capabilities. Therefore, it is important to keep on developing AI algorithms that have an understanding of the context, since this can improve the proactive measures and automated features of security systems. Another recurring obstacle is the number of false alerts and false positives. It is still complicated to generate accurate warnings and to completely rely on the algorithm.

The complexity of networks and the huge amount of network traffic that companies encounter daily are contributing to this problem. A solution for this is to improve data collection and develop AI systems that can distinguish relevant metadata.

Additionally, AI is also used by malicious actors [18]. The AI techniques used by attacks are sometimes more sophisticated than the countermeasures and in the future the challenge will be to keep up with the novel methods attackers use. Most datasets used in training and testing AI systems are not up to date and do not include enough malicious datapoints. Therefore, datasets need to be updated so that the adopted AI systems can learn what normal activity really looks like and vice versa [18]. Protecting the integrity of AI algorithms is crucial as they can be hacked as well. The combination of a tampered data set and a hacked AI algorithm yields false results and leads to the security system working completely wrong [76].

The knowledge and knowhow in the field of cybersecurity are factors for advancement in this field. There is a shortage of skills in the field of cybersecurity and there is a high demand for people for these positions. In order to develop detection systems, improve AI algorithms, and come up with security measures, more people need to be educated and trained in cybersecurity [73]. This is linked to the interpretability of AI functions since analysts need to have an understanding of how the AI component of the system has come to its conclusion [76]. The development of better AI is also dependent on a fundamental understanding of existing AI algorithms, and a general lack of skill stalls the progress. An aspect to consider when implementing AI security systems is to make them easy to use and visually compelling for companies and enterprises to adopt them in a larger scale.

As mentioned, the term AI is used in a broad sense and there can be some ambiguity in what certain AI functions really do. There is a distinction between AI-assisted and AI-driven machines [76]. The main purpose of most AI solutions is to enable analysts to make better decisions, and a fully automated AI-driven solution is still an idea for the future. To avoid blind trust in vendor solutions, industry standards could be set to measure the autonomy of the system and to benchmark the AI systems [76]. Another challenge facing AI deployment is transparency in product testing. There are few industry frameworks and standards to rely on and this can cause and has already caused some human resistance to the paradigm shift to AI.

8. Conclusion

The idea of tackling future cyber attacks with AI solutions is a double-edged sword. As attackers and malicious actors constantly refine their attack techniques there is an urgency to react. On the other hand, there is a slight misconception about what AI actually can do for cybersecurity. AI solutions are marketed as the answer to most or all security problems, but the industry is not there yet. The majority of the existing IDS, NDR, and SOAR technologies still require much human intervention.

However, adoption of AI tools is the way forward as the traditional reactive, rule-based prevention measures are not sufficient to tackle the number of attacks. Cyber-attacks today can circumvent many older security mechanisms and go totally undetected, causing serious financial and reputational damage to enterprises and companies. AI, and specifically ML, provide detection and response systems with the capacity to be more proactive and form real-time actions. This also improves data collection and interpretation of network traffic, which advances security operations and the efficiency of the security teams. The AI techniques used today are limited to specific tasks in cyber and network security. In future there needs to be more research into how human interaction could be decreased and how the AI solutions could be more automated. To achieve this goal there need to be more specific tests, training datasets, and industry standards. ML and DL approaches also need to become better at understanding context in datasets in order to make humanlike decisions with a low rate of false results. Most of the material giving insight into AI products comes from vendors and this, of course, causes bias when gauging the functionality of the solution. In the near future AI solutions will be a valuable tool for security analysts as the tasks of data collection and root-cause investigation will become faster and more efficient. A fully automated AI-driven security infrastructure may never be possible and that is important to remember when discussing and reviewing the role of AI in cybersecurity in the coming years.

SVENSKT SAMMANDRAG

9. Artificiell intelligens i cybersäkerhet

Cyberhot och nätverksattacker är mer sofistikerade och mer oförutsägbara nuförtiden. Antalet dagliga attacker ökar globalt och sätten och teknikerna som används för att infiltrera företagssystem och personliga enheter är mångsidiga. Bakom attackerna är ensamma angripare, organiserade grupper och statsorganisationer. Resurserna som attackerarna har till förfogande har också ökat och idag kan cyberattacker ha allvarliga effekter och konsekvenser. De nya attackmetoderna har visat att våra nuvarande metoder inom cybersäkerhet inte längre klarar av denna utveckling, vilket leder till att vi måste hitta nya tekniker för att minimera cyberattackerna.

I denna avhandling eftersträvas att ge en översikt över hur artificiell intelligens (AI) och underdomäner som maskininlärning och djupinlärning kan tillämpas i cybersäkerhetsfrågor. Denna avhandling strävar också efter att ta upp befintliga AI-tekniker och hur de förbättrar såväl cybersäkerhet som nätverkssäkerhet. Med AI kan säkerhetssystemen bli mer proaktiva, bättre på att förutspå kommande attacker och upptäcka korrelationer i nätverksdata som människor inte kan. AI kan också ersätta vissa manuella repetitiva säkerhetsuppgifter, vilket ger mer tid till cybersäkerhetsanalytiker att fokusera på mer relevanta arbetsuppgifter.

Flera cybersäkerhetsmetoder som används idag är så kallade regelbaserade system. För att dessa system ska kunna upptäcka en attack eller ett hot måste man i förväg definiera vilken typs attribut systemet ska försöka hitta. Detta innebär att regler för systemet måste uppdateras för varje ny attack. I takt med att cyberkriminaliteten utvecklas uppdateras också attacktyperna konstant. Detta leder till mycket manuellt arbete för att upprätthålla säkerhetsinfrastrukturen och -systemen. Under det senaste året har cyberkriminalitet och den ekonomiska skada den åstadkommit globalt överskridit en biljon dollar [19]. De mest förekommande och mest skadliga cyberattacker som påträffas i dagens läge är skadlig kod, nätfiske och identitetsstöld. En oroväckande och växande trend inom dessa attacker är att de blir mer och mer riktade mot enskilda, privata personer.

Fördelarna med att ta AI-metoder i bruk är flera. Med hjälp av AI kan man snabbare göra beräkningar och prognoser på stora mängder av data. Man kan alltså lättare upptäcka datamönster och gömda korrelationer i nätverksflödet. Flera uppgifter som är repetitiva kan ersättas med AI-baserade algoritmer som både kan utföra uppgiften snabbare och dessutom utföra den mer korrekt. Maskininlärning (ML) har utnyttjats mycket inom cybersäkerhet och den har visat goda resultat i upptäckandet av attacker. Typiska maskininlärningsalgoritmer som testats inom cybersäkerhet och nätverkssäkerhet är neurala nätverk, stödvektormaskiner, beslutsträd och klusteranalys. Ifall attributen och reglerna som algoritmen jobbar med är väldefinierade har dessa tekniker en mycket hög precision när det kommer till att upptäcka skadlig nätverkstrafik. En vidare utveckling inom maskininlärning är djupinlärning, som är en metod där man med hjälp av algoritmer och komplexa neurala nätverk modellerar data och strävar efter att upptäcka abstraktioner ur data. Största skillnaden med maskininlärning och djupinlärning är att djupinlärning kan träna sig själv för att utföra en uppgift. Inom cybersäkerhet kommer detta till nytta eftersom attackmetoderna utvecklas och förändras konstant. Djupinlärningstekniker som testats inom cybersäkerhet är bland annat DBN-nätverk (*eng. deep belief networks*), CNN-nätverk (*eng. convolutional neural network*) och RNN-nätverk (*eng. recurrent neural network.*).

Den praktiska delen av denna avhandling är gjord i samarbete med If försäkringsbolag. Uppdraget gick ut på att intervjua anställda på If som jobbar på avdelningen för nätverksmonitorering och -säkerhet. Uppgiften var att se vilken typs problem som uppkommer, hur dessa skulle kunna lösas med AI-metoder och vilka lösningar det finns till förfogande på marknaden. De största problemen som framkommer är att Ifs system inte är tillräckligt proaktiva, det saknas bra visibilitet inom nätverksdata och att kapacitet inom NDR-system (*eng. network detection and responce*) fattas. Dessutom undersöktes vad som behöver tas i beaktande för att implementera ett automatiserat system som upptäcker avvikelser i ett nätverk och vilka principer som gäller för designen av det.

Designprinciperna för ett automatiskt system för upptäckande av avvikelser består av fem delar. Man bör definiera hur snabbt ett system ska reagera och hur stora mängder data delarna i systemet ska arbeta med. Dessa aspekter påverkar hurdana ML- och DL-algoritmer man ska välja för systemet. Dataflödet och dess innehåll kan fluktuera och detta innebär att systemet måste kunna anpassa sig till dessa ändringar. Dessutom måste man ta i beaktande vilka parametrar systemet monitorerar och vilken typs beteenden inom nätverkstrafiken som klassas som avvikelser eller onormala.

Typiska teknologier och system som används inom nätverksmonitorering och för att upptäcka hot är intrångsdetekteringssystem (*eng. intrusion detection system, IDS*) och intrångsskyddsystem (*eng. intrusion prevention system, IPS*). Dessa fungerar bra för att upptäcka attacker och intrång på nätverk så länge som det finns klara regler och signaturer systemet ska upptäcka. Detta kan skapa problem när det gäller att utveckla ett mer proaktivt system. Ytterligare kan IDS och IPS inte upptäcka cyberattacker som redan finns inne i det interna nätverket.

Ett relativt nytt teknologiskt koncept som ger lösningar på detektering av intrång, upptäckande av interna näthot och erbjuder bättre visibilitet på nätverksdata är NDR teknologi. NDR tillsammans med SIEM och EDR ger monitoreringssystem mer realtidsinformation och möjliggör snabbare respons till hot och attacker. I dessa system automatiseras samarbetet mellan NDR, SIEM och EDR med hjälp av AI-baserade algoritmer och det finns NDR-lösningar på marknaden som redan kunnat upptäcka allvarliga attacker och även eliminerat dem totalt.

Framtiden inom cybersäkerhet behöver AI-baserade lösningar men det finns en del problem som måste beaktas innan de kan nå sin fulla potential. Ett missförstånd med AI-lösningar är att de ska kunna ersätta människor totalt och att man ska kunna blint lita på AI-baserade säkerhetssystem. AI fungerar för tillfället som ett hjälpverktyg för cyberexperter så att de kan vara snabbare och effektivare i sitt arbete. Ett fullständigt automatiserat system är inte möjligt inom en snar framtid. För att uppnå en större grad av självständighet inom dessa system måste algoritmerna förses med bättre träningsdata för att automatisera inlärningsprocessen. Industristandarder och regleringar måste även införas för att kunna ge företag en klar uppfattning om vad kommande AI-system gör och för att kunna vidareutveckla dem.

BIBLIOGRAPHY

References:

1. P. H. Winston. (1992). *Artificial Intelligence.* Addison-Wesley Publishing Company. [Online] Available at: https://courses.csail.mit.edu/6.034f/ai3/rest.pdf
2. J. McCarthy. (1998). *What is Artificial Intelligence?.*. Stanford University. [Online] Available at: http://cogprints.org/412/2/whatisai.ps
3. J. Copeland. (1993). *Artificial Intelligence: A Philosophical Introduction.* Blackwell Publisher. [Online] Available at: https://books.google.fi/books?hl=fi&lr=&id=T05ICgAAQBAJ&oi=fnd&pg=PP1&dq=artificial+intelligence&ots=NVMoYtbXkF&sig=Ps6rxPvMRUuMI4cIx1rh4utmhNA&redir_esc=y#v=onepage&q=artificial%20intelligence&f=false
4. J. N. Kok, E. Boers et al. (2001). *Artificial Intelligence: Definitions, Trends, Techniques and Cases.* EOLSS. [Online] Available at: https://www.eolss.net/Sample-Chapters/C15/E6-44.pdf
5. A. Barr, E. A. Feigenbaum. (1982). *Handbook of Artificial Intelligence, Volume 2.* William Kaufmann, Inc. [Online] Available at: https://books.google.fi/books?hl=fi&lr=&id=xP7iBQAAQBAJ&oi=fnd&pg=PP1&dq=Barr,+A.%3B+Cohen,+P.+R.%3B+Feigenbaum,+E.+A.+(eds.)+1989.+Handbook+of+Artificial+Intelligence&ots=KbiUWZ0qWf&sig=Y2B00iQIexlscp3Q7jQZt1epWuU&redir_esc=y#v=onepage&q=Barr%2C%20A.%3B%20Cohen%2C%20P.%20R.%3B%20Feigenbaum%2C%20E.%20A.%20(eds.)%201989.%20Handbook%20of%20Artificial%20Intelligence&f=false
6. N. J. Nilsson. (1980). *Principles of Artificial Intelligence.* Stanford University. Morgan Kaufman Publishers, Inc. [Online] Available at: https://books.google.fi/books?hl=fi&lr=&id=mT-jBQAAQBAJ&oi=fnd&pg=PP1&dq=artificial+intelligence+&ots=hMWi8K2E9o&sig=vehuplMNMMcquD9cJzS6KvofDcc&redir_esc=y#v=onepage&q=artificial%20intelligence&f=false

7. M. Haenlein, A. Kaplan. (2019). *A Brief History of Artificial Intelligence: On the Past, Present, and Future of Artificial Intelligence.* Sage Journals. [Online] Available at: https://journals.sagepub.com/doi/10.1177/0008125619864925
8. Cybersecurity & Infrastructure Security Agency (CISA). (2019). *What is Cybersecurity.* [Online] Available at: https://us-cert.cisa.gov/ncas/tips/ST04-001
9. J. De Groot. (2020). *What is Cybersecurity? Definitions, Best Practices and More.* Digital Guardian. [Online] Available at: https://digitalguardian.com/blog/what-cyber-security
10. A. Henderson. (2019). *The CIA Triad: confidentiality, integrity, availability.* Panamore Institute. [Online] Available at: http://panmore.com/the-cia-triad-confidentiality-integrity-availability
11. Marin, G. A. (2005). *Network Security Basics. IEEE Security and Privacy Magazine, 3(6), 68–72.* doi:10.1109/msp.2005.153. [Online] Available at: https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=1556540
12. Swanson, G. (2014). *3 Application Lifecycle Phases You Must Security Test.* Trustwave. [Online] Available at: https://www.trustwave.com/en-us/resources/blogs/trustwave-blog/3-application-lifecycle-phases-you-must-security-test/
13. Singh, A., & Chatterjee, K. (2017). *Cloud security issues and challenges: A survey*. Journal of Network and Computer Applications, 79, 88–115. doi:10.1016/j.jnca.2016.11.027
14. McAfee. 2021. *What is Endpoint Security.* [Online] Available at: https://www.mcafee.com/enterprise/en-us/security-awareness/endpoint.html
15. Risk Based Security. (2020). *2019 Year End Report. Data Breach QuickView.* [Online] Available at: https://pages.riskbasedsecurity.com/hubfs/Reports/2019/2019%20Year%20End%20Data%20Breach%20QuickView%20Report.pdf
16. National Technology Security Coalition (NTSC). (2020). *Cybersecurity Report 2020.* [Online] Available at: https://www.ntsc.org/assets/pdfs/cyber-security-report-2020.pdf

17. Reed Smith. (2020). *Corona virus is now possibly the largest-ever security threat.* [Online] Available at: https://www.reedsmith.com/en/perspectives/2020/03/coronavirus-is-now-possibly-the-largest-ever-security-threat
18. Segal, E. (2020). *The Impact of AI on Cybersecurity.* IEEE Computer Society. [Online] Available at: https://www.computer.org/publications/tech-news/trends/the-impact-of-ai-on-cybersecurity
19. Smith, Z. M., Lostri, E. (2020). *The Hidden Costs of Cybercrime.* McAfee. [Online] Available at: https://www.mcafee.com/enterprise/en-us/assets/reports/rp-hidden-costs-of-cybercrime.pdf
20. ServReality. (2020). *Artificial Intelligence in Cybersecurity.* [Online] Available at: https://servreality.com/blog/artificial-intelligence-in-cybersecurity-pros-and-cons/
21. Brook, C., (2020). *What is User and Entity Behavior Analytics? A Definition of UEBA Benefits, how it Works, and More.* Data Insider. [Online] Available at: https://digitalguardian.com/blog/what-user-and-entity-behavior-analytics-definition-ueba-benefits-how-it-works-and-more
22. ENISA. (2020). *ENISA Threat Landscape 2020 – List of Top 15 Threats.* DOI: 10.2824/552242 [Online] Available at: https://www.enisa.europa.eu/publications/enisa-threat-landscape-2020-list-of-top-15-threats
23. Imperva. (2020). *Credential Stuffing.* [Online] Available at: https://www.imperva.com/learn/application-security/credential-stuffing/
24. ENISA. (2020). *ENISA Threat Landscape 2020 – Malware.* DOI: 10.2824/552242 [Online] Available at: https://www.enisa.europa.eu/publications/malware
25. WhatIsMyIPAddress. (2020). *MaaS Chaos. Malware-as-a-Service is Growing.* [Online] Available at: https://whatismyipaddress.com/maas
26. ENISA. (2020). *ENISA Threat Landscape 2020 – Web-based Attacks.* DOI: 10.2824/552242 [Online] Available at: https://www.enisa.europa.eu/publications/web-based-attacks

27. McAfee. (2013). *What is a "Drive-by" Download?* [Online] Available at: https://www.mcafee.com/blogs/consumer/drive-by-download/
28. ENISA. (2020). *ENISA Threat Landscape 2020 – Phishing.*
    DOI: 10.2824/552242
    [Online] Available at: https://www.enisa.europa.eu/publications/phishing
29. Jolera. (2020). *3 Ways AI Prevents Phishing Attacks.* [Online] Available at: https://www.jolera.com/3-ways-ai-prevents-phishing-attacks/
30. ENISA. (2020). *ENISA Threat Landscape 2020 – Data Breach.*
    DOI: 10.2824/552242
    [Online] Available at: https://www.enisa.europa.eu/publications/enisa-threat-landscape-2020-data-breach
31. IBM. (2020). *Cost of a Data Breach Report 2020.* [Online] Available at: https://www.ibm.com/security/data-breach
32. ENISA. (2020). *ENISA Threat Landscape 2020 – Web Application Attack.*
    DOI: 10.2824/552242
    [Online] Available at: https://www.enisa.europa.eu/publications/web-application-attacks
33. Imperva. (2020). *SQL (Structured query language).* [Online] Available at: https://www.imperva.com/learn/application-security/sql-injection-sqli/
34. Acunetix. (2019). *Cross-site Scripting (XSS).* [Online] Available at: https://www.acunetix.com/websitesecurity/cross-site-scripting/
35. CISA. (2009). *Understanding Denial-of-service attacks.* [Online] Available at: https://us-cert.cisa.gov/ncas/tips/ST04-015
36. ENISA. (2020). *ENISA Threat Landscape 2020 – Distributed denial of service.*
    DOI: 10.2824/552242
    [Online] Available at: https://www.enisa.europa.eu/publications/enisa-threat-landscape-2020-distributed-denial-of-service
37. Expert.ai. (2020). *What is Machine Learning?* [Online] Available at*:* https://www.expert.ai/blog/machine-learning-definition/
38. Li, J. (2018). *Cyber security meets artificial intelligence: a survey. Frontiers of Information Technology & Electronic Engineering, 19(12), 1462–1474.* doi:10.1631/fitee.1800573. [Online] Available at: https://link.springer.com/content/pdf/10.1631/FITEE.1800573.pdf

39. Xin, Y., Kong, L., Liu, Z., Chen, Y., Li, Y., Zhu, H., … Wang, C. (2018). *Machine Learning and Deep Learning Methods for Cybersecurity. IEEE Access, 6, 35365–35381.* doi:10.1109/access.2018.2836950. [Online] Available at: https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=8359287
40. Kumar, G., Kumar, K., & Sachdeva, M. (2010). *The use of artificial intelligence based techniques for intrusion detection: a review. Artificial Intelligence Review, 34(4), 369–387.* doi:10.1007/s10462-010-9179-5. [Online] Available at: https://link.springer.com/content/pdf/10.1007/s10462-010-9179-5.pdf
41. Pachghare, V. K., Kulkarni, P., & Nikam, D. M. (2009). *Intrusion Detection System using Self Organizing Maps. 2009 International Conference on Intelligent Agent & Multi-Agent Systems.* doi:10.1109/iama.2009.5228074. [Online] Available at: https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=5228074
42. Chen, W.-H., Hsu, S.-H., & Shen, H.-P. (2005). *Application of SVM and ANN for intrusion detection. Computers & Operations Research, 32(10), 2617–2634.* doi:10.1016/j.cor.2004.03.019. [Online] Available at: https://reader.elsevier.com/reader/sd/pii/S0305054804000711?token=158EC39D73E52E64E55AAE2A9702B8387B85ABDBAC0C991CE586E3841B1C8F2B98B4D672C9E7FE940B7A2310F240EF5B&originRegion=eu-west-1&originCreation=20210401140345
43. Hixon, R., & Gruenbacher, D. M. (n.d.). *Markov chains in network intrusion detection.* Proceedings from the Fifth Annual IEEE SMC Information Assurance Workshop, 2004. doi:10.1109/iaw.2004.1437849. [Online] Available at: https://ieeexplore.ieee.org/document/1437849
44. Sabaliauskas, D. (2020). *Hidden Markov Model (HMM).* Towards Science. [Online] Available at: https://towardsdatascience.com/hidden-markov-model-hmm-simple-explanation-in-high-level-b8722fa1a0d5
45. Mahoney, M.V., Chan, P.K. (2001) *PHAD: packet header anomaly detection for identifying hostile network traffic*. Department of Computer Sciences, Florida Institute of Technology, Melbourne, FL, USA, Technical Report CS-2001-4. [Online] Available at:

https://www.researchgate.net/publication/2834600_PHAD_Packet_Header_Anomaly_Detection_for_Identifying_Hostile_Network_Traffic

46. Vuong, T. P., Loukas, G., Gan, D., & Bezemskij, A. (2015). *Decision tree-based detection of denial of service and command injection attacks on robotic vehicles. 2015 IEEE International Workshop on Information Forensics and Security (WIFS).* doi:10.1109/wifs.2015.7368559. [Online] Available at: https://ieeexplore.ieee.org/document/7368559
47. Moon, D., Im, H., Kim, I., & Park, J. H. (2015). *DTB-IDS: an intrusion detection system based on decision tree using behavior analysis for preventing APT attacks. The Journal of Supercomputing, 73(7), 2881–2895.* doi:10.1007/s11227-015-1604-8. [Online] Available at: https://link.springer.com/content/pdf/10.1007/s11227-015-1604-8.pdf
48. CrowdStrike. (2021). *Advanced Persistent Threat Definition.* [Online] Available at: https://www.crowdstrike.com/cybersecurity-101/advanced-persistent-threat-apt/
49. Liao, Y., & Vemuri, V. R. (2002). *Use of K-Nearest Neighbor classifier for intrusion detection. Computers & Security, 21(5), 439–448.* doi:10.1016/s0167-4048(02)00514-x. [Online] Available at: https://www.sciencedirect.com/science/article/pii/S016740480200514X
50. DeepLearning.TV. (2015). *Deep Belief Nets – Ep.7.* YouTube. [Online video] Available at: https://www.youtube.com/watch?v=E2Mt_7qked0
51. Ding, Y., Chen, S., & Xu, J. (2016). *Application of Deep Belief Networks for opcode based malware detection. 2016 International Joint Conference on Neural Networks (IJCNN).* doi:10.1109/ijcnn.2016.7727705. [Online] Available at: https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=7727705
52. Yin, C., Zhu, Y., Fei, J., & He, X. (2017). *A Deep Learning Approach for Intrusion Detection Using Recurrent Neural Networks. IEEE Access, 5, 21954–21961.* doi:10.1109/access.2017.2762418 https://www.researchgate.net/publication/320366926_A_Deep_Learning_Approach_for_Intrusion_Detection_Using_Recurrent_Neural_Networks
53. Bansari, S. (2019). *Introduction to how CNNs work.* Data Driven Investor. [Online] Available at:

https://medium.datadriveninvestor.com/introduction-to-how-cnns-work-77e0e4cde99b

54. B. Kolosnjaji, G. Eraisha, G. Webster, A. Zarras, and C. Eckert, ‘‘Empow- ering convolutional networks for malware classification and analysis,’’ in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, 2017, pp. 3838–3845.
55. Mujtaba, H. (2020). *Introduction to Autoencoders.* Great Learning. [Online] Available at: https://www.mygreatlearning.com/blog/autoencoder/
56. Y. Yu, J. Long, and Z. Cai, ‘‘Network intrusion detection through stacking dilated convolutional autoencoders,’’ *Secur. Commun. Netw.*, vol. 2, no. 3, pp. 1–10, 2017.
57. Ahmed, M., Naser Mahmood, A., & Hu, J. (2016). *A survey of network anomaly detection techniques. Journal of Network and Computer Applications, 60, 19–31.* doi:10.1016/j.jnca.2015.11.016. [Online] Available at: https://reader.elsevier.com/reader/sd/pii/S1084804515002891?token=5ABFFC26C0D944A7A72B79311B33E0D4EC9583EB0EA3C51C4F1EC63BC6EDED41554F33157D0F012236F210B6D3282B36&originRegion=eu-west-1&originCreation=20210418084303
58. Anodot. (2019). *Ultimate Guide to Build a Machine Learning Anomaly Detection System, Part 1: Design Principals.* [Online] Available at: https://mail.google.com/mail/u/0?ui=2&ik=aef918248f&attid=0.7&permmsgid=msg-f:1697108300793539028&th=178d5702062a69d4&view=att&disp=safe
59. Anodot. (2019). *Build a Large-Scale Machine Learning Anomaly Detection System, Part 2: Learning the Normal Behavior of Time Series Data.* [Online] Available at: https://mail.google.com/mail/u/0?ui=2&ik=aef918248f&attid=0.8&permmsgid=msg-f:1697108300793539028&th=178d5702062a69d4&view=att&disp=safe
60. Anodot. (2019). *Ultimate Guide to Build a Machine Learning Anomaly Detection System, Part 3: Correlating Abnormal Behavior.* [Online] Available at:

https://mail.google.com/mail/u/0?ui=2&ik=aef918248f&attid=0.6&permmsgid=msg-f:1697108300793539028&th=178d5702062a69d4&view=att&disp=safe

61. Huang, J., Kalbarczyk, Z., & Nicol, D. M. (2014). *Knowledge Discovery from Big Data for Intrusion Detection Using LDA. 2014 IEEE International Congress on Big Data.* doi:10.1109/bigdata.congress.2014.111
62. Amrollahi M., Hadayeghparast S., Karimipour H., Derakhshan F., Srivastava G. (2020) *Enhancing Network Security Via Machine Learning: Opportunities and Challenges. In: Choo KK., Dehghantanha A. (eds) Handbook of Big Data Privacy*. Springer, Cham. https://doi.org/10.1007/978-3-030-38557-6_8
63. Bhardwaj R., (2018). *IDS vs IPS vs Firewall.* IPWITHEASE. [Online] Available at: https://ipwithease.com/firewall-vs-ips-vs-ids/
64. N-Able. (2021). *Intrusion Detection System (IDS): Signature vs Anomaly-Based.* [Online] Available at: https://www.n-able.com/blog/intrusion-detection-system
65. Snyder C., (2019). *Exploring the Frontier of Enterprise Security.* ExtraHop [Online] Available at: https://www.extrahop.com/company/blog/2019/network-detection-response-vs-intrusion-detection-systems/
66. Snyder C., (2019). *Of Fables and False Alters.* ExtraHop. [Online] Available at: https://www.extrahop.com/company/blog/2019/network-detection-response-vs-intrusion-prevention-systems/
67. Norris D., (2021). *NDR and the SOC Visibility Triad.* ExtraHop. [Online] Available at: https://www.extrahop.com/company/blog/2019/ndr-and-the-soc-visibility-triad/
68. Oakley C., (2020). *The SOC Visibility Triad-SIEM, EDR & NDR.* Nettitude. [Online] Available at: https://blog.nettitude.com/the-soc-visibility-triad
69. IronNet Cybersecurity Inc. (2020). *Dynamic detection for dynamic threats.* [Online] Available at: https://f.hubspotusercontent20.net/hubfs/6306975/IronNet-NDR-eBook-Dynamic-detection-for-dynamic-

threats.pdf?__hstc=173192060.2619f70fa6d4c2eb17336ef596a40042.1619086119807.1619086119807.1619172611329.2&__hssc=173192060.1.1619172611329&__hsfp=1686731293&hsutk=2619f70fa6d4c2eb17336ef596a40042&contentType=standard-page

70. Darktrace. (2020). *Darktrace Immune System. Self-learning Detection & Response.* [Online] Available at: https://www.darktrace.com/en/resources/wp-platform.pdf
71. Darktrace. *Detecting and Fighting Ransomware in Real Time.* [Online] Available at: https://mail.google.com/mail/u/0?ui=2&ik=aef918248f&attid=0.2&permmsgid=msg-f:1697108300793539028&th=178d5702062a69d4&view=att&disp=safe
72. Vectra AI. *About Us.* [Online] Available at: https://www.vectra.ai/about/company
73. Vectra AI. (2017). *Minding the cybersecurity gap.* [Online] Available at: https://mail.google.com/mail/u/0?ui=2&ik=aef918248f&attid=0.4&permmsgid=msg-f:1697108300793539028&th=178d5702062a69d4&view=att&disp=safe
74. De Bari Y., (2019). *The Future of Tomorrow: Automation of Cybersecurity.* Infosys. [Online] Available at: https://www.infosys.com/about/knowledge-institute/insights/documents/future-tomorrow.pdf
75. FireEye. (2019). *What is SOAR. Definition and Benefits.* [Online] Available at: https://www.fireeye.com/products/helix/what-is-soar.html
76. Donegan P., (2019). *AI in Cybersecurity: Filtering out the Noise.* [Online] Available at: https://www.hardenstance.com/wp-content/uploads/2019/02/HardenStance-White-Paper-on-AI-Final-Version.pdf
77. Hackmageddon. (2021). *2020 Cyber Attacks Statistics.* [Online] Available at: https://www.hackmageddon.com/2021/01/13/2020-cyber-attacks-statistics/
78. Bambrick N., (2020). *Support Vector Machines.* [Online] Available at: https://www.kdnuggets.com/2016/07/support-vector-machines-simple-explanation.html

79. Wikipedia. *Restricted Boltzmann machine.* [Online] Available at: https://en.wikipedia.org/wiki/Restricted_Boltzmann_machine

80. IBM. (2020). Recurrent Neural Networks. [Online] Available at: https://www.ibm.com/cloud/learn/recurrent-neural-networks

81. Ibragimov, B., & Xing, L. (2017). Segmentation of organs-at-risks in head and neck CT images using convolutional neural networks. Medical Physics, 44(2), 547–557. doi:10.1002/mp.12045. [Online] Available at: https://www.researchgate.net/publication/311564222_Segmentation_of_organs-at-risks_in_head_and_neck_CT_images_using_convolutional_neural_networks/figures?lo=1

82. Mujtaba H., (2020). *Introduction to Autoencoders.* My Great Learning. [Online] Available at: https://www.mygreatlearning.com/blog/autoencoder/

83. Shashanka, M., Shen, M.-Y., & Wang, J. (2016). *User and entity behavior analytics for enterprise security. 2016 IEEE International Conference on Big Data (Big Data).* doi:10.1109/bigdata.2016.7840805. [Online] Available at: https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=7840805

84. Mohri, M., Rostamizadeh, A. (2018). *Foundation of Machine Learning, second edition.* Masachusetts Institute of Technology. [Online] Available at: https://books.google.fi/books?hl=fi&lr=&id=dWB9DwAAQBAJ&oi=fnd&pg=PR5&dq=machine+learning+&ots=AypPXOrZk6&sig=CwS1EGYyGDO1bzSugHqeSlquC1o&redir_esc=y#v=onepage&q=machine%20learning&f=false

85. Quantdare. (2019). *What is the Difference between feature extraction and feature selection?.* [Online] Available at: https://quantdare.com/what-is-the-difference-between-feature-extraction-and-feature-selection/

86. Roy, R. (2020). *AI, ML, and DL: How not get them mixed!.* Towards Data Science. [Online] Available at: https://towardsdatascience.com/understanding-the-difference-between-ai-ml-and-dl-cceb63252a6c

87. Yegnanarayana, B. (2006). *Artificial Neural Networks.* Prentice-Hall of India Private Limited. [Online] Available at: https://books.google.fi/books?hl=fi&lr=&id=RTtvUVU_xL4C&oi=fnd&pg=PR9&dq=neural+networks&ots=Gd8YyjzIPy&sig=8KComDmLKsstPpie7RBhHRbtCdE&redir_esc=y#v=onepage&q=neural%20networks&f=false

88. Hsu, H. (2020). *How Do Neural Network Systems Work?.* Medium. [Online] Available at: https://medium.com/chmcore/how-do-neural-network-systems-work-dbe1bc0c4226

89. Nielsen, M. (2019). *Using neural nets to recognize handwritten digits.* Neural Networks and Deep Learning. [Online] Available at: http://neuralnetworksanddeeplearning.com/chap1.html

90. Elisha, O. (2020). *Transfer function for machine learning, simplified.* Heartbeat. [Online] Available at: https://heartbeat.fritz.ai/transfer-functions-for-machine-learning-simplified-eff2fddd133b

91. Bol, L., Hacker, D. J., Mattarella-Micke, A., Beilock, S. L., Seel, N. M., Rosenstand, C. A. F., … Ell, S. (2012). Competitive Learning. Encyclopedia of the Sciences of Learning, 671–677. doi:10.1007/978-1-4419-1428-6_175. [Online] Available at: https://link.springer.com/referenceworkentry/10.1007%2F978-1-4419-1428-6_175

92. Wu, D., Ren, J., & Sheng, L. (2015). *Representative points clustering algorithm based on density factor and relevant degree. International Journal of Machine Learning and Cybernetics, 8(2), 641–649.* doi:10.1007/s13042-015-0451-5. [Online] Available at: https://www.researchgate.net/publication/284233083_Representative_points_clustering_algorithm_based_on_density_factor_and_relevant_degree

93. Kuper, A. Reeves, S. & Levinson, W. (2021). *Qualitative research: An Introduction to Reading and Appraising Qualitative Research.* [Online] Available at: https://www.jstor.org/stable/pdf/20510591.pdf?refreqid=excelsior%3A199b81cf37f67d16411675a370b14a20

94. Fossey, E., Harvey, C., Mcdermott, F., & Davidson, L. (2002). *Understanding and Evaluating Qualitative Research. Australian*

& New Zealand Journal of Psychiatry, 36(6), 717–732. doi:10.1046/j.1440-1614.2002.01100.x [Online] Available at: https://journals.sagepub.com/doi/full/10.1046/j.1440-1614.2002.01100.x

95. Canary, A. (2019). *How to Analyze Interview Transcript in Qualitative Research.* [Online] Available at: https://www.rev.com/blog/analyze-interview-transcripts-in-qualitative-research
96. TechTerms. (2021). *SMB.* [Online] Available at: https://techterms.com/definition/smb
97. IronNet. (2021) *What is Network Detection and Response.* [Online] Available at: https://www.ironnet.com/what-is-network-detection-and-response#how-did-ndr-evolve
98. Upadhya, V., & Sastry, P. S. (2019). *An Overview of Restricted Boltzmann Machines. Journal of the Indian Institute of Science.* doi:10.1007/s41745-019-0102-z.