Anton Kelo

DATA SECURITY AND AUTOMATIC THREAT DETECTION

JYVÄSKYLÄN YLIOPISTO
INFORMAATIOTEKNOLOGIAN TIEDEKUNTA
2024

TIIVISTELMÄ

Kelo, Anton
Dataturvallisuus ja automaattinen uhkahavainnointi
Jyväskylä: Jyväskylän yliopisto, 2024, 26 s.
Tietojärjestelmätiede, kandidaatintutkielma
Ohjaaja: Vuorinen, Jukka

Teknologian nopeassa kehityksessä turvallisuus ja sääntely kehittyvät jäljessä, erityisesti tieto-, sekä dataturvallisuuteen liittyvät ongelmat ovat puhuttaneet teknologia-alalla viimevuosien aikana. Tämä tutkimus on tärkeä teknologisen kehityksen havainnollistamiseksi teknologian hallintaan ja turvallisuuteen liittyvien näkökulmien kannalta, jotka ovat tärkeitä alan tutkimuksessa ja kehityksessä. Tutkimus tarkasteli dataturvallisuuden kehitystä erityisesti kybermaailman turvallisuuteen liittyvien elementtien yhteydessä kirjallisuuskatsauksen keinoin, käyttäen lähteinä tutkimuksia, kirjoja, sekä yksittäisiä alan kattotason organisaatioiden internetjulkaisuja. Tutkimus keskittyi kyberturvallisuuden elementeistä erityisesti uhkien havainnointiin liittyvien järjestelmien dataturvallisuuteen, sekä tarkasteli niiden kykyä toteuttaa dataturvallisuuden vaatimuksia. Dataturvallisuutta tarkisteltiin alalla yleisesti käytetyn mallin, CIA-triadin avulla, sen osa-alueita ovat luottamuksellisuus, eheys, ja saavutettavuus. Data- ja kyberturvallisuuden antaminen tekoälysovellusten käsiin on herättänyt keskustelua alalla. Tutkimus keskittyi myös tarkastelemaan, kuinka luotettavana havainnointijärjestelmiä voi pitää koostamalla tietoja järjestelmien suorituskyvyistä. Tutkimus tuo kyberturvallisuutta ja uhkahavainnointia lähemmäs lukijan ymmärrystä, jonka avulla lukijan on helpompi ymmärtää ja nähdä automaattisen uhkahavainnoinnin käyttötapauksia tai turvallisuutta. Tutkimus osoitti sen, että koneoppimisen tarjoamat mahdollisuudet säilyttävät dataturvallisuuden hyvin, tuottaen samalla erinomaisia tuloksia uhkien havainnoinnissa.

Asiasanat: Koneoppiminen, CIA-triad, kyberturvallisuus, uhkahavainnointi

ABSTRACT

Kelo, Anton
Data Security and Automated Threat Detection
University of Jyvaskyla, 2024, 26 pp.
Information Systems, bachelor's thesis
Supervisor: Vuorinen, Jukka

With the rapid development of technology, security and regulation are lagging, and in particular information and data security issues have been a major concern for the technology sector in recent years. This study is important to illustrate the technological developments in terms of technology governance and security aspects, which are important for research and development in the field. The study examined the evolution of data security, specifically in the context of the security elements of the cyber world, through a literature review, using studies, books, and individual Internet publications of umbrella organizations in the field as sources. Among the elements of cybersecurity, the study focused specifically on the data security of threat detection systems and examined their ability to implement data security requirements. Data security was reviewed using a commonly used model in the industry, the CIA triad, with the components of confidentiality, integrity, and accessibility. Putting data and cybersecurity in the hands of AI applications has sparked debate in the industry. The study also focused on examining how reliable observation systems can be considered by compiling data on system performance. The study brings cybersecurity and threat intelligence closer to the reader's understanding, making it easier for the reader to understand and see the use cases or security of automated threat intelligence. The study showed that the capabilities offered by machine learning preserve data security well while delivering excellent results in threat detection.

Keywords: Machine Learning, CIA-triad, Cyber Security, threat detection

TABLES

TABLE 1, Sections of machine learning……...9

TABLE 2, Technology Characteristics ……..…….19

TABLE OF CONTENTS

TIIVISTELMÄ
ABSTRACT
TABLES

1 INTRODUCTION....6

2 CONTEMPORARY TRENDS OF MACHINE LEARNING AND CYBER SECURITY....8
2.1 Artificial intelligence and machine learning....8
2.2 Cyber security....10
2.3 CIA triad and AI in threat detection....10
2.3.1 CIA Triad....11
2.3.2 Automated threat detection....11

3 DATA SECURITY IN AUTOMATED THREAT DETECTION....13
3.1 Shallow machine learning....14
3.2 Machine learning combination techniques....15
3.3 Deep learning methods....16

4 DISCUSSION AND FINDINGS....18
4.1 Answers to research question....18
4.2 Additional notes and further research....20

5 CONCLUSION....21

SOURCES....23

1 INTRODUCTION

The development of technology has been fast for the past ten years. In recent years, there have been some notable events changing the path of the development of technology, for example, COVID-19. The COVID pandemic increased remote working and forced many organizations to adapt to new working environments (Leonardi, 2020; Kudyba, 2020). The release of ChatGPT and other large-language models as well as generative Artificial Intelligence (AI) tools in the year 2022 to the public brought AI to the mouths of the general population by bringing potential to transform industries (Dwivedi et al., 2019). AI has been researched for a long time before large language models were released, but in a way that does not appear in the everyday life of the public. For example, a certain sub-category of AI has been researched for use in cyber security threat detection for more than ten years. The interest in AI development and AI-driven business has increased in recent years since there are a lot of new possibilities to utilize. There are threat detection solutions to integrate into cyber security in the markets already, but leaving such a critical element as security to be handled by technology itself raises concerns because of possible vulnerabilities (Li, 2018). The number of internet-connected devices is expected to increase (Gandotra et al., 2017), and the risks of the prominent trends of cyber-attack automatization and sophistication are rising. This is the reason why there is a need for more capable and sophisticated threat detection systems. During the last few years, threats have been increasing as the absolute number of attacks increased (Albahar, 2019). Trends of the spreading, targeting, and sophistication of attacks can be seen (Sarker, 2023). More and exceedingly diversified devices are connected to the internet, for example, more mobile devices, fridges, televisions, and cars, which are called IoT devices (Oweis et al., 2016)

The main question this research studies and attempts to answer can be divided into two different questions:

- What are the main characteristics of developments in automated threat detection techniques?

- How do these techniques affect the data security of a system?

These questions aim to create a foundation for the research of the performance of technologies. The first question focuses on the data security and the system relationship, while the second question focuses on the developments of technologies. Additionally, these questions shed light on the field of study and governance of data security. These topics will be discussed briefly, however, these are not the main focus points of the paper.

This topic is important to study because the fast-moving development of advanced technology might feel overwhelming and difficult to grasp. This research provides a comprehensive look into the development of automatic intrusion detection in a readable and understandable format. This paper does not focus on the development or deployment of other AI-based technology than threat detection. This paper does not investigate any other major parts of cyber security than threat detection, for example, cryptography or authentication.

In the introduction, the study briefly introduces the research and investigates the need and foundation of this research. In this research, the second section investigates advanced technologies, such as automated threat detection and machine learning to provide readers with a good background of the technologies. The third section provides the reader with a historical view of the development of technologies investigated in this paper, including sources from the last decade. Section four is for findings and discussion about the development of these technologies, it summarizes the answers to the research questions. The final section is the conclusion, it includes the findings as a summary, as well as future research recommendations.

This research focuses on the development of ML and deep learning-based automated intrusion detection from the point of view of data security, data security is reviewed with the CIA triad, a model that is comprised of "Confidentiality, Integrity, Availability". This paper assigns a reader with a review of the development of threat detection systems, which provides the reader with a clear view of how the technology developed to this point and how efficient it is. Review source queries focused on IEEE Xplore, ACM digital library, and Springer. The main queries and keywords for database searches were different combinations of the following words: "cyber security", "threat detection" "CIA-triad", "intrusion detection" and "machine learning". The review also provides readers with a shallow understanding of machine learning methods, cyber security, intrusion detection, and data security as individual concepts. In this study, forty-eight different sources, including research papers, books, and certain internet publications from major organizations in the IT field were screened and used as a source for the study.

2 CONTEMPORARY TRENDS OF MACHINE LEARNING AND CYBER SECURITY

This section is an overall look into the key technologies to provide the reader with a better understanding of the concepts of this paper. The two most important high-level concepts are machine learning and cyber security. These concepts act as a top-level term and the focus of the paper is a more detailed view of certain technologies within these domains. After we have looked into top-level concepts, we will investigate technologies enabled by these, like automated threat detection. In this section, contemporary Machine Learning (ML) and Artificial Intelligence (AI) technologies will be investigated to understand the principles that are available for use in cyber security. This section also explores cyber security at a high level to understand what it includes and how it can be perceived, so more context can be added to the topic.

2.1 Artificial intelligence and machine learning

Machine learning is an important concept for this paper, it provides a broad selection of use cases that can be found useful for data security and seems to be beneficial for threat detection as well. There are quite a few definitions and approaches to artificial intelligence AI. Ertel (2018) provides readers with three broad-scale definitions for the term. To summarize these, AI can be defined as technology, that can solve complex problems and show human-like intelligence (Ertel, 2018). AI is a term that includes a lot of technologies, but in the context of this text, the most important one is Machine Learning (ML). ML is a term, that describes a selection of algorithms and technologies that provide behavior that is described in the definition of AI. According to Mohri et al. (2018), Machine learning is a set of computational methods, that can be used to do or enhance the performance of predictions, however, that is a broad definition. Machine learning is capable of tackling problems like text or document classification,

natural language processing, speech processing, computer vision, and many other challenges, like the ones with cyber security (Mohri et al., 2018). One interesting part of machine learning is that according to Mohri et al. (2018), the success of the algorithm depends on the data that is used in the training of the model. This is interesting, because now the data comes to be a crucial part of the system function, and this affects the way the model might perceive and analyze data.

To understand machine learning better, readers should understand what kinds of methods there are. According to Mahesh (2020), there are eight different sections of machine learning: supervised learning, unsupervised learning, semi-supervised learning, reinforcement learning, multi-task learning, ensemble learning, neural network, and instance-based learning. As shown in Table 1 (TABLE 1), these sections have different characteristics. These techniques include different kinds of algorithms that perform differently in different situations (Mahesh, 2020). Knowing previous "shallow learning" techniques is important when studying one of the most prominent machine learning models, deep learning. It can be argued that the development of deep learning is one of the most interesting developments in machine learning. According to Apruzzese et al. (2018), deep learning is essentially a neural network in multiple layers, also called a "multilayer neural network". According to IBM (2024), deep learning is especially useful because it can handle large amounts of data and does not need as much human intervention as machine learning does. Currently popular technology companies develop and use deep learning in their functions, for example, Nvidia and OpenAI. (Nvidia, 2024; OpenAI, 2024)

Supervised learning	Maps input to output based on example.
Unsupervised learning	No correct answer to input data, learn features from the data.
Semi-supervised learning	Combination of the supervised and unsupervised learning.
Reinforcement learning	Learns by gaining maximal reward signal from environment.
Multi-task learning	Aims to take advantage of similarities between tasks.
Ensemble learning	Aims to improve the performance of current model.
Neural network	Mimics human brain, neurons that aims to recognize relationships in dataset.
Instance based learning	Aims to be efficient in classification and regression.

TABLE 1, Sections of machine learning, (Mahesh 2020; Ernst & Louette 2024)

2.2 Cyber security

Cyber security is a major part of modern information technology. Major entities and organizations are taking measures to make their cyber security legislation and preparedness to the level, where it can keep up with rapidly evolving technology. For example, according to the European Commission (EU, 2024), the European Union has started to implement broad modernization of cyber security legislation and regulation. EU has implemented or is implementing four different acts regarding cyber security.

Defining cyber security can be tricky since it covers a broad collection of different devices and networks. According to the National Cyber Security Center of the UK (NCSC, 2024), the definition "Cyber security is how individuals and organizations reduce the risk of cyber-attack." However, there is no single and comprehensive definition for cyber security, for example, The US National Institute of Standards and Technology (NIST, 2024) gives six different definitions of cyber security. From these different definitions, we can find that there are multiple ways of defining cyber security. It emphasizes the fact that it includes many different approaches and views. According to Singer and Friedman (2014), it is difficult to define the concept in a single sentence, but it is security regarding cyberspace. From these views, it can be concluded that cyber security is security in different systems within cyberspace. We can say that users and humans are part of cyberspace, it is backed by Dalal and Rele (2018) who argue that cyber security is a domain where difference between irregularities is found, it needs human expertise alongside machine learning, and it an intersection of network-, computer-, and information security. In the context of this research, it is reasonable to focus on the systems working in the digital realm. One approach to cyber security is a data-centered focus that focuses on the information in the digital space. This approach is called "CIA-triad", and it is a de-facto guideline for data security.

2.3 CIA triad and AI in threat detection

The main concepts of this text mentioned previously are machine learning and cybersecurity. However, these technologies include a huge number of sub-technologies and applications. For example, ML can be seen as an enabler, from which new use cases can be derived. Cyber security is also an umbrella term, we will investigate the main concepts regarding the topic, for example, data security, threat, and intrusion detection, and CIA Triad (Confidentiality, Integrity, Availability).

2.3.1 CIA Triad

The point of view of this paper is data security. Data security comes down to the CIA triad model. Data security and the CIA triad is important in cyber security because it create a cornerstone for security that directs actions of organizations to objectively control the three aspects for optimal efficiency.

According to Samonas and Coss (2014), the CIA triad refers to the information systems security controls fundamental elements. CIA triad consists of three elements, Confidentiality, Integrity, and Availability. According to previous authors, confidentiality means the ability to keep the knowledge within the access of desired users and machines only. In the future where amounts of data increase and knowledge becomes more valuable, it is important to hold critical data confidential. In critical industries, like healthcare and the military, it is also extremely important to actively protect sensitive information. Integrity means "improper information modification or destruction". Integrity of data is important, because altering data could be altered without a specific person knowing about it, and working with improper data might lead to issues (Samonas & Coss, 2014). Availability means that the data is available when it is wanted to be. If data is not available because of a cyber-attack, it cannot be accessed and used for the occurring need (Samonas & Coss, 2014). These three elements can be extended with authentication and non-repudiation, which creates a more comprehensive setting (Chowhudry et al., 2023).

CIA triad has also been criticized for not necessarily being up to its time. For example, Covert et al. (2020) compared data privacy models by two organizations and researched an expanded model by ULD data protection authority that includes transparency, intervenability, and unlinkability. However, even though the CIA triad has been expanded and modified it sums up the holistic picture of information and data security. It creates a bedrock for this kind of paper. According to Kumar et al. (2014), it can be described as a fundamental concept.

2.3.2 Automated threat detection

In cyber security and data protection, it is optimal to detect and identify threats and intrusion before harm is done. Sewak et al. (2023) argue that modern threat detection systems are a combination of detection and response, and they provide a comprehensive approach to a broad scale of threats. Threat detection systems combine data from different intrusion detection systems (IDS) and endpoint detection & response systems (EDR). IDS consists for example of network-based IDS. EDR could consist of an endpoint protection platform which includes for example host-based intrusion detection and antivirus programs (Sewak et al., 2023). Especially EDR seems to be important from the point of view of the CIA triad, since the host-level of computing usually secures the outgoing traffic, hence affecting integrity and confidentiality.

Automated threat detection has progressed in a fast manner since the integration of machine learning algorithms, which offer the potential to act more proactively towards threats. However, challenges persist in cyber security, due to the ever-changing and dynamic nature of the field (Nitesh et al., 2023). Previously anti-virus programs and firewalls have been seen as sufficient measures to protect against threats, however, threats have changed, and new ways of protection are needed (Sewak et al., 2023). Firewalls and security policies are still important in today's cyberspace as well, but additional tools could be useful for a more comprehensive approach. According to Sornsuwit and Jaiyen (2019), new forms of attacks have surfaced, including ones using weaknesses in operating systems and settings in communication. According to Narayanan et al. (2018) a class of attacks, advanced persistent threats (APT) is one of the reasons why comprehensive threat detection is needed. APT attacks tend to be persistent and sophisticated. Advanced persistent threat creates a major threat to organizations because these attacks tend to focus on critical data like personal information, trade secrets, and other valuable data (Chen et al., 2014). When investigating the focus and nature of these attacks, APT attacks likely intend to compromise all aspects of the CIA triad. It can be argued that due to the rise in cyber-attacks, automated threat detection development is needed to battle this development. Dynamic and comprehensive threat detection is important from the perspective of the CIA triad for finding threats in all stages of the data processing cycle. While designing, planning, and implementing cyber security and threat detection processes, the CIA triad is a useful tool to remember what the main functions of the data are used.

3 DATA SECURITY IN AUTOMATED THREAT DETECTION

Machine learning and cyber security have been steadily gaining popularity as a subject of cyber security research since 2010. The focus and popularity of technologies have changed during the last 15 years, from machine learning technologies like SVM and Random Forest to combinations and deep learning (Shaukat et al., 2020). According to their study, until 2016 simple Machine Learning (ML) techniques were the most researched technology, but in 2016 Deep Learning (DL) started to rapidly gain popularity in cyber security applications. This section reviews the development of automated threat detection and its relationship with data security. This section will provide a set of examples of ML-based detection technologies, that might not themselves be whole threat detection systems but are possibly usable as a part of threat detection systems. The section will also investigate these examples from the point of view of the CIA triad after each technology section. The focus is especially on the attention to data security within these papers. It can be argued that these systems aim to increase the data security of the system, this study investigates if these tools and technologies can accomplish the confidentiality, integrity, and availability dimensions. Systems mentioned in this study are tested with different datasets. Popular datasets include NSL-KDD, spambase, and DARPA. These datasets feature different characteristics, and it might affect the functionality of the threat detection model (Shaukat et al., 2020). For this reason, the accuracy rate percentage which is used in this research to illustrate effectiveness of a system is not a strict measure of quality, rather more of a general illustration about approximate capabilities of the system. From the literature review, we can see that threat detection developed rapidly in the past 15 years, from simple machine learning techniques to complex, combined deep learning methods, with little to no negative effect on data security.

3.1 Shallow machine learning

In the beginning, security-focused systems focused on intrusion detection systems (IDS). In such a system, the approach is not as holistic as in a broader threat detection system and focuses on recognizing intrusions in the targeted system.

One of the first machine learning techniques used for threat detection was the naïve Bayes classifier and ID3 algorithm. Farid et al. (2010) used it to detect intrusions in the targeted system. With this approach, Farid et al. (2010) reached over 99% detection accuracy. However, false positives produced by the system should be reduced and that requires future research. Another machine learning technique, support vector machine (SVM) gained popularity rapidly in threat detection. SVM is solid for purposes of classification of cyber-attacks and selecting security features (Sarker 2023). Amer et al. (2013) researched a way with SVM to increase security in a system by comparing it with two other methods and using three different datasets. The system was found to be working and outperforming compared systems in two out of four datasets. This is done by giving a dataset to the SVM and making it find anomalies in the given dataset. This can be seen to investigate the integrity of the data in a sense, where the data is valid, but it may contain something that is not preferable.

Kevric et al. (2017) researched tree algorithms for intrusion detection. At this point, these detection systems already presented solid accuracy numbers, but in this case, the solution reached under 90% with testing datasets. However, the authors mention that with a certain dataset, accuracy can reach over 99%. It needs to be considered that these techniques are tested with predetermined datasets, and it does affect the results. The final traditional machine learning technique researched in this paper is by Zhou et al. (2020), where they achieved 99.89% accuracy for the IDS. In a decade of machine learning development, the robustness and accuracy of these systems were developed to provide over 99% accuracy for IDS.

From the nature of machine learning-based techniques, we can argue, that it is a significant development for confidentiality, integrity, and availability of the data. Shallow machine learning algorithms are relatively simple and do not have a notable impact on the surveillance data. These systems are based on an idea where system data is run through the detection system and anomalies are recognized. For that reason, the data stays intact and the found anomalies are reported to the user, however, these studies often mention that false positives are a reported problem with these systems. This is an issue regarding the integrity of the data if false assumptions from the data are made and the data is concluded into something, that is a false positive for the use of the user. Previously mentioned “shallow” machine learning methods might also affect the availability of the data. After implementation, the data might not be accessible before it goes to the model. It might affect the availability of the data for example in a situation, where the system is down due to a malfunction of the detection system.

3.2 Machine learning combination techniques

From the point where standalone ML techniques were used to make threat detection more efficient, it was noted that these techniques can be changed and evaluated for certain use cases. Sornsuwit and Jaiyen (2019) propose a method, where five different machine learning techniques are combined and evaluated, and one is chosen by Adaboost.M1 to evaluate, which technique is the best for the particular task. This is because different machine learning methods are more effective in certain types of intrusions. The data used in this example is a preprocessed dataset with multiple types of data. It seems that this system uses the gathered data only inside of the detection system and does not create have notable effect on the CIA triad aspects. However, it could have a significant boost in threat detection and hence might prove to be a useful tool for the protection of CIA triad aspects. The combinatory approach can also be used in different ways, Dalal and Rele (2018) presented a machine learning-based architecture model that is based on machine learning and requires the use of a sandbox environment. Authors list at least malware, watering hole, webshell, and spearfishing as possible threats that ML can help against. For example, malware attacks can affect all sides of the CIA triad and underscores that these systems can provide more help regarding the CIA triad than risks. On the other hand, for example, spear phishing does not necessarily pose a risk to the CIA triad.

Narayanan et al. (2018) found out that currently commercially available threat detection systems are usually noisy, difficult to use, and lack actionable details. The authors set up multiple agents producing different kinds of information and provide the data to the Cognitive Cyber Security (CCS) module that reasons with the imported data. The system was tested with custom-built ransomware that uses an SMB vulnerability. The test setting consisted of multiple phases and the system managed to detect the related events properly. It can be argued that for the importance of the CIA triad, these kinds of systems might prove to be more important. Systems that gather data from multiple agents and compile it in an understandable form. Even though this might boost the overall security of a system, it might also affect the integrity of data. A system that joins data from multiple domains might alter it in a way, that does not present complete data to a user.

In the 2020s it is starting to be obvious that threat mitigation and risk must be addressed by a portfolio of different tools. When working with multiple tools utilizing different technologies, it is important to have something to conclude the information. Kumar et al. (2021) have researched the use of monitoring platforms and its efficiency in improving cyber security, finding out that the semantic machine-learning technique is the most efficient solution. Compared to the semantic approach, the deep learning approach could perform

better, McElwee et al. (2017) reached a better accuracy of 98 %, four years earlier with the deep learning approach.

When investigating CIA triad aspects in combined machine learning techniques, the baseline is the same as with shallow techniques. It does have a positive impact on the overall security of a system as a tool. However, it does add extra steps to the data processing, that might have an impact on data. It is generally said that more complex systems are usually more vulnerable.

3.3 Deep learning methods

In the 2020s machine learning algorithms used for threat detection are still developing, but the basic, shallow machine learning algorithms have started to be established and tested. Nitesh et al. (2023) researched ten different machine learning methods and seven of the models reached over 98 % accuracy and precision. Even though there were good results for traditional machine learning models, the authors mention that applying deep learning methods holds great promise. It can be argued that while machine learning methods are established and easily provide accuracies of over 98%, deep learning methods that are still in development can provide accuracies of over 98%. DL does have the potential for even more, this is backed by McElwee et al. (2017), who reached 98 % accuracy.

Lately, different DL methods have been increasing in numbers with ML techniques. DL has shown promising results and could improve efficiency. The first DL technique investigated in this paper is the long, short-term memory (LSTM)-based approach focusing on anomalies by Kim et al. (2016). They had a good start for deep learning methods since they achieved a 99.8% detection rate for a 5.5% false alarm rate. Jiang et al. (2018) presented one of the first long, short-term memory recurrent neural networks (LSTM-RNNs) for threat detection purposes. In practice, LSTM-RNNs means a deep learning approach, Deep neural networks are state-of-the-art techniques for pattern recognition (Moosavi-Dezfooli et al., 2016). According to Jiang et al. (2018), it appears to be the first multi-channel approach based on deep learning. The model was compared with seven different machine-learning techniques for threat detection and it seemed to be the best with 98.94 % accuracy (Jiang et al., 2018).

There are different accuracies reached with different DL techniques and datasets. For example, Liu et al. (2020) reached only 95.4 % accuracy with a more traditional deep neural network. Ullah et al. (2019) researched how deep learning-based threat detection can be utilized on the Internet of Things (IoT) network. The authors researched a way to use deep learning in piracy- and malware detection and in that way reduce cyber threats in IoT networks. This approach mostly focused on network-level vulnerabilities and threat detection, instead of endpoint-level threats (Ullah et al., 2019). When compared with other state-of-the-art piracy detection techniques, the approach was the second most effective solution. Cosma and Joy (2011) reached better accuracy with the

semantic approach already in 2011, which means that the DL approach is not necessarily better in this use case.

The three previous DL approaches do not necessarily affect the CIA triad directly. The mentioned systems do improve data security by improving the overall level of cyber security in the target system. However, in DL approaches users need a significantly bigger amount of data to train the model, and extract features from this data, to make their conclusions about it (Karatas et al., 2018). This might affect the confidentiality and integrity of the data going through the systems because the DL system can learn from the user data it processes.

Over the past five years, we have found a rise in the number of deep learning solutions, which might create data security issues. Issues with deep learning data security arise not from the fact that it is efficient in threat detection, but from the fact that users can fool these systems. According to Moosavi-Dezfooli et al. (2016), a deep learning model can be fooled by an algorithm called DeepFool. This does not directly affect cyber security because data is not generated by attackers on all occasions, but by a machine and is interpreted by a threat detection system. However, this proves that deep learning methods might be crucially vulnerable to being fooled, and it could prove to be an issue for the data security of these systems. This is backed by Heaven (2019), in a study where it is stated that even the state-of-the-art deep neural networks can be fooled easily. For example, a hacker could use these weaknesses to hijack an AI-based online system. In addition to the possible weakness of fooling, Chowdhury et al. (2023) argued, that OpenAI ChatGPT compromises parts of the CIA triad by saving its user prompts with no plans to erase it, with a possibility of this data including sensitive information. This also underlines the responsibility of the organization and its security, not the technology used to conduct the business.

4 DISCUSSION AND FINDINGS

In this section, findings from previous sections will be discussed. It can be found that threat detection has developed quickly on par with the general development of technology. The result that can be found from this study can be divided into two different answers based on the questions mentioned in the introduction. Additionally, to the answers to the research question, this section provides additional notes that were found to be of importance within the field of research. These additional notes include findings about upcoming legislation that does have a substantial impact on cyber security and Machine Learning (ML) usage, as well as notes on the research field and its restrictions. This section also provides ideas and thoughts about further research.

4.1 Answers to research question

Technological development finding can be divided into three different main developments. Automated threat detection developed from simple machine learning techniques to deep learning and combining different techniques with different characteristics. Sarker (2023) reviewed different ML methods for different use cases and listed the following: classifying cyber-attacks, selecting security features, developing models, building models, reducing false alarm rate, network intrusion detection, detecting anomalies, detecting XSS attacks, detecting Denial of Service attacks, detecting botnets, predicting impacts, and distributed threat detection. As mentioned in section 2.3.2, a thread detection system is a combination of data by different systems and sensors, the mentioned use cases do fulfill the need for a threat detection system when the technologies are selected correctly. Traditional machine learning techniques tend to focus on intrusion detection and do not provide a comprehensive approach to threat detection, but it does perform well in anomaly detection and hence is a working technique for a part of threat detection. From “shallow” machine learning techniques the technology developed to DL techniques, partly

overlapping with the development of other technologies. This is backed by the review by Shaukat et al. (2020) where the authors researched the development of these technologies in the past decade. These answers come together when weighing different technologies for use as a part of a threat detection system. As mentioned earlier, threat detection systems are holistic systems and require correct tools for specific situations. Findings for the question of the main characteristics of development can be found in Table 2 (TABLE 2).

Data security-related finding is that machine learning-based capabilities are prominent and capable of threat detection. However, data security within the threat detection system, especially with deep learning methods might be questionable in some situations. The finding drawn from this paper when comparing the downsides and upsides of automatic threat detection is that automatic threat detection systems, especially IDS can improve the overall security of a system while not critically compromising data security. While these threat detection systems provide over 99% accuracy, they do not have a critical impact on the protection of the system data, even though in some cases DL models have been misused, the natural development of these technologies has been reliable. The answer to the research question of "How do these techniques affect the data security of a system" is that ML techniques do not generally hurt the data security of a system.

It is also important to acknowledge that ML systems do have downsides when ensuring security, for example, supervised learning-based algorithms cannot detect threats that are not thought to be within training data but can provide good data security for the detection tool itself. Shallow technologies reach a solid accuracy rate; however, deep learning is more flexible, but might also be more vulnerable. Which system provides better confidentiality, integrity, and availability in the end? Deep learning methods learn from unlabeled data, and thus it might be fooled while providing better overall security for the system. This has been proven with a couple of examples in previous sections, but how does it affect the security of a system? In a security system, the threat detection system is generally not accessible to be interacted with for the attacker. The main job of the security system is to recognize anomalies in the data produced by the system, from the input of the user. If deep learning systems can provide a solid detection rate with a robust design that cannot fooled, it can increase overall security.

Technology	**Detection rate**	**Weakness**	**Additional notes**
"Shallow" ML	>99%	Limited learning possibilities	Solid research background
Combined ML	>99%	Complexity	-
Deep Learning	>98%	Possibility for misuse	Need further research, prominent technology

TABLE 2, Technology Characteristics

4.2 Additional notes and further research

Outside of the main research questions, two different closely related notes came up, the complicated research of the field and the changing regulatory environment, after the literature review, notes about the research field came up as one of the main contributions of this paper. It is difficult to use the result data of previous research because test settings are different in most cases. Different datasets and test configurations output different kinds of results and these cannot be easily compared with each other, or the real world. These datasets are usually made up of real-world data but usually do not represent current internet traffic and new types of cyber-attacks. This creates a difficult set to review the efficiency of these systems in a real set. Reviewed threat detection systems are also a subject of research, vulnerabilities of these systems are not necessarily found yet. If vulnerabilities are recognized, these might not be mentioned because the test setting does not require a completely safe system. This creates limitations for the research of the field.

Another important note is changing the regulatory side of the field. These systems are part of a bigger system of systems. These systems are affected by legislative manners within EU borders, such as GDPR, NIS2, and CRA. GDPR defines the ways of gathering and using user data. This creates certain borderlines for the use of data gathered by a threat detection system and how it is used. For example, data gathered within the EU should not be stored outside the EU without European Commission approval for the country. GDPR also requires target organizations to follow integrity and confidentiality, as well as store the data in a way so it can be accessed and deleted when needed (Tietosuoja.fi, 2024). NIS2 focuses especially on critical actors, defining scope, risk management, reporting, and enforcement of cyber actions. With NIS2 as well, automated threat detection can be a strength for organizations, to conclude finding threats and reporting about findings automatically. CRA is an EU Cyber resilience act, it requires a certain level of security for a broader scope with devices with access to the internet. According to the European Commission (EU, 2023), these regulations are expected to enter in 2024. These regulations do have a notable impact on data management and planning on cyber actions and organization defenses, showing that the CIA triad is not yet obsolete.

Further research in the field is important for a better holistic picture of the research field and improvement of threat detection systems. Especially two important subjects, which are Advanced Persistent Threats (APT), and deep learning characteristics in threat detection. APT needs to be studied further because that type of attack is increasing in numbers and these types of attacks generally have a better probability of penetrating defensive systems. Deep learning model reliability improvement research has focused on computer vision and text processing, with at least eight different defensive strategies found (Alshemali & Kalita, 2020). Some of these strategies can be used in threat detection, but the subject requires additional research.

5 CONCLUSION

In conclusion, the rapid evolution of technology over the past decade has led to working environments shifting to more remote and technology intensive setting, this has affected cybersecurity and its developments of research. This study focused on investigating the developments of these advancements from the data security perspective, examining various Machine Learning (ML) and Deep Learning (DL) techniques used in threat detection, with a focus on the Confidentiality, Integrity, and Availability that comprises the CIA-triad. The introduction of large-language models to markets was a turning point, bringing AI available for anyone and highlighting the potential for cybersecurity. As organizations adapt to remote working environments and the increasing amount of Internet of Things (IoT) devices expands attack surfaces, the need for sophisticated threat detection mechanisms becomes increasingly important.

This study was conducted as a literature review. Forty-eight different sources were screened as material for this study. Sources include mostly peer-reviewed research articles. Some books, as well as internet publications by top technology companies, were used as a source for general technology knowledge. Material was sourced from different databases and search engines, for example, ACM, Google Scholar, and IEEE Xplore. Sources were used to gather information about the contemporary techniques for threat detection and understanding the data management and security of detection systems.

The need for this research comes from the fact that technology and its practical applications develop rapidly, while legislation and regulation strive to keep up with the phase. For the understanding of the development, it is necessary to screen the development and the contemporary technology details. With that understanding, professionals can utilize and implement technology that can prevent and optimize functions safeguarding the organization. The goal of this research was to find out how threat detection systems as a whole system can carry out the needs of data security. The topic does set some borderlines for the study, as the subject is complex and CIA-triad is abstract as a framework. The goals of this research were achieved by finding the current strengths and weaknesses in the data security of threat detection systems.

The answers to two different research questions can be divided into two answers. The first question about characteristics of technology developments of different ML and DL methods, it's evident that while traditional "shallow" ML techniques excel in certain aspects like anomaly detection, they may lack comprehensive coverage in threat identification. On the other hand, DL methods, with their ability to learn from unlabeled data, hold promise for the future but raise concerns regarding potential vulnerabilities to manipulation. The answer to the second question about data security is that new ML and DL technologies are generally capable of providing solid data security.

Furthermore, the study notes the necessary connection between technological advancements and regulations, such as GDPR, NIS2, and CRA, which shape data management practices and cybersecurity strategies. Compliance with these regulations not only mandates stringent measures for data protection but also underscores the relevance of the CIA triad in contemporary cybersecurity.

In navigating the complexities of automated threat detection, organizations must strike a balance between leveraging the latest technologies for enhanced security while ensuring the integrity confidentiality, and availability of their data. As the threat landscape continues to evolve with technology, further research into the robustness and resilience of ML and DL models is essential to fortify defenses against emerging cyber threats. In essence, while automated threat detection systems hold the potential for increasing cybersecurity posture, their deployment must be accompanied by rigid oversight and continuous evaluation to mitigate risks and hold the principles of the CIA triad. As we look towards the future, interdisciplinary collaboration and ongoing innovation will be key in safeguarding the cyber world against evolving threats. This requires future research in measures against advanced persistent threats and weaknesses of deep learning technologies.

SOURCES

Albahar, M. (2019). Cyber Attacks and Terrorism: A Twenty-First Century Conundrum. Science and Engineering Ethics, 1-14. https://doi.org/10.1007/s11948-016-9864-0.

Alshemali, B., & Kalita, J. (2020). Improving the reliability of deep neural networks in NLP: A review. *Knowledge-Based Systems*, *191*, 105210.

Amer, M., Goldstein, M., & Abdennadher, S. (2013, August). Enhancing one-class support vector machines for unsupervised anomaly detection. In *Proceedings of the ACM SIGKDD workshop on outlier detection and description* (pp. 8-15).

Apruzzese, G., Colajanni, M., Ferretti, L., Guido, A., & Marchetti, M. (2018, May). On the effectiveness of machine and deep learning for cyber security. In *2018 10th international conference on cyber Conflict (Cy Con)* (pp. 371-390). IEEE.

Cosma, G., & Joy, M. (2011). An approach to source-code plagiarism detection and investigation using latent semantic analysis. *IEEE transactions on computers*, *61*(3), 379-394.

Covert, Q., Steinhagen, D., Francis, M., & Streff, K. (2020). Towards a triad for data privacy.

Chen, P., Desmet, L., Huygens, C. (2014). A Study on Advanced Persistent Threats. In: De Decker, B., Zúquete, A. (eds) Communications and Multimedia Security. CMS 2014. Lecture Notes in Computer Science, vol 8735. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-662-44885-4_5

Chowdhury, M. M., Rifat, N., Ahsan, M., Latif, S., Gomes, R., & Rahman, M. S. (2023, May). ChatGPT: A threat against the CIA triad of cyber security. In *2023 IEEE International Conference on Electro Information Technology (eIT)* (pp. 1-6). IEEE.

Dalal, K. R., & Rele, M. (2018, October). Cyber Security: Threat Detection Model based on Machine learning Algorithm. In *2018 3rd International Conference on Communication and Electronics Systems (ICCES)* (pp. 239-243). IEEE.

Dwivedi, Y., Hughes, L., Ismagilova, E., Aarts, G., Coombs, C., Crick, T., Duan, Y., Dwivedi, R., Edwards, J., Eirug, A., Galanos, V., Ilavarasan, P., Janssen, M., Jones, P., Kar, A., Kizgin, H., Kronemann, B., Lal, B., Lucini, B., Medaglia, R., Meunier-FitzHugh, K., Meunier-FitzHugh, L., Misra, S., Mogaji, E., Sharma, S., Singh, J., Raghavan, V., Raman, R., Rana, N., Samothrakis, S., Spencer, J., Tamilmani, K., Tubadji, A., Walton, P., & Williams, M. (2019). Artificial Intelligence (AI): Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for

research, practice and policy. International Journal of Information Management. https://doi.org/10.1016/J.IJINFOMGT.2019.08.002.

Ernst, D., & Louette, A. (2024). Introduction to reinforcement learning.

Ertel, W. (2018). *Introduction to artificial intelligence*. Springer.

European Commission. (EU, 2024). Shaping Europe's Digital Future. https://digital-strategy.ec.europa.eu/en/policies/cybersecurity-policies

European Commission, EU Cyber Resilience Act. (2023). Shaping Europe's Digital Future. https://digital-strategy.ec.europa.eu/en/policies/cyber-resilience-act

Farid, D. M., Harbi, N., & Rahman, M. Z. (2010). Combining naive bayes and decision tree for adaptive intrusion detection. *arXiv preprint arXiv:1005.4496*.

Gandotra, P., Jha, R., & Jain, S. (2017). A survey on device-to-device (D2D) communication: Architecture and security issues. J. Netw. Comput. Appl., 78, 9-29. https://doi.org/10.1016/j.jnca.2016.11.002.

Heaven, D. (2019). Why deep-learning AIs are so easy to fool. *Nature*, *574*(7777), 163–166.

IBM, What is machine learning (ML)? | IBM. (Sourced 10.04.2024). https://www.ibm.com/topics/machine-learning

Jiang, F., Fu, Y., Gupta, B. B., Liang, Y., Rho, S., Lou, F., ... & Tian, Z. (2018). Deep learning based multi-channel intelligent attack detection for data security. *IEEE transactions on Sustainable Computing*, *5*(2), 204–212.

Karatas, G., Demir, O., & Sahingoz, O. K. (2018, December). Deep learning in intrusion detection systems. In *2018 International Congress on Big Data, Deep Learning and Fighting Cyber Terrorism (IBIGDELFT)* (pp. 113-116). IEEE.

Kevric, J., Jukic, S., & Subasi, A. (2017). An effective combining classifier approach using tree algorithms for network intrusion detection. *Neural Computing and Applications*, *28*(Suppl 1), 1051-1058.

Kudyba, S. (2020). COVID-19 and the Acceleration of Digital Transformation and the Future of Work. Information Systems Management, 37, 284 - 287. https://doi.org/10.1080/10580530.2020.1818903.

Kumar, S., Singh, B. P., & Kumar, V. (2021, December). A semantic machine learning algorithm for cyber threat detection and monitoring security. In *2021 3rd International Conference on Advances in Computing, Communication Control and Networking (ICAC3N)* (pp. 1963-1967). IEEE.

Kumar, G., Kaur, A., & Sethi, S. (2014). Computer network attacks-a study. *Int. J. Comput. Sci. Mob. Appl*, *2*, 24-32.

Kim, G., Yi, H., Lee, J., Paek, Y., & Yoon, S. (2016). LSTM-based system-call language modeling and robust ensemble method for designing host-based intrusion detection systems. *arXiv preprint arXiv:1611.01726*.

Leonardi, P. (2020). COVID-19 and the New Technologies of Organizing: Digital Exhaust, Digital Footprints, and Artificial Intelligence in the Wake of Remote Work. Journal of Management Studies, 58, 249 - 253. https://doi.org/10.1111/JOMS.12648.

Li, J. (2018). Cyber security meets artificial intelligence: a survey. Frontiers of Information Technology & Electronic Engineering, 19, 1462-1474. https://doi.org/10.1631/FITEE.1800573.

Liu, Z., Ghulam, M. U. D., Zhu, Y., Yan, X., Wang, L., Jiang, Z., & Luo, J. (2020). Deep learning approach for IDS: using DNN for network anomaly detection. In *Fourth International Congress on Information and Communication Technology: ICICT 2019, London, Volume 1* (pp. 471-479). Springer Singapore.

Mahesh, B. (2020). Machine learning algorithms-a review. *International Journal of Science and Research (IJSR).[Internet]*, *9*(1), 381-386.

McElwee, S., Heaton, J., Fraley, J., & Cannady, J. (2017, October). Deep learning for prioritizing and responding to intrusion detection alerts. In *MILCOM 2017-2017 IEEE Military Communications Conference (MILCOM)* (pp. 1-5). IEEE.

Mohri, M., Rostamizadeh, A., & Talwalkar, A. (2018). *Foundations of machine learning*. The MIT Press.

Moosavi-Dezfooli, S. M., Fawzi, A., & Frossard, P. (2016). Deepfool: a simple and accurate method to fool deep neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 2574-2582).

Narayanan, S. N., Ganesan, A., Joshi, K., Oates, T., Joshi, A., & Finin, T. (2018, October). Early detection of cybersecurity threats using collaborative cognition. In *2018 IEEE 4th international conference on collaboration and internet computing (CIC)* (pp. 354-363). IEEE.

NCSC, UK, sourced 15.02.2024 https://www.ncsc.gov.uk/section/about-ncsc/what-is-cyber-security

NIST, CSRC Content Editor. (Sourced 10.04.2024). *cybersecurity - Glossary | CSRC*. https://csrc.nist.gov/glossary/term/cybersecurity

Nitesh, K. T., Thirumala, A. K., Mohammed, U. F., & Ahmed, M. R. (2023, August). Network Security Threat Detection: Leveraging Machine Learning Algorithms for Effective Prediction. In *2023 12th International Conference on Advanced Computing (ICoAC)* (pp. 1-5). IEEE.

NVIDIA solutions for deep learning training. (Sourced 2024). NVIDIA. https://www.nvidia.com/en-eu/deep-learning-ai/solutions/training/

OpenAI, GPT-4. (Sourced 2024). https://openai.com/research/gpt-4

Oweis, N. E., Aracenay, C., George, W., Oweis, M., Soori, H., & Snasel, V. (2016). Internet of Things: overview, sources, applications and challenges. In *Proceedings of the Second International Afro-European Conference for Industrial Advancement AECIA 2015* (pp. 57-67). Springer International Publishing.

Samonas, S., & Coss, D. (2014). The CIA strikes back: Redefining confidentiality, integrity and availability in security. *Journal of Information System Security*, *10*(3).

Sarker, I. H. (2023). Machine learning for intelligent data analysis and automation in cybersecurity: current and future prospects. *Annals of Data Science*, *10*(6), 1473-1498.

Sewak, M., Sahay, S.K. & Rathore, H. Deep Reinforcement Learning in the Advanced Cybersecurity Threat Detection and Protection. *Inf Syst Front* **25**, 589–611 (2023). https://doi-org.ezproxy.jyu.fi/10.1007/s10796-022-10333-x

Shaukat, K., Luo, S., Varadharajan, V., Hameed, I. A., & Xu, M. (2020). A survey on machine learning techniques for cyber security in the last decade. *IEEE access*, *8*, 222310-222354.

Singer, P. W., & Friedman, A. (2014). *Cybersecurity: What everyone needs to know*. oup usa.

Sornsuwit, P., & Jaiyen, S. (2019). A new hybrid machine learning for cybersecurity threat detection based on adaptive boosting. *Applied Artificial Intelligence*, *33*(5), 462-482.

Tietosuoja.fi, Tietosuojavaltuutetun toimisto, Sourced 22.04.2024, https://tietosuoja.fi/tietojen-siirto-muihin-maihin

Ullah, F., Naeem, H., Jabbar, S., Khalid, S., Latif, M. A., Al-Turjman, F., & Mostarda, L. (2019). Cyber security threats detection in internet of things using deep learning approach. *IEEE access*, *7*, 124379–124389.

Zhou, Y., Cheng, G., Jiang, S., & Dai, M. (2020). Building an efficient intrusion detection system based on feature selection and ensemble classifier. *Computer networks*, *174*, 107247.