TALLINN UNIVERSITY OF TECHNOLOGY
School of Information Technologies

Georgios Kontis 2016831IVCM

Technical and Organizational Challenges on Enhancing Cybersecurity with Artificial Intelligence

Master's Thesis

Supervisor: Hayretdin Bahşi
Professor

Tallinn 2022

TALLINNA TEHNIKAÜLIKOOL
Infotehnoloogia teaduskond

Georgios Kontis 2016831IVCM

Tehnilised ja Organisatsioonilised Väljakutsed Küberturvalisusese suurendamisel Tehisintellekti abil.

Magistritöö

Juhendaja: Hayretdin Bahşi
Professor

Tallinn 2022

Author’s declaration of originality

I hereby certify that I am the sole author of this thesis. All the used materials, references to the literature and the work of others have been referred to. This thesis has not been presented for examination anywhere else.

Author: Georgios Kontis

16.05.2022

Abstract

Technical and Organizational Challenges on Enhancing Cybersecurity with Artificial Intelligence.

In an evolving cyber threat landscape with advanced and sophisticated threats, Artificial Intelligence (AI) steps into the cyber arena providing new opportunities and challenges to both “opponents”. Applying AI to Cybersecurity (AICS) can increase the security posture of an organization, overcoming the limited capabilities of traditional detection tools and human capacity. AICS can enhance the cognitive skills of the security analysts in terms of analysing large amounts of data as well as focusing on responses. But AICS technology comes with limitations and new challenges. AICS bears its own inbuilt weaknesses and enables new forms of attacks on the AI. These constitute both technical and organizational challenges for the organizations. AICS with the ability to alter an organization’s attack surface is challenging cybersecurity procedures and management in place. Additionally, cybersecurity frameworks and standards, drafted around classical information security and static security controls, pose another challenge. They have not addressed yet the AI and decision capable algorithms.

This research is an empirical study based on semi-structured interviews with ten cybersecurity experts with experience in the deployment of AICS tools. The research aims to explore the technological and organizational challenges of applying AICS. The research studies the different approaches chosen by the cybersecurity experts in regard to deploying AICS and the impact on their procedures and processes.

The results of the study highlight the cognitive advantage of AICS and the staggering amount of information that needs to be analysed and correlated in modern IT environments, the security controls for AICS tools, and the inability of the current versions of cybersecurity standards and frameworks to address the threats to AI.

This thesis is written in English and is 75 pages long, including 6 chapters, 7 figures and 5 tables.

Annotatsioon

Tehnilised ja Organisatsioonilised Väljakutsed Küberturvalisusese suurendamisel Tehisintellekti abil.

Areneval küberohtude maastikul koos kasvavate ja keerukate ohtudega, astub areenile tehisintellekt (AI) pakkudes uusi võimalusi ja väljakutseid mõlemale "vastasele". Tehisintellekti rakendamine küberturbes (AICS) võib suurendada organisatsiooni turvalisust ületades traditsiooniliste tuvastamisvahendite ja inimvõimekuse piiratud võimalused. AICS võib parandada turbeanalüütikute kognitiivseid oskusi nii suurte andmemahtude analüüsimisel kui ka reageerimisele keskendumisel. Aga AICS tehnoloogiaga kaasnevad piirangud ja uued väljakutsed. AICS-l on oma sisseehitatud nõrkused ja see võimaldab uusi rünnakuvorme tehisintellektile. Organisatsioonidele seab see tehnilisi ja organisatsioonilisi väljakutseid. AICS, mis võib muuta organisatsiooni ründepinda, on väljakutseks olemasolevale küberturbe protseduuridele ja haldamisele. Lisaks, küberturbe raamistikud ja standardid, mis on tehtud klassikalise teabeturbe ja staatiliste turvakontrolli jaoks, on täiendavaks väljakutseks. Neis ei ole veel käsitletud tehisintellekti ja otsustusvõimelisi algoritme.

Käesolev uuring on empiiriline uurimus, mis põhineb poolstruktureeritud intervjuudel kümne AICS-i tööriistade juurutamise kogemusega küberturbe eksperdiga. Töö eesmärk on uurida AICS tehnoloogiate rakendamisega kaasnevaid tehnoloogilisi ja organisatsioonilisi väljakutseid. Selles uuritakse erinevaid lähenemisviise, mida küberturbe eksperdid on valinud AICS kasutuselevõtmisel ning selle mõju nende protseduuridele ja protsessidele.

Uurimuse tulemused toovad esile AICS-i kognitiivse eelise ja analüüsimist ning seostamist vajava informatsiooni erakordselt suure mahu kaasaegses IT-keskkonnas, turvakontrollid AICS-i tööriistadele ja küberturvalisuse standardite ja raamistike praeguste versioonide suutmatuse adresseerida tehisintellektiga kaasnevaid ohte.

See magistritöõe on kirjutatud inglise keeles ja on 75 lehekülge, sisaldab 6 peatükki, 7 joonist ja 5 tabelit.

List of abbreviations and terms

AI	Artificial Intelligence
AICS	Artificial Intelligence in Cybersecurity
AVC	Application Visibility and Control
CDM	Continuous Diagnostics and Mitigation
CERT	Computer Emergency Response Team
CIIP	Critical Information Infrastructure Protection
CIS	Center for Internet Security
CISO	Chief Information Security Officers
CSF	Cyber Security Framework
CTA	Cognitive Task Analysis
DARPA	Defence Advanced Research Projects Agency
DC	Domain Controller
DDoS	Distributed Denial of Service
DHCP	Dynamic Host Configuration Protocol
DNS	Domain Name System
EDR	Endpoint Detection and Response
ENISA	European Union Agency for Cybersecurity
GDPR	General Data Protection Regulation
IDS	Intrusion Detection Systems
InfoSec	Information Security
IoC	Indication of Compromise
IP	Internet Protocol
IPS	Intrusion Prevention Systems
ISO	International Standards Organization
IT	Information Technology
ML	Machine Learning
MSSP	Managed Security Service Provider
MTTD	Mean Time to Detect
MTTR	Mean Time to Respond
NDR	Network Detection and Response
NIST	National Institute of Standards and Technology
OS	Operating System
OT	Operation Technology
PCI DSS	Payment Card Industry Data Security Standard

PoC	Proof of Concept
RL	Reinforcement Learning
RQ	Research Question
SCAS	Stream Clustering Algorithm for Suricata
SIEM	Security Information and Event Management
SL	Supervised Learning
SME	Small Medium Enterprise
SOC	Security Operation Center
TTP	Tactics Techniques Procedures
UEBA	User and Entity Behaviour Analysis
UL	Unsupervised Learning
URL	Uniform Resource Locator
UTM	Unified Threat Management
VM	Virtual Machine
ZTA	Zero Trust Architecture

Table of contents

List of figures

List of tables

1 Introduction

Cyber threat landscape keeps evolving rapidly with cyber threats becoming more advanced and sophisticated [1]. The modern IT/OT environments become more complex and the new trends of remote working, mobile devices, IoT, cloud computing, and dispersed systems expand their attack surface further. Adversaries have already started exploiting the use of Artificial Intelligence (AI) in their attacks, bringing new challenges to the cyber defenders and the organizations and companies in general [2], [3]. Cybersecurity experts and analysts, find themselves overwhelmed with large volumes of security alerts [4] generated from various sources, and the workload of triaging and response.

1.1 Research motivation

Artificial Intelligence (AI) or *Machine Learning* (ML) is stepping into the cyber arena providing new opportunities and challenges to both "opponents". Implementing ***Artificial Intelligence in Cybersecurity*** **(AICS)** can increase the security posture of an organization, overcoming the limited capabilities of traditional detection tools and human capacity with advantages as:

- Processing large volumes of data from a range of sources,
- Maintaining the level of cybersecurity of the organizations by rapid identification of threat factors and allowing security teams to focus on strategic tasks,
- Identifying Cyber Threats and suspicious Behaviours by slight movements and changes undetected to humans,
- Improving and automating the detection of cyber threats as AICS is adapted and learns from experience and patterns without human involvement,
- Accelerating detection and response time by handling multiple security alerts and allowing faster response.

As Andrade et al. [5] emphasize, AICS can enhance the cognitive skills of security analysts to comprehend and respond appropriately to security operations. ML-based tools allow them to analyse large amounts of data with enriched information and solve the limitations of human capabilities to process volumes of data within a reduced time. But like with any other technology, AI has its own list of failures as has been summarised by Yampolskiy [6]. In areas beyond cybersecurity with further experience in AI implementations, one may find extensive examples of AI weaknesses and failures, with some of them even deadly. E.g., AI systems have failed to recognise the same picture used for the model training when presented upside down [7], or have led to the first fatal crash, involving Tesla's Autopilot system [8].

The introduction of AICS comes with limitations and new **technological challenges** [3] that cyber defenders should consider too: The categorization models need training data with detailed datasets of anomalies, and malicious or non-malicious code that covers most use cases. Obtaining these efficient datasets requires time and resources that most organizations cannot afford [9]. Data quality is of critical importance as low quality results in poor decisions and vulnerable systems. When it comes to deploying AICS the increased possibility of false-positive, the creation of blind security spots [10], and the automated actions that can alter the attack surface and the usability of the system, may add extra burdens to the defenders [11]. Further, as another system, it can potentially be exploited by the adversaries, either in the training or the development phase [12]. As Anderson demonstrated [13], AI agents designed to probe ML-based detection systems, can reveal weak spots. From the end-users[1] perspective, if the system fails, that will happen in unpredictable and unexplainable ways [11].

AI can also enable **new forms of attacks** to the AI system itself as:

- Data poisoning by injecting false training data to corrupt the learning model and impact the future behaviour of the system.
- Adversarial inputs with minor changes to the original input, undetected by humans but not to the algorithm, can fool the classification model and tamper the system for future exploitation.
- The categorization model constitutes the most confidential asset of an AI system and Model stealing is a massive security breach as the output becomes predictable and systems manipulatable.
- Establishing backdoors triggers by injecting carefully crafted training data, could allow the adversary to dictate the future system's responses.

Additionally, to the above, AI technologies have inbuilt weaknesses as:

- Explainability and reasoning [14]. Due to the complexity of the decision models, AI systems have the reputation of being "black boxes", for not providing explanations and reasons for the decisions made.
- Lack of transparency and lack of reasoning in the AI decisions is an obstacle to better understanding the threat landscape and managing the risk. This is even more crucial when auditing the system.
- Trustworthiness. Lack of transparency, inexplainable decisions, and an evolving learning environment make it hard to evaluate whether the system will continue to behave as expected in any given context.

Summing up the above, one must thoroughly control the datasets provided to the AI system, especially during the model training, to safeguard the decision-making models, and at the same time to constantly question the system's reliability.

The above constitute not only technical but **organizational challenges** also for the organizations. Beyond adding one more IT asset to the risk management, AI with the ability to alter an organization's attack surface, challenges the cybersecurity procedures and management. Either having a customized framework to the individual's needs or following one of the international standards and frameworks as e.g., NIST Cybersecurity Framework or NIST Risk Management Framework, and ISO/IEC 27000-series, AI interacts with all stages of the risk management process. According to Wagner [15], cybersecurity is, after all, a process rather than a product, which should ideally strive to

[1] Throughout the text the term end-user refers to all the possible users of AICS solutions as cybersecurity experts, analysts, practitioners, AI researchers, SOC members, etc.

be consistent with broader organizational risk management practices and objectives in order to be maximally effective.

On the organizational side, cybersecurity frameworks and standards pose another challenge to AICS. Although a plethora exists, they are fragmented between various applications deriving from specific needs of individual industry sectors. The existence of an *industry-specific cybersecurity framework, or lack thereof, hinders the realization of cybersecurity goals in a wide range of industries* [16]. Additional to the above, there is no standardized method for evaluating cybersecurity technologies resulting in many disparate and non-repeatable evaluation processes [17]. Further, current frameworks and standards have been drafted around classical information security and static security controls controlled by humans and do not support the new paradigm of decision-capable algorithms [18].

1.2 Research novelty

Related studies on applying AI in the field of cybersecurity have focused on the usability assessment of ML-based tools [19], on functionality issues of SOC analysts and how they handle large amount of false-positive alerts from the AICS tools [20], on assessment of AI-based UEBA analytics in network security from not-seen before attacks [21], on the cognitive awareness of the cyber analysts and the elements they seek to perform their tasks [22], on the evaluation of organizational risk when deploying emerging technologies as AI and whether the executives are aware of the engendered risk [23], on the awareness of SME enterprises in UK about the AICS tools and the comparison between outcome and implementation costs of three specific tools [24], on the area of adversarial ML how industry practitioners protect, detect, and respond to attacks on their ML systems [25], on the impact of AI on the human aspects of information and cybersecurity [26], on the applied methods when deploying AICS technologies and the technical outcomes [27], or in evaluating cybersecurity risks in applications of AI in Government-to-Citizen e-services [28].

The above studies have focused to the AICS technologies and tools, the technical outcomes, the human aspects of the cybersecurity teams, and on the evaluation of the organizational risk when deploying AI technologies. This research aims to explore the impacts of the AICS technologies on the organization and management of cybersecurity. Also considering the inadequacy of the cybersecurity standards and frameworks, in their current versions, to directly address the threats to AI technologies, it is the author's belief that a gap exists in the literature. This research aims to explore the approaches chosen by the cybersecurity experts to secure their AICS deployments and the impact on their procedures and processes and their overall cybersecurity policies.

1.3 Research questions

The topic of the research is summarised in the phrase: **Technical and Organizational Challenges on Enhancing Cybersecurity with Artificial Intelligence.**

The research investigates the role and fit of AI-based tools into the cybersecurity management of organizations through the following research questions:

RQ: What are the technical and organizational challenges to enhancing cybersecurity with AI?

RQ1. What has been the main motivation for the use of AICS solutions?

RQ2. How have AICS solutions been implemented?

RQ2a. Have the organizations considered the possible cyber-risks as expressed above?

RQ2b. Do the organizations consider the implementation successful and what is their experience so far?

RQ3. Has the design or implementation phase of an AICS system affected the current cybersecurity policy of an organization and how?

RQ4. Is the cybersecurity framework used for cybersecurity management still applicable with the integration of an AICS solution?

The research was designed as an empirical qualitative study around a questionnaire with open-ended questions. As an inquisitive study having open-ended questions was considered appropriate in order to record a wide range of possible results. The research embarks not from an initial hypothesis that remains to be proved or not, but from a general notion that automation in cybersecurity comes *naturally* with the evolvement of the technology. It aims to understand the *whys* and *hows* are happening in the field of cybersecurity and to investigate the impacts and outcomes.

1.4 Research scope

This research scopes to understand how organizations' cybersecurity management is affected by AICS solutions and to what extent. For that, firstly, examines the stance of the organizations and the cybersecurity experts in deploying AICS solutions. The research focuses on the motivations and the expectations of the participants. Secondly, investigates the various deployment approaches and what challenges may arise with the introduction of AICS into productional environments. Investigates the cybersecurity provisions taken by the organizations and the necessary organizational changes that had to be made. Finally, the research records the participants' experience with the use of AICS and delves into the symbiosis with cybersecurity teams and analysts.

1.5 Thesis overview

Chapter 2 provides the background knowledge of the research. It is an overview of the basics of AI, the need to introduce AI in cybersecurity, the technology used by AICS, the AICS tools already available in the market, the inherent vulnerabilities and challenges of the AI systems, and the threats to AI and how the systems can be secure. Also overviews the current version of popular cybersecurity management frameworks in the concept of AI technologies. Section 2.4 describes the literature review and summarises the results of similar surveys in the literature. Chapter 3 presents the methodology of the research and the thematic analysis of the results. Chapter 4 presents the results of the survey grouped by the identified themes. Chapter 5 discusses further the results and presents the findings of the research. A conclusion of the research is given in chapter 6 followed by appendixes and a reference list.

2 Background information on AI and Cybersecurity Management

According to Trend Micro’s 2021 Annual Cybersecurity Report “*the digital migrations and transformations that had enabled organizations to continue their operations amid the Covid-19 pandemic continued to usher in significant shifts in the threat landscape in 2021”* [1]. The modern IT environments are becoming complex and distributed beyond the traditional boundaries of an organization’s workspace. With the proliferation of cloud services and remote working, traditional firewalls based on signature and rule-based methods, have been proven inadequate to protect the business in the modern hybrid cloud environment [29].

Intrusion Detection Systems (IDS) and Intrusion Prevention Systems (IPS) similarly use signature-based detection rules to analyse the network traffic. In a constantly shifting cyber threat landscape, it is difficult to constantly adapt the rules for abnormal behaviour, leading to staggering numbers of alerts [30]. The overwhelming number of the generated alerts and the high ratio of false-positive alerts, need to be manually reviewed by the cybersecurity analysts in a time-consuming and tedious process that exceeds SOC’s capacity to handle them [31], [32].

Security Information and Event Management (SIEM) systems provide the ability to correlate events from various sources according to predefined correlation rules and can provide context as visualisation of the results in a central platform. But as Sheridan [33] describes SIEM systems are excessively dependent on the analysts’ competence to comprehend the information. As McElwee et al [32] described, SIEM systems are open to human errors too, and an overload of non-important alerts overwhelms the cybersecurity analysts. Further, the cybersecurity teams need to constantly craft new correlation rules to keep the SIEM systems updated with the evolution of the cyber threats.

For the above reasons is imperative the need of intelligent agents to handle the tedious tasks in a 24/7 basis, allowing cyber analysts to overcome the *alert fatigue* and assist them to stay focused on the bigger picture of cybersecurity. This chapter provides the academic background for the need for AI in cybersecurity and the challenges to its implementation.

2.1 AI in Cybersecurity

AI is a term that collectively describes systems perceived to have intelligence and ability to make decisions or predictions based on previous experiences and learnings. Typical AI applications nowadays include speech, image and video recognition, autonomous vehicles, natural language processing, conversational agents (chat-bots), as well as smart automation and advanced simulation. ML as a branch of AI is the science of writing computer code that has the ability to learn and adapt itself with the use of algorithms and data. With the use of statistical models, ML can analyse data and discover patterns and inferences.

Organizations’ attack surface can be massive and with the latest technology trends also complex. Cloud computing, remote working, mobile devices, different technologies and legacy systems, dispersed systems, IoT devices, and IT/OT landscape constitute a complex IT ecosystem. For the cybersecurity teams responsible for the protection and detection of cyber threats, generated alert notifications and log entries are crucial for risk calculation and threat detection. But the analysis of thousands, if not more, of systems' alerts and timely response within seconds, is a problem beyond human scale anymore [34].

For such an unprecedented challenge and to assist cybersecurity teams to improve their posture and efficiency, cybersecurity tools have started to implement the use of AI. AICS technologies are critical to quickly analyse events and alerts of multiple sources and identifying various types of threats and anomalous behaviours. From malicious code exploiting zero-day vulnerabilities to phishing attacks and suspicious data movements. AI technologies learn over time, building profiles of the organization's IT assets, networks, and users, being able to recognise deviations from standard normal behaviours.

A distinction needs to be made between pure AI-enabled systems and data analysis tools. The latest analyse volumes of data and based on pre-set rules drive certain responses. AI tools on the other hand automate their actions based on reproduced cognitive abilities. The main distinctions are:

- AI tools become more intelligent by time and the data we feed them. They possess the ability to self-tech themselves from past experiences and increase their capabilities. They could be characterised as dynamic and evolving systems.
- Data analysis examines big sets of data to return conclusions through pre-defined rules and instructions. Data analysis is neither self-learning nor iterative. In the area of cybersecurity, tools as Zeek[1], Snort[2] or Suricata[3] provide that level of automation to identify traces of threats and intrusions. Similarly, modern antivirus software [35] check suspicious files against known signatures in malware databases.

Another clarification needed is between the terms of AI and ML. Although both definitions are closely related and can be used interchangeably are not the same as ML is considered a subset of AI. AI with the use of math and logic, allows a system to mimic human learning and problem-solving functions and to make decisions or perform task on its own. ML is an application of AI and allows the computer to learn on its own without direct instruction through analysis of given data sets and to improve its knowledge based on experience. For the current study, we examine the use of AI-enabled software in cybersecurity tools and systems, aiming at enhancing traditional cybersecurity controls. Throughout the document, AI is used as a general term that also includes ML technologies.

The basics of AI

AI, and its subset ML, are the science – or art – of crafting computers to learn. It exists at the intersection of computer science, statistics, and linear algebra, with insights from neuroscience and other fields as well [36]. Contrary to the traditional programming of instructing computers to act on specific rules in fixed conditions, AI is focusing on

[1] https://zeek.org/

[2] https://www.snort.org/

[3] https://suricata.io/

programming computers with the ability to teach themselves. It uses algorithms that study data for patterns and correlations and focus on creating predictions. To understand the vulnerabilities and the threats against AI a cursory understanding of the technology is provided.

One of the core components of an AI system is the **classification model**. This could be as simple as an extensive list of IF…ELSE… statements, advanced technology as neural networks, or a future technology not existing yet. These constitute the critical parameters of the model that need to be tuned with the most accuracy possible. Whatever the underline technology being used, the critical principle for the cybersecurity practitioners is one: the classification model's accuracy depends on the tuning of the parameters which is determined by the training of the AI system.

AI is fuelled by data. The necessary data needs to be collected from somewhere in structured or unstructured form. In the **training phase**, the classification model extracts patterns and inferences from the acquired dataset. The AI system itself, learns from the data by calibrating the parameters of the classification model to match these patterns. Several approaches are used but the main idea remains the same, the system adjusts the model to the training data. It is important to understand that the system has no previous knowledge of the desired pattern, it simply teaches itself from the dataset in the best possible way.

In the **deployment phase** (Figure 1) the AI component becomes part of a larger system with several other models. In deployment, several models may be combined to perform similar tasks, and some might keep being updated and trained while others to remain in the initial deployment phase. Additionally, human interaction might be involved to approve a system's decision before acting. Essentially, the AI system translates real-world input data, which is hopefully similar to the training data, into decisions and actions[37].

AI categorization

According to the level of sophistication, the required tasks, and the learning techniques, AI systems can be categorised in three different ways [38]. As of their sophistication or their cognitive abilities to understand, learn, and act on the provided information, AI technologies can be characterised as [39]:

- **Narrow or Weak AI**. The term refers to AI already in use today, e.g., such as smart applications (e.g., Alexa), recommendation systems (e.g., online shopping), and spam filtering (e.g., gmail). It refers to AI systems with a narrow focus created for specific tasks and unable to perform other tasks beyond them. In general, this level of AI although has specific analytical capabilities beyond humans, when it comes to the overall cognition is still inferior. The learning process is mostly supervised through labelled data.
- **General or Strong AI**, refers to future advancements of AI, not present today, when its perceptual and cognitive abilities are expected to reach humans. The learning process should also be unsupervised.
- **Super AI,** refers to AI advancements beyond the human abilities and the far-off development stage.

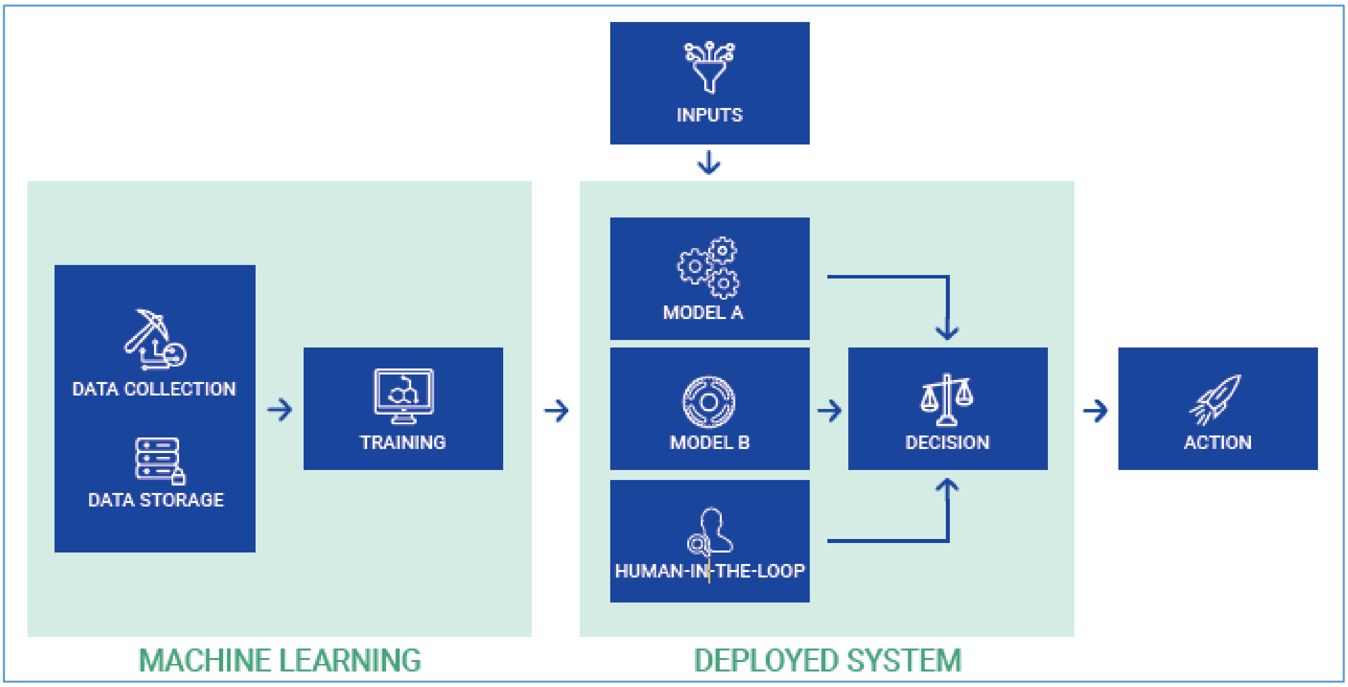

Figure 1. Development stages on an AI system. Once the classification model is trained, it becomes part of a larger system [37]

AI training methods

The three main methods to train ML algorithms are:

- Supervised Learning or SL is training the system with the use of pairs of labelled data as input objects and desired output values. It is a task-driven process where data are provided as input–output pairs, and the system learns a function[1] that maps them together [40] [2].
- Unsupervised Learning or UL is a data-driven process. The system is provided with a large amount of unstructured data and tries to identify patterns and key structures without explicit instructions. The most common method used is clustering the data into groups that share meaningful characteristics [40][3].
- Reinforcement Learning or RL, is training the system through providing feedback in the form of rewards or punishments. The algorithm introduces an ML agent into the environment and through repetitive reinforcement teaches it how to act. It is up to the agent to decide which prior action led to the reinforcement [40].

[1] A cost function enables the algorithm to make predictions based on previous data and to test itself by calculating the error or how far the algorithm is from perfect performance

[2] Popular supervised training algorithms are k-Nearest Neighbor (kNN), decision trees, Multi-Layer Perceptron (MLP) and Support Vector Machines (SVM).

[3] Popular unsupervised training algorithms are k-means clustering, single linkage clustering and y-algorithm

Applying AI to cybersecurity

AI is suitable for complex problems solving and cybersecurity fits perfectly into that. With the evolving situation in the cyberspace AI can be used to automate threat detection and even assist in automated responses better than the traditional software tools with the “hard-wired” approaches. Factors that support the use of AI are:

- the vast attack surface of the organizations
- the plethora of devices and technologies in the IT ecosystem
- the evolving number of cyber-attacks
- the lack of skilled professionals in cybersecurity [41] [42].
- the volume of data that needs to be evaluated has moved beyond human capacity

An AI solution in Cybersecurity with self-learning capacity can solve the above challenges enabling security posture advancements for the organizations. Key advantages of AI solutions to consider are:

- AI can correlate automatically various potential incidents. By overcoming human shortcomings (inexperience, fatigue, shift turnover[1] or time) and combining various sources of threat intelligence AI can highlight commonalities across vast volumes of alerts, current or past, and provide feedback to the security analysts.
- Operating on a 24/7 basis AI provides correlations and assists to make decisions on a knowledge built on the individual threat profile of every organization.
- AI performs investigations in both structured and unstructured data, at times faster than humans. It may provide data-driven analytics to security analysts to detect known and unknown threats and reduce key indicators as MTTD (Mean Time to Detect) and MTTR (Mean Time to Respond).
- AI can assist security analysts to conduct more thorough and consistent investigations in a shorter time and focus on strategic threat investigations and threat hunting.
- It can assist security teams to prioritise the most critical alerts first.
- Certain AI tools can correlate the cyber incidents to known cybersecurity frameworks such as MITRE ATT&CK, providing a visualization of the event.

AI enablers to cybersecurity architecture and policymaking:

- AI can assist organizations to draft an accurate **IT asset inventory** and develop better **visibility[2]** over their IT environments. Having full knowledge of the IT environment is challenging and almost impossible for large organizations. AI can assist to detect unknown traffic, devices, services, and applications, and properly applying cybersecurity policies. With better visibility, cybersecurity architects can design robust security schemes and place security controls in critical areas.
- AI can keep up with the latest global and industry-specific external threat information. Security teams can shape a better view of the organization’s **threat exposure** and make critical decisions on possible attacks.

[1] Referring to SOCs personnel and the rotational working shifts

[2] Application Visibility and Control (AVC)

- AI can assist security teams to maintain their security posture and assess the **effectiveness of security controls** and tools indicating potential security gaps.
- Assuming the above, AI can assist security teams in better assessing their risk of a potential breach and improve an organization's resilience.
- Allows security teams to design robust incident responses by providing context for prioritization of their responses and root cause analysis to mitigate vulnerabilities and future events.

Early uses of AI in Cybersecurity

Google has employed ML to filter emails since 2000, while currently, ML applications support various services provided by the company. "*Before we were in a world where the more data you had, the more problems you had. Now with deep learning, the more data the better*". Elie Bursztein, head of the anti-abuse research team at Google [43].

IBM has been developing its AI solution, IBM QRadar with Watson, for years, as a "cognitive learning platform for knowledge consolidation" based on ML. "*A lot of work that's happening in a security operation center today is routine or repetitive, so what if we can automate some of that using machine learning or just make it easier for the analyst?*" – Koos Lodewijkx, vice president and chief technology officer of security operations and response at IBM Security [43].

Commercially available AI solutions for Cybersecurity

IBM QRadar Advisor with Watson[1]

IBM has been a pioneer in AI research and "*IBM QRadar Advisor with Watson*" is its proposal for applied AI in cybersecurity. The solution is focused on security operation centres (SOC) and functions as a platform combining SIEM tools and automated investigations. With the use of cognitive reasoning provides the security analysts with critical insights and assists them to drive consistent and deeper investigations, and accelerating response and incident handling. Security analysts can assess threats and reduces the risk of missing them by combining current and past incidents and adding context to the threat analysis.

Darktrace[2]

Since 2013, Darktrace is developing its autonomous cyber-AI solution. Darktrace's "*Enterprise Immune System*" it employs correlation techniques to understand the normal activity and behaviour of an organization's network within a particular IT environment. From that knowledge, it detects anomalies and potential threats in real-time. Darktrace's "*Antigena*" with the help of ML technology is an autonomous response and self-defence solution that allows the network to react in machine time against in-progress cyber-attacks. Darktrace's "*Cyber AI Analyst*" combines the above solution into a single platform with AI investigation technology which automatically triages, interprets, and reports on the full scope of security incidents.

[1] https://www.ibm.com/products/cognitive-security-analytics
[2] https://www.darktrace.com/en/

Siemens Energy, Managed Detection and Response[1]

Cybersecurity is crucial to Operational Technology (OT) as in Information Technology (IT). In a partnership with Darktrace, Siemens Energy offers an industrial cybersecurity service, Managed Detection and Response (MDR), to help small and medium-sized energy companies defend critical infrastructure against cyberattacks. It employs AI and ML methodologies for real-time anomaly detection in industrial OT environments.

Blackberry Cylance AI[2]

Cylance, currently owned by Blackberry, was one of the first antivirus developers to apply ML algorithms in cybersecurity. Its product of "*Blackberry Cylance AI*", already in the seventh generation, is trained on billions of diverse threat data sets over several years of real-world operation and tested across an array of cybersecurity.

Sophos Intercept X tool[3]

Sophos Intercept X uses AI neural network empowered by deep learning techniques. Intercept X extracts millions of features from a file until a file executes, performs an in-depth inspection, and decides whether a file is benign or malicious in milliseconds. The model is trained by access to millions of samples on real-world feedback and bidirectional threat intelligence exchange. Additionally, to limit new ransomware and boot-record threats, Intercept X utilizes behavioural analysis on the network and the user.

Acalvio ShadowPlex[4]

Alcavio presents a different approach with its "*ShadowPlex*" product. It uses AI technology to actively defend against adversaries with an autonomous deception solution. The tool addresses the limitations of manually creating and maintaining traps and honeypots in an organization's environment with the introduction of AI technology. The algorithms based on network discoveries and specified playbooks autonomously create deceptions to divert attacks in safe and contained sandboxes to observe and collect TTPs.

Balbix[5]

The Balbix Platform uses AI algorithms to discover and analyse the organization's attack surface to give an accurate view of breach risk. The tool provides a prioritized set of actions that the security team can take to enhance the cybersecurity posture and reduce the cyber risk. The core component, "*Balbix Brain*", runs in the cloud and leverages advanced AI and self-learning algorithms to calculate risk for every network entity, unify cybersecurity data, predict likely breach scenarios, prioritize vulnerabilities, and prescribe necessary risk mitigation actions.

[1] https://press.siemens-energy.com/global/en/pressrelease/siemens-energy-announces-new-ai-driven-cybersecurity-monitor-and-detection-service

[2] https://www.blackberry.com/us/en/products/unified-endpoint-security/cylance-ai

[3] https://www.sophos.com/en-us

[4] https://www.acalvio.com/

[5] https://www.balbix.com/

Securonix UEBA[1]

The Securonix User and Entity Behavior Analytics (UEBA) tool uses ML and behaviour analytics to analyse and correlate interactions between users, systems, applications, IP addresses, and data. The tool detects advanced insider threats, cyber threats, fraud, cloud data compromise, and non-compliance and includes built-in automated response playbooks and customizable case management workflows for immediate response to cyber events.

SCAS[2]

SCAS (Stream Clustering Algorithm for Suricata) [44] is a stream clustering algorithm designed for classifying Suricata IDS alerts in real-time, and mining frequent alert patterns that represent common attack scenarios of low importance. It is an open-source solution, developed by one of the participants of the research, P2b, and is widely available.

The reader may find an extensive review of ML techniques applied in cybersecurity in the works of Torres [45], Husák [46], and Sokolov [47]. Various researchers have presented ML implementations in cybersecurity for SDN [48], Connected Autonomous Vehicles [49], Brain to Computer Interface for IoT [50], smart grid [51], healthcare critical infrastructures [52] and [53], network traffic analysis [54], smart homes [55], insider threat detection [56], or defensive deception [57].

2.2 Barriers and challenges of AI

As discussed already, AI fits perfectly into today's complex cybersecurity landscape. However, there are challenges in the use of AICS. The challenges arise for reasons discussed below inherent to the AI technology itself as cybersecurity has not been the focus of developing AI and ML models [58] so far. There is a need to develop powerful and robust ML techniques focused on cybersecurity.

- AI systems depend heavily on the receiving data either during the training / testing of the model or during the deployment. The quality of the data is of importance as is the integrity. Poor data quality or erroneous data lead to poor models and random decisions [59].
- AI systems are trained and designed usually against specific and known cyber-attacks [58]. Most publicly available data sets usually do not include the latest attacks and crafting customised data sets is not possible for many organizations. Additionally, detecting attacks not seen before can be challenging for the AI models.
- Hindy et al. [60] in their review of the availability of proper training datasets have calculated that around 75% of highly cited academic research on ML-based IDS systems rely on datasets generated between 1998 and 2000 (mostly KDD '99 and DARPA datasets) or variants. Further, approximately 11% of the studies use

[1] https://www.securonix.com/products/ueba/

[2] https://github.com/ristov/scas

recent or real-life generated / simulated datasets. Although flows have been identified in these datasets, as Sommer et al. [34] highlight on applying ML in network IDS, still exist few updated public datasets available to academia for research.

- Ibrahim et al.[61], stress further the threat intelligence problem for training and updating the AI systems. Most of the existing AI systems train on specific security threat scenarios with well-defined inputs (i.e., data streams and logs) and provide well-defined outputs (security indicators) that can then be integrated into other security metrics and tools, or manually investigated by security analysts [61]. There is a lack of threat intelligence that can be generalized across different architectures or businesses without complex development and customization [61].
- Related to the data sets, Quionero et al. [62] pointed out the data shift problem where AI techniques failed where the joint distribution of inputs and outputs differs between the training and test stages of the algorithms [62].
- Evaluation of the classification models is not uniform and standardised. As Shaukat et al. proved [58], various researchers evaluate classification models on different parameters. An agreed standard set of metrics is needed for the model's comparison [58].
- One of the biggest challenges for AI is the "**black box**" characterization [63]. It is challenging even for the developers to understand the inner workings of the algorithms [64] and explain the output. This lack of transparency might be deliberate when it comes to commoditized software. But as Burrell [65] points can be simply attributed to the complexity between the algorithm and the data on which the systems base its decisions, which is too complex for humans to understand [65]. It is not only the volume of the data as input but mainly the nature of the AI algorithm as is altered when learning on training data. AI learning models possess a degree of unavoidable complexity [65] arising the issue of **explainability** of AI applications. The explainability of AI is an important problem as algorithms often fail to provide insight to their behaviour and decision process. These explanations are key to establishing **trust** in the operation of the AI, ensuring the openness of the technology, and identifying potential bias/problems in the training datasets, Gilpin et al. [66]. Phillips et al.[67] set the four principles of explainable AI as: explanation or accompanying evidence for all outputs, meaningfulness or understandable explanations to the users, explanation accuracy reflecting the system's process for generating the output, and knowledge limits as the system should operate within the scope it has been designed. Various researchers have worked on the issue of explainable ML algorithms with the work of Gilpin et al. [66] providing a summary.
- **Adversarial robustness** of an AI system is defined as the ability of the AI to return the same output when presented with slightly modified input [18]. Carlini et al. stress that despite the number of studies attempting to design defences that withstand adaptive attacks, most papers quickly shown to be incorrect [68]. In the same study, they point out that to evaluate the defence mechanisms defenders should test them according to the worst scenario against it [68].
- The lack of **transparency**, provable preservation of the system's properties, and the complexity of the learning process, raise a concern on the trustworthiness of the AI systems. "*Records of past behaviour are neither predictive of the systems' robustness to future attacks, nor are they an indication that the system has not been corrupted by a dormant attack (for example, has a backdoor) or by an attack that has not yet been detected"*[11].

The above challenges are characterised more distinctively by Buchanan and Miller in [36]. They consider ML to be still comparatively immature with the ML applications in cybersecurity to be nascent (as Najafabadi agrees in [69]). The authors argue that as ML gains more role in society, these concerns must be accorded paramount importance, security principles must be built in from the start, and operators must be able to adapt flexibly to emerging threats [36].

Malicious use of AI by adversaries

As AI technologies become more mature, cheap, and widespread[1] [70], adversaries can employ the advantages of AI tools to advance their attacks. State-sponsored attackers, criminal cyber-gangs, and hacktivist groups can employ the same publicly available AI technologies to defeat defences and avoid detection. Researchers expect [71] AI-enabled attacks to expand existing attacks to become more effective, more finely targeted, and more difficult to attribute. To introduce new threats and more likely to exploit vulnerabilities and inherent characteristics in AI systems. And to alter the typical characteristics of threats [71]:

- As security teams use AI to scan an organization's attack surface for vulnerable areas and security gaps, adversaries can do the same and automate the reconnaissance stage if not the attack process [2].
- AI gives the potential to the adversaries to leverage traditional attacks to unprecedented dimensions. E.g., phishing can become more sophisticated mimicking the victim's style more accurately.
- Adversaries can develop AI-powered attack tools that adapt to our defences and responses in machine speed, e.g., AI-powered malware.
- The same AI tools designed to protect our organizations can be used to develop new stealthier forms of attacks.
- Or even to create new forms of attacks such as Deep Fake and Misinformation Campaigns[2]
- AI systems, like any other technology, comes with their own specific vulnerabilities (explained in the next paragraph "Applying Cybersecurity to AI").
- Cheap technology will enable the increase of threat actors[3] [70] as the technology will become more available.

[1] As J.-M. Rickli puts it, "*artificial intelligence relies on algorithms that are easily replicable and therefore facilitate proliferation. While developing the algorithm takes some time, once it is operational, it can be very quickly and easily copied and replicated as algorithms are lines of code*", J.-M. Rickli (2018), "The impact of autonomy and artificial intelligence on strategic stability", UN Special, July-August, pp. 32-33.

[2] A portmanteau of "deep learning" and "fake media," deep fakes involve the use of AI techniques to manipulate or generate visual and audio content that is difficult for humans or even technological solutions to immediately distinguish from authentic ones [12]

[3] M.C Horowitz et al. give the example of the 'script kiddies', e.g., "…*relatively unsophisticated programmers, (...) who are not skilled enough to develop their own cyber-attack programs but can effectively mix, match, and execute code developed by others? Narrow AI will increase the capabilities available to such actors, lowering the bar for attacks by individuals and non-state groups and increasing the scale of potential attacks for all actors*.", M.C Horowitz et al. (2018), "Artificial Intelligence and International Security", Center for a New American Security, p. 13.

Applying cybersecurity for the AI

ENISA, at the 2020 report on "AI Cybersecurity Challenges" [72] points out that "*the AI threat landscape is vast and dynamic, since it evolves alongside the innovations observed in the AI field and the continuous integration of numerous other technologies in the AI quiver*" [72]. AI systems have a larger attack surface. As an IT system, are exposed to typical cybersecurity risks stemming from hardware / software bugs to vulnerabilities and zero-day attacks. But the distinctive features of AI systems can be attacked in non-traditional ways, raising new cybersecurity questions. This includes the phases of training the models, the interaction of the system with the external environment[1], and its future evolvement during runtime[2]:

- Model theft. Although an attack to steal the classification model belongs to "traditional information security", is worth noting for two reasons. Firstly, classification models, if are mostly developed by researchers, can be publicly available as "open-source software" to anyone. Secondly, the model can be stolen as any other data stored in an IT system. As models are not always seen as highly sensitive assets, the systems holding these models may not have high levels of cybersecurity protections [73]. Judging by the commercial software in the market already, Comiter remarks [73] that commoditized software is often handled insecurely, as demonstrated by the prevalence of the root passwords, and expects the same to happen for AI.
- The attack doesn't even have to be sophisticated or illegitimate. In an AI-enabled system with the authority to perform actions such as e.g., blocking network traffic, all that is needed by the adversaries is to discover the correct pattern to trigger the system's response. With continuous probing and without interfering with the model or the dataset they can achieve their goal to turn the system against its organization.
- Input attacks attempt to manipulate the input data of the systems to alter the system's output, decision, or prediction. As shown in Figure 2 and Figure 3, an "attack pattern" is added to the input data causing the model to make the incorrect decision. The attack pattern doesn't necessarily have to be noticeable as shown in Figure 3. Further, the adversaries do not need the knowledge of the classification model itself. By having access to the training dataset, who usually are publicly available as the models can be, adversaries can develop "copy models" and prepare their attacks [74] on the original models.

- Attacks against the AI systems often aim to acquire the control of the targeted system and change its behaviour [11]. Pupillo et. Al [70] provide four major possible types of attacks: data poisoning, tempering of the categorisation models, backdoors, and reverse engineering of the AI models:

  1. Data poisoning [11]: Attackers may add carefully crafted erroneous data among the legitimate dataset used to train the system to modify its behaviour.
  2. Tampering of categorisation models [11]: By manipulating the categorisation models, attackers could modify the final outcome. For instance, researchers using pictures of 3D printed turtles, obtained using a specific algorithm, were

[1] Referring to the operational environment of the AI system

[2] Especially for the ML systems were past behaviour affects future decisions and predictions

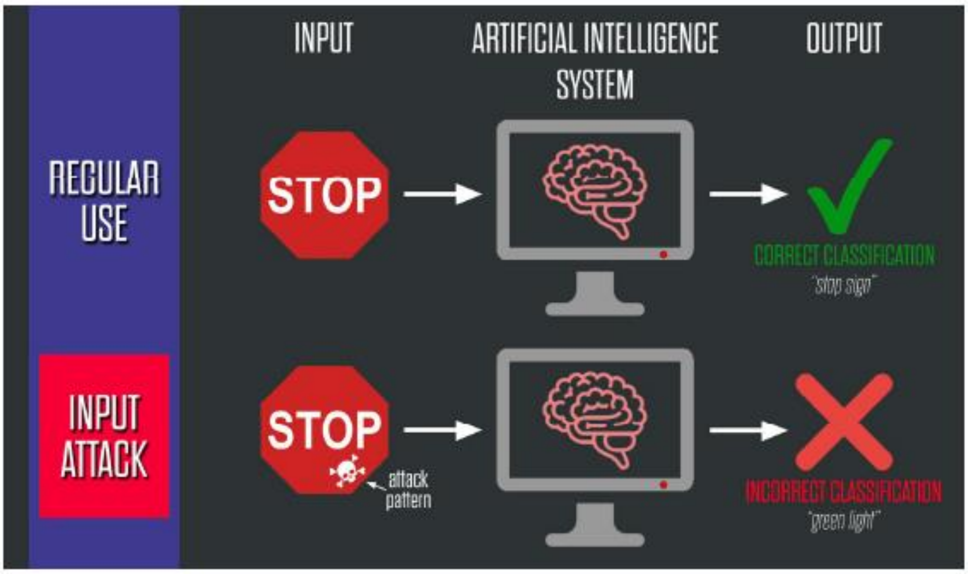

Figure 3. On top, the AI system processes a valid input and makes a decision. In the bottom, the input has been modified with an attack pattern causing the systems to make a wrong decision.

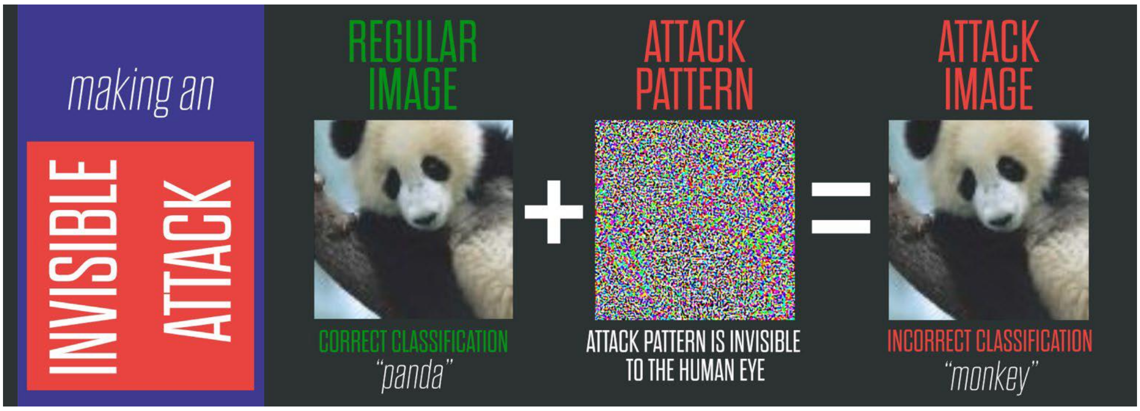

Figure 2. By adding an invisible to human eye noise to the image the AI system misclassifies the image. Image concept showing how attack is formed from Goodfellow, Ian J., Jonathon Shlens, and Christian Szegedy. "Explaining and harnessing adversarial examples." arXiv preprint arXiv:1412.6572 (2014)

able to deceive the learning method of an AI system and classify turtles as rifles [75].

3. Backdoors [11]: The AI systems can be compromised through backdoor injection attacks. For such an attack, the adversaries inject a "*customized perturbation mask applied to selected images*" via data poisoning of the training dataset and override the correct classifications. "*The backdoor is injected into the victim model via data poisoning of the training set, with a small poisoning fraction, and thus does not undermine the normal functioning of the learned deep neural net*". Hence, such attacks, once triggered, "*can exploit the vulnerability of a deep learning system in a stealthy fashion, and potentially cause great mayhem in many realistic applications − such as sabotaging an autonomous vehicle or impersonating another person to gain unauthorized access*" [76]. This case is illustrated in Figure 4.

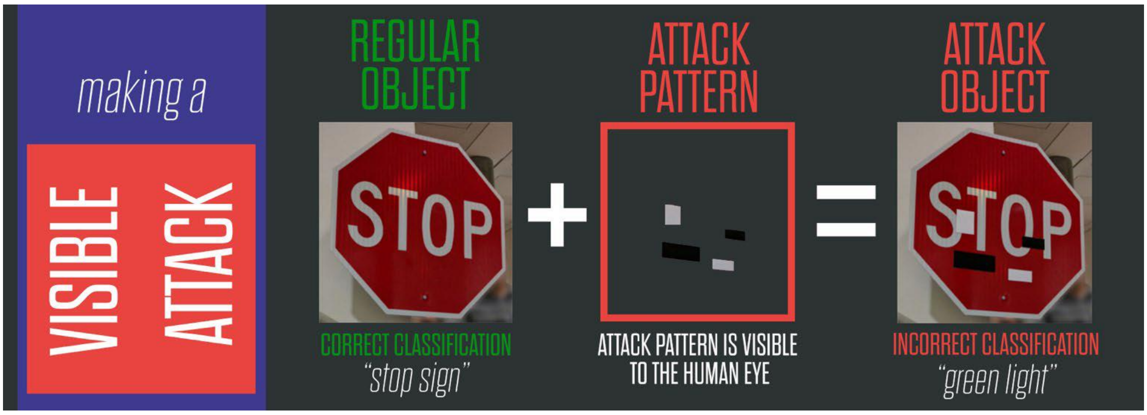

Figure 4. A visible input attack. A small attack pattern is affixed.

4. Reverse engineering the AI model [70]: The adversaries gain access to the AI model through reverse engineering and perform more targeted and successful attacks. As Dubey et al. proved in [77], with the use of Differential Power Analysis, the adversary can target the ML inference, assuming the training phase is trusted, learn the secret model parameters, and develop a "copy model".

Table 1 gives an overview of possible attacks.

Table 1. Intentionally motivated ML failure modes[1]

Attack	**Overview**
Perturbation attack	Attacker modifies the query to get appropriate response
Poisoning attack	Attacker contaminates the training phase of ML systems to get intended result
Model Inversion	Attacker recovers the secret features used in the model through careful queries
Membership Inference	Attacker can infer whether a given data record was part of the model's training dataset
Model Stealing	Attacker can recover the model through carefully crafted queries
Reprogramming ML system	Repurpose the ML system to perform an activity it was not programmed for
Adversarial Example in Physical Domain	Attacker brings adversarial examples into the physical domain to subvert ML system e.g., 3D printing special eyewear to fool facial recognition system
Malicious ML provider recovering training data	Malicious ML provider can query the model used by customer and recover customer's training data

[1] Source: R. Shankar, S. Kumar, D. O'Brien, J. Snover, K. Albert and S. Viljoen (2019), "Failure Modes in Machine Learning", Microsoft, November (https://docs.microsoft.com/en-us/security/engineering/failure-modes-in-machine-learning#intentionally-motivated-failures-summary).

Attack	Overview
Attacking the ML supply chain	Attacker compromises the ML models as it is being downloaded for use
Backdoor ML	Malicious ML provider backdoors algorithm to activate with a specific trigger
Exploit Software Dependencies	Attacker uses traditional software exploits like buffer overflow to confuse/control ML systems

2.3 Organizing cybersecurity through risk management frameworks

Cybersecurity, as Information Security in general[1], spans many areas when comes to protect the IT infrastructure and the data within. Areas such as perimeter protection from potential intruders or DDoS attacks, encryption of sensitive and crucial information, application security from malicious software, etc. But asides from the technological aspects of the chosen security solutions and controls, organizations should give importance to procedures, plans, and policies in place, and address the organizational aspects of cybersecurity [79]. The protection of the digital assets becomes even more complicated by variant legislations such as GDPR[2] for personal data or internationally accepted standards such as PCI DSS[3] for digital payments. In extension, organizations should be aware and prepared to manage cyber incidents and crises and qualitatively assess their risk exposure and appetite. To address the above-said considerations, they should adopt efficient cybersecurity management that includes effective processes and methodologies beyond the technological solutions.

Cybersecurity frameworks, standards, and best practices contribute to understanding the different types of attacks and managing cyberattacks [80]. Also, provide the necessary tools to the organizations to identify and protect their critical IT assets from the evolving cyberthreats. Cybersecurity frameworks refer to defined guidelines, typically voluntary, containing processes and practices to assist organizations to secure IT assets and manage the cybersecurity risk. The frameworks can consist of security standards, best practices, or implementations. There are a plethora of different frameworks proposed by various bodies[4], but few constitute the most popular and widely accepted. These include the US National Institute of Standards and Technology Cybersecurity Framework (NIST CSF)[5], the Center for Internet Security Critical (CIS)[6] Security Controls, and the International Standards Organisation (ISO) frameworks ISO/IEC 27001 and ISO/IEC 27002[7]. It should be noted that the said Cybersecurity Frameworks are mostly voluntary guidelines for the organizations to design their cybersecurity and architecture and manage their risk exposure. As Klahr et al. [81] found in their survey in 2017, most of the businesses were not aware of the said frameworks, unless they were involved in projects and tenders with

[1] According to Basie von Solms [78] "cybersecurity is a subset of information security, and therefore, cybersecurity governance is a subset of information security governance"

[2] https://gdpr.eu/

[3] https://www.pcisecuritystandards.org/

[4] government agencies, industry consortia, professional societies and others

[5] https://www.nist.gov/cyberframework

[6] https://www.cisecurity.org/

[7] https://www.iso.org/isoiec-27001-information-security.html

mandated compliances, although younger companies were more cybersecurity aware, [81] [82], and willing to implement one. For the current study is important to mention a major disadvantage of mandated certifications, as organizations may achieve compliance without necessarily having a robust enterprise systems security guideline to reduce risk,[83] [84] [85].

NIST Cybersecurity Framework

NIST Cybersecurity Framework is voluntary guidance, based on existing standards, guidelines, and best practices to help organizations to better manage and reduce cybersecurity risk[1] and increase their security posture. The framework was compiled in 2014, after the Cybersecurity Enhancement Act of 2014[2], and consists of three main parts: 1) the core framework, 2) the implementation tiers of the framework, and 3) the framework profile. While this framework was developed to improve cybersecurity risk management in Critical Infrastructure, it can be used by organizations in any sector[3]. The core framework establishes a common language for communication across an organization from the operation / implementation level to the executive level. It has a five-functions approach: Identify, Protect, Detect, Respond, and Recover as illustrated in Figure 5.

These functions provide a strategic view of the cyber risk exposure and cover the lifecycle of the risk management for the organizations. Within these five functions are a total of 23 activities or "Categories", under which numerous activities and controls ("Subcategories") are listed. The intent is that organizations will evaluate each of the controls within the Core against their assets and risk assess the tolerance, to arrive at their desired cybersecurity posture. The following Table 2 combines the basic definitions of each function with the contribution of AI in each one, according to Szychter [86]:

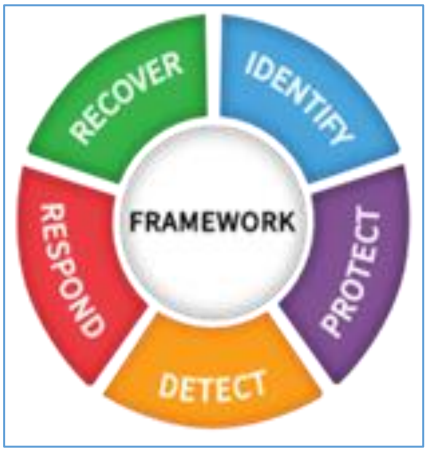

Figure 5. The 5-step approach of NIST Cybersecurity Framework. Image taken from the official web page [9].

[1] https://csrc.nist.gov/Projects/cybersecurity-framework/nist-cybersecurity-framework-a-quick-start-guide
[2] https://www.congress.gov/bill/113th-congress/senate-bill/1353/text
[3] https://nvlpubs.nist.gov/nistpubs/CSWP/NIST.CSWP.04162018.pdf

Table 2. How AI can be used in every step of the NIST Cybersecurity framework [86]

Step		NIST functions	NIST activity categories	AI contribution
1	Identify	Develops and organizational understanding of the IT assets, critical processes, threats, and risk exposure.	• Asset Management • Business Environment • Governance • Risk Assessment • Risk Management Strategy • Supply Chain Risk Management	AI-assisted risk analysis to determine the most valuable and vulnerable parts of a system
2	Protect	Develop and implement the necessary security controls and policies to ensure the service delivery	• Identity Management and Access Control • Awareness and Training • Data Security • Information Protection Processes and Procedures • Maintenance • Protective Technology	Pattern recognition and building effective counter-attack systems
3	Detect	Develop and implement the necessary tools to monitor and detect cybersecurity events	• Anomalies and Events • Security Continuous Monitoring • Detection Processes	Detect anomalies and signs of new unidentified threats
4	Respond	Develop and implement the necessary tools and plans to respond in detected cybersecurity events	• Response Planning • Communications • Analysis • Mitigation • Improvements	Assisting and training operators with AI
5	Recover	Develop and implement the necessary tools and plans to maintain resilience and to restore capabilities or services impaired due to cybersecurity events	• Recovery Planning • Improvements • Communications	

The publication of the NIST Institute, NIST.SP.800-207 [87], August 2020, on Zero Trust Architecture (ZTA), although does not refer directly to ML, it acknowledges that ZTA paradigm relies on continuous monitoring of all the organization's resources over the network, for evaluation and control of their behaviour. *"An enterprise implementing a ZTA should establish a continuous diagnostics and mitigation (CDM) or similar system."* [87], where the CDM functionality can be provided by NDR and relevant ML-based tools. The NIST guidelines and the ZTA practice are still not widely adopted by the community, but the executive order on cybersecurity[1] of the President of the United States, on May

[1] https://www.whitehouse.gov/briefing-room/presidential-actions/2021/05/12/executive-order-on-improving-the-nations-cybersecurity/

2021, is expected to accelerate the proliferation of ZTA paradigms at least for the large organizations in the United States.

CIS Security Controls

The CIS Security Controls, issued by the Center for Internet Security, started as an essential list of the most common and important steps that should be taken to defend against critical attacks. Over years and with the contribution of an international community of volunteer IT experts, the CIS Controls have matured to a prioritized set of best practices to mitigate the most prevalent cyber-attacks against systems and networks. The controls are mapped to and referenced by multiple legal, regulatory, and policy frameworks[1].

Cyber Kill Chain

Developed by Lockheed Martin, the seven stages Cyber Kill Chain framework is used widely in cybersecurity to assist practitioners to understand what objective adversaries need to achieve during their attack. It helps the defender retain a strategic overview of cyber incidents and identifies chain links in every stage to stop the malicious activity. It

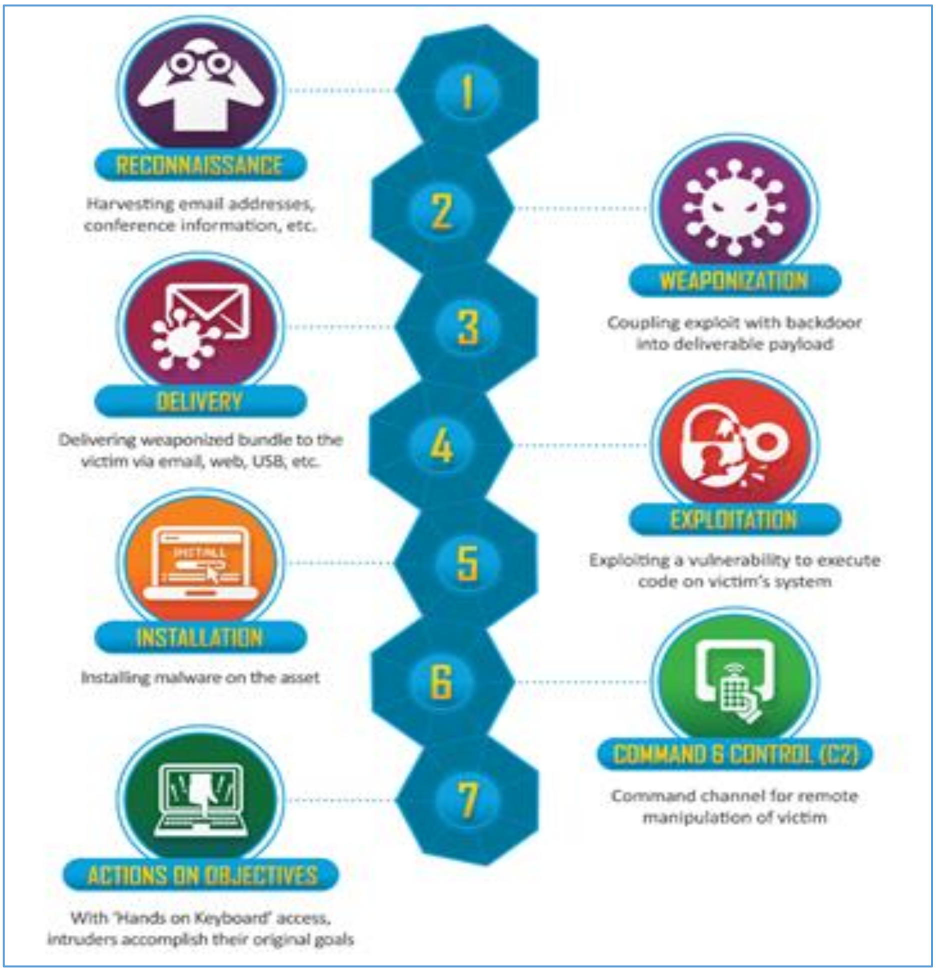
1
RECONNAISSANCE
Harvesting email addresses, conference information, etc.
2
WEAPONIZATION
Coupling exploit with backdoor into deliverable payload
3
DELIVERY
Delivering weaponized bundle to the victim via email, web, USB, etc.
4
EXPLOITATION
Exploiting a vulnerability to execute code on victim's system
5
INSTALLATION
INSTALL
Installing malware on the asset
6
COMMAND & CONTROL (C2)
Command channel for remote manipulation of victim
7
ACTIONS ON OBJECTIVES
With 'Hands on Keyboard' access, intruders accomplish their original goals

Figure 6. The 7 stages of the Cyber Kill Chain Framework

[1] https://www.cisecurity.org/controls/v8

assumes an organised activity on the adversaries' side and focuses solely on external attacks without addressing human errors or inside threats [88], Figure 6.

ISO/IEC 27032:2012

The ISO/IEC 27032:2012 standard was published in 2012 by the International Standards Organization (ISO) and is part of ISO/IEC 27000-series standards for information security. ISO/IEC 27032:2012 focuses on cybersecurity and provides guidelines and implementations, and dependencies on other security domains, as information security, network security, internet security, and critical information infrastructure protection (CIIP). ISO/IEC 27032:2012 is not part of a certification scheme but rather a framework for collaboration between different security domains in resolving cybersecurity issues.

ISKE

ISKE is an information security standard developed for the Estonian public sector that, since 2004, is compulsory for state and local government organisations that handle databases / registers. The standard is based on the German information security standard (IT Baseline Protection Manual or IT- Grundschutz in German) which was adapted to Estonian requirements. The goal of ISKE is to ensure a sufficient level of security for the data processed in IT systems. This is achieved by implementing three different sets of security measures (organisational, infrastructural/physical, and technical) for three different security requirements (different databases and information systems may have different security levels).[1]

GDPR

Although not directly related to cybersecurity but mostly focused on the protection of privately identified information, there is a need to mention the European General Data Protection Regulation (GDPR). As of May 2018, GDPR[2], sets a single set of rules for all companies operating with the personal information of EU citizens. GDPR regulates the processing and circulation of personal data related to natural and legal persons, identifying roles and responsibilities. The regulation requires organizations to demonstrate that they have embedded the principle of data protection by design and by default. For example, Article 8 requires that data controllers "*shall implement appropriate technical and organizational measures to ensure that processing of data is performed in accordance with the Regulation*". The regulation puts data security in a central role in every organization's cybersecurity policy as a wider compliance obligation (mainly considered in articles 5 and 32 of GDPR).

The reader may find a detailed list of Cybersecurity Standards and Frameworks and a relevant literature review and analysis in the work of Syafrizal, [82], in Table 6.

[1] https://www.ria.ee/en/cyber-security/it-baseline-security-system-iske.html

[2] European Union: Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), May 2016.
http://eurlex.europa.eu/legal-content/EN/TXT/?uri=OJ:L:2016:119:TOC

2.4 Literature research

The literature review on similar studies was conducted in two stages, in November of 2021 and in March of 2022. The review started, firstly by searching Google with main keywords such as "*Artificial Intelligence*" and "*Cybersecurity*" and combinations of words such as "*hardening*", "*implementation*", "*security standard*", "*NIST*", "*ISO*", etc. The same approach was followed in specific academic databases of Web of Science, ResearchGate, and finally Scopus. The initial abstract way also provided the author with a general understanding of the different research studies in the area and helped to form more concretely the related to the research questions cyber-risks.

The literature research proceeded to a systematic review through the online database Scopus. Firstly, a general pool of documents was created with the main keywords "*Artificial Intelligence*" or *"Machine Learning"*, and "*Cybersecurity*", in the fields of Article Title, Abstract, and Keywords. The initial list was shortened to results only in English and documents published since 2016. That returned a list of 3241 documents (March 2022) which were recycled in the following searches. On the initial collection of 3241 articles several research queries were applied as illustrated in Table 3:

Table 3. Literature research applied keywords' combinations, (*): not reviewed

Search queries		Results
0	**Initial query:** ({*Artificial Intelligence}* OR {M*achine Learning}*) AND ({cybersecurity} OR {cyber security}) published in English since 2016	3241*
1	Initial query AND ({cybersecurity framework} OR {cyber security framework})	52
2	Initial query AND (organizational AND aspect)	30
3	Initial query AND keyword 'NIST'	205
4	Initial query AND keyword 'interview'	40
5	Initial query AND keyword '27001'	11
6	Initial query AND keyword '27032'	2
7	Initial query AND keyword 'SANS'	91
8	Initial query AND keyword '270'	53
9	Initial query AND keyword 'ISO' excluding keyword 'NIST'	60
10	Initial query AND ({SOC} OR {security operation centre})	235

Except for the results of the initial query, the rest of the results were reviewed twice. In the first review, as the focus was on the novelty of the research, the exclusion criteria were the similarity or not with the research questions. The review included the article title, the abstract, and in some cases the paper if it was not clear already. The review of the paper was in the order of abstract – introduction - conclusion or discussion – research method – other interesting parts of the document. For some of the articles, such as the works of Nespoli [89], Meier et al. [38], Oesch et al. [19], Rawindaran [24], Akinrolabu et al. [20], the review included the citations and the references too. The second review had two objectives. To verify the result of the first and to create a collection of articles that support the considerations of the study.

Several researchers have performed in situ surveys with cybersecurity practitioners on the use of AI solutions in cybersecurity. Oesch et al. [19] have performed a usability assessment of two ML-based tools using SOC analysts. Although their research focuses on usability issues, the authors discovered that analysts lacked a clear mental model of how these tools generate scores, resulting in mistrust and/or misuse of the tools themselves. *"The benefit of ML is lost if analysts cannot understand the meaning of the scores produced"* [19].

Akinrolabu et al. [20] have also interviewed experienced SOC analysts to better understand obstacles to detecting sophisticated attacks. Their research focused on how the analysts handle the large amount of false-positive alerts from the existing ML-based tools. Their findings support the notion that malware detection should not be based solely on individual events but on correlated network and application traffic in the IT system. In their document, the authors suggest that involving the user in the creation of the ML model can provide significant benefits.

Salitin and Zolait [21] performed a self-administrated survey and interviews to assess the value and success of using UEBA analytics in securing the network from not-before-seen attacks, and to verify the effectiveness of the said solutions based on behaviour analytics.

Gutzwiller, Hunt, and Lange [22], focused on the human cognitive awareness in cyberspace. They have conducted a cognitive task analysis (CTA) to determine the goals and abstracted elements of awareness that cyber analysts seek in order to perform their tasks. They proved that data visualizations are most useful when combined with situated knowledge of the network from the analyst to make accurate decisions.

Griffy-Brown [23] has researched whether the organizational risk is being considered or evaluated from a governance perspective within an organization when deploying emerging technologies like AI. Based on interviews with executives facing these new environments, they have concluded that firms need support in understanding and making decisions around the risk engendered in the emerging technologies they deploy.

Kumar et al [25], have interviewed industry practitioners in adversarial ML. Their research found that ML engineers and cyber incident responders are "*unequipped to secure industry-grade ML systems against adversarial attacks"* [25]. According to the researchers, they lack the tactical and strategic tools to protect and detect attacks on their ML systems.

Masombuka [27] in his thesis performed an empirical study on cybersecurity practitioners applying ML in cybersecurity but his focus had been mostly on the applied methods when deploying the technology and the technical outcomes of the deployments.

Rawindaran et al [24] have also researched the use of ML tools in the area of cybersecurity in SME enterprises in the UK. Their research has been focused on the awareness of the SMEs about ML, the factors that contribute to adopting such solutions, and a comparison between outcome and implementation costs for three specific detection tools with ML capabilities.

Malatji et al. [26] have researched the impact of AI on the human aspects of information and cybersecurity. Their findings revealed that AI is currently utilised only for augmenting human capacity in information and cybersecurity activities whereas the future trends are unknown [26].

In broader use areas of AI, as in Government-to-Citizen e-services, Dreyling et al.[28], in their empirical research have explored a particular method for determining cybersecurity risks for a virtual assistant enabled in government service.

The above-identified studies have focused beyond the AICS technologies, the technical outcomes, and the human aspects of the cybersecurity teams, also on the evaluation of the organizational risk when deploying AI technologies. This research aims to explore the impacts of the AICS technologies on the organization and management of cybersecurity. Considering the inadequacy of the cybersecurity standards and frameworks, in their current versions, to directly address the threats to AI technologies, it is the author's belief that a gap exists in the literature. This research aims to explore the approaches chosen by the cybersecurity experts to secure their AICS deployments and the impact on their procedures and processes and their overall cybersecurity policies.

3 Research Method

The chosen research method for this study was to conduct interviews with the end-users of the said technology. Surveys are suitable for gathering structured information but when you are interested to collect the impressions, thoughts, and experiences of individuals or groups, interviews are a better collection tool [90]. Hence, the study was performed as empirical research, based on a qualitative study of semi-structured interviews with open-ended questions. A questionnaire, Appendix 3 - Questionnaire, was used during the interviews with the participants of the research, to direct the conversation, collect answers to common questions, and encourage the interviewees to share their insights and indicate interesting points according to their experiences. To enhance the interview process, the informants were presented with an introductory summary of the concept of this study, Appendix 2 - Introductory note sent to the interviewees. The summary included an overview of the classification of the AI systems as described in the literature, a description of the cyber-risks describing the research's considerations, and explanatory notes on the questionnaire.

To be better prepared for the interview, the author asked the participants beforehand to indicate if they already use an AICS solution and the vendor of the product.

3.1 Interviewees' selection

According to the definition given by Bogner [91], an expert is a person with specific practical or experimental knowledge about a particular problem area or subject area and is able to structure this knowledge in a meaningful and action-guiding way for others. Participants for the interview process were selected according to their experience in the field of cybersecurity and their functioning role within their organization. The author, through recommendations by the University and personal contacts, approached 15 companies and organizations, 11 in Estonia, 3 in Greece, and 1 in Finland, that operate cybersecurity teams and contacted leading officers as Chief Information Security Officers, Chief Information Officers, or cybersecurity experts in cybersecurity architecture. Additionally, the author also approached researchers in the field of AI with lengthy experience in cybersecurity.

3.2 Qualitative interview

The questionnaire was designed on the basis of the research questions in two versions. The first version formed the questionnaire sent to the participants accompanying the interview invitation and the introductory summary, Appendix 3. An effort was put to include a minimum number of questions as the aim was to provoke an open conversation. The second version included a list of targeted keywords and topics, identified by the literature research, and was used by the author to facilitate the interview process. In the design phase of the interview, a desk review was performed by the author to identify key concepts of the research and to properly formulate the questionnaire and the interview process. The questionnaire was tested in a pilot interview with a cybersecurity expert and AI researcher (participant P1 of the research), experienced with SIEM tools and cybersecurity management. The pilot interview provided insight into the value of the targeted keywords and topics and helped the author to adjust them.

3.3 Research limitations

The timeframe of the thesis preparation was a limitation for the research as a certain number of interviews was possible to conduct (in total 11 organizations in Estonia, 3 in Greece, and 1 in Finland were approached). The war in Ukraine that started in March 2022 and the escalated hybrid warfare of Russia in the cyberspace, affecting businesses, especially in Estonia, had been a negative factor too, as cybersecurity teams and experts approached by the author were extremely occupied during the research period. The author was able to conduct 7 interviews with 10 participants. Of them, two experts have experience in deployments of AICS solutions for their customers functioning as *resellers*, seven are experts in cybersecurity and have experience in using AICS solutions as *end-users* of the technology, and one of the participants is specialised in information security and assurance.

3.4 Analysis method

Thematic analysis, as described by Braun and Clarke [92] and Maguire and Delahunt [93], was employed to analyse the data set of the interviews. With the informants' consent, the interviews were audio-recorded, and transcribed during the analysis of the results. Braun and Clarke [92] propose data analysis should follow six main steps summarised as in Table 4:

Table 4. Phases of thematic analysis according to Braun and Clarke [92].

1. Familiarizing yourself with your data:	Transcribing data (if necessary), reading and re-reading the data, noting down initial ideas.
2. Generating initial codes:	Coding interesting features of the data in a systematic fashion across the entire data set, collating data relevant to each code.
3. Searching for themes:	Collating codes into potential themes, gathering all data relevant to each potential theme.
4. Reviewing themes:	Checking if the themes work in relation to the coded extracts (Level 1) and the entire data set (Level 2), generating a thematic 'map' of the analysis.
5. Defining and naming themes:	Ongoing analysis to refine the specifics of each theme, and the overall story the analysis tells, generating clear definitions and names for each theme.
6. Producing the report:	The final opportunity for analysis. Selection of vivid, compelling extract examples, final analysis of selected extracts, relating back of the analysis to the research question and literature, producing a scholarly report of the analysis.

The order of the steps should not necessarily be followed linearly with Braun and Clarke [92] proposing two different approaches to the process of searching for the themes. A top-down theoretical analysis, driven mostly by the research questions and using the latest as the initial themes to label the data with various codes during the analysis. Or a bottom-up inductive analysis driven by the data in the dataset and labelling the whole dataset with various codes and extracting the major themes in the later stages of the analysis process. The author of the this study chose the second approach as he sensed that labelling data prior to themes' search and grouping, is the more suitable approach when interviews include open-ended questions, as any implied meanings could be better understood by the analyst. Additionally, by using the research questions as the main themes to analyse the data, one risks simply summarising and organising the data set rather than analysing it [92].

Following the above framework proposed by Braun and Clarke [92] and further to the interviews' transcription, the author started organising the data in a systematic way with the use of excel worksheets. The transcripted texts of the interviews were corrected from mistakes created by the automatic transcription[1], scattered phrases were collected in distinctive paragraphs of questions and answers, transformed the text into a two-column table, keeping the timestamps in the first column and the actual text to the second, and transferred it into an excel sheet. For a bottom-up inductive analysis of the text, the author went through the interview's text, and for every interesting answer he created a short highlighted topic in the adjacent cell. If the answer highlighted more than one topic, a separate note was made in another adjacent cell. After this process, the author went through the list of highlighted topics and codified each row in one or more *first-layer themes*. Having done so for all the transcriptions, the highlighted topics were collected into a single excel sheet, by keeping the topics of each participant in separate columns and the highlighted topics of similar *first-layer themes* on the same row. Further analysis of the results was done by moving rows of similar *first-layer themes* together and grouping them into *second* and *third-layer themes*. The last phase of the analysis was to combine the themes that emerged on the *third-layer* grouping with the research questions of the study and highlight the similarities or differences of the participants' responses.

3.5 Ethics

Aware of how sensitive security issues are and the fact that the topic of the current research touches on core elements of an organization's security architecture, the author was extra cautious regarding the ethics of the research, the confidentiality, and the anonymity of collected information. Prior to the interviews, participants were informed about the author's status as a student of Tallin's Technical University, and a Certification of Studies was sent together with the interview's invitation. Participants were asked to consent to the recordings prior to the interviews. All the collected material (recordings and transcriptions) were anonymous and remained anonymized during the analysis and the final presentation of the results. Any information given to the author with the request to remain anonymous, was anonymised for the analysis and the presentation of the results. By the end of the research and the final presentation of the master thesis project, all related files will be deleted, and the storage media will be securely sanitized according to NIST guidelines [94].

[1] Microsoft Word 365, online version, "Transcribe" feature

4 Results

This chapter provides an analysis of the data collected from the performed interviews with the cybersecurity experts and explains the insights of the qualitative research.

The interviews with the experts, were conducted with the use of a semi-structured questionnaire, as explained already in the previous section. With regards to the limitations mentioned before, a total of seven interviews were conducted with ten participants. The author conducted the interviews, five through web meetings, two physically, and one participant submitted his answers to the questionnaire by email. Of these, the first interview served as a pilot interview for the questionnaire and the interview process. The interviews were conducted in Greek for participants P1 and P6, and in English for the rest. Out of the seven interviews, two of the participants, P3 and P6, are in the business of providing cybersecurity services and AICS solutions to their customers, have experience with AICS deployments, and P3 even uses the same solution for the protection of his company too. For the presentation and analysis of the results, P3 and P6 participants, are characterized as *resellers* while other participants as *end-users* of the technology.

Information on the participants' profiles, the interview process, their experiences with AICS, and their contribution to the research:

- Participant P1, is an AI researcher located in Estonia of Greek origin. He is experienced in SIEM tools and cybersecurity management and has a previous experience with a deployment of an AICS solution for a government agency in Greece. The interview with P1 was physical, conducted in Greek, recorded with the participant's consent, the audio file transcripted in Greek, and the thematic analysis conducted in English, as explained in the previous section. Being the first interview, helped the author to identify complex questions and prepare the keywords with more details for the next interviews. The participant was approached also with a later email for additional information.

  His experience is with the software Enorasys SIEM system[1], a next-generation SIEM tool that provides real-time security intelligence, based on IBM's QRadar ML component. The tool was deployed in the production environment for monitoring and detecting anomalies in web traffic. The deployment was performed by the vendor, as the modelling phase, over a period of several months and the participant has the experience of overseeing the deployment and operating with the tool later.

- The second interview was with two members of a large educational institute in Estonia. The interview was physical, conducted in English, recorded with the participants consent, transcripted, and analysed thematically as above. Both participants have years of experience in cybersecurity with P2a participant being the CISO officer of the campus and P2b a member of the academia, professor in

[1] https://www.encodegroup.com/product/enorasys-siemt

the cybersecurity department, he has working experience in large financial organizations, and author of one of the AICS tools used currently by the institute.

The institute operates two AICS tools. The first, Microsoft Sentinel, comes as a part of the institute's subscription to Microsoft Cloud services, and the second, SCAS, is an open-source stream clustering algorithm developed by P2b participant as part of a research initiative. The participant was approached also with a later email for additional information.

The Microsoft Sentinel solution provides EDR functionality to the institute's IT systems but is provided as a service for the organization with the security team having limited ability for modifications. On the other hand, on the SCAS tool, the participants have full control, and the tool is employed for the analysis and classification of network IDS alerts. SCAS assigns scores to certain alert groups and provides context and prioritization to the security analysts. The tool operates in the production environment for over two years and P2b participant has been involved in the installation and fine-tuning of the tool's numerical parameters.

Participant P2a, as CISO of the campus, provided more insight into the Microsoft Sentinel EDR solution and the organizational aspects of the campus cybersecurity, while participant P2b provided insight on the SCAS solution.

- Participant P3, works as a cybersecurity expert for an Estonian MSSP and has experience with the design and operation of SOC. The interview was conducted in English through a web meeting, recorded with the participant's consent, transcripted, and analysed thematically as above. The participant was approached also with a later email for additional information.

  He has experience with EDR, NDR, and SIEM tools and since September 2021, together with his security team, has tested and deployed in production an ML-based component for their SIEM platform that is based on IBM QRadar Advisor. He has the experience also of running the tuning phase for a period of one month. The tool's functionality is the analysis of user's behaviour, UEBA, through the log files collected by the SIEM platform. The said tool is used also by the company's SOC for its own use and for the managed services offered to their customers.

- Participant P4, is the CISO officer of a high technology and blockchain Estonian company. He has more than 23 years of experience in cybersecurity, mostly in banking institutions, prior to his current position. The interview was conducted in English through a web meeting, recorded with the participant's consent, transcripted, and analysed thematically as above.

  From his current position, participant P4 has experience with an endpoint solution in EDR functionality[1] that employs ML capabilities and has performed a proof of concept (PoC) with Darktrace software for several months but without proceeding to final deployment. Darktrace was tested by monitoring the network aggregation points and the internet gateway to detect external intrusion. Due to the

[1] The name of the tool is not disclosed as of participants' request

organization's security architecture, the participant with his team, tested the tool separate from the production environment without integration. They acknowledged to the tool the legitimate traffic and the network's core elements and they simulated malicious traffic with an external command-and-control centre. In their experiments, Darktrace failed to distinguish the legitimate traffic continuously and to detect the intrusion passing through its sensors.

- Participant P5, is the head of InfoSec of a major cloud-based software development Estonian company. The interview was conducted in English by email, transcripted, and analysed thematically as above. The participant was approached also with a later email for additional information. Unfortunately, the information collected from participant P5 has been limited.

  The participant has the experience of testing an AICS tool[1] into his organization but without proceeding to final deployment. The tool was tested on network flow analysis, identity and access management, IoC detection, and access and data extraction. In the test run, the tool returned a very high rate of false-positive alerts, faulty findings and detection rates of possible IoC's. In his remarks, the participant pointed out that he stands negative on his experience with the technology and considers that currently similar tools are not practical or beneficial to his organization, due to their size, risk profile, and structure.

- Participant P6, is a cybersecurity expert, located in Greece, with years of experience as a security architect in major relevant companies in Greece. The interview was conducted in Greek through a web meeting, recorded with the participant's consent, transcripted in Greek, and analysed thematically in English, as above.

  He has experience in multiple deployments of at least two AICS tools offered by major international vendors[2] and is currently employed by one of the said vendors in Greece. He presented his opinion from the reseller's / installer's side, having a collective experience with deployments of EDR, NDR, and SIEM tools in large scale enterprises and organizations. He provided valuable insight also to one of the major problems described by the other participants, the large rate of false-positive alerts, especially during the modelling phase.

- The seventh interview was with three members of a European agency, having its headquarters in Estonia, Tallinn, and the operational sites in France, Strasbourg. The interview was conducted in English through a web meeting, recorded with the participants consent, transcripted, and analysed thematically as above.

  All three participants have experience in information security and cybersecurity. Participant P7a is the Head of the organization's Secure Operations Sector, within the Security Unit, participant P7b is the Head of Information Security and Assurance sector, and participant P7c is an AI researcher working close with the in-house SOC team of the organization.

[1] Tool's name was not provided by the participant
[2] Fortinet's FortiEDR and Cisco's StealthWatch

> They presented the side of a large organization, spanning in multiple countries and operating complex IT systems, with multiple stakeholders. The organization for the last seven months has implemented an AICS tool into the production environment and is currently evaluating the first use case. The AICS tool comes as an ML component of the existing SIEM platform, Splunk, and is employed in behaviour analysis of the same log files fed to the SIEM. Its functionality is to provide judges and prioritization of alerts to the cybersecurity analysts of the organization's SOC. The organization is also preparing a PoC on the use case of detecting suspicious outgoing connections with the ML tool.

Table 5 summarises the above information of the participants and provides a comparative overview to their profiles:

Table 5. Summarised information about the participants and their experience with AICS.

Participant	P1 (Pilot interview)	P2a & P2b	P3	P4	P5	P6	P7
Industry of the participant	Government agency	Educational institute	MSSP	High technology and blockchain company	Cloud app and software development	Cyber security services	European agency
Participant has experience on AICS tools as	*End user*	*End user*	*Reseller & end user*	*End user*	*End user*	*Reseller*	*End user*
Participant's opinion is based on single or multiple deployments	Single	At least two	Single	At least two	Single	Multiple	Single
Use of commercial or open-source tool	Commercial	Commercial & Open-Source	Commercial	Commercial	Commercial	Commercial	Commercial
Name of AICS tool used by the participant	Enorasys SIEM platform that includes ML-based tool based on IBM QRadar Advisor	Microsoft Sentinel[1] & Open-Source SCAS	SIEM platform that includes ML-based tool based on IBM QRadar Advisor	Darktrace	*Not provided*	FortiEDR & Cisco Stealth Watch	Splunk SIEM platform that includes ML toolkit
Is the tool deployed in production environment	Yes	Yes	Yes	No	No	Yes	Yes

[1] As part of participants Microsoft Cloud services' subscription with limited ability for modifications, the research focused mostly on the second tool used by the participant

Participant	P1 (Pilot interview)	P2a & P2b	P3	P4	P5	P6	P7
Time in production environment	*Not provided*	2 years	Since September 2021	PoC for several months	PoC only	*Not applicable*	7 months
Functionality and use cases of the AICS tool	Traffic monitoring Anomaly detection of web traffic	IDS alerts monitoring Analysis of IDS alarms Prioritization by assigning scores to certain alarm groups	UEBA (User entity behaviour analytics)	Monitoring of network aggregation points Detection of external intrusion	Network flow analysis IAM Management & control IoC detection Access & Data extraction	NDR EDR SIEM	Behaviour analysis of the feed sent to the SIEM platform

4.1 Introducing experts' opinion

4.1.1 On the motivation to use AICS

Handling the large number of alerts and the imposing workload to the cyber analysts had been pointed as the main motivation for the introduction of AICS tools by the participants. As is pointed by Kokulu et al [4], Chowdhury et al [95], and Akinrolabu et al. [20], cyber analysts today need to cope with a plethora of alerts generated by various systems making it difficult for them to correlate the events and focus on the bigger picture. As pointed out by participant P4, *"...old fashioned approach is touching the limits of human capability to work with events and grasp the bigger picture. So, we are on the limit, if not beyond already, we might miss something in our current setup in current infrastructure"*. AICS tools have the ability to correlate present and past events from different sources with external information (e.g., threat intelligence, geolocation, etc.). This provides the cyber analysts why valuable context when investigating the cyber incidents and as pointed out by participant P7c assists the analysts in triaging and focusing on real malicious events (true positive). Seven of the participants consider AICS as an enhancement to the organization's cybersecurity posture while participant P7a emphasised the improvement of the analysis capacity beyond the human limits. Participant P7c marked also that SIEM's static correlation rules are inadequate to detect real events and adapt to the dynamic behaviour of the adversaries. In support to the above, participant P6, stressed that AICS tools provide organizations with *application visibility and control* (AVC) over their networks. The shift to cloud-based applications has transformed the network traffic to mostly HTTP-based (ports 80/443), as illustrated in Figure 7 from Juniper Networks[1]. Further, nowadays most of the users' traffic is encrypted with the use of encrypted pages

[1] https://www.juniper.net/us/en/research-topics/what-is-application-visibility-and-control.html

in Chrome browser being over 90%[1] on Windows and Android devices. Applications e.g., for instant messaging and social networking change dynamically their protocols and ports or use tunnelling through commonly used protocols like HTTP or HTTPS. Attackers take advantage of this landscape to obfuscate their attacks and their malicious communications. Decrypting the traffic for inspection is not always feasible due to performance or regulatory impact. NDR tools based on ML, inspect the traffic metadata from routers and switches, and correlate threat behaviour in the organization with global threat intelligence.

In regards to the selection of the AICS tool deployed, previous experience with the same vendor was the main factor for participants P1, P3, and P7c, as the ML tool was provided as an additional component of their existing SIEM platform.

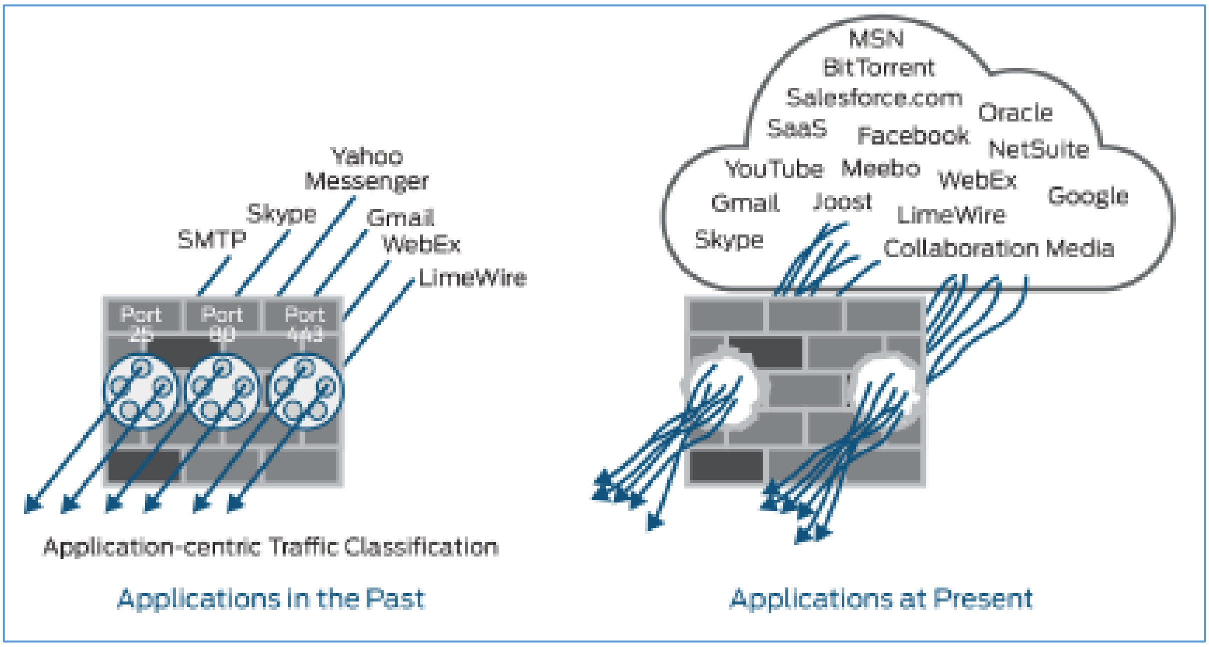

Figure 7. Juniper Networks, Learn about AVC, Applications Landscape – Past and Present.

4.1.2 On the initial period of training and modelling the AICS

As explained by participant P6, AICS tools are allowed a period to self-teach themselves in the operational environment. The learning process is mostly unsupervised with the installer checking regularly and providing input to the tool about legitimate traffic and infrastructure's core elements such as DNS servers, DHCP servers, Domain Controllers etc. The duration depends on the operational environment and the specific use case of the tool, and can last even a year for large and complex IT environments. During this time the tool creates the baseline configuration, and the installer fine-tunes the system's thresholds that trigger the alerts according to the desired outcome and the organization's policy. It is therefore common to have a high rate of false-positive alerts during the training period but for some of the participants this has been discouraging. For participant

[1] Google Transparency Report, *HTTPS usage in Chrome worldwide*, January 2022

P5 the high false-positive rate, together with faulty findings of possible IoCs, was a reason to reject the technology for the moment. Participant P3, who operates a SOC in Estonia and has been working with AICS on analysing users' behaviour, remains sceptical about the tool's capability in his current setup. Participant P1 experienced a similarly high rate of false-positive in the training period but according to his use case, detected anomalies beyond the baseline configuration remain true positive although they might originate from legitimate traffic or events[1].

The use case of the AICS seems to determine the importance of the false-positive alerts for the analyst and the future usage of the tool. For participant P2b, who uses AICS to analyse and prioritise IDS alerts, the high rate of false-positive can be explained also by the evolving activity of botnets in the outer perimeters of the networks. IDS correctly responds with alerts to this malicious activity but as some of them are old and search for outdated vulnerabilities, they represent a harmless activity to an updated system. AICS tool will correlate and prioritise these events and for the training period, the scoring will be high. But as long as the analysts do not react and the botnets' behaviour does not change, the tool assigns a behaviour pattern with low scoring to this activity. In other words, the activity remains present in the network and visible to the IDS and the AICS tool, but is assigned a low priority and constitutes a residual noise for the AICS.

Participant P7c approach to tool's training is different due to the dynamic IT environment of his organization. He considers the training period as an ongoing process for the tool, which is done on a daily basis in order to keep his baseline configuration updated.

4.1.3 On the learning method of the ML component

The proposed learning method for the AICS tools is unsupervised as the ML component is intelligent enough to discover the necessary information from the operational environment. Depending on the employed functionality and the environment, this process can be lengthy in time. To accelerate the learning process, the core elements of the infrastructure can be provided to the tool. During the training period, additional fine-tuning of thresholds and numerical parameters can be provided to the tool regarding the legitimate traffic and events. According to participant P6, in practice the tool discovers by itself 90% of the necessary information about the operational environment while the rest 10% is provided as fine-tuning from the installer in regular periods after the tool has developed the baseline configuration. Both participants, P2b and P6, are strong supporters of unsupervised learning and the least effort put into the deployment of AICS.

The notion of the least effort put into the deployment phase seems to be a confusing point for the *end-users* of the technology. For participant P2b in order to provide supervised learning to the ML, there is the need for labelled datasets which, and especially in the field of cybersecurity, are difficult to craft, as pointed out by Ibrahim et al. [61] and ENISA [72]. According to the experience of participant P6, his clients expect the tool to *magically* teach itself and discover the adversaries in record time with the least effort. But as P6 pointed it, it needs time to "*train the algorithm and the end-user of the technology too".* As application visibility being one of the advantages of AICS tools, typically the

[1] A simple example from the participant for his point, is the case of a user checking his work email on the weekend, outside of his usual working habits. Although for the user is a legitimate action for the AICS remains an anomaly, different to the established behaviour for the user.

organizations lack a full picture of their IT environments in terms of applications used and generated traffic. This can be due to malicious use but in most of the cases is legitimate, but simply unknown to the operators.[1] If not acknowledged differently by the analyst this traffic will remain an alert for the tool leading to false-positives alerts. On the proviso of the use case for the tool, the alerts described as false-positive by the *end-user* may constitute true-positive alerts for the AICS tool representing simply unknown so far traffic or applications due to the lack of visibility to the organization's IT environment by the analysts.

4.1.4 On the case of deploying AICS in a compromised network

Apart from participant P4 who performed the PoC for the software of Darktrace in a test environment, the other participants deployed the AICS tool into their production environments. On the author's question on the hypothesis that the environment is already compromised and the possibility of ML characterising the malicious activity as legitimate, participants' answer was based on the functionality of their tools. As they have deployments solely in supportive roles that provide alerts to the analysts the risk is low. They acknowledged, especially participants P1 and P3 who employ AICS for anomalies detection, the possibility of malicious activity being learnt as legitimate by the ML and registered in the baseline configuration. Participant P2b, who operates AICS for analysing IDS alerts, expanded the case, so that any malicious activity ongoing over a long period will be characterised as normality or given low scoring by the algorithm, if the security analysts do not mark it differently. To his opinion, this poses a level of risk although low. As long as the security analysts decide not to escalate it to an incident, and the behaviour pattern of the activity does not change, the alert gets a low score of priority and remains as background noise, and the algorithm is simply unanimous with the analysts' actions. Any changes to the behaviour pattern will be re-evaluated by the tool as a new behaviour.

For participant P7c one should assume always that he is compromised but at the same time he should try to identify actions that are beyond the established baseline configuration. Analysts should try to identify the adversaries on the next step of their attack, e.g., a lateral movement, that will trigger a violation on the baseline configuration of the tool. As he remarked: "*you need the adversaries to make one mistake and then you can go back and close the chain of the attack".*

4.1.5 On the case of disabling the learning process to protect the algorithm

Participant P6 presented a different approach through his experience, that of disabling or degrading[2] the learning process of the ML algorithm once the baseline configuration has been established. The reasoning behind his approach is to protect the algorithm against manipulation by the adversaries. In a compromised environment, the final baseline

[1] Participant P6 provided an interesting example of legitimate but unknown traffic created by modern printers and scanners. As most of the printing services nowadays are provided in the form of leased machines, it is common practice for the companies operating these machines to collect information about the machine consumption by regular reporting sent by the machines back to the company's servers. This traffic is mostly unknown to the installers and the system administrators but can easily be detected by NDR tools.

[2] Participant P6 pointed out that the process can be done in several ways, either by disabling the learning process or by raising the existing thresholds. Further, the practice can be enforced partially for certain network segments or applications.

configuration after the deployment and tuning phases, will have recorded the malicious activity as normality. If the attackers have the knowledge of the existence of an AICS tool they will attempt to misdirect it before attempting to launch their attack or change their behaviour. To achieve that they will escalate their malicious activity gradually over a long period raising the thresholds one at a time. With this practice, the algorithm will constantly record the minor malicious activity as new normality, and in the long run, the adversaries will remain undetected and will succeed to train the algorithm to their needs before performing their attack. By locking the baseline configuration to higher thresholds against these minor changes, malicious activity can easily be detected.

On follow-up communication with the other participants, participant P1 expressed the opinion that this practice poses the risk of recording malware activity as legitimate and remains undetected as long as the adversary does not change his behaviour[1]. Participants P3 and P7c pointed out that the approach is not valid in their case since the operational environment of their AICS implementation changes on a regular basis as they are onboarding constantly new log sources. Further, they questioned how this practice can keep the algorithm updated with the new environmental changes. Participants commented also that such practice is not recommended by their vendors, too. Participant's P2b opinion was that automatic adaptation of the tool is possible only if the algorithm is continuously learning all the new regularities that appear among the IDS alerts.

4.1.6 On keeping the system updated and informed with the latest threat intelligence

Depending on their use cases and the AICS tool used, participants follow different approaches for keeping the AICS tool updated and informed with threat intelligence. For the commercial solutions the system's updates are provided by the vendor and the *end-users* have a change management process in place. In the case of participants P1 and P2b, due to their use case, there is no need for threat intelligence information. For participant P1 the tool must *"detect the unknown - the anomaly"*. And for participant P2b the threat intelligence is provided to the IDS that generates the alerts. For participants P3, P5, and P6 the threat intelligence is provided by the vendors, e.g., FortiGuard Labs[2] from Fortinet or TalosIntelligence from Cisco[3], and other commercial sources and CERT feeds. Participant P7c for the moment is receiving open-source feeds from MISP project[4] but is in search of a professional tool. As explained the malicious feeds can contain extended lists of malicious IPs, URLs, domains, file hashes, or applications, and in the case are fed directly to the AICS tool, are used complementary by the ML for cross-checking the alerts.

In the context of enhanced visibility offered to the organizations by the AICS tools, participant P6 pointed out the aspect of keeping the system informed also with the organization's security policy. Therefore, discovering what applications and services are used in an IT environment and restricting those not desired by the organization such as e.g., crypto miners, due to performance impact or internal regulations.

[1] Which might not happen if the attack is already successful in its objectives!

[2] https://www.fortinet.com/fortiguard/labs

[3] https://talosintelligence.com/

[4] https://www.misp-project.org/

4.1.7 On the ML competence of the security analyst

All the participants expressed the opinion that no ML competence is required by the *end-user* of the AICS tool, the cybersecurity analyst or expert. Although, it was expressed that a general understanding of how in principle the software functions and how the set thresholds are triggered, can be beneficial to the analyst. Participant P6 stressed that analysts should be able to understand the impact of the alerts and their correlation with the security architecture in place. Participant P4, although he remains in the stage of PoC, does not foresee such a need and expects the final product and its vendor's support, to determine the need for a specialist in ML on the premises.

4.1.8 On the black box reputation of the AI

Participants' opinion on the fame of AI applications being characterised as *"black boxes"* was that typically vendors are reluctant to share information and details on their software. On the concept that extra features or information could have been helpful to the analysts, especially when in need to question the tool's decisions, they stand that does not affect the tool's usability and their tasks in general.

4.1.9 On the interpretability and explainability of the AICS

Delving further into the concept of interpretable and explainable AI, participants agreed that as long as the tool provides evidence of which signature / threshold violation has triggered the alerts in correlation to the corresponding input, the tool is sufficient interpretable to them. As participant P4 phrased it, *"a certain level of understanding on how it works, operates, and how attacks are identified is needed".* Participant P6 remarked that according to his experience and the specific tools he has worked with, for any event characterised as an anomaly by the algorithm, the analyst is provided with the relevant information of the input that triggered the alert and the incompatibility to the baseline behaviour e.g., the violated thresholds. Participants P3 and P7c, for whom the AICS tool is provided as an extra component of their existing SIEM platform, consider an advantage to the interpretability of the tool, the integration with the existing SIEM platform.

4.1.10 On the trustworthiness of the AICS

Regarding the establishment of trust in the AICS tool, participants expressed various opinions. Participants P1, P2b, and P3, marked that trust can be achieved if the analyst understands the tool's functions and if the tool is interpretable enough to explain which threshold was triggered or what anomaly violated the baseline configuration. Both stressed that trust is built over time confirming the previous remarks of participant P6 about the need to train the analyst on the tool also (paragraph 4.1.3). Participants P4 and P5, who did not proceed beyond the test phase, agreed that a general understanding of how the tool functions and how the attacks are identified is needed to achieve a level of confidence. Beyond that, as for any other system in their IT environment, additional monitoring should be in place for the AICS tool also. Participant P7c connected trust to the tool with the false-positive alerts stressing that carefully crafted log files, that feed the AICS tool in his case, can significantly reduce the number of false-positive alerts and assist in the establishment of trust in the tool. Participant P6, with the experience of the installer, acknowledged vendors' reluctance to provide detailed information about their models, as explained earlier, but he pointed out that at the same time vendors provide tested and detailed use cases that cover *end-users'* needs and the majority of the deployed scenarios.

4.1.11 On the responsibility in case of failure

Kumar et al [25] in their survey of industry practitioners in adversarial ML discovered a mismatch between reality and expectations by the security analysts. Security analysts, accustomed to traditional software practices, expect that AI algorithms available in platforms such as Keras[1], TensorFlow[2], or PyTorch[3], are secure and stress-tested against adversarial manipulations. Kumar even discovered that organizations push the security responsibility "*upstream*" [25] to the MSSP providers and expect them to provide robust algorithms and platforms.

Taking into account the inherent complexity of AI/ML algorithms, the lack of transparency as explained before, the dynamic nature as they self-adopt and self-teach themselves into the operational environments, the ever-evolving state of the algorithms, and the additional complexity of being trained according to the *end-users'* feedback, the author addressed a similar question to the participants. What are their expectations regarding the robustness of the systems when so many factors are involved in the systems' knowledge and performance, and who they would consider responsible in the case of a failure?

Participants P2a, P3, P4, and P6 tried to come up with an answer but with no concrete justification. Depending on the use case of their tool, their opinions shifted between being shared with the vendor or solely theirs. Participant P3, with the identity of the MSSP provider, shifted to also include the customer too[4]. Participants P1, P2b, P5, P7a, P7b, and P7c did not participate in this question.

Considering the proliferation of AI technology the author believes that this topic needs further investigation with separate research in the future.

4.1.12 On protecting the AICS

All the participants expressed the opinion that no specific security controls are needed for the AICS tools providing the organization has already established concrete cybersecurity practices. For the participants, the attack surface is considered minimal and the risk relatively low[5]. According to his own cybersecurity practice, participant P3 commented that the concept of layered security applies here too. Participant P5 commented that treatment can be similar to the secure engineered VMs used by his organization already. Participant P6 noted also the multi-layered approach with a combination of firewall rules, UTM rules, and layer 7 inspection. He added that his recommended practice is keeping the main component of the tool, the supervisor, isolated, distributing data collectors and sensors in the IT environment, and allowing the communication of the supervisor only with the distributed sensors. For participants P1, P3, and P7c, to whom the AICS tool is

[1] https://keras.io/

[2] https://www.tensorflow.org/

[3] https://pytorch.org/

[4] The author will point to the work of Akinrolabu et al. [20] on the common failures of SOC and the shared responsibility of the customers to provide valuable data

[5] since the AICS's functionality is the analysis of log files and alerting

integrated into their SIEM platform, the security controls cannot be different, and they follow the same cybersecurity approach as the rest of the core elements of their infrastructure.

4.1.13 On the impact of AICS on the organization’s cybersecurity policy

As explained above, participants agreed that AICS has an impact on an organization’s cybersecurity architecture, as would any other new IT system. Depending on the functionality employed by the tool, and the additional software / hardware that will be introduced into the IT environment, the impact can be minimal or more. Participants agreed also that the overall cybersecurity policy of the organization cannot be affected by the addition of another security control. But the established procedures of the organization, such as e.g., the change management policy, or the validation procedures during incident handling, can be impacted as the new monitoring solution should be taken into account.

In regards to the impact of AICS on an organisation’s risk management, participants couldn’t provide a concrete answer as they were lacking the relevant experience in assessing the risk for similar technology. Participants P1 and P2a, according to their experience in risk assessing traditional software, commented that as another IT asset for their organization, AICS should be calculated in the risk assessment. Although P2a participant wondered how this could be done with cloud services such as Microsoft Sentinel since is provided as a service to his organization and is beyond his control. P3 participant, who currently is in the progress of preparing for ISO27001 certification, considers that as a component of the SIEM platform, the ML-based tool should be risk assessed for the whole platform. Participant P7b has also not proceeded with risk assessment of the AICS implementation or any other AI technology in his organization. In his experience, he considers that since the AICS supports the incident management process by monitoring and incident analysis, affects the overall risk management of his organization. For the moment, his organization is in the process to identify the particular security controls that need to enhance in their control baseline, and to address the threats to AI (algorithms, data, learning methods). He considers also that the risk assessment should be done in two parallel streams, one according to the use case of the AI application, and another in combination with the systems that AICS complements (in his case the SIEM platform).

4.1.14 On the competence of the cybersecurity management frameworks to include AICS

To the knowledge of the participants, AI/ML-based tools are not addressed by information security management standards such as ISO27001 and NIST CSF do not define specific controls for AI technologies. To their verification, the author will note that the most relevant publication of the NIST Institute to such technology, is the publication NIST.SP.800-207 [87] on Zero Trust Architecture (ZTA), finalised in August 2020. Although the publication does not refer directly to ML, it acknowledges that ZTA paradigm relies on continuous monitoring of all the organization’s resources over the network, for evaluation and control of their behaviour. *“An enterprise implementing a ZTA should establish a continuous diagnostics and mitigation (CDM) or similar system.”* [87], where the CDM functionality can be provided by NDR and relevant ML-based tools. The NIST guidelines and the ZTA practice are still not widely adopted by the community, but the executive order on cybersecurity[1] of the President of the United States, in May 2021, is expected to accelerate the proliferation of ZTA paradigms at least for the large organizations in the United States.

It is also valuable to note the comments of the participants on the topic of the cybersecurity management frameworks:

- Participant P2a connected the increasing need for standardization around AI technologies, with the cybersecurity management frameworks. Although frameworks are recommended guidelines, when not required by regulations, assist the industry to establish a common language for comparing different architectures and technologies, address complex implementations, and facilitate the communication with management and executive levels. As cybersecurity experts lack a deep understanding of AI/ML technologies, participant P2a thinks that AI should also be included in the recommendations as a step towards standardization.
- On the other hand, participant P6 pointed out the problem that has been created with the Cybersecurity Frameworks, especially when it comes to compliance certificates requested by various regulations, which may affect also AICS. As has been pointed out in [83] [84] [85], an organization being compliant with a certain standard or framework does not necessarily imply that is also secure. Participant marked that frameworks, and by extending the notion the auditors responsible for the certifications, typically examine the cybersecurity architecture macroscopically without getting (or even being able to get) into details. In his words, an organization can be compliant simply by having a firewall plugged in without anyone wondering about the firewall rules.
- Participants' P7b and P7c organization is working closely with ENISA to identify how the existing cybersecurity management frameworks can involve and address threats to AICS.

4.1.15 On the participants' experience from the AICS so far

Participants P1, P2a, and P2b, characterised as positive their experience in working with the AICS tool so far, while participants P7b and P7c characterised it as challenging. Participant P3 characterised it as neutral although he recognizes his limited time using the tool. On the contrary participant P5, after the PoC he run, characterised it as negative.

4.1.16 On the participants' expectations from the AICS tools

Regardless of their answer to the previous question, most of the participants stand optimistic about the future potential of the AICS. They consider AICS as a much-needed technology even in the current perspective of passive and assistive functionality of security monitoring and alerting. In participant's P7c words "*is the only way forward". *In combination with their answers about their motivation for implementing AICS (see paragraph 4.1.1), their expectation is ML to assist in Tier 1 related tedious and repetitive

[1] https://www.whitehouse.gov/briefing-room/presidential-actions/2021/05/12/executive-order-on-improving-the-nations-cybersecurity/

tasks[1] or to assist the cybersecurity teams to cope with the alerts and the event correlation by providing context to the incidents.

4.1.17 On the participants' further remarks

Holding interviews with open-ended questions, can elicit additional information from the participants. It is valuable to mention some of their remarks on the AICS in general.

- Participant P1, from the perspective of employing AICS for web traffic monitoring, considers that the tool can assists the analysts identify issues that might escape "the human eye" which comes in unison with the findings of Akinrolabu et al. [20] and Malajti et al. [26] on cybersecurity analysts' fatigue.
- Participants P2a and P3, commented on the issue of standardization of the AI technology, as they think that non-ML experts can easily get confused with the technology and the terminology. Due to the AI hype nowadays, according to the participants, and as Kroll et al. have remarked in [96], a lot of technologies are characterised as AI-based when they are not actually. They think that more needs to be done from the vendors' side.
- Especially on the AICS technologies and the standardization, participant P7c commented on the need for a similar to MITRE ATT&CK[2] platform that would allow experts and practitioners to choose the right tool and (AI) technology according to their use case.
- Participant P2a, beyond the issue of standardization of AI technology (paragraph 4.1.14), stressed the need for the AI to be interpretable as several times systems have to be audited for action justification or even for testimony (depending on the industry).
- Participant P4, although stands optimistic about the future possibilities of the AICS, for the time he thinks that the technology is not mature enough to deliver what it promises. From his experience with an EDR solution with ML component, beyond the PoC with the software of Darktrace, he has noted differences in the performance of the software over different operating systems. Although the difference in performance is not mentioned by the vendor, he considers the technology still lagging to address the full spectrum of OSs in an IT environment. Further, he thinks that the overreliance of an organization on an intelligent agent such as the AICS, can create additional problems for the organization if not sound security practices are in place.
- Participant P5, presented a very interesting negative impact of the AICS. He considers that AICS might have an impact on an organization's compliances, especially with GDPR regulation, as unnecessary processing of Personal Identified Information (PII) has been recorded in the past[3]. Depending on the use case of the AICS, and with the proliferation of MSSP services the author considers this as a big gap in AICS that should be researched in the future, especially when

[1] Participant P3 referred to Layer 1 of incident handling procedures of established SOC

[2] https://attack.mitre.org/

[3] The participant P5 has not elaborated on his answer further, but the author considers him to refer in general applications of AI technology and potential processing of PII information as e.g., in the case of Government-to-Citizen e-services, Dreyling et al.[28]. In the case of log monitoring of IDS alerts no PII is involved but the monitoring of an endpoint might include, depending on the use case.

SOC can fail cause of the customers’ fear of data privacy as Akinrolabu et al. [20] discovered. The topic is analysed further in the discussion section.

- Participant P6, presented a better insight on the issue of false-positive alerts, as according to his experience a lengthy training period, with regular fine-tuning, is crucial to the systems learning process, although the *staggering* initial number of the false-positive might discourage the *end-users*. Further, on the same issue, he stressed the importance to train the analyst parallel to the ML, as large organizations lack the full picture, *visibility or AVC,* of their IT environment. Participants P7c comment on the issue of the false-positive alerts was that the loss (volume) is balanced by the gain (detection of the unknown).
- Participant P6, he further emphasised the need for constant monitoring of the AICS tools on 24/7 basis, in order to be effective. Although he acknowledges that for small companies and organizations this may not be feasible, he, therefore, expects a shift in the market from in-house solutions to managed security service providers (MSSP).

5 Discussion

This chapter discusses further the results of the qualitative research and aims to answer the research questions:

5.1 The motivation for AICS

Analysing further the results of the survey about the motivations of the participants to deploy AICS, a pattern common to the majority of the participants became apparent. In unison with the findings of Shah [30] and Feng [31], the majority of the participants confirmed the need for automating the analysis of generated alert messages. As explained in these documents, the cybersecurity teams and analysts find themselves overwhelmed by the sheer volume of the information that must analyse. Participant P7c described the experience as ***alert fatigue***. The various sources of log files and alerts and the need for correlation with current and previous incidents in a fast-evolving threat landscape, have driven already the cybersecurity teams to their human limits. The latest expressed also by participants P2b, P4, P6, and P7a who consider that cybersecurity teams have already reached their capacities.

AI is the perfect fit for such an application, as it has the ability to *understand, interpret, and find patterns in vast amounts of data that can be used to provide in-depth analysis and to create targeted exploration processes by overcoming human limitations.*[2] Depending on the use case, traffic monitoring, behaviour analysis, etc., AICS can provide integrated information to the analysts, relating current and past events with external threat intelligence and additional information. AICS can provide prioritization and assist the analysts to stay focused on the most critical events. It can allow them to detect low malicious activity spanning longer periods, offering them the ability to contain an overview of an organization's cybersecurity posture. Further, as expressed by the participant's P3 motivation, AICS can assist analysts with the tedious and repetitive tasks in a SOC's Tier 1 functionality. For participant P7c is "*the only way forward"* to monitor an IT environment. The success of AICS in the demanding environment of SOC is something that future research can investigate.

5.2 The cognitive impact of AICS

As explained, the modern IT environments have become complex, and for large organizations less transparent, distributed between several business units. This complexity leads to knowledge gaps for the analysts on the infrastructure and the services running. Kokulu et al [4] in their survey about failures of SOCs, have found that low visibility into the network infrastructure and the endpoints is the most acknowledged issue that prevents SOCs to be effective. AICS by monitoring large volumes of network traffic and correlating the traffic with detected behaviours, can detect unknown (malicious or legitimate) activity in the IT environment. Especially regarding the legitimate traffic that can be unknown to the analysts, AICS offers a cognitive advantage to the organizations of **application visibility and control** (AVC) over their infrastructure. This cognitive impact enables organizations to enhance and better enforce their cybersecurity policies.

5.3 Technical challenges of AICS

A notion that seems to confuse, or to a certain extent, discourage the participants of the survey is the initial phase of training and tuning the AICS tool. As explained the deployment phase is followed by the training phase of the tool in the operational environment. Depending on the tool's use case and the size of the organization's IT system, this period can be lengthy and last up to a year, as pointed out by participant P6. The tool is allowed this time to self-teach itself of the operating infrastructure and to develop the baseline configuration. In order to accelerate the process, the installer can provide information[1] and additional feedback. Then follows the tuning phase where the installer inspects the tool's output and calibrates the thresholds and the numerical parameters in order to reduce the false-positive alerts generated by the tool. It is obvious that this period leads to an excessive number of false-positive alerts which the installer tries to reduce with the fine-tuning of the tool. The level of acceptance and tolerance was something that varied between the participants of the survey. For participant P5 was a major reason to reject the technology for the moment. For participant P4 was a reason to be sceptical about the tool's performance that together with the tool's inability to detect malicious activity led to a similar conclusion. For participant P3 is a reason to remain sceptical about the tool's deliverables and expects its future performance. Participants P3 and P7c who operate in a dynamic production environment with constant changes understand the reason for the volume of the false-positive alerts but consider the gain more valuable. Participant P2b on the other hand, who deployed an open-source solution for IDS monitoring, was comfortable with the initial period and the number of false-positives. For participant P1, who expects the tool to discover *the unknown*, had not been a major issue also.

It seems that three factors affect the tolerance of the participants in the number of the false-positives. The first and obvious is the employed functionality of the tool that also determines the expectations of the participants from the tool. Especially in the users' behavioural analysis, the notion of *legitimate but unknown* behaviour seems to confuse the end-users of the technology. The second is related to employing an open-source or commercial product, that seems to shape the participants' expectations regarding the tool's performance. As had been pointed also by Kumar et al. [25], cybersecurity analysts, and their organizations, expect commercial tools to be stress-tested and proven technologies that will enable them to put the *least effort* into their implementations. The latest was stressed also by participant P6 as his personal conclusion on the expectations of *magical solutions* by his customers on the said technologies. The third factor is related to the cognitive impact of AICS, see section 5.2, and the *end-users* level of knowledge regarding the organization's IT environment and assets. It is the combination of lack of full knowledge of the infrastructure (network, applications, services)[2] and the ability of AICS to detect the unknown, that makes them underestimate the number of false-positive alerts.

It is valuable to this point to refer to the findings of Kokulu et al [4] and Sommer et al. [34]. Kokulu et al in their survey, on SOCs functionality and issues that might affect them, have discovered, that the false-positive alerts do not have a majority impact onto SOCs

[1] Information about core infrastructure elements as DNS, DHCP, DC, etc.

[2] Combined of course with the complexity of the modern IT systems and the end users' behaviour

effectiveness in opposition to academia's belief. The interviewees of their study have stressed more concern about the unfiltered, unrelated, and uncorrelated alert data.

Sommer et al. [34] in their survey highlight that applying ML in Network IDS is fundamentally different from other application areas making it difficult for effective implementations especially in presenting results with context understandable and actionable by the analysts.

5.4 The cybersecurity for the AI

The participants of the survey expressed the opinion that no specific provisions and security controls are needed for the AICS, on the proviso that the organization has already established concrete cybersecurity practices. It is the author's result also, that following the best cybersecurity practices, a multi-layered and zero-trust cybersecurity architecture (ZTA), AICS tools should be treated with the same precaution as any other core IT asset. Depending on the use case, in supportive roles such as monitoring the tools do not interact with the adversaries directly and have a relatively low risk. The critical component of the AICS, the supervisor, should be isolated from the rest of the network, and allowed to communicate only with the distributed sensors and data collectors in the endpoints and the infrastructure. Monitoring should be in place even for the AICS as for any other IT asset.

A special focus was given in the interviews on the case of deploying the AICS in a compromised system. Due to the tool's nature, it is expected that any malicious activity present during the deployment and training phase, to be learned as legitimate behaviour. Five of the participants confirmed it but remarked that any other activity of the adversaries will result to changes in their behaviour and therefore will violate the AICS established baseline configuration. It seems that in an IT system designed according to concrete cybersecurity practices, AICS will be able to detect not only the escalation of malicious activity but also the lateral movement of the adversaries.

5.5 Working with the AICS

The participants expressed the opinion that no ML competence is needed by the cybersecurity analysts to work with the AICS. It was also common the belief that a sufficient level of understanding is needed of how the tool operates and how attacks are identified. As participants P1, P2b, P3, and P7c marked, the analyst should be able to understand the tool's features and functions, and the tool should be interpretable enough to correlate which input triggered what threshold or what anomaly violated the baseline configuration.

These two factors contribute to the **usability** of the AICS. The first, relates to both vendor and end-user, but is mostly up to the end-user to be well informed about the tool. The second calls for the tool to be sufficiently informative to the analyst about the alerts and to correlate the output with the input that triggered the alert.

The author will add a third factor directly related to the cognitive advantage offered by AICS in section 5.2, that of the detailed knowledge of the deployment environment by the end-users. The ability of the AICS to detect legitimate traffic that is unknown to the end-users seems to be underestimated by the end-users and considered as a tool's inability to perform, while is simply a lack of detailed knowledge from their side. As explained, organisations may lack full knowledge of their IT environments and even legitimate traffic over their networks. This lack of knowledge by the analysts can only prevent them from properly evaluating the alerts and understanding the correlation between input and output of the AICS.

The above three factors constitute for AICS technology to be **interpretable** by the end-users. On the contrary, the reluctance of the vendors to provide detailed information about their models and the internal functions of the ML algorithms, may keep AICS tools non-explainable, but with no major impact on the usability.

The interviews have been focused also to the responsibility in the case of failure with the participants' answers shifting between being shared with the vendor or solely belonging to the end-user. Considering the proliferation of AI technology it seems that the topic needs further investigation with separate research.

Overall the participants' experience with the AICS had been positive, with six characterising it as positive, one as neutral due to his short experience, and two as negative.

5.6 The impact of the AICS on cybersecurity management

The participants agreed that AICS has no direct impact to an organization's cybersecurity policy, but as an additional security control system, that enters the IT environment, it will impact beyond the security architecture, the established procedures and processes in place, as e.g., the change management, or incident handling. Depending on the functionality employed by the tool, and the additional software / hardware that will be introduced into the IT environment, the impact can be minimal or more.

In the concept, that policies describe an organization's attempt to be proactive regarding their cybersecurity, e.g., by forbidding / requesting the usage of certain applications, the author will add, that AICS can only have a positive impact on the ability of an organization to enforce its cybersecurity policy. The cognitive advantage of application visibility and control of the AICS, can assist an organization to detect and deter undesired services and applications over its IT environment.

Regarding the impact of AICS on an organization's risk assessment, participants couldn't provide a concrete answer as they were lacking the relevant experience in risk assessment for similar technology. According to their previous risk assessment experience, four participants commented that as AICS supports incident management processes for their organization, affects the overall risk management and it should be risk assessed. It seems that exists minor experience in risk assessing AICS technologies. Future research to identify particular threats and security controls to AICS, depending on the use case, would be of the most value for the practitioners.

5.7 The competence of cybersecurity frameworks to address AICS

As described in the section of the results, information security management standards like ISO27001 and NIST Cybersecurity Framework do not define specific controls for AI/ML-based tools. The author will comment, as before, that the most relevant publication of the NIST Institute on such technology, publication NIST.SP.800-207 [87] on Zero Trust Architecture (ZTA), since August 2020, does not refer directly to ML. It acknowledges that ZTA relies on continuous monitoring of all the organization's resources over the network, for evaluation and control of their behaviour. This continuous monitoring functionality can be provided by NDR and relevant ML-based tools.

Since the said NIST guidelines and ZTA practice are still not widely adopted, future research of established deployments remains to elaborate on that. The author will add that the insufficiency to address the AI technologies by the said standards and frameworks, has its impact on the practitioners of the technology, as they seek individually ways to evaluate risk properly, and determine the appropriate controls. With the proliferation of AI technologies, and the accompanying threats to AI, the need for common guidelines and standardised procedures for risk assessment is more crucial than before.

5.8 Additional aspects of enhancing cybersecurity with the use of AI

As pointed out by participant P6, AICS tools to be effective, need constative monitoring on 24/7 basis for immediate response to potential cyber incidents. For smaller organizations and companies, this cannot be easily feasible. Therefore, the market for managed security service providers or MSSP is expected to grow together with the awareness of the public on cybersecurity. But as Akinrolabu et al. [20] found in their research, one of the reasons for SOCs' failure is the customer's inadequacy to provide quality network and application logs. Data confidentiality or fear of overexposure of an organization's asset, impacts the detection of sophisticated attacks or the automation of the detection process. With the proliferation of the AICS solutions, future research on the MSSP could explore more on that.

An interesting topic pointed also by participant P4, through his experience with an EDR solution with ML component, beyond the PoC with the software of Darktrace. The participant observed inconsistencies in the tool's performance over different operating systems. Although the vendor of the tool claims that it can operate through various OS, participant P4 observed a staggering number of false-positive alerts when operating in Linux OS rather than in Windows OS. Although is a sole reference of such inconsistency recorded in the research, as participant P4 is the only one with experience with ML-based EDR tool, the author considers that future research should elaborate more on the performance and the consistency of the technology across various OSs.

5.9 Conclusion

The staggering amount of information that needs to be analysed and correlated seems to be the driving motivation for the deployment of AICS tools by the organizations. The AICS technology offers a cognitive advantage to the organizations to discover unknown traffic and services in their IT environments but seems to create confusion for the end-users. Depending on the employed functionality and the environment, for the deployment to be successful, detailed knowledge of the environment is required. Otherwise, the staggering number of alerts and the inability to correlate the AICS output with the proper trigger will discourage the end-users from the technology. Regarding the cybersecurity controls, no dedicated controls are needed, as long as sound cybersecurity practices are in place. AICS can assist organizations to enhance and better enforce their cybersecurity policies. But regarding the risk assessment of the AICS technology the end-users are without guidance as the current versions of the cybersecurity standards and frameworks do not address AI-related threats directly.

6 Summary

The goal of this study had been to research the challenges of applying AI-based technologies in cybersecurity in the terms of organizational and technological aspects. The research tried to present the facts from both sides of technology practitioners, the installer / reseller of the technology, and the end-user of the technology who will operate also with the tool. While the AICS tools are emerging in the market, the available studies on them are limited, in various topics, due to the complexities of the solutions and the use cases.

The results highlight that one of the main motivations for the introduction of the AICS technology is the alert fatigue of cybersecurity analysts. The beyond the human capacity volume of information that needs to be analysed and correlated is the main motivation for the organizations in an evolving cyber threat landscape.

The research delved further into one of the major problems that end-users face with the technology, the number of false-positive alerts. The discussion of the results matches the number of false-positive alerts with the visibility offered by the technology to the IT environments and the lack of detailed knowledge that most organizations face nowadays over their networks. This knowledge gap prevents them from properly evaluating the outcome of AICS and correlating it to the proper trigger.

The research discovered, also, that in most of the deployments there is no need for specific security controls as long as the AICS tools are treated as core elements of the infrastructure and sound cybersecurity practices are in place. The findings highlight that no ML competence is needed by the end-users but in order for the technology to be effective, it should be interpretable enough to the end-users. The technology was found to have a positive impact on the organizations' cybersecurity policies allowing them to properly apply them to their IT environments.

The research, finally, highlighted the gap that exists in the literature regarding the risk assessment of AI technologies. The inadequacy of the current versions of the cybersecurity frameworks and standards to address the threats on AI deters the end-users from properly evaluating risk and determining the appropriate security controls.

References

[1] Trend Micro, “Navigating New Frontiers, Trend Micro 2021 Annual Cybersecurity Report,” Mar. 17, 2022. https://www.trendmicro.com/vinfo/us/security/research-and-analysis/threat-reports/roundup/navigating-new-frontiers-trend-micro-2021-annual-cybersecurity-report (accessed Apr. 24, 2022).

[2] N. Kaloudi and J. Li, “The AI-Based Cyber Threat Landscape: A Survey,” *ACM Comput. Surv.*, vol. 53, no. 1, pp. 1–34, Jan. 2021, doi: 10.1145/3372823.

[3] M. Blowers and J. Williams, “Artificial intelligence presents new challenges in cybersecurity,” in *Disruptive Technologies in Information Sciences IV*, Online Only, United States, May 2020, p. 19. doi: 10.1117/12.2560002.

[4] F. B. Kokulu *et al.*, “Matched and Mismatched SOCs: A Qualitative Study on Security Operations Center Issues,” in *Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security*, London United Kingdom, Nov. 2019, pp. 1955–1970. doi: 10.1145/3319535.3354239.

[5] R. Andrade, J. Torres, and L. Tello-Oquendo, “Cognitive Security Tasks Using Big Data Tools,” in *2018 International Conference on Computational Science and Computational Intelligence (CSCI)*, Las Vegas, NV, USA, Dec. 2018, pp. 100–105. doi: 10.1109/CSCI46756.2018.00026.

[6] R. V. Yampolskiy, “Predicting future AI failures from historic examples,” *FS*, vol. 21, no. 1, pp. 138–152, Mar. 2019, doi: 10.1108/FS-04-2018-0034.

[7] M. A. Alcorn *et al.*, “Strike (with) a pose: Neural networks are easily fooled by strange poses of familiar objects,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 4845–4854.

[8] Evan Ackerman, “Fatal Tesla Self-Driving Car Crash Reminds Us That Robots Aren’t Perfect,” *IEEE Spectrum*, Jan. 07, 2016. https://spectrum.ieee.org/fatal-tesla-autopilot-crash-reminds-us-that-robots-arent-perfect (accessed Apr. 24, 2022).

[9] J. Bieniasz and K. Szczypiorski, “Dataset Generation for Development of Multi-Node Cyber Threat Detection Systems,” *Electronics*, vol. 10, no. 21, p. 2711, Nov. 2021, doi: 10.3390/electronics10212711.

[10] N. Kshetri, “Economics of Artificial Intelligence in Cybersecurity,” *IT Prof.*, vol. 23, no. 5, pp. 73–77, Sep. 2021, doi: 10.1109/MITP.2021.3100177.

[11] M. Taddeo, T. McCutcheon, and L. Floridi, “Trusting artificial intelligence in cybersecurity is a double-edged sword,” *Nat Mach Intell*, vol. 1, no. 12, pp. 557–560, Dec. 2019, doi: 10.1038/s42256-019-0109-1.

[12] Trend Micro Research, United Nations Interregional Crime and Justice Research Institute (UNICRI), and Europol’s European Cybercrime Centre (EC3), “Malicious Uses and Abuses of Artificial Intelligence,” p. 80.

[13] H. S. Anderson, A. Kharkar, and B. Filar, “Evading Machine Learning Malware Detection,” p. 6.

[14] J. Li, “Cyber security meets artificial intelligence: a survey,” *Frontiers Inf Technol Electronic Eng*, vol. 19, no. 12, pp. 1462–1474, Dec. 2018, doi: 10.1631/FITEE.1800573.

[15] D. Wagner, “Building More Resilient Cybersecurity Solutions for Infrastructure Systems,” in *Systems Engineering in the Fourth Industrial Revolution*, John Wiley & Sons, Ltd, 2019, pp. 415–443. doi: https://doi.org/10.1002/9781119513957.ch16.

[16] S. Zeadally, E. Adi, Z. Baig, and I. A. Khan, “Harnessing Artificial Intelligence Capabilities to Improve Cybersecurity,” *IEEE Access*, vol. 8, pp. 23817–23837, 2020, doi: 10.1109/ACCESS.2020.2968045.

[17] J. Romero-Mariona *et al.*, “An Approach to Organizational Cybersecurity,” in *Enterprise Security*, vol. 10131, V. Chang, M. Ramachandran, R. J. Walters, and G. Wills, Eds. Cham: Springer International Publishing, 2017, pp. 203–222. doi: 10.1007/978-3-319-54380-2_9.

[18] G. Vidot, C. Gabreau, I. Ober, and I. Ober, “Certification of embedded systems based on Machine Learning: A survey,” *arXiv:2106.07221 [cs, stat]*, Jul. 2021, Accessed: Apr. 24, 2022. [Online]. Available: http://arxiv.org/abs/2106.07221

[19] S. Oesch *et al.*, “An Assessment of the Usability of Machine Learning Based Tools for the Security Operations Center,” in *2020 International Conferences on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData) and IEEE Congress on Cybermatics (Cybermatics)*, Rhodes, Greece, Nov. 2020, pp. 634–641. doi: 10.1109/iThings-GreenCom-CPSCom-SmartData-Cybermatics50389.2020.00111.

[20] O. Akinrolabu, I. Agrafiotis, and A. Erola, “The challenge of detecting sophisticated attacks: Insights from SOC Analysts,” in *Proceedings of the 13th International Conference on Availability, Reliability and Security*, Hamburg Germany, Aug. 2018, pp. 1–9. doi: 10.1145/3230833.3233280.

[21] M. A. Salitin and A. H. Zolait, “The role of User Entity Behavior Analytics to detect network attacks in real time,” in *2018 International Conference on Innovation and Intelligence for Informatics, Computing, and Technologies (3ICT)*, Sakhier, Bahrain, Nov. 2018, pp. 1–5. doi: 10.1109/3ICT.2018.8855782.

[22] R. S. Gutzwiller, S. M. Hunt, and D. S. Lange, “A task analysis toward characterizing cyber-cognitive situation awareness (CCSA) in cyber defense analysts,” in *2016 IEEE International Multi-Disciplinary Conference on Cognitive Methods in Situation Awareness and Decision Support (CogSIMA)*, San Diego, CA, USA, Mar. 2016, pp. 14–20. doi: 10.1109/COGSIMA.2016.7497780.

[23] C. Griffy-Brown, H. Miller, V. Zhao, D. Lazarikos, and M. Chun, “Emerging Technologies and Risk: How Do We Optimize Enterprise Risk When Deploying Emerging Technologies?,” in *2019 IEEE Technology & Engineering Management Conference (TEMSCON)*, Atlanta, GA, USA, Jun. 2019, pp. 1–5. doi: 10.1109/TEMSCON.2019.8813743.

[24] N. Rawindaran, A. Jayal, and E. Prakash, “Machine Learning Cybersecurity Adoption in Small and Medium Enterprises in Developed Countries,” *Computers*, vol. 10, no. 11, p. 150, Nov. 2021, doi: 10.3390/computers10110150.

[25] R. S. Siva Kumar *et al.*, “Adversarial Machine Learning-Industry Perspectives,” in *2020 IEEE Security and Privacy Workshops (SPW)*, San Francisco, CA, USA, May 2020, pp. 69–75. doi: 10.1109/SPW50608.2020.00028.

[26] M. Malatji and A. Marnewick, “The Impact of Artificial Intelligence on the Human Aspects of Information and Cybersecurity,” p. 12, 2018.

[27] M. Masombuka, “Towards an Artificial Intelligence Framework to Actively Defend Cyberspace in South Africa,” p. 99.

[28] R. Dreyling, E. Jackson, and I. Pappel, “Cyber Security Risk Analysis for a Virtual Assistant G2C Digital Service Using FAIR Model,” in *2021 Eighth International Conference on eDemocracy & eGovernment (ICEDEG)*, Quito, Ecuador, Jul. 2021, pp. 33–40. doi: 10.1109/ICEDEG52154.2021.9530938.

[29] B.Arshia, M.Gayathri, and P.Manaswini, “AI in Cyber Security,” vol. 4, no. 9, Sep. 2017.

[30] S. A. R. Shah and B. Issac, “Performance comparison of intrusion detection systems and application of machine learning to Snort system,” *Future Generation Computer Systems*, vol. 80, pp. 157–170, Mar. 2018, doi: 10.1016/j.future.2017.10.016.

[31] C. Feng, S. Wu, and N. Liu, “A user-centric machine learning framework for cyber security operations center,” in *2017 IEEE International Conference on Intelligence and Security Informatics (ISI)*, Beijing, China, Jul. 2017, pp. 173–175. doi: 10.1109/ISI.2017.8004902.

[32] S. McElwee, J. Heaton, J. Fraley, and J. Cannady, “Deep learning for prioritizing and responding to intrusion detection alerts,” in *MILCOM 2017 - 2017 IEEE Military Communications Conference (MILCOM)*, Baltimore, MD, Oct. 2017, pp. 1–5. doi: 10.1109/MILCOM.2017.8170757.

[33] Kelly Sheridan, “Future of the SIEM,” *DARKReading*, Mar. 22, 2017. https://www.darkreading.com/threat-intelligence/future-of-the-siem (accessed Apr. 24, 2022).

[34] R. Sommer and V. Paxson, “Outside the Closed World: On Using Machine Learning for Network Intrusion Detection,” in *2010 IEEE Symposium on Security and Privacy*, Oakland, CA, USA, 2010, pp. 305–316. doi: 10.1109/SP.2010.25.

[35] “How Anti-Virus Software Works,” *Virus - A retrospective*. https://cs.stanford.edu/people/eroberts/cs201/projects/2000-01/viruses/anti-virus.html (accessed Apr. 24, 2022).

[36] B. Buchanan and T. Miller, “Machine Learning for Policymakers,” *Machine Learning*, p. 58, 2017.

[37] Center for Security and Emerging Technology and A. Lohn, “Hacking AI: A Primer for Policymakers on Machine Learning Cybersecurity,” Center for Security and Emerging Technology, Dec. 2020. doi: 10.51593/2020CA006.

[38] R. Meier, A. Lavrenovs, K. Heinaaro, L. Gambazzi, and V. Lenders, “Towards an AI-powered Player in Cyber Defence Exercises,” 2021, vol. 2021-May, pp. 309–326. doi: 10.23919/CyCon51939.2021.9467801.

[39] Vijay Kanade, “What Is Machine Learning? Definition, Types, Applications, and Trends for 2022,” *TOOLBOX tech*, May 04, 2022. https://www.toolbox.com/tech/artificial-intelligence/articles/what-is-ml/#_003 (accessed Apr. 24, 2022).

[40] S. J. Russell, P. Norvig, and E. Davis, *Artificial intelligence: a modern approach*, 3rd ed. Upper Saddle River: Prentice Hall, 2010.

[41] (ISC)2, “A Resilient Cybersecurity Profession Charts the Path Forward, (ISC)2 CYBERSECURITY WORKFORCE STUDY, 2021.” [Online]. Available: https://www.isc2.org//-/media/ISC2/Research/2021/ISC2-Cybersecurity-Workforce-Study-2021.ashx

[42] Klaus Schwab and Børge Brende, “Global Risks Report 2022, World Economic Forum,” 17th Edition. [Online]. Available: https://www.weforum.org/reports/global-risks-report-2022

[43] Lily Hay Newman, “AI Can Help Cybersecurity—If It Can Fight Through the Hype,” *Wired*, Apr. 29, 2018. https://www.wired.com/story/ai-machine-learning-cybersecurity/ (accessed Apr. 24, 2022).

[44] R. Vaarandi, “A Stream Clustering Algorithm for Classifying Network IDS Alerts,” in *2021 IEEE International Conference on Cyber Security and Resilience (CSR)*, Rhodes, Greece, Jul. 2021, pp. 14–19. doi: 10.1109/CSR51186.2021.9527926.

[45] J. Martínez Torres, C. Iglesias Comesaña, and P. J. García-Nieto, “Review: machine learning techniques applied to cybersecurity,” *Int. J. Mach. Learn. & Cyber.*, vol. 10, no. 10, pp. 2823–2836, Oct. 2019, doi: 10.1007/s13042-018-00906-1.

[46] M. Husák, J. Komárková, E. Bou-Harb, and P. Čeleda, “Survey of Attack Projection, Prediction, and Forecasting in Cyber Security,” *IEEE Communications Surveys Tutorials*, vol. 21, no. 1, pp. 640–660, 2019, doi: 10.1109/COMST.2018.2871866.

[47] S. A. Sokolov, T. B. Iliev, and I. S. Stoyanov, “Analysis of Cybersecurity Threats in Cloud Applications Using Deep Learning Techniques,” in *2019 42nd International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, Opatija, Croatia, May 2019, pp. 441–446. doi: 10.23919/MIPRO.2019.8756755.

[48] Y. Nikoloudakis *et al.*, “Towards a Machine Learning Based Situational Awareness Framework for Cybersecurity: An SDN Implementation,” *Sensors*, vol. 21, no. 14, p. 4939, Jul. 2021, doi: 10.3390/s21144939.

[49] Q. He, X. Meng, R. Qu, and R. Xi, “Machine Learning-Based Detection for Cyber Security Attacks on Connected and Autonomous Vehicles,” *Mathematics*, vol. 8, no. 8, p. 1311, Aug. 2020, doi: 10.3390/math8081311.

[50] G. Mezzina, V. F. Annese, and D. De Venuto, “A Cybersecure P300-Based Brain-to-Computer Interface against Noise-Based and Fake P300 Cyberattacks,” *Sensors*, vol. 21, no. 24, p. 8280, Dec. 2021, doi: 10.3390/s21248280.

[51] K. Hasan, S. Shetty, and S. Ullah, “Artificial Intelligence Empowered Cyber Threat Detection and Protection for Power Utilities,” in *2019 IEEE 5th International Conference on Collaboration and Internet Computing (CIC)*, Los Angeles, CA, USA, Dec. 2019, pp. 354–359. doi: 10.1109/CIC48465.2019.00049.

[52] S. Boudko and H. Abie, “Adaptive Cybersecurity Framework for Healthcare Internet of Things,” in *2019 13th International Symposium on Medical Information and Communication Technology (ISMICT)*, Oslo, Norway, May 2019, pp. 1–6. doi: 10.1109/ISMICT.2019.8743905.

[53] H. Abie, “Cognitive Cybersecurity for CPS-IoT Enabled Healthcare Ecosystems,” in *2019 13th International Symposium on Medical Information and Communication Technology (ISMICT)*, Oslo, Norway, May 2019, pp. 1–6. doi: 10.1109/ISMICT.2019.8743670.

[54] D. Bruns-Smith, M. M. Baskaran, J. Ezick, T. Henretty, and R. Lethin, “Cyber Security through Multidimensional Data Decompositions,” in *2016 Cybersecurity Symposium (CYBERSEC)*, Coeur d’Alene, ID, USA, Apr. 2016, pp. 59–67. doi: 10.1109/CYBERSEC.2016.017.

[55] G. Spanos, K. M. Giannoutakis, K. Votis, and D. Tzovaras, “Combining Statistical and Machine Learning Techniques in IoT Anomaly Detection for Smart Homes,” in *2019 IEEE 24th International Workshop on Computer Aided Modeling and Design of Communication Links and Networks (CAMAD)*, 2019, pp. 1–6. doi: 10.1109/CAMAD.2019.8858490.

[56] M. N. Al-Mhiqani *et al.*, “A new intelligent multilayer framework for insider threat detection,” *Computers & Electrical Engineering*, vol. 97, p. 107597, Jan. 2022, doi: 10.1016/j.compeleceng.2021.107597.

[57] P. V. Mohan, S. Dixit, A. Gyaneshwar, U. Chadha, K. Srinivasan, and J. T. Seo, “Leveraging Computational Intelligence Techniques for Defensive Deception: A Review, Recent Advances, Open Problems and Future Directions,” *Sensors*, vol. 22, no. 6, p. 2194, Mar. 2022, doi: 10.3390/s22062194.

[58] K. Shaukat, S. Luo, V. Varadharajan, I. A. Hameed, and M. Xu, “A Survey on Machine Learning Techniques for Cyber Security in the Last Decade,” *IEEE Access*, vol. 8, pp. 222310–222354, 2020, doi: 10.1109/ACCESS.2020.3041951.

[59] S. Studer *et al.*, “Towards CRISP-ML(Q): A Machine Learning Process Model with Quality Assurance Methodology,” *arXiv:2003.05155 [cs, stat]*, Feb. 2021, Accessed: Apr. 24, 2022. [Online]. Available: http://arxiv.org/abs/2003.05155

[60] H. Hindy *et al.*, “A Taxonomy of Network Threats and the Effect of Current Datasets on Intrusion Detection Systems,” *IEEE Access*, vol. 8, pp. 104650–104675, 2020, doi: 10.1109/ACCESS.2020.3000179.

[61] A. Ibrahim, D. Thiruvady, J.-G. Schneider, and M. Abdelrazek, “The Challenges of Leveraging Threat Intelligence to Stop Data Breaches,” *Front. Comput. Sci.*, vol. 2, p. 36, Aug. 2020, doi: 10.3389/fcomp.2020.00036.

[62] J. Quiñonero-Candela, Ed., *Dataset shift in machine learning*. Cambridge, Mass: MIT Press, 2009.

[63] B. Brevini and F. Pasquale, “Revisiting the Black Box Society by rethinking the political economy of big data,” *Big Data & Society*, vol. 7, no. 2, p. 205395172093514, Jul. 2020, doi: 10.1177/2053951720935146.

[64] M. Craglia and Europäische Gemeinschaften, Eds., *Artificial intelligence: a european perspective*. Luxembourg: Publications Office of the European Union, 2018. doi: 10.2760/936974.

[65] J. Burrell, “How the machine ‘thinks’: Understanding opacity in machine learning algorithms,” *Big Data & Society*, vol. 3, no. 1, p. 205395171562251, Jun. 2016, doi: 10.1177/2053951715622512.

[66] L. H. Gilpin, D. Bau, B. Z. Yuan, A. Bajwa, M. Specter, and L. Kagal, “Explaining Explanations: An Overview of Interpretability of Machine Learning,” in *2018 IEEE 5th International Conference on Data Science and Advanced Analytics (DSAA)*, Turin, Italy, Oct. 2018, pp. 80–89. doi: 10.1109/DSAA.2018.00018.

[67] P. J. Phillips, C. A. Hahn, P. C. Fontana, D. A. Broniatowski, and M. A. Przybocki, “Four Principles of Explainable Artificial Intelligence,” preprint, Aug. 2020. doi: 10.6028/NIST.IR.8312-draft.

[68] N. Carlini *et al.*, “On Evaluating Adversarial Robustness,” p. 25.

[69] M. M. Najafabadi, F. Villanustre, T. M. Khoshgoftaar, N. Seliya, R. Wald, and E. Muharemagic, “Deep learning applications and challenges in big data analytics,” *Journal of Big Data*, vol. 2, no. 1, p. 1, Feb. 2015, doi: 10.1186/s40537-014-0007-7.

[70] L. Pupillo, S. Fantin, A. Ferreira, C. Polito, and Centre for European Policy Studies, *Artificial intelligence and cybersecurity technology, governance and policy challenges: final report of a CEPS Task Force*. 2021. Accessed: Apr. 24, 2022. [Online]. Available: https://www.ceps.eu/download/publication/?id=33262&pdf=CEPS-TFR-Artificial-Intelligence-and-Cybersecurity.pdf

[71] M. Brundage *et al.*, “The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation,” *ArXiv*, vol. abs/1802.07228, 2018.

[72] European Union Agency for Cybersecurity., *AI cybersecurity challenges: threat landscape for artificial intelligence.* LU: Publications Office, 2020. Accessed: Apr. 25, 2022. [Online]. Available: https://data.europa.eu/doi/10.2824/238222

[73] M. Comiter, “Attacking Artificial Intelligence,” p. 90, 2019.

[74] Y. Liu, X. Chen, C. Liu, and D. X. Song, “Delving into Transferable Adversarial Examples and Black-box Attacks,” *ArXiv*, vol. abs/1611.02770, 2017.

[75] A. Athalye, L. Engstrom, A. Ilyas, and K. Kwok, “Synthesizing Robust Adversarial Examples,” *ArXiv*, vol. abs/1707.07397, 2018.

[76] C. Liao, H. Zhong, A. C. Squicciarini, S. Zhu, and D. J. Miller, “Backdoor Embedding in Convolutional Neural Network Models via Invisible Perturbation,” *Proceedings of the Tenth ACM Conference on Data and Application Security and Privacy*, 2020.

[77] A. Dubey, R. Cammarota, and A. Aysu, “MaskedNet: The First Hardware Inference Engine Aiming Power Side-Channel Protection,” in *2020 IEEE International Symposium on Hardware Oriented Security and Trust (HOST)*, 2020, pp. 197–208. doi: 10.1109/HOST45689.2020.9300276.

[78] B. von Solms and R. von Solms, “Cybersecurity and information security – what goes where?,” *ICS*, vol. 26, no. 1, pp. 2–9, Mar. 2018, doi: 10.1108/ICS-04-2017-0025.

[79] H. Mokalled, D. Debertol, E. Meda, and C. Pragliola, *The Importance to Manage Data Protection in the Right Way: Problems and Solutions*. 2017, p. 82. doi: 10.1007/978-3-319-67308-0_8.

[80] J. L. Carlson *et al.*, “Resilience: Theory and Application.,” Feb. 2012, doi: 10.2172/1044521.

[81] R. Klahr *et al.*, *Cyber security breaches survey 2017 (main report)*. 2017.

[82] M. Syafrizal, S. R. Selamat, and N. A. Zakaria, “Analysis of Cybersecurity Standard and Framework Components,” vol. 12, no. 3, p. 16, 2020.

[83] A. Dedeke and K. Masterson, “Contrasting cybersecurity implementation frameworks (CIF) from three countries,” *ICS*, vol. 27, no. 3, pp. 373–392, Jul. 2019, doi: 10.1108/ICS-10-2018-0122.

[84] R. Diesch, M. Pfaff, and H. Krcmar, “A comprehensive model of information security factors for decision-makers,” *Computers & Security*, vol. 92, p. 101747, May 2020, doi: 10.1016/j.cose.2020.101747.

[85] M. Malatji, A. Marnewick, and S. von Solms, “Validation of a socio-technical management process for optimising cybersecurity practices,” *Computers & Security*, vol. 95, p. 101846, Aug. 2020, doi: 10.1016/j.cose.2020.101846.

[86] A. Szychter, H. Ameur, and A. Kung, “The Impact of Artificial Intelligence on Security: a Dual Perspective,” p. 14.

[87] Scott Rose, Oliver Borchert, Stu Mitchell, and Sean Connelly, “Zero Trust Architecture, NIST Special Publication 800-207.” National Institute of Standards and Technology, Aug. 2020. [Online]. Available: https://doi.org/10.6028/NIST.SP.800-207

[88] E. M. Hutchins, M. J. Cloppert, and R. M. Amin, “Intelligence-Driven Computer Network Defense Informed by Analysis of Adversary Campaigns and Intrusion Kill Chains,” p. 14.

[89] P. Nespoli, D. Papamartzivanos, F. G. Mármol, and G. Kambourakis, “Optimal Countermeasures Selection Against Cyber Attacks: A Comprehensive Survey on

Reaction Frameworks," *IEEE Communications Surveys and Tutorials*, vol. 20, no. 2, pp. 1361–1396, 2018, doi: 10.1109/COMST.2017.2781126.

[90] T. W. Edgar and D. O. Manz, "Part IV. Experimental Research Methods," in *Research Methods for Cyber Security*, T. W. Edgar and D. O. Manz, Eds. Syngress, 2017, p. 213. doi: 10.1016/B978-0-12-805349-2.00034-0.

[91] A. Bogner, B. Littig, and W. Menz, *Interviews mit Experten: eine praxisorientierte Einführung*. Springer-Verlag, 2014.

[92] V. Braun and V. Clarke, "Using thematic analysis in psychology," *Qualitative Research in Psychology*, vol. 3, no. 2, pp. 77–101, Jan. 2006, doi: 10.1191/1478088706qp063oa.

[93] M. Maguire and B. Delahunt, "Doing a Thematic Analysis: A Practical, Step-by-Step Guide for Learning and Teaching Scholars.," vol. 8, no. 3, p. 14, 2017.

[94] R. Kissel, A. Regenscheid, M. Scholl, and K. Stine, "Guidelines for Media Sanitization," National Institute of Standards and Technology, NIST SP 800-88r1, Dec. 2014. doi: 10.6028/NIST.SP.800-88r1.

[95] N. H. Chowdhury, M. T. P. Adam, and G. Skinner, "The impact of time pressure on cybersecurity behaviour: a systematic literature review," *Behaviour & Information Technology*, vol. 38, no. 12, pp. 1290–1308, Dec. 2019, doi: 10.1080/0144929X.2019.1583769.

[96] J. A. Kroll, J. B. Michael, and D. B. Thaw, "Enhancing Cybersecurity via Artificial Intelligence: Risks, Rewards, and Frameworks," *Computer*, vol. 54, no. 6, pp. 64–71, Jun. 2021, doi: 10.1109/MC.2021.3055703.

Appendix 1 – Non-exclusive licence for reproduction and publication of a graduation thesis[1]

I Georgios Kontis

1. Grant Tallinn University of Technology free licence (non-exclusive licence) for my thesis “**Technical and Organizational Challenges on Enhancing Cybersecurity with Artificial Intelligence”**, supervised by Professor Hayretdin Bahşi.
   1.1. to be reproduced for the purposes of preservation and electronic publication of the graduation thesis, incl. to be entered in the digital collection of the library of Tallinn University of Technology until expiry of the term of copyright;
   1.2. to be published via the web of Tallinn University of Technology, incl. to be entered in the digital collection of the library of Tallinn University of Technology until expiry of the term of copyright.
2. I am aware that the author also retains the rights specified in clause 1 of the non-exclusive licence.
3. I confirm that granting the non-exclusive licence does not infringe other persons' intellectual property rights, the rights arising from the Personal Data Protection Act or rights arising from other legislation.

16.05.2022

[1] The non-exclusive licence is not valid during the validity of access restriction indicated in the student's application for restriction on access to the graduation thesis that has been signed by the school's dean, except in case of the university's right to reproduce the thesis for preservation purposes only. If a graduation thesis is based on the joint creative activity of two or more persons and the co-author(s) has/have not granted, by the set deadline, the student defending his/her graduation thesis consent to reproduce and publish the graduation thesis in compliance with clauses 1.1 and 1.2 of the non-exclusive licence, the non-exclusive license shall not be valid for the period.

Appendix 2 - Introductory note sent to the interviewees

Organizational aspects of enhancing Cybersecurity with the use of Artificial Intelligence

Research's introductory note

Modern cyber threat landscape is evolving rapidly with cyber threats becoming more advanced and sophisticated. Artificial Intelligence (AI) is stepping into the cyber arena providing new opportunities and challenges to both "opponents". Implementing Artificial Intelligence in Cybersecurity (AICS) can increase the security posture of an organization, overcoming the limited capabilities of traditional security tools and humans with advantages such as: Processing large volumes of data from a range of sources, Maintaining the level of cybersecurity of the organization by rapid identification of threat factors and allowing security teams to focus on strategic tasks, Identifying cyber threats and suspicious behaviour by slight movements and changes to behaviours undetected to human, Improving and automating the detection of cyber threats as AICS is adapted and learns from experience and standards without human involvement, Accelerating detection and response time by handling multiple security alerts allowing faster response.

But the introduction of AICS comes with some limitations and new challenges that cyber defenders should consider too: The categorization models need training data with detailed datasets of anomalies, and malicious or non-malicious code that covers most use cases. Obtaining these efficient datasets requires time and resources that most organizations cannot afford. Data quality is of critical importance as low quality results in poor decisions and vulnerable systems. AI can also enable new forms of attacks to the AI system itself such as: Data poisoning by injecting false training data to corrupt the learning model and impact the future behaviour of the system. Adversarial inputs with minor changes to the original input, undetected by humans but not to the algorithm, can fool the classification model and tamper the system for future exploitation. The categorization model is the most confidential asset of an AI system and Model stealing is a massive security breach as the output becomes predictable and systems manipulatable. Establishing backdoors triggers by injecting carefully crafted training data that would allow the adversary to dictate the future system's response.

Additionally, to the above, AI technologies have inbuilt weaknesses such as: Explainability and reasoning. Due to the complexity of the decision models, AI systems have the reputation of being "black boxes", for not providing explanations and reasons for the decisions made. Lack of transparency, lack of reasoning in the AI decisions is an obstacle to better understanding the threat landscape and managing the risk. This is even more crucial when auditing the system. Trustworthiness. Lack of transparency, inexplainable decisions, and an evolving learning environment make it hard to evaluate whether the system will continue to behave as expected in any given context.

Summing up the above, one must thoroughly control the datasets provided to the AI system, especially during the training, to safeguard the decision-making models, and at the same time to constantly question the system's reliability. The above constitute a technical challenge to the AI developers and cybersecurity experts, but also an organizational challenge to the organizations. Beyond adding one more IT asset to the risk assessment, AI with the ability to alter an organization's attack surface, challenges the cybersecurity policies and procedures. Either having a customized policy to the individual's needs or following one of the international Standards and Frameworks such as NIST CSF or RMF and ISO/IEC 27000-series, AI interacts with all phases of Risk Management.

This research aims to understand how organizations' cybersecurity policies are affected by AICS solutions and to what extent. For that, it firstly examines the stance of the organizations and the experts in implementing AICS solutions. Either one has been chosen already or not, the research focuses on the motivations and the expectations of the participants from the software. Secondly, investigates the various deployment approaches and what challenges may arise with the introduction of AICS. Investigates the cybersecurity provisions taken by the organizations and the necessary organizational changes that had to be made. Finally, the research records the participants' experience with the use of AI and delves into the symbiosis with humans, especially in the heart of an organization's cybersecurity, the cybersecurity teams and analysts.

About questionnaire, confidentiality, and anonymity

The questionnaire is mostly composed of open-end questions, aiming to motivate a productive discussion. The collected information of the participant and his organization is anonymous, and the author of the research is committed to the confidentiality of the information.

Appendix 3 - Questionnaire

A. Information about the organization and its cyber awareness.

1. Do you consent to the recording of this interview?

2. Please state the industry sector and the size of your organization

3. Please state your position within the organization.

4. How would you describe the knowledge of the organization regarding the IT infrastructure is using (hardware and software)? The organization:
    a. is not aware,
    b. has limited knowledge, IT infrastructure is provided as a service by third parties,
    c. has knowledge of the infrastructure besides services / systems provided by third parties,
    d. has knowledge of the infrastructure besides legacy applications / systems that operate as "black boxes",
    e. has full knowledge of the infrastructure is using and the specifications of the services / systems provided by third parties.

5. How would you describe the knowledge of the organization regarding the data management? The organization:
    a. is not aware,
    b. has the knowledge only of the data located inside the organization,
    c. has the knowledge of the data located inside the organization and in transition.

6. How would you describe the level of security awareness of your organization regarding cybersecurity and related cyber risks?
    a. Not aware b. Limited c. Concerned d. Conscious

7. What provisions have been made to mitigate cyber risks in the organization? Provisions as e.g.:
    a. Dedicated or designated cybersecurity software / packages
    b. Dedicated or designated cybersecurity team
    c. Active SOC – Security Operation Center
    d. Formal cybersecurity policy in place
    e. Other, please describe

8. Does the organization have a formal cybersecurity policy? If yes, this is:
    a. Compliant to an international acknowledged Risk Management Framework or Information Security Standard. Please name it.
    b. Customised to your needs or business.
    c. No security policy in place

9. On regards to the digital assets, is there a requirement for compliance for the organization as e.g. GDPR or industry specific?

B. Information about the use of an Artificial Intelligence solution in Cybersecurity (AICS)

1. Has the organization adopted an AICS solution already?

   1.1. **If yes**, is it a commercial product available in the market? Please state the product and the vendor.

   1.2. What had been the motivation for the use of an AICS solution?

   1.3. What had been your criteria for choosing the specific software?

   1.4. In regards of the implementation, have your expectations been met?

2. **If no,** what is the organization's opinion of enhancing its cybersecurity capacity with an AICS?

   2.1. What would you describe as a positive motivation for implementing an AICS solution?

   2.2. What would be your choosing criteria?

   2.3. What would your expectations from the software?

3. In case your organization is **not in favour of** using AI in cybersecurity, can you elaborate more?

C. Information about the utilization of Artificial Intelligence in cybersecurity

1. What **use areas** would you consider for an AICS solution in the organization?
   (e.g., network security, data security, identification, access security, etc.)

2. What **utilization and functionalities** would you consider for an AICS solution in the organization?
   (e.g., detection of cyber threats, prediction of cyber risks, etc.)

D. Information about the security controls related to the protection of AICS solution itself

1. What security controls would you consider protecting the AICS solution against cyber-attacks?

2. What parts of the AICS solution would you consider protecting and how?

E. Information about the governance of the AICS solution by the organization

1. In regard to your answer about the use area and the functionalities given to the AICS solution, what **roles** would you consider for the AICS solution?

Passive (e.g., monitoring, reporting, threat intelligence) or Active (e.g., defence, response)

2. What **method** would you consider for the AICS solution, Supervised or Unsupervised?

3. How would you keep the AICS solution updated with the latest information? (e.g., threat intelligence, retrain the model to the latest threat detection)

4. What measures would you include to establish **trust** in the AICS solution?

5. In regard to your previous answers in this section, would you consider the organization’s security architecture affected by the implementation of the AICS solution?

6. How would you establish a beneficial cooperation between human and the AICS solution?

F. Information about the Management of Risk and Cybersecurity

1. Would the AICS solution be included in the organization’s formal cybersecurity policy and what would be the changes to the existing one?

2. Which stages of the policy would include the AICS solution? (e.g., risk assessment, incident management, etc.)

3. What would be the impact of the AICS solution on your Risk Analysis? Would you consider the risk would be reduced?

4. With reference to your answer in question A8a, would you consider that the applied framework / standard in its current format, is capable to assess the AICS solution?

5. Would you consider that the AICS solution impacts your certification/compliance, if any?

G. Information about the experience with an AICS solution

1. How would you describe the implementation of the AICS solution regarding the organization of cybersecurity?

2. What has been the acceptance of the AICS solution by the security team? Have there been any complaints or objections to the symbiosis with the AICS solution?

3. Would you consider your security has been improved?

4. How would you rate the organization’s experience with the AICS solution so far? Please choose:

Problematic	the solution has failed to our expectations by creating additional problems which have affected the organization’s cybersecurity capabilities.
Negative	the solution has created additional problems, but the organization’s cybersecurity capabilities have not been affected.
Neutral	the solution has minimal impact on the organization’s cybersecurity capabilities.
Positive	the solution has improved the organization’s cybersecurity capabilities.
Strongly positive	the solution has improved substantially the organization’s cybersecurity capabilities.

5. Is there something you’d like to share that could improve our awareness of the AICS solution and your experiences?

6. Would you recommend this research project to your peers in the industry?

Thank you for your time and effort.