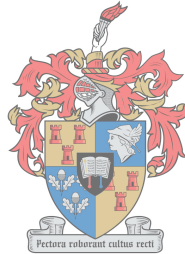


# Towards an Artificial Intelligence Framework to Actively Defend Cyberspace in South Africa

By

Mmalerato Masombuka



UNIVERSITEIT  
iYUNIVESITHI  
STELLENBOSCH  
UNIVERSITY

Thesis presented in fulfilment of the requirements for the degree of Master of Philosophy (Information and Knowledge Management) in the Faculty of Arts and Social Sciences at the University of Stellenbosch

Supervisors: Dr Marthie Grobler and Prof. Bruce W. Watson  
December 2018.

## **Declaration**

By submitting this report electronically, I declare that the entirety of the work contained therein is my own original work that I am the sole author thereof (save to the extent explicitly otherwise stated), that reproduction and publication thereof by Stellenbosch University will not infringe any third party rights and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

Date: December 2018.

Copyright © 2018 Stellenbosch University

All rights reserved

## **Acknowledgement**

I would first like to thank Dr Marthie Grobler for her support, guidance and patience throughout my research study. I have been extremely blessed to have a supervisor who cared so much about my work, who played such an active role and provided me with all the support I needed throughout this process. I would also like to express my heartfelt gratitude to the head of department and supervisor Prof. Bruce Watson for his support. This research study was a true demonstration of teamwork.

I would also like to give a special thanks to all the interviewees for their time, effort and for providing me with valuable information. I must acknowledge and appreciate the role played by Dr Ntandazo Sifolo, Pierre Jacobs, Dr Petrus “Beer” Duvenage and Mr Siphwe Xaba in my life and throughout this study. I was truly standing in the shoulders of giants.

Lastly, I would like to express profound gratitude to my partner and family for providing me with continued support and encouragement throughout my years of study. Your love and prayers have carried me through.

## Executive summary

Cyberattacks pose a great threat to users, including private corporations, academia and government institutions, as they embrace and rely on technology for competence, service provision and other daily routines. Furthermore, the expansion of ICT has introduced an unprecedented magnitude of convenience, efficiency and effectiveness to its users. Similarly, the expansion of ICT has also seen an increase in accompanying risks. Innovation and novelty in areas such as mobile and banking applications, cloud computing and the Internet of Things (IoT) are increasing, culminating in cumulative security challenges as they increase. Thus, in this digital age, safeguarding the privacy and security of information is critical. The countering of advanced adversaries requires an active approach to cybersecurity

Therefore, innovative approaches such as the application of AI tools that have a learning capacity and are adaptable, analysis-driven and able to detect user behaviour, make intelligent and real-time decisions will assist in fighting the cyber threat. To demonstrate the need to defend the cyberspace using AI and to show current progress by the South African private sector in terms of AI-driven tools, four companies were interviewed. The companies were selected based on their cybersecurity approach that gravitates towards demonstrating the significance of using AI for cybersecurity, and because their future prospects of using AI for cybersecurity were fitting for this particular research.

The cyberspace comprises diversified aggressors with varied motivations; thus, this research study proposes a shift in defence surface within the South African context, a shift that is inclusive of AI for cybersecurity. The research study proposes an AI framework aimed at demonstrating the significance of combining AI and cybersecurity. The proposed framework has prioritised 9 elements that will promote the protection and enhance the cyber resilience of information systems and other critical infrastructures that have an impact on national security. The proposed framework is called CAIBER Framework and the name is pronounced as C-Y-B-E-R. The CAIBER Framework is inspired by the core functions of the National Institute of Standards and Technology's Cyber Security Framework for cyber defence. Moreover, the core elements that have been prioritised by the CAIBER Framework emanated from the limitations that the four companies have demonstrated in their cyber defence system.

The application of the CAIBER Framework is demonstrated through its mapping to the AI-enabled tools used by the participant companies. Moreover, the application of the proposed framework is demonstrated through the mapping of the core elements to the Cyber Kill Chain. The significance of the CAIBER Framework is also demonstrated through its application to four case studies of cyberattacks experienced by the companies. The aim of the case studies is to demonstrate how the application of the proposed CAIBER Framework could help remediate cyber threats and enhance cyber resilience.

<b>Contents .....</b>	
CHAPTER 1: RESEARCH OVERVIEW .....	1
1.1 Introduction .....	1
1.2 Background.....	1
1.3 Terminology.....	2
1.3.1 Cyberspace.....	2
1.3.2 Cybersecurity.....	3
1.3.3 Artificial intelligence .....	3
1.3.4 Internet of Things.....	4
1.4 Sources of Cyberattacks .....	5
1.5 Problem Statement.....	6
1.6 Aim and Research Questions.....	7
1.6.1 Research mission .....	7
1.6.2 Research questions .....	7
1.7 Importance of Study .....	8
1.8 Limitations.....	9
1.9 Chapter Outline .....	9
1.10 Conclusion .....	11
CHAPTER 2: LITERATURE REVIEW.....	12
2.1 Introduction .....	12
2.2 Current State of South African Cyber Community.....	12
2.2.1 Status of cyber challenges in South Africa .....	12
2.2.2 Current approaches to cyber defence .....	13
2.3 Interconnectivity of Technology and Associated Challenges .....	15
2.3.2 Internet insecurity .....	16
2.4 Application of AI to Cybersecurity.....	17
2.4.1 Securing cyberspace through AI .....	17
2.4.2 Examples of AI application in cyber defence.....	18
2.4.2.1 Machine Learning .....	19
2.4.2.2 Artificial Neural Network.....	20
2.4.2.3 Deep Neural Network.....	20
2.4.2.4 Intelligent Agent .....	21
2.4.2.5 Artificial Immune System .....	21
2.4.2.6 Generic Algorithms .....	22
2.4.2.7 Fuzzy Logic.....	22
2.5 Conclusion .....	23

CHAPTER 3: RESEARCH DESIGN AND METHODOLOGY .....	25
3.1 Introduction .....	25
3.2 Research Design .....	25
3.3 Research Methodology.....	25
3.3.1 Qualitative method.....	26
3.3.2 Quantitative method .....	26
3.3.3 Mixed method .....	26
3.4 Data Collection .....	27
3.4.1 Methods for data collection.....	27
3.4.1.1 In-depth interviews.....	29
3.4.1.2 Observation.....	30
3.4.1.3 Questionnaire.....	31
3.5 Sampling.....	31
3.6 Ethical and Trustworthiness Considerations .....	32
3.7 Limitations.....	33
3.8 Data Analysis.....	33
3.8.1 Overview of companies .....	33
3.8.1.1 Company_Magix .....	34
3.8.1.2 Company_Pillar.....	35
3.8.1.3 Company_De_Link .....	37
3.8.1.4 Company_Geo.....	38
3.9 Conclusion .....	39
CHAPTER 4: PROPOSED CAIBER FRAMEWORK.....	38
4.1 Introduction .....	38
4.1.1 Background.....	38
4.2 NIST Cybersecurity Framework .....	39
4.2.1 Function 1: Identify .....	40
4.2.2 Function 2: Protect .....	40
4.2.3 Function 3: Detect .....	40
4.2.4 Function 4: Respond .....	40
4.2.5 Function 5: Recover .....	41
4.3 Proposed CAIBER Framework.....	41
4.3.1 Element 1: Monitoring.....	42
4.3.2 Element 2: Identify.....	43
4.3.3 Element 3: Discover .....	43
4.3.4 Element 4: Detect .....	43
4.3.5 Element 5: Investigate .....	43

4.3.6	Element 6: Analyse.....	43
4.3.7	Element 7: Respond.....	44
4.3.8	Element 8: Prevent.....	44
4.3.9	Element 9: Predict.....	44
4.4	Mapping of AI Tools to CAIBER Framework.....	44
4.5	Conclusion.....	46
CHAPTER 5: APPLICATION OF CAIBER FRAMEWORK.....		46
5.1	Introduction.....	46
5.2	Cyber Kill Chain.....	46
5.2.1	Phases of Cyber Kill Chain.....	46
5.2.2.	Evolution of cyberattacks and Cyber Kill Chain.....	48
5.2.3.	Application of AI to Cyber Kill Chain.....	49
5.2.3.1	Stage 1: Preparation (Reconnaissance and Weaponisation).....	50
5.2.3.2	Stage 2: Incident (Delivery, Exploitation and Installation).....	50
5.2.3.3	Stage 3: Active Intrusion (C2 and Actions on Objectives).....	51
5.3	Case Studies: Application of CAIBER Framework.....	52
5.3.1	Case study 1: Attempt by an insider to harvest data.....	52
5.3.2	Case study 2: Storage on cloud server threatens intellectual property.....	54
5.3.3	Case study 3: Email document containing ransomware.....	57
5.3.4	Case study 4: Insecurity of IoT.....	59
5.4	Conclusion.....	62
CHAPTER 6: CONCLUSION.....		63
6.1	Introduction.....	63
6.2	Need for CAIBER Framework to Defend Cyberspace.....	63
6.3	Summative Overview of Research.....	64
6.4	Reflection on Achievement of Research Study.....	65
6.5	Contribution of Study.....	68
6.6	Recommendations for Future Research.....	68
6.7	Closure.....	69

**List of figures**

Fig 1: IoT security risks..... 16

Fig 2: AI for cybersecurity..... 19

Fig 3: Primary research method ..... 27

Fig 4: Methods of data collection ..... 29

Fig 5: Extended sources of cyberattacks ..... 34

Fig 6: NIST CSF ..... 40

Fig 7: Core elements of CAIBER framework ..... 42

Fig 8: Cyber Kill Chain ..... 47

Fig 9: CAIBER framework mapped onto the Cyber Kill Chain ..... 50



**List of tables**

Table 1: Sources of cyberattacks .....	5
Table 2: Interviews overview .....	30
Table 3: Observation overview .....	30
Table 4: List of participants.....	31
Table 5: Cyberthreat remediation guideline .....	36
Table 6: Mapping of AI-enabled tools to CAIBER Framework elements .....	45
Table 7: <b>Case study 1</b> : Mapping of CAIBER Framework to Cyber Kill Chain .....	52
Table 8: <b>Case study 2</b> : Mapping of CAIBER Framework to Cyber Kill Chain .....	55
Table 9: <b>Case study 3</b> : Mapping of CAIBER Framework to Cyber Kill Chain .....	57
Table 10: <b>Case study 4</b> : Mapping of CAIBER Framework to Cyber Kill Chain .....	60

**Acronyms**

ACD	Active Cyber Defence
ACL	Agent Communication Language
AI	Artificial Intelligence
AIS	Artificial Immune System
AISIR	Artificial Immune System Intrusion Response System
APT	Advanced Persistent Threats
C&C	Command and Control
CB	Carbon Black
CNN	Convolutional Neural Networks
CSAIL	Computer Science and Artificial Intelligence Laboratory
CSF	Cyber Security Framework
CSIR	Council for Scientific and Industrial Research
CSIRT	Cyber Security Incident Response Team
CSOC	Cyber Security Operations Centre
DDoS	Distributed Denial of Service
DNN	Deep Neural Network
DNS	Domain Name System
DoD	Department of Defence
DoS	Denial of Service
ESM	Enterprise Security Manager
GAIDS	Generic Algorithm Rule-based Intrusion Detection System
ICT	Information and Communications Technology
IDS	Intrusion Detection System
IDPS	Intrusion Detection Prevention System
INSA	Intelligent and National Security Alliance
IoC	Indicator of Compromise
IoT	Internet of Things
IPS	Intrusion Prevention Systems
IT	Information Technology
KDD	Knowledge Discovery in Databases
KNN	k-Nearest Neighbour
MANET	Mobile ad hoc Network
MIT	Massachusetts Institute of Technology
ML	Machine Learning
MLP	Multilayer Perceptron Neural Network
NGO	Non-Governmental Organisations

NIST	National Institute of Standards and Technology
NLP	Natural Language Processing
NN	Neural Network
POS	Point of Sale
R&D	Research and Development
SA	South Africa
SAE	Stacked Autoencoders
SIEM	Security Information and Event Management
SMB	Server Message Block
SOC	Security Operation Centre
SVM	Support Vector Machine
TOR	The Onion Router
UBA	User Behaviour Analytics
UK	United Kingdom
US	United States
USSD	Unstructured Supplementary Service Data
VPN	Virtual Private Network

## **CHAPTER 1: RESEARCH OVERVIEW**

### **1.1 Introduction**

The Internet's expansion as a new power domain and the development of Information and Communications Technology (ICT) has introduced an unprecedented magnitude of convenience and efficiency to its users. However, as the innovation and novelty increase, so are security challenges and accompanying risks. Cyber attackers are developing Artificial Intelligence (AI)-enabled malware that is adaptive, understand the target environment, have the ability to evade detection, continue to learn and make advanced decisions. In this regard, malware is getting smarter and cyber threats are evolving and becoming more sophisticated and complex. Thus, human intervention and capacity are not enough to sufficiently deal with advanced threats, the speed of processes, the amount of data, and the vulnerability of intrusion. This research study proposes the use of an AI framework to address these advanced threats.

This study explores how AI tools can be used to actively defend the cyberspace. It presents empirical and theoretical knowledge on the prospects of enhancing cyber defence capabilities by means of increasing the intelligence of defence systems with AI tools. The study proposes a framework aimed at enhancing the security posture of organisations and demonstrates the significance of combining AI and cybersecurity. The proposed framework is inspired by the core functions of the National Institute of Standards and Technology (NIST) Cyber Security Framework (CSF) (NIST, 2018:6). The proposed framework is further enhanced from research conducted on the four South African companies interviewed as part of the study.

The proposed framework is also aimed at laying a foundation for future research and investigation on the significance of defending cyberspace through AI. The relevance of the proposed framework is demonstrated by mapping the framework elements on the AI-enabled cybersecurity space of the four South African companies interviewed. This research study was published under the 17<sup>th</sup> European Conference on Cyber Warfare and Security.

### **1.2 Background**

The interconnectedness of technology devices combined with the proliferation of hacker tools demonstrates how computer systems are becoming more prone to security risks. This digital era is not only dominated by smart machines, it is also fuelled by exponential growth and coverage of multiple scientific and technological fields that include big data, Internet of Things (IoT), self-computing hardware, cloud computing, wearable devices, digital currencies, Blockchain distributed ledger systems as well as mobile computing (Talwar and Koury, 2017: 14). The explosion of modern technologies, the growth of users' reliance on universally interconnected technology, and the automation and commoditisation of cyberattack tools demonstrate the complexity of the cybersecurity

landscape (Weber and Studer, 2016: 716).

Cybercrime is a global threat and, despite many years of increased developments in cyber defence, it is still a challenge to manage (Devlin, 2016: online). The cyberattack landscape has shifted considerably (Deloitte, 2017:8) and attackers are constantly changing their cyber campaigns and expanding their range of tools to attack. Moreover, cyber attackers are developing AI-based attacks that increase the speed, scale, sophistication and frequency of their attacks (Talwar and Koury, 2017: 16). AI-enabled malware is adaptive and continues to learn to be more efficient and successful in its attacks. According to Yampolskiy (2017: online), the rise of AI-enabled attacks will cause more automated and increasingly sophisticated social engineering attacks as well as an explosion of network penetrations, personal data thefts and an epidemic-level spread of intelligent computer viruses.

The countering of such advanced adversaries require an active approach that will place an emphasis on proactive measures to security, real-time detection, as well as active monitoring and mitigation of key threats. Therefore, innovative approaches such as the application of learning capable AI tools that are adaptable, analysis-driven and able to detect user behaviour, make intelligent and real-time decisions would assist in fighting the cyber threat. Moreover, the cyberspace comprises diversified aggressors with varied motivations; thus this study proposes a shift in defence surface, one that is inclusive of AI for cybersecurity within the South African context.

### **1.3 Terminology**

This section is intended to assist and enhance the reader's understanding of key concepts that will be used throughout the study. It is also aimed at circumventing any form of ambiguity, confusion or misunderstandings.

#### **1.3.1 Cyberspace**

An early definition of cyberspace was proposed by William Gibson as "a consensual hallucination experienced daily by billions of legitimate operators, in every nation, by children being taught mathematical concepts... A graphic representation of data abstracted from banks of every computer in the human system" (Gibson, 1984). Different versions of the term have evolved over the years. Dorman (2011: 2) argues that cyberspace is both a physical and non-physical environment that encompasses mobile devices, data, servers, routers, fibre optic cables, computer systems, networks, and users. Cyberspace is a dynamic and evolving system that entails physical infrastructure, software, regulations, notions, innovations, and interactions influenced by an increasing number of contributors who represent the range of human intentions (Craig, Diakun-Thibault, and Purse, 2014: 14).

This study has adopted the following definition for cyberspace: *Cyberspace is not just software or*

*networks or computers, rather it is (1) a dynamic operational space used by people/organisations/states to act and create effects either on the cyberspace or across into other domains (2) a domain that is made up of electromagnetic activity (3) information base that allows users to create, store, transform, share and use information in a variety of ways and lastly (4) interconnected networks that carry information (Reveron, 2012: 5; Robinson, Jones and Janicke, 2015: 74).*

### **1.3.2 Cybersecurity**

Cybersecurity is a multidimensional, complex and interdependent concept (Craig, et al, 2014:15). The term has been broadly used; however, its definition often varies since it is context-bound and subjective. There are arguably multiple interlocking discourses around the field of cybersecurity and this lack of a commonly-agreed definition complicates discussions. Von Solms and Van Niekerk (2013:100) described how the concept of cybersecurity has evolved over time from securing information or information systems resources (e.g. defending against malware) to the more general integration of physical and digital domains (e.g. national infrastructure) and securing users who function in cyberspace, whether individuals, organisations or nations.

In support of the above definition, Harel, Gal, and Elovici (2017: 2) proposed a comprehensive definition of cybersecurity that relates to Cyber+Security and a larger Cyber Phenomenon. Cybersecurity refers to all activities that can take place on a computerised platform, with or without the knowledge of the user/owner of the platform, as well as all of the technologies, products, and efforts that can be used to defend against such actions.

This study has adopted the following definition: *Cybersecurity is a collection of tools, policies, security concepts, security safeguards, guidelines, risk management approaches, assurances and technologies that can be used to protect ICT, users and their information from unauthorised access, harm or misuse. Cybersecurity is aimed at ensuring the confidentiality, integrity, and availability of information and communication systems. Additionally, cybersecurity is aimed at deterring, denying and defending information and critical infrastructure from malicious activities of adversaries (Craig, et al, 2014:18).*

### **1.3.3 Artificial intelligence**

The definition of AI has changed over time due to continuous developments in technology and development in the field. In its broadest sense, AI has been described as the study of the computations that make it possible to reason, perceive, learn, make decisions, adapt and act, or the automation of intelligent behaviour (De Spiegeleire, Maas, and Sweijs, 2017: 26). AI has also been described in two ways: (i) as a science that aims to discover the essence of intelligence and develop intelligent machines; and (ii) as a science of finding methods for solving complex problems that cannot be solved without applying some intelligence (Dilek, Cakır and Aydın, 2015: 23).

There are three tiers of AI, which can also be seen as three generations of AI (De Spiegeleire, et al, 2017: 30; Urban, 2015: online):

1. **Artificial Narrow Intelligence** – machine intelligence that equals or exceeds human intelligence on specific tasks. This type of AI is designed to deliver exceptional performance for a specific task, e.g. objects recognition in images. Examples include search engines, High-Frequency Trading Algorithms, IBM's Deep Blue (Chess) and Watson's 'Jeopardy!', Google Translate, spam filters, etc.
2. **Artificial General Intelligence** – machine intelligence meeting the full range of human performance across any task. The goal of Artificial General Intelligence is to create a platform that simulates human cognitive tasks and that generalises across a broad range of circumstances. This type of AI has the ability to reason, plan, solve problems, think abstractly, comprehend complex ideas and learn quickly and from experience.
3. **Artificial Super Intelligence** – machine intelligence that exceeds human intelligence in practically every field, including scientific creativity, general wisdom and social skill, etc. This type of AI has its own consciousness and self-awareness.

In relation to the study *AI is defined as nonhuman intelligence that is measured by its ability to replicate human mental skills, such as pattern recognition, manipulation and Natural Language Processing (NLP), adaptive learning from experience, thinking, planning, strategizing, deduction or reasoning about others (Russell and Norvig, 1997:5). AI is not only a study of computations or a science that develops methods to solve complex problems but it is also a branch of computer science that is focused on discovering the essence of intelligence and develop intelligent machines and software (De Spiegeleire, et al, 2017: 28). In essence, AI is that activity devoted to making machines intelligent.*

#### **1.3.4 Internet of Things**

Internet of Things (IoT) is a term that describes scenarios in which Internet connectivity and computing capability extends to a variety of objects, devices, sensors, and everyday items that include cars, refrigerators, thermostats, health monitors and roads (Weber and Studer, 2016: 717). The study adopted an *IoT definition, which describes it as an extension of the Internet by integrating mobile networks, Internet, social networks and intelligent things to provide better services or applications to users (Li, Li, and Tryfonas, 2016: 338). Moreover, it is a network of interconnected objects and people providing services and sharing data in order to fulfil a certain task in various applications. The range of IoT applications is rapidly increasing and already covers several domains, those include environmental monitoring, healthcare, education, surveillance, smart environment (home, offices, cities) and transportation (Lanotte, and Merro, 2018: 259). However, this study will focus on IoT and its implication on cybersecurity.*

## 1.4 Sources of Cyberattacks

Cyber threats emanate from a wide range of prospective perpetrators and their motives are often diversified. This section is aimed at providing the reader with an overview of different sources of cyberattacks, as well as the motivation described and simplified in Table 1 below.

**Table 1: Sources of cyberattacks**

Sources	Description
Crackers	Crackers are often motivated by fun and the possibility to test their skills or display their capabilities. In some instances, they conduct cyberattacks for bragging rights in their hacker society (De Bruijne, Van Eeten, Gañán and Pieters, 2017: 60).
Cybercriminal(s)/organised crime	These cyber actors include individuals that conduct malicious activities that include stealing and/or distorting sensitive information most often for monetary reasons. These actors often target personally identifiable information of users that include health records, credit cards or banking information (De Bruijne, et al, 2017: 60).
Cyberterrorists	Cyberterrorists are individuals that conduct unlawful acts using cyber systems. Their unlawful and malicious activities are often aimed at instilling fear into their targets or citizens. Moreover, the activities often result in violence, destruction and/or disruption of services and weakening of the economy (Rudner, 2013:460-455). Their motive often also includes influencing the government or population to conform to a particular political, social or ideological agenda.
Hacktivist	Hacktivism use cyber tools to aggravate, provoke and challenge government sites, companies and non-governmental organisations that oppose their moral stance. They often conduct politically motivated attacks. Their modes of attack also include defacing government sites, stealing information or simply intruding just to point out a security flaw of the government (Rudner, 2013:460-463).
Insider threat	Cyberattacks orchestrated by a member of that organisation who could be disgruntled (De Bruijne, et al, 2017: 61). Objectives of these attacks range from financial gains, double agents, religious zeal or manipulation, etc. (Rudner, 2013:460-467).
Script kiddies	These actors are generally less skilled and lack funding. They are not assessed as posing a substantive threat to the wider economy; however, they can cause alarming damage (Enisa, 2017: 96).
State actors	Individuals who are funded by a certain country to penetrate government



Sources	Description
	networks for political, diplomatic, commercial and strategic gain. Their cyberattacks are often aimed at strategic government departments or critical information infrastructures (De Bruijne, et al, 2017: 61). These actors are often well resourced and have advanced technical capabilities to employ sophisticated attacks.

## 1.5 Problem Statement

The South African government is slowly shifting towards the digital space as technology continues to develop; its dependency on it for efficiency is evolving (Nyirenda-Jere and Biru, 2015: 10). However, the reliance on technology also increases the risks of cyberattacks. Vicente (2016: online) avers that various adversaries (including hackers, cybercriminal enterprises, state-sponsored groups, foreign intelligence services and political adversaries) are increasingly targeting government institutions. The protection of cyberspace is challenging for the government because cyberattacks have a low entry barrier, are less costly, simple to conduct, and there is provision for anonymity (Denning, 2009:6). In addition, the South African government does not have monopoly over the cyberspace nor does it regulate it (Reveron, 2012:6). The cyberspace is not an entity that is owned by any individual, state or organisation; consequently, any individual with a mobile device, computer or Internet connection can operate in cyberspace.

South Africa is one of the leading targets of cybercriminals on the African continent due to its relatively high rate of Internet connectivity in relation to other African countries (Davies, 2018: online). The South African government, industry, academia, and organisations depend on cyber systems for the provision of essential services, economic prosperity and communication. Therefore, not only will attacks on the systems threaten the provision of those services, but it will also have a major impact on critical infrastructure, intellectual property, and the privacy of users' data, sensitive national security information, as well as government personnel data. Governments all over the world are constantly under the threat of complex, sophisticated attacks launched by rival nation-states, terrorist groups, hackers, and cybercriminals.

Not only has South Africa been targeted by cyberattacks aimed at state institutions, but the country has also grappled with terrorist organisations trying to recruit its citizens through the use of cyberspace (Davies, 2018: online). The securing of government systems and networks against adversaries remains one of the main challenges faced by the South African government because not only are cyberattacks intensifying, they are also evolving and are becoming sophisticated and complex. Furthermore, the proliferation and increased reliance on mobile phones with Internet access also add another dimension to cybersecurity (Dimension Data, 2017:8).

**Problem statement**

The increased reliance on cyberspace for efficiency, convenience, communication and interaction, as well as the interconnectedness of ICT devices has introduced users to multiple cyber threats. Human intervention alone is not sufficient to manage cyber threats that evolve in sophistication, speed, intelligence and complexity. Therefore, the study proposes an AI framework that is aimed at enhancing cyber defence capabilities and resilience. Research and development of security approaches and measures that are aimed at enhancing cyber resilience will ensure that users and critical information systems are protected; it will facilitate and enhance the availability of information and will ensure its integrity.

**1.6 Aim and Research Questions**

This section will inform what the research study aims to achieve, its objectives and its future goals. It will also deal with the questions the study sets out to address.

**1.6.1 Research mission**

This research is aimed at exploring new, novel and innovative methods of combating cyberattacks and improving cybersecurity in South Africa. It is also aimed at demonstrating how AI can be used as an active defence mechanism to combat cyberattacks. Furthermore, the research will provide empirical results on the prospects of enhancing cyber defence capabilities by means of increasing the intelligence of defence systems with AI tools. The proposed study will be exploratory in nature, with the aim of gaining new insight and depth into the use of AI tools to actively defend South African citizens' cyberspace. Furthermore, this research study will lay a foundation for future research, provide research direction and develop an AI framework that will contribute to a better theoretical and conceptual understanding of using AI to combat cyberattacks.

**1.6.2 Research questions**

The study is focused on exploring how AI can be used to secure cyberspace in South Africa. It is also aimed at expanding understanding and knowledge on the application of AI to cybersecurity, as well as risks the South African citizens can be exposed to in cyberspace. The questions below are outlined to address the objective of this research study.

The primary research question to be addressed by this study is:

Will the proposed AI framework effectively contribute to, and enhance the current South Africa's cyberdefence capabilities?

The secondary research questions to be addressed in this study include:

1. What is South Africa's current approach to cybersecurity?
2. What can be done to improve the current cyber defence employed in South Africa?
3. What kind of AI tools can be used to actively defend cyberspace in South Africa?

4. How will developing and implementing an AI-based framework help enhance cyber resilience in South Africa?
5. How will the proposed AI framework enhance cyber defences currently employed?

## 1.7 Importance of Study

The study is aimed at expanding the understanding and knowledge of AI for cybersecurity and the risks South African citizens could be exposed to in the cyberspace. The lack of maintaining and providing improved, secure, resilient and trustworthy cyber systems undermines confidence in the information society. Public trust in the integrity of financial systems, information networks, and other critical information infrastructure systems is essential for continued economic growth, public safety, and innovation (Gagliardi, et al, 2016:2). Achieving strong cyber resilience requires coordination between academia, government and the public sector. Thus, research into advanced methods for cyber defences, such as AI techniques, to support the development of new capabilities for policy direction and cyber projects instigation is required. Coordination of academia, public and private investments in research and development will help spur the necessary scope and scale of research vital to developing next-generation AI technologies that could be implemented in cybersecurity.

Understanding the current and predicted landscape of cyber threats lays the groundwork for future exploration of options to reduce the risks to which South African citizens are exposed. The translation of novel ideas and approaches stemming from this research study will not only create a strong supply of reliable, tested solutions to cybersecurity threats, but it will also contribute to building a strong security posture. The recommendations of the research will be focused on building a trustworthy, self-improving and resilient cyberspace that will thrive in the face of unanticipated, complex, advanced and sophisticated cyber threats.

South Africa is currently a magnet for cyberattacks, with hackers set on stealing data (TimesLive, 2018: online). According to Dimension Data's (2017:8) *Global Threat Intelligence Report*, attacks on government organisations increased sharply in 2016 compared to 2015 and this is mainly because government agencies hold vast amounts of sensitive information, ranging from personnel records, budgetary data, and sensitive communications to intelligence findings. For this reason, this research study is aimed at promoting research and development into AI techniques that will enhance cyber resilience in South Africa. This study will also generate advances that will assist cybersecurity to keep up with the evolving cyber risks.

Cyber resilience will foster better communication among cyberspace users and will create and enhance cybersecurity culture in South Africa. The research study will also aid in creating a trusted and resilient digital environment. Moreover, it will aid in building a digital society that is not only resilient to cyber threats but one that is equipped with capabilities required to maximise opportunities and manage cyber risks.

## 1.8 Limitations

The use of AI in cybersecurity is dual and relates to two opposing objectives. The first objective is focused on AI-controlled systems as potential targets of cyberspace, mainly due to their increasing role in controlling significant and complex systems. The second objective addresses AI as an imperative field to assist in identifying, combating and countering cyberattacks or cyber-related risks. The scope of this study is limited to addressing AI as an imperative field to enhance cyber resilience.

Information that relates to cyber threats and cyber defence mechanisms (of the companies to be interviewed) may be considered as sensitive/private and thus restricted to the public view. Therefore, this form of restriction will limit the researcher's access to information. In relation to government entities, certain information that relates to cyber incidents could be of national interest, creating another limit to the kind of information that one can access or utilise for the research study. This research is AI and cybersecurity-driven; therefore, it might be a challenge to find companies that have advanced AI capabilities for their cybersecurity. Moreover, it might be a challenge for the researcher to find companies in South Africa that are at an advanced stage in terms of developing and implementing their own AI for cyberdefence.

Other limitations that might potentially affect the study include time, skills, capabilities and lack of reliable and available data. When conducting a research study using mixed methods, the supplementary time has to be allocated to the research study for multiple unforeseeable interferences. In relation to resources, there is limited literature or research that has been conducted on AI and cybersecurity combined; thus, resources and literature study is limited. This research is largely dependent on interviews and observation and therefore this form of data collection method might potentially limit the researcher. For instance, the researcher could be restricted by the availability of participants to be interviewed. The withdrawal of participants in the middle of the study might be a hindrance, as that would require the researcher to restart the process of recruiting participants.

## 1.9 Chapter Outline

This section will present a summary of the different chapters of the research study. It will also present a table that will summarise the chapters and their intended outcomes. The section will demonstrate how these chapters are aligned and how they relate to the research questions.

- Chapter 1 – Research overview: This chapter serves as a guideline for the study. It also enlightens readers about the motivation of the study, problem statement, terminology that will be mostly used throughout the document, as well as the significance and value that the study will add to the literature and cyber sector.

- Chapter 2 – Literature review: The undertaking of the literature review will allow the researcher to acknowledge and appreciate in the literature conducted on AI, cybersecurity and IoT. The literature review will outline ideas, theories and significant literature published within the cybersecurity and AI field. It will also give the researcher a platform to relate other research or articles written on cybersecurity in South Africa with the study. The researcher will be in a position to identify data sources that other researchers have used; identify the relationship between concepts; gain insight on other notions of AI and cybersecurity; understand and highlight significant concepts within the field, both AI and cybersecurity; and identify research methods used in previous research.
- Chapter 3 – Research design and methodology: This chapter will detail the research methodology employed in the study. The objective of this chapter is to reveal why certain methods are selected as the most suitable for the research over many other alternative methods. The chapter will also detail data collection and data analysis methods used the research study. It will reveal why other methods in the study were selected over others. Additionally, it will investigate different tools utilised by the four companies and the extent to which AI is utilised to defend the cyberspace. It will also identify other gaps and limitations that could potentially form part of future research.
- Chapter 4 – Proposed CAIBER Framework: Chapter 4 will provide a brief overview of the National Institute of Standards and Technology (NIST) Cyber Security Framework, which is key in the formation of the proposed framework for the study. Subsequently, it will also provide a comprehensive description of the proposed AI framework for cybersecurity, as well as its elements. The proposed framework will demonstrate the significance of applying AI to enhance cyber resilience. Lastly, the chapter will demonstrate the application of the proposed framework through mapping of elements (of the proposed framework) to AI-enabled tools used by the four companies interviewed.
- Chapter 5 – Application of CAIBER Framework: In this chapter, an AI framework for the use of AI in cyberdefence will be developed. The framework will entail different concepts, assumptions, key factors, expectations, variables, beliefs, and theories that will support and inform the research study. The framework will be presented in both graphics and narrative forms. The application of the proposed framework will allow the researcher to transition from simply describing a phenomenon observed to its actual application. It will aid in identifying future work that could be conducted in order to enhance the findings of the study.
- Chapter 6 – Conclusion and Recommendations: This chapter will summarise the key findings of the study. It will also provide recommendations for future research. It will also provide a summative overview of the study.

## **1.10 Conclusion**

Chapter 1 addressed the key concepts of the study and also provided the reader with an overview of the different sources of attackers and diversified motives. The problem statement was defined in this chapter, the purpose of conducting the study, as well as research questions that the study sets out to address. Moreover, the reader was introduced to the significance of conducting the study, as well as foreseeable limitations that could possibly restrain or delay the research study.

Chapter 2 will discuss and analyse the literary body of knowledge with the aim of determining what is known and not known about cybersecurity and AI. This chapter will focus on the current state of cyber community in South Africa; it will look at newer technologies that include IoT and the impact on cybersecurity. Furthermore, it will focus on the significance of using AI for cybersecurity and lastly detail some existing applications of AI techniques for cyber defence.

## **CHAPTER 2: LITERATURE REVIEW**

### **2.1 Introduction**

The purpose of this chapter is to investigate and present evidence on advances made thus far in the application of AI for cybersecurity. The information will demonstrate how AI tools can be effectively applied for detection, active monitoring and prevention. It will also provide scope for future work in the application of AI on cyberspace defence. This chapter will, moreover, provide an analysis of the existing literature and demonstrate how the literature informs this research.

The first section of the chapter is focused on providing an overview of different approaches currently employed to cyber defence. The second section details challenges within the cybersecurity field caused by the expansion of the Internet, IoT, and interconnectivity of technology. The third section details the significance of using AI to enhance cyber resilience. The section also presents related work and some existing applications of AI techniques for cyber defence.

### **2.2 Current State of South African Cyber Community**

This section will examine the status of cyber challenges in South Africa. This will be done by discussing the most prominent attacks launched against specific entities in the country. The section will also focus on the different approaches used for cyber defence.

#### **2.2.1 Status of cyber challenges in South Africa**

Cyberattacks like malware and ransomware continue to pose a major challenge for most commercial, government and academic institutions (Fralely and Cannady, 2017: 1) in South Africa. The cyber environment, coupled with ICT developments, has turned the space into an attack space, which extends the conventional landscape to a virtual domain where key economic and national security assets are exposed to significant threats.

Cyberattacks are a massive threat to the South African business sector, as this was demonstrated by the recent ransomware attack on Liberty Holdings (Mahlaka, 2018: online). According to Liberty, the company became aware of the incident after the attacker alerted them that they had seized their data and were demanding payment for it (Mahlaka, 2018: online). With regards to the public sector, the official website of The Presidency, [thepresidency.gov.za](http://thepresidency.gov.za), was defaced by a group called the "Black Team" on 07 July 2018. The group left a message on the site that reads, "Hacked by Black Team. Sahara is Moroccan. And Morocco is ur Lord!" The Presidency's website is where most South African citizens access statements from the President and the government (Mngadi, 2018: online).

In 2017, the WannaCry Ransomware attack targeted Telkom service systems, including the Unstructured Supplementary Service Data (USSD) menu, smartphone app and call centre

(Vermeulen, 2017: online). The WannaCry ransomware attack infected Telkom computers that were running Windows and had not been patched with the latest updates. One of South Africa's largest web hosting companies, Hetzner SA, was hacked exposing millions of customers' information that include banking information, domain names and back-end logins to websites (Venktess, 2017: Online).

In South Africa, cybercriminals have increased their attacks (Duncan, 2016: online) due to lack of jurisdiction and anonymity among others, but the most critical is due to inadequate security control. Information Technology (IT) News Africa (2016: online) revealed that over 8.8 million South Africans fell victim to cybercrime in 2015, while it was estimated that cybercrime cost South African companies around R5.8 billion in 2014. IT has largely increased online activities and opportunities; however, these opportunities have also introduced great risks to users. Cybercrime is the fastest-growing economic crime, with a third of companies affected, and South Africa is said to be a global leader in economic crime, with 69% of companies affected (Van der Merwe, 2017: online).

Africa as a continent is moving into a virtual world that is becoming less protected. A report by the United Nations Economic Commission for Africa (UNECA 2014:2) noted that Africa was prone to cyber-related threats. Moreover, it is prone to cyberattacks because it has an increased number of domains coupled with weak networks and information security (UNECA, 2014:2). In support of the above, Wolden, Valverde, and Talla (2015: 1846) stated that cyberattacks pose a great danger to many organisations, particularly those that embrace the use of modern technology.

### **2.2.2 Current approaches to cyber defence**

The increased number of businesses embracing digitisation, use of IoT, mobile devices, and cloud technology have been the driving forces for the expansion of the attack surface (Cisco, 2017:10). However, in this constantly evolving digital threat landscape, firewalls and antiviruses are considered tools of antiquity as they are unable to keep pace with the rapid development and mutation of new threat vectors. Firewalls are no longer enough to protect the content of systems and rules cannot pre-emptively defend against all possible attack vectors. Signature-based detection methods fail repeatedly. Moreover, the exfiltrated data is typically encrypted, rendering rule-based network intrusion tools and firewalls to be ineffective. Some adversaries also use an anonymous network for Command and Control (C&C) which makes it difficult for the security analysts to trace the traffic. For instance, Onion Ransomware uses The Onion Router (TOR) network to communicate with its C&C. Firewalls have been able to form a baseline of what normal activities are across the network and then use certain traffic pattern rules to decide whether or not a certain traffic pattern or network flow fits that rule. However, these traditional firewalls are no longer effective in the hybrid cloud environments business now operate in, where users are connecting from any number of locations or devices on the Internet (Arshia, Gayathri, and Manaswini, 2017: 52).



Some organisations secure their assets through the use of **Intrusion Detection Systems (IDS)**, among other cyber defences (McElwee, Heaton, Fraley and Cannady, 2017: 1). However, the shortfall of the IDS is that they use signature-based detection which focuses on known traffic data in order to analyse unwanted traffic. Moreover, IDSs generate large numbers of security alerts that require manual review, which can be overwhelming and time-consuming (McElwee, et al, 2017: 1). For this reason, potential events and compromised hosts could be missed. Thus, machine learning is a viable approach to reducing the false positive rate and improving the productivity of security analysts (Feng, Wu, and Liu, 2017:173).

An **Intrusion Detection Prevention System (IDPS)** has limited capabilities in providing enhanced cyber resilience. The network environment is continuously changing and this makes it difficult for an IDPS to accurately define patterns of normal and abnormal systems behaviour. This, in turn, leads to false negative detection, failure to detect threats in advance, as well as false classification of malicious network activity (Shah and Issac, 2018:13). An IDPS is not adaptable, it is unable to process and analyse large amounts of data quickly; it requires constant human supervision and it also lacks automation. Analysts have to manually analyse log data, readjust systems to changes in the network environment and also determine the appropriate reaction for cyber threats (Wirkuttis and Klein, 2017: 109). The drawback of IDPS also includes its inability to identify and characterise new attacks and to respond to these in an intelligent manner.

In the cybersecurity environment, endpoint security methods, **Security Information and Event Management (SIEM)**, and sandboxes are deployed to enforce specific policies and provide protection against certain threats. SIEM tools are used to detect suspicious activity on their networks for analysis and incident investigation. These tools form an important part of an organisation's cyber defence strategy, but they are insufficient in the new age of dynamic cyber threats (Darktrace, 2017b: 3). SIEM tools have limitations, which make them incompetent especially without additional tools and personnel (Murzina, 2016:10). One of the shortcomings of SIEM technology is its dependence on human capacity as it is rule-based and expert-described (Sheridan, 2017: online). As a result, SIEM is prone to human error and unimportant information overload, which can become a burden for analysts as it requires them to spend hours sifting through millions of unimportant alerts to find the one threat to act upon (Maher, 2017:9). Moreover, security teams need to continually build processes and make correlation rules to truly benefit from SIEM capabilities as the cyber threat landscape evolves.

Businesses ordinarily have numerous information security tools that operate in silos and SIEM technology will need to connect these silos, as well as automate processes and investigations across these tools. SIEM solutions often lack granular details about events hiding behind a load of raw logs, which makes it challenging for analysts to quickly identify incidents, pivot through indicators and retrieve appropriate data (McElwee, et al, 2017: 1). Furthermore, SIEM solutions produce reports that contain too much noise (Sheridan, 2017: online).

Businesses often use more than one security vendor to address their cybersecurity needs. For instance, some business employ SIEM technology, which they integrate with endpoint protection technology and User Behaviour Analytics (UBA) solutions that use AI, while others employs SIEM technology and integrate it with a cloud-based security platform for IoT and a security intelligence platform (including UBA) that complements SIEM with machine learning capabilities. Hence, the role of AI for cyber defence in this regard is imperative.

## **2.3 Interconnectivity of Technology and Associated Challenges**

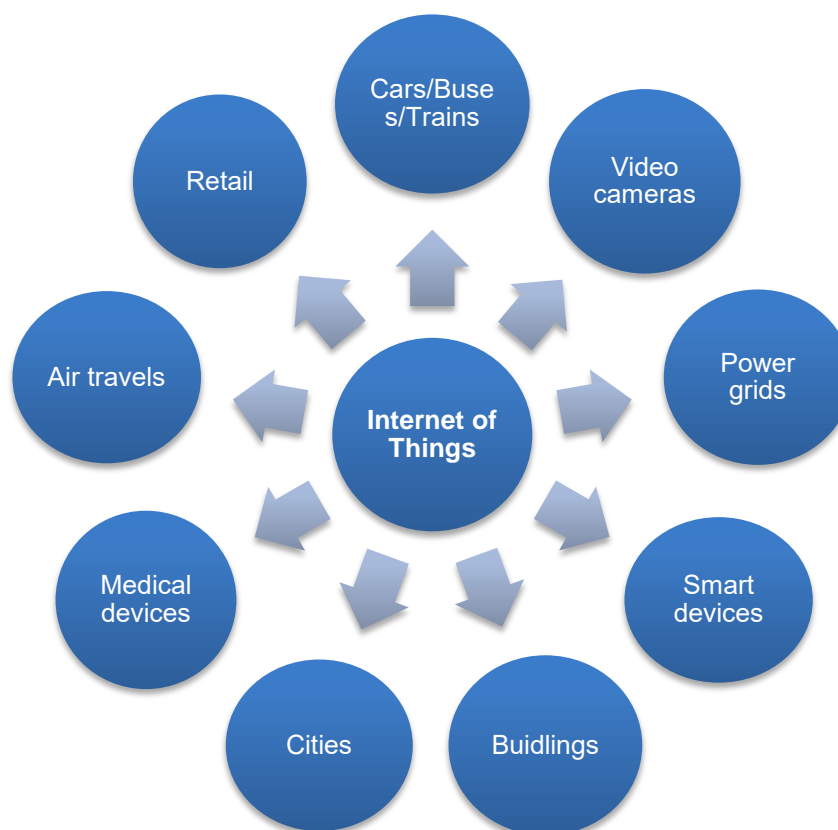
The cyberspace is arguably difficult to secure due to the use and expansion of the Internet, continuous technological developments, interconnectivity between cyberspace and physical systems, as well as the development and growing use of mobile devices. The following section will detail the impact of the aforementioned technologies in the security of cyberspace.

### **2.3.1 IoT insecurity**

IoT technology permits universal and abundant connectivity of different types of devices at a given place and time. However, this connectivity has negative and positive consequences, the negative being the compromise of privacy, while the positive being the efficiency and productivity enabled by these devices. For instance, in Rwanda, SIM cards are connected to Point of Sale (POS) terminals in areas that are isolated in order to accommodate the use of credit card payments. In South Africa, smart meters are already being installed to measure energy consumption. IoT technology is also being used in Eastern and Central Africa to protect endangered Black Rhinoceroses from poachers. This IoT technology (embedded into the Rhino's horn and ankle) is used to locate the animals and can also be used to monitor the Rhino's vitals (Symantec, Online: 2016: 8).

IoT has given rise to improved digitisation of personal information, networking of technologies, increased global connectedness and the networked society but it has also exposed users to exploitation. The network heterogeneity and ubiquity of IoT devices with capacity for surveillance, communication, storage and retrieval of user data have amplified demands on both security and privacy protection. IoT devices have a significant impact on the privacy of users, as there is great potential for surveillance without the users' knowledge or consent. IoT further allows the organisations and third parties to collect, store and analyse information of users and their environment for their own benefit (Caron, et al, 2016: 6).

Caron, Bosua, Maynard, and Ahmad (2016:4) stated that IoT heralds a new era of computing where almost every imaginable object is integrated or interconnected to a smart device that automates the sharing of information and communication. Figure 1 below demonstrates the hyper-connectivity introduced by IoT technology and also demonstrates how the cyber risk landscape is evolving with newer technologies. Moreover, the figure provides a broader structure for identifying and managing a greater range of risks that will arise from the implementation of IoT.



**Fig 1: IoT security risks (Buntz, 2017: online)**

The interconnectivity of people, devices, and organisations in this digital world opens up a new field of vulnerabilities and access points that cybercriminals can utilise. The explosion of new technologies, the growth of users' reliance on universally interconnected technology together with the automation and commoditisation of cyberattack tools demonstrates the complexity of the cybersecurity landscape (Weber and Studer, 2016:716). Cyberattacks on digitally-connected devices pose a risk of information misuse or damage, unauthorised access to multiple devices and a risk of attackers virtually controlling the devices.

### **2.3.2 Internet insecurity**

The Internet is a critical infrastructure in its own right and it is embedded in almost all other critical infrastructures used for daily functioning. The expansion of the Internet beyond computers and mobile phones into other cyber-physical devices or smart systems has extended the threat of remote exploitation to a host of new technologies. Smart device ownership is growing exponentially in Africa, so is the use of social media and the use of IoT technology. This ubiquity of mobile phones has transformed communication in Africa and has allowed the continent's communication networks to advance to the digital age. However, the steady rise of mobile malware (that mostly targets Android operation systems) is concerning given that 89% of the smartphone market share in Africa runs on

that platform. For example, according to Symantec data, more than one out of every seven mobile devices in Nigeria using an Android operating system is currently infected with mobile malware (Symantec, Online: 2016:8)

Smart devices have become an integral part of the lives of users and users also rely on them to store significant sensitive information. Therefore, any cyberattack to such devices could result in information loss and potentially lead to identity loss. The increasing use of the Internet and smart devices means that the boundary of any private corporation or government is disappearing; as a result, the risk landscape also becomes unbounded. The Oxford Analytica (2017:160) predicted that cyber criminality would increase by 2022 due to dependency on the Internet and lack of security of Internet-connected devices.

The Internet is inherently insecure; moreover, the anonymity that comes with it makes cybercriminals feel more secure (Serianu, 2017: 27). The expansion of the Internet has also given rise to services that are being provided by cybercriminals to groups or individuals with ambitions to conduct cybercrimes. Their services include selling of stolen credit details, intellectual property, malware and other tools. Subsequently, this has made it easy for cybercriminals to outsource skills and tools they do not possess.

## **2.4 Application of AI to Cybersecurity**

This section will examine the objectives of applying AI in cybersecurity, as well as the significance thereof. It will also examine related work and some existing applications of AI techniques for securing cyberspace.

### **2.4.1 Securing cyberspace through AI**

Cyberattacks have advanced to an extent that human intervention is not sufficient to detect, monitor, prevent or even handle the volume of data that needs to be analysed in order to formulate an appropriate response and ultimately remedy the attack (Dilek, et al, 2015: 33). In this digital age, organisations create infinitive amounts of data, both internally and externally (through their partners, stakeholders, suppliers and customers). Human capability is limited to securing or monitoring data breaches or potential threats as systems have now become too widespread, data-laden and unwieldy (Talwar and Koury, 2017: 16).

Information systems have millions of potential combinations of irregularities to detect and timeously remedy, but humans lack the time and capacity to check every single one of them (Tuvey, 2017: online). Thus, the application of AI that is flexible, efficient and can identify cyber threats in real time and with improved accuracy (in contrast to human) is integral to cyber resilience in South Africa. However, the use of AI to actively defend the cyberspace does not mean that the human element will be replaced (White House, 2016:32). Instead, the introduction of AI into cybersecurity reduces the

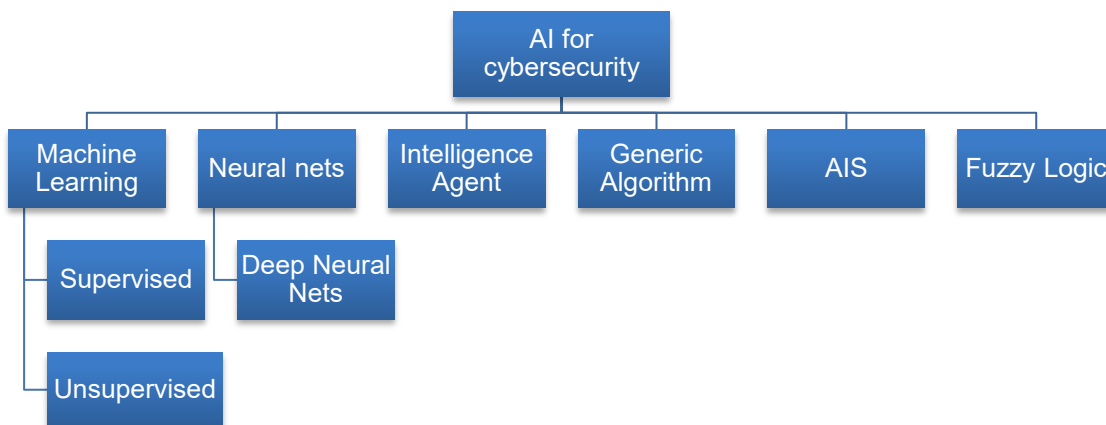
workload of security analysts and allows them to focus on less common events and new social engineering attack vectors, which in turn helps to identify new categories of threats (Tuvey: 2017: online). AI tools are able to analyse the network in real time without human oversight, providing insight and a critical level of accuracy and speed, as cybercriminals get smarter. With malicious software becoming more capable of adapting to linear traditional security solutions, the application of AI-enabled tools will provide users with a system that is able to determine how malware looks like, how it acts and how it may evolve.

One of the reasons for the employment of AI is that cyber attackers are developing AI-based attacks that increase the speed, scale, sophistication, frequency and breadth of their attacks (Talwar and Koury, 2017: 16). In support of the aforementioned, Fraley and Cannady (2017:6) stated that adversaries are now using AI tools that include machine learning to advance their cyberattacks. According to Yampolskiy (2017: online), the rise of AI enables attacks will cause more automated and increasingly sophisticated social engineering attacks, as well as an explosion of network penetrations, personal data thefts and an epidemic-level spread of intelligent computer viruses. Cyberattacks, through AI techniques, have the potential to create new and unprecedented dangers for personal privacy, financial details, social security information, free speech, and any number of human rights (Yampolskiy, 2017: online). Hence, research and development of AI cyber defences is empirical and will enhance cyber resilience.

#### **2.4.2 Examples of AI application in cyber defence**

Information systems (with the advancement of technology and widespread connectivity) are constantly being updated, modified and extended to serve new users and new business functions. Therefore, in such as a fluid environment, it is critical that organisations employ AI-enabled tools that can cut through the noise, detect anomalies and provide other functionalities that will enhance businesses cyber resilience. Moreover, businesses operate in a complex environment where their attack surface is getting larger because of the large data volumes that they produce. Organisations are inundated with masses of network connections and traffic flows, the disappearance of traditional parameters due to the rise of cloud and mobile technologies and cybersecurity events that require thorough investigation and analysis as well as remediation.

The large volume of traffic and events as well the complexity of hybrids cloud networks make it difficult for human beings to monitor, analyse, investigate and make a timely decision, thus the application of AI for enhanced cyber resilience. AI's predictive analytics, detection and UBA provide a powerful use case for network and cybersecurity application. The application of AI in cybersecurity is also ideal for achieving cyber hygiene, reduction of the attack surface at scale and enhancing businesses cyber resilience; however, that application requires a clear understanding of the intended state of an application. Thus, the information below details how AI techniques could be applied for cybersecurity. Fig 2 below provides an overview of the application of AI to cybersecurity, which will be discussed below.



**Fig 2: AI for cybersecurity (own compilation)**

#### **2.4.2.1 Machine Learning**

Machine Learning (ML) comprises computational methods for acquiring new knowledge, new skills and new ways to organise existing knowledge (Tyugu, 2011:6). ML is an application of AI, which provides computers with the ability to reason, solve problems, adapt to the environment, make decisions, etc. ML has been mostly responsible for the recent advances in AI system implementations and this includes enabling AI systems to learn to identify deep, hidden patterns in existing datasets, or learn to match specific features in data with specific responses or outputs (De Spiegeleire, et al, 2017: 40).

In cybersecurity, ML methods promise to enhance network visibility, improve detection levels, enhanced analysis and learning pattern of life, resolve complex and sophisticated problems, as well as discover previously unknown relationships. It also promises to detect patterns in big data and then use the uncovered patterns to predict future patterns of data or detect other kinds of decision-making under uncertainty (Husain and Muhammad, 2013: 31).

#### **Examples of application of ML for cybersecurity:**

The Massachusetts Institute of Technology's (MIT) Computer Science and Artificial Intelligence Laboratory (CSAIL), as well as machine-learning PattenEx, developed an AI platform called AI2. AI2 is able to predict cyberattacks significantly better than existing systems by continuously incorporating input from human experts. The tool uses unsupervised machine learning to monitor, identify and prevent cyber threats (AI.Business, 2016: online). Different companies such as Darktrace (2017) are applying machine learning in the solution in order to increase efficiency and provide enhance cyber resilience for their customer's environment.

Shah and Issac (2018:13) described a study conducted by Firdausi, Lim, Erwin and Nugroho that relates to the use of machine learning techniques to analyse behaviour based malware detection. The malware behaviours were analysed with five machine learning algorithms k-Nearest Neighbour (kNN), NaiveBayes, Decision Tree, Support Vector Machine (SVM) and Multilayer Perceptron Neural

Network (MLP). The analysis of experimental results showed that Decision Tree performs well with 95.9% a false positive rate of 2.4%, a precision of 97.3% and an accuracy of 96.8%.

Shah and Issac (2018:13) further conducted a comparison study of Snort and Suricata, both Intrusion Prevention Systems (IPS) for detection and cyber defence. They discovered that the continued increase in network speed and malicious traffic caused significant challenges for both systems. The IPSs use rules to detect known malicious traffic, however, this was a challenge because both systems could not detect or take any action against unknown malicious traffic. Snort and Suricata both had a common problem, which was triggering false positive alarms. For instance, a legitimate network traffic consisting of Domain Name System (DNS) or web requests could lead the IPS's to trigger a false positive alarm.

However, in order to improve the efficiency and detection accuracy of both Snort and Suricata, a machine learning algorithm was applied. The results demonstrated that the detection rate of both Snort and Suricata increased to 96% with a low false positive ate average of 3%. The application of machine learning reduced the workload of security analysts, duplication of tasks, increased their detection rate of events, and enhanced cyber resilience.

#### **2.4.2.2 Artificial Neural Network**

ANN is a biologically inspired computation device that simulates the structural and functional aspects of neural networks as existing in biological nervous systems like the human brain (Wirkuttis and Klein, 2017: 111). ANN is intuitive, flexible and capable of learning and processing large volumes of data. It can also handle complex nonlinear functions; it is resilient to noise and uncertainty. Neural nets are ideal for situations that require prediction, classifications or control in a dynamic and complex computer environment (Bitter, Elizondo and Watson 2010: 3). They are well suited for learning pattern recognition, for classification, selection of responses to attacks, detection and prevention of potentially dangerous activity (Tygu, 2011: 4).

#### ***Examples of application of ANN for cybersecurity:***

Barika, Hadjar, and El-Kadhi (2009: 4) presented a detailed architecture of a distributed IDS-based on neural nets for enhanced intrusion detection on networks. Bitter et al. (2010: 6) presented host-based and network-based intrusion detection systems with a special focus on systems that employ artificial neural networks to detect suspicious and potentially malicious traffic. In cybersecurity, ANNs have been used successfully in all stages of the cyberattack lifecycle. In contrast to conventional methods used for cybersecurity, the advantage of using ANNs is their ability to learn from past network activities and attacks in order to prevent future ones from occurring (Wirkuttis and Klein, 2017: 111).

#### **2.4.2.3 Deep Neural Network**

Deep Neural Network (DNN) is a more elaborative and computationally complex form of ANN. DNN has been used not only for cyberdefence but also to predict cyberattacks (Wirkuttis and Klein, 2017:

111). DNN has the ability to learn complex functions by mapping the input to the output directly from data, without depending completely on human input.

***Examples of application of DNN for cybersecurity:***

A company called Deep Instinct introduced security software that uses DNN to digest huge volumes of data and prevent against zero-day and Advanced Persistent Threat (APT) attacks. It is also able to process a multitude of data sources, which provides them with timely detection, prediction and prevention abilities of known and unknown cyber threats (AI.Business, 2016: online).

Lotfollahi, Shirali, Siavoshani, and Saberian (2018) presented a Deep Packet framework that automatically extracts features from network traffic using deep learning algorithms to classify traffic. Deep Packet uses deep learning algorithms namely Stacked Autoencoders (SAE) and one-dimensional Convolutional Neural Networks (CNN) to handle both application identification and traffic characterisations tasks. Moreover, it could be modified to handle more complex tasks like multi-channel classification that includes distinguishing between different types of Skype traffic (such as chats, voice call, and video call) and accurate classification of TOR traffic, etc. Contrary to most methods that relate to the use of DNN, Deep Packet can identify encrypted traffic and distinguish between (Virtual Private Network) VPN and non-VPN network traffic.

**2.4.2.4 Intelligent Agent**

Intelligent Agent is a branch of AI that possesses some features of intelligent behaviour that include the ability to learn and make some decisions, as well as adapt understand the Agent Communication Language (ACL) (Tyugu, 2011: 5). Intelligent agents are not only proactive but they also have reactive behaviour; they understand and respond to changes in their environment and are able to interact with it and other agents as well (Wirkuttis and Klein, 2017: 110). Their collaborative nature, mobility and ability to self-adapt to dynamic changes in their environment also make them suitable for cyberspace defence (Dilek, et al, 2015: 32).

***Examples of application of intelligent agents for cybersecurity:***

Ye and Li (2010:1) presented a Mobile Ad hoc Network (MANET), a network security architecture that was aimed at improving security and protecting mobile ad hoc networks. The network security architecture was based on an Artificial Immune System (AIS) and it used two types of mobile multi-agents, i.e. detection agents and counterattacks agents. Their technology combined advantages of the AIS with intelligent agent technology and also had traits of distribution, adaptation, learning, and expandability. Intelligent agents have also been applied to detect and prevent against DDoS attacks.

**2.4.2.5 Artificial Immune System**

Artificial Immune System (AIS) is a subfield of computational intelligent systems, which imitates the biological immune system (Jyothsna and Nilina, 2013:1720). AISs are adaptable, self-organising and have the ability to solve complex and sophisticated problems (Dilek, et al, 2015: 28). The dynamic



structure of AIS allows it to remove malicious activity by the best means available. The AIS has a multi-layered structure. This means that multiple layers of different structures are in charge of monitoring a single point, making it difficult for attackers to succeed with their malicious activities (Qiang and Yiqian, 2010:42).

#### ***Examples of application of AIS for cybersecurity:***

Rui and Wanbo (2010: 86) proposed an Artificial Immune System Intrusion Response System (AISIR) model that would be able to recognise and classify unknown attacks. Their AIS model had a dynamic decision-making mechanism that could adjust its defence tactics in accordance to its changing environment. Their model was able to provide efficient intrusion response and it also had qualities that included rationality, self-learning and quantitative calculation. Kumar and Reddy (2014: 43) developed a unique agent-based intrusion detection system for wireless networks that collect information from various nodes and uses this information with evolutionary AIS to detect and prevent intrusion via bypassing or delaying the transmission over the intrusive paths. The experimental results showed that the system is well suited for intrusion detection and prevention in wireless networks. Zhang, Wang, Sun, Green II, and Alam (2011:796) proposed a distributed intrusion detection system for smart grids (SGDIDS) in order to improve the cybersecurity of the smart grid. This was done by developing and deploying an intelligent module, the Analysing Module (AM), in multiple layers of the smart grid where they used the support vector machine (SVM) and AIS to detect and classify malicious data and possible cyber attacks.

#### **2.4.2.6 Generic Algorithms**

Generic Algorithms provide optimal solutions for complex computing problems and they also adapt very easily to their environment. They are extremely flexible, adaptable, have a robust global search and they have proven to be efficient in solving complex problems (Sharma, Kumar, and Kaur, 2014: 6474). Generic Algorithms are applied to network traffic in order to accurately identify IoT and mobile devices (Harel, et al, 2017:7).

#### ***Examples of application of Generic Algorithms for cybersecurity:***

Ojugo, Eboka, Okonta, Yoro and Aghware (2012:1184) presented a Generic Algorithm Rule-based Intrusion Detection System (GAIDS) aimed at improving systems security, integrity, confidentiality and resource availability in networked settings. The proposed system used a set of classification rules obtained from networked audit data and support-confidence framework, which was used as a fitness function to evaluate the quality of each rule. Zamani and Movahedi (2015:6) noted the use of Genetic Algorithm and decision trees to create rules for an intrusion expert system, which aids and supports security analyst's job by differentiating anomalous and normal activity in the network.

#### **2.4.2.7 Fuzzy Logic**

Fuzzy Logic-based approaches are often used for detecting network intrusions. The main characteristic of Fuzzy Logic-based approaches is the robustness of its interpolative reasoning

mechanism (Sharma, et al, 2014: 6475). It has the ability to reason and to solve complex problems. In contrast to ANN, Fuzzy Logic does not try to mimic the human brain, rather it extracts the essence of the decision-making process of the human (Husain and Muhammad, 2013: 30).

***Examples of application of Fuzzy Logic for cybersecurity:***

Jongsuebsuk, Wattanapongsakorn, and Charnsripinyo, (2013: 2) proposed a network IDS-based on a fuzzy genetic algorithm. Fuzzy rules were used to classify network attack data, whereas genetic algorithms were used to optimise finding the appropriate fuzzy rule in order to obtain the optimal solution. The evaluation results showed that the proposed IDS could detect network attacks in real time (or within 2-3 seconds). Moreover, the system had a detection rate of over 97.5% (Jongsuebsuk, et al, 2013:5).

Zamani and Movahedi (2015:8) described how fuzzy logic could be used to reduce the false alarm rate in determining intrusive activities. This was done by defining a set of fuzzy rules to define the abnormal and normal behaviour in a network and a fuzzy inference engine to determine intrusions. Moreover, they use a genetic algorithm to generate fuzzy classifiers, which is a set of fuzzy rules in the form defined above. Each fuzzy rule was represented by a genome and the Genetic Algorithm was used to find the best genomes (fuzzy rules) to be added to the fuzzy classifier. The results demonstrated that their algorithm achieved an overall true positive rate of 98.95% and a false positive rate of 7%.

Shanmuham and Idris (2009:213) used the Fuzzy Logic machine learning algorithm to accurately detect the anomaly or any form of misuse of information. The Fuzzy Logic was analysed with Knowledge Discovery in Databases (KDD) 1999 dataset and as a result of the research, the IDS Framework was proposed. The IDS Framework improved the Apriori Algorithm that yielded faster rule generation, detection rates for malicious attacks and reduction in false positives.

## **2.5 Conclusion**

This digital age is witnessing the intelligent machine revolution where machines (and humans) are using unpredictable, varied, complex and sophisticated methods for cyberattacks. Attack methods have become more varied and are now specifically individualised. Cyber defences that include firewalls, antiviruses, SIEM technology, IDPSs, and sandboxes are no longer enough to protect the content of systems. Rules cannot pre-emptively defend against all possible attack vectors and signature-based detection methods fail repeatedly. AI techniques as noted in this chapter are more flexible, adaptable, resilient, dynamic, and robust and have the ability to make decisions. Thus AI application to cybersecurity will provide users with enhanced cyber resilience. It will also improve security performance and better protect information systems from an increasing number of sophisticated and automated cyber threats.

The following chapter will outline the research methodology that was used in the study. It will also outline the research approach and study design. It will also detail the data collection methods, sampling procedures and further provide an analysis of the data collected. Lastly, it will look at the ethical considerations, limitations of research and instruments used to maintain validity and reliability.

## **CHAPTER 3: RESEARCH DESIGN AND METHODOLOGY**

### **3.1 Introduction**

In this chapter, the research methodology and design used in the study are described. The chapter will also detail the rationale for the methodology. In more detail, the chapter will present the methods of data collection and analysis, selection of the sample and methods used to maintain the validity. Ethical considerations and research limitations will also be described. The following section will focus on the research design and research methodology

### **3.2 Research Design**

The research design for this study is exploratory and is analysed largely through qualitative methods with a quantitative component. The research design and methodology used were selected in order to address the research questions and mission of the research. Burns and Grove (1999:38) define exploratory research as research conducted to discover new ideas and/or increase knowledge of a phenomenon. Exploratory research can be defined as a form of research that produces initial insight into the nature of certain phenomena and subsequently develop questions or findings that will result in a more extensive study in the future (Marlow 2005:334). This research is aimed at exploring novel and innovative measures for enhancing cyber resilience in South Africa through the application of AI to cybersecurity.

The primary objective of choosing an exploratory design is to gain new insights, in-depth knowledge, fill a knowledge gap and gain a broader understanding of AI for cyber defence. Exploratory research is broad in focus and provides infinite answers to specific questions, which leave room for further research.

### **3.3 Research Methodology**

The research methodology of the research study will be defined by three basic research methodologies: qualitative method, quantitative method, and the mixed method. Each of these methods has their exclusive tools and techniques that a researcher can apply in their research study. A mixed method approach will be used in this study as that will allow the researcher to answer broader, more dynamic and complex range of research questions. Additionally, it will allow the researcher to use the strengths of one method to overcome the weaknesses of another method (Johnson and Onwuegbuzie, 2004:21). It will provide the researcher with richer and comprehensive insights and understanding that might be missed when only a single method is used. The methodologies will briefly be discussed next.

### 3.3.1 Qualitative method

Qualitative research can be defined as methods and techniques of observing, documenting, analysing, and interpreting attributes, patterns, characteristics and meanings of specific, contextual or specific features of a phenomenon (Leininger, 1985:5). The qualitative research part of the study is aimed at exploring and discovering the complexities, differences, unknown dimensions and characteristics of the problem (Philip, 1998: 267). Qualitative methods, (which are usually inductive in nature), are often utilised to gather data for exploratory studies (Babbie, 2010:92).

Qualitative research derives meaning from the participant's perspective and also regards reality as subjective. The use of qualitative research will allow the researcher to generate an in-depth description and narrative that will display a dynamic picture of the respondents' reality. Qualitative research presents data as a descriptive narration through the use of words, in contrast to quantitative research, which presents its results by means of numerical or statistical data. Moreover, qualitative research is inductive, in contrast to quantitative research which is deductive, and it attempts to understand phenomena in natural settings.

### 3.3.2 Quantitative method

Quantitative traditionalists maintain that the researcher should be objective; meaning the researcher should eliminate any biases, and they must remain emotionally detached and uninvolved with the objects of their study (Johnson, Onwuegbuzie, and Turner, 2007:125). One of the major characteristics of quantitative research is its focus on deduction, confirmation, hypothesis or theory testing, explanation, prediction, standardised data collection, and statistical analysis (Johnson and Onwuegbuzie, 2004:18). However, quantitative research alone is limited in nature as it only focuses on one small portion of a reality that cannot be fragmented or unitised without losing the significance of the whole phenomenon (Krauss, 2005: 759).

### 3.3.3 Mixed method

The objective of a mixed method research approach is not to replace the quantitative or qualitative approaches to research, rather draw from the strengths of these approaches and to minimise possible weaknesses (Johnson and Onwuegbuzie, 2004:14). Mixed method researches offer great potential for practising researchers who would like to see methodologists describe and develop techniques that are closer to what researchers actually use in practice.

The rationale for choosing a mixed method research design for this research was to:

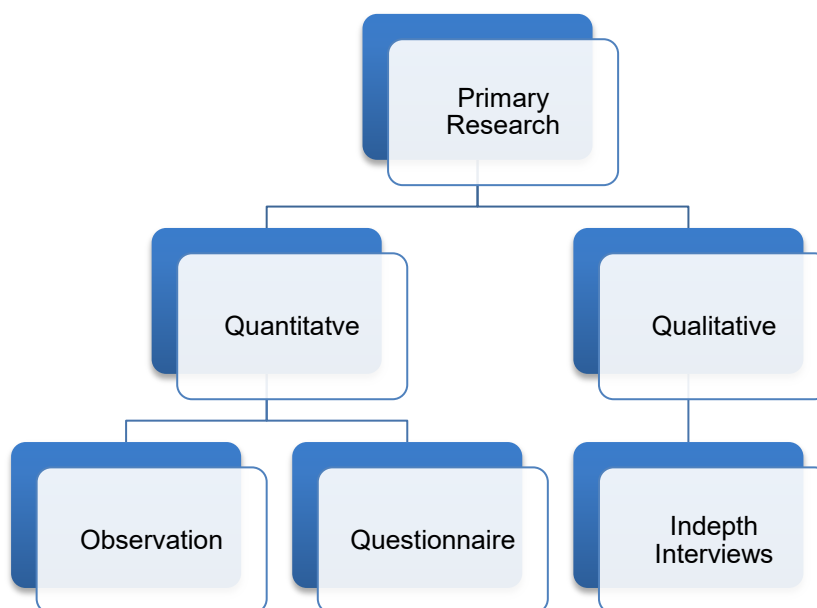
- Generate deeper insights into the application of AI for cybersecurity;
- Facilitate an enhanced understanding of the relationship between AI and cybersecurity;
- Explore distinctive approaches, perspectives, and practices of different participants to be interviewed;
- Obtain comprehensive evidence on the application of AI tools for cyber resilience through the application of both qualitative and quantitative methods;

- Gain a broader understanding of cybersecurity in South Africa and thus get a fuller research picture;
- Allow for unpredicted developments that might lead to future studies regarding AI and cybersecurity; and
- Provide a more elaborated understanding of the AI and cybersecurity within the South African context.

The mixed method was chosen because the researcher wanted to gain greater confidence in the conclusions produced by the research study (Andrew, and Halcomb, 2009: 37 and Johnson and Onwuegbuzie 2004:17). The section to follow will describe in detail the methods used to collect data.

### 3.4 Data Collection

Burns and Grove (1999:744) describe data collection as a systematic way of gathering information that is relevant and significant to the study through the use of diverse methods that range from interviews, observation, case studies and focus groups. The main data collection techniques used in this study were semi-structured interviews, participant observation, and a questionnaire. Fig 3 below is a representation of the primary research methods used in this study and also demonstrates the classification of the respective data collection methods.



**Fig 3: Primary research method (Kumar, 2014: online)**

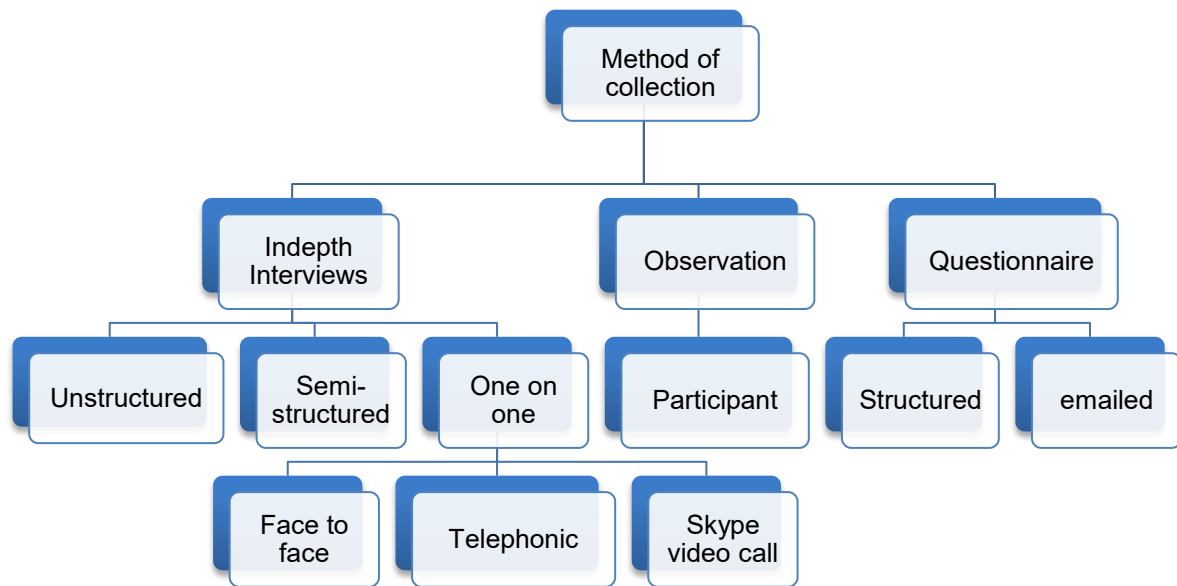
#### 3.4.1 Methods for data collection

The research was conducted on four South Africa-based companies and one virtual company in the United Kingdom (UK). The companies specialise in different sectors; for instance, the first company is

a law and audit firm with a focus on cybersecurity from a risk angle and the second company is a company that provides comprehensive ICT systems and end-to-end solutions to corporate and public sector organisations. The third company offers security monitoring services on security devices and outputs on a 24/7 basis while the fourth company provides managed services to clients and 24/7 Cyber Security Operation Centre (CSOC) technical support to ensure that clients have control over their cyber threat landscape.

The companies were selected based on their cybersecurity approach, which gravitates towards demonstrating the significance of using AI for cybersecurity. Another element of significance is that their cybersecurity approach and scope of services are not limited to South Africa, rather they are global. The companies were also selected because of their lead in cyber defence and because they are slightly advanced in terms of their cyberdefence capabilities. Moreover, their future prospects of using AI for cybersecurity were fitting for this particular research. The companies, however, have different employee sizes, cyber defence capabilities, tools employed for cyber defence and advances on AI plans and implementation for cybersecurity.

To respect the confidentiality and anonymity of the companies, pseudonyms were used in the study. Moreover, pseudonyms were used because the information the researcher was exposed to was sensitive in nature and could jeopardise the integrity and business operation of both the companies and their clients. The companies have a wider variety of clientele, i.e. their clientele is neither restricted to a specific industry nor is it limited to the private and public sector. Moreover, these companies have both international and local clients which include motoring industries, financial services, mining industries, government departments, manufacturing, retail shops and fast food outlets whose information and integrity they want to secure at all cost. The pseudonyms that were allocated to the companies are Company\_Magix, Company\_Pillar (South Africa and the UK), Company\_De\_Link as well as Company\_Geo. Fig 4 describes in detail the methods of collection that will be used in this study.



**Fig 4: Methods of data collection (Kumar, 2014: online)**

### **3.4.1.1 In-depth interviews**

The purpose of collecting data through unstructured and semi-structured interviews was to explore the views of different professionals from across industries focused on cybersecurity. The questions for the interviews were prepared and shared with the interviewees beforehand. The questions were not only prepared to guide the researcher towards the satisfaction of the research objectives but to also make the process seamless and comfortable for the interviewee (a detailed form of the questions is presented in Appendix A). The interviews enabled the researcher to gain deeper insight and understanding of the environment in which the companies operate, acquire more knowledge about the tools they use to defend the cyberspace and extent to which they use AI. Moreover, the researcher was able to uncover or identify the limitations of the AI tools and techniques that they use.

The unstructured and semi-structured interviews allowed the researcher to pose open-ended questions and stimulate discussions around cybersecurity and AI. Unstructured interviews provided the researcher with the possibility of exploring richer ideas of participants. Moreover, semi-structured interviews are flexible and allowed the researcher to explore the interviewee's opinions or perspectives about the use of AI for cybersecurity. One of the disadvantages of both these methods is that they are time-consuming and difficult to analyse. However, they are insightful and encouraged two-way communication. Table 2 below describes the number of interviews for each company, as well as their length.



**Table 2: Interviews overview**

<b>Companies</b>	<b>Number of interviews</b>	<b>Length of interview</b>
Company_Pillar_SA	2 interviews	2hrs
Company_Pillar_UK	1 interview via Skype	2hrs
Company_Magix	1 interview	1hr 30min
Company_De Link	3 interviews	4hrs (on separate days)
Company_Geo	1 interview	1hr 30min

### **3.4.1.2 Observation**

According to Leedy and Ormrod (2013: 152), the primary advantage of conducting observations is flexibility. The researcher was allowed into the Security Operation Centre (SOC) located in the premises of the company for observation. The researcher was exposed to the organisations' daily operations, daily routines, different cybersecurity processes, as well as the mechanisms they apply for cybersecurity. The observation allowed the researcher to learn about the activities of the participants in their natural setting through observing and participating in those activities (Kawulich, 2005: online). This collection method heightened the researcher's awareness of significant processes and events.

The observation process and exposure to the participants' natural setting or environment provided context for further development of interview questions. Observation also eased the facilitation of the research process and interviews as the observer easily assimilated into the participant's environment. Thus, it was easy for the participants to inform the researcher about their future endeavours into AI for cybersecurity. Moreover, the observation process allowed the researcher to understand the definitions of terms that participants use in their environment, observe events that participants were unable or unwilling to share during interviews and observe situations that they had described in interviews. Through observation, the researcher was also able to identify exaggerations, distortions or inaccuracies in the description of certain events provided by participants during the interviews.

Observation among the four companies varied in terms of the time they permitted the researcher to be in their SOC. Table 3 details the overview of the observation process. In relation to Company\_Pillar UK, no observation was conducted because they utilise the SOC that is at the head office of Company\_Pillar\_SA.

**Table 3: Observation overview**

<b>Company</b>	<b>Observation</b>	<b>Length of interview</b>
Company_Pillar_SA	SOC	3 hours 30mins

Company	Observation	Length of interview
Company_Magix	SOC	3 hours (on separate days)
Company_De Link	SOC	4 hours (on separate days)
Company_Geo	SOC	1 hour 30mins

### 3.4.1.3 Questionnaire

A comprehensive questionnaire was only forwarded to participants when an information gap was identified or when the participant was unable to schedule a follow-up interview. This method of data collection is cost effective and the participants could complete the questionnaires at their convenience. The questionnaire was less time consuming and it required less analysis as the questions were straightforward and unambiguous (a detailed form of the questions is presented in Appendix B).

Questionnaires were sent to Company\_Geo and Company\_Pillar (SA and UK). In relation to Company\_Geo, the questionnaires were sent via email to both the General Manager and Business Development Manager. Questionnaires were also sent via email to the Global Head of Cyber Defence at Company\_Pillar\_UK and to the SOC Manager at Company\_Pillar\_SA. There was no deadline allocated for the return of the questionnaires mainly because the researcher wanted them to complete the questionnaire in the comfort of their own time and space. Through the interviews, the participants were made aware of the timeline of the research study. The following section will provide a thorough content analysis of the four companies interviewed.

## 3.5 Sampling

Purposive sampling was used to define the research sample. The participants were selected based on their positions at their companies, as well as the experience they have on cybersecurity and AI. They were also selected based on their ability to enhance the study and from whom the most could be learnt. The individuals that were selected had relevant work experience in cybersecurity. They also had a comprehensive understanding of their environment, their clients and cyber defence systems and capabilities employed in their company. Table 4 below details the individuals that were interviewed, their positions, the company and location of their place of operation.

**Table 4: List of participants**

Entities	Participants	Department	Location
Company_Pillar	1X SOC Manager 1X Data Analyst	Cyber Defence	South Africa
Company_Pillar_UK	1X Global Head of Cyber Defence	Cyber Defence	United Kingdom

Entities	Participants	Department	Location
	1X Data scientist		
Company_Magix	1X Associate Director in Cyber and Technology Risk within Risk Advisory Africa	Cyber Intelligence Centre	South Africa
Company Delink	1X SOC Manager 1X SOC Assistant Manager	Security Operations	South Africa
Company_Geo	1X General Manager 1X Business Development Manager	Cyber Centre	South Africa

### 3.6 Ethical and Trustworthiness Considerations

The participants were informed about the study prior to their interviews and were also given assurance about the research's ethical principles, which included anonymity and confidentiality. The anonymity and confidentiality of the participants were maintained through the removal of any identifying characteristics (such as their names and surnames) preliminary to the propagation of any information. Confidentiality and anonymity were also guaranteed by ensuring that the information provided and participants' details were not made accessible to parties other than the researcher and supervisors. Moreover, the sensitivity of the information being handled by the participants was considered, hence information that includes the technology they are currently working on and the names of their clients were not revealed in the study.

Prior to the interview, the participants received a document that included questions and discussion points. This was done so that they could familiarise themselves with what would be discussed. The participants were also informed about the purpose of the study, the methods being used, as well as objectives. They were also informed about associated demands or inconveniences that might arise from their participation, for instance, taking time off work and helping in facilitating the observation process. It was communicated prior to the participants that the research was only for academic purposes and that their contribution was purely voluntary.

Prior to the observation, the researcher obtained permission from relevant authorities as the observation process was done in their areas of daily operation. The researcher respected the participant' rights and privileges to withdraw from the study at any time. Moreover, the information provided was treated with the highest confidentiality and regard in terms of storage, analysis and handling.

### **3.7 Limitations**

The sample size for all companies was limiting, consequently making it a challenge to explore other areas of the companies' cyber defence capabilities. It was a challenge for the researcher to obtain more participants for the study mainly because the information required for the study was considered private and confidential for some companies. Moreover, key players in industries that include banking, internet providers, telecommunications and government department responsible for cybersecurity were approached for the study; however, the researcher was unable to secure their participation.

The in-depth interviews were both time consuming and costly as the researcher had to travel to the different companies for interviews and observation. However, these processes allowed the researcher to elicit more information and explore a greater depth of meaning and understating. In addition, the observation process was time-consuming but it was rich in information and provided the researcher with a deeper understanding of the context the companies operate in. The vast amount of data collected created ordering and interpretation challenges. The content analysis was also challenging, as the vast amount of data was unstructured. Scheduling follow-up interviews was a bit of a setback mainly because of the time factor and considering the nature of the participants' daily functions. Thus the researcher forwarded a comprehensive questionnaire to the participants.

### **3.8 Data Analysis**

Content analysis was used to analyse the data gathered from in-depth interviews and observation. The analysis process is aimed at presenting data in a comprehensible and interpretable form. Marshall and Rossman (1999:150) describe data analysis as the process of bringing order, structure, and meaning to the mass of collected data. It is the activity of making sense of, interpreting and theorising data that signifies a search for general statements among categories of data (Schwandt, 2007:6). The following section will focus on the analysis of data, and will also provide an overview of AI-enabled tools used by the companies.

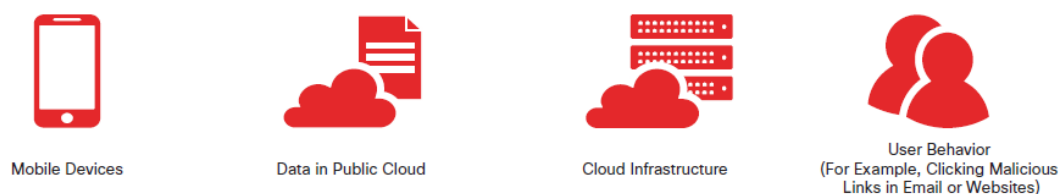
#### **3.8.1 Overview of companies**

The companies participating in this study believe that every user is exposed to cyber threats, with the main motivation being monetisation, distorting the integrity of information, hindering of service delivery and stealing of sensitive information for personal identity. The following section will focus on the cyber threat overview of all the companies interviewed, classification of the threats, as well the attack surface in their environment. All the companies use more than one security vendor to address their cybersecurity needs. They use different security vendors for their capabilities, strengths and talent. It is worth noting the size, capacity and capabilities of all four companies interviewed differ, hence other companies use more tools for cyber defence over other companies. Additionally, these companies serve different clients, with unique cyberthreats and risks, unique security capabilities and diverse budgets thus the use of multiple tools. Company\_Magix and Company\_Pillar echoed their frustration

at times of using multiple tools for cybersecurity because of duplication of activities however, the companies stated that they are working towards implementing their own integrated AI-enabled solution for cyber resilience.

### 3.8.1.1 Company\_Magix

Company\_Magix stated that their South African clients were often victims of targeted attacks. They regard the financial services and a smaller percentage of retail clients as the most targeted sectors followed by the manufacturing and mining industries. According to Company\_Magix, some of the cyberattacks their clients experienced include distortion of sensitive information, theft of intellectual property, phishing attacks, and DDoS attacks. The company noted an increase in ransomware attacks on their clients between 2016 and 2017. While ransomware continues to pose the biggest threat to their clients, other threats that include IoT and smartphone malware are starting to be prominent. The increased use of smart mobile devices and mainstream adoption of cloud and IoT technologies have opened up new platforms and users for attackers to target. Fig 5 below demonstrates the company's biggest source of concern in relation to cyberattacks.



**Fig 5: Extended sources of cyberattacks (Cisco, 2017:10)**

Company\_Magix has established a Malware Information System (MIS) platform that allows them to share information such as Indicators of Compromise (IoC) with their clients. Information that is on the MIS is validated and updated on a daily basis. The company has different MIS platforms for every sector, as well as a central MIS platform that connects their entire client spectrum, as some sectors are interlinked.

Company\_Magix uses different security vendors for their capabilities, strengths, and talent. For instance, the company uses Archsight<sup>1</sup> as its SIEM technology and it has deployed Cybereason<sup>2</sup>, Exabeam<sup>3</sup> and SNYPR<sup>4</sup> on service endpoints, IoT, cloud and other information systems mainly for detection, analysis, investigation, real-time response, prevention of cyber threats, enhanced management of cyber risks, as well as prediction.

<sup>1</sup> <https://software.microfocus.com/en-us/software/siem-security-information-event-management>

<sup>2</sup> <https://www.cybereason.com/>

<sup>3</sup> <https://www.exabeam.com>

<sup>4</sup> <https://www.securonix.com>

Company\_Magix has deployed three AI-enabled tools because their SIEM technology was not provided with a comprehensive security picture within their environment. In addition, the company stated that their network perimeters expanded and now includes newer technologies such as tablets, mobile devices, wearable technologies, as well as cloud and IoT devices. The company also indicated that their cyber risks and trends were constantly increasing and evolving; thus, they needed cybersecurity measures that would provide them with enhanced cyber resilience, and AI-enabled tools provide them with that. Therefore, the AI tools they deployed have demonstrated to them the significant role of AI in cybersecurity; hence, they are central to their cyber defence approach. Company\_Magix emphasised their use for the Cyber Kill Chain for cybersecurity. They also stated that their use was motivated by their need to understand the steps that an adversary takes in order to eventually launch an attack.

Company\_Magix uses a variety of commercial security products and the first one is Exabeam which complements the company's SIEM with machine learning, algorithm and automation. This AI tool monitors events and employees through time, thus making it possible for the organisation to uncover and investigate suspicious events when they occur. Exabeam's user behaviour analytics solution leverages existing log data to detect advanced attacks, prioritise incidents and guide effective response. Additionally, it uses unsupervised machine learning to automatically and continuously learn employee's behaviour over time.

Cybereason is a security platform that allows Company\_Magix to continuously monitor their systems with the aim of detecting the adversary's actions or intentions. It is also a solution that is focused on gathering and analysing behavioural data. Cybereason is an endpoint protection tool with a strong focus on malware (Stephenson, 2017a: online).

Securonix SNYPR is an analytics platform that uses a combination of context enrichment, machine learning and threat modelling to predict, detect and contain advanced threats, anywhere in real-time. SNYPR leverages sophisticated machine learning algorithms to accurately identify the most hard-to-detect cyber threats, insider threats and fraud. SNYPR enables detection of privilege abuse, data exfiltration, sophisticated malware and Advanced Persistent Threats (APT).

### **3.8.1.2 Company\_Pillar**

Company\_Pillar has a diverse clientele that ranges from the automobile and financial sector to the government. The cyber threats vary from industry to industry; for instance, they would target the Chief Executive Officer in some industries while in other cases they would target tellers within the bank and offer them money in exchange for something that might help them with their operation. The company noted that some of the government departments they serviced often became victims of spear phishing and DDoS attacks. The main motive of these attacks is preventing the government from rendering its services. According to the company, government departments often became victims of cyberattacks

because of lack of visibility in their environment, lack of proper security measures, as well as a poor security posture.

Company\_Pillar is part of a global cyber community that often updates them on the latest cyber threats. For instance, they knew about the WannaCry attack a few months before it occurred. As a result, they scanned their client's environment to see if there were any clients that were vulnerable, then fixed patches, and put in place other security measures before the actual attack occurred.

The organisations classify their cyber threats based on severities, which range from Priority 1 (P1), being the most critical, to P4, the lowest. Table 5 below is an example of Company\_Pillar's cyber threat remediation guideline. The severity of the threats also differs in time and, as such, the table is a guideline since some P1 threats are resolved in less than four hours. The attacks can occur asymmetrically or on a day-to-day basis.

**Table 5: Cyber threat remediation guideline**

<b>Meantime to restore service (MTTR)</b>	<b>Service request</b>	<b>Medium (Severity 3)</b>	<b>High (Severity 2)</b>	<b>Critical (Severity 1)</b>
Priority 1 (P1)	24 hours (24+X7X365)	16 hours (24+X7X365)	8 hours (24+X7X365)	4 hours (24+X7X365)
Priority 2 (P1)	3 business days	16 business hours	12 hours (24x7x365)	8 hours (24x7x365)
Priority 3 (P1)	5 business days	24 business hours	24 business hours	16 business hours
Priority 4 (P1)	5 business days	48 business hours	48 business hours	48 business hours

Company\_Pillar (UK and SA) uses Qradar<sup>5</sup> as their SIEM technology and has deployed, in addition to the SIEM, CrowdStrike and Carbon Black (Cb)<sup>6</sup>. The company stated that their SIEM technology had different capabilities and strengths; however, that is not sufficient to protect against the advancing nature of cyberattacks faced by their clients and employees. According to Company\_Pillar, their SIEM technology presented them with multiple limitations and vulnerabilities, and it was failing to keep up with the rate of security events they were faced with. Their SIEM technology was limited to providing comprehensive visibility to their environment, which in turn prevented them from better thwarting persistent and determined cyberattacks.

<sup>5</sup> <https://www.ibm.com/security/security-intelligence/qradar/>

<sup>6</sup> <https://www.carbonblack.com>

Therefore, Company\_Pillar employed AI-enabled tools for enhanced cyber reliance and to help them better manage their expanding cyber threat vector and evolving cyber risks. According to Company\_Pillar, these AI-enabled tools provided them with enhanced visibility on their environment and have also reduced response times, as well as the time their analysts used to invest in analysis and investigation. Moreover, these tools have the ability to adapt to their evolving cyber landscape. Company\_Pillar emphasised its use of the Cyber Kill Chain for cybersecurity and to reduce the likelihood of adversary from being successful.

Cb is a cloud-based security platform that runs on the company's endpoints to monitor the environment in real time (Shenk, 2017:2). Cb uses machine learning to monitor, detect, analyse, respond and remediate the threat. It relies on a combination of process analysis, threat intelligence feeds, traffic analysis, IoC, antivirus engines and rules to provide a broader picture. This tool provides Company\_Pillar with a robust protection platform, greater visibility and faster response to an incident all while reducing the response time (Zeigler, 2017:1).

CrowdStrike Falcon<sup>7</sup> is a security platform that is custom-built to deter cyberattacks using a unified set of cloud-delivered technologies that prevent all types of attacks including malware, zero-day attacks, advanced malware-free attacks, APT, as well as IoC. CrowdStrike uses machine learning for detection, behavioural analysis and custom whitelisting/blacklisting, as well as response and prevention. With CrowdStrike, analysts are exposed to an in-depth and historical understanding of adversaries, their operations, any attempts to spread to other endpoints and their motivations in the form of intelligence reports that provide real-time analysis.

### **3.8.1.3 Company De\_Link**

Company De\_Link provides services to both corporate and public sectors. According to Company De\_Link, ransomware continued to threaten the majority of their clients. The company discovered that cyberattacks that include ransomware were spread using a number of vectors that included spam emails and malicious attachments. Some ransomware, for instance, was spread through brute forcing login credentials, targeted vulnerabilities running on ICT infrastructure and untrusted third-party app stores. They also noted that attackers had begun to change their campaigns or tactics and were making use of operating system features, off-the-shelf tools, and cloud services. The Company De\_Link SOC runs a 24/7 operation where their analysts constantly update their clients about their security posture and also provides other security services.

De\_Link also uses different security vendors for their capabilities and strengths. The company uses McAfee Enterprise Security Manager (ESM) as its SIEM technology and also uses Darktrace in addition to SIEM. McAfee ESM is used for detection, monitoring, event analysis, correlation and mitigation of cyber threats. The company stated that it deployed Darktrace because their SIEM technology was limited to providing comprehensive visibility to their environment, which in turn

---

<sup>7</sup> <https://www.crowdstrike.com>



prevented them from better thwarting persistent and determined cyberattacks. De\_Link stated that their SIEM technology proved to be an average security measure as their cyber threats landscape is always changing due to the advancement of technology, increased number of cyber adversaries, rising number of both local and international clientele and the heavy reliance on technology for operation and communication by both their organisation and clients. Thus, the company deployed an AI-tool in addition to their SIEM to enhance cyber resilience. The company emphasised the use and significance of the Cyber Kill Chain in their cyber defence approach.

Darktrace<sup>8</sup> uses machine learning and AI algorithms to detect, respond to and prevent cyber threats in real time (Darktrace, 2016: 1). According to Company\_De\_Link, this tool understands its use, employees and the network's pattern of life. Darktrace has the ability to identify and detect subtle, stealthy and previously unknown threats. Darktrace automatically defends Company\_De\_Link's network with digital "antibodies" that take measured and targeted action to neutralise in-progress cyber threats (Darktrace, 2016: 5).

#### **3.8.1.4 Company\_Geo**

Company\_Geo has a diverse clientele that ranges from financial services providers, insurance companies, share trading, and loan providers and government. The company stated that cyberattacks were largely directed to anyone who is connected to the Internet. Thus, they deploy sensors and also conduct vulnerability assessments on their clients' environments. Company\_Geo also builds honeypot sensors that they distribute across their client's networks. These honeypot sensors provide indicators of possible perpetrators already acting inside their client's networks. The honeypots enable them to detect, deflect and counteract attempts of unauthorised use of systems and, moreover, enables them to build better defences for their clients. Company\_Geo emphasised its use of the Cyber Kill Chain for cybersecurity and how it has been one of the main models they use to understand the attacker and the steps they take to achieve their objective.

In addition to their cybersecurity technologies, Company\_Geo uses Splunk for cybersecurity; however, the company has already initiated the process of implementing its own AI-enabled technology into Splunk. The company stated that they realised that Splunk was inefficient to provide them with the enhanced cyber resilience they required without additional investments in AI-enabled technology after they fell victim to multiple cyberattacks. Thus, they initiated the process of developing and slowly implementing their AI-enabled technology in their networks, which they are slowly integrating with Splunk.

Company\_Geo uses Splunk<sup>9</sup> for enhanced cyber resilience. Splunk uses a Machine Learning Toolkit that allows the company to detect incidents, reduce resolution times and predict and prevent undesired outcomes. Splunk uses predictive analytics that continually learn from historical data to

---

<sup>8</sup> <https://www.darktrace.com>

<sup>9</sup> <https://www.splunk.com>

detect cyber threats. The Machine Learning Toolkit helps their security team to build solutions or customise their cybersecurity solution to the needs of a specific client. Splunk has the ability to automatically detect anomalies and patterns in data to help investigators identify and resolve incidents (Merritt, 2017: online). Company\_Geo indicated that they have also integrated Splunk into the AI-tool they are developing.

### **3.9 Conclusion**

This chapter described the research methodology and design of the study. It also detailed methods of data collection, as well as measures used to ensure ethical consideration. The chapter also provided an overview of the cyber threat landscape of all four companies, as well as the various AI-enabled tools they utilise for cybersecurity.

Chapter 4 will provide a brief overview of the National Institute of Standards and Technology (NIST) Cyber Security Framework, which is pivotal in the formation of the proposed framework for the study. Subsequently, it will also provide a comprehensive description of the proposed AI framework for cybersecurity, as well as its elements. The chapter will culminate in the mapping of the proposed AI framework to AI-enabled tools used by the four companies interviewed.

## **CHAPTER 4: PROPOSED CAIBER FRAMEWORK**

### **4.1 Introduction**

This chapter proposes a framework that will demonstrate the significance of defending the cyberspace through AI. The elements of the proposed framework are inspired by the core functions of the National Institute of Standards and Technology (NIST) Cyber Security Framework (CSF), which include Identify, Protect, Detect, Respond and Recover. However, this research proposes supplementary elements, which could be leveraged and integrated to cyberdefence with the use of AI. Those supplementary elements include discovering of cyberthreats, investigation, analysis, prediction and continuous monitoring of the environment for any vulnerabilities and cyber risks. Additionally, these prioritised core elements have emanated from the limitations that the four interviewed companies' cyberdefence system experienced. The proposed framework is called the CAIBER framework and the name was motivated by the research study's main objective, which is to demonstrate the significance of combining AI, and cybersecurity to enhance protection and cyber reliance within the cyber user's environment. The proposed framework is also aimed at laying a foundation for future research and investigations on the significance of defending the cyberspace through AI. The chapter will conclude by demonstrating the significance of the proposed CAIBER framework through mapping AI-enabled tools used by the four companies interviewed.

#### **4.1.1 Background**

The South African government is slowly shifting towards the digital space as technology continues to develop. Moreover, the public and private sector as well as academia, depend on cyber systems for the provision of essential services, economic prosperity and communication. This dependency on technology has great benefits but so are the accompanying risks. The cybersecurity landscape is constantly evolving; cyber attacks are increasingly becoming sophisticated and automated. With an increase in cyberattacks that aim to disrupt critical infrastructure or core business operations, there is an amplified need for cyber resilient systems that are able to withstand cyber incidents in this evolving digital environment (Deloitte, 2017: 8). Thus, the framework aims to demonstrate the significance of using AI for cybersecurity for enhanced cyber resilience. Moreover, the objective of the framework is to demonstrate the need to deploy cyberdefence that have a deep learning capability, effectively monitor, and automatically detect unusual patterns (in the network as well as the IoT environments), conduct thorough analysis and subsequently prevent cyber events.

The proposed framework is, however, not a one-size-fits-all solution to managing cybersecurity, as different cyber users (private companies, government, individual and academia) will be faced with distinctive cyber risks; they will have different vulnerabilities and different capabilities. The elements of the proposed framework are not intended to lead to a static process; rather, they can be performed

concurrently and continuously in order to address the cyber risks. The ultimate aim of the framework is reducing cyber risks, enhancing security levels and better managing cyber threats.

## 4.2 NIST Cybersecurity Framework

To better address cyber risks, the NIST developed the CSF aimed at enhancing the security and resilience of the nation's critical infrastructure. The NIST CSF consists of standards, guidelines, and best practices to manage cybersecurity-related risks (NIST, 2018:1). The framework was developed after a thorough engagement between NIST and other government departments, the public, as well as private companies (McCafferty, 2017: online).

The framework is aimed at helping organisations manage and reduce cybersecurity risks. Moreover, it helps protect and promote the resilience of critical infrastructure and other sectors that are important to the economy and national security (Barrett: 2017: online). It helps prioritise investments and establish the right level of security for an organisation based on business requirements. The NIST CSF will enable organisations to harmonise cybersecurity approaches and provide a common language for discussing cybersecurity risks within and across organisations and industries. The framework is technology neutral, meaning it continues to support technical innovation, while also referencing a variety of existing standards, guidelines, and practices that evolve with technology (NIST, 2018:2).

A clear understanding of an organisation's business drivers and security considerations specific to its use of technology is required in order to effectively defend against cyber threats and risks. Owing to the uniqueness of each organisation's risks, priorities and systems, the tools and methods used to achieve the outcomes described by the framework will vary (NIST, 2018:1). The framework complements and does not replace an organisation's risk management process and cybersecurity programme. It is a tool for aligning policy, business and technological approaches to managing that cyber risk (McCafferty, 2017: online). However, organisations with no formal security programme can leverage the framework as a roadmap to identify business security needs and take the necessary steps to address cybersecurity risks to their data, operations, systems and employees.

The NIST CSF provides functions that could be applied for the protection of critical infrastructure assets against cyber risks. The framework functions are not a checklist of action, rather they present key cybersecurity outcomes. These functions can be performed continuously to form an operational culture that addresses the dynamic cybersecurity risks (NIST, 2018:7). The functions as depicted in Figure 6 below include *Identify*, *Protect*, *Detect*, *Respond* and *Recover*.



**Fig 6: NIST CSF (McCafferty, 2017: online)**

#### **4.2.1 Function 1: Identify**

The *Identify* function calls for organisations to recognise the potential risks that could impact the information systems they use to support their daily operations and critical corporate activities. This function is aimed at developing and enhancing organisational understanding to manage cybersecurity risks (McCafferty, 2017: online). It is critical that the organisation understands its business context, resources that support critical functions, as well as related cybersecurity risks, as that will enable the organisations to focus and prioritise its efforts for cyber defence (NIST: 2018: 7).

#### **4.2.2 Function 2: Protect**

Subsequent to identifying the cyber risk, the organisation has to develop and implement appropriate cyber defences in order to ensure delivery of critical infrastructure services (McCafferty, 2017: online). These protection measures are aimed at limiting or reducing the impact of a possible cybersecurity event by leveraging best practices for data protection and overall security. The *Protect* function supports the ability to limit or contain the impact of a potential cybersecurity event (NIST: 2018: 7).

#### **4.2.3 Function 3: Detect**

The *Detect* function includes the development and implementation of appropriate activities that will enable organisations to identify the occurrences of cybersecurity events. This function is aimed at enhancing the timely discovery of cybersecurity events. Examples of outcomes within this function include the detection of anomalies and events, continuous monitoring of security and development of detection processes (NIST: 2018: 7).

#### **4.2.4 Function 4: Respond**

The *Respond* function supports the ability to contain the impact of a potential cyber incident. This includes developing and implementing appropriate measures and activities to undertake regarding a detected cybersecurity event (McCafferty, 2017: online). According to NIST (2018: 8), this includes establishing response plans, communications and mitigation plans.

#### 4.2.5 Function 5: Recover

Similar to the response function, the *Recover* function is a post-event or incident reactive function. The *Recover* function includes the development and implementation of activities that will maintain plans for resilience and to restore any capabilities or services that were impaired due to a cybersecurity incident (McCafferty, 2017: online). The recovery plan could include processes and procedures that will aid in restoring confidence in the recovered systems and data. The Recover function supports timely recovery to normal operations to reduce the impact of a cyberattack (NIST: 2018: 8).

The NIST CSF provides a set of security measures that businesses can use to assess the degree to which their organisation has implemented these core activities, which can be used as a gauge to assess how prepared the organisation's systems are against an attack. The proposed AI framework will expand on the NIST framework by adding elements which include the discovery of cyber threats, as well as investigation, analysis, prediction and continuous monitoring. AI will provide enhanced cyber defence to evolving and sophisticated cyber threats. AI will provide prevented and predictive methods that are not addressed by NIST. AI will identify the patterns in both potential and real threats, as well as discover and detect both known and unknown threats. AI also addresses the insider threats, which is not sufficiently addressed in the NIST framework.

The following section will provide an overview of the proposed framework for AI in the cyberspace and describe its core elements.

#### 4.3 Proposed CAIBER Framework

This research study proposes a shift in defence surface within the South African context to include AI for cybersecurity. In this regard, the research proposes the CAIBER Framework. The name CAIBER was motivated by the research study's main objective, which is to demonstrate the significance of combining AI, and cybersecurity to enhance protection and cyber reliance within the cyber user's environment. The name is pronounced C-Y-B-E-R.

The proposed CAIBER framework has prioritised different elements that will promote the protection and resilience of information systems and other critical infrastructures that will affect national security. The proposed framework is aimed at improving organisations' security posture and resilience to cyber threats. It is aimed at providing knowledge and a better understanding that will allow users of cyberspace to act in more informed and effective ways. This proposed framework serves as means for identifying and defining research problems and for prescribing and evaluating solutions to research problems.

The nine core elements of the framework are demonstrated in Figure 7 below and are inspired by the core functions of the NIST CSF. The core functions outlined by the NIST CSF for the protection of critical infrastructure assets against cyber risks include *Identify*, *Protect*, *Detect*, *Respond* and

Recover as described in Section 4.3. However, this research study proposes supplementary elements which could be leveraged and integrated to cyber defence with the use of AI. Those supplementary elements include discovering cyber threats, investigation, analysis, prediction and continuous monitoring of the environment for any vulnerabilities and cyber risks. The prioritised core elements emanated from the limitations that the four participant companies have in their cyber defence systems.



**Fig 7: Core elements of the CAIBER Framework (own compilation)**

#### **4.3.1 Element 1: Monitoring**

Monitoring is an overarching and continuous process, as depicted by the arrows in Figure 7. The purpose of continuous monitoring is to alert security teams of suspicious or malicious activities and behaviour within the network. Organisations require a system that will continuously monitor their systems with the aim of detecting adversarial actions or intentions. This element is also focused on discovering new threat vectors, as well as the known and unknown threats. Monitoring will enable the organisation to understand the motive of attacks, their capability, as well as opportunities available for the attacker. Such continuous monitoring will eliminate security blind spots and ultimately provide organisations with a holistic security approach.

#### **4.3.2 Element 2: Identify**

Identification of threats and their dynamics in cyberspace is key to understanding what is at risk and potential attacks an organisation might be exposed to. Machine learning can distinguish between normal network traffic and abnormal traffic; for instance, if an employee swipes their security card in the office but then logs on to a company's computer remotely from a different country, the machine learning tool should flag it instantly, alerting the security analysts of a potential breach. This element will enable security teams to identify relationships between events and then discover malicious activities, cyber risks, and trends.

#### **4.3.3 Element 3: Discover**

AI will proactively hunt for attackers, which in turn will aid businesses to discover adversaries that are targeting their users (Gladen, 2017: online). This element includes the discovery of both known and unknown attacks; malicious activities of internal and external actors, and anomalous behaviour patterns. This will also aid in identifying gaps in the defence program and improve response time. This element includes the discovery of subtle but malicious activities of unassuming insider actors.

#### **4.3.4 Element 4: Detect**

Threat detection is among the key focuses of cyber defence; thus the use of AI to defend against threats in the cyberspace will aid in mitigating security incidents. Detection of cyber threats includes the ability to detect the presence of a malicious act prior to any exploitation or damage. AI can use the knowledge it gains to detect threats, including those that are yet to be discovered, by identifying shared characteristics within families of threats.

#### **4.3.5 Element 5: Investigate**

The significance of the investigation is that analysts are exposed to rich and comprehensive details of cyber events. The investigation of events will help determine the attack vector, status of attack, uncover exploits/vulnerabilities that undermined the system and aid in gathering the maximum amount of information about the attacker.

#### **4.3.6 Element 6: Analyse**

Analysis of cyber threats will enable analysts to evaluate and examine new threats, vulnerabilities and attack vectors, which in turn will aid in preventing future attacks. The analysis element will address a number of questions, for instance, how and why the attacker reached a certain point in the network, what elements or devices participated in the attacks, as well as elements which failed to prevent the attacks and how that happened. AI will enable analysts to analyse each cyberattack phase so that they can determine what action should be taken in the future.



#### **4.3.7 Element 7: Respond**

In response to affected machines, AI will be able to isolate infected machines and prevent the attack from moving forward along the kill chain. Moreover, it will enable security teams to respond in a timely manner to detected threats by preventing any form of progression (Gladen, 2017: online). AI will combine analytical intelligence and machine learning techniques to not only detect new threats but also reduce the time lapse between detection, response and successful prevention (Arshia, et al, 2017: 55).

#### **4.3.8 Element 8: Prevent**

AI will proactively prevent advanced attacks before they get into their environment (Arshia, et al, 2017:53). AI has a self-learning capability that allows it to adapt and evolve in an intelligent manner, defending against stealthy, and never-before-seen threats. Not only will the use of AI to predict cyber threats give organisations a competitive edge but it will also improve the end-user experience while increasing the level of security and trust (Jyothsna and Nilina, 2013: 1721).

#### **4.3.9 Element 9: Predict**

With the use of IA, organisations can deploy deep learning and machine learning for the ever-rising volume of data as these AI tools have greater potential to better predict cyber threats. AI can excavate relevant data that will enable analysts to make better decisions and also identify patterns that may indicate an imminent attack. For instance, AI has the capability to crawl the web and download large volumes of text for natural language processing. That application can identify potential threatening intent on social media for instance (even if indications are subtle), by analysing the words, tone, and content of posts as they are circulated. The predictive element of AI offers cyber users a distinctive weapon in the fight against cybercrime.

### **4.4 Mapping of AI Tools to CAIBER Framework**

The purpose of mapping the AI-enabled tools used by the four participant companies is to demonstrate the significance of applying the proposed CAIBER framework in real-world cyber defence. Moreover, the mapping is illustrative of the strides that the South African private sector has made in defending their cyberspace through AI-enabled tools. This mapping is presented in Table 6 below.

**Table 6: Mapping of AI-enabled tools to CAIBER Framework elements**

CAIBER Framework Elements	AI-enabled Tools						
	Exabeam	Cybereason	Securonix SNYPR	Carbon Black	CrowdStrike	Darktrace	Splunk
<i>Identify</i>	X	X	X	X	X	X	X
<i>Discover</i>	X	X		X		X	X
<i>Detect</i>	X	X	X	X	X	X	X
<i>Investigate</i>				X	X	X	X
<i>Analysis</i>	X	X	X	X	X	X	X
<i>Prevent</i>	X	X	X	X	X	X	X
<i>Respond</i>	X	X	X	X		X	
<i>Predict</i>	X					X	X
<i>Monitor</i>	X	X	X	X	X	X	X

All the AI-enabled tools have placed emphasis and significance on the *identification* and *discovery* of cyber threats, mainly because this allows users to thwart cyber threats in the early stages of the attack. Other elements that have been prioritised by all the tools include *analysis* and *prevention* of cyber threats. Continuous *monitoring* of the internal and external network for cyber threats is another element that is critical to cybersecurity, hence it is addressed by all tools. A major setback of some of the tools is that they do not have an effective and accurate predictive element, which in turn makes it a challenge for users to proactively defend against future threats.

The implementation of defence layers that include identifying, discovering, detecting, analysing and eventually preventing an attack does not only increase the level of security and trust but it also improves the end-user experience and allows organisations to have a proactive approach to cybersecurity. AI-enabled tools that include Exabeam, Cybereason, Securonix, Carbon Black and Darktrace demonstrate the need to have cybersecurity tools that will not only focus on discovering and identifying external threats will also identify and respond to insider threats. The significance of employing AI-enabled tools is that they have the ability to adapt and learn over time, meaning they can detect both known and unknown attacks. Tools that are AI-enabled provide advances in sophisticated analytics capabilities, powerful cognitive computing and deep learning that identify and detect malware, attack patterns and prevent attacks in near real time.

A tool that allows users, for instance, to only detect, analyse, respond to and prevent attacks provides, to a certain degree, a relatively sufficient mitigation measure at a tolerant risk level. However, this will not provide users with a comprehensive defence as cyberattacks evolve, advance in sophistication and are becoming more autonomous. The significance of employing tools that encompass all the elements of the framework is that they provide enhanced visibility into the network. Additionally, deploying an AI-enabled tool with all the elements of the framework will allow organisations to improve their cyber resiliency, reinforce their security posture and shift their cybersecurity capabilities. This provides organisations with a framework for aggregation, early detection, in-depth investigation and analysis, accurate, as well as timely response and prediction of threats.

#### **4.5 Conclusion**

This chapter provided an overview of the NIST CSF. It also described the core functions of the framework, which include *Identify*, *Protect*, *Detect*, *Respond* and *Recover*. Additionally, a proposed CAIBER Framework that is focused on demonstrating the significance of using AI for cyber resilience was developed. The core elements of the CAIBER Framework as noted in this chapter were inspired by the NIST CSF and emanated from the limitations discovered in the cyber defence system of the four participant companies. This chapter concluded by mapping the AI-enabled tools identified in chapter 3 (see 3.8.1 Overview of companies) to the CAIBER Framework.

The next chapter will focus on the application of the proposed framework on Cyber Kill Chain. It will also demonstrate the significance of the proposed framework by applying it to case studies of the four companies interviewed.

## **CHAPTER 5: APPLICATION OF CAIBER FRAMEWORK**

### **5.1 Introduction**

The chapter is focused on the application of the proposed CAIBER framework that was discussed in Chapter 4. This chapter will draw upon previous research pertaining to cyber threat modelling with the Cyber Kill Chain. The chapter will begin by providing a comprehensive description of the Cyber Kill Chain and it will also look at different phases of the chain. In relation to the application of the proposed CAIBER Framework, the chapter will map out the proposed framework for the Cyber Kill Chain to demonstrate its application to better thwart cyberattacks. Subsequently, the chapter will look at case studies of different cyberattacks experienced by the four participant companies. The aim of the case studies is to demonstrate how the application of the proposed CAIBER Framework would help remediate cyber threats and enhance resilience cyberdefence. The following section will focus on the Cyber Kill Chain and its different phases.

### **5.2 Cyber Kill Chain**

The Cyber Kill Chain is a model that describes the sequence of events that an attacker must perform in order to achieve success during an attack (Hutchins, Cloppert, and Amin, 2011: 4). It was developed by Lockheed Martin as a representation of a sequence of actions that an attacker will go through to achieve their ultimate objectives (Dalziel, 2015: 7). Hutchins et al (2011:4) described the Cyber Kill Chain as an intrusion-based methodology that allows organisations to focus on the various phases of an attack. The model identifies what the adversaries must complete in order to achieve their objective.

The Cyber Kill Chain is a tool aimed at helping cyber defenders to better identify the stages at which they can detect the adversary activity to mitigate or to prevent against it and to place other defensive controls or mitigation actions in place. It also aids cyber defenders to better describe uses of certain tools, as well as illustrate the investment of the attacker at each phase. The end goal of this model is to reduce the likelihood of adversary success, prioritisation of resources and increasing performance and effectiveness of cyber defence by enabling security analysts to understand the threat, its intent, capability, doctrine and patterns of operation (Hutchins et al, 2011: 12). The Cyber Kill Chain is focused primarily on intrusions, malware and external types of incidents and attacks; however, it has a limitation in terms of identifying and detecting insiders (Korolov and Myers, 2017: online). The next section introduces the seven phases of the Cyber Kill Chain.

#### **5.2.1 Phases of Cyber Kill Chain**

The Cyber Kill Chain consists of seven steps that a malicious actor has to accomplish in order to successfully launch their attack or achieve their desired objective. The use of the model can aid defenders to develop resilient cyber approaches. Moreover, each of these steps is mapped to the

progressions of detection, response and prevention (Hutchins, et al 2011: 2). The seven steps of the Cyber Kill Chain are Reconnaissance, Weaponisation, Delivery, Exploitation, Installation, Command and Control (C2), and Actions on Objectives as represented in Figure 8.



**Fig 8: Cyber Kill Chain (Sager, 2014:3)**

Below is a description of Cyber Kill Chain phases:

- a) **Reconnaissance** – This is the initial phase where an attacker gathers information about their target (Hutchins et al, 2011: 4). There are a variety of methods that the adversary can use to achieve this. These include examining the target's social media accounts, harvesting email addresses, collecting press releases and contract awards, as well as using conference attendee lists and other public information. Moreover, this step also includes technical tactics such as scanning ports for vulnerabilities, services and applications to exploit (Sager, 2014: 2).
- b) **Weaponisation** – This phase is largely dependent on the accuracy and amount of *reconnaissance* performed by the attacker (Velazquez, 2015:5). The attacker analyses the data gathered in the reconnaissance phase to determine which mode of attack to apply. For instance, they might target the organisation's operating systems and firewalls or even use client application data files such as Adobe Portable Document Format (PDF) or Microsoft documents as the weaponised deliverable (Hutchins et al, 2011: 4).
- c) **Delivery** – In this phase, the attacker delivers the malicious attack identified in the *Weaponisation* phase. This can be done through different vectors of attack, which include social

media, malicious email or USB or even luring them to a website that might look authentic (Sager, 2014: 2).

- d) **Exploitation** – Subsequent to the *delivery* of the weapon, a targeted user or an employee in that targeted organisation might interact with the weaponised deliverable through opening an attachment or clicking on a malicious link or even leveraging an operating system feature that auto-executes code (Hutchins et al, 2011: 4). In this particular phase, the target becomes a victim.
- e) **Installation** – The installation phase is when the attacker uses their chosen attack vector (weaponised deliverable) to *exploit* the target by installing the malware and establishing privileged operations on the victim's environment. For instance, the installation of a remote access Trojan horse or backdoor in the victim's system will allow the attacker to have persistent existence in the victim's environment.
- f) **Command and Control (C2)** – During this phase, the attacker has established a control channel through the malicious *installation* that enables them to remotely manipulate their victim's system. This could include backdoors, unauthorised accounts, opened ports or services on particular systems or DNS and email protocols (Sager, 2014: 2).
- g) **Actions on objectives** – During this phase, the intruder executes actions through the established C2 to achieve their main objectives or goal (Hutchins et al, 2011: 5). Those objectives, for instance, could include exfiltration of data, distortion of data, theft of sensitive information or intellectual property or further intrusion into the network to infect further systems (Velazquez, 2015:5).

The application of correct tools and security measures within the correct Cyber Kill Chain phase will help defenders to disrupt or deny the adversary's ability to perform their malicious activities. Each phase in the Cyber Kill Chain is an opportunity to stop the attack in its tracks (Korolov and Myers, 2017: online). It is imperative to have defences for every phase of the kill chain and each phase is equally important. Howarth (2016: 1) stated that if an attack were stopped near the beginning, cleaning up any attack would be less costly and time-consuming. However, that requires that an organisation to have a comprehensive picture and visibility on their network. The following section will address the evolving nature of cyberattacks and their implication on Cyber Kill Chain.

### 5.2.2. Evolution of cyberattacks and Cyber Kill Chain

The original interpretation of the Cyber Kill Chain as developed by Lockheed Martin assumes a traditional perimeter defence where a firewall is the main impediment to intruders (Greene, 2016: online). However, the insider threat does not fit within this perimeter and is, therefore, a representation of the evolution of the Cyber Kill Chain. To remedy this, it was suggested that the Cyber Kill Chain include the word "internal" in each step of the attack. Furthermore, the Cyber Kill Chain assumes that an attacker has to go through all the phases of attack before an attacker is successful. However, the evolution of cyberattacks is demonstrating that attackers are not following

the cyberattack life cycle; rather they skip steps, add others and backtrack some (Tarnowski, 2017: 3).

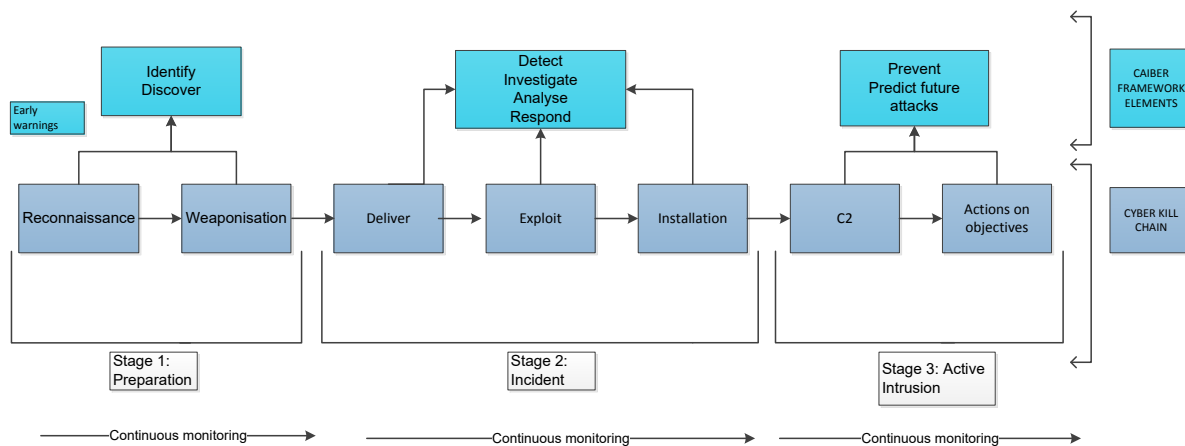
### **5.2.3. Application of AI to Cyber Kill Chain**

Three of the companies that were interviewed as mentioned in chapter 3 emphasised their use of the Cyber Kill Chain to track cyberattacks and to better secure their cyberspace (see 3.8.1 Overview of companies). Thus, the proposed CAIBER framework was mapped onto the Cyber Kill Chain. The main objective of applying AI to the Cyber Kill Chain is to enhance visibility on all seven steps, break the chain of attack and subsequently prevent the attacks. For instance, ANN and machine learning have the ability to learn, adapt, monitor and solve complex problems; thus, they can be applied successfully to all phases of the Cyber Kill Chain (Rajbanshi, Bhimrajka, and Raina, 2017: 132).

Cyber users need to constantly adjust and improve their security systems in the face of the changing cyber environment and evolution of cyberattacks in order to be resilient and provide continuous protection. Thus, employing AI could improve overall security performance and provide better protection from an increasing number of sophisticated cyber threats (Wirkuttis and Klein, 2017: 109). It is imperative to have multiple layers of defence to ensure that if one of the defences is bypassed, another line to protect one's information is present, as well as other critical infrastructure. AI has the ability to solve problems and execute tasks mimicking the human cognitive process. These abilities include understanding the scope of the problem, knowing where to find sources of information to solve the problem and ingesting data from the outside.

Organisations in this digital age are not only porous but they are also categorised by high usage of mobile devices and cloud-based services, as well as the proliferation of smart devices, IoT devices, and expansion of networks (Howarth, 2016: 1). Moreover, in order to be able to interrupt or significantly impede the cyberattacks, it is necessary to have defences such as AI that will continuously monitor, identify threats, learn users' behaviours, determine anomalous activities and prevent cyber threats in the early stages of the Cyber Kill Chain (Tarnowski, 2017: 6).

As noted above monitoring is an overarching and continuous process. In essence, AI will provide proactive defences by continuously monitoring all information systems and devices that connect to the network. The aim of monitoring is to provide analysts with an insightful picture of their environment and to provide a comprehensive picture of the progression of attacks on every phase of the Cyber Kill Chain. This will in turn enhance the organisation's security posture. Figure 9 below shows how the elements of the proposed framework would be applied to the Cyber Kill Chain for enhanced cybersecurity.



**Fig 9: Mapping of CAIBER Framework to Cyber Kill Chain (Own compilation)**

The objective of this section is to demonstrate how the mapping of the CAIBER Framework elements onto the Cyber Kill Chain can aid organisations in implementing the proposed framework for enhanced cyber resilience.

#### **5.2.3.1 Stage 1: Preparation (Reconnaissance and Weaponisation)**

In the application of the CAIBER framework to the Cyber Kill Chain, the first stage, *Preparation*, comprises the *Identify* and *Discover* elements of the CAIBER framework mapped onto the *Reconnaissance* and *Weaponisation* phases of the Cyber Kill Chain. This stage relates to the reconnaissance of security vulnerabilities and the development of tools to exploit those vulnerabilities before attempting to launch a cyberattack. By identifying and discovering the adversary's action in the *Reconnaissance* and *Weaponisation* phases, the AI can respond accordingly by deploying resilience measures that will prevent the malicious actions from progressing and causing damage.

This stage is critical in the attack and thus using AI tools that include ANN, Intelligent Agent, deep learning and AIS could be used to better identify and discover malicious activities. In this stage, early and accurate warnings that would translate newly gathered threat data into actionable tasks will also be generated. Moreover, AI can be used to alert security teams about any anomalous behaviour, malicious activities and deviation on the pattern of life. AI will channel organisations to have a user-centric approach to their security approach, which will, in turn, help them understand the users' normal behaviour. Lastly, AI could be applied to identify and discover cyber threats before they affect the entire network.

#### **5.2.3.2 Stage 2: Incident (Delivery, Exploitation and Installation)**

In relation to the application of the CAIBER framework to the Cyber Kill Chain, the second stage, *Incident*, is inclusive of *Detect*, *Investigate*, *Analysis* and *Respond* elements of the CAIBER framework mapped onto the *Delivery*, *Exploitation* and *Installation* phases of the Cyber Kill Chain. The incident stage is also critical in the cyber incident lifecycle because an attacker can deliver the attack, exploit the user's vulnerability and subsequently launch the attack. These vulnerabilities may exist in



the form of technical or non-technical components of a victim's network; for instance, public information pertaining to the identities of executive leadership or job postings for individuals trained on specific information systems, details of employees, etc. As discussed in Section 5.2.2, the Cyber Kill Chain has expanded and now encompasses threats such as insider threats that are more difficult to detect (Howarth, 2017: 2). Thus, the implementation of defence layers in the early steps of the attack is critical to detecting such subtle risks.

Through the application of AI, analysts will spend less time sifting through large volumes of data and using manual methods to find traces and evidence of evolving and stealthy attacks. Security teams will be exposed to comprehensive details of cyber events, which in turn can help prevent future attacks. An investigation and analysis of an event will help security teams determine the effectiveness of the response. For instance, the analysis and investigation will include identification of indicators of compromise, determining where the indicators were observed in the Cyber Kill Chain, which will be compared with other threat intelligence and also determine if the event has spread to other parts of the network. An in-depth understanding of how a malicious actor operates in the Cyber Kill Chain empowers the defender to determine how to respond and to also devise more effective and resilience defensive strategies.

### **5.2.3.3 Stage 3: Active Intrusion (C2 and Actions on Objectives)**

In the application of the CAIBER Framework to the Cyber Kill Chain, the last stage, *Active Intrusion*, comprises the *Prevent and Predict* elements of the CAIBER framework mapped onto the C2 and Actions on Objectives phases of the Cyber Kill Chain. This phase outlines how AI will effectively respond to attacks. Moreover, it demonstrates how AI will predict future attacks, based on its experience, knowledge, learned behaviour and insight. In order to prevent an attack from infecting other systems or to further cause damage, the attack needs to be detected and stopped at the beginning. However, that requires an organisation to have a clear view of what is happening on its network as a whole.

The use of AI will enable the organisation to have complete visibility of its entire environment. Moreover, the understanding and gained insight into the adversary's targets, actions, behaviour and strategy within the Cyber Kill Chain (which will have been derived/achieved in stage 1 and 2) will enable AI to better prevent and predict the next attack. This will not only increase the preventative and predictive measures but it will also enhance cyber resilience within the organisation. The application of AI will enable the organisation to prioritise events, anticipate, and better prevent incidents before they occur.

The mapping of the CAIBER Framework onto the Cyber Kill Chain will be demonstrated in the next section as an effective way of using AI to counter cyberattacks and improve cyber defences. The following section presents case studies that support the mapping noted in all three stages (refer to

5.2.3. Application of AI to Cyber Kill Chain). Additionally, the section demonstrates the application of the proposed CAIBER Framework for cyber resilience.

### 5.3 Case Studies: Application of CAIBER Framework

Organisations vary considerably in terms of the level of maturity in their cybersecurity incident response capability, but also in the way in which they need to respond. Thus, the following case studies are aimed at sketching scenarios of cyberattacks and demonstrating how the application of the proposed framework could aid in remediating those attacks. These case studies are loosely based on the information obtained during data collection by means of interviews and observations (refer to 3.4.1 Methods for data collection), with additional detail to illustrate the full application of the CAIBER Framework.

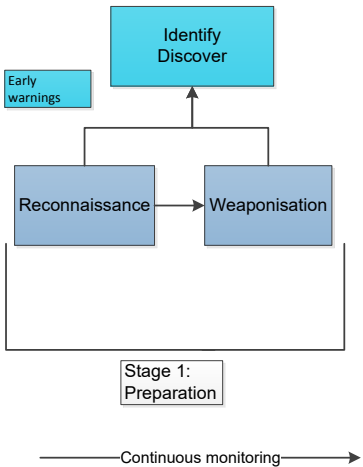
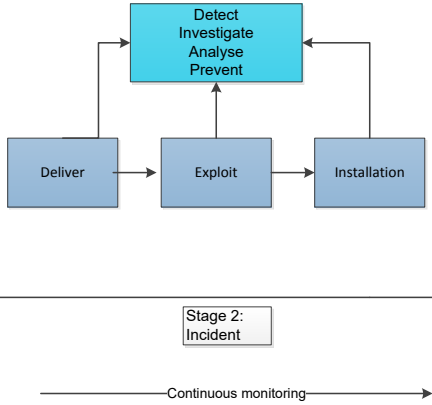
#### 5.3.1 Case study 1: Attempt by an insider to harvest data

During the interview with Company\_Geo, it was mentioned that indicators of malicious activities conducted by an insider are often subtle. For instance, an employee of the company installed their own personal Banana Pi M3s onto the company's information system without being detected by the IT department. This device was set to act as a gateway, redirecting network traffic to a destination pre-determined by the malicious employee. These activities represented a significant deviation from Company\_Geo's pattern of life. The device was communicating with a suspicious website (hosted on an alternative server) that was set up to look like it belonged to Company\_Geo. The aim of the employee's actions was to harvest other staff members' personal information within the organisation and then escalate his privileges to that of an administrator where he could have unlimited access to critical information. This was done by redirecting other staff members to a fake login page where they were required to enter their usernames, passwords and company numbers.

The company later discovered that more than 60% of their employees' data and other sensitive information had been stolen and they since have not been able to recover it. This also resulted in a disruption of services for 3 days, which resulted in Company\_Geo's losing a large portion of its client and revenue. Company\_Geo suspected that the perpetrator was most likely using the devices to profile the cyber defence strategy employed by the company so that he could launch a more targeted attack in the future. Table 7 below presents the application of the CAIBER framework on the Cyber Kill Chain mapping on this malicious insider case study.

**Table 7: Case study 1: Mapping of CAIBER Framework to Cyber Kill Chain**

<b><i>Cyber Kill Chain</i></b>	<b><i>Application of CAIBER Framework</i></b>
<i>Stage 1: Preparation</i>	The application of the proposed CAIBER framework would have enabled Company_Geo to <i>identify</i> insider threats through monitoring employees' activities on the system. AI

Cyber Kill Chain	Application of CAIBER Framework
 <p data-bbox="359 607 464 651">Stage 1: Preparation</p>	<p data-bbox="660 241 1396 640">would have been able to discover Banana Pi M3 after it was lodged into the network and then alert the security team about a foreign device in the network. AI has the ability to understand what represents normal behaviour for every employee and device connected to the network; hence, it would have identified any malicious activity being conducted by the Banana Pi M3. Moreover, it would have been able to identify any form of a deviation of an employees' behaviour and any device in the network.</p> <p data-bbox="660 701 1396 913">The <i>monitoring</i> element of the CAIBER framework would allow Company_Geo to have visibility across the company network, which in turn would provide the security team with an overall oversight of the company's system and also protect them from emerging cyber threats.</p>
<p data-bbox="209 931 411 965"><i>Stage 2: Incident</i></p>  <p data-bbox="384 1384 472 1429">Stage 2: Incident</p>	<p data-bbox="660 931 1396 1234">The application of the proposed CAIBER framework would have enabled Company_Geo to <i>detect</i> suspicious behaviour. This would include the uploading and downloading of larger than usual amounts of data, sending of packets to unusual locations or in an unusual pattern, such as that of the Banana Pi M3 communicating with a website hosted on an alternative server.</p> <p data-bbox="660 1249 1396 1648">Instead of relying on knowledge of past threats for vulnerabilities, AI is able to independently classify data and detect compelling patterns that define what may be considered to be normal behaviour (Darktrace, 2017b: 7). The <i>investigation</i> element would have exposed security analysts to comprehensive details of the security event. This may include the status of the attack, specific time periods, event severity, triggering files, privileged accounts and other information.</p> <p data-bbox="660 1664 1396 1921">The application of AI could assist Company_Geo in <i>analysing</i> new threats, vulnerabilities and attack vectors, which in turn can help, <i>prevent</i> future attacks. In <i>response</i> to the attack, AI would have prevented the Banana Pi M3 from acting as a gateway to another website, which in turn would have prevented the internal traffic from diverting to a malicious site.</p>

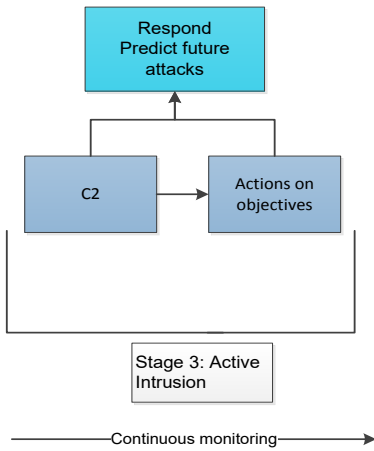
<b>Cyber Kill Chain</b>	<b>Application of CAIBER Framework</b>
<p data-bbox="204 237 501 271"><i>Stage 3: Active Intrusion</i></p>  <pre> graph TD     C2[C2] --&gt; AO[Actions on objectives]     AO --&gt; RP[Respond Predict future attacks]     RP --&gt; C2     subgraph Stage3 [Stage 3: Active Intrusion]         C2         AO     end     CM[Continuous monitoring] --&gt; C2     </pre>	<p data-bbox="659 237 1388 405">AI would have identified vulnerabilities in the system and subsequently, respond by applying resilient security measures that would <i>prevent</i> any exploitation. Subsequently, AI would then be able to predict future attacks.</p>

Table 7 presented the application of the CAIBER framework onto the Cyber Kill Chain phases in a malicious insider cyber incident. The application of the CAIBER framework by Company\_Geo would have increased the company's overall cyber resilience by providing early identification of cyber threats and discovering foreign and malicious devices in real time. Additionally, the application of the CAIBER Framework would have enabled Company\_Geo to detect the subtle anomalous activities in their network. Company\_Geo would have also been able detected the deviation from the normal pattern of life and behaviour within its network. The investigation and analysis would have provided the company with an overall picture of the attack within the Cyber Kill Chain. Lastly, the attack would have been prevented as the AI would have responded and prevented the malicious activities from causing any damage to the device in real time.

### 5.3.2 Case study 2: Storage on cloud server threatens intellectual property

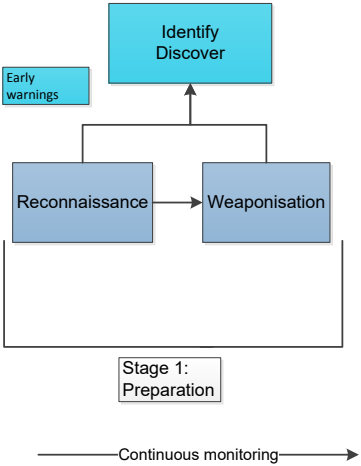
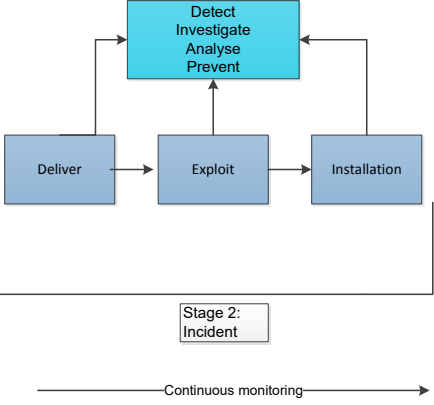
During the interview with Company\_Pillar, the company mentioned that the organisation used a third-party cloud server to store their most sensitive information, including their intellectual property. However, the only security measure that they have employed to that cloud server was a username and a weak password, which was updated on a quarterly basis. The password was made weak because there were multiple people using it, thus they chose an easy-to-remember the password. The individuals who have access to this username and password are Company\_Pillar's top management, middle management and security team. In essence, the data on the cloud server was available without further restrictions or any form of encryption.

By intercepting Company\_Pillar's network communications while the data on the server is in this unsecured state, a malicious actor can easily discover the address with insignificant effort. In another variation of this scenario, a determined Company\_Pillar employee would face almost no barriers if they wanted to download the intellectual property and sell it to Company\_Pillar's competitors. In this

scenario, a malicious actor detected this vulnerability and attempted to download a ZIP file that contained sensitive data from an IP address outside of Company\_Pillar's network.

The file contained information that included patent, details of copy writes, as well as trade secret, clients' information, business plans, proprietary software, and hardware. According to Company\_Pillar, the attack went on for weeks and the company was unable to detect it. The company was unable to recover the information or even prevent the adversaries from causing further damage. The estimation of this attack by Company\_Pillar was well over a billion rand. This attack also affected the relationship and trust between the company and its customers. It also resulted in the company losing its current contracts and other future business ventures. Table 8 below presents the application of the CAIBER Framework on the Cyber Kill Chain through mapping it to cyber threat on a cloud server that contained intellectual property.

**Table 8: Case study 2: Mapping of CAIBER Framework to Cyber Kill Chain**

<b>Cyber Kill Chain</b>	<b>Application of CAIBER Framework</b>
<p data-bbox="204 927 454 958"><i>Stage 1: Preparation</i></p>  <pre> graph TD     EW[Early warnings] --&gt; ID[Identify Discover]     R[Reconnaissance] --&gt; ID     W[Weaponisation] --&gt; ID     R --&gt; W     subgraph Stage1 [Stage 1: Preparation]         R         W     end     CM1[Continuous monitoring] --&gt; CM1     </pre>	<p data-bbox="659 927 1409 1368">Primarily, AI would <i>identify</i> and flag out the insubstantial security measures (which is the weak password) employed in the cloud server. AI would immediately alert the security team of emerging threats that range from subtle insiders to low and slow attacks, including automated viruses. The organisations would have been able to <i>identify</i> and <i>discover</i> undesirable access to their intellectual property data and also received accurate alerts when the information was exposed to unauthorised insiders or when an adversary was trying to access it.</p>
<p data-bbox="204 1442 496 1473"><i>Stage 2: Attacker phase</i></p>  <pre> graph TD     D[Deliver] --&gt; E[Exploit]     E --&gt; I[Installation]     E --&gt; DAIAP[Detect Investigate Analyse Prevent]     I --&gt; DAIAP     subgraph Stage2 [Stage 2: Incident]         D         E         I     end     CM2[Continuous monitoring] --&gt; CM2     </pre>	<p data-bbox="659 1442 1409 2018">The attempt to download a ZIP file could have been <i>detected</i> and flagged as an anomalous behaviour by the system. In order to provide security analysts with an insightful picture of the history and progression of attacks on every stage of the attack, activities within the cloud server would have been recorded and easily accessible. The <i>investigation</i> and <i>analysis</i> phase would have enabled Company_Pillar to analyse raw data network traffic and then formulate an evolving understanding of what is normal for different users, devices and the cloud server as a whole. AI would have aided Company_Pillar to determine the chain of custody and help them determine the type of information that was stolen, how, why and by whom. In <i>response</i> to the attack, AI would have</p>

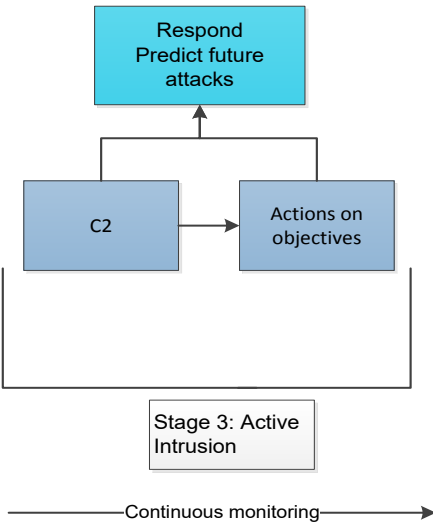
Cyber Kill Chain	Application of CAIBER Framework
	<p>isolated the compromised information systems and also identify patterns of attack to determine the extent of the compromise. The security team would have been able to easily pick up anyone accessing their network, what they are accessing and whether that individual displays normal behaviour or not.</p>
<p>Stage 3: Active Intrusion phase</p>  <pre> graph TD     C2[C2] --&gt; AO[Actions on objectives]     AO --&gt; RP[Respond Predict future attacks]     RP --&gt; C2     subgraph Stage3 [Stage 3: Active Intrusion]         C2         AO     end     CM[Continuous monitoring] --&gt; Right[ ]     </pre>	<p>AI has an inherent ability to continuously learn and study large pools of knowledge in order to anticipate future threats and an appropriate response (White House, 2016:30). Therefore, it would proactively <i>prevent</i> advanced any form of attacks in the cloud server before they get into the user's environment. Thus, it has the ability to not only prevent both known and unknown attacks, but it could also <i>predict</i> anomalies that could have led to data leakages, distortion or theft in the cloud server.</p>

Table 8 above presents the application of the CAIBER Framework on the Cyber Kill Chain mapping for cyber threat on a cloud server that contained intellectual property. The application of the proposed CAIBER framework with all the elements would have helped Company\_Pillar to implement stronger security measures for such critical information. Through the application of the CAIBER Framework, Company\_Pillar would have identified and detected malicious activities in real time and not weeks after the damage was already done. Subsequent to identifying and detecting malicious activities, a thorough analysis and investigation would have been conducted by AI in order to determine the extent of the compromise.

Additionally, the analysis and investigation element would have provided Company\_Pillar with a comprehensive picture of the compromise. In response to the discovering and detection of these activities, AI would have prevented the adversaries from stealing more intellectual property. In essence, the application of a CAIBER Framework with all elements would have helped Company\_Pillar to prevent the loss of information, clients and revenue, as well as reputational

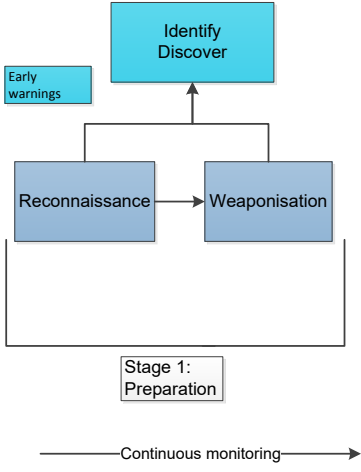
damage. The application of an AI framework with all elements enables organisations to extend visibility into otherwise unseen parts of their network, including the activities in the cloud. This, in turn, helps eliminate blind spots and protect data, regardless of where it resides. Moreover, the CAIBER Framework would have enhanced Company\_Pillar's cyber resilience.

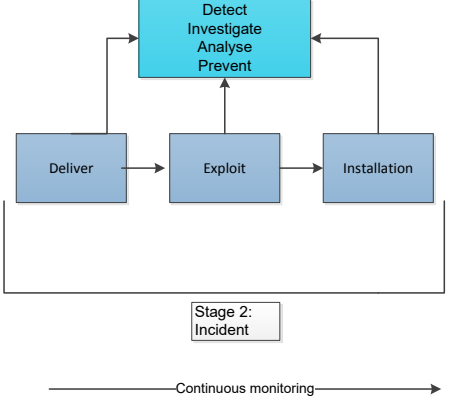
### 5.3.3 Case study 3: Email document containing ransomware

During the interview with Company\_De\_Link, the company mentioned that one of their employees had a personal emergency and had to access her personal email account using the company's desktop. This act meant that she was circumventing the company's security policy. Owing to the nature of the emergency, the employee opened an email attachment that she believed was a Microsoft Word document. However, the document was actually a malicious ZIP file that contained a ransomware payload. This payload caused the laptop to contact a malicious domain to download a suspicious EXE file.

The ransomware then began to search for available Server Message Block (SMB) shares. By the time the IT department noticed the malicious activity on the company network, the ransomware executable had already bypassed multiple security perimeter protocols on the employee's infected desktop. The ransomware had encrypted all files on the employees' desktop and across the entire network. Company\_De\_Link stated that their attacker demanded a payment of 50 million dollars in less than 48 hours or they would never get to access their information systems. This attack did not only cause major operational disruption but it also resulted in revenue losses, loss of client trust and reputational damage. Table 9 below presents the application of the CAIBER framework on the Cyber Kill Chain mapping on an incident similar to that of the employee that downloaded a Microsoft Word document containing ransomware.

**Table 9: Case study 3: Mapping of CAIBER Framework to Cyber Kill Chain**

<b>Cyber Kill Chain</b>	<b>Application of CAIBER Framework</b>
<p data-bbox="204 1487 454 1518"><i>Stage 1: Preparation</i></p>  <pre> graph TD     Recon[Reconnaissance] --&gt; Weapon[Weaponisation]     Weapon --&gt; ID[Identify Discover]     EW[Early warnings] --&gt; ID     CM[Continuous monitoring] --&gt; Right[ ]     subgraph Stage1 [Stage 1: Preparation]         Recon         Weapon     end             </pre>	<p data-bbox="662 1487 1412 1848">The application of the framework would have enabled the organisation to <i>identify</i> and <i>discover</i> in real time the highly anomalous activity. In addition, AI would have identified the circumvention of security policies within the organisation by the employee and then alert the security team. The security team would then have taken the necessary action; however, failure of security to act upon this anomaly, AI would automate the necessary response.</p>
<p data-bbox="204 2011 494 2042"><i>Stage 2: Attacker phase</i></p>	<p data-bbox="662 2011 1412 2042">AI has the ability to adapt and learn users' behaviour over time,</p>

<b>Cyber Kill Chain</b>	<b>Application of CAIBER Framework</b>
	<p>meaning it would have been able to detect primarily the deviation and transgression of the security policy by the employee. Moreover, AI would have proactively <i>detected</i> a slight deviation from normal behaviour when, for instance, the executable file began to encrypt SMB shares. AI would have been able to conduct a thorough <i>investigation</i> and <i>analysis</i> that would enable the organisation to understand who the sender of the email was, the motive behind the attempted attack and also determine the extent of the attack (for instance, uncover compromised credentials before they result in loss of intellectual property or some other form of cybersecurity risk). AI would have applied the new found knowledge of the unknown threat to other systems in the network in order to <i>investigate</i> whether other machines exhibit evidence of the threat or threat type.</p> <p>Company_De_Link’s security team would also have been able to have access to the timeline of the incident to the point it was detected. For instance, the team would have been able to learn when the attack began, how it began and which vector of attack was exploited. This kind of in-depth analysis would provide analysts with a high-level oversight of threat levels and, moreover, allow them to excavate into granular details of the attack (Darktrace, 2017a: 6). In <i>response</i> to such an attack, AI would have autonomously interrupted all attempts to write encrypted files to network shares.</p>
<p><i>Stage 3: Active Intrusion phase</i></p>	<p>This real-time response would have <i>prevented</i> the attack from spreading to the whole network in a matter of minutes, reducing the overall impact of this compromise. In a case where other information systems were affected, AI would have identified that compromise and prioritised responses based on impact to the affected assets. The application of the proposed framework with all the elements would have enabled the organisation to anticipate and better <i>predict</i> incidents before they occurred.</p>



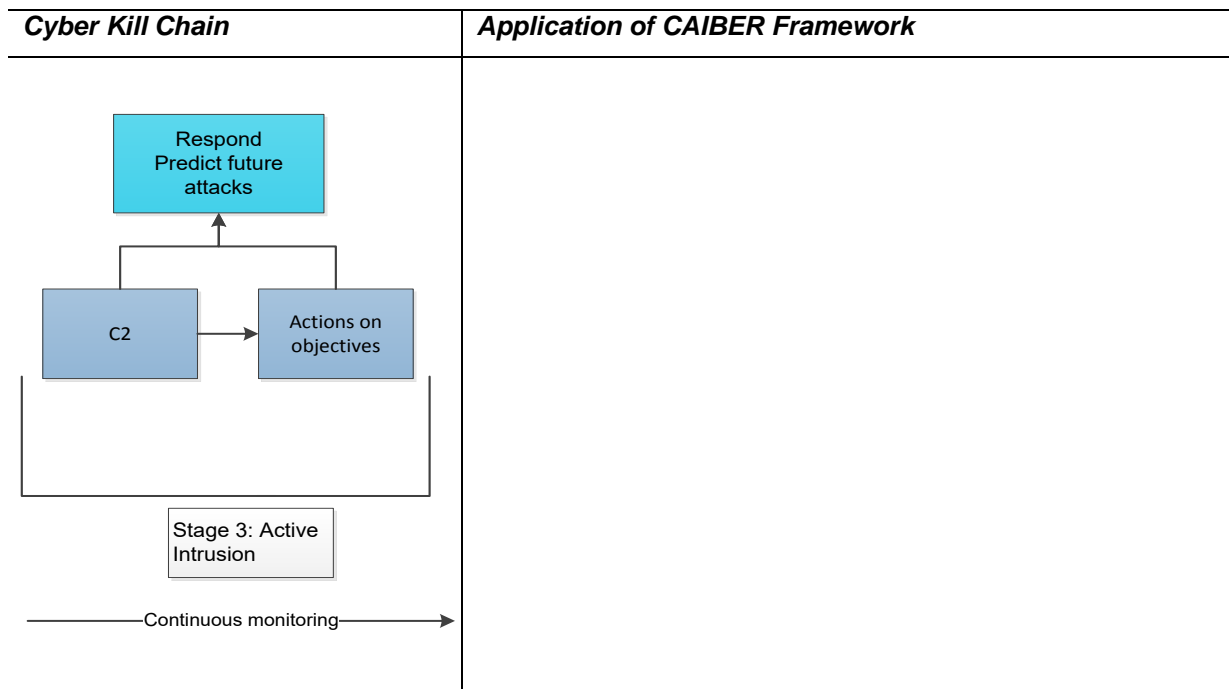


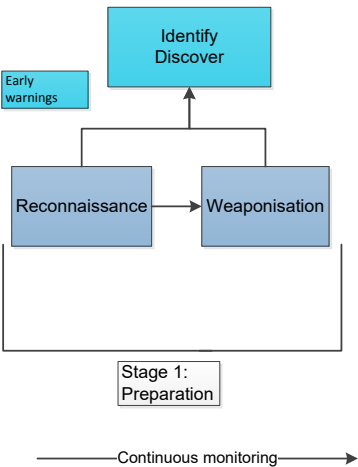
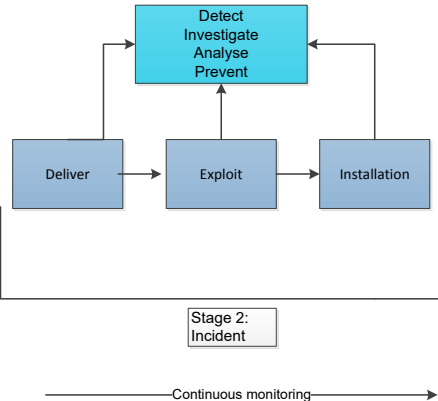
Table 9 above presents the application of the CAIBER Framework on the Cyber Kill Chain mapping on an incident at Company\_De\_Link where an employee circumvented the company's security policy and subsequently downloaded a Word document with ransomware. The application of the CAIBER framework with all the elements would have enabled Company\_De\_Link to identify and detect the threat in real time. Moreover, information discovered during the analysis and investigation would have provided them with full visibility into the attacker's life cycle, i.e. from where the attack began to the point it was detected by the company. The application of the CAIBER Framework with all the elements would have provided Company\_De\_Link with complete visibility of its environment and also enhance its cyber resilience.

#### 5.3.4 Case study 4: Insecurity of IoT

During the interview with Company\_Magix, the company mentioned that it had purchased smart devices (writing and drawing pads) for employees that were part of a working project in the company. However, the purchase of these pads was done without the involvement of Company\_Magix's IT department and the information security teams. These pads improved employee productivity and morale. They enabled the users to work competently and diligently, as they were able to send their plans and drawings to clients and other colleagues. Moreover, the pads were connected to the users' smartwatches, which enabled them to work efficiently. For instance, employees would be able to receive email alerts, alerts about meetings and appointments, etc. on their smartwatches. instantly on their smartwatches despite their locations. In essence, these devices enabled Company\_Magix employees to drastically change the way in which they interacted with their clients and with one another.

Unbeknown to Company\_Magix, the devices were connected to the company's Wi-Fi router without the configurations of default settings (which also included the changing of the default login credentials). These devices created a wave of vulnerabilities, as they were widely accessible and opened a wide range of channels that malicious actors could exploit. A week into the network, a malicious actor scanning the Internet identified the vulnerable devices and exploited them by sending large volumes of data to the devices and other devices connected to the pads through the Wi-Fi. This was later identified as a DoS attack by the company's IT department. This attack prevented the project team and the entire organisation from providing services to their clients for 2 days, which was a major challenge for the organisation as they have a large clientele that includes local and international clients. The project that the team was working on collapsed due to this incident. In addition, the reputational damage to the company caused paramount revenue losses and clientele. Table 10 below presents the application of the CAIBER framework on the Cyber Kill Chain mapping on the insecurity of the IoT device case study.

**Table 10: Case study 4: Mapping of CAIBER Framework to Cyber Kill Chain**

<b>Cyber Kill Chain</b>	<b>Application of CAIBER Framework</b>
<p><i>Phase 1: Preparation</i></p> 	<p>AI is self-organising, flexible, intuitive, and dynamic and has the ability to learn the behaviour of devices and of a network as a whole over a period of time. AI would have been able to <i>discover</i> the foreign device and, over time, learn its behaviour and pattern of life. Primarily, AI would have alerted the security team about foreign devices that are connected to the Wi-Fi (as they would not be part of the regular device). Moreover, the use of the proposed framework would have enabled the organisation to discover and <i>identify</i> cyber threats across networks that include endpoints, mobile devices, virtual systems, as well as cloud and IoT devices.</p>
<p><i>Phase 2: Attacker phase</i></p> 	<p>Subsequently, AI would have <i>detected</i> anomalous activity from the pads as soon as it began and also identify interruptions in the data transfer and invalid data points. The investigation phase would have helped analysts determine a number of things, including when the attack occurred, how it occurred and the attack vector used, as well as the area(s) they mostly invested their time and resources on within the Cyber Kill Chain. Through <i>investigation</i> and <i>analysis</i>, AI would have provided security analysts with in-depth details on the nature of the attack, motivation, as well as the underlying vulnerability. This would, in turn, allow the security analysts to manage the cyber risk and reduce future risks of such incidents. As a way</p>

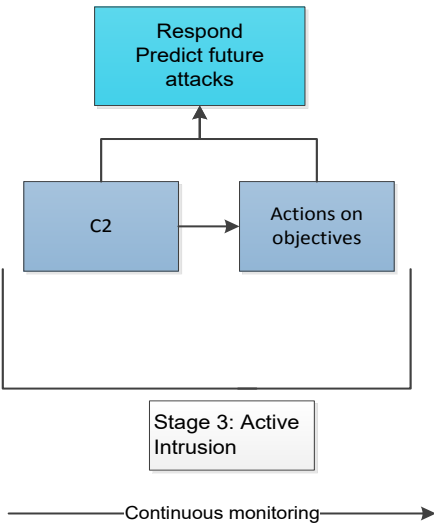
<b>Cyber Kill Chain</b>	<b>Application of CAIBER Framework</b>
	<p>of ensuring cyber resilience of IoT devices, AI constantly monitors activities, risks and also make informed decisions. AI would have made it easy for analysts to sift through millions of events in a way that is impossible using manual methods to find IoC and prevent persistent attacks such as DoSs. In <i>response</i> to the DoS attack, AI would have isolated other machines that have already been infected, and prevent the attack from running on any endpoint throughout the organisation. Prior to the prevention of the attack, AI would have isolated the pads from the organisation's network as they would have been foreign devices and an immediate investigation on the pads and their activities would have ensued.</p>
<p><i>Phase 3: Active Intrusion phase</i></p>  <pre> graph TD     A[Respond Predict future attacks]     B[C2]     C[Actions on objectives]     D[Stage 3: Active Intrusion]     E[Continuous monitoring]      B --&gt; A     C --&gt; A     B --&gt; C     D --- E     </pre>	<p>AI would have <i>prevented</i> the treat from progressing and compromising other devices on the network. The use of AI would have aided in <i>predicting</i> cyber threats, which in turn would have given Company_Magix a competitive edge, improved the end-user experience and increased the level of trust by its clients.</p>

Table 10 above presents the application of the CAIBER framework on the Cyber Kill Chain mapping of the case study that relates to IoT devices that caused a DoS attack. The application of the CAIBER framework with all the elements would have enabled Company\_Magix to discover and identify foreign devices in their networks. AI is self-learning, adaptive and flexible; it would have been to detect, in real time, any malicious activities and deviation of normal traffic or any anomalous activities across its entire networks. AI would have continuously monitored Company\_Magix's environment for both unknown and unknown cyber threats and trends. Any deviation from a pattern of life within the network, AI would have indicated that as a threat or compromise. The investigation element would have provided the company's security team with in-depth and rich details of the event. AI would have thoroughly analysed network traffic data and intelligently handled the unexpected cyber threats in

order to accurately detect unauthorised access. In essence, the application of the CAIBER framework with all the elements would have enhanced Company\_Magix's cyber resilience and the visibility necessary to defend IoT devices in the increasingly hostile, evolving and unpredictable cyberspace.

#### **5.4 Conclusion**

The mapping of the Cyber Kill Chain demonstrated how the proposed CAIBER framework could enhance cyber resilience through its application to the attack lifecycle. The application of the proposed framework was demonstrated through the case studies of the four companies that were interviewed. The case studies demonstrated how the application of the proposed framework with all the elements would enable organisations to prioritise events, anticipate and better prevent incidents before they occur. It also demonstrated how the application of the proposed framework with all the outlined elements would increase the visibility of the network and enable organisations to identify threats in the early stages of an attack.

The next chapter will conclude the study and also summarise the key findings of the study. The chapter will make recommendations and address possible areas for future research on cybersecurity, especially within the South African context.

## **CHAPTER 6: CONCLUSION**

### **6.1 Introduction**

This chapter will conclude the study and also provide a brief summary of the chapters entailed in the study. In brief, this chapter will provide an overview of all the chapters, reflect on the key questions the study detailed in chapter 1 and also provide an overview of the contribution made by the study. Lastly, the chapter will make recommendations and address possible areas for future research on cybersecurity, especially within the South African context. The following section will address the need to have the CAIBER framework.

### **6.2 Need for CAIBER Framework to Defend Cyberspace**

Cybersecurity is an emerging challenge for the South African national security. This has been demonstrated by multiple cyberattacks to both private and public sector (which had a knock-on effect on ordinary citizens). These attacks included the defacing of government websites, including that of The Presidency by hacktivists, ransomware attacks such as WannaCry and data breaches to one of South Africa's largest insurance companies, Liberty Holdings (refer to Status of cyber challenges in South Africa). Some of these attacks were politically motivated, while others were aimed at pointing out the lack of security measures within government departments. Some attacks were for financial gains and some were aimed at undermining the security of its citizens (refer to Current State of South African Cyber Community).

The technological advancement, together with business opportunities in cyberspace, have resulted in increased cybercrime. The introduction of modern technologies has given rise to improved digitisation of personal information, networking of technologies, increased global connectedness and the networked society but it has also exposed users to exploitation. Cyber threats and actors are evolving over time and are becoming faster, more frequent and sophisticated, thus the application of technological developments for cyber resilience is important. Human capability is limited to detect cyber threats in real time, to monitor activities in the network, determine known and unknown threats in the system, analyse large sets of data, investigate cyber events and subsequently prevent those attacks. Thus, the application of AI that is flexible, efficient and can identify cyber threats in real time and with improved accuracy (in contrast to human) is integral to cyber resilience in South Africa. However, the application of AI for enhanced cyber resilience does not mean the human element will be replaced rather this will reduce the workload of security analysts, time spent on tasks and allow them to focus on other critical tasks.

One of the cumulative and challenging risks within the cyber domain is the use of automation by cybercriminals to launch sophisticated, seamless and faster attacks. Thus, in order to better thwart these attacks and improve cyber resilience, it is critical to continue researching multiple ways in which

AI could be used for cyber defence. The following section will provide a summary of the research study.

### **6.3 Summative Overview of Research**

This research study comprises six chapters. This section will briefly introduce each chapter and highlight the respective contributions.

#### **Chapter 1: Research overview**

Chapter 1 presented key concepts that dominated the study. It also looked at different sources of cyberattacks and their diversified motives. The chapter addressed the problem statement and subsequently provided a roadmap detailing how the study would go about deriving a solution. It outlined questions that the study set out to address, the research mission, the contribution of the study and the limitations associated with the study. Lastly, the chapter enlightened the readers with regard to the structure of the research study.

#### **Chapter 2: Literature study**

Chapter 2 focused on the research that has already been conducted on cybersecurity and AI. The chapter looked at the status of the South African cyber community, as well as challenges faced by cyber users. Having discussed the cyber challenges, a brief discussion relating to the benefits introduced by the Internet and IoT, as well as accompanying risks associated with these technological developments were also discussed. Furthermore, the chapter addressed current approaches employed in cyber defence. A literature review on the significance of combining AI and cybersecurity was also conducted. Lastly, the chapter presented existing research and applications of AI techniques to cybersecurity.

#### **Chapter 3: Research design and methodology**

Chapter 3 outlined research design and methodology as well as the empirical and non-empirical techniques applied in the study. The chapter also described data collection methods, the sampling procedure and provided a brief analysis of the data collected. Additionally, it described ethical considerations and the limitations of the study.

#### **Chapter 4: Proposed CAIBER Framework**

Chapter 4 focused on establishing the CAIBER framework. It proposed a framework that is aimed at demonstrating the significance of enhancing cyber resilience through AI. Chapter 4 described how the CAIBER framework was developed after in-person interviews with four South African companies operating within the cyberspace (see 4.1.1 Background). The CAIBER framework was also inspired by the NIST CSF. The relevance of the framework was demonstrated by mapping out its elements to the AI-enabled tools that are currently used by the companies interviewed.

## **Chapter 5: Application of the CAIBER Framework**

Chapter 5 focused on the application of the proposed CAIBER framework. The chapter began by describing the Cyber Kill Chain and its phases. The chapter demonstrated the application of the proposed framework by mapping it to the Cyber Kill Chain. The application of the proposed framework was further demonstrated through its application to four different case studies based on knowledge obtained from the four participant companies. The case studies demonstrated how the application of the proposed framework would aid in remediating cyber threats, thus enhancing the cyber resilience of organisations. Lastly, the chapter took the reader through the phases of the Cyber Kill Chain and how the application of the proposed CAIBER framework could help identify and detect threats in the early stages of the Cyber Kill Chain, thus thwarting the attack before it further persists to cause damage.

## **Chapter 6: Conclusion**

Chapter 6 concludes this study by emphasising the goal of the research, highlighting how the study was structured and explaining some of the key findings to be taken forward. This chapter also reflects on the achievement of the research questions, provides an overview of the contribution made by the study and, lastly, it makes recommendations for future research. The following section will reflect on the research questions of the study.

### **6.4 Reflection on Achievement of Research Study**

This section will reflect on the research questions (refer to 1.6.2 Research questions) that the study aimed to achieve as noted in chapter 1.

#### **6.4.1 What is South Africa's current approach to cybersecurity?**

The research reflected on how the digital landscape and attack surface is evolving with the development of technologies. Therefore, some cyber defence approaches still used by some South African cyber users (which include firewalls and antiviruses) are failing to keep pace with the mutation of new cyber threats. Some organisations still use IDS technology, which as noted in the study in chapter 2 (see 2.2.2 Current approaches to cyber defence), has shortfalls and limitations that include prohibiting it from enhancing the users' cyber resilience required for the ever-evolving cyber threats.

As stated in chapter 2, (see 2.2.2 Current approaches to cyber defence) SIEM tools form an important part of a cyber-defence strategy, but these tools provide insufficient cyber defence, particularly in this new age of dynamic cyber threats. The limitations of SIEM tools include its heavy reliance on the human element; and it requires security teams to continually build processes, correlation rules and uses cases. SIEM solutions often lack granular details about events, lack real-time identification of cyber threats, produce reports that contain too much noise and security teams are required to sift through millions of alerts. Lastly, in order to enhance cyber resilience, organisations are required to integrate SIEM tools with other tools such as AI-enabled tools.

#### **6.4.2 What can be done to improve the current cyber defence employed in South Africa?**

There are multiple approaches that can be applied to improve cybersecurity in South Africa. However, this particular study focused on demonstrating the significance of enhancing cyber resilience through the application of AI to cybersecurity. The study presented empirical and theoretical research on the prospects of enhancing cyber defence capabilities by means of increasing the intelligence of defence systems with AI tools.

#### **6.4.3 What kind of AI tools can be used to actively defend cyberspace in South Africa?**

AI tools that were identified in the study as depicted in Chapter 2 (refer to 2.4.2 Examples of AI application in cyber defence), include ML, NN, DNN, AIS, Intelligent Agents, Generic Algorithms and Fuzzy Logic. AI tools are best suited to secure the cyberspace as they have been increasingly applied in the area of information security, information assurance, and cybersecurity measures. These tools have been used in multiple cyber defences and they have been proven to be able to enhance cyber resilience for cyber users.

For instance, ML has been the most popular application of AI in general and specifically for cybersecurity. It has been applied by Darktrace, Exabeam, CrowdStrike (refer to 3.8.1 Overview of companies) to improve network visibility, enhance detection capabilities, enhanced analysis, and learning pattern of everyday life, resolve complex problems as well as discover previously-unknown cyber threats and patterns, as well as prediction of threats. ANN has been used for intrusion detection and prevention; moreover, it has been used for DoS attacks, computer worm detection, spam detection, malware classification and forensic investigation. Intelligent Agents have been used to uncover suspicious cyber activities, detect cyber threats and verify properties of cybercrimes and to prevent cyberattacks.

AI tools that include Fuzzy Logic, Intelligent Agent Systems, Genetic algorithms and AIS have been key in solving cybersecurity complex problems, decision making, monitoring of cyber threats and detection, preventing and also predicting cyber threats. However, the identification of these tools by this research study does not suggest that these are the only AI tools that can be used for cyber resilience (refer to 2.4.2 Examples of AI application in cyber defence).

#### **6.4.4 How will developing and implementing an AI-based framework enhance cyber resilience in South Africa?**

Information security and the protection of cyber infrastructure require flexible, adaptable and active cyber defence systems that can make intelligent decisions in real time, detecting a wide variety of threats and attacks. Thus, this study developed an AI framework that is aimed at enhancing cyber defence capabilities. Among the significant benefits that an AI-driven solution will provide are automation of tasks, behaviour analytics, as well as the provision of intelligence that will result in an actionable decision and ultimately help enable continuous security improvements for cyber users.



The application of AI to cyber defence will help organisations to discover and identify cyber threats in real time and with much accuracy in the early stages of the attack. It will further provide an organisation with technology that is self-organising, flexible, intuitive, adaptable, and dynamic and has the ability to learn new behaviour, cyber threats and trends. AI can use the knowledge it gains to detect threats, including those that are yet to be discovered, by identifying shared characteristics within families of threats. AI will also aid in identifying gaps in the defence program and improve response time by proactively hunting for attackers.

The use of AI in the cyberspace can help security teams to identify patterns that may indicate a threat that ordinarily would be missed by conventional cybersecurity tools that mentioned in Question 1. This can in turn help analysts to spend less time studying "false positive" alerts and investigating blank walls while missing genuine malicious activities. Moreover, AI will allow analysts to correlate attacks or events across time and geography in order to develop a comprehensive picture within the network.

#### **6.4.5 How will the proposed AI framework enhance cyber defences currently employed?**

Organisations require a system that would continuously monitor their environment with the aim of detecting adversarial actions or intentions, thus the monitoring element within the CAIBER Framework is an overarching and continuous process. The use of AI for cyberdefence will provide security teams with in-depth, rich and comprehensive details of cyber events. It will help analysts determine varied information, including the attack vector, status of attack and, moreover, help uncover known and unknown vulnerabilities, as well as aid in the gathering of comprehensive information about the attacker.

The cyberspace comprises diversified aggressors with varied motivations; thus, the CAIBER Framework proposes a shift in defence surface within the South African context, a shift that is inclusive of AI for cybersecurity. The application of the proposed framework with all the elements will enable the organisation to prioritise cyber risks and better prevent incidents before they occur or cause damage. Furthermore, it will enable the organisation to have a complete view of its entire environment. The application of the proposed framework with all the elements will increase the level of security and trust and enhance users' experience. Moreover, it will place the organisation at a competitive edge. Moreover, the application of the proposed framework will enable security teams to respond in a timely manner in detecting threats, thus preventing them from any form of progression to the organisation's network.

In essence, the application of the proposed AI framework with all the elements will help shift the company's cybersecurity capabilities from reactive to more proactive; it will enable an organisation to have a complete view of its entire environment and it will help shift the security approach to be more user-centric. In that way, the company will understand the user's behaviour. Ultimately, this will enable organisations to be more resilient to new and evolving cyber threats and also help them to better

understand their clients and business environments. The next section will describe the contribution of this research study.

## **6.5 Contribution of Study**

Not only will the proposed framework lay the foundation for future research on the significance of defending cyberspace through AI, but it will also enhance the users' understanding of their environment. Innovative approaches that include the application of AI in order to provide cyber users with a secure, reliable, interoperable cyberspace are a necessity for cyber defence. Using a system that incorporates all the elements outlined in the framework will increase the visibility of the network environment and enable organisations to identify threats in the early stages of the attack. It will increase the level of security and trust and enhance the experience of users. The framework is, however, not a one-size-fits-all approach to managing cybersecurity as different cyber users (private companies, government, individual and academia) face distinctive cyber risks as they have different vulnerabilities and capabilities.

The application of AI in cyberspace will enable organisations to proactively detect unknown threats and diminutive nuances and changes in data. Moreover, AI will be more conducive to the protection of newer and evolving technologies that include IoT, cloud service and smart mobile devices. The use of technology that allows the organisation to monitor user behaviour and predict threats not only gives them a competitive edge, but it also reduces cyberattacks and also improves the end-user experience while increasing the level of security and trust. The following section will discuss recommendations that relate to the study.

## **6.6 Recommendations for Future Research**

Cybersecurity is a multi-dimensional, crosscutting and cross-disciplinary challenge that requires a more comprehensive and inclusive coordination between state departments responsible for cybersecurity, as well as collaborative efforts between the private sector, academia, and citizens. The South African government and private sector need to be at the forefront of AI and cybersecurity initiatives in order to provide its citizens with a secure, reliable, interoperable cyberspace. The South African government together with the private sector and academia needs to invest in cybersecurity research and development related to AI. Moreover, academia and research institutions are the seedbeds of AI development and they offer fertile ground for scientists and engineers to explore their ideas.

The implementation of AI for cybersecurity should not be limited to any government department or particular industry in the private sector; rather, it should be employed across different public and private sectors for more enhanced cybersecurity in South Africa. However, it should be noted that different businesses and sectors have diversified cyber risks, requirements for cybersecurity,

resources and budgets and different exposure to cyber threats. Therefore, more research on the development and implementation of AI measures for cybersecurity for the varied players should be conducted. For instance, the cybersecurity needs of the health industry are different from the needs of the mining industry, while the needs of mining are also different from those of the financial sector. More studies should be conducted regarding the skills set and capacity required in South Africa in order for the country to take full advantage of implementing AI for cybersecurity. More research within the South African cyber domain needs to be conducted on the application of different AI tools for cybersecurity.

Additionally, the South Africa government is in the process of signing the Cyber Security and Cyber Crimes Bill, which will aid in dealing with some of the cybersecurity challenges faced by the country. However, the Bill does not address how the law will regulate machine versus machine attacks. Thus, there should be investments in research and development that will focus on, for instance, the advancements of AI in the cybersecurity field for malicious activities, as well as implications on policy.

The advancement of AI will give rise to a range of intelligent implementations within the cybersecurity field that will potentially transform the nature of daily operations of cyber users and the structure of the workplace in all industries that are active on the cyber domain in South Africa. These implementations will be delivered as a new class of intelligent apps, IoT, etc. and provide intelligence to a wide range of mesh devices. Such implementations will enhance the attack vector, thus more research on such developments should be conducted for cyber users to be best secured.

Another technology development that requires research in terms of the application of AI for cyber resilience is Blockchain and distributed-ledger concepts. Multiple cyber users have embraced the use of Blockchain as it is transforming multiple industries, including the financial services industry, music distribution, etc. However, this technology is also enhancing the cyberattack vector. Therefore, research into AI application for cyber resilience within this domain should be considered. Data protection and the protection of users' privacy within the cyberspace should be an ongoing conversation within research institutions, and the application of technology advancement such AI should enhance those conversations. Additionally, more research into the banking, automobile and health sector with specific focus on AI and cybersecurity should be conducted. The following section will not only conclude the study but will provide recommendations that relate to cybersecurity in South Africa.

## **6.7 Closure**

There are no silver bullets for solving the cybersecurity threat. The cybersecurity landscape is in a constant state of flux, meaning that cyberattacks are constantly improving in sophistication and complexity. The introduction of modern technologies has given rise to improved digitisation of personal information, networking of technologies such IoT, cloud, smart devices, Blockchain, etc. and

has increased global connectedness and the networked society. However, this has also exposed users to multiple cyber risks. Countering unconventional and determined adversaries require an active approach to security.

The implementation of the CAIBER Framework will not only lay a foundation for future research on the significance of defending cyberspace through AI, but it will also enhance the users' understanding of their environment. Moreover, innovative approaches that include the application of AI in order to provide cyber users with a secure, reliable and an interoperable cyberspace are a necessity for cyber defence in South Africa. Cyber resilience in South Africa will increase the level of security and trust and enhance users' experiences. The objective of the proposed framework is to enhance cyber resilience and aid in providing measures that will ensure that cyber users' information, their integrity and confidentiality are properly protected.

## References

- AI.Business. 2016. *Artificial Intelligence in Defence and Security Industry*. AI Business, Retrieved from: <http://ai.business/2016/06/21/artificial-intelligence-in-defence-and-security-industry/> (Accessed on 15 March 2017).
- Andrew, S. and Halcomb, E.J., 2009. *Mixed methods research for nursing and the health sciences*. John Wiley & Sons.
- Arshia, B., Gayathri, M. and Manaswini, P. 2017. AI in Cyber Security. *International Journal of Engineering Research in Computer Science and Engineering*, 4, (9), pp 51-55.
- Avira. 2017. *The Application of AI to Cybersecurity: An Avira White Paper*. Retrieved from: [https://oem.avira.com/resources/whitepaper\\_AI\\_EN\\_20180306.pdf](https://oem.avira.com/resources/whitepaper_AI_EN_20180306.pdf). (Accessed on 27 November 2017).
- Babbie, E. 2010. *The Practice of Social Research*. London: Wadsworth Cengage learning.
- Barika, F., Hadjar, K. and El-Kadhi, N., 2009. Artificial neural network for mobile IDS solution. *Security and Management*, pp.271-277.
- Barrett, M. 2017. *A Framework for Protecting Our Critical Infrastructure*. Retrieved from: <https://www.nist.gov/blogs/taking-measure/framework-protecting-our-critical-infrastructure> (Accessed on 21 May 2018).
- Bitter, C., Elizondo, D.A and Watson, T. 2010. "Application of Artificial Neural Networks and Related Techniques to Intrusion Detection", IEEE World Congress on Computational Intelligence (WCCI 2010), pp. 949 – 954.
- Buntz, B. 2016. *The 10 Most Vulnerable IoT Security Targets*. Retrieved from: <http://www.ioti.com/security/10-most-vulnerable-iot-security-targets>. (Accessed on 10 July 2018).
- Burns, N. and Grove, S. (2009) *The practice of nursing research: Appraisal, synthesis and generation of evidence*. 6th Edition, Saunders Elsevier, St. Louis.
- Caron, X., Bosua, R., Maynard, S.B. and Ahmad, A., 2016. The Internet of Things (IoT) and its impact on individual privacy: An Australian perspective. *Computer law & security review*, 32(1), pp 4-15.

- Cisco. 2017. *Cisco 2017 Annual Cybersecurity Report*. Retrieved from: <https://www.cisco.com/c/en/us/products/security/security-images-acr2017.html> (Accessed 15 on October 2017).
- Craigen, D., DAlkun-Thibault, N. and Purse, R. (2014). Defining cybersecurity. *Technology Innovation Management Review*, 4(10).
- Dalziel, H., 2015. *Securing Social Media in the Enterprise*. Syngress.
- Darktrace. 2016. *The Enterprise Immune System: Proven Mathematics and Machine Learning for Cyber Defence*. Retrieved from: [https://microstrat.com/sites/default/files/Enterprise%20Immune%20System%20—%20US%20\(2\)\\_0.pdf](https://microstrat.com/sites/default/files/Enterprise%20Immune%20System%20—%20US%20(2)_0.pdf) (Accessed on 16 March 2018).
- Darktrace. 2017a. *General Data Protection Regulation (GAP)*. Retrieved from: <https://www.groveis.com/docs/darktrace-gdpr-whitepaper.pdf> (Accessed on 19 May 2018).
- Darktrace. 2017b. *Machine Learning: A Higher Level of Automation*. Retrieved from: [https://www.ciosummits.com/Online\\_Asset\\_Darktrace\\_Whitepaper-Machine\\_Learning.pdf](https://www.ciosummits.com/Online_Asset_Darktrace_Whitepaper-Machine_Learning.pdf) (Accessed on February 2017).
- Davies, S. 2018. *The ongoing battle against cybercrime in South Africa*. Retrieved from: <https://www.iafrikan.com/2018/04/19/the-ongoing-battle-against-cybercrime-in-south-africa/> (Accessed on 10 May 2018).
- De Bruijne, M., Van Eeten, M., Gañán, C. and Pieters W. 2017. *Towards a new cyber threat actor typology: A hybrid method for the NCSC cyber security assessment*. Retrieved from: [https://www.wodc.nl/binaries/2740\\_Volledige\\_Tekst\\_tcm28-273243.pdf](https://www.wodc.nl/binaries/2740_Volledige_Tekst_tcm28-273243.pdf) (Accessed on 19 February 2018).
- De Spiegeleire, S., Maas, M. and Sweijts, T., 2017. *Artificial Intelligence and the Future of Defence: Strategic Implications for Small-and Medium-Sized Force Providers*. The Hague Centre for Strategic Studies.
- Deloitte. 2017. *The cyber security imperative: Protect your organisation from cyberthreats*. Retrieved from: <https://www2.deloitte.com/content/dam/Deloitte/ca/Documents/risk/ca-en-13-3694-cyber-security-pov-fin.pdf> (Accessed on 05 October 2017).
- Denning, D. E., "Barriers to Entry: Are They Lower for Cyber Warfare?" *IO Journal*, pp 6-10.

Devlin, C. 2016. *Global Threat Intelligence report ahead of Government Cyber Security Summit*. Retrieved from: <http://www.stuff.co.nz/business/79488994/global-threat-intelligence-report-ahead-of-government-cyber-security-summit>. (Accessed on 05 October 2017).

Dimension Data. 2017. *The Executive's Guide to the 2017 Global Threat Intelligence Report*. Retrieved from <https://www.dimensiondata.com/-/media/dd/corporate/global/pdf/the-executives-guide-to-the-2017-global-threat-intelligence-report.pdf>. (Accessed on 18 November 2017).

Dilek, S., Cakır, H. and Aydın, M. 2015. Applications of Artificial Intelligence Techniques to Combating Cyber Crimes: A Review. *International Journal of Artificial Intelligence & Applications*, 6(1), pp 21-39.

Dorman, N. (2012). *To defend the web: using artificial intelligence as online security*. Retrieved from: <http://www.pitt.edu/~nad59/nadwa3.docx>. (Accessed on 09 August 2016).

Duncan A. 2016 SA business 'unprepared' for cybercrime. Retrieved from <http://www.fin24.com/Tech/Cyber-Security/sa-business-unprepared-for-cybercrime-20160609>. (Accessed on 20 February 2017).

Enisa, 2017. *ENISA Threat Landscape Report 2017*. Retrieved from: <https://www.enisa.europa.eu/publications/enisa-threat-landscape-report-2017> (Accessed on 19 July 2018).

Feng, C., Wu, S. and Liu, N. 2017. A user-centric machine learning framework for cyber security operations center. In *Intelligence and Security Informatics (ISI), International Conference on* (pp. 173-175). IEEE

Fralely, J.B. and Cannady, J. 2017. The promise of machine learning in cybersecurity. In *SoutheastCon, 2017* (pp. 1-6). IEEE.

Gagliardi, F., Hankin, C., Gal-Ezer, J., McGettrick, A. and Meitern, M. (2016). *Advancing Cybersecurity Research and Education in Europe*. Retrieved from: [https://www.acm.org/binaries/content/assets/publicpolicy/2016\\_euacm\\_cybersecurity\\_wHITE\\_paper.pdf](https://www.acm.org/binaries/content/assets/publicpolicy/2016_euacm_cybersecurity_wHITE_paper.pdf) (Accessed on 30 May 2017).

Gibson, W. (1984) *Neuromancer*. New York, Ace Books.

Gladen, E. (2017). *Trend for 2017 & Beyond – Artificial Intelligence Fuelled Human Threat Hunting*. Retrieved from: <https://www.cyberseer.net/blog-artificial-intelligence-fuelled-human-threat-hunting-cyber-security/> (Accessed on 19 January 2017).

Greene, T. 2017. *Why the 'cyber kill chain' needs an upgrade*. Retrieved from: <https://www.networkworld.com/article/3104542/security/why-the-cyber-kill-chain-needs-an-upgradesecurity-pros-need-to-focus-more-on-catching-attackers-aft.html> (Accessed on 19 May 2018).

Harel, Y., Gal, I.B. and Elovici, Y. 2017. Cyber Security and the Role of Intelligent Systems in Addressing its Challenges. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 8(4), pp 1-12.

Howarth, F. 2016. *Evolving uses of the kill chain framework*. Retrieved from: <https://logrhythm.com/pdfs/3rd-party-whitepaper/uk-bloor-evolving-uses-of-the-kill-chain-framework-independent-white-paper.pdf> (Accessed on 19 May 2018).

Husain, R. and Muhammad, S. 2013. A Survey on soft computing techniques in Network Security. *Scholarly Journal of Mathematics and Computer Science*, 2(3), pp. 28-32.

Hutchins, E.M., Cloppert, M.J. and Amin, R.M. 2011. *Intelligence-driven computer network defense informed by analysis of adversary campaigns and intrusion kill chains*. *Leading Issues in Information Warfare & Security Research*, 1(1), pp. 1-12.

Jacobs P., Arnab, A and Irwin B. 2013. Classification of security operation centres. In *Information Security for South Africa, 2013* (pp. 1-7). IEEE.

Jan Vermeulen. 2017. *How did WannaCry traffic take down Telkom's app, call centre, and website*. Retrieved from: <https://mybroadband.co.za/news/security/212442-how-did-wannacry-traffic-take-down-telkoms-app-call-centre-and-website.html> (Accessed on 01 July 2017).

Johnson, R.B. and Onwuegbuzie, A.J., 2004. Mixed methods research: A research paradigm whose time has come. *Educational researcher*, 33(7), pp.14-26.

Johnson, R.B., Onwuegbuzie, A.J. and Turner, L.A., 2007. Toward a definition of mixed methods research. *Journal of mixed methods research*, 1(2), pp.112-133.

Jongsuebsuk, P., Wattanapongsakorn, N. and Charnsripinyo, C.2013. Real-time intrusion detection with fuzzy genetic algorithm. In *Electrical Engineering/Electronics, Computer,*



*Telecommunications and Information Technology (ECTI-CON), 2013 10th International Conference on* (pp. 1-6). IEEE.

Jyothsna S., M. and Nilina, T. 2013. Prospects of Artificial Intelligence in Tackling Cyber Crimes. *International Journal of Science and Research*, 6, (14), pp 1717-1723.

Kawulich, B.B., 2005. Participant observation as a data collection method. In *Forum Qualitative Sozialforschung/Forum: Qualitative Social Research*, 6 (2) pp 1-33.

Korolov, M. and Myers L. 2017. *What is the cyber kill chain? Why it's not always the right approach to cyber attacks*. Retrieved from: <https://www.csoonline.com/article/2134037/cyber-attacks-espionage/strategic-planning-erm-the-practicality-of-the-cyber-kill-chain-approach-to-security.html> (Accessed on 12 February 2018).

Krauss, S.E. 2005. Research paradigms and meaning making: A primer. *The qualitative Report*, 10(4), pp.758-770.

Kumar, G.P. and Reddy, D.K. 2014. An agent based intrusion detection system for wireless network with artificial immune system (AIS) and negative clone selection. In *Electronic Systems, Signal Processing and Computing Technologies (ICESC), 2014 International Conference on* (pp. 429-433). IEEE.

Kumar, V. 2014. Interview Methods in Research. Retrieved from: <https://www.slideshare.net/VinayKumar49/interview-method-in-research> (Accessed on 19 July 2018).

Lanotte, R. and Merro, M., 2018. A semantic theory of the Internet of things. *Information and Computation*, 259, pp. 72-101.

Leedy, P. D., & Ormrod, J. E. (2013). *Practical research: Planning and design*. Boston: Pearson.

Leininger, M. M. (1985). *Qualitative research methods in nursing*. Orlando, Fla: Grune & Stratton.

Li, H., Li, S. and Tryfonas, T. 2016. The Internet of Things: a security point of view. *Internet Research*, 26(2), pp.337-359.

Lotfollahi, M., Shirali, R., Siavoshani, M.J. and Saberian, M. 2018. Deep Packet: A Novel Approach For Encrypted Traffic Classification Using Deep Learning. arXiv preprint arXiv:1709.02656. (pp 1-13). IEEE.

- Maier, D., 2017. Can artificial intelligence help in the war on cybercrime?. *Computer Fraud & Security*, 2017(8), pp.7-9.
- Mahlaka, R. 2018. *Counting the cost of Liberty's cyber attack*. Retrieved from: <https://www.moneyweb.co.za/news/companies-and-deals/counting-the-cost-of-libertys-cyber-attack/> (Accessed on 10 July 2018).
- Marlow, C.R. 2005. *Research Methods for Generalist Social Work*. New York: Thomson Brooks/Cole.
- Marshall, C. and Rossman, G.B., (1999) *Designing qualitative research*. 3rd ed. London: Sage Publications.
- McCafferty, J. 2017. *NIST Publishes Update to Its Cybersecurity Framework*. Retrieved from: <https://misti.com/internal-audit-insights/nist-publishes-update-to-its-cybersecurity-framework> (Accessed on 21 May 2018).
- McElwee, S., Heaton, J., Fraley, F. and Cannady, J. (2017). Deep Learning for Prioritizing and Responding to Intrusion Detection Alerts. *Cyber Security and Trusted Computing*, 3, pp 1-5.
- Mngadi, M. 2018. *Presidency website up and running after hacking attack*. Retrieved from: <https://www.news24.com/SouthAfrica/News/breaking-presidency-website-hacked-20180707>. (Accessed on 10 July 2018).
- Murzina, A. (2016). *SIEM efficiency Survey*. Retrieved from: [https://www.netwrix.com/2016\\_siem\\_efficiency\\_survey\\_report.html](https://www.netwrix.com/2016_siem_efficiency_survey_report.html) (Accessed on 18 January 2018).
- NIST. 2018. *Framework for Improving Critical Infrastructure Cybersecurity*. Retrieved from: <https://nvlpubs.nist.gov/nistpubs/CSWP/NIST.CSWP.04162018.pdf> (Accessed on 17 April 2018).
- Nyirenda-Jere, T. and Biru, T., 2015. *Internet development and Internet governance in Africa. ISOC Report*. Retrieved from: <https://www.sbs.ox.ac.uk/cybersecurity-capacity/system/files/Internet%20development%20and%20Internet%20governance%20in%200Africa.pdf> (Accessed on 10 March 2017).
- Ojugo, A.A., Eboka, A.O., Okonta, O.E., Yoro, R.E. and Aghware, F.O. 2012. Genetic algorithm rule-based intrusion detection system (GAIDS). *Journal of Emerging Trends in Computing and Information Sciences*, 3(8), pp.1182-1194.

- Onwubiko, C. 2015. Cyber security operations centre: Security monitoring for protecting business and supporting cyber defense strategy. In *Cyber Situational Awareness, Data Analytics and Assessment (CyberSA)*, 2015 International Conference on (pp. 1-10). IEEE.
- Oxford Analytica. 2017. *Prospects 2017 forecasts key issues that will shape policy, economic and strategic issues globally in the year ahead*. Retrieved from: <https://www.oxan.com/media/1814/oxford-analytica-2017-briefing.pdf>. (Accessed 26 June 2017).
- Philip, L.J.1998. Combining quantitative and qualitative approaches to social research in human geography—an impossible mixture? *Environment and planning journal*, 30(2), pp.261-276.
- Qiang, H. and Yiqian, T. 2010. A Network Security Evaluate Method Base on AIS. In *Information Technology and Applications (IFITA), International Forum*, 2, pp. 42-45.
- Rajbanshi, A, Bhimrajka, S and Raina C. K. 2017. Artificial Intelligence in Cyber Security. *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*, 2(3), pp 132-137.
- Reveron, D., 2012. *Cyberspace and National Security*. Washington DC: Georgetown University Press.
- Robinson, M., Jones, K. and Janicke, H. 2015. Cyber warfare: Issues and challenges. *Computers & security*, 49, pp.70-94.
- Rudner, M., 2013. Cyber threats to critical national infrastructure: An Intelligence challenge. *International Journal of Intelligence and Counterintelligence*, 26(3), pp. 453-481.
- Rui, L. and Wanbo, L. 2010. Intrusion Response Model based on AIS. In *Information Technology and Applications (IFITA), International Forum*, 1, pp. 86-90.
- Russell, S.J. and Norvig, P. (1997). *Artificial Intelligence: A modern Approach*. Prentice Hall, Englewood Cliffs, New Jersey.
- Sager, T. 2014. Killing Advanced Threats in Their Tracks: An Intelligent Approach to Attack Prevention. Retrieved from: <https://www.sans.org/reading-room/whitepapers/analyst/killing-advanced-threats-tracks-intelligent-approach-attack-prevention-35302> (Accessed on 12 February 2018).

Schwandt, t. A. (2007). *The Sage Dictionary of Qualitative Inquiry* (Third ed.), Thousand Oaks, California: sage Publication, Inc.

Serianu. 2017. *Africa's cybersecurity report: demystifying Africa's cybersecurity poverty line*.

Retried from:

<http://www.serianu.com/downloads/AfricaCyberSecurityReport2017.pdf> (Accessed on 01 April 2018)

Shah, S.A.R. and Issac, B., 2018. Performance comparison of intrusion detection systems and application of machine learning to Snort system. *Future Generation Computer Systems*, 80, pp 157-170.

Shanmugam, B. and Idris, N.B., 2009, December. Improved intrusion detection system using fuzzy logic for detecting anomaly and misuse type of attacks. In *Soft Computing and Pattern Recognition, 2009. SOCPAR'09. International Conference of* (pp. 212-217). IEEE.

Sharma, S., Kumar, S. and Kaur, M., 2014. Recent trend in Intrusion detection using Fuzzy-Genetic algorithm. *International Journal of Advanced Research in Computer and Communication Engineering*, 3(5).

Sheridan, K. 2017. *Future of the SIEM*. Retrieved from: <https://www.darkreading.com/threat-intelligence/future-of-the-siem-/d/d-id/1328457?> (Accessed on 21 January 2018).

Skierka, I., Morgus, R., Hohmann, M. and Maurer, T. 2015. *CSIRT Basics for Policy-Makers*. Retrieved from: [https://static.newamerica.org/attachments/2943-csirt-basics-for-policy-makers/CSIRT\\_Basics\\_for\\_Policy-Makers.493bfc8eb0ef4caa90869a4db30f47ce.pdf](https://static.newamerica.org/attachments/2943-csirt-basics-for-policy-makers/CSIRT_Basics_for_Policy-Makers.493bfc8eb0ef4caa90869a4db30f47ce.pdf) (Accessed on 18 December 2016).

Symantec. 2016. *Cyber crime & cyber security Trends in Africa*. Retried from: [https://www.thehaguesecuritydelta.com/media/com\\_hsd/report/135/document/Cyber-security-trends-report-Africa-en.pdf](https://www.thehaguesecuritydelta.com/media/com_hsd/report/135/document/Cyber-security-trends-report-Africa-en.pdf). (Accessed on 01 April 2018)

Symantec. 2017. *Internet Security Threat Report*. Retrieved from: <https://www.symantec.com/content/dam/symantec/docs/reports/istr-22-2017-en.pdf>. (Accessed 26 February 2018).

Talwar, R. and Koury, A., 2017. Artificial intelligence—the next frontier in IT security?. *Network Security*, 2017(4), pp.14-17.

Tarnowski, I. 2017. *How to use cyber kill chain model to build cybersecurity?* Retrieved from: <https://tnc17.geant.org> (Accessed on 12 February 2018).

TimesLIVE. 2018. *South Africa is top target for cyberattacks.* Retrieved from: <https://www.timeslive.co.za/sunday-times/business/2017-11-02-south-africa-is-top-target-for-cyberattacks/> (Accessed on 10 May 2018).

Tuvey, E. 2017. *Fighting Fire with Fire – AI, Cyber Security, and Roles of the Future.* Retrieved from: <https://www.infosecurity-magazine.com/opinions/fighting-fire-ai-cyber-security/> (Accessed on 10 July 2017).

Tyugu, E. 2011. Artificial Intelligence in cyber defence, 3rd International Conference on Cyber Conflict (ICCC 2011), pp. 1–11.

UNECA. 2014. *Tackling the challenges of cybersecurity in Africa.* Retrieved from [https://www.uneca.org/sites/default/files/PublicationFiles/ntis\\_policy\\_brief\\_1.pdf](https://www.uneca.org/sites/default/files/PublicationFiles/ntis_policy_brief_1.pdf). (Accessed on 27 April 2017)

Urban, T. 2015. *The AI Revolution: The Road to Superintelligence.* Retrieved from: <https://waitbutwhy.com/2015/01/artificial-intelligence-revolution-1.html>. (Accessed on 15 February 2018)

Van der Merwe, M. 2017. *Ransomware: Prepare for more, bigger, worse – and closer to home.* Retrieved from: <https://www.dailymaverick.co.za/article/2017-05-22-ransomware-prepare-for-more-bigger-worse-and-closer-to-home/#.WW3uMTWxUdU>. (Accessed on 18 June 2017).

Velazquez, C. 2015. *Detecting and Preventing Attacks Earlier in the Kill Chain.* Retrieved from: <https://www.giac.org/paper/gsec/36774/detecting-preventing-attacks-earlier-kill-chain/145219>. (Accessed on 23 October 2017).

Venktesh, K. 2017. *'Explosive time' in SA for cybercrime after Hetzner breach.* Retrieved on <https://www.fin24.com/Tech/News/explosive-time-in-sa-for-cybercrime-after-hetzner-breach-20171102> (Accessed on 09 December 2017).

Vicente, A. (2016). *SA is top cybercrime target in Africa.* Retrieved from: [http://www.itweb.co.za/index.php?option=com\\_content&view=article&id=150566](http://www.itweb.co.za/index.php?option=com_content&view=article&id=150566) (Accessed on 11 December 2016).

- Von Solms, R. and Van Niekerk, J. 2013. From information security to cyber security. *Computers and Security* 38, pp 97–102.
- Weber, R.H. and Studer, E. 2016. Cybersecurity in the Internet of Things: Legal aspects. *Computer Law & Security Review*, 32(5), pp.715-728.
- White House, Executive office of the president. 2016. *Artificial intelligence, automation, and the economy*. Retrieved from:  
<https://www.whitehouse.gov/sites/whitehouse.gov/files/images/EMBARGOED%20AI%20Economy%20Report.pdf>. (Accessed on 18 June 2017).
- Wirkuttis, N. and Klein, H. 2017. Artificial Intelligence in Cybersecurity. *Cyber Intelligence, and Security Journal*, 1, pp.21-3.
- Wolden. M., Valverde, R. and Talla, M. 2015. The effectiveness of COBIT 5 Information Security Framework for reducing Cyber Attacks on Supply Chain Management System. *IFAC-Papers OnLine*, 48 (3), pp 1846–1852.
- Yampolskiy, R. V. 2017. AI Is the Future of Cybersecurity, for Better and for Worse. Harvard business school.
- Ye, X. and Li, J. 2010. A security architecture based on immune agents for MANET. *International Conference on Wireless Communication and Sensor Computing*, (pp. 1-5). IEEE.
- Zamani, M. and Movahedi, M., 2013. Machine learning techniques for intrusion detection. arXiv preprint arXiv:1312.2177. (pp 1-13). IEEE
- Zeigler, C. 2017. Carbon Black® Cb Defense - A Solution Review. Retrieved on <https://www.carbonblack.com/wp-content/uploads/2017/09/CB-Defense-NextGenAV.pdf> (Accessed on 07 October 2017).
- Zhang, Y., Wang, L., Sun, W., Green II, R.C. and Alam, M. 2011. Distributed intrusion detection system in a multi-layer network architecture of smart grids. *IEEE Trans. Smart Grid*, 2(4), pp 796-808.

## Appendix A

**Paper title:** Towards an artificial intelligence framework to actively defend cyberspace

**Paper authors:** Mmalerato Masombuka<sup>1</sup>, Marthie Grobler<sup>2</sup> and Bruce W. Watson<sup>1, 3</sup>

**Conference name:** 17<sup>th</sup> European Conference on Cyber Warfare and Security

**Conference website:** <https://www.academic-conferences.org/conferences/eccws/>

**Conference date:** 28-29 June 2018

**Conference location:** Oslo, Norway

### **Paper abstract:**

The Internet's expansion as a new power domain and the development of Information and Communications Technology has introduced an unprecedented magnitude of convenience and efficiency to its users. In addition, the innovation and novelty in many areas is increasing; likewise, the security challenges and accompanying risks are also accumulating. Cyberattackers are developing Artificial Intelligence (AI)-enabled malware that are adaptive, understand the target environment, have the ability to evade detection, continue to learn and make advanced decisions. In this regard, malware is getting smarter and cyberthreats are evolving and becoming more sophisticated and complex. Thus, human intervention and capacity is not enough to sufficiently deal with advanced threats, speed of processes, the amount of data, and the vulnerability of intrusion. This paper proposes the use of an AI framework to address these advanced threats.

The countering of advanced adversaries requires an active approach to security that will place an emphasis on proactive measures, real time detection, active monitoring and mitigation of key threats. Therefore, innovative approaches such as the application of AI tools that have a learning capacity, are adaptable, analysis driven and able to detect user behaviour make intelligent and real time decisions that would assist in fighting the cyberthreat. To demonstrate the need to defend the cyberspace using AI and show current progress by the South African private sector in terms of AI driven tools, four companies were interviewed. The cyberspace comprises of diversified aggressors with varied motivations; thus this paper proposes a shift in defence surface within the South African context, a shift that is inclusive of AI for cybersecurity.

## Appendix B

The questions below guided the collection process and they were also provided to the participants before the interview. These questions are open-ended and unstructured meaning the answers, opinions and perspectives provided by the interviewees guided the answers and the interview process.

### Question for interviews

1. What systems or software does the company for cybersecurity currently use?
2. Does the company use any AI tools to for cyber resilience? If so which tools?
3. If not, are there plans or intention to use AI to secure cyberspace?
4. Which tools do they use for monitoring?
5. What tools does the company use for identification, identifying and detecting cyberthreats? What about analysis and investigation?
6. What tools does the company use for prevention and prediction?
7. What are the limitations of the AI employed?
8. Can your AI adapt to new attacks and unknowns
9. Does the automation increase productivity?
10. What is the company's classification of cyber attacks?
11. Which are the most frequent cyber threats?
12. How often do they occur?
13. To who are they mostly directed? Civilians? Government?
14. What kind of data is being targeted? Personal data, Intellectual capital etc.
15. What are the main causes of cyber threats? Human error, lack of proper security measures?
16. How often does the SOC scan for threats?
17. What systems do they use to detect cyber threats?
18. What are the limitations to real time detection of threats?
19. How are cyber incidents analysed?
20. What kind of analysis approach do they use for cyber threats?
21. Is your system able to provide enhanced cyber reliance on other technologies that include IoT, cloud, wearable devices and other smart devices



## Questionnaires

What type of **prevention** capabilities is provided by your organisation? Please indicate whether these services are provided by an internal SOC, a SOC service (including cloud-based) or both. Leave blank those that don't apply and add those that apply to your SOC but not listed below.

How does your SOC **correlate and analyse** event data, (IOCs and other security- and threat-related data?) Select those that most apply and add those that apply to your SOC but not listed below.

Don't always know—it all happens in the cloud	
Through a threat intelligence platform	
Through our SIEM (security information event manager)	
Through home-developed APIs and dashboards	
Through our aggregated log management systems	
Other	

What type of **detective** capabilities are provided by your SOC? Please indicate whether these services are provided by an internal SOC, a SOC service (including cloud-based) or both.

Leave blank those that don't apply and add those that apply to your SOC.

Network intrusion detection and prevention	
Web application firewall (WAF)	
SIEM reporting and analytics	
Egress filtering	
Risk analysis and assessment	
Threat hunting	
Application log monitoring	
Deception technologies to use against attackers	
Customized or tailored SIEM use-case monitoring	
AI or machine learning	
Other	
DoS or DDoS protection	
Log management	
Endpoint or host-based detection and response (EDR)	
Windows event log monitoring	
Netflow analysis	
Threat Intelligence	

What **response** services does your organisation perform? Please indicate whether these services are performed by an internal staff, an outsourced service, or both.

Endpoint detection and response (EDR)	
Reverse engineering of malware	
Threat attribution	
Adversary containment	
Adversary deception	
Playbook-based response actions	
Other	
Adversary interruption	
Threat neutralization	
Network forensic analysis	
Command centre	
Threat campaign tracking	
Customer interaction (call centre)	
Hardware reverse engineering	

Which of these most closely resembles your organization's **definition of a security incident**?

We have no formal definition of a security incident.	
An incident is any adverse event or the threat of an adverse event above the normal level of noise.	
There are multiple specific incident types and impact levels that are formally defined as an incident in our organization.	
The organization doesn't use the term incident, because that would trigger regulatory or industrial reporting requirements it wants to avoid.	
We haven't sorted out the difference in the definitions of an incident, a breach and a threat, but we are hoping to.	
Other – please add your definition	

Which of the following **characteristics does your AI** comprise of? Only select those that apply to your organization (if not listed please do add)

Learning	
Is the learning supervised or unsupervised	
Adaptable to their environment	
Capable of acting intelligently	
It's doing things normally done by people	
Monitor a network's data	

Predict future attacks/threats	
Real-time risk assessment	
Real-time detection (potential and suspicious malicious traffic)	
Ability to identify and prevent both known and unknown cyber threats	
Self-learning and accumulate knowledge	
Studies behaviour of threats	
Minimised false alerts	
Prevention/Initiate countermeasures	
Classify various attacks/malicious data	
Flexibility and mobility	
Decision making mechanism	
Anomaly detection	
Investigation of cyber threats/attacks	