

The perceived impact of Artificial Intelligence on cyber security operations in the South African banking sector

Edward van Bosch

2419308

Supervisor: Dr Ayanda Magida

**Submitted to the Faculty of Commerce, Law and Management, University of
the Witwatersrand, in fulfilment for the degree of Master of Management in the
field of Digital Business**

Declaration

I, Edward John van Bosch, declare that this research report is my own work except as indicated in the references and acknowledgements. It is submitted in partial fulfilment of the requirements for the degree of Master of Management in the field of Digital Business at the University of the Witwatersrand, Johannesburg. It has not been submitted before for any degree or examination in this or any other university.

Name: Edward John van Bosch Signature: Edward John van Bosch

Signed at ...Randburg.....

On the 28th..... day of April..... 2025....

Abstract

Cybersecurity threats are increasing exponentially, and South Africa is one of the top countries targeted by cyber criminals. Banks are adopting AI-powered cybersecurity solutions to fortify themselves against the ever-increasing cyber threat landscape, but this presents opportunities and challenges. To ascertain the efficacy and impact of AI in cybercrime eradication, the TOE framework and Perception Theory were adopted to cover the operating landscape of CS officers and guided the cognition and the rationalisation of AI adoption. The data from the literature review was posited against qualitative findings from semi-structured interviews of key CS personnel and was analysed using thematic analysis. The findings indicate a positive perception of AI as a cybersecurity force multiplier, particularly in enhancing real-time threat detection and automating response mechanisms. However, concerns were also raised about overreliance on AI, data privacy and the potential exploitation of vulnerabilities and algorithms within AI-powered cybersecurity solutions. The findings clearly demonstrated that AI is not a turnkey solution and does not possess the power and ability to eradicate cybercrime in totality but is the only solid viable defensive solution for cyber-attacks. AI cannot eradicate crime completely but is the most efficient tool and system in cybercrime detection and prevention. There was a marked disparity in perception of AI between direct AI users and literary conjecture on the impact of AI. Given the vastness of AI, a novel bespoke framework was created to guide new users on the necessity, risk, benefits and scale of AI implementation intra-organisationally.

Keywords: AI, cybersecurity, cyberattacks, force multiplier, overreliance, solution.

Acknowledgements

To my creator, without whom nothing is impossible, I praise you for giving me the opportunities, abilities and persistence to complete this research.

To my wife Chantelle and daughters Kyla and Talor, thank you for always motivating and supporting me. We sacrificed a lot of time together. Thank you for being part of my life and your inspiration on this journey.

To my friend Adelia: thank you for your encouragement, your belief in me, and your motivation through very tough and challenging times to see this through and to finish my studies and research. You always inspired me to be more and to be better, and for that I will always be grateful. Thank you.

Thank you to all the participants who participated and who took time away from their very busy schedules.

Dr Ayanda Magida, thank you for guiding me through this process.

To everyone who assisted me on this journey — there are so many of you. Thank you very much for your support.

Table of Contents

Declaration.....	2
Abstract.....	3
Acknowledgements.....	4
List of Figures	9
List of Tables	10
List of Abbreviations.....	11
Chapter 1: Introduction.....	13
1.1 Statement of Purpose	13
1.2 Background of the Study	14
1.3 Research Problem.....	21
1.4 Research Questions.....	21
1.5 Justification for the study.....	22
1.6 Delimitations of the Study	23
1.7 Report Outline	24
Chapter 2: Literature Review.....	26
2.1 Introduction	26
2.2 Defining the TOE Framework	28
2.3 Technological component of TOE affecting CS operations	29
2.3.1 The necessity of Cybersecurity	30
2.3.2 AI CS Anthropomorphism	33
2.4 Organisational component of TOE affecting CS operations	40

2.5 Environmental component of TOE affecting CS operations	42
2.6 Challenges to AI CS operations	43
2.6.1 Technological Challenges.....	43
2.6.2 Organisational Challenges	46
2.6.3 Environmental Challenges	47
2.7 Defining Perception Theory.....	47
2.7.2 Cognition, Science and Perception Flow	51
2.8 AI-CS and PT – A Synthesis.....	52
2.9 Literature Review Findings Viz Research Questions.....	55
2.10 Conclusion.....	56
Chapter 3: Research Methodology	58
3.1 Introduction	58
3.2 Research Philosophy.....	59
3.3 Research Assumptions.....	61
3.4 Research Design	62
3.5 Research Strategy and Method	63
3.6 Population.....	64
3.7 Target Population and Sample Group.....	65
3.8 Unit of Analysis	66
3.9 Data Collection	67
3.10 Data Analysis.....	69
3.11 Quality Assurance	72
3.12 Ethical Considerations.....	73

Chapter 4: Findings and Discussion of Studies	75
4.1 Introduction	75
4.2 Summary Background of Participants.....	75
4.3 Interview Themes and Rationale	77
4.4 Interview Findings.....	79
4.4.1 Current use of AI in CS teams.....	79
4.4.2 Benefits of AI	82
4.4.3 Risks of AI in CS.....	85
4.4.4 CS-related Risk Mitigation Strategies	90
4.4.5 Impact on Security Posture	95
4.4.6 Automation on Activities	97
4.4.7 AI's Impact on CS Engineers.....	98
4.4.8 Decision-Making Visibility	103
4.4.9 Data Management and Governance	107
4.4.10 Protecting AI Systems	111
4.4.11 Protecting non-AI Systems	113
4.4.12 Does AI give companies a competitive edge	113
4.5 Conclusion.....	114
Chapter 5: Conclusion & Recommendations	116
5.1 Introduction	116
5.2 Conclusions regarding Research Question 1	117
5.3 Conclusions regarding Research Question 2	118
5.4 Conclusions regarding Research Question 3	119
5.5 Limitations of this study.....	120
5.6 Recommendations	121

5.7 Suggestions for further studies	124
References	127
APPENDIX A	139
APPENDIX B	141
APPENDIX C	143
APPENDIX D	145

List of Figures

Figure 1.1:Top three cyber risks in the financial services sector	17
Figure 1.2: Percentage of Fin companies underinvesting on CS capabilities.....	19
Figure 2.1: Difference between Simple and Deep Learning Neural Networks.....	37
Figure 2.2: Four Types of Metaheuristic Algorithms in CS operations	38
Figure 2.3: Difference between Internalists and Externalists.....	49
Figure 2.4: Difference between PT's Top-down Bottom-up Approach	52
Figure 2.5: Locating the subject on Perception Theory's Quadrants	53
Figure 3.1: Map of Research Methodology	59
Figure 4.1: Top 4 benefits of AI within CS operations	82
Figure 4.2: Benefits of AI in CS operations.....	84
Figure 4.3: Summary of Participants Risks in AI	86
Figure 4.4 CS Risks Across TOE and Protection Channels	89
Figure 4.5: List of 14 Risk Mitigation Tactics within CS Operations	90
Figure 4.6: Participants outlook on AI improving the banks security posture.....	95
Figure 4.7: AI Automation Scale	98
Figure 4.8: Participants Response to AI's Impact on Engineers in CS Operations...	99
Figure 4.9: Participants Results of AI's Decision-Making Visibility.....	104
Figure 4.10: Summary themes of participants feedback to AI Protection Measures	111

List of Tables

Table 2.1: Types of Cyber-Attacks (non-exhaustive)	32
Table 3.2: CS participants profiles	69
Table 4-1: Summary of Participants Background and Experience.....	77
Table 4.2: Interview Themes and Rationale	79
Table 4.3: AI tools and its Uses in CS teams.....	80
Table 4.4: Description of Risks to AI and its systems	94

List of Abbreviations

AAI – Applied Artificial Intelligence

AI – Artificial Intelligence

APT - Advanced Persistent Threats

BEC - Business Email Compromise

CPU - Central Processing Unit

CS – Cyber Security

DoS - Denial of Service

DDoS - Distributed Denial of Service

DL – Deep Learning

GAI – Generative Artificial Intelligence

GenAI – Generative Artificial Intelligence

GDP – Gross Domestic Product

HR – Human Resource

IIF- Institute of International Finance

IT – Information Technology

KPI – Key Performance Indicator

MA – Metaheuristic Algorithms

MitM - Man-in-the-Middle

ML – Machine Learning

NAS - Network Analysis Systems

NPS - Network Protection Systems

PAM – Privileged Access Management

POPIA - Protection of Personal Information Act

SaaS - Software as a Service

SABRIC - South African Banking Risk Centre

SA – South Africa

TAM - Technology Acceptance Model

TOE - Technological Organisational and Environmental Framework

XAI - Explainable AI

XXS - Cross-site Scripting

Chapter 1: Introduction

1.1 Statement of Purpose

The financial services industry in South Africa (SA) is at the helm of the digital revolution. The banking sector prioritises digital transformation at the heart of its strategy within the financial services sector. In SA alone, the Top 4 banks, in order of asset holding size, Standard Bank, First Rand, ABSA and Nedbank, account for over 50 percent of the continent's assets as Tier 1 banks. They have a combined user base of 20 million and contribute around 88 percent to the country's Gross Domestic Product (GDP) (Everington, 2024). The ease of internet and mobile banking across all platforms and devices has made transferring capital and data easy and efficient. However, the underside of digital transformation is a massive surge in cybercrime, resulting in financial losses year-on-year. SA's banking sector is one of the primary targets of cyberattacks. This was witnessed by SA having the third most cyber-crimes globally since 2020 (Bank, 2022). The rise in cybercrime warrants an increase in cyber-defence and cybersecurity (CS).

Four main technology trends are employed within cyber-security: Cloud and edge computing, Applied Artificial Intelligence (AAI), Next-generation software development and Trust architecture and digital identity (Atkins L. B., 2024). Of the four, AAI is most relied upon in the South African context and will form the focal point of this paper. As such, this paper seeks to ascertain the concatenation of AAI and cybersecurity and whether the former acts as a solution to the latter, an enabler to cyber-crimes or if it has a benign position but is adapted and implemented due to seeming populus consensus and the necessity of being on par or ahead of the digital curve and ahead of competitors. This will be achieved via a bi-pronged process.

1.2 Background of the Study

The deployment and use of AI in cybersecurity developed alongside and in conjunction with key AI disruptors in the banking sector. Today, AI is a part of the DNA of cybersecurity itself to the point of almost unconscious relegation. Innovation and strategic intent promote the interplay of critical cognitive functioning to support and advance machine learning. This paper seeks to question the efficacy of the impact of AI in CS. Are AI, AAI, and Generative AI (GAI) trends, necessities, or impediments in CS banking operations today? The impetus for the questioning lay within personal experience in the banking sector and personally witnessing the radical adoption of AI without testing its impact against higher intent, that is, safeguarding and protecting data and minimising threats. To ascertain the efficacy, the vantage point is unpacking the evolution of AI itself. How did AI become the ethos of the banking sector in SA, how did cybersecurity develop to the digital boom, and what challenges AI presents to CS operations? These questions will form the basis of the background and precede the core and supplementary research questions that follow.

AI has been one of the biggest disruptors in the banking sector. It has radically evolved from work automation and data analysis to machine learning, helping to create a unique personalised customer experience, to risk management and fraud prevention (Atkins L. et al, 2014). "The emergence of AI is disrupting the physics of the industry, weakening the bonds that have held together the components of the traditional financial institutions and opening the door to more innovations and new operating models" (Deloitte, 2025, p. 1). There have been four marked disruptions in the banking sector due to AI. Understanding the four points of disruption will assist in identifying the needs and importance of cybersecurity within banks.

Firstly, the impact of Big Data changed the experience and expectations of customers as an increasing number of people shifted towards the use of the internet and mobile devices. *Big data analytics* was able to collate unstructured data across multiple user platforms systematically and offer highly personalised experiences viz-a-viz traditional blanket offerings (Deloitte, 2025). Banks now adopt a 360-degree view of customers' interactions with brands, social media and the interfacing of personal data to inform their decision-making processes. The second disruptor came from the availability of fast computers, hardware, software and *cloud infrastructure* (Deloitte, 2025). Banks could quickly compute, process and store large amounts of data and Software as a Service (SaaS) became a core part of all infrastructure. Third, AI was able to disrupt the banking sector by assisting in *regulatory obligations*. Data collection, storage and processing aided in banks being able to quickly meet regulatory requirements, radically improving the efficiency and output of back-office functions (Deloitte, 2025). The last and fourth disruptor appeared via *inter-bank competition*. Banks leverage AI and other tools to customise solutions via harvesting large amounts of data they possess. As a result, there is a race amongst banks to adopt and implementing the last tech innovations to create solutions that have a stand-out factor and provided a value-added service internally and externally (Deloitte, 2025). Parallel to these disruptors was the cultivation of talent management and hiring of Information Technology (IT) professional which also gave rise to the gig-economy. Emerging technologies offer significant benefits but simultaneously introduce new cyber threats and entrench existing ones.

CS risk management and operations are integral to a bank's core infrastructure and support. The importance of a comprehensive strategy has never been more prudent

as companies aim to increase their digital and technological footprint. With data being the new currency, banks are not only combating highly trained, well-funded cyber criminals but cyber terrorists as well. Cyber criminals are also adopting emerging technologies to support their attacks. According to the CrowdStrike (2024) global threat report, “Electronic Crime (eCrime) continues to rise and is the most pervasive threat in 2023. Data-theft extortion also continues to rise, and 2023 saw a 76 percent increase in victims named on eCrime dedicated leak sites compared with 2022. As companies increase their use of technology, they are also increasing the number of avenues for a potential cyberattack by mature threat actors.” As banks increase their use of digital technologies, they also expand the scope for cyberattacks, especially by mature threat actors. Due to the slow implementation of legislation and regulations within the AI space, combined with the lack of morality and ethics of AI itself, it allows the layman to utilise AI platforms to engage in criminal activity. Anyone with a phone can use AI platforms to try to hack systems. According to the 2023 Future of Cybersecurity Survey (2024), completed in collaboration between McKinsey and the Institute of International Finance (IIF), wherein they interviewed banks globally, the biggest risks presented included cyberattacks, AI, talent management, third-party and supply chain management, and data security as seen in Figure 1:

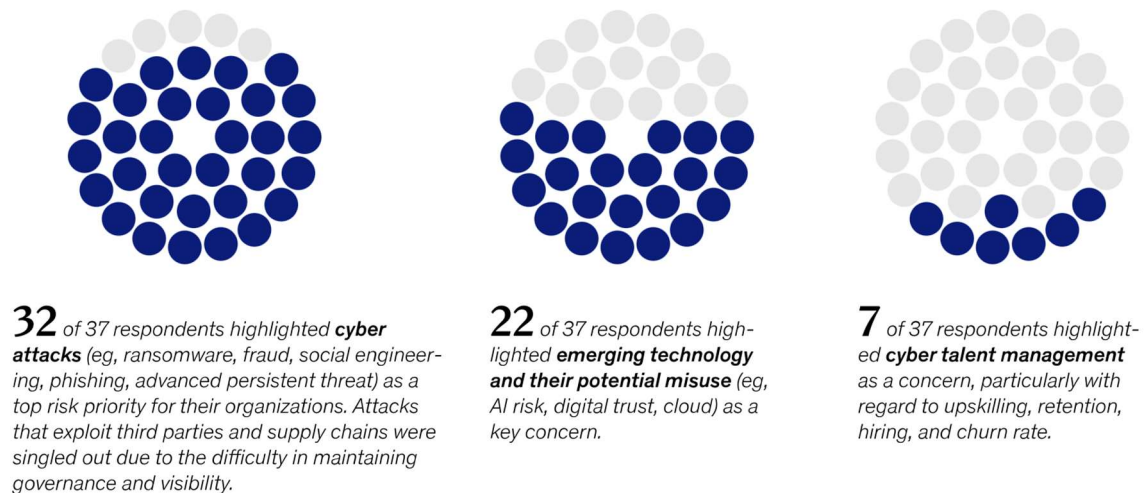


Figure 1.1: Top three cyber risks in the financial services sector

The dynamic interconnectedness of the threats is tied to the four digital disruptors mentioned earlier. What risk does big data and cloud storage present to banks' CS operational support? While banks adopt the latest talent and technological capability in creating complex firewalls to prevent breaches, banks are also reliant on third-party data, and any weaknesses in their system can result in data breaches and data theft. This is corroborated by the above-mentioned survey, wherein most survey respondents "Recognised the need to strengthen critical cybersecurity capabilities, including third-party or supply chain management and privileged access management (PAM)." The importance of cloud storage cannot be underscored, given that data is the era's currency. Any disruption of files, poisoning of data, and leakages could have massive legal ramifications for banks and financial losses for customers. SA has seen a persistent annual increase in cybersecurity attacks since 2019 (Mugwagwa, 2024). According to the South African Banking Risk Centre (SABRIC), South Africa lost more than 3 billion Rand in 2023 due to financial crime. Reports from the South African Cyber Security Hub and cybersecurity companies indicate that attacks targeting

individuals, businesses and government departments are on the increase (SABRIC, 2023).

Before banks introduce newer and newer technologies, there needs to be measures to ensure the risk management of a capability is exercised to the superlative, the ramifications of which, when not performed, result in the risks outweighing the benefits. Additional risks include corporate espionage and headhunting/talent management, which can easily lead to some banks monopolising the market. Threats may not only come from external sources but also internally and from other banks. It can also result in the prevention of centralising key functions and limit cross-pollination for the advancement of the industry. This also puts pressure on internal teams to always adopt the latest tech instead of creating vertical growth. Lastly, the speed at which regulators create new regulations to adapt to changes in the industry is far slower than the fintech evolution. Resultantly, leveraging AI will always result in banks being further ahead than regulators or managing ways to outmanoeuvre regulators. Banks are willing to pay the fines regulators impose because the outcome and financial reward far outweigh the cost of fines. This protects banks, gives more liberty to AI and causes regulators to lose credibility. Banks also hire the best talent in AI, and regulators struggle to secure the personnel needed to effectuate regulations promptly.

While CS is a critical part of business operations, as technology evolves and the ethical concerns around AI are magnified, CS operation officers are questioning the use of AI and GAI in preventing cyber-attacks. Naturally, without any CS support, the banks' entire structure and well-being will collapse, so the expulsion of AI in CS is not an option. However, it is becoming increasingly visible that AI as a solution is not

preventative and only superficial. Furthermore, CS teams are underfunded and face challenges when requesting additional funds to set up the infrastructure to limit cyber-criminal activity. Acknowledgement of underspending in capabilities has grown. According to the Future of Cyber Security Survey (2024), 70 per cent of the survey respondents believe they are underspending and should spend more. The marked shift between the 58 per cent who acknowledged underspending in 2020 and the 70 per cent in 2023 alludes to companies suffering the consequences of underspending, as depicted below.

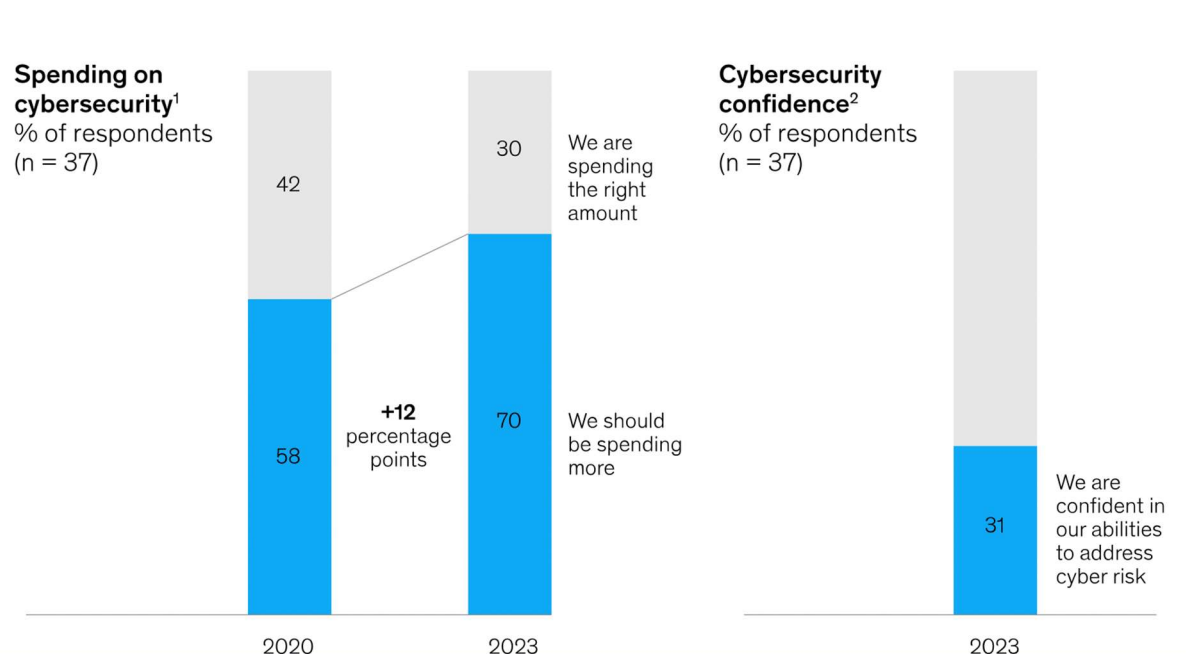


Figure 1.2: Percentage of Fin companies underinvesting on CS capabilities

Financial services institutions allocate approximately 13 percent of their IT budget to CS.

Banks are not conducting exercises to ascertain how much capital will be lost with weak, fair and strong CS protection and ratioing it to the capital gains incurred. The

rapid pace of technological evolution is stripping time away from teams to focus on quantitative metrics to assess the impact – financial, ethical; operational, talent retention and upskilling and so forth – of AI of CS functionality and operationality. As banks continue investing heavily in technologies, there is now a warranted need to consider the short, medium and long-term implications of these technologies for CS to maintain protection of their environments and to question novel capabilities outside of AI that can prevent cyber security theft and enable improved protection of data and resources (Atkins L. B., 2024).

Thus, financial institutions face various cyber threats that are both sophisticated and diverse. The volume of attacks and the increased complexity pose a significant risk. The increased number of successful cybersecurity attacks has led to increased sensitive data, critical infrastructure and financial asset compromises (Kayode-Ajala, 2024). Perpetrators of these attacks range from individual criminals to state-sponsored hackers. Threat actors employ a wide range of methods to attack financial institutions. Attacks can vary from basic phishing attacks to more sophisticated ransomware campaigns. As banks are using AI to innovate and increase their technological footprint, profitability and scalability, it is equally being adopted by cyber criminals to leverage key data points and attack resources and capabilities. These actions form causative questions: Is the AI used by the banks CS operational teams effectively combating the root cause of criminal cyber-attacks? Is a technological AI trimming needed to improve operations? What non-AI solutions and capabilities can be adopted to improve CS? This thinking form will formulate the basis of the research problem, questions, and the following justification.

1.3 Research Problem

The use of AI has the potential to solve and address many of the cybersecurity concerns and challenges which the banking sector in South Africa is currently facing. However, AI, like any other technology, is also vulnerable to exploitation and cyberattacks (De Azambuja, 2023). Knowing the new threats and risks introduced when AI-powered technologies are used remains paramount.

An analysis of the literature indicates that there is a research gap regarding the use and impact of AI in cybersecurity and South African Banks. Therefore, it is necessary to investigate the impact of AI-powered cybersecurity solutions on the number and severity of cyber threats in South African Banks year-on-year.

The purpose of this research is to explore the impact AI has on improving the operational efficiency of cybersecurity in banks. While the adoption and implementation of AI across banks is rampant, does AI actually decrease the CS team's workload and, based on machine-based learning and predictability, engage in smart learning to prevent attacks, or is it merely a tool/resource for responding to threats? The uncertainty and lack of clarity on the highly pertinent and impactful matter necessitated the investigation.

1.4 Research Questions

- To what extent can artificial intelligence serve as an effective solution for eradicating and preventing cybercrime within the South African banking sector?
- What are the factors that contribute to the decrease of the CS staff complement in banks to improve 'organisational effectiveness'?

- What evaluation criteria or assessment frameworks are employed by banks to assess the effectiveness of AI-CS capabilities in reducing cybercrime through preventative measures?

1.5 Justification for the study

The impetus and justification for the study are based upon personal experience in cybersecurity in the banking sector in SA. While AI offers radical approaches to computing, processing and storage, it also offers innumerable challenges, lacks ethical and moral considerations and legislation, and provides little to no ramifications to cyber criminals during data theft and breaches. CS teams are often left in isolation when breaches occur due to AI. The justification is predicated upon understanding both ends of the AI coin. There are three main reasons why this topic has been chosen to explore.

Firstly, AI is seen as a turnkey solution; that is, implementing AI capabilities in CS will eliminate any challenges and issues. AI does possess several advantages, especially regarding process, synthesising and collating email, endpoint, mobile, network perimeter security, for example, Distributed Denial of Service (DDoS), but most importantly, it does not eradicate threats from happening, which is the main function of CS. Secondly, machine learning is causing job losses, with only 50 per cent of a team needed compared to what was. Given the thin regulation around ethics in AI, with a smaller CS staff complement, data quality, protection and sanctity are highly comprisable compared to an in-person staff complement that can exercise ethical considerations and make decisions in line with business outcomes. From a human capital perspective, the impact on people needs to be considered in the medium-to-

long term, and not solely in the short term, as most people/organisations are currently doing.

Third, Security AI systems are SaaS, which are sitting in the Cloud. This begs the question: Are banks fully aware of the telemetry that the AI system is collecting, where it is stored and who has access to it? Do they understand the potential impact if the data collected by these AI platforms land in the wrong hands? An ongoing supply chain and third-party risk need to be considered. Using highly specialised capabilities for criminal activity will always be open and an option. AI can be used to compromise security AI systems and to poison data used by AI algorithms. AI, like any other technology is vulnerable to misuse and attacks.

The adoption of AI in CS is a trend. But once an organisation employs AI to a certain degree, there is a point of no return. This paper adopts the view that AI CS platforms do not improve a bank's cybersecurity posture as popular consensus dictates, but AI becomes the oxygen for digital firms to a point of cessation. Today, big banks have no choice but to adopt AI technologies. With an already laid-out trajectory, will AI/CS operate on a loop-based continuum *ad infinitum*?

1.6 Delimitations of the Study

The key limitation of the study emerged in the disparity between access to private quantitative data within banks' CS teams and the qualitative methodological approach adopted to ascertain results. The primary intention of this paper is to ascertain if AI has a real impact in preventing cyber-crimes, saving organisations money and developing critical infrastructure to limit the attacks against banks due to impenetrable

protections. Unfortunately, access to this data amongst all big banks is highly confidential and public knowledge or it will jeopardise their overall CS. As a mitigatory response and methodological substitute, this paper employed interview questions with key personnel in CS in the banking sector to gauge their sentiment on efficacy, impacts, and solutions.

What follows is the Literature Review, which explores a holistic view of the interplay between AI and CS in a macro- and micro-economic landscape.

1.7 Report Outline

This paper will adopt the following chronological structure to answer the research questions.

Chapter 1: Introduction

This chapter contextualises the research questions and provides a background and landscape to the operating dynamic of cybersecurity professionals. It laid the foundation for the rationale for the research topic and positioned the research questions and their concomitant justifications.

Chapter 2: Literature Review

The Literature Review offers both a topographical and in-depth view of the body of literature that exists on the subject matter. It will specifically look at the evolution of AI adoption within SA banks with the key tools, processes and people that change along the way. The literature is categorised and analysed via the Technological Organisational and Environmental (TOE) framework as well as a psychological-

qualitative approach called Perception Theory. The usage of Perception Theory is novel to this paper and the industry as traditionally only technologically oriented frameworks have been used for technological analysis.

Chapter 3: Research Methodology

This paper has adopted a qualitative phenomenological approach to analysis. Twelve participants were chosen as the sample group, and they occupy diverse and various roles in CS operations within SA banks. Via semi-structured interviews, the participant's experiences and tenure in AI-CS provide an insightful, rich and accessible view on the intricacies and sensitivity of AI-CS operations.

Chapter 4: Findings of the Study

Chapter 4 offers an in-depth view of the answers provided in the interviews. The participants were asked questions that broadly but succinctly covered the internal operations of a CS team within a big bank. The data was categorised into themes, and the views of TOE and Perception Theory were used to analyse the findings. The questions were all curated to cover the research questions and very interesting outlooks that both confirmed and challenged the narrative were discovered.

Chapter 5: Conclusions and Recommendations

The paper concludes with an overview of how the research questions were answered. It presents a sobering view on the reality of AI-CS operations and mentions gaps within academia and the technical CS teams that future research can explore. Lastly, a framework was created that showed how data travels across CS operations and the risks and benefits that teams and organisations face in the data journey.

Chapter 2: Literature Review

2.1 Introduction

This Literature Review aims to review the existing body of work on the subject matter delineated. While there is an exceptionally large body of academic and technical literature on AI itself across all sectors globally, the interplay of AI in cybersecurity operations is relatively niche. Furthermore, the determination in assessing the 'perceived impact' of AI in CS operations within banks was remarkably thin and infantile. This presented two core challenges. The first challenge was the ultra-technical nature of AI CS data. While the body of literature was thin, the journal articles and online sources offered a very in-depth view of the software being used, but it solely mentioned the solutions being used and the rationale – a digital evolutionary mapping. This information was vetted by the researcher, who has direct experience in bank CS operations. The TOE framework will be used for its holistic outlook in structuring the ideology of AI adoption in CS operations. The framework also assists in answering RQ2.

The second challenge, which was more acute, was the subjective nature of assessing the impact and adoption of AI in CS within banks in SA. There was a lack of succinct theories and frameworks to draw upon that assisted in answering the research questions. The Technology Acceptance Model (TAM) was initially explored for its acceptance usage framework. Still, it did not suffice to guide the core research questions of whether or not AI is supporting CS operations. Looking outside of technological approaches and frameworks, this paper adopted Perception Theory as the guiding framework in gauging the efficacy of AI solutions in CS operations.

Perception Theory, with its qualitative phenomenological outlook, has dramatically guided the narrative and opened the channel to inter-disciplinary pollination and provided a unique philosophical-psychological outlook completely unexplored in digital and AI approaches and frameworks. This has also informed the framework-come-theory postulated and positioned in Chapter 5, respectively (RQ3).

The structure of the Literature Review is head-first to explore the vicissitudes and advancements of AI in cyber-security, followed by a critical analysis of Perception Theory. The Literature Review will conclude with a synthesis of AI-CS and Perception Theory.

Before continuing, it is important to mention that the taxonomy of AI can be dissected across themes and applied to all sectors of the economy, from military advancements, determination of cases in the healthcare sector and so forth. The AI lexicon is vast. The term AI encompasses smartphone usage to metaheuristics, and a definition of the term itself is evasive despite standardisation attempts by governments, industry leaders and digital organisations. Inspired by Kelly, Kaye and Oviedo-Trespalacios (2023), the following definition of AI will be used in this paper: “An unnatural object or entity that possesses the ability and capacity to meet or exceed the requirements of the task it is assigned when considering cultural and demographic circumstances.” (Kelly, 2023, p. 2). There is an exceptionally large body of literature available on AI itself, should the reader wish to explore, but given the vastness of AI, for the purposes of this paper, AI will be contextualised in the cybersecurity operations within South African banks and will not be explored any further.

2.2 Defining the TOE Framework

TOE examines the adoption of technology from the viewpoint of an organisation. It outlines the various aspects that impact the adoption of technology and the probability of its implementation. The theory suggests that the technology adoption processes of an organisation are influenced by its technological, organisational and environmental settings (Hasani, 2023).

The technological context encompasses the internal and external technology pertinent to the company. Technologies encompass both physical equipment and procedural methods. The organisational context refers to the attributes and assets of the company, such as its dimensions, level of centralisation, level of formalisation, managerial framework, workforce, quantity of surplus resources, and connections among employees. The environmental context encompasses the dimensions and composition of the industry, the firm's rivals, the overall economic conditions, and the governing framework. The TOE framework offers a comprehensive viewpoint on technology adoption, adaptability and application (Baker, 2011).

The TOE framework is relevant to this study from a technology perspective because AI, as a technology, can help transform CS operations and help assess the readiness and compatibility of AI-CS solutions within local banks in SA. From an organisational level, the TOE framework evaluates the factors that influence the adoption and impact of AI on banks, such as resource constraints, skills shortages and legacy systems, to name a few. South African banks are highly regulated and operate in an environment with a growing need for financial inclusion. From an environmental perspective, the TOE framework considers external pressures such as regulators, legislations and

policies, customer expectations and inter-bank competitiveness. All three features will be discussed chronologically.

2.3 Technological component of TOE affecting CS operations

AI has the potential to transform cybersecurity and improve its efficiency and effectiveness. AI-based security software enables faster detection and response to CS threats. It automates repetitive tasks, freeing security professionals to concentrate on more intricate security-related tasks (Sontan, 2024). AI systems can filter through massive amounts of data and identify patterns that point to possible threats. The adoption and use of AI-powered systems in CS is expected to significantly increase globally. Markets and Markets predicts that the global market for AI in cybersecurity will increase from \$8.8 billion in 2019 to \$38.2 billion in 2026 (a 23.3 per cent increase) (Ibrahim, 2023). Organisations use AI to improve and strengthen their cyber defence systems against cyberattacks. AI benefits businesses that use it in their cyber defence operations. The advantages include automation, threat intelligence, and enhanced CS. AI-based CS tools are increasingly important for identifying and averting cyberattacks (Ibrahim, 2023).

The growing use of online services has seen an increase in the frequency and sophistication of cyberattacks, resulting in an obsolescence of traditional CS solutions (Kaur, 2023). Networks and system complexity are increasing, and AI-powered cybersecurity solutions have become essential. The adoption and use of AI tools provide considerable advantages to organisations in assisting them in protecting their networks from cyberattacks (Jada, 2024).

2.3.1 The necessity of Cybersecurity

In today's rapidly advancing digital age, the CS landscape confronts unprecedented cyber threats. Traditional security measures, such as firewalls and intrusion detection systems, have fallen short in their advancement as they require manual processes and systems drastically slower than the automation and speed of data processing offered by AI software and solutions. The COVID-19 pandemic exacerbated the adoption and scaling of AI as businesses of all sizes shifted to an online presence and banks experienced a surge in activity to cater to customer needs and improve competitiveness. Simultaneously, this led to a surge in cybercrimes, which is predicted to reach \$10.5 trillion in 2025 (Eian, 2020). To mitigate these losses, there is a poignant drive to improve AI-related resources and tools to improve banks' cyber defence. AI mimics and simulates human behaviour, resulting in the speed of automation beyond manual human capability and with profound benefits (Xu, 2021). The evolution of AI from Machine-Learning (ML) to Deep-Learning (DL) has minimised human error and effort. It has become an invaluable tool to the myriad forms of cyber-attacks banks face daily.

How society functions has also rapidly changed. Customers would traditionally only go to a brick-and-mortar bank or telephonically discuss certain mandated matters, but today, all forms of banking can be performed online. Resultantly, cyber-attacks disrupt and target computer systems, networks and data, and stringent security measures exist along these triadic platforms (Salem, 2024). Tabulated below is a non-exhaustive list of cyberattacks:

Type of Attack	Description
Phishing	Tricking individuals into divulging sensitive information such as login credentials and financial data by masquerading as trustworthy electronic communication (Forbes, 2024).
Malware	Malware/ 'malicious software' includes viruses, worms, trojans, and ransomware that are designed to damage or unauthorised access to computer systems and data (Forbes, 2024).
DDoS	Denial of Service (DoS)/Distributed Denial of Service (DDoS) attacks overwhelm a system's resources, making it unable to respond to legitimate service requests. DDoS attacks use multiple compromised computer systems as sources of attacks, exacerbating the scale and impact of assault (Forbes, 2024).
MitM	Man-in-the-Middle (Mitm) attacks that are invisible, intercept communication between two parties, resulting in theft, alterations or fabrication of messages, leading to data breaches and eavesdropping (Forbes, 2024).
SQL Injection	Exploits vulnerabilities in a database-driven website by injecting malicious SQL statements, resulting in access to sensitive data that can be tampered with and manipulated (Forbes, 2024).
Zero-Day Exploit	On the 'zeroth' day, a vendor becomes aware of a vulnerability. The attack exploits vulnerabilities before they can be addressed, exacerbating the ability to defend (Forbes, 2024).
Ransomware	A form of malware that encrypts a victim's files, making them inaccessible and holding the victim hostage until they pay for the decryption key (Forbes, 2024).
XSS	Cross-site Scripting (XSS) attacks third-party web resources to run scripts in the victim's web browser or app (Forbes, 2024).

APT	Advanced Persistent Threats (APT) occur when an individual or group acquires unauthorised access to a network for an unnoticed period of time, resulting in the exportation of data (Forbes, 2024).
BEC	Business Email Compromise (BEC) attacks target employees with financial authority to trick them into sending money into the attacker's account (Forbes, 2024).
Crypto-jacking	Confidentially uses a victim's computing resources to mine cryptocurrency.
Password Attacks	Attempts to crack a user's password and phishing for passwords. Hacking tools include Brute-Force, Rainbow Table, Credential Stuffing, Password Spraying and Keylogger attacks (Forbes, 2024).
Insider Threat	Individuals within an organisation who misuse the company's systems and data.
Botnet Attack	A network of compromised computers controlled remotely by an attacker to coordinate malicious attacks (Forbes, 2024).

Table 2.1: Types of Cyber-Attacks (non-exhaustive)

“Intending to cause damage, unauthorized access, or service interruptions that cause severe data loss or financial damage and often lead to long-lasting consequences, these are the insider threats that represent a significant and growing segment of these attacks, usually committed by disgruntled or rogue employees who exploit their authorized access to steal data or cause harm. These threats can also emerge from intrusive applications that users accidentally install on their devices, allowing these apps to access and misuse sensitive information. Advanced behavioural anomaly detection and auto-resiliency mechanisms are being developed to combat these

threats by proactively identifying and mitigating malicious actions at both the employee and application levels.” (Salem, et al, 2024, p.4).

Irrespective of the nature of the attacks, AI can learn from data, compute, process and defend against threats and attacks, enabling the protection of infrastructure, software and data. The benefits of integrating AI with CS include improved decision-making due to concise quantitative data being easily accessible, enhanced detection of network intrusions, overall management of cyberattacks, and its impact on systems and software. This evolution, which allows for real-time detection and response, has significantly reduced the ‘false positives’ rate, which is common in traditional models of CS. ML and DL have enabled predictive analysis to be made of datasets and to identify potential vulnerabilities before they are exploited. This has enabled a preventative, proactive approach, as opposed to a reactive one (Welukar, 2021). For example, phishing attacks have become increasingly sophisticated, and traditional methods of detection have been unable to identify and prevent them effectively. Conventional methods relied on blacklists, heuristics, and content analysis but attackers are now using AI and social engineering resulting in convincing phishing emails that easily evade traditional detection methods (Eze C. S., 2024).

The three main types of AI in CS that will be explored are ML, DL and Metaheuristic Algorithms (MA).

2.3.2 AI CS Anthropomorphism

AI integrates vast quantities of data with fast, iterative processing and sophisticated algorithms. This allows the software to acquire knowledge automatically by analysing

patterns or characteristics in the data. AI is a discipline within computer science that spans several domains, such as ML, DL, data analytics, languages, and software engineering (Shrivastava, 2024). These fields frequently entail the creation of AI algorithms that mimic the decision-making mechanisms of the human brain. These algorithms can learn and remember information from available data, resulting in more accurate categorisation (Shaveta, 2023).

Vulnerability management and remediation of vulnerabilities in operating systems and applications have become increasingly difficult, with organisations struggling to manage the ever-increasing number of vulnerabilities due to traditional models (Kumar, 2023). AI assists banks in better managing and remediating vulnerabilities across their networks. AI-powered vulnerability management systems can scan networks proactively for potential vulnerabilities and can detect and classify network vulnerabilities (Rizvi, 2023). AI-driven incident response platforms utilise sophisticated algorithms, such as ML and natural language processing, to examine security alerts, rank incidents based on importance and carry out predetermined reaction steps independently. These platforms can process high numbers of real-time security alerts and make data-driven decisions to promptly and efficiently resolve security events (Sontan, 2024). Using AI-powered incident response systems reduces human error, and the speed at which incidents are responded to is improved. The organisation's overall security posture is improved as AI-powered incident response systems are more flexible and scalable. AI-powered incident response systems can also assist with anticipating and preventing future attacks and improve situational awareness and threat visibility (Rizvi, 2023). The first shift from traditional AI-related vulnerability management and remediation was to Machine Learning.

2.3.2.1 Machine learning

ML empowers computers to solve and interpret issues without extensive programming. The learning methodology analyses past data, consolidates various algorithms, identifies them and demarcates them into categories/typologies, and performs tasks based on the desired outcome (Kuntla, 2021). The algorithms are tri-pronged (see Figure 2.1 below), and cluster data to defend against the attacks mentioned in Table 2.1 above:

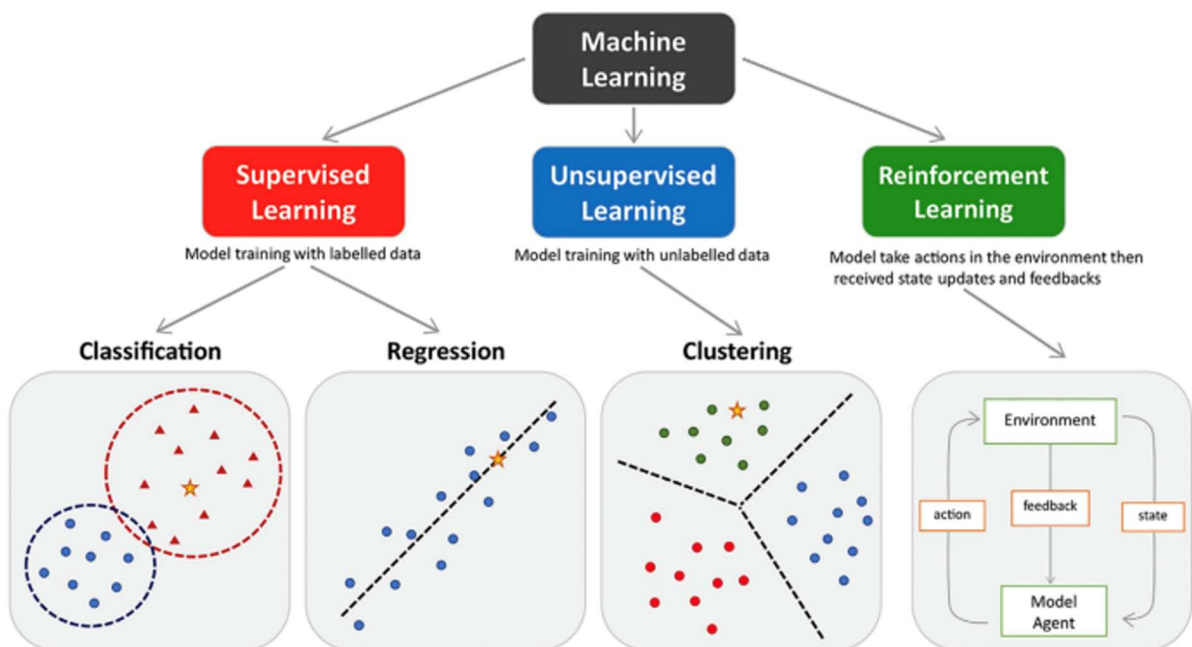


Figure 2.1: Types of Machine Learning

ML architectures are broadly scoped and focus on identifying false data to censor outcome manipulation. They detect network anomalies and internally function like an immune system by defending against alien or foreign data and software.

While ML was promising for its duration as the forerunner of AI in CS, its reliance on manual feature extraction lay too close to traditional modelling, which limited the

efficiency and accuracy of threat detection and became a limitation within CS. ML's functionality is significantly dependent on the precision of feature extraction and recognition (Eswaran, 2023). This is extremely important as biometrics has become the access format of customers within banks.

2.3.2.2 Deep Learning

Disruptions and growth of technologies within AI are almost always a result of shifts in the global/macro-operating environment. Akin to CS responding to threats in a reactive way, the development of AI has produced novel solutions reactively. The IoT, the explosion of network connectivity, positioned CS at the forefront of operations due to the large amounts of data that needed to be stored on the cloud. Malware and third-party cloud computing and data storage suppliers became the new target for cyber-attackers. Attackers would look at networks and malware points of vulnerability, and where ML had reached its zenith in protectionism. Herewith, DL presented an anthropomorphic response to the challenges, CS methodologies, and operations as a whole (Morovat, 2020).

DL grew from some of the challenges of ML. It revolutionised the AI landscape by creating "Advanced neural networks that can process large amounts of data and learn from experience, mimicking human brain functions to recognize complex patterns" (Barik et al, 2022). It mimicked images, voices and behaviours and open pathways for changes in robotics, speech and facial recognition. In the sphere of CS, DL plays a critical role in detecting intrusions and monitoring for malware (Musa, 2024). It was propelled by advancements in preventing APT attacks by recognising different scales of attacks used in the most evasive tactics (Salem et al, 2024). As presented in Figure

2.2 below, the neural network has several layers, all highly complex and hidden in nature. These barriers prevent the breach of attacks, help systematise and categorise the attack type and configure and learn based on the inputs to adapt to the rampant and diverse types of attacks across all three malware platforms, network and data.

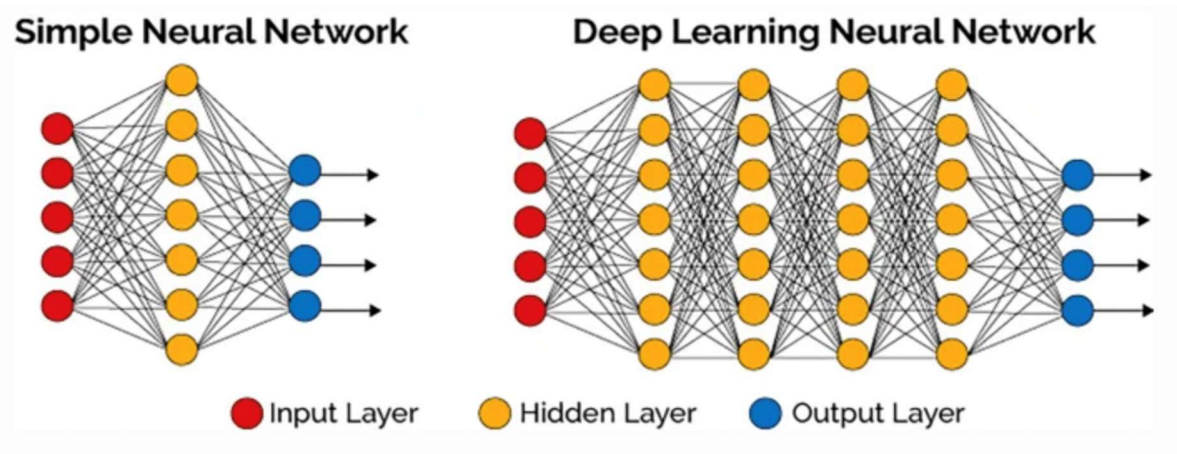


Figure 2.1: Difference between Simple and Deep Learning Neural Networks

Naturally, given that customers are providing access to their biometrics, there are stringent requirements for data protection, as loss of data could result in a customer's breach of privacy and hackers and cyber thieves using the data to gain access to customers personal information resulting in identity theft, citizenship mimicking, bank account information and social media accesses. A significant and tight regulation is needed on this account. The necessity of CS was never more prudent. As the bank's CS teams improve their solutions and offerings, cyber thieves also use the latest AI to engage in criminal activity, creating a hostile environment, but also being the impetus for ongoing progress and development. The next type of AI comes in the form of MA.

2.3.2.3 Metaheuristic Algorithms

MA addresses complicated problems that cannot be solved using standard processes.

MA are optimisation strategies inspired by animals' problem-solving activities and abilities in their natural habitats. The algorithms have shown remarkable optimisation-related success when applied. MA is delineated into four categories, namely evolution-based, swarm intelligence-based, physics-based, and human-related algorithms as shown below:

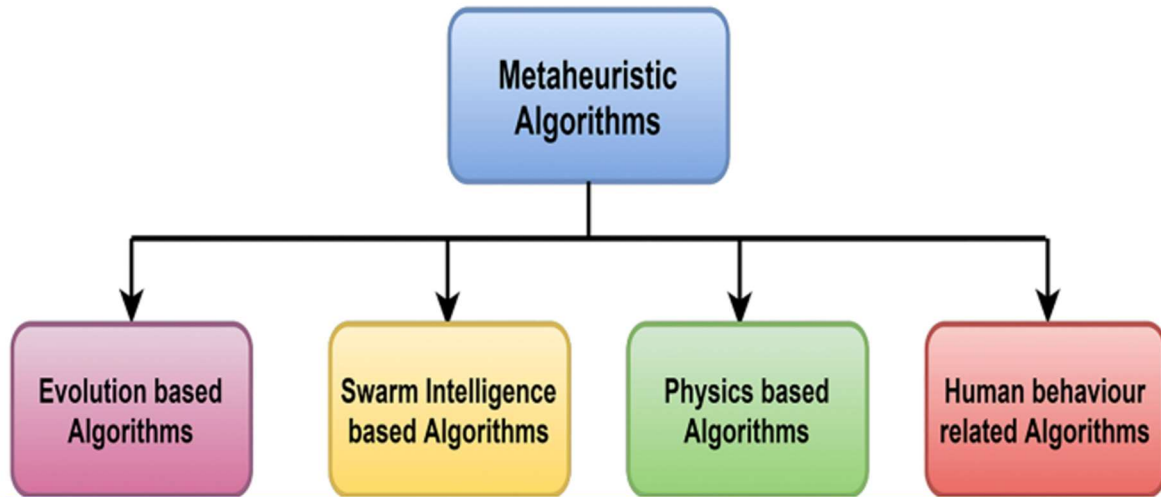


Figure 2.2: Four Types of Metaheuristic Algorithms in CS operations

MA is vital in improving the efficiency and accuracy of various detection learning by enhancing the learning as they expand the search space explored during model training, potentially uncovering superior solutions that traditional methods might miss. This is particularly beneficial in CS, where the landscape and attack patterns change rapidly. The advantages of MA in cyber-attack detection and defence include, according to Nordin et al (2021) in *A comparative analysis of metaheuristic algorithms in fuzzy modelling for phishing attack detection*, an improvement in optimisation, automation and speed.

- Optimisation: MA's are better at finding optimal solutions to complex problems that are otherwise too challenging for conventional methods;

- Automation: By automating the tuning of detection parameters, the algorithms minimise the need for human intervention, making the detection process fast and reliable;
- Speed: MA achieves faster convergence to effective solutions, essential in time-sensitive cybersecurity environments where threats must be quickly identified and mitigated.

Collectively, these AI-based technologies address the challenges in CS more effectively than traditional methods. They do so by automating the detection and response processes and enhancing the speed and accuracy of threat detection. Understanding the mechanisms and operations behind these AI models offers more profound insights into their application, highlighting their critical role in developing more resilient cybersecurity systems. MA supports the robustness and adaptability of cyber-attack detection systems, making them more effective against various cyber threats. They ensure that detection systems are accurate and remain relevant as attack methods evolve.

Interestingly, the results of AI were experienced similarly in Qatar's banking sector. In a study conducted by Al-Dosari et al (2022) it was revealed that AI is a crucial technology for enhancing the cybersecurity measures and posture of banks in Qatar. In another study done by Ibrahim and Hassan (2023) it was found that established businesses were using AI to boost their defences against new emerging threats. AI-powered technologies is rapidly changing cybersecurity, making it more effective and efficient. AI can utilize predictive modelling to proactively prevent threats, identify threats in advance and proactively prevent them by analysing previous attacks and by

identifying patterns. AI can quickly assess security risks and act on potential threats in real time (Rizvi, 2023). AI is now used for real-time threat detection and prevention, as it can continuously monitor network traffic and identify suspicious user behaviour through AI-powered behaviour and anomaly detection systems (Sontan, 2024). AI CS technologies can help enhance threat detection, vulnerability analysis, and incident response.

2.4 Organisational component of TOE affecting CS operations

Every organisation, including banks, has certain goals and objectives to achieve year-on-year. In SA alone, the top-tier banks possess a healthy level of competition, which promotes the advancement of AI and digital tools and solutions. The banks have become the entry point and testing vehicle of anthropomorphic AI deployment. The influence of AI cuts across all sections of the bank, from the increase in productivity in legal departments due to AI storage, processing and data tools, to brick-and-mortar stores increasing the number of digital features in the stores to improve customer service. AI has infiltrated all parts of the bank as an ideology and way of operating. In this context, four points will be analysed: the interconnectivity of AI, third-party vendors, CS security personnel and the gig-economy.

AI as a tool has a scalable feature. Depending on the nature of work to be completed, anything from data processing and storage can be performed, up to highly complex malware and Central Processing Units (CPUS). The scalability and interconnectivity of AI allow for a centralisation functionality that accords CS officers oversight over data, network, and malware intra-organisational protection (Ahmed et al., 2022). “Network protection using AI-driven network analysis systems (NAS) and network

protection systems (NPS) can guarantee the safety and availability of computer networks within an organisation, not just for a single computer but for an entire computer network system simultaneously.” (Xu, 2021). AI solutions can be adopted at any juncture within the life cycle, allowing for a complete interconnected solution. Each feature exists as a totalised, individual whole, while belonging to a bigger totalised centralised system (Raimundo, 2021). The usage of third-party vendors has necessitated the need for greater protection. While internal threats can be detected and managed, external threats from vendors has the ability of crippling the system. CS defence and protection must also extend beyond the parameters of the banks, including the vendors.

The use of AI has an impact on the people supporting cybersecurity platforms. Due to its rapid growth, AI technology has created gaps by displacing the market’s knowledge and understanding of operational procedures. To stay employable and competitive in the field of CS, professionals need to learn about AI and gain expertise on AI-powered platforms continuously. Having the right people to implement and support AI-powered cybersecurity platforms in an organisation is critical. Additionally, and sub-tangentially, employee-training programs are essential, focusing on the significance of cybersecurity knowledge to deter phishing attempts and other forms of social engineering approaches (Andriu, 2023). In a study conducted by AL-Dosari et al. (2022) and Ibrahim and Hassan (2023) it was found that there is a lack of skilled employees who can deploy, and support AI-powered cybersecurity platforms. The gig-economy, the notion of a global citizen and the outsourcing of IT specialists were common practices between 2010 and 2020; the COVID-19 pandemic magnified the practice. While retaining key talent and finding the right people in person may be

challenging, remote working has allowed banks to rely on personnel in other countries for the experience and expertise in managing the systems and processes.

Lastly, there is the looming concern of relevancy and machines replacing humans. Time-consuming and manual tasks are now being performed by AI, displacing human labour. There is an ongoing debate and thoughts of regulation to prevent the wholesale machine-replacing-man narrative, the results of which have massive macro socio-economic consequences.

2.5 Environmental component of TOE affecting CS operations

AI collects and processes large volumes of data, and personal data usage in AI models is growing, which presents significant challenges to sensitive data protection. The data protection used by AI has become increasingly concerning (Sontan, 2024). Protecting this data is critical, as misuse or exposure can impact privacy and individuals' rights. There is a mandatory need to protect AI privacy as a fundamental right. Organisations must address the associated privacy challenges and take steps to protect personal information. AI regulatory and ethical frameworks can guide to ensure the ethical and responsible use of AI (Ferrara, 2023). However, on par with international standards, there is no regulation on AI. Presently, there are only four Acts that enforce and protect customers' data in SA - the Protection of Personal Information Act (POPIA), the Copyright Act, the Patents Act, and the Competition Act.

After coming to power in 2019, President Cyril Ramaphosa initiated commissions and steps to help regulate AI in the fintech industry, but no formal enforcement has been issued to date. Given the absence of standardized regulations for AI usage across

countries, prospective research can delve into diverse legal frameworks that effectively oversee the application of AI in cybersecurity within different nations. These studies can potentially mitigate the risk of malicious AI usage in organisations by contributing to the establishment of more comprehensive and effective regulatory measures.

2.6 Challenges to AI CS operations

The challenges of AI in CS will follow the same format as TOE.

2.6.1 Technological Challenges

Human error is the main factor contributing to CS breaches (Amoresano, 2023). Humans are considered the most vulnerable point in the cybersecurity system. Whether the human error is a decision-based error or an error in performing the task, whether that fault is intentional or not, or skill-related, the initial stage in establishing a highly secure cyber environment is by eliminating the possibility for human error. Automating CS tasks by leveraging AI minimises the requirement for human involvement and interaction, reducing potential human mistakes throughout the security lifecycle. What sometimes takes hours to do, AI can do effectively and efficiently in seconds (Jada, 2024).

The use of AI still faces challenges, as there are harmful activities from telemetry that AI does not identify as threats. Adversarial actions targeting ML can manipulate ML training samples to influence the ML model's accuracy (De Azambuja, 2023). These attacks aim to disrupt the sample data's quality, undermining the model's trustworthiness. The goal is to generate malicious behaviour that appears legitimate.

AI's potential replacement of security analyst roles, normally responsible for detecting and responding to these events, is hindering its effectiveness, leading to uncertainty in the risk mitigation strategies of financial institutions like banks. Furthermore, the algorithms used by AI tools may raise concerns about their testing, potentially leading to bias and some threats not being detected (Ibrahim, 2023).

Vulnerabilities and risks associated with AI platform implementations have grown in concern. The manipulation of training data, the initial design of the system, the lack of robustness in AI algorithms, and the abuse of implementation weaknesses all lead to vulnerabilities in AI (Villegas-Ch, 2023). These risks lead to adversarial attacks, unwanted biases, information leaks, and AI model manipulations. Implementing a robust security framework is necessary to mitigate these risks (Villegas-Ch, 2023). Hackers can exploit AI-based protection systems and evade the AI from detecting them. AI-driven attacks use neural fuzzing, allowing them to progress more rapidly than AI-powered protective tools. Neural fuzzing employs neural networks to detect vulnerabilities in AI-powered systems, which enables threat actors to acquire knowledge from existing AI-powered systems (Jada, 2024).

Explainable AI (XAI) - knowing why and how algorithms make their decisions, is critical in CS, and current AI algorithms lack that transparency, which erodes and questions their trustworthiness (Tiwari, 2023). The matter is sometimes exacerbated by the intricate neural networks and DL algorithms, as they function in ways that are not easily comprehensible to humans. XAI allows security analysts to comprehend the reasoning behind AI-generated alerts, suggestions, or choices by revealing the inner workings of AI algorithms (Sontan, 2024). Considerable progress has been achieved in the domain

of XAI, as researchers have devised strategies and procedures to augment the transparency of AI systems (Tiwari, 2023).

The deployment of AI in cybersecurity is challenging. According to Aggarwal (2023), AI systems require extensive data to function effectively, and processing such volumes can be resource intensive. The risk of false alarms can undermine user trust in AI systems, and delayed responses to threats may compromise system effectiveness. AI models trained on biased or unrepresentative datasets have the potential to acquire and magnify existing biases, resulting in outcomes that are unjust or discriminatory. (Ferrara, 2023). Training data biases can arise from historical inequities, cultural prejudices, or sample biases, leading to biased predictions or choices by AI systems that disproportionately affect specific groups or individuals. It is crucial to prioritise establishing impartial data and model development to advance equitable and ethical AI. Methods such as dataset augmentation, bias-aware algorithms, and user feedback can be used to address the issue of bias in AI (Ferrara, 2023).

There is no one-size-fits-all solution like traditional firewalls, anti-virus or anti-malware software (Cucu, 2019). Each AI solution must be tailored to a specific organisation using organisational-specific internal and external data (Raimundo, 2021). This requires greater financial and manpower commitment, and greater hardware and infrastructure commitment to implement. Whilst traditional network security tools such as firewalls and anti-virus software are universal, they are seen today as border-based protection (Xu, 2021). AI-driven NASs have substituted these traditional network protection mechanisms as they are faster and more efficient (Aliyari, 2021).

2.6.2 Organisational Challenges

The future of AI evolution involves tackling emerging threats, scalability, deployment challenges, and organisational readiness, including necessary training for cybersecurity professionals.

One of the main obstacles to multi-lateral and multi-vertical adoption of AI is infrastructure and hardware requirements, as significant CPU and manpower are needed, skills which are currently not present in the industry (Dawson, 2021). Implementing AI requires a range of specialised professionals, such as data scientists, data analysts, AI experts, machine learning specialists, developers, cybersecurity specialists, and project managers, each with different levels of technical expertise. (Attaran, 2018). This bars the high costs required for once-off implementation and maintenance (Wilkins, 2018). Without the manpower and budget, pilot testing and implementation could take years across an organisation. This excludes the starting point of systems, some of which are legacy and require a complete technological overhaul.

Another challenge lies in the compatibility issues caused by the continued use of outdated systems, programming languages, and overall technological infrastructure in many organisations. Legacy systems fails to adequately support the requirements of AI and ML technologies. For instance, the analysis of vast amounts of complex data, a critical step in successful AI and ML deployment, is hindered by the lack of scalability offered by legacy databases and obsolete systems. Implementing AI solutions in organisations is not a simple task, as it often necessitates a complete overhaul of the technological infrastructure (Arasada, 2021).

The literature consistently highlights a recurring theme of insufficient high-quality, error-free, clean data availability. AI solutions rely on extensive datasets for training models and achieving accurate results. As a result, obtaining a large quantity of data is essential for training AI models effectively (Sun, 2022). Implementing a cyber-AI solution necessitates a more complex organisational data management process due to the diverse volumes and types of data stored, the speed at which data is accumulated, the need to maintain data confidentiality, and the constant requirement for additional data (Raimundo, 2021). This aspect is particularly crucial because the intelligence of AI solutions relies solely on the quality of the datasets used to train the models.

2.6.3 Environmental Challenges

There are no global, national, or industry-specific regulations to manage the enforcement of AI, let alone CS operations. The ethical and moral challenges lie with individuals and banks themselves in controlling data. The lack of regulation enables insider breaches and attacks for financial gain.

2.7 Defining Perception Theory

Andrew Demuth in *Perception Theories* (2013) Analyses the consequences of higher cognitive processes and its link to perception. What is perception? Where is the origin point of perception? What informs what we perceive and how we process the information we perceive? Perception Theory walks the bridge between general and cognitive psychology and philosophical epistemology and prompts one to explore the episteme and ontology of perception. The notion of perception is important to this paper because the core question is to understand the perception of AI in CS operations. As witnessed in the literature above, AI is anthropomorphic, it has a

scalable feature and operates through a centralised system that can easily proliferate the essence, DNA and fabric of an organisation. It also has the propensity to alter the ethos of a department and organisation. The perception of the effect AI has on the technological resources, on the intra and inter-organisational dynamics for the promotion of growth and well-being, as well as paying heed to regulations to protect the integrity of data conferred by customers, is crucial. While AI is in vogue and at the forefront of the digital revolution, the underside of the belly is dark and dangerous. Thus, the perception of AI needs to be an ongoing question to purview the landscape and determine its success or lack thereof. Are the challenges of AI to CS outweighing the benefits, and is there a non-AI solution? Are banks getting carried away by the current of AI? What is the social impact due to a mammoth disparity between AI-user organisations and non-AI user organisations? Is the continuous implementation of new AI features necessary? Does AI truly support CS operations or make operations more challenging to manage? Perception Theory will provide a frame of reference for exploring the mapping of how to answer these cutting-edge questions.

The weightiest question in Perception studies is 'What is the source of our cognition?' The perspective is bifurcated into internalists, who propound a top-down approach and externalists, who espouse a bottom-up approach (Demuth, 2013). The former school and body of thought proclaim that knowledge, its sources and principles can be found within the subject itself, and cognition is but a mere discovery of *a priori* pieces of knowledge. The internalists are championed by Descartes and Plato, who believe that knowledge is about the collection of already acquired contents. All knowledge is pre-existing and present; our perception and cognition bring realisation to what is (for example, the source of MA is nature). The second school looks at cognition through

an external sense, that being – experience. The externalists asserts that the mind is a blank sheet of paper (*tabula rasa*), and an external reality prints all knowledge. The externalists adopt a phenomenological stance in that they view cognition as a composition of social constructs created by personal, familial, cultural and socio-economic conditions, and the on-going overlaying of experiences constantly alters one's perception of their environment and world (Demuth, 2013, p. 15). Perception is the rationalisation of one's personal experiences.

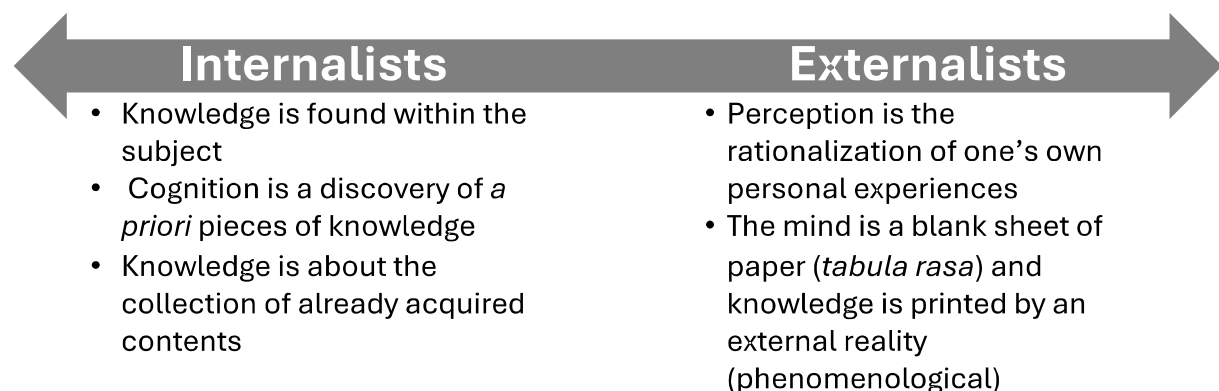


Figure 2.3: Difference between Internalists and Externalists

A question confronting authors and supporters of both schools asks the next question of how one can explain the awareness of a new reality in our consciousness. How does consciousness/the mind meet the world? The point that both schools agree on is the trustworthiness. The trustworthiness of a statement is maintained by the trustworthiness of the source/authority that postulates it (akin to a non-manipulative subject-matter expert). As Descartes articulated, 'we must verify the knowledge principles themselves.' The verification of testimonies takes three forms (Demuth, 2013):

1. I verify data myself, but it may not be possible when one needs to test the data using an instrument. For example, using a microscope as opposed to optical vision can alter or skew information
2. Testing the testimonies by mutual confrontation – Populus data testing on a large target sample. If most people share the same sentiment, it must be correct. However, the majority is not always correct
3. Critically reassess the reliability and competence of each witness and their testimonies—In the judicial system, the verdict of the witnesses is important based on the background and sanctity of their truthfulness. This is tested by assessing how a sample witness arrives at his own testimony by verifying what he has already seen (his perception). Point three satisfies the paradox of one and two.

Thus the synthesis of the internalist view (all knowledge is already pre-existing it is only rationalisation and awareness to it that is required) and the externalists view (one's perception is based on personal experience and to each individual that will differ based on the matrix of social constructs he operates from) is a collective acknowledgement from a large body of people who all have experience on a subject matter and can verify data based on their own internal and external sense of perception.

Now that the horizontal paradigm has been established, the vertical paradigm will examine the top-down and bottom-up approaches to perception.

2.7.2 Cognition, Science and Perception Flow

Philosophers were the first to explore perception. As centuries passed and the Scientific Revolution took hold, there was a need to methodologically test every idea and perception to assess if it fell within reality. Philosophy provided the fodder for the inquisitive nature of science. "Philosophy enriched science not only via concepts and ideas, but also through distinctive methodology and ways of thinking. As good examples, we can mention the introspective method, the phenomenological reduction, and the description of phenomenological experience." (Demuth, 2013, p. 20).

According to the direction of data acquisition and processing, information can be positioned into a 'top-up' or 'bottom-down' approach. On the one hand, the bottom-up approach (herewith the neurological-cognitive stance will not be looked at) views data from a distance first and, upon closer inspection and engagement, can see and experience the complexity of processes. On the other hand, the vantage point for the top-down approach (again, the neurological-cognitive stance will not be looked at) is feeling. There is an inner sight and inner feeling about a subject matter that probes the exploration into determining action. It is top-down because the feeling appears as a thought in the mind and trickles into bodily action via the exploratory process. The essence of the top-down approach is tied to the view that one needs to have knowledge and experience in a field to help organise the cognitive contents (Demuth, 2013, p. 23).

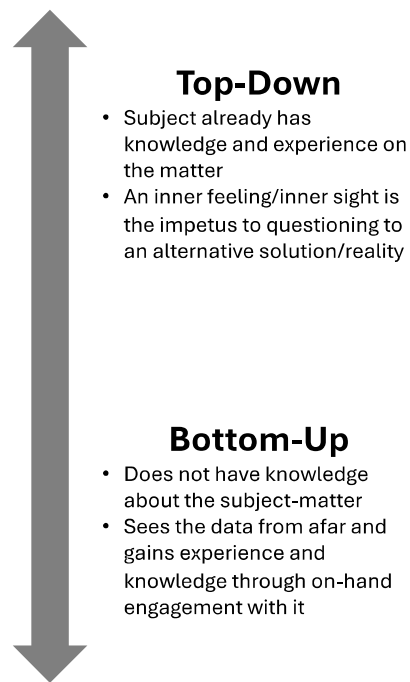


Figure 2.4: Difference between PT's Top-down Bottom-up Approach

The internalist-externalist view and top-down bottom-up perspectives are significantly richer and detailed. While this paper wishes to explore the exactities of the ethos on a grander scale due to constraints, the door will be left open for future explorations and determinations. In what follows is the adaptation of Perception Studies to AI and CS.

2.8 AI-CS and PT – A Synthesis

The author's personal experience will act as the basis of the investigative eye in this synthesis/processing juncture, with its zenith being explored in semi-structured interviews and presented in the findings (Chapter 4).

From the outset, there is an already systematic placement of the subject on both the internalist and externalist spectrum. Regarding the former, the author/researcher/subject has *a priori* experience in CS operations in the banking

sector and has first-hand experience in witnessing the evolution of AI tools and software from traditional firewall and malware tools to AAI and GenAI in the CS realm. The questioning of efficacy emerges from the ongoing exposure to the dynamics and effects of AI in CS. Positioned on the *tabula rasa* part of the externalists' view, the experience and almost AI-saturation in CS operations warranted the questioning of 'what else' and 'what next'. Apart from the anthropomorphic nature of AI, which is in the infant stage of its revolution, what other technological pursuits can be created, gauged and utilised to prevent an AI burn-out that has deep-seated ramifications for organisations and society as a whole given the lack of regulation in the industry? However, there is a leaning towards the externalists' position given the basis of perception questioning that emerged from tenure and experience in the industry.

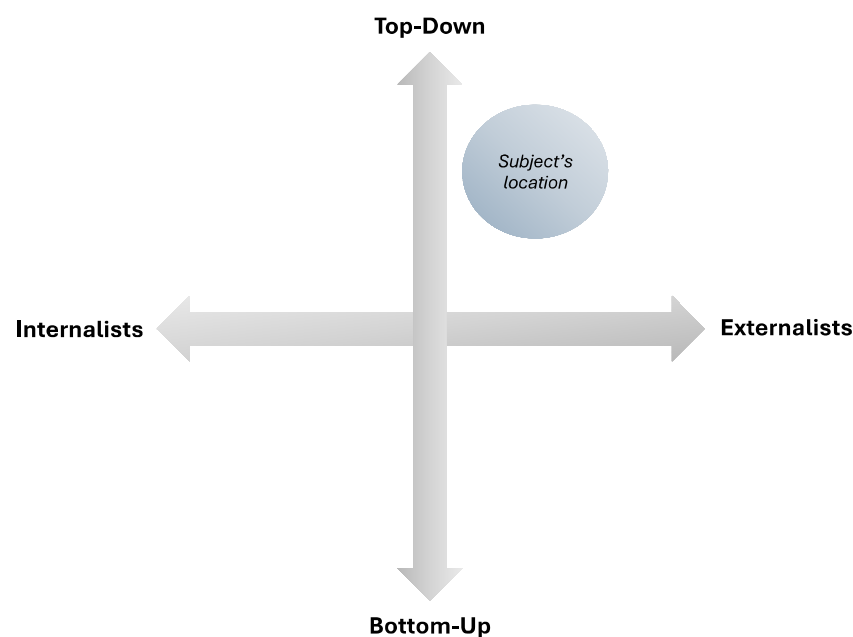


Figure 2.5: Locating the subject on Perception Theory's Quadrants

Perception according to the horizontal paradigm is that experience within AI is a critical necessity when deciding what features to implement or not implement. The resources – personnel, infrastructure, and financial - are all needed to create a supporting

environment for AI implementation, without which the system has a way of failing the organisation. This is specific to CS operations within SA banks and broadly applicable.

Across the vertical axis of bottom-up and top-down, the author/researcher/subject aligned predominantly towards the top-down approach as the genesis of the thought-piece is the semi-mal-adaptation of AI in CS due to organisational pressures. Interestingly, the proximity to aligning with the top-down approach emerged from the bottom-up perspective. An analogy will be used: in the bottom-up perspective, a subject views an object from a distance, and with closer and closer engagement, the nuances and complexities can be seen and experienced (Demuth, 2013). The example used will be that of an apple. An apple, based on the outside covering, can be understood based on form, colour and size. Its texture can be gauged through touch. However, to the onlooker, the contents of the inside of the apple are unknown. The subject can bite into the apple and taste it, a step closer to understanding the apple – the taste, the texture, the juice, and so forth- but still, the contents are unknown to him. Upon the slicing the apple in half, the subject sees the core and the penultimation in the experience is the realisation that the apple's seed contains many apple trees. The researcher's engagement with AI and CS affords him the trustworthiness, credibility and validity.

Secondly, the vertical and horizontal axes are currently seen as static. To make them dynamic, they are viewed in 3D. This 3d form and credibility of perception of AI improving the efficacy of CS operations is tested by taking a subset of personnel, like the researcher, and testing their perception. As discussed in Chapter 3, 12 samples with varying levels of experience (2-28 years) were used in semi-structured interviews

to gauge their perception, sentiment and outlook of the efficacy of AI in CS operations in banks in SA.

The articulation of the totality of the impetus and sentiment of the paper related to the ongoing effectiveness of AI in CS operations in SA banks can be experienced in the following quotation: *“Where in the world is a metaphysical subject to be found? You say here it is just as with the eye and the field of vision. But you do not really see the eye. And nothing in the field of vision allows the conclusion that it is seen by an eye.”*
- Ludwig Wittgenstein, Tractatus Logico–philosophicus, sentence 5.633 (2010).

2.9 Literature Review Findings Viz Research Questions

Research Question1: To what extent can artificial intelligence serve as an effective solution for eradicating and preventing cybercrime within the South African banking sector? From the literature, AI is not seen as a turn-key solution and is not viewed and used to eradicate cybercrime wholly. It is a mitigatory, reactive response that is anthropomorphic , and as new threats and attacks arise from complex, sophisticated attackers and software, the CS teams are in a position of response. The use of AI is to advance the banks position, and defence strategies are developed based on the views of potential threats and when threats and attacks occur in real-time.

Research Question 2: What are the factors that contribute to the decrease of the CS staff complement in banks to improve ‘organisational effectiveness’? There is an overwhelming perception that the introduction of AI results in an almost immediate lay-off of staff, be it within or outside of CS teams.

Research Question 3: What evaluation criteria or assessment frameworks are employed by banks to assess the effectiveness of AI-CS capabilities in reducing cybercrime through preventative measures? The body of literature is exceptionally thin and infantile in nature on CS operations and AI and CS. A lexicon has just been created, and most studies (academic and technical) do not cover internal assessment matrices, banks, and organisations are using vis-à-vis to defend software and tools to monitor output. Presently no template/dashboard exists that is publicly available. This is not to say internal CS operating teams do not have one, but based on the available literature, none exists as AI tools are custom and augmented to a team's needs and scale requirements.

2.10 Conclusion

The TOE framework and Perception Theories created a scaffolding to identify and critically assess the body of literature of AI in CS operations in South African banks. Based on the analysis thus far, the approach and findings were both novel. Firstly, through the TOE framework, it was revealed that AI is generally viewed as a single solution that can either be integrated and adapted or not, and that the latest solution is the correct solution, precisely because of the perception that the latest update and upgrade is the best. However, the anthropomorphic nature of AI revealed that, according to an organisation's needs, certain elements, features and infrastructures can be used and others not utilised at all. A legal firm may need simple protections and not the full GenAI feature. The Organisational modality revealed the lack of expert personnel resources needed to manage the system, and the scope for human error is

large; therefore, the quality of CS officers is of utmost importance. The environmental component revealed that the speed and velocity of the absorption of AI in SA banks far supersedes the time it takes to pass regulation. This places banks and CS officers at the helm of safeguarding and defending highly personal data of millions of individuals. The amount of power banks has accrued is becoming unprecedented and borderline dangerous. Via the literary discourse, it became clear that once an organisation starts on the path to advanced AI features, there is a point of no return. This, to a degree struck off and negated the notion that perception even counts, for once on the single road path, one must stay on it till the end.

Secondly, through the analysis of Perception Theories and locating where and how perception forms, the process and outcome was one of empowerment. It located the author/researcher/subject as the focal point of the analysis, given his tenure and experience of working with AI in CS operations within South African banks, and positioned him on the quadrant of 'externalist-top-down'. This view becomes a point of empowerment for the questioning of efficacy of AI in CS operations itself will prevent a herd-mentality of ongoing AI absorption and implementation, and place cognitive checks and balances in place before future AI uptake.

Chapter 3: Research Methodology

3.1 Introduction

The purpose of this research is to explore the impact AI has on improving the operational efficiency of cybersecurity in banks. While adopting and implementing AI across banks is rampant, does AI actually decrease the CS team's workload and, based on machine-based learning and predictability, engage in smart learning to prevent attacks, or is it merely a tool/resource for responding to threats?

Two things are occurring concurrently. Firstly, there is not a large body of literature focused on the impact of AI in CS, enabling this research to fill the gap and guide potential users in deciding whether AI benefits them. Secondly, given the devastating impact leaked information could have on institutions, no prevalent data is existing that banks and other financial institutions are willing to readily share severely limiting the knowledge base of the true impact of AI and AAI on CS operations. Knowledge of AI and CS, both in operating and academic circles, is both slim and infantile in nature. Given the highly specialized nature of roles of CS officers and any operating tools being public knowledge, it will give fodder to cyber criminals hence the data available is scarce. To mitigate this challenge a qualitative approach was adopted.

A qualitative approach guided by interpretivism as the research philosophy and the adoption of the Phenomenological Approach, combined with the researchers' personal experience, allowed for highly applicable questions and ease of questioning to security officers in banks in SA. A sample of 12 participants were interviewed sequentially, allowing for healthy engagement and dialogue and creating space for nuanced

information to flow. Given the very thin body of literature on the subject matter, the findings of the analysis will inform the creation of a new model/theory that CS officers across sectors can use as a litmus test and an implementation framework to ascertain if AI is the most viable solution for their platforms and goals. The process is mapped below:

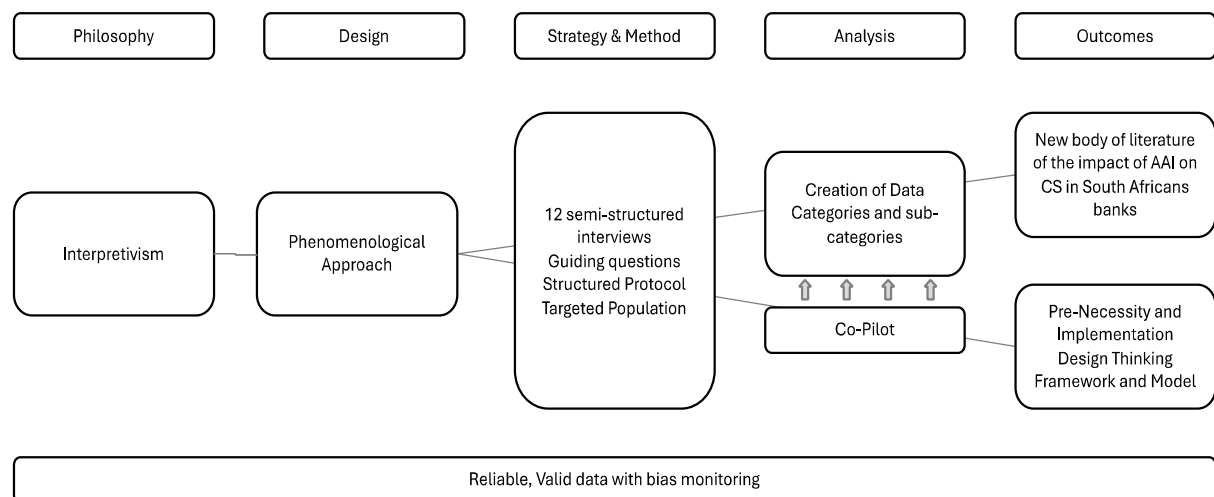


Figure 3.1: Map of Research Methodology

3.2 Research Philosophy

In a study by Kingsley Ofosu-Ampong (2024), he scoped the exact amount of literature available in academic and financial services domains concerning data on AI. Of the content published and produced over three years (2020-2023), only 14 per cent was qualitative. By adopting a qualitative approach, the paper will contribute significantly to the infantile nature of the topic itself and drastically aid in the body of literature and nature of findings in both academic and digital circles. Apart from the novel nature of the topic and given the key limitation of privacy of data to conduct a quantitative study, the author has strategically adopted a qualitative approach.

A qualitative approach aims to facilitate the acquisition of understanding through the experiences of others. Qualitative research designs possess several important qualities. These include a focus on comprehending meaning, utilising non-numerical data, employing inductive analysis, including subjectivity, utilising small sample sizes, maintaining flexibility, gathering rich and thorough data, and placing emphasis on context (Sharma, 2023). This method was selected as the research investigates the perceived impact of AI on cybersecurity operations in the South African banking sector.

According to Ghafar (2024), qualitative research data offers several advantages: It is a detailed and comprehensive description of the participants' ideas, beliefs, and experiences; It most accurately represents the human experience within specific circumstances; It provides various viewpoints, research methodologies, and tools to understand an individual's experiences and; It gives an accurate and thorough assessment of a subject because they enable participants to ascertain which study aspects hold the highest significance for them. Interpretivism is the research philosophy adopted to answer the qualitative paradigm's core and subsidiary research questions.

Four main research philosophies exist: positivism, interpretivism, critical realism and pragmatism. Interpretivism was the most appropriate given that the research methodology is qualitative, with the main approach being semi-structured interviews for a particular target and sample group (Jain, 2023). The results are naturally subjective in nature but given the infantile schematic of qualitative data on AI in general and AAI-come-CS, a qualitative view will offer an ontological view that a purely positivistic approach would be unable to generate. Interpretivism recognises that

“humans are different from physical phenomena because they create meaning” (Saunders, 2023, p. 150) and that reality is a navigatory matrix composed of the intersections of social constructs, the outcomes of which shape experiences, perceptions and interactions.

By adopting interpretivism, which is inductive, the study explored the true lived experience of people with knowledge of specific phenomena. It uncovered the dynamic layers that not only constitute perceptions of AI and CS but also promote healthy cognition towards the approaches taken to the betterment of roles and resources. The aim was not to blanket participants' responses but to understand the rationale and perception behind the participant's interactions and engagements with AI, given their roles in CS.

3.3 Research Assumptions

The ontological nature underlying interpretivism navigates the intersectionality of constructs and the outcomes produced because of. The nature of semi-structured interviews combined with the authors personal experience with AAI in CS enabled a healthy dialogue that unpacked the nuances of how the theological nature of AI is embedded into the ethos of banks, as well as AI being multifaceted in nature and a blanket answer to its perception and efficacy will not do justice to its impact in protection and digital development.

As will be seen, depending on one's tenure within AI and one's roles within CS, the outlooks of the participants have a natural similarity and natural variance. Interestingly, the outcomes tallied with the AI themes of Ampong (2024); that is, AI and AAI fall into

the typologies of technological issues, contextual knowledge co-creation issues, conceptualisation, and domains and applications. The typologies correspond with the interpretative, exploratory nature of the study, which recognises that AI in CS will have multiple constructs. However, the constructs exist within a system, and the nature of all ecosystems is the natural development of an order and hierarchy, which also has an unpredictable angle. The simultaneous interplay of agency within structure guided the qualitative approach and design.

3.4 Research Design

There is a plethora of qualitative research designs, each with its unique characteristics and purposes. According to Among (2024), there are presently no frameworks on CS and AI and AAI given the generic nature of AI as well as the scant data produced on AAI and CS (There are general qualitative and quantitative AI frameworks such as TAM; Fuzzy logic; Self-determination Theory; Critical Theory and Antromorphism Theory, but they are not fully applicable to CS). As such, a phenomenological research design will be utilised.

According to Jain (2023, p. 3) phenomenological research “Aims to understand the essence and meaning of human experiences related to a particular phenomenon. Researchers explore participants’ subjective experiences through in-depth interviews or observations to uncover the underlying structures and patterns of their lived experiences.” This approach successfully marries with the exploratory interpretive guiding philosophy. Data was gathered via comprehensive semi-structured interviews with individuals who are experiencing the phenomenon.

The key characteristic of the design is that it is exploratory in nature, the open-endedness allows for the discovery of new insights, patterns and paradigms. It has a contextual understanding which underscores the socio-cultural and historical contexts that shapes individuals' experiences. The specificities that emerge from the participants' responses offer deep insight into the syntagmatic and paradigmatic outlooks for future design thinking. There is an acknowledgement of the subjective and reflexive nature of qualitative methods due to the inherent bias of the researcher. However, there is a poignant refrain to let biases inform the trajectory of the interviews. Given the highly specialized niche of AI and CS, a small sample group was chosen. In this case, irrespective of sample size, a vibrant rich body of literature was able to be generated due to the caliber of the participants chosen. Lastly, the data analysis itself was an iterative process of "coding, categorizing, and interpreting data to identify patterns, themes, and relationships" (Jain, 2023, p. 6) across an intense three-month period with breaks in between to review, revise and interpret data through a different lens. Thus, the goal of the design approach is to discover and understand the views and perspectives of the studies participants (Cooper, 2007).

3.5 Research Strategy and Method

Data was collected via semi-structured interviews with 12 CS personnel across four tier 1 banks in SA to explore their perception of the impact of AI-CS outcomes. The interviews were guided by an interview protocol which is a schedule of open-ended questions to guide the interview (Edmondson, 2007). Based on the researchers' personal experience, twelve main themes were pre-identified that form the end-to-end technological, contextual, and conceptualisation typologies of AI in CS and aided in answering the research questions.

These were tabulated in an Excel document, and participants' responses were recorded and categorised according to sub-themes. These will be discussed in Chapter 4.

According to Josselson (2013), successful interviews depend on effective structuring; the protocol helped the interviewer (and in this instance, the author/researcher) structure the interview, ensure all relevant topics were discussed and that the interview remained focused on the research objectives. Terms of confidentiality were addressed, and interviews were recorded using the MS Teams platform. The virtual platform has benefits in interview efficiency (Bergelson, 2022). After the interviews, the researcher analysed the data, looking for subthemes across the main themes discussed in the protocol/structure.

3.6 Population

According to Saunders et al. (2023, p. 289), "The full set of cases or elements from which a sample is taken is called a population." As discussed earlier, an extremely thin body of literature is available on the efficacy and impact of technological, contextual, conceptualisation and domain and application themes in AI across all sectors, including banking. There was no support to lean on for paradigmatic guidance. Thus, the population for this study consisted of CS employees who have experience working with AI in CS. As the research aimed to capture a broad range of perspectives and experiences, employees were chosen with diverse backgrounds, and on different rungs of the organisational hierarchy to give deeper insight and richness to the research questions and results.

3.7 Target Population and Sample Group

This study's population consisted of cybersecurity managers, heads of function, and the chief security officer of a South African bank. The banks are among the top four in South Africa and have fully incorporated AI in their CS operations. This was purposefully done for the research questions to be answered and for an expected diverse array of answers. The research used purposive sampling/judgmental sampling. With purposive sampling, researchers deliberately choose individuals with specific characteristics or unique experiences relevant to the research subject. The objective is to locate and enlist individuals who can offer comprehensive and varied data to augment the research findings (Campbell, 2020).

The researcher compiled a list of the requirements of the participants who would be included in the research from the banks' cybersecurity offices. This was done to confirm the participants' relevance to the research problem, considering their real-life experiences.

Population to be interviewed	Position in the Security Office	Sampled Number
Cybersecurity office leaders	Managers in the Cybersecurity Office	8
	Heads of Cybersecurity Functions	3
	Bank Chief Security Officer	1
Number of respondents:		12

Table 3.1: Profile of participants

The total number of participants was 12, as this is the minimum number of participants required for a heterogeneous sample where it is important to maximise variation (Cooper, 2007).

3.8 Unit of Analysis

Conducting interviews is the most popular method of collecting data for qualitative research (Laryeafio, 2023). The researcher informed the participants before the planned interview and scheduled the interviews in their calendars. To ensure organic engagement and to prepare the participants, the researcher shared the interview questions upfront with the participants. The interviews were semi-structured, with a prepared list of questions that allowed for further clarification on the responses given or answers already provided. The advantage of using semi-structured interviews is that the interview is interactive, and it combines structure with flexibility, reflexivity and agency during the interview process (Balushi, 2018).

The units of analysis were the participants' experience and perception of AI, its impact and efficacy on the technological foray, its broader skills-based impact, and so forth. By focusing on each employee at a time and giving each the requisite space to explain their personal experience and idiosyncratic views, the researcher was able to develop unique insights that quantitative analysis would preclude.

3.9 Data Collection

Data for this study were collected by conducting semi-structured interviews with targeted, identified participants. The interview guide was prepared, considering the literature reviewed and the authors' experience working in a bank, on a CS team, and deploying AI and its concomitant digital tools and infrastructure. The purpose and format of the interview were explained to each participant. Terms of confidentiality were addressed, and interviews were recorded using the MS Teams platform.

The participants were contacted telephonically to arrange in-person or virtual interviews. Interviews were scheduled in timeslots suitable for the participant and interviewer. A list of the managers and heads of department who would be interviewed from the bank's security office was obtained from the bank's Human Resources (HR) department. A research data collection log containing the details of the participant's name, the interview number, and the interview's date, time, and place was kept as a method of tracking the data collection. The log and notes made during the interview were stored in a secure location where it is easy to retrieve the data should it be required.

The expected duration of the interviews was shared with these individuals, as they have limited time available during the day. During the interviews, the researcher ensured that the length of the interviews adhered to the prearranged and agreed-upon schedule (60 minutes). A pilot interview was conducted to gauge the interview duration before starting the field work. The pilot interview test aided the researcher in identifying deficiencies, constraints, and other vulnerabilities in the interview design to enable the

researcher to make essential modifications before carrying out the study. The sentiment analysis contributed towards the inductive interpretivism.

Participant	Gender	Participants Profile	Experience in CS role (Years)	Years working in AI	Types of AI and AAI	Meeting Type	Interview duration (mins)
1	Male	Manager Directory Service	2	2	Threat and Anomaly detection	On-line	45
2	Male	Head of Security Architecture	28	2	ChatGPT and CS AI solutions	On-line	40
3	Male	Head of Security Operations	6	4	EDR treat detection and Email Phishing detection	On-line	38
4	Male	Manager Platforms security	2	2	EDR, Email and Database security	On-line	40
5	Male	Manager Security Architecture	20	4	AI SC platforms, including Copilot, MS MDO and email security	On-line	40
6	Male	Manger Detection Engineering and Automation	15	5	EDR, Security Copilot and Charlotte	On-line	48
7	Male	Manager Technical and specialist services	18	2	Copilot and ChatGPT	On-line	30
8	Male	Chief Security Officer	3	3	CS prevention and detection	On-line	40
9	Male	Head Business Advisory Services	6	2	LLM for MI and Analytics	On-line	37

10	Male	Manager Threat hunting and intelligence	14	1	Security Copilot, Charlotte and AI BOT	On-line	40
11	Male	Manager Identity and Access management	7	1	EDR and Email Security	On-line	35
12	Female	Head of Security Risk	20	1	Risk Qualification	On-line	47

Table 3.2: CS participants profiles

Importantly, ten research themes were identified given the nature of work across AI/CS and its typologies. However, it was not pertinent to ask all participants questions based on the themes depending on their experience and role in the workplace. As such, the themes and subsequent questions were adapted per participant for a fruitful interview. The dynamic open-ended nature of interviews caters within its modality for improvisation and adaptation without radically altering the results (Brinkmann, 2018). Themes and questions were adapted, and Participants were interviewed on the full-suite of eight themes and questions as per Appendix B.

The data was collected and saved on a participant-to-participant basis and collated for analysis in an excel workbook.

3.10 Data Analysis

The data collected will be analysed using the thematic analysis technique. Thematic analysis systematically identifies and organises data to provide meaning within a dataset. This method allows for a rich and detailed understanding of participants'

experiences and perspectives. By focusing on themes across a dataset, the thematic analysis technique allows the researcher to understand the shared experiences and meanings across the dataset (Braun, 2017). Thematic analysis follows six steps. (Naeem, 2023).

Step 1: Familiarise yourself with the data - This is the preliminary stage of the thematic analysis process. It entails the transcription of material and the acquisition of familiarity with it. Researchers thoroughly analyse the content to identify preliminary themes and significant portions. They subsequently choose quotations that vividly illustrate the data and effectively embody various perspectives and trends relevant to the research aims.

Step 2: Identify key words and themes - This phase entails meticulously analysing the data, whether derived from interviews, focus groups, or visual materials. Researchers recognise recurring patterns, terminology or visual components and classify them as keywords. These keywords encapsulate the experiences and perspectives of participants, directly drawn from the data.

Step 3: Generate codes for data - In the third phase, coding/'brief phrases or words' are allocated to data segments that encapsulate the material's essential message, importance, or subject. This stage streamlines intricate textual data by converting it into a theoretical framework and aids in pinpointing features pertinent to the research topics. Keywords are fundamental in coding, serving as the foundation for analysis and facilitating the transformation of raw data into informative, manageable units.

Step 4: Develop and review themes—Theme development entails categorising codes into coherent groups to discern patterns and correlations, providing insights into the study subject. In this phase, the researcher shifts from an exhaustive examination of codes and categories to a more abstract interpretation by formulating themes. These themes transcend mere recurrence, encapsulating patterned meanings that connect the research questions and data.

Step 5: Conceptualise and define themes – Conceptualisation entails comprehending and delineating concepts derived from the data. Researchers identify societal trends and distil them into classifications that correspond with their studies. They employ instruments such as diagrams or models to comprehend the interrelations among these notions. The quality of these definitions is evaluated based on clarity, precision, dependability, applicability, and their contribution to theory and practice.

Step 6: Develop a conceptual model—The concluding phase of the topic analysis involves formulating a conceptual model. This method generates a distinctive representation of the data, frequently informed by established theories. The model addresses the research questions and highlights the study's contribution to knowledge. This stage is the conclusion of the analysis, summarising the findings and insights obtained from the data.

The 6 steps listed will be used to analyse and interpret the data obtained during the research participant interview process.

Microsoft Copilot was used to perform the analysis and generate the key themes from the interviews. The key themes were then coded and summarised using MS Excel.

The researcher collected data through interviews, which allowed for a thorough understanding of the phenomenon through the participants' direct accounts. (Patton, 1999). The researcher considers the information to be the intellectual property of the participants. Strict confidentiality and anonymity were guaranteed to ensure no information leakage would occur that would benefit or disadvantage any of the participants in their respective places of work and careers. The consent letters and data are stored without identifiers and are therefore anonymous. The data is stored privately and has strict access controls. A back-up exists on the cloud and is scheduled for deletion at a specific time period.

3.11 Quality Assurance

Reliability and validity are the cornerstones of a successful qualitative body of literature. Indiscriminate research bias and skewed results pollute the sanctity of the findings, rendering the initiative benign. According to Cope (2014), credibility, dependability, confirmability, transferability and authenticity are the five criteria that should be used to ensure credibility and trustworthiness of qualitative research.

Credibility refers to the credibility of the facts on the participants' perspectives, as well as the researcher's interpretation and depiction of them. Dependability reflects the consistency of the data under comparable conditions. Confirmability refers to the researcher's ability to establish that the data accurately reflect the participant's responses rather than the researcher's biases or perspectives. Transferability refers

to findings that can be employed in different contexts or groups. Authenticity refers to the ability and degree to which the researcher accurately conveys the feelings and emotions inherent in the participant's experiences.

In addition to the five key data-checking mechanisms, member checking was also employed to enhance the validity and reliability of the researchers' findings. Member-checking was performed by sharing the finding of the results with the participants so that they can provide feedback on the findings. However, this was not performed during the data collection and analysis process due to time-sensitivity, but all reliability and validity measures were strictly performed.

3.12 Ethical Considerations

The researcher requested the participants' permission to record and utilize any sensitive information they might have disclosed during the data collection interviews, respecting their rights, opinions, beliefs, principles, and preferences. Laryeafio (2023) postulates ethical considerations to follow to safeguard the rights of the participants and this was performed through:

- Research objectives were communicated both verbally and in writing to participants
- Participation in the research was voluntary
- Participants were notified of all the data collection activities
- Participants will remain anonymous
- The final report will refer to a bank and will not contain any of the bank's names
- The researcher agrees to the terms of the plagiarism declaration form, which includes acknowledging other researchers where necessary

- The collection of data only commenced once an Ethics Clearance Certificate was received from the Wits Business School.

Chapter 4: Findings and Discussion of Studies

4.1 Introduction

This chapter aims to analyse the findings of the answers given by the twelve participants in the semi-structured interviews. The interviews were centred around twelve main themes and questions that broadly encompass the operating structures and model of CS operations within large banks in SA.

The chapter will first recap the participants, followed by the responses to the interview questions. Secondly, the format will offer a summarized version of the answers provided with a discussion on key findings. This will form the bulk of the chapter. The findings are coagulated answers positioned graphically for ease of understanding. All the charts, tables, and diagrams are the authors' own.

4.2 Summary Background of Participants

The data collected by the research was conducted via semi-structured interviews. Twelve participants, who possess a vast array of experience in banking security and cybersecurity, were included. They all work in different parts of CS operations, and their perspectives form a rich tapestry of data. The heterogeneous sample included 8 Cybersecurity Managers, 3 Heads of Function, and a Chief Security Officer of a South African bank. All participants have knowledge of AI and experience using AI in various roles within the bank's cybersecurity office. The average cybersecurity experience among the participants was 11 years and 9 months, with a range from 2 to 28 years of experience in cybersecurity itself. Nine participants have been using AI-powered cybersecurity solutions actively in their current roles for between two and five years.

Importantly, cybersecurity teams work in silo's depending on the nature of work. As mentioned in the literature review, cybersecurity covers data protection, infrastructure and software protection. As such, each channel does its best not to cross into, learn from, and impede the other. There are strict internal protocols to protect separation and division and ensure security measures are upheld to the highest degree. Whilst this may seem like a constraint to the findings the opposite holds. Each participant's subjective personal experience based on their role and experience offers a particular insight that a homogeneous group would be unable to. The scale and depth of answers, as well as the cross-sectional similarities and differences also speak to the nuanced nature of implementing and managing AI.

Below is a summary of the participants' background, role, and experience with AI. Notably, Participants 2, 5 and 12 hold the most experience and their insights contributed significantly to many of the points that formulate outliers in the findings below but are phenomenal pillars of knowledge in and off themselves.

P	Role	Total Experience	Experience in AI
1	Manager Directory Service	2	2
2	Head of Security Architecture	28	2
3	Head of Security Operations	6	4
4	Manager Platforms security	2	2
5	Manager Security Architecture	20	4
6	Manger Detection Engineering and Automation	15	5
7	Manager Technical and specialist services	18	2

8	Chief Security Officer	3	3
9	Head Business Advisory Services	6	2
10	Manager Threat hunting and intelligence	14	1
11	Manager Identity and Access management	7	1
12	Head of Security Risk	20	1

Table 4-1: Summary of Participants Background and Experience

4.3 Interview Themes and Rationale

The interviews were guided by the following themes/facets that cover the end-to-end operations of a CS team within a big bank. Most importantly, the questions were derived from the operating model and articulated to feed the scope of the research questions. The themes, the researcher's rationale, and their link to the research questions per question are presented below.

Guiding Questions	Researchers Rationale	Link to RQ
Current Use of AI in Cybersecurity	To understand if they are using AI in CS and for what AI is used.	RQ1
Benefits of AI in Cybersecurity	To ascertain the benefits of using AI in CS from their perspectives.	RQ1
Risks Associated with AI in Cybersecurity	To ascertain if there are any potential risks that they are aware of in using AI in CS.	RQ1
Mitigating AI Risks	To understand if there are risks associated with the use of AI in CS, how those risks could potentially be avoided or mitigated.	RQ1, RQ3

Impact of AI on the bank's Security Posture	To understand if the use of AI in CS has provided better protection against cyber attacks than older traditional security systems.	RQ1, RQ3
Automation of Activities	Was any work previously done by people now automated and done by AI.	RQ1
Impact on Cybersecurity Engineers	Was there an impact on team sizes, structures and the daily work which CS engineers are performing.	RQ2
Decision-Making Visibility	Is there visibility on how AI makes its decisions and is it important to understand how the decision was made or are they just focused on the output from the AI decision.	RQ2, RQ3
Data Management and Governance	Do they know what security data from systems are collected by AI, where it is stored and who have access to that collected system security data.	RQ1, RQ2, RQ3
Protecting AI Systems	To get a better understanding of what they do to secure the CS AI platforms to ensure the AI itself is safe from cyberattacks.	RQ3

Protection of Non-AI Systems ¹	To get an understanding of what they do to protect legacy systems where CS AI systems are not compatible or able to function.	RQ3
AI providing a Competitive Advantage	To get an understanding if the use of AI in CS improves the overall competitiveness of the organization.	RQ1

Table 4.2: Interview Themes and Rationale

Section 4.4 covers the findings of the interviews comprehensively.

4.4 Interview Findings

4.4.1 Current use of AI in CS teams

AI has become an integral part of cybersecurity operations, offering numerous benefits and transforming how organizations protect their digital assets. AI is utilized to enhance threat detection, automate responses, and improve overall security posture.

As mentioned earlier, the participants occupy different roles within the bank's CS teams. They each have varying levels of specialisation and expertise, and their background and experience provide incredible insight into the world of AI-CS. Participants added a disclaimer that the full end-to-end AI tools cannot be mentioned due to the privacy and sensitivity of their work. Participants respectfully shared what was within the ambit of public AI-CS deference. Furthermore, many of the teams were in the pilot testing and experimental stages of launching new AI tools, but were unable

¹ The protection of non-AI systems was only asked to the participants who have scope outside of the CS division and the interviewer determined the necessity in asking the question.

to share the full scope of tools and software due to internal protocols and for want of not providing any causative information resulting in data breaches. Cumulatively and in summation the tools, and uses of AI are as follows:

AI CS Tools	Uses
CrowdStrike EDR	AI-Powered endpoint detection and threat response
Microsoft Defender for Office 365 &Endpoint Defender	Endpoint security platform that prevents, detect, investigate and responds to advanced threats.
Microsoft Copilot	A generative AI assistant that assists with productivity and helps to get more done faster and in less time.
Large language models	Enhance user experience
Vendor solutions	Data optimization
Copilot and ChatGPT	Daily work, primarily for analysing information, creating documents, and communication
Charlotte AI	ML for threat detection
Other AI tools	• <i>Protect and detect spaces and AI for CS that are not covered and prone to vulnerabilities.</i> • <i>Dealing with self-mutating malware and superior IQ threat actors</i> • <i>Used in behavioural analytics to detect potential malicious activities and focus incident response teams</i>

Table 4.3: AI tools and its Uses in CS teams

Three main themes emerged from the responses: Firstly, ‘Enhancing Threat Detection and Response’, AI-powered platforms significantly improve the speed and accuracy of

threat detection and response. For instance, Participant 3, mentioned that AI allows for responding to specific threats with machine speed, which is crucial given the rapid pace of cyberattacks, “AI helps in responding to threats with machine speed and enhances the capabilities of junior analysts” (2025). Similarly, Participant 2 highlighted the benefits of AI in threat detection and response, wherein he noted, “We see a lot of benefit in terms of threat detection and response, being able to automate a lot of our threat detection”.

The second theme was that of AI ‘Automating mundane tasks’. AI is being used to automate repetitive and mundane tasks, freeing up cybersecurity professionals to focus on more complex issues. Participant 10 emphasized the benefits of AI in report writing and interpreting technical topics. He said, “AI helps in summarizing long reports into short segments and interpreting technical concepts into simpler terms”. The third theme was ‘Improving Email Security’. Email security is another area where AI is making a significant impact. Participant 5, who focuses on email security within the cybersecurity office, mentioned that AI is used for analysing email patterns and detecting threats. He explained, “AI analyses bulk email in real-time and surfaces pertinent information much faster than in the past”.

In summary, the AI-CS employed stemmed from AI for performing routine tasks to prevent human monitoring fatigue, email threat security software, bot creation, and multi-layered AI tools that integrate and centralise certain data feeds.

4.4.2 Benefits of AI

In terms of the benefits of AI, eighteen core points emerged. Six of the participants mentioned four recurring points. Nine aspects became the outlier with <four mentions.

There was an overwhelming view on the benefits of AI within CS operations:

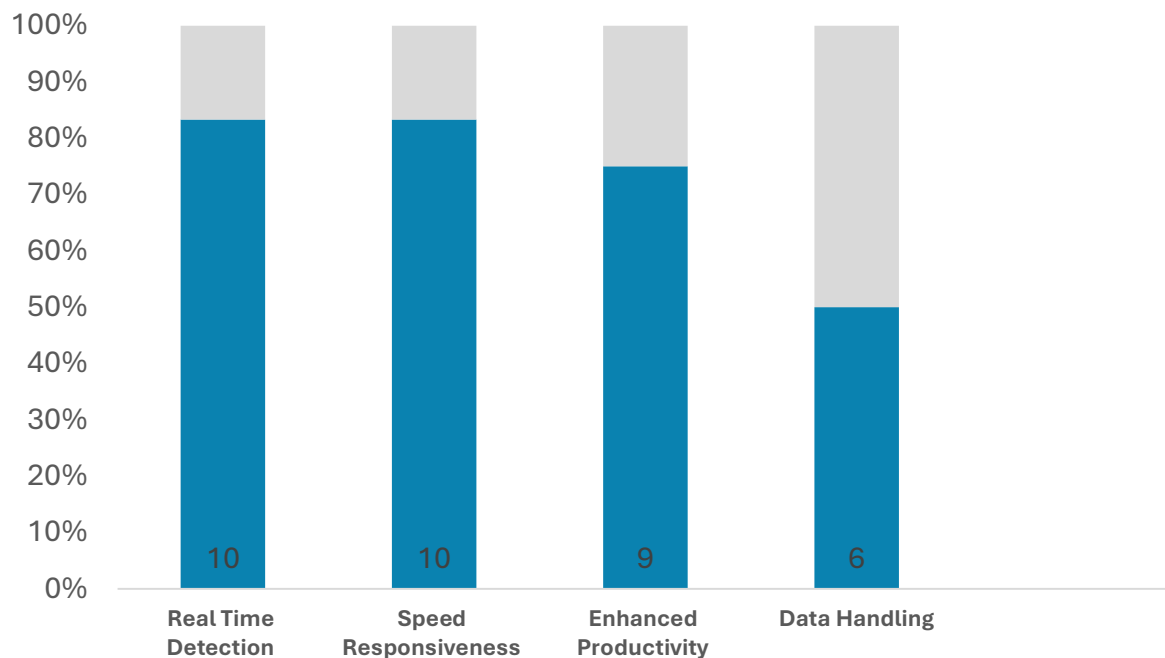


Figure 4.1: Top 4 benefits of AI within CS operations

Ten participants mentioned ‘Real-time threat detection’ as a benefit. Here, AI tools provide real-time threat detection by continuously monitoring network traffic and sensors on desktops, allowing for immediate action and faster response times. The speed of response is also critical to threat detection, as attackers use AI, and the slower the banks respond, the faster attackers can breach a system. Nine participants cited an ‘Increase in Productivity’ due to AI tools automating routine tasks, allowing CS professionals to focus on more complex issues. Half of the participants mentioned ‘Data Handling’ as AI can handle large amounts of data and identifying anomalies that human analysts might miss.

Recalling the establishment of credibility and trustworthiness of Perception Theory, while the majority of participants did mention similar benefits, and they all have validity and truth, henceforth there will be priority weighting attached to three of the participants who have +15 years of experience in CS operations and work with the introduction and implementation of new versions of AI as they evolve. Importantly, as mentioned by Participant 2, the nature of data, threats and breaches is highly confidential, and centralised access is non-permissible. Each team within CS operations only has access to enough CS data for their roles and responsibilities. A criticism lunged at banks is that there is no centralised data access point and teams work in silos preventing cross-pollination, usage and learnings of AI resources, people resources and cognition/solution capability. However, since there is a lack of regulation with regard to AI, there is a necessity to install a separation of powers for the purposes of organisational and financial health of the bank and its customers, respectively.

Based on the cross-sectional diversity of the sample, the remainder of the responses are presented below:

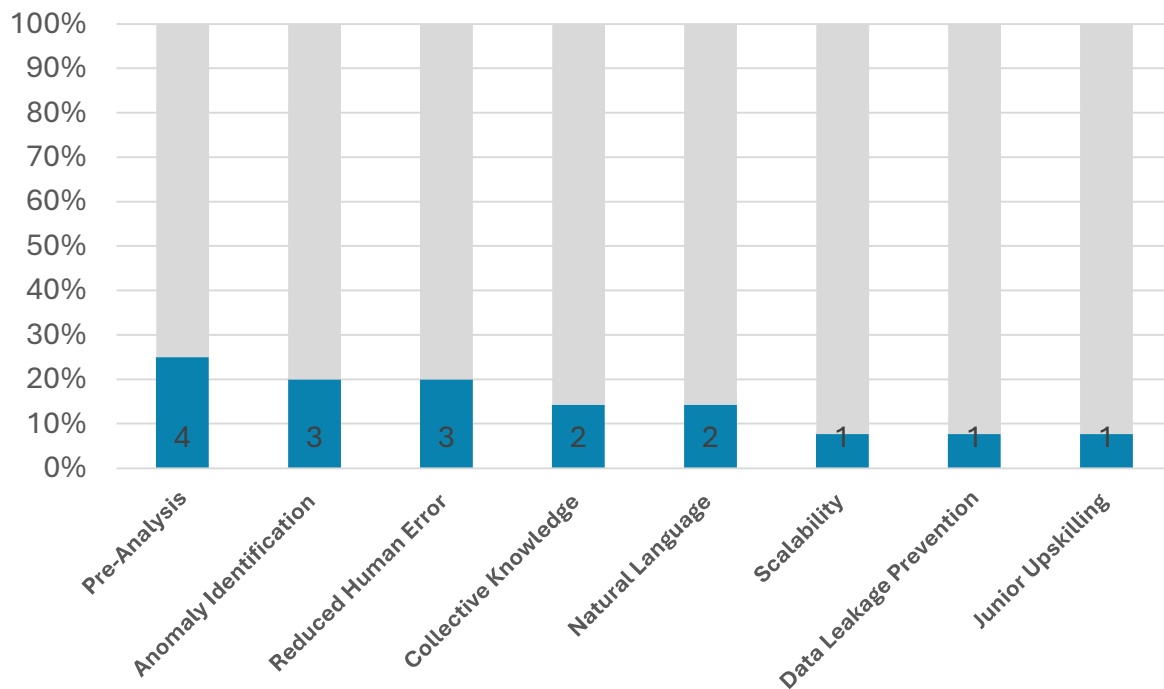


Figure 4.2: Benefits of AI in CS operations

‘Pre-analysis’ is an AI-powered platform that provides a pre-analysis of incidents, which helps in faster investigation and response. It is beneficial in handling the large volume of alerts and incidents that teams confront. Anomaly Identification refers to the detection of threats that fall outside of the scope of normal threats. With the correct data input, AI reduces human error by handling basic detections accurately and consistently without fatigue. AI offers access to collective knowledge and data, helping teams learn from attacks experienced by others. AI facilitates easier data integration from multiple security layers, providing contextual feedback on detections. AI has a good response to natural language, enabling ease of use.

The following three points were only single inputs: AI models offer better scalability by integrating data from various sources, such as EDR and network telemetry, to provide comprehensive context for detections; AI solutions are employed to prevent data leakage, ensuring sensitive information is protected from unauthorised access or

exposure and; AI reduces the need for extensive training for email engineers as it can analyse and surface pertinent information quickly. Notably, this does not render analysts redundant but frees up their time to focus on more complex tasks. Providing them with pre-analysis of incidents and explaining complex issues helps junior analysts perform at a level closer to that of senior analysts. Notably, while artificial intelligence substantially enhances efficiency and facilitates decision-making, it concurrently engenders ethical concerns, such as potential job displacement and the reduction of human interaction in customer service (Shankar, 2024).

While AI offers CS teams a plethora of benefits, it is also highly risky. We will now discuss these, followed by risk mitigation strategies employed.

4.4.3 Risks of AI in CS

AI has become an integral part of modern cybersecurity operations, offering numerous benefits such as enhanced threat detection, automation of mundane tasks, and improved efficiency. However, using AI also introduces several risks that must be carefully managed to ensure the security and integrity of systems. Participants in the research project have shared their insights and experiences, highlighting the various risks associated with using AI in cybersecurity. Human error is the primary factor contributing to CS breaches (Amoresano, 2023).

The incorporation of AI into computer science presents several risks, such as issues related to data quality and bias, challenges with interpretability and transparency, as well as concerns about privacy and security (Božić, 2023). Twelve risks emerged from the interviews. The risks offered an insightful view into what CS teams have to

encounter daily. The risks will be detailed below and categorised according to TOE and the scope of CS operations. Despite the size of the CS teams being relatively small compared to other divisions within banks, the scale of operations is vast and its impact deep as demonstrated in Figure 4.3. The summary findings of the risks were:

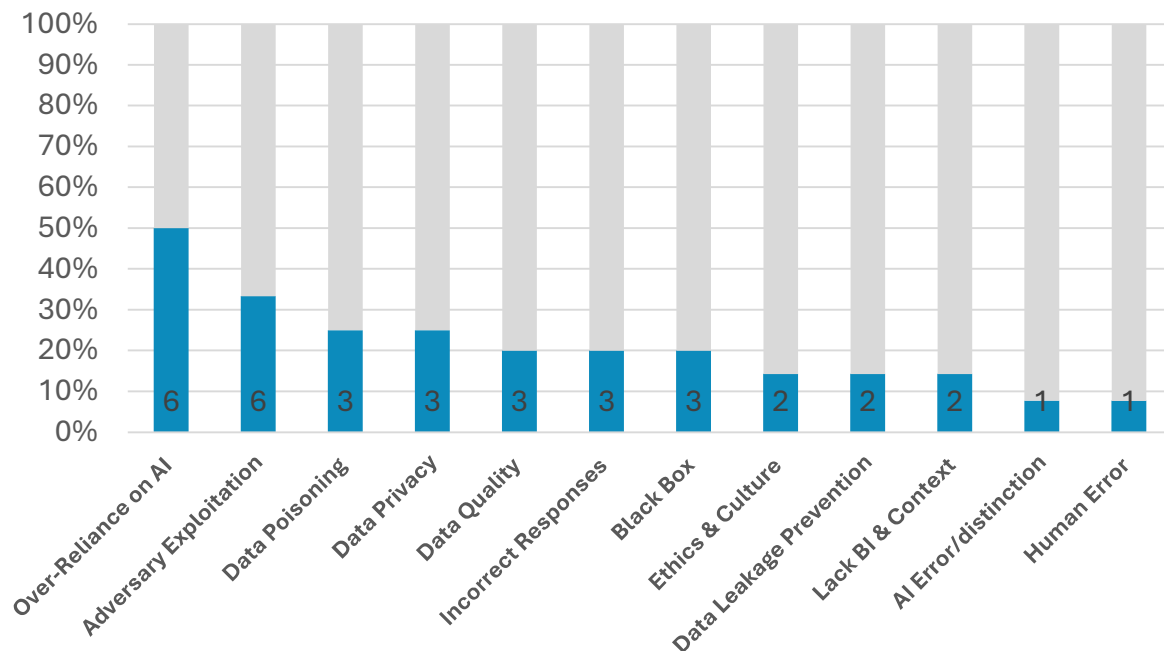


Figure 4.3: Summary of Participants Risks in AI

Six of the participants cited ‘over-reliance on AI’ and ‘Adversary Exploitation’ for the following reasons. Over-Reliance on AI leads to CS professionals overlooking malicious activities as they develop a dependency on AI for threat detection. Adversary Exploitation of AI vulnerabilities, such as introducing noise into images to bypass security measures, make it difficult for AI to recognise explicit content. AI can be tricked by adversaries, which can lead to incorrect recognition and potential security breaches. Participant 2 mentioned “Our adversaries and how they use these cybersecurity platforms because as much as we could use it for defence, many of

these platforms are open to abuse from an attack perspective as well". This is consistent with the findings (Kumar, 2023; Salem et al, 2024).

Three of the participants mentioned the following five themes. One, 'Data Privacy' due to the large amounts of security data processed by AI tools, including the sensitivity and protection of the data. Two, 'Data Poisoning' wherein attackers use AI to corrupt data, compromising the system's security and affecting AI decisions. Participant 3 pointed out this risk, noting, "There is a risk around data leakage and data exposure that comes with AI". Malicious actors intentionally feed incorrect data into AI systems to manipulate their outputs. Three, 'Data Quality' impacts AI decision-making capability with decisions being only as good as the data quality. Data quality used to train AI models is crucial for their accuracy and reliability. Poor quality data can lead to incorrect decisions and increased risks. Participant 12, stated, "If the quality of data is bad, the decisions ultimately are going to be bad or incorrect". Fourth, 'Incorrect Responses' wastes time and resources as analysts focus on irrelevant issues while real threats move through the environment. One of the primary risks associated with AI is the potential for incorrect responses and misinterpretations. AI systems are not infallible and can sometimes provide inaccurate or misleading information. Participant 3 stated, "The biggest issue is the risk of incorrect responses, which can lead to wasted time and missed threats". This sentiment was echoed by Participant 10, who mentioned, "AI can sometimes misinterpret technical topics or fabricate events, which can be problematic if trusted blindly". Five, 'Black Box' - Experts do not fully grasp how AI evolves, which leads to unforeseen risks. The proprietary lay with the vendor and there is no visibility on the internal AI processing.

The following risks represented the remainder of the feedback. Two of the participants mentioned 'Ethics and Culture' as AI must be aligned with organisational ethics and culture to avoid decisions that conflict with the organisation's values. If not properly trained, AI can make decisions that conflict with these standards. Participant 12 emphasized the importance of aligning AI with organizational ethics, stating, "You need to make sure that AI is taught correctly in the sense of what is your ethics and morals in the organization". 'Data Leakage' as there is concern about how data is used to train models and whether it is exposed or shared globally, which could lead to unintended data exposure and; 'AI systems lack business intelligence and environmental context'. Here AI treats all information equally without understanding specific business contexts and the rationale behind priority weighting. Without this context, the accuracy and effectiveness of AI responses can be compromised, leading to potential security risks. AI systems often lack the context and understanding that human analysts possess, which can lead to incorrect or incomplete responses. Participant 5 mentioned, "AI lacks the business intelligence in the system normally, so it does not have real intelligence on how you like to operate". This lack of context can result in AI systems making decisions not aligned with the organisation's goals or ethics. Within the field of generative AI, additional ethical concerns emerge regarding disseminating misinformation and reinforcing biases present in the training data (Zlateva et al., 2024).

Single risks included 'AI Error/Distinction', as sometimes no distinctions are made between large language model AI, like Copilot, and other types of AI, such as ML, which is more statistical and processes numeric data, and 'Human Error', which is a risk but a natural component of functionality. As a result, cyber-attacks interfere with

and aim at computer systems, networks, and data, making it essential to implement strict security measures across these areas (Salem, 2024).

The risks affect the technology, organisation, and environment in the following ways and impact the channels needing protection.

 Risk affliction
  Risk non-affliction

	Themes in CS Risks	Framework			Channels needing Protection		
		T	O	E	Data	Infrastruc- ture	Software
1	Over-Reliance on AI						
2	Adversary Exploitation						
3	Data Privacy						
4	Data Poisoning						
5	Data Quality						
6	Incorrect Responses						
7	Black Box						
8	Ethics and Culture						
9	Data Leakage						
10	Business Intelligence						
11	AI Distinction/Error						
12	Human Error						

Figure 4.4 CS Risks Across TOE and Protection Channels

While AI offers significant benefits in cybersecurity, it also introduces several risks that must be carefully managed. By understanding these risks and implementing appropriate mitigation strategies, organisations can leverage the power of AI while

minimising potential negative impacts. According to Palakurti (2025), the creation of ethical AI/ML systems should prioritize transparency, accountability, and inclusivity.

4.4.4 CS-related Risk Mitigation Strategies

The risk mitigation strategies were not asked to be answered in parallel and in congruence with the risks mentioned above, but were mentioned on an ad hoc basis, depending on role and experience applicability. There were fourteen risk mitigation tactics mentioned. A fundamental step in mitigating AI risks is to ensure a thorough understanding of these risks across all levels of the organisation. Participant 12 emphasized the importance of understanding AI risks, stating, “A better understanding of the AI risks is already a mitigant because if you understand something, you can manage it”. This involves educating staff about the potential risks and ensuring they are equipped to handle AI-related challenges. The findings are presented graphically below:

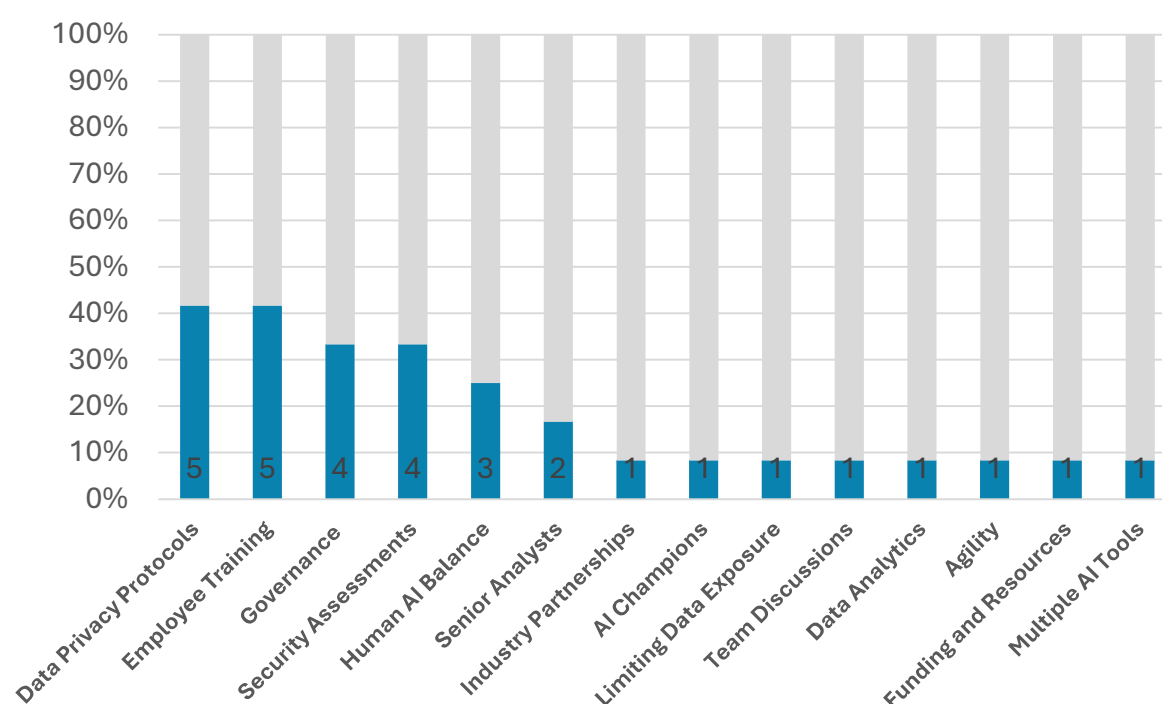


Figure 4.5: List of 14 Risk Mitigation Tactics within CS Operations

'Data Privacy Protocols and Employee Training' were the most mentioned tactics to prevent AI-related risks from arising. In terms of the former, implementing robust data privacy protocols is essential to protect sensitive information processed by AI systems. Participant 3 mentioned the importance of basic security hygiene, stating, "You have to make sure then that proper hygiene is in place and that systems are locked down to their bare minimum to only what they need to function. Also, ensure Patching is done right". This includes regular patching, access control, and other security practices to safeguard AI systems. Participant 12 stated, "Codifying information into guardrails helps prevent AI from acting in ways that would not be acceptable if done by humans." Participant 4 suggested a layered defence strategy to mitigate AI risks, ensuring multiple layers of security to catch threats that may bypass initial defences. Ensuring robust security measures are critical to protecting AI systems from potential vulnerabilities.

In respect to the latter, training employees on best practices for using AI systems is crucial to mitigate risks and ensure proper handling of AI tools. All members of the CS operations are hired on the basis on pre-existing skills and knowledge required for the role. The nature of work demands newly hired members to be proficient on AI tools and systems. However, AI is a completely customised and bespoke tool and, in some teams, the staffs role remains the same irrespective of evolving AI, and in other teams, roles change depending on the malleability required for the purpose of work. In this regard, employee training is conducted for novel features to a tool/solution/software for optimal performance and results. There is ongoing educational and on-the-job training. However, the training falls outside the traditional setae and Udemy-related trainings. Thus, proper training and skill development are crucial in mitigating AI

risks. Participant 12 highlighted the need for training, stating, “I think training and development plans for people around what is AI really and not the hype, but the actual... How do we skill ourselves up to utilize it in the right way so that you actually don't introduce risk but actually opportunities”. This ensures that employees are well-versed in AI technologies and can use them effectively without introducing new risks. ‘Governance’ and ‘Security Assessments’ were the next bundle of risk mitigation strategies that were recurrently noted. In terms of Governance, a better understanding of AI risks is cited as being crucial for effective management. This is in specific regard to business executives who demand transparency and explainability in AI outputs to ensure decisions are well-understood and justified. Secondly, regular security assessments and audits are necessary to ensure the effectiveness and security of AI-powered platforms. This includes Data Tuning (regularly tune and refine data to ensure AI models provide accurate results).

‘Human-AI Balance’ and ‘Senior Analysts’ role were deemed mid-level points of criticality. Maintaining a balance between human expertise and AI capabilities is crucial to mitigate risks as AI-to-fight-AI can often evade context. Furthermore, as will be mentioned shortly, the quality of data and lack of visibility over AI tools given by proprietors prevent augmentation in defence. Thus, a human view is needed in certain circumstances. It was mentioned that senior analysts should test AI solutions due to their experience, capability and oversight compared to junior analysts. This will posit more robust protocols and solutions, prevent lesser augmentation in future and decrease the labour of iteration. This approach also helps avoid the risk of teaching incorrect methods to junior analysts.

The eight additional risk mitigation tactics are tabulated below for ease of readability:

Risk Mitigation Tactics	Rationale
Industry Partnerships	Partnering with industry leaders and using tried and tested AI models to mitigate risks. Participant 12 emphasized the importance of partnering with industry leaders, stating, “Partnering with industry leaders, let’s not learn on our own... make use of industry tools”. This ensures that organizations are using tried and tested AI models and solutions, reducing the likelihood of errors and vulnerabilities.
AI Champions	Working with AI champions in different areas, such as vulnerability management and threat hunting to test the accuracy and responsiveness of AI solutions. These champions run specific use cases and multiple iterations to ensure consistency and reliability. Participant 3 mentioned the use of AI champions and senior analysts to test and validate AI responses, stating, “We are working with AI champions in each area to test the accuracy and consistency of AI responses”. This helps in identifying and rectifying any errors or inconsistencies in AI outputs.

Limiting Data Exposure	Limiting data exposure by using specific models and platforms. There is caution about which XAI platforms are used to minimize the risk of unintended data exposure. Limiting data exposure and ensuring that AI systems do not inadvertently leak sensitive information is another crucial aspect of risk mitigation. Participant 3 highlighted the importance of data exposure management, stating, “We still trying to understand data leakage exposure that comes with a lot of this and how is our data being used to train particular models then being used globally”. Implementing strict data governance policies can help in minimizing the risk of data leaks.
Team Discussions	Fostering team discussions to share experiences and insights helps in recognizing AI limitations and improves its use.
Data Analytics	Using data analytics to monitor AI performance and validate its decisions can help mitigate risks and ensure accuracy.
Agility	Agility in responding to AI-driven attacks, ensuring the organization can quickly adapt and address emerging threats.
Funding & Resources	Position for significant funding and resources to effectively mitigate AI-related threats.
Multiple AI Tools	Use different AI tools working together in layers to offer a better risk posture than relying on a single platform.

Table 4.4: Description of Risks to AI and its systems

While AI offers significant benefits in cybersecurity, it also introduces several risks that must be carefully managed. By understanding these risks, providing proper training, implementing oversight mechanisms, partnering with industry leaders, ensuring robust security measures, conducting thorough testing, and limiting data exposure, organizations can effectively mitigate the risks associated with using AI in cybersecurity.

4.4.5 Impact on Security Posture

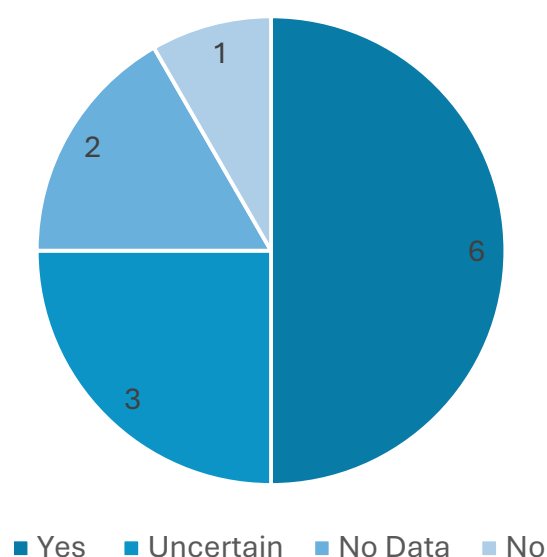


Figure 4.6: Participants outlook on AI improving the banks security posture

Two of the participants did not provide any data on the security posture. 50 percent of the participants believed AI improved the banks security posture for the following reasons: Significantly improved email filtering, reduces the number of spam emails and enhances the bank's ability to detect and prevent email-based threats; Enhanced threat detection capabilities, allowing for more efficient identification and response to potential security threats; Improved the bank's response capabilities enabling quicker

and more effective actions against identified threats; It serves as an extension to existing security measures; AI has significantly reduced the risk of ransomware attacks by analysing behaviour patterns and intent leading to a 300 percent reduction in emails affected by malware year-on-year; Increased productivity within the CSO as AI is used to manage meetings, create and read documents. AI tools make the bank more efficient in handling vulnerabilities by providing timely and accurate threat intelligence reports.

Of the three uncertain participants and one who emphatically affirmed no, they cited benefits but believed the risks outweighed the benefits. This is due to AI developing a false sense of security by assuming AI alone can significantly improve security posture without considering adversary advancements; other teams witnessing an incremental improvement in AI and believe AI only works well when combined with strong cybersecurity basics and expertise. The CS personnel bring the effects of AI to life and create results and improvements; AI is necessary to keep up with AI-powered threats, even if it does not necessarily improve security – AI is a reactive agent but not a cyber-criminal killer.

Interestingly, while AI increases productivity and eases certain routine tasks, more complex systems require advanced and skilled analysts and engineers to manage the AI, without which the AI will not be used to its full extent. In order for the bank's security posture to improve, it is necessary to invest in the right skills and right people first instead of placing the proverbial AI cart before the robotic horse. Human-machine interaction and augmentation are critical to seeing digital and CS advancements to their plenitude.

Overall, the use of AI in cybersecurity has positively impacted the security posture of organizations. Participant 2 noted that AI has allowed for better threat detection and response, improving the organization's security capabilities. He stated, "AI has definitely improved our ability to prevent certain activities and has enhanced our security posture". AI has significantly improved the security posture of the organization by enhancing threat detection, improving response capabilities, and automating mundane tasks.

4.4.6 Automation on Activities

There was a limited and thin response in assessing whether AI has automated activities across different CS teams. Based on the responses, automation existed across a scale. In some teams, AI-powered platforms did not automate activities. In contrast, in others a playbook was created for threat response exclusively (this is also linked the notion of formal training via best practices depending on role and requirements). This is contrary to what was found by Jada (2024). In most teams where there was a level of automation present, there was also a level of human involvement in the decision-making capabilities. The team were the primary decision-makers on if, how and to what extent automation will occur. Participant 7 mentioned the need for change management as AI is always augmenting internal roles and staff demands and a level of pliability is needed within the cognitive functioning and career safety of staff before acting. The looming fear of machine-replacing-man will prevent employees from implementing certain features if they know their place within the organisation is at risk. The 'perception of AI' in creating safety or displacement for employees always plays a subconscious role in its deployment.

Interestingly, the less automation there is, the more man-power is needed for tasks. But AI also reaches a saturation point and beyond a juncture of complexity cannot shed man-power completely and persistently needs the right people to manage the tools and software to mitigate AI error, to monitor for anomalies and to provide overall defence and security. Whereas other industries may use digital solutions with minimal manpower, there will always be a need for humans with the right skills to manage AI, its automation and functioning.

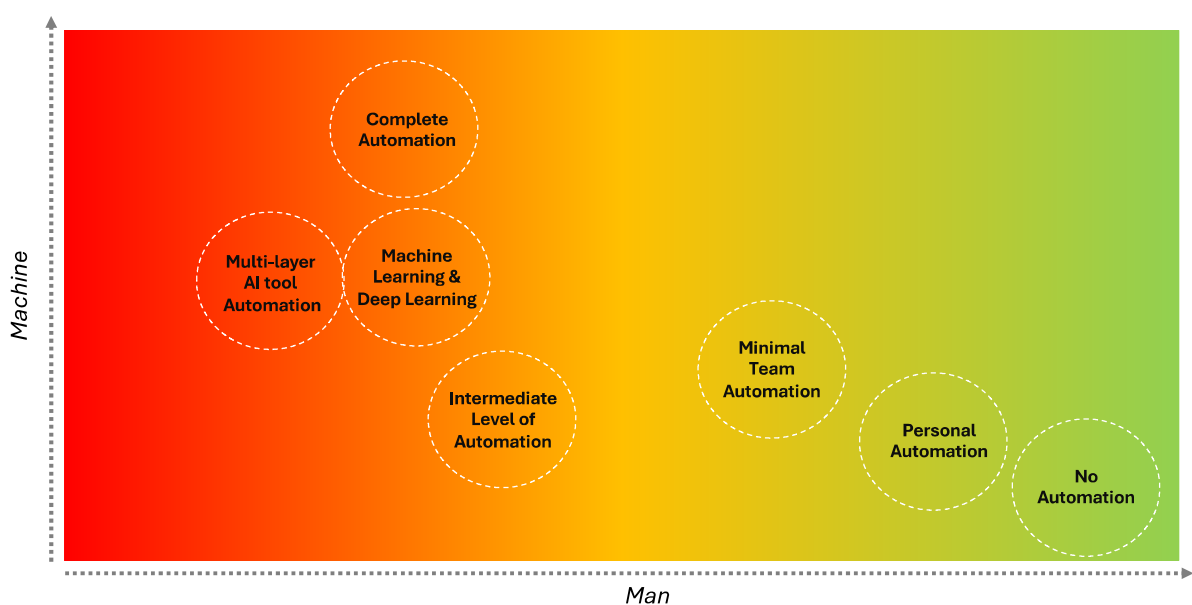


Figure 4.7: AI Automation Scale

Automation is needed to advance CS protection, but it is not mandatory in all teams and functions.

4.4.7 AI's Impact on CS Engineers

The use of AI impacts the people who support and work with AI technologies. This section of the paper explores how AI has influenced cybersecurity professionals, supported by insights and feedback from industry experts.

Participants were asked about the AI's role of CS engineers in the banks CS operations. CS engineers are the security product specialists who support the security technology(ies) in question. For example, in email security the application which protects the organisation against email security threats will be supported by 'email security engineers'. Engineers were chosen because of the 'perception' that AI would be displacing their roles and work specifically, in comparison to any other security-related roles. The participants perception of CS engineers are presented in the graph below:

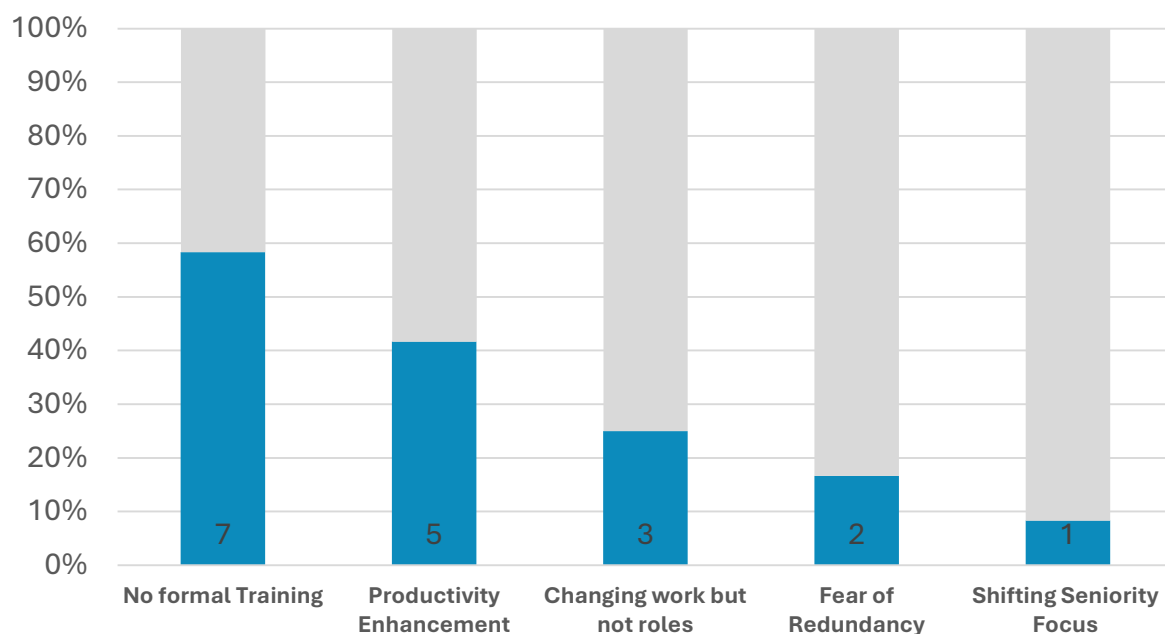


Figure 4.8: Participants Response to AI's Impact on Engineers in CS Operations

AI has significantly improved threat detection and response capabilities, enabling cybersecurity personnel to identify and mitigate threats more effectively. Participant 2 emphasized the benefits of AI in threat detection, stating, "We see a lot of benefit in terms of threat detection and response, being able to automate a lot of our threat detection". This automation allows cybersecurity professionals to respond to threats

faster and more accurately, reducing the risk of potential breaches. AI has also played a crucial role in enhancing cybersecurity personnel's skill development and training. By providing tools that can analyse and interpret data, AI enables professionals to gain deeper insights into cybersecurity threats and vulnerabilities. Participant 1 mentioned the importance of continuous investment in training engineers to use, configure, and support AI cybersecurity platforms, "From an organisation point of view, we need to make sure that our engineers are skilled in how to use and configure and how to interpret the data produced by the AI cybersecurity platform tools... there should be a continuous investment on the training to make sure that we have skilled engineers and be able to use and configure and support the AI service security platforms". This investment ensures that cybersecurity personnel are well-equipped to handle the complexities of AI-powered tools and platforms. Notwithstanding, the perception of engineers is that they receive no formal training on the AI capabilities on some of the tools. CS engineers have been trained on CrowdStrike and Microsoft Defender, but not specifically on the AI capabilities of these tools and the internal workings of AI in these toolsets. The majority of the participants shared that they believe the engineers need formal AI training, and there is scope for future training.

As mentioned earlier, all new members who are hired already possess AI software and tool knowledge but the landscape they operate in is in a constant state of flux and evolution. As security demands increase, focus has shifted from needing and training junior engineers to employing senior engineers. Junior engineers are being displaced as AI itself creates playbooks, senior engineers and analysts are wanted to run diagnostics and testing as they have the capability of implementing new solutions and the speed of response to fix issues is significantly faster. This increases productivity

immensely and frees up the team to focus on other issues and deployments. Senior engineers also do not require as much training and the training that is needed is around ethics, regulations, internal department changes, product enhancements and so forth. The use of AI has shifted the focus to more senior engineers, reducing the need for junior engineers to handle routine tasks and freeing them to handle more complex tasks. There is displacement for upskilling but not for machine-replacing humans.

One of the most notable impacts of AI on cybersecurity employees is the enhancement of efficiency and productivity. AI-powered tools and platforms can automate repetitive and mundane tasks, allowing cybersecurity professionals to focus on more complex and strategic activities. Participant 7 highlighted this benefit, stating, "AI helps in summarising long reports into short segments and interpreting technical concepts into simpler terms". This capability not only saves time but also ensures that cybersecurity personnel can dedicate their efforts to more critical tasks (Jada, 2024).

This is tied to the third main perception that AI changes the nature of work but not necessarily roles. Once again, the anchor of this change is organisational effectiveness and engineers' fear of facing redundancy. When the participants were questioned if staff have been displaced, the answer was that no engineers have been replaced because there is a skills and man-power shortage, but as AI tools evolve in departments, there is also a required adaptation by some staff requiring to upskill to meet work demands. Participant 9 stated, "The adoption of AI has not yet led to changes in job roles and responsibilities, but in future it will, and it will require cybersecurity personnel to develop new skills to stay relevant."

The AI tools are new, and human resource capability is required to learn alongside the AI. Irrespective of seniority, AI has displaced no engineers, but roles and team structures have altered. Notably and importantly, these changes occur over a period of time and there is never a severity and strain in the role-related changes. The adoption of AI in cybersecurity has not yet led to changes in job roles and responsibilities. Participant 11 stated “Not at this point but I think in the future it will definitely influence changing these roles, in future we might not need junior engineers but will rely more on senior engineers..”. Participant 9 noted, “I don't think yet, but I do see that it will happen in the future”. The perception of the role, and impact on engineers of AI is bifurcated. On the one hand, growing bodies of literature point out that machines will replace the role of human resources and decrease headcount. And on the other hand, participants shared that no people are getting displaced in CS operations but are solely being upskilled to be fit-for-purpose. This naturally answers RQ2, ‘What are the factors that contribute to the decrease of the CS staff complement in banks to improve ‘organisational effectiveness’?

With AI handling routine tasks, cybersecurity personnel can focus more on strategic activities such as threat hunting, incident response, and policy development (Villegas-Ch, 2023). Participant 3 highlighted the importance of AI in enabling cybersecurity professionals to respond to threats with machine speed, stating, "AI allows us to have the ability to respond to specific threats with machine speed". This shift towards strategic activities ensures cybersecurity personnel can proactively address emerging threats and vulnerabilities.

Cybersecurity personnel must adapt to new technologies and methodologies as AI continues to evolve. This continuous adaptation ensures that cybersecurity personnel remain at the forefront of technological advancements and can effectively leverage AI to enhance security operations. The integration of AI into cybersecurity operations has profoundly impacted cybersecurity personnel. From enhancing efficiency and productivity to improving threat detection and response, AI has transformed how cybersecurity professionals operate. However, it also introduces new challenges and risks that must be carefully managed. By understanding these impacts and continuously skilling up and adapting to new technologies, cybersecurity personnel can effectively leverage AI to enhance their capabilities and ensure the security and integrity of their organizations (Kayode-Ajala, 2024).

4.4.8 Decision-Making Visibility

Transparency in AI decision-making processes ensures trust, accountability, and ethical use of AI technologies. In cybersecurity, where AI is increasingly being used to enhance threat detection, automate responses, and improve overall security posture, transparency becomes even more critical. One of the key aspects of transparency is understanding how AI systems make decisions. This involves having visibility into the data and algorithms used by AI systems and the rationale behind their decisions. Participant 1 stated that “AI systems often operate as black boxes, making it difficult to understand how decisions are made. This lack of transparency can be a major barrier to trust and adoption”. Participant 3, emphasized the importance of understanding AI decisions, stating, “The biggest issue is the risk of incorrect responses, which can lead to wasted time and missed threats”. This highlights the need for transparency to ensure that AI systems are making accurate and reliable

decisions. A comprehensive approach is crucial for the effective protection of AI systems. This approach should involve establishing robust data governance, minimizing bias in AI algorithms, ensuring transparency and accountability, and adhering to ethical principles (Božić, 2023).

Participants in the research project have shared their insights and experiences, highlighting the importance of transparency and the challenges associated with achieving it. All 12 participants were asked if they had visibility of AI's decision-making capability. The results were:

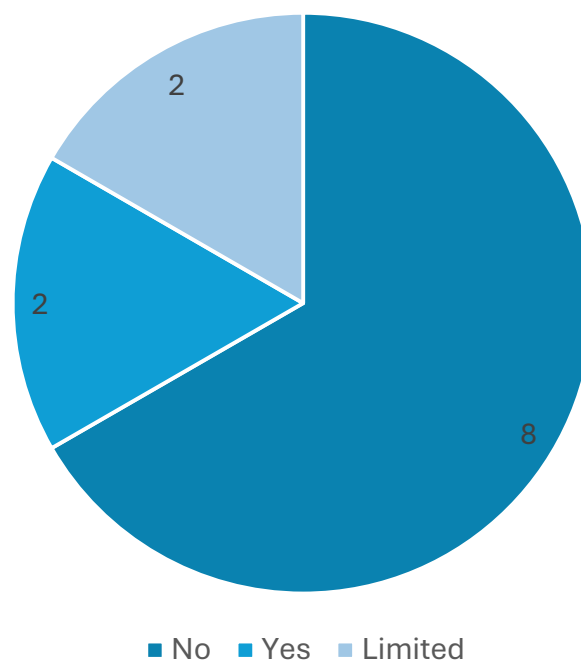


Figure 4.9: Participants Results of AI's Decision-Making Visibility

Sixty-six percent of the participants said they do not have visibility. Despite the importance of transparency, many AI systems still operate as 'black boxes' making it difficult for users to understand how decisions are made. Participant 4 stated that

“Transparency in AI decision-making is essential for building trust and ensuring that we can hold AI systems accountable for their actions”. Participant 10 pointed out the limited visibility into AI decision-making processes, stating, “There's a bit of concern there. So you don't have that visibility”. This lack of transparency can lead to mistrust and hesitation in fully relying on AI systems. Participant 8 highlighted the importance of transparency and stated that “It's essential to have transparency in AI decision-making to ensure that the decisions made by AI systems are aligned with our organisational goals and ethical standards”. This emphasizes the role of transparency in risk management and also showcases why XAI emerged.

Seventeen per cent of participants said they have limited visibility, and seventeen per cent said they possess visibility. The access to visibility of AI's decision-making capability is role-specific and private and confidential on complex AI tools, rendering the findings contextually accurate. Participants 3 and 5, who said they do have visibility, cited, “have visibility into the decisions made by AI-powered platforms, especially regarding detection and mitigation actions”. They can see the output and the reasons behind the decisions, such as identifying malicious behaviour.

Participants 7 and 11 who said they limited visibility said “ While we have visibility into the decisions, the underlying rules and models that support these decisions are proprietary”, meaning they do not have full access to the details of how the AI models work. Visibility of AI decisions is limited due to the proprietary nature of vendor-provided AI models. Vendors play a crucial role in ensuring transparency in AI decision-making processes. They need to provide clear explanations and documentation on how their AI systems work and the data they use. Participant 2

mentioned the reliance on vendors for transparency, stating, “We do expect our vendors to take things seriously. So the expectation is that they always looking into and the tracing these vulnerabilities”. This underscores the importance of vendor accountability in maintaining transparency. Almost all the participants stated that the banks rely on vendor guarantees about the data and CS teams rely on vendor communication for feedback on AI decisions. In the Literature Review there was mention of the disparity between AI users and non-AI users. The proprietary nature of AI visibility is linked to the systemic creation of AI dependency and disparity, also known as “cyber-poverty”. In *Cyber Poverty: Unraveling its Complex Impact on Individuals, Companies, Society, Countries, and Global Trade*,

“Cyber Poverty embodies a complex and multifaceted challenge, stemming from the glaring lack of access, skills, and resources essential for individuals to fully and effectively engage in the digital domain. This issue reverberates across a vast spectrum, impacting individuals, companies, societies, nations, and the intricate tapestry of global trade. This impact is particularly pronounced in today’s interconnected world, where technological advancement is the bedrock of economic and social progress.” (2025, para 1).

There is a very large disparity in the cost of AI infrastructure emerging from the US and Europe compared to India and China. The vendors creating and licensing AI have control over how the algorithms and machines respond and to what extent users and consumers have visibility on the inner workings of the AI. Additionally, banks have to trust that the vendors storing security data are upholding’s the scope and terms of their

agreement. Apart from the cost of AI, the place of origin of the vendors dictates potential access to data, creating a potential war on data.

While banks are the forerunners of AI in SA, this also places banks at the mercy of vendors, creating an exceptionally unhealthy relationship of dependency. The implementation of AI becomes a double-edged sword for it can provide innumerable growth when applied with the right resources and supporting structures, but once the path to AI has been paved, there is a point of no return and organisations and teams are at the mercy of vendors. This is exacerbated by a lack of global regulation on AI, which also begs the question, is the lack of global regulation on AI intentional? The rise of XAI as a solution to the visibility is thus warranted (Tiwari, 2023).

4.4.9 Data Management and Governance

The use of AI necessitates robust data management and governance practices to ensure the security, privacy, and ethical use of data. Participants in the research project have shared their insights and experiences, highlighting the importance of data management and governance practices when using AI in cybersecurity.

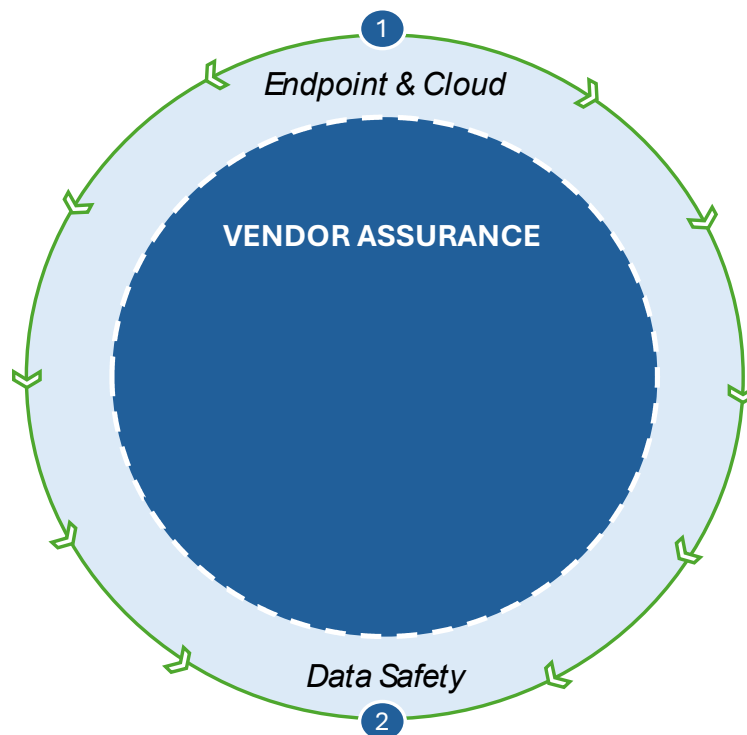


Figure:4.10: Data Management and Governance Structure

AI-powered platforms collect data from sensors on endpoints, which is then stored in cloud environments managed by the vendors. They rely on vendor assurances regarding data security. Vendors claim that the data is stored securely and is only accessible to them. However, there is no visibility into whether any metadata is moved out for model training or shared. Additionally, the team must trust that the data collected and stored by AI-powered platforms is only accessible to authorised personnel, and strict protocol is in place to ensure only members with access can access data. There are thus five main themes on data management and governance.

Data Collection and Storage

One of the fundamental aspects of data management is understanding what types of data are being collected by AI-powered platforms, where this data is stored, and who

has access to it. Participant 12 emphasized the importance of knowing the data being collected, stating, "If the quality of data is bad, the decisions ultimately are going to be bad or incorrect". This underscores the need for high-quality data to ensure accurate AI decisions. Participant 5 also highlighted the importance of understanding data collection, mentioning, "We know where they're stored as well, and we've got guarantees from the vendor that the information used. The data seen won't be used outside of our realm". This indicates the reliance on vendor guarantees to ensure data privacy and security.

Ensuring Data Privacy and Security

Data privacy and security are paramount when using AI in cybersecurity. Participant 3 pointed out the risks associated with data exposure, stating, "We are still trying to understand data leakage exposure that comes with a lot of this, and how our data is being used to train particular models, then being used globally". This highlights the need for stringent data governance practices to prevent unintended data exposure. Participant 10 echoed similar concerns, mentioning, "We don't know, like you know, as organisations, we don't really have a handle on who's putting what information into what AI agent. And yeah, might be confidential data". This indicates the challenges organizations face in ensuring data privacy and security when using AI.

Vendor Accountability and Trust

Organizations often rely on vendors to provide AI-powered solutions, making vendor accountability a critical aspect of data management and governance. Participant 2 emphasised the importance of vendor accountability, stating, "We do expect our vendors to take things seriously. So the expectation is that they are always looking into tracing these vulnerabilities". This underscores the need for vendors to maintain

high data security and privacy standards. Participant 5 also mentioned the reliance on vendor guarantees, stating, "We have to trust the vendor". This indicates the great trust organisations place in vendors to ensure data privacy and security.

Implementing Robust Security Measures

Ensuring robust security measures are in place is critical to protect AI systems from potential vulnerabilities. Participant 3 mentioned the importance of basic security hygiene, stating, "You have to make sure that proper hygiene and that system is locked down to the bare minimum of hard needs to function. Patching is done right". This includes regular patching, access control, and other security practices to safeguard AI systems.

Limiting Data Exposure

Limiting data exposure and ensuring that AI systems do not inadvertently leak sensitive information is another crucial aspect of data management and governance. Participant 3 highlighted the importance of data exposure management, stating, "We still trying to understand data leakage exposure that comes with a lot of this and how is our data being used to train particular models then being used globally". Implementing strict data governance policies can help in minimizing the risk of data leaks. There is also a concerted effort to not centralise datapoints and keep borders between intra-divisions within banks so that if one section of the banks operations is breached, it does not spread easily. Furthermore, within the internal team there is no oversight over customers personal data.

Effective data management and governance practices are essential when using AI in cybersecurity. By understanding data collection and storage, ensuring data privacy and security, holding vendors accountable, implementing oversight and validation mechanisms, providing proper training, aligning AI with ethical standards, implementing robust security measures, conducting thorough testing, and limiting data exposure, organizations can leverage the power of AI while minimizing potential risks. These practices ensure that AI systems are used responsibly and ethically, ultimately enhancing the security posture of the organization.

4.4.10 Protecting AI Systems

Protecting AI systems had the highest number of similar responses by the participants. Given that each interview was done individually and participant's did not cross-check data, the response speaks to the protocols needed to protect AI systems.

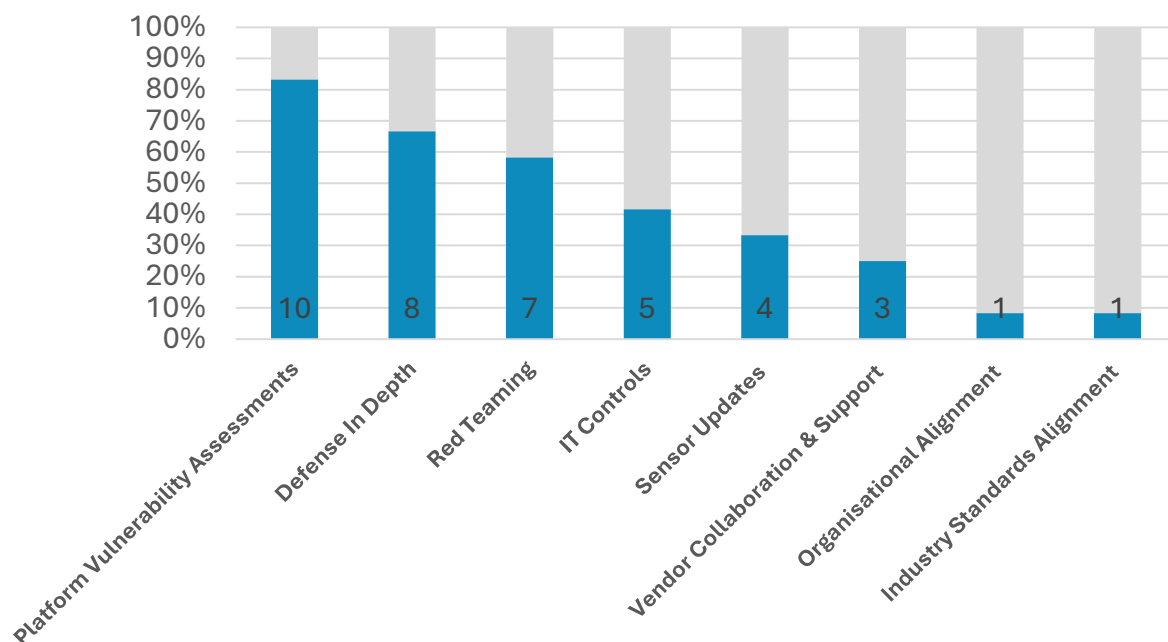


Figure 4.10: Summary themes of participants feedback to AI Protection Measures

Ten of participants stated that regular assessments are necessary to identify and address vulnerabilities in AI-powered platforms. This is over and above the banks defence-in-depth strategy. Defence in Depth Strategy' is a concept developed by the US National security agency that targets high level security of the nation's critical infrastructures.

"This approach is an information assurance concept in which multiple layers target the high level security controls, which are placed throughout an information technology (IT) system. It is also known as the (castle approach). Its intent is to provide redundancy in the event a security control fails or vulnerability is exploited. The idea behind the defense-in-depth strategy is to defend a system installed in the industrial infrastructure against any particular attack using several independent methods. It is a layering tactic, as a comprehensive approach to information and electronic security that ensure multi-layer high-level protection to avoid damage of industrial infrastructures, each layer protects the other layers and the attacker must spend more time and effort at each transition." (Abdelghani, 2020:18).

Red Teaming is a novel response to prompt injecting and data poisoning (Prompt Injection refers to adversaries prompt injections which alter the inputs to AI systems, leading to unintended or harmful outputs and AI platforms can be vulnerable to) and has been cited seven times to protect and tests AI models. General IT controls, such as change management and access controls, were noted as important for protecting AI platforms and keeping sensors updated with the latest versions to mitigate vulnerabilities and ensure effective protection. Collaboration with vendors is essential to address potential threats and ensure the security of AI-powered platforms, especially when there is no visibility. Lastly, Participant 8 and 11 mentioned that it is

crucial to validate AI decisions to ensure they align with the organisational strategy and culture and that staying aligned with industry standards is essential to ensure the security and effectiveness of AI-powered cybersecurity platforms.

4.4.11 Protecting non-AI Systems

Only four participants spoke on the protection needed on non-AI systems due to their tenure at the banks and marginal oversight over the impact AI breaches have on traditional systems. Three points emerged.

Firstly, AI-powered platforms do not protect all systems, particularly legacy applications and unsupported platforms. They rely on best-effort methods and older tools to protect themselves, which may result in security gaps. Although artificial intelligence (AI) enhances cybersecurity capabilities, it simultaneously introduces new risks. AI-driven attacks are a significant threat, necessitating further research and regulation (Folorunso et al., 2024). Secondly, traditional methods/systems not covered by AI-powered cybersecurity platforms should be protected using traditional methods, such as traditional antivirus protection, malware detection and user access management. Third, there was a marked emphasis on continued monitoring, specifically the monitoring of privileged and elevated user access, because these systems are even more susceptible to threats and attacks.

4.4.12 Does AI give companies a competitive edge

The last question asked the participants whether or not AI gives companies a competitive edge. All 12 participants agreed that AI does provide a competitive advantage and listed the following reasons:

Organizations that do not adopt AI may face higher risks and costs to catch up; AI is important in responding to modern, fast-paced cyber threats; AI-powered platforms allow for more efficient handling of large volumes of data, which is humanly impossible to manage manually; AI reduces stress on teams by handling large volumes of data and surfacing pertinent information quickly; Using AI-powered platforms provides a significant advantage over organizations that do not use AI creating a competitive advantage, as it enhances the ability to detect and respond to threats efficiently and; AI is critical for staying ahead of emerging threats and maintaining a strong security posture. Artificial intelligence transforms market strategies by facilitating personalised marketing, optimising supply chain operations, and enhancing decision-making processes through predictive analytics. Organisations use AI to automate customer service, improve product development, and identify novel business models (Hossain et al., 2024).

4.5 Conclusion

The interviews highlight the significant impact of AI on cybersecurity operations in the South African banking sector. AI has enhanced threat detection and response capabilities, automated mundane tasks, and improved the overall security posture of organisations. However, implementing AI also introduces several risks, including incorrect responses, data quality issues, and ethical concerns. To mitigate these risks, organisations must ensure proper training, implement robust security measures, and maintain transparency in AI decision-making processes. The study underscores the

importance of leveraging AI to enhance cybersecurity while carefully managing the associated risks.

Chapter 5: Conclusion & Recommendations

5.1 Introduction

The aim of this paper was to test the perception versus reality of the impact of AI in cybersecurity teams in South African big banks. To reach a stable, unbiased, informed and comprehensive conclusion a rigorous process of critically assessing the literature on the application of AI tools in banks was analysed. This was performed through the structured lens of the TOE framework and Perception Theory. The former framework covered the aspects of an internal organisation and some of the external environment. It was found that AI has a significant anthropomorphic impact on the technological aspects of banks' CS teams. They curate their strategies and outlooks in tandem with organisational goals and limit conflicts of interest to the greatest degree. The environmental context proved interesting in that there was no international and national regulation (both legislation and policies) on the direction, capability and parameters of operability for AI in general.

Interestingly, via Perception Theory, a disparity between academic literature and the experience of professionals within CS operations utilising AI tools, software and infrastructure emerged. Their internal views challenged many of the notions such as, AI displacing the role of CS staff and engineers; AI being a one-stop-shop solution for security; AI requiring a blanket approach in application; the oversight over data access and safety; and the strict protocols required internally to ensure defence mechanisms function at their optimum. Cybersecurity was both a defence and protection against external forces, and an overlap into internal ones to ensure internal malice does not affect future outlooks and prospects. Based on the predilections of Perception Theory

that grounded itself in phenomenological and ontological credibility and reliability, the direct experience of the participants held greater weight viz the quantitative conjecture and deductions as per the critical and comprehensive findings of the Literature Review. The adoption of Perception Theory was novel to AI-related literature as most technology-linked frameworks are technological, and there has not been an interdisciplinary and cross-pollination of ideologies before. The approach and the insights demonstrate the richness that inter-disciplinary studies can offer.

This chapter provides conclusions, recommendations, limitations, and suggestions for further research based on the study's outcomes. It will first cover conclusions per research question and then conclude in relation to the conceptual framework (linked to step 6 of the research framework). Research limitations are discussed, and recommendations for future research are made.

5.2 Conclusions regarding Research Question 1

Concerning the first research question, *'To what extent can artificial intelligence serve as an effective solution for eradicating and preventing cybercrime within the South African banking sector?'* Via a deductive analysis in the Literature Review and personal experience of the interviewees, it becomes markedly and emphatically clear that AI fights cyber-crime but does not eradicate it. AI is an anthropomorphic tool that is able to grow, develop and advance based on the data input, but parallel to this advancement is the deployment of AI, sometimes further in advancement to AI solutions, for cybercrime. AI is a reactive tool to cyber-crime, and it develops 'in-response-to'.

AI is used for cyber-defence to protect data, to protect networks, software of both the banks infrastructure and that of the customers. AI is a scalable tool and not a one-size-fits-all solution, and, as such, even within CS operations, different teams employ varying degrees of automation depending on function, processes and strategic direction and intent. Similar to the findings in the Literature Review, AI is not a turn-key solution and does not possess the power and ability to eradicate cyber-crime in totality. However, based on the benefits and risks Folorunso et al., 2024, it is evident that AI is a solution for cyberattacks. Without AI-related tools, banks and institutions are prone to trillions of dollars in losses. While AI is not a 'silver-bullet' in the paradigm that it eliminates cybercrime at the root, being an anthropomorphic entity and system of functioning, as digitisation evolves, so will crimes and so will defence. AI cannot eradicate crime completely but is the most efficient tool and system in cyber-crime detection and prevention.

5.3 Conclusions regarding Research Question 2

In response to the RQ2, *What are the factors that contribute to the decrease of the CS staff complement in banks to improve 'organisational effectiveness'?* The common perception of AI is that AI will replace manual human labour. This view also upheld from the findings in the Literature Review but as witnessed in the interviews, there is a skills shortage of well-trained skilled AI-users, and the banks and CS team have not laid off any personnel. Depending on the nature of work and on the team, either the work itself changes or the roles change, but thus far, there has not been any incidences of AI shedding roles as a result of itself.

The cybersecurity team itself may be small and specialized but as AI develops and evolves so do the team skills requirements, roles and dynamics. Whilst this evolutionary process within banks is seen as rapid from a topographical perspective, internally, CS spend years creating, testing, piloting and launching new solutions. These solutions are iterative and require constant upkeep. CS teams must defend across all these paradigms simultaneously, with an *ad infinitum* timer and skilled staff will always be in demand. The findings from the interviews demonstrate that AI did not job-shed in the CS team but infact creates an on-going demand for specialized skills.

5.4 Conclusions regarding Research Question 3

Interestingly, the findings of RQ3 (*What evaluation criteria or assessment frameworks are employed by banks to assess the effectiveness of AI-CS capabilities in reducing cybercrime through preventative measures?*) speak to the magnitude and scale of the work being complete in this paper due to no literature being available on data collection and none of the interview participants speaking about how data is collected, where exactly is it stored and how it is analysed for future positioning. Given that SA is the country with the third highest number of cyberattacks, there is an assumption that stronger protocols and tools for measuring efficacy are certainly at play. However, it is not available for internal and external consumption. As such, the answer to the research problem will have to remain suspended indefinitely.

In response to the absence of data and a clear, glaring knowledge gap, a framework has been created to ascertain the scale of the adoption of AI and to analyse the risks-reward ratio via a holistic divisional and organisational view. This is posited in the Recommendations (5.6).

5.5 Limitations of this study

Despite efforts to mitigate bias throughout the interview process and when analysing the data, the subjective nature of qualitative research cannot remove the natural bias itself in totality. Fragments of bias will always inform the articulation of questions, the perceptions of participants responses during interviews and the curation of findings during analysis. The researcher was mindful that social desirability could have influenced participants (Dodou & de Winter, 2014). Social desirability can skew participants' responses if the participants want to provide socially acceptable answers, viz-a-viz expressing their true sentiment and feelings based on their lived experience. The limitation of social desirability was off set by the researcher sending the bespoke questions to the participants beforehand and having privacy, confidentiality and anonymity declarations signed.

A third limitation came in the form of time constraints. The interviews were conducted within a time frame, and this disallowed the researcher from engaging in any follow-up questions post-analysis to enrich the body of research. Whilst they could have been further iterations of the process to develop an extra layer of theoretical, technical and ideological richness, the ever-evolving nature of AI and its interface with CS will provide a substantive starting point in the literary journey for the impact of AI on CS operations in South African banks (with the findings applying to the financial services sector as a whole).

A fourth limitation is that the research does not consider how AI assists CS and the broader organisational goals. Whilst CS operations may endure a benign or active

system, the ramifications on organisational goals were not looked at. What could be a challenge within CS operations may count as a success within the bank's larger operating and business model and outcomes. Therefore a top-down approach in ushering alternative technologies may not be welcome. Additionally, the impact of a decreased CS staff complement may affect CS operations but assist the bank in reaching its staff goal and annual Key Performance Indicator (KPI). This, notably, is a very broad limitation, but should also be taken into account given the extent of the specificity of the subject-matter.

5.6 Recommendations

The perception of AI has a marked influence on its uptake and implementation within a division or organisation. One of the alarming and looming gaps was the lack of assessments that exist on AI implementation, especially within CS. Accordingly, this paper has created a bespoke framework and assessment criteria to assist users and help the scant and thin body of literature on the subject-matter. The framework is based on conclusive findings from the literature review and was curated based on the actual experience of CS officers within banks, adding a dimension of credibility and reliability that conjecture cannot. It also encapsulates the findings from TOE and PT.

To fill this knowledge gap, this paper proposes a novel JAI-Perception Theory framework that shows the start-to-end journey of data input and storage and the tools and parties needed to use AI effectively. Depending on the nature of an organization, department, team, and person, AI is a scalable solution with associated risks and benefits. At every point in the AI journey, data interfaces with new forms of AI tools, software, and solutions.

The JAI-Perception Theory framework has two parts: part one, which explains the location of AI within a department's processes and operations, and part two, which analyses the use of AI across the division to ascertain the risk/reward. It is a clear-cut system that helps users easily locate the type of AI needed and ascertain the impact that AI will have should they wish to implement certain tools and features.

JAI-Perception Framework

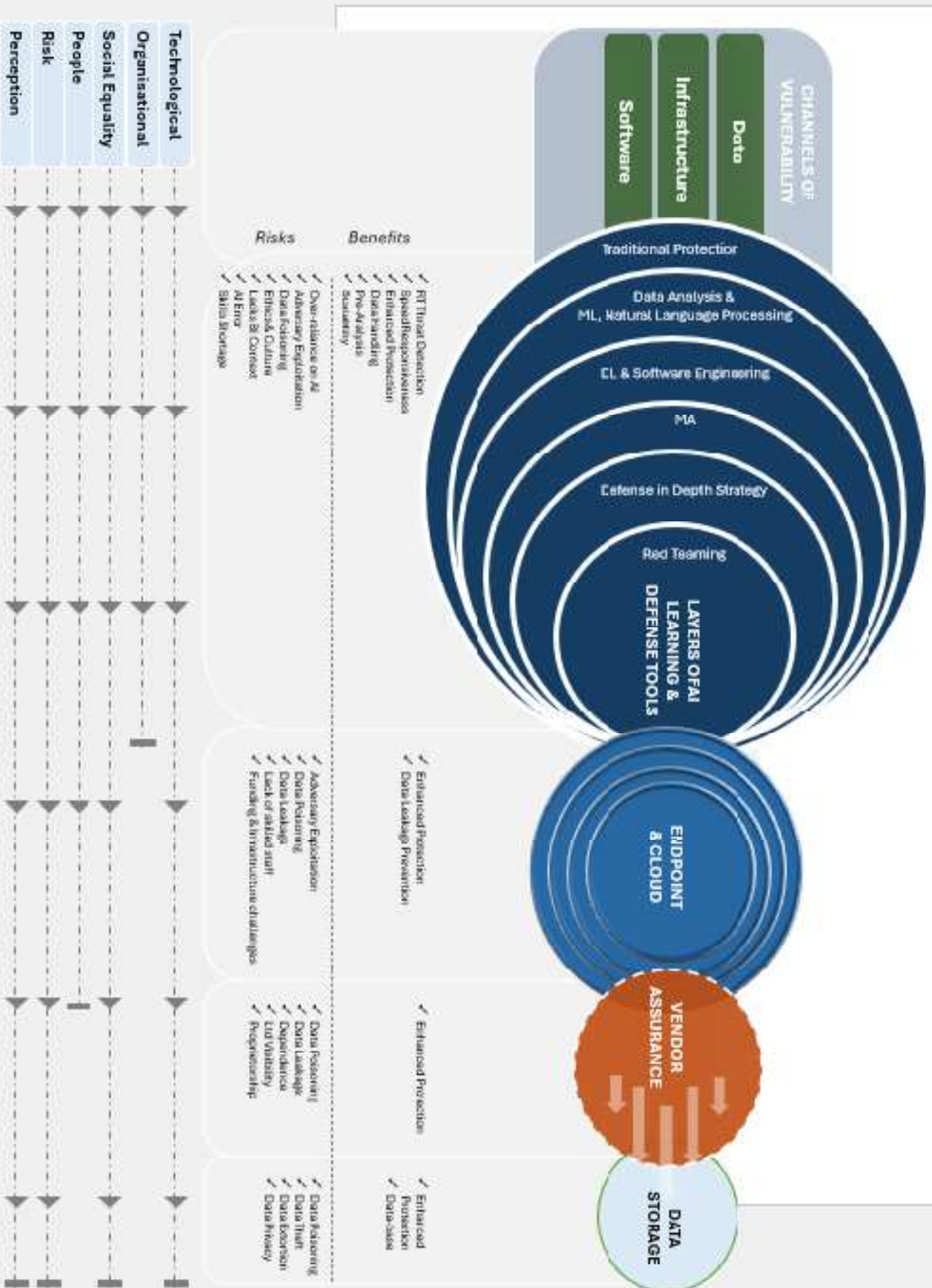


Figure 5.1: JAI Perception Framework

JAI-Perception Framework											
TECHNOLOGICAL		ORGANISATIONAL		SOCIAL EQUALITY		PEOPLE		RISK		PERCEPTION	
<i>Considerations that affect the rationale for implementation, size of organization/level of defense needed and skilled staff availability</i>		<i>CS aligning with the organisations goals and strategy direction – people, budget, business and operating model, risk appetite, human resource</i>		<i>The external disparities and inequality AI creates depending on level of AI applications</i>		<i>The seniority of staff required and willingness of CS officers to changes specialization and roles</i>		<i>The multiple types of risk that exist from human resources, the AI itself, breach of data and systems, third party weaknesses</i>		<i>The views users and non-users have of AI that affects the adoption, implementation, use and management of AI</i>	
	Y/N		Y/N		Y/N		Y/N		Y/N		Y/N
What is the size of the organization		Is the company technologically driven		Can users keep up with the pace of AI implementation		Do CS officers feel safe in their roles or do they feel replaceable by AI		Is there a cyber attack insurance		Is there a perception vs reality guiding module	
What level of protections exist		Where does the company stand in comparison to competitors		Will there be stark disparity between banks and other organizations who lag in AI implementation and prevent collaboration		Is there a large disparity between junior and senior roles		Do risk managers understand cyber risk		Do staff understand the risks associated with the implementation of management tools	
Is there scope to employ junior and senior engineers		Are there governance charters in place		Is there is disparity between the AI and the engineers		Are all engineers capable and willing to constantly mutate in their roles		Is there a creation of risk mitigation tools that are manageable by the company			
Does the organization have budget for AI-implementation scoping		Is there is a risk management structure in place		Will cyber poverty be supported							
What defense systems are currently being used		Will the company support the growth of the CS team									
Has any third-party vendors / service providers been identified		Does the implementation of AI align with the companies' values and growth trajectory									
Are competitors using AI and what AI		Is there greater budget allowance									
Will AI improve the company's security posture											
What level of automation currently exists											

Figure 5.2: JAI Perception Framework

The above Perception Framework is a template users can use to guide the efficacy of AI before and during implementation. The structure can be adapted to any size team or organisation and populated according to any industry. Importantly, it merely provides a guiding pathway to navigate the AI journey and, like AI itself, is fully customisable and malleable.

5.7 Suggestions for further studies

What does the future of CS operations look like with evolving AI? Technologically, there is a shift towards wanting to not depend on vendors to address vulnerabilities, and for

senior CS engineers to understand the telemetry and internal rationalising of how the algorithm works - XAI. Organisationally, there is a demand to understand if data is being kept safely and not leaked. Environmentally, there are legislations and policies in the pipeline for implementation, but there is no concerted initiative on capping the ongoing deployment of AI within banks. However, specifically in South Africa, this exacerbates an already severely distinct divide between developed and undeveloped sectors of the economy, and other financial institutions and non-financial institutions will have to pay a premium price to 'catch-up' and stay relevant. There is a large cohort of businesses and people in society will be rendered uniform, redundant and technologically inactive because of the advances. Banks will need to adopt a bi-pronged approach of catering to technologically advanced customers and traditional legacy-style customers. Alongside this potential shift, the role of CS officers will too become bifurcated. There will continue to be perceptions about the impact of AI, but from the findings of the study, security personnel with actual experience in the industry are more reliable, trustworthy sources, vis-à-vis online conjecture. Interpretivism, which explores the true lived experience of people with knowledge of specific phenomena, is the suggested approach, especially in niche, technical topics, spheres and environments.

In summation:

- How can banks cultivate long-term talent management programs to minimise third-party vendor risks and ensure continuity
- How can banks capitalise on telemetry to create custom solutions, software and hardware that offers full oversight of a solution

- What is the socio-economic impact of AI on the social inequity divide in South Africa
- What is the impact of cyber-poverty on CS teams in SA
- How can South African banks buffer themselves from the data hegemony fight by US-China
- What public standardised policies should be in place to ensure the separation of powers and the prevention and misuse of AI data intelligence to limit crime within banks
- What novel change management solutions need to be created to aid engineers and CS personnel navigate a ever-evolving environment.

References

- Abdelghani, T. (2020). Implementation in Depth Strategy to Secure Industrial Control system in Critical Infrastructure. *American Journal of Artificial Intelligence*, 3(2), 17-22. <https://doi.org/10.11648/j.ajai.20190302.11>
- Aggarwal, D. S. (2023). Role of AI in cyber security through Anomaly detection and Predictive analysis. *Journal of Informatics Education and Research*, 3(2). <https://doi.org/https://doi.org/10.52783/jier.v3i2.314>
- AL-Dosari, K. F. (2022, August). Artificial Intelligence and Cyber Defense System for Banking Industry: A Qualitative Study of AI Applications and Challenges. *Cybernetics and Systems: An International Journal*, 55(2), 30. <https://doi.org/https://doi.org/10.1080/01969722.2022.2112539>
- Aliyari, M. (2021). Securing Industrial infrastructure against cyber-attacks using machine learning and artificial intelligence at the Age of industry 4.0. *Turkish Journal of Computer and Mathematics Education*, 12(11), 6581-6594. <https://doi.org/https://doi.org/10.17762/turcomat.v12i11.7074>
- Alsamiri, J. a. (2019). Internet of things cyber attacks detection using machine learning. *International Journal of Advanced Computational Science Application*, 10(12), 627–34. <https://doi.org/https://doi.org/10.14569/ijacsa.2019.0101280>
- Amoresano, K. &. (2023). Human Error - A Critical Contributing Factor to the Rise in Data Breaches: A Case Study of Higher Education. *HOLISTICA – Journal of Business and Public Administration*, 14(1), 23. <https://doi.org/http://dx.doi.org/10.2478/hjbpa-2023-0007>
- Ampong, K. (2024). Artificial intelligence research: A review on dominant themes, methods, frameworks and future research directions in Telematics and

- Informatics Reports. *Telematics and Informatics Reports*, 14(8), 100127.
<https://doi.org/https://doi.org/10.1016/j.teler.2024.100127>
- Andriu, A. V. (2023). Adaptive Phishing Detection: Harnessing the Power of Artificial Intelligence for Enhanced Email Security. *Romanian Cyber Security Journal*, 5(1), 7. <https://doi.org/https://doi.org/10.54851/v5i1y202301>
- Arasada, S. (2021, 8 31). *These four challenges in adopting machine learning can lower your ROI and sabotage success*.
<https://doi.org/https://www.forbes.com/sites/forbestechcouncil/2021/08/31/these-four-challenges-in-adopting-machine-learning-can-lower-your-roi-and-sabotage-success/?sh=2818cdf47c4a>
- Asiri, M. (2023). Hybrid metaheuristics feature selection with stacked deep learning-enabled cyber-attack detection model. *Computer Systems Science and Engineering*, 45(2), 1679–1694.
<https://doi.org/https://doi.org/10.32604/csse.2023.031063>
- Asiri, S. X. (2024). PhishingRTDS: a real-time detection system for phishing attacks using a Deep Learning model. *Computers & Security*, 141, 103843.
<https://doi.org/https://doi.org/10.1016/j.cose.2024.103843>
- Atkins, L. B. (2024, March). *The cyber clock is ticking: Derisking emerging technologies in financial services*. Retrieved from McKinsey & Company:
<https://www.mckinsey.com/capabilities/risk-and-resilience/our-insights/the-cyber-clock-is-ticking-derisking-emerging-technologies-in-financial-services>
- Atkins, L. B. (2024). *The cyber clock is ticking: Derisking emerging technologies in financial services*. Institute of International Finance and McKinsey & Company.

- Attaran, M. &. (2018). Machine learning: The new 'big thing' for competitive advantage. *International Journal of Knowledge Engineering and Data Mining*, 5(4), 277-305. <https://doi.org/10.1504/IJKEDM.2018.10015621>
- Baker, J. (2011). The Technology–Organization–Environment Framework. *Information Systems Theory. Integrated Series in Information Systems*, 28, 231 - 245. https://doi.org/https://doi.org/10.1007/978-1-4419-6108-2_12
- Balushi, K. A. (2018). The Use of Online Semi-Structured Interviews in Interpretive Research. *International Journal of Science and Research (IJSR)*, 7(4), 1 - 7. <https://doi.org/10.21275/ART20181393>
- Bank, S. A. (2022, December). *Consultation paper on cyber-resilience*. <https://doi.org/https://www.resbank.co.za/content/dam/sarb/what-we-do/payments-and-settlements/consultation-documents/Cyber%20Resilience%20consultation%20paper%20published%20December%202022.pdf>
- Barik, K. M.-S. (2022). Cybersecurity deep: approaches, attacks dataset, and comparative study. *Applied Artificial Intelligence*, 36(1), 1 - 24. <https://doi.org/https://doi.org/10.1080/08839514.2022.2055399>
- Bearman, M. (2019). Focus on Methodology: Eliciting rich data: A practical approach to writing semi-structured interview schedules. *Focus on Health Professional Education A Multi-Professional Journal*, 20(3), 1 - 11. <https://doi.org/https://doi.org/10.11157/fohpe.v20i3.387>
- Benhang, H. F. (2024). The Safety Measures on Electronic Banking Operations and Its Cybersecurity: Basis for Enhancement Plan. *Journal of Business and Management Studies*, 6(4), 24. <https://doi.org/10.32996/jbms.2024.6.4.20>

- Bergelson, I. T. (2022). Best Practices for Reducing Bias in the Interview Process, *Current Urology Reports. National Library of Medicine*, 23, 319 - 325.
<https://doi.org/10.1007/s11934-022-01116-7>
- Božić, V. &. (2023). Measuring organizational performance. *Researchgate*.
<https://doi.org/10.13140/RG.2.2.20225.12642>
- Braun, V. &. (2017). Thematic analysis. *The Journal of Positive Psychology* , 12(3), 297-298. <https://doi.org/https://doi.org/10.1080/17439760.2016.1262613>
- Brinkmann, S. &. (2018). Doing Interviews. *Sage Research Methods* , 1 - 45.
<https://doi.org/https://doi.org/10.4135/9781529716665>
- Campbell, S. G. (2020). Purposive sampling: complex or simple? Research caseexamples. *Sage Journals*, 25(8), 10.
<https://doi.org/https://doi.org/10.1177/1744987120927206>
- Cooper, S. &. (2007). Generic qualitative research: a design for qualitative research in emergency care? *Emergency Medicine Journal*, 24(12), 816 - 819.
<https://doi.org/10.1136/emj.2007.050641>
- Cope, D. G. (2014). Methods and Meanings: Credibility and Trustworthiness of Qualitative Research. *Oncology Nursing Forum*, 41(1), 89 - 91.
<https://doi.org/http://dx.doi.org/10.1188/14.ONF.89-91>
- Cowling, N. (2024, March 27). *Total assets of the banking sector as percentage of gross domestic product (GDP) in South Africa from 2013 to 2020*. Retrieved from Statista: <https://www.statista.com/statistics/1349337/bank-assets-as-share-of-gdp-in-south-africa/>
- CrowdStrike. (2024). *CrowdStrike 2024 Global Threat Report*. Retrieved from [www.crowdstrike.com: https://www.crowdstrike.com/en-us/resources/reports/crowdstrike-2024-global-threat-report/](https://www.crowdstrike.com/en-us/resources/reports/crowdstrike-2024-global-threat-report/)

- Cucu, C. G. (2019). Current technologies and trends in cybersecurity and the impact of artificial intelligence. *eLearning & Software for Education*, 2, 208 - 214.
<https://doi.org/10.12753/2066-026X-19-099>
- Dawson, M. (2021). Cybersecurity impacts for artificial intelligence use within industry 4.0. *Scientific Bulletin*, 26(1), 1 - 8. <https://doi.org/10.2478/bsaft-2021-0003>
- De Azambuja, A. J. (2023). Artificial Intelligence-Based Cyber Security in the Context of Industry 4.0—A Survey. *Electronics* 2023, 12(8), 18.
<https://doi.org/https://doi.org/10.3390/electronics12081920>
- Deloitte. (2025). *How Artificial Intelligence is Transforming the Financial Services Industry*. Retrieved from Deloitte: <https://www.deloitte.com/ng/en/services/risk-advisory/services/how-artificial-intelligence-is-transforming-the-financial-services-industry.html>
- Demuth, A. (2013). *Perception Theories*. Faculty of Philosophy and Arts · Trnava University in Trnava.
https://doi.org/https://ff.truni.sk/sites/default/files/publikacie/demuth_perception_theories_1.1.pdf
- Dodou, D. &. (2014). Social desirability is the same in offline, online, and paper surveys: A meta-analysis. *American Psychological Association*, 36, 487 - 495.
<https://doi.org/https://doi.org/10.1016/j.chb.2014.04.005>
- Edmondson, A. C. (2007). Methodological fit in management field research. *Academy of Management Review*, 32(4), 1155-1179.
<https://doi.org/https://doi.org/10.5465/amr.2007.26586086>
- Eian, I. Y. (2020, September 26). *Cyber-attacks in the era of COVID-19 and possible solution domains*. Retrieved from Europe PMC:
<https://europepmc.org/article/ppr/ppr219480>

- Eswaran, M. H. (2023). Survey of Cyber security approaches for Attack Detection and Prevention. *Turkish Journal of Computer and Mathematics Education (TURCOMAT)*, 12(2), 3436–3441.
<https://doi.org/https://turcomat.org/index.php/turkbilmat/article/view/2406>
- Everington, J. (2024, July 9). *Standard Bank continues dominance of South African banking Weaker rand and sluggish domestic economy impact local lenders' rankings*. Retrieved from The Banker:
<https://www.thebanker.com/content/b2e6edc8-3d63-569a-81b7-1cf67f786687#:~:text=The%20six%20lenders%20listed%20in,and%20Tier%201%20capital%20base>
- Eze, C. S. (2024). Analysis and Prevention of AI-Based Phishing Email Attacks. *Electronics* 2024, 13(10), 13.
<https://doi.org/https://doi.org/10.3390/electronics13101839>
- Eze, C. S. (2024). Analysis and Prevention of AI-Based Phishing Email Attacks. *Electronics*, 13(10), 1893.
<https://doi.org/https://doi.org/10.3390/electronics13101839>
- Ferrara, E. (2023). Fairness and Bias in Artificial Intelligence: A Brief Survey of Sources, Impacts, and Mitigation Strategies. *Sci* 2024, 6(1), 15. <https://doi.org/https://doi.org/10.3390/sci6010003>
- Ghafar, Z. N. (2024). The evaluation research: A Comparative Analysis of Qualitative and Quantitative Research Methods. *Journal of Language, Literature, Social and Cultural Studies*, 2(1), 1 - 10.
<https://doi.org/https://doi.org/10.58881/jllscs.v2i1.122>

- Hasani, T. O. (2023). Evaluating the adoption of cybersecurity and its influence on organizational performance. *SN Business & Economics* , 3(97), 38.
<https://doi.org/https://doi.org/10.1007/s43546-023-00477-6>
- Hoory, L. A. (2024, June 5). *What Is A Cyber Attack? Definition, Types & Prevention*. Retrieved from https://www.forbes.com/https://www.forbes.com/advisor/business/what-is-cyber-attack/#common_types_of_cyber_attacks_section
- Ibrahim, M. A. (2023, December). Incorporating the Future: Optimizing Cybersecurity through Seamless Integration of Artificial Intelligence. *International Journal for Electronic Crime Investigation*, 7(4), 10.
<https://doi.org/https://doi.org/10.54692/ijeci.2023.0704166>
- Institute, C. P. (n.d.). *The Issue*. Retrieved from Cyber Poverty: Unraveling its Complex Impact on Individuals, Companies, Society, Countries, and Global Trade: <https://www.cyberpovertyline.org/the-issue>
- Jada, I. &. (2024). The impact of artificial intelligence on organisational cyber security: An outcome of a systematic literature review. *Data and Information Management*, 8(2), 15.
<https://doi.org/https://doi.org/10.1016/j.dim.2023.100063>
- Jain, N. (2023, July 7). *What is Qualitative Research Design? Definition, Types, Methods and Best Practices*. Retrieved from IDEASCALE: <https://ideascale.com/blog/qualitative-research-design/#:~:text=Qualitative%20research%20design%20typically%20involves,perspectives%2C%20experiences%2C%20and%20contexts>
- Josselson, R. (2013). *Interviewing for qualitative inquiry: A relational approach*. Guilford Press.

- Kaur, R. G. (2023). Artificial intelligence for cybersecurity: literature review and future research directions. *Information Fusion*, 97, 1 - 29.
<https://doi.org/https://doi.org/10.1016/j.inffus.2023.101804>
- Kayode-Ajala, O. (2024). Applications of Cyber Threat Intelligence (CTI) in Financial Institutions and Challenges in Its Adoption. *Applied Research in Artificial Intelligence and Cloud Computing*, 7(6), 21.
<https://doi.org/https://researchberg.com/index.php/araic>
- Kelly, S. K.-T. (2023). What factors contribute to the acceptance of artificial intelligence? A systematic review. *Telematics and Informatics*, 77, 1 - 33.
<https://doi.org/https://doi.org/10.1016/j.tele.2022.101925>
- Kumar, S. G. (2023, August 31). Artificial Intelligence: Revolutionizing Cyber Security in the Digital Era. *Journal of Computers, Mechanical and Management*, 2(3), 12. <https://doi.org/https://doi.org/10.57159/gadl.jcmm.2.3.23064>
- Kuntla, G. T. (2021). Security and privacy in machine learning: A survey. *Issues in Information Systems*, 22(3), 224 - 240.
https://doi.org/https://doi.org/10.48009/3_iis_2021_242-258
- Laryeafio, M. N. (2023). Ethical consideration dilemma: systematic review of ethics in qualitative data collection through interviews. *Journal of Ethics in Entrepreneurship and Technology*, 3(2), 17.
<https://doi.org/https://www.emerald.com/insight/publication/issn/2633-7436>
- Mohamed, N. (2023). Current trends in AI and ML for cybersecurity: A state-of-the-art survey. *Cogent Engineering*, 10(2), 30.
<https://doi.org/https://doi.org/10.1080/23311916.2023.2272358>
- Morovat, K. &. (2020). A survey of artificial intelligence in cybersecurity. Proceedings. *2020 International Conference on Computational Science and Computational*

- Intelligence (CSCI)* (pp. 109–115). Las Vegas: IEEE.
<https://doi.org/https://doi.org/10.1109/CSCI51800.2020.00026>
- Mugwagwa, A. B. (2024). Cybersecurity strategy: future proof cybersecurity for small to medium enterprises in South Africa. *INTERNATIONAL JOURNAL OF RESEARCH IN BUSINESS AND SOCIAL SCIENCE*, 13(4), 10.
<https://doi.org/https://doi.org/https://doi.org/10.20525/ijrbs.v13i4.3308>
- Musa, N. M. (2024). Machine learning and deep learning techniques for distributed denial of service anomaly detection in software defined networks—current research solutions. *IEEE Access*, 12, 17982-18011.
<https://doi.org/https://doi.org/10.1109/ACCESS.2024.3360868>
- Naeem, M. O. (2023). A Step-by-Step Process of Thematic Analysis to Develop a Conceptual Model in Qualitative Research. *International Journal of Qualitative Methods*, 22, 15. <https://doi.org/https://doi.org/10.1177/16094069231205789>
- Nordin, N. I. (2021). A comparative analysis of metaheuristic algorithms in fuzzy modelling for phishing attack detection. *Indonesian Journal of Electronic Engineering and Computational Science*, 23(2), 1146 –1158.
<https://doi.org/https://doi.org/10.11591/ijeecs.v23.i2.pp1146-1158>
- Ofosu-Ampong, K. (2024). Artificial intelligence research: A review on dominant themes, methods, frameworks and future research directions. *Telematics and Informatics Reports*, 14, 1 - 11.
<https://doi.org/https://doi.org/10.1016/j.teler.2024.100127>
- Patton, M. (1999). Enhancing the quality and credibility of qualitative analysis. *Health Services Research*, 34, 1189 - 1208.
- Paul, E. O. (2023). Cybersecurity strategies for safeguarding customers data and preventing financial fraud in the united states financial sectors. *International*

- Journal on Soft Computing*, 14(3), 16.
<https://doi.org/https://aircconline.com/ijsc/V14N3/14323ijsc01.pdf>
- Radanliev, P. (2025). AI Ethics: Integrating Transparency, Fairness, and Privacy in AI Development. *Applied Artificial Intelligence*, 39(1), 2 - 42.
<https://doi.org/https://doi.org/10.1080/08839514.2025.2463722>
- Raimundo, R. &. (2021). The impact of artificial intelligence on data system security: A literature review. *Sensors*, 21(21), 7029.
<https://doi.org/https://doi.org/10.3390/s21217029>
- Rizvi, M. (2023). Enhancing cybersecurity: The power of artificial intelligence in threat detection and prevention. *International Journal of Advanced Engineering Research and Science (IJAERS)*, 10(5), 55 - 60.
<https://doi.org/https://dx.doi.org/10.22161/ijaers.105.8>
- SABRIC. (2023). *Annual crime statistics 2023*. Retrieved from <https://sabric.co.za/https://sabric.co.za/media/vjyn5f4d/sabric-annual-crime-stats-2023-2.pdf>
- Salem, A. A. (2024). Advancing cybersecurity: a comprehensive review of AI-driven detection techniques. *Journal of Big Data*, 11(105), 1 - 38.
<https://doi.org/https://doi.org/10.1186/s40537-024-00957-y>
- Saunders, N. L. (2023). *Research Methods for Business Students*. PEARSON.
- Sharma, L. J. (2023). Navigating the Research Landscape: A Guide to the Selection of the Right Research Design. *International Research Journal of MMC (IRJMMC)*, 4(1), 64 - 78.
<https://doi.org/https://doi.org/10.3126/irjmmc.v4i1.51863>
- Shaveta. (2023). A review on machine learning. *International Journal of Science and Research Archive*, 9(1), 281–285.
<https://doi.org/https://doi.org/10.30574/ijsra.2023.9.1.0410>

- Shrivastava, A. P. (2024). Artificial Intelligence (AI): Evolution, Methodologies, and Applications. *International Journal for Research in Applied Science & Engineering Technology (IJRASET)*, 12(4), 5501 - 5505.
<https://doi.org/https://doi.org/10.22214/ijraset.2024.61241>
- Sontan, A. D. (2024). The intersection of Artificial Intelligence and cybersecurity: Challenges and opportunities. *World Journal of Advanced Research and Reviews*, 21(02), 1720–1736.
<https://doi.org/https://doi.org/10.30574/wjarr.2024.21.2.0607>
- Sun, H. &. (2022). Enterprise Information Security Management Using Internet of Things Combined with Artificial Intelligence Technology. *Computational Intelligence and Neuroscience*, 1 - 16.
<https://doi.org/https://doi.org/10.1155/2022/7138515>
- Tiwari, R. (2023). Explainable AI (XAI) and its Applications in Building Trust and Understanding in AI Decision Making. *International Journal of Scientific Research in Engineering and Management (IJSREM)*, 7(1), 13.
<https://doi.org/10.55041/IJSREM17592>
- Villegas-Ch, W. &.-O. (2023, May 10). Toward a Comprehensive Framework for Ensuring Security and Privacy in Artificial Intelligence. *Electronics* 2023, 12(18), 17. <https://doi.org/https://doi.org/10.3390/electronics12183786>
- Welukar, J. &. (2021). Artificial Intelligence in Cyber Security - A Review. *International Journal of Scientific Research in Science and Technology*, 8(6), 488-491.
<https://doi.org/https://doi.org/10.32628/IJSRST218675>
- Wilkins, J. (2018). Is artificial intelligence a help or hindrance? *Network Security*(5), 18-19. [https://doi.org/https://doi.org/10.1016/S1353-4858\(18\)30046-1](https://doi.org/https://doi.org/10.1016/S1353-4858(18)30046-1)

Wittgenstein, L. (2010). *Tractatus Logico-Philosophicus*. KEGAN PAUL, TRENCH,
TRUBNER & CO., LTD.

<https://doi.org/https://www.gutenberg.org/files/5740/5740-pdf.pdf>

Xu, Y. e. (2021). Artificial intelligence: A powerful paradigm for scientific research. *The
Innovation*, 2(4), 2666-6758.

<https://doi.org/https://doi.org/10.1016/j.xinn.2021.100179>

Zlateva, P., Steshina, L., Velez, D., & Petukhov, I. (2024). *A Conceptual Framework for
Solving Ethical Issues in Generative Artificial Intelligence*. institut f r ost und s
dosteuropaforschung. <https://doi.org/10.3233/faia231182>

APPENDIX A

Interview Guide

Standard Participant Questions

1. What is your current role in the bank?
2. Which department in the security office do you work at?
3. How long have you worked in cybersecurity?
4. In your department, have you adopted any AI-powered platforms?
5. How long have you been working with AI powered solutions in your department?

Question formulation

1. What are the benefits in using AI-Powered cybersecurity platforms?
2. In what capacity are you using these AI-Powered cybersecurity platforms?
3. Are there any risks associated with the use of AI-Powered cybersecurity platforms?
4. If any, what are those risk posed by AI-Powered cybersecurity platforms?
5. How can the risks / challenges posed by AI-Powered cybersecurity platforms be mitigated?
6. Did the use of AI in cybersecurity improve the security posture of the bank?
7. Were any activities automated by using AI-Powered cybersecurity platforms?
8. What were the impact of using AI-Powered cybersecurity platforms on your cybersecurity engineers?
9. Have cybersecurity engineers been formally trained to use AI-Powered cybersecurity platforms?
10. Did the adoption and use of AI-Powered cybersecurity platforms change any of your cybersecurity engineers job roles and functions?

11. Do you have visibility of some of the decisions made by the AI-Powered cybersecurity platforms and why the decision was made?
12. Do you know what type of data the AI-Powered cybersecurity platforms are collecting, where it is stored and who have access to it?
13. Can the AI-Powered cybersecurity platforms be exposed to vulnerabilities?
14. If yes, how do you protect the AI-Powered cybersecurity platforms from these potential vulnerabilities?
15. Do the AI-Powered cybersecurity platforms protect all your in-scope systems or applications and is there any exclusions?
16. How do you protect the platforms and applications where the AI-Powered cybersecurity platforms are not able to work or function on?
17. Do you think the use of AI-Powered cybersecurity platforms gives your organization an advantage over those that does not use AI?

APPENDIX B

Participants Interview questions according to Research and CS themes.

Research Questions		Themes	P1	Guiding Questions
Is AI a true solution in the eradication and prevention of cyber-crime in the South African banking sector?	1	Benefits of AI in CS		What are the benefits in using AI-Powered cybersecurity platforms?
	2	Current Use of AI in CS		In what capacity are you using these AI-Powered cybersecurity platforms?
	3	Risks Associated with AI in CS		- Are there any risks associated with the use of AI-Powered cybersecurity platforms? - If any, what are those risks posed by AI-powered cybersecurity platforms? - Can the AI-Powered cybersecurity platforms be exposed to vulnerabilities? - Do the AI-Powered cybersecurity platforms protect all your in-scope systems or applications and is there any exclusions? - How do you protect the platforms and applications where the AI-Powered cybersecurity platforms are not able to work or function on?
	4	Mitigating AI Risks		- How can the risks / challenges posed by AI-Powered cybersecurity platforms be mitigated? - If yes, how do you protect the AI-Powered cybersecurity platforms from these potential vulnerabilities? - How do you protect the platforms and applications where the AI-Powered cybersecurity platforms are not able to work or function on?
	5	Impact of AI on Security Posture		- Did the use of AI in cybersecurity improve the security posture of the bank? - Do you think the use of AI-Powered cybersecurity platforms gives your organization an advantage over those that does not use AI?

Is it wise and feasible to decrease the CS staff complement in banks given the absence of ethics in AI?	6	Impact on CS personnel		○ Were any activities automated by using AI-Powered cybersecurity platforms? ○ What were the impact of using AI-Powered cybersecurity platforms on your cybersecurity engineers? ○ Have cybersecurity engineers been formally trained to use AI-Powered cybersecurity platforms? ○ Did the adoption and use of AI-Powered cybersecurity platforms change any of your cybersecurity engineers job roles and functions?
	7	AI Decision-Making Visibility		○ Do you have visibility of some of the decisions made by the AI-Powered cybersecurity platforms and why the decision was made? ○
Is AI a true solution in the eradication and prevention of cyber-crime in the South African banking sector?	8	Data Management and Governance		○ Do you know what type of data the AI-Powered cybersecurity platforms are collecting, where it is stored and who have access to it?

APPENDIX C

Consent Form

Title of project “The perceived impact of Artificial Intelligence on cyber security operations in the South African banking sector “

Name of researcher: Edward van Bosch

I, ..., agree to participate in this research project.

I agree to the following:

(Please circle the relevant options below)

The research study was explained to me. I understand what this study is about.	YES	NO
--	-----	----

I understand that I can volunteer to take part in the study	YES	NO
---	-----	----

I agree that the interview/focus group/other activity may be audio recorded	YES	NO
---	-----	----

I agree that direct quotations from my interview/focus group/other activity may be used by the researcher in their research report/manuscript/book chapter	YES	NO
--	-----	----

I agree that my participation will remain anonymous (my name or other identifying data will not be used by the researcher in their research report/manuscript/book chapter)	YES	NO
---	-----	----

I agree that other researchers may use the information I provide in my interview/focus group/other activity (depending on their own ethics clearance being obtained) but my name and any personal information will not be used or passed on	YES	NO
---	-----	----

..... (signature)
..... (name of participant)
..... (date)

..... (signature)
..... (name of researcher/person seeking consent)
..... (date)

APPENDIX D

Graduate School of Business Administration
University of the Witwatersrand, Johannesburg



Wits Business School Ethics Committee
Constituted under the University Human Research Ethics Committee (Non-Medical)

Ethics Clearance Certificate

Ethics protocol number: WBS/DB2419308/110

This certificate is only valid with a legitimate ethics protocol number and signed by the Researcher (below).

This certificate is only valid if accompanied by formal permission from the relevant stakeholder(s).

Project title	The perceived impact of Artificial Intelligence on cyber security operations in the South African banking sector
Investigator / Researcher	Mr. Edward Van Bosch
Nature of Project	MM (Digital Business)
Decision of the Committee	Approved, provided stakeholders and participants are guaranteed confidentiality.
Issue Date of Certificate	23/10/2024
Expiry date	Date of submission of the project / research report
Chairperson	Dr Ayanda Magida ☎ +27 11 717 3953 ✉ ayanda.magida@wits.ac.za

Declaration by Researcher

One copy must be signed by the Researcher and returned to the Chairperson of the Wits Business School Ethics Committee.

I fully understand the conditions under which I am authorized to carry out the abovementioned research and I guarantee to ensure compliance with these conditions. Should any departure to be contemplated from the research procedure as approved I undertake to resubmit the protocol to the Committee.

Signature

Date: