2020

Intelligence aspects of big data analytics for Kenya national security

Njoroge, Ann Wangechi
School of Humanities and Social Sciences
Strathmore University

INTELLIGENCE ASPECTS OF BIG DATA ANALYTICS FOR KENYA NATIONAL SECURITY

NJOROGE ANN WANGUI

PF NO: 119117

SUBMITTED IN PARTIAL FULFILMENT OF THE REQUIREMENTS OF MASTERS IN DIPLOMACY INTELLIGENCE AND SECURITY AT STRATHMORE UNIVERSITY

SCHOOL OF HUMANITIES AND SOCIAL SCIENCES

STRATHMORE UNIVERSITY

NAIROBI, KENYA

NOVEMBER 2020

DECLARATION

I declare that this work has not been previously submitted and approved for the award of a degree by this or any other University. To the best of my knowledge, the Thesis Proposal contains no material previously published or written by another person except where due references is made in the thesis itself.

Njoroge Ann Wangui

…….

……

Approval

The thesis of Ann Wangui Njoroge was reviewed and approved for examination by the following:

Professor Robert Mudida

……

……

Dr. Magdalene Dimba,

Dean, School of Humanities and Social Sciences,

Strathmore University.

Dr. Bernard Shibwabo,

Director of School of Graduate Studies,

Strathmore University.

ABSTRACT

In the age of rapid technological advancement, the range of threats to national security have evolved, becoming more complex and diverse, while the duty of the government remains that of securing its country. With the exponential rise in amount of structured, semi-structured and unstructured data generated every day, big data has become a core competence for the government which is linked to national security and the operations of the intelligence community. Big data provides intelligence organizations the opportunity to increase their investigative capabilities to combat threats to national security by enabling them to collect, analyze and disseminate information at a pace which could not be as effective in the traditional era. Governments will however be faced with the challenge of developing new capabilities to exploit and manage big data, which will require a rigorous review of the existing intelligence models and processes. The aim of this study is to examine the scope of applicability of big data and analytics in the functioning of the intelligence community, with a focus on the intelligence cycle. It also looks at the ethical and technical issues that limit the use of big data and analytics for national security. An exploratory research design was used to provide insight from the national security organs and intelligence community in Kenya on the applicability of Big Data and Analytics for national security. The study established that big data and analytics have a statistically significant effect on the intelligence cycle and national security organizations should embrace this new technology since it provides a lot of actionable insights. Based on this study it is recommended that the government enact more legislation to help develop an efficient and effective policy infrastructure for the various stakeholders in the intelligence community.

Table of Contents

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

IEEE:	Institute of Electrical and Electronics Engineers
APT:	Advanced Persistent Threats
ANOVA:	Analysis of variance
AI:	Artificial Intelligence
CIA:	Central Intelligence Agency
CI:	Collective Intelligence
DARPA:	Defense Advanced Research Projects Agency
DCI:	Department of Criminal Investigations
GLM:	Generalized Linear Model
GPDC:	Generalized Partial Directed Coherence
HRSC:	Homeland Security Research Corp
HUMINT:	Human Intelligence
IC:	Intelligence Community
ICT:	Information and Communication Technology
IT:	Information Technology
ICTAS:	Institute for Critical Technology and Applied Science
IARPA:	Intelligence Advanced Research Projects Activity
IBM:	International Business Machines Corporation
IGI:	International Gemological Institute
IOT:	Internet Of Things
IDS:	Intrusion Detection System
JNIM:	Jama'at Nusrat al-Islam wal-Muslimin
MIT:	Massachusetts Institute of Technology
MLE:	Maximum Likelihood Estimation
MANOVA:	Multivariate analysis of variance
NCIC:	National Cohesion and Integration Commission
NHTRC:	National Human Trafficking Resource Center
NIS:	National Intelligence Service

NSA:	National Security Agency
NLP:	Natural Language Processing
NGO:	Non-Governmental Organization
NATO:	North Atlantic Treaty Organization
OSINT:	Open Source Intelligence
PDC:	Partial Directed Coherence
SMS:	short message service
SMACT:	Social Media Analysis for Combating Terrorism
SOCMINT:	Social Media Intelligence
SPSS:	Statistical Product and Service Solutions
TECHINT:	Technical Intelligence
UN:	United Nations
US:	United States
USA:	United States of America

ACKNOWLEDGEMENT

First, I thank God the Almighty for His grace and blessings throughout the duration of my Masters coursework up to the successful completion of this thesis.

I would like to express my deep and sincere gratitude to my supervisor Professor Robert Mudida for his patience, enthusiasm and immense intellectual knowledge and support. His invaluable guidance and motivation helped me complete this thesis in time and it was a great honor and privilege to work and study under his guidance. I also extend my gratitude to all the lecturers took us through the coursework and members of faculty who facilitated the learning.

Lastly, I will be eternally grateful to my parents without whom I would not be where I am today, my brothers and sister as well as friends and classmates who encouraged me to complete my Masters.

DEDICATION

I dedicate this Master Thesis to my family. My lovely parents, John Njoroge Maku, and Lucy Nduta Njoroge who are the reason I am here today. You are my pillars in this life and your prayers keep me going and give me a reason for working hard to make you proud. My siblings; Humphrey Maku, Paul Ng'ang'a, Faith Muthoni; my nephew, Robi Killian Maku and niece Ivanna Nduta Maku. You are my true angels. Your love, prayers, encouragement and total acceptance give me the hope to face every day. And above all else, I thank the ALMIGHTY GOD for His grace and blessings over my life.

CHAPTER ONE

INTRODUCTION

1.1 Background of the Study

We currently live in a knowledge economy, where data is almost like a new currency. For instance the quantity of data that is interpreted per day globally through a smart phone surpasses the text data of all the English literature of the 20th century.[1] Every day, 2.5 quintillion of data is produced, with 90% of all data having been produced in the last 2 years. It is estimated by 2020, 40 zettabytes (43 trillion gigabytes) of data will have been generated, a 300% rise from 2015.[2] For a country's intelligence community, data and its analytics is key for its mission to protect the country's national interest and to anticipate any surprise. With the exponential increase in open source data from various areas such as social media, big data has become an area that is linked to national security and consequently to the intelligence community. The term big data is used here to mean the exponentially growing amount of digital material that is being created by new information technologies such as the online transactions, emails, search queries, social media interactions, the internet of things (IOT), and the advanced analytic methods to process this data.[3]

Big data does not only yield a quantitative increase in information, but also a qualitative change in how new knowledge is created and understood in the world. The more we capture large amounts of data from humans and the environment, the more the need to analyze them grows and this phenomenon is known as datafication.[4] Data related information technologies have grown and

[1] Karan Jani, "The Promise and Prejudice of Big Data in Intelligence Community," *ArXiv Preprint ArXiv:1610.08629*, 2016.
[2] IBM, "Big Data for the Intelligence Community," 2019, 20.
[3] Paul B. Symon and Arzan Tarapore, "Defense Intelligence Analysis in the Age of Big Data," *Joint Forces Quarterly—JFQ* 79 (2015): 4–11.
[4] Kevjn Lim, "Big Data and Strategic Intelligence," *Intelligence and National Security* 31, no. 4 (July 3, 2015): 619–35, https://doi.org/10.1080/02684527.2015.1062321.

have made revolutionary changes in fields such as commerce and science while transforming the economy and being an enabler of emergent technological. In defense intelligence communities, some of these technologies have been adopted for tasks, including technical collection and operational intelligence fusion among many other national security functions.

The intelligence community in the US for instance, has institutionalized big data with the creation of advanced analytics units in both civilian and military intelligence agencies, and a growing number of data analytics projects funded through organizations like the Intelligence Advanced Research Project Activities (IARPA), the Defense Advanced Research Project Agency (DARPA) and the Central Intelligence Agency (CIA).[5] This was as a result of the controversies caused by the 9/11 attack which cast a shadow of doubt on the methods of intelligence collection and analytic capabilities of the agencies in the US government.

Global national security threats have since evolved to include not only terrorism but also cyber security, transnational organized crimes and even election interference through cyber espionage as was seen in the recent US election. According to the report on Worldwide Threat Assessment of the US Intelligence Community by Coats, Russia's social media efforts such as spreading disinformation, hacking and leaking operations and data manipulation have been used to influence US policy and elections hence compromising the national security of the country. [6] Additionally, the report purports that Iran, which has a history of using social media campaigns as a means of

[5] Thomas Kean and Lee Hamilton, *The 9/11 Commission Report: Final Report of the National Commission on Terrorist Attacks upon the United States* (Washington, DC: National Commission on Terrorist Attacks upon the United States, 2004).

[6] Daniel R. Coats, "Statement for the Record: Worldwide Threat Assessment of the US Intelligence Community," ed. Office of the Director of National Intelligence, (January 29, 2019), https://www.dni.gov/files/ODNI/documents/2019-ATA-SFR---SSCI.pdf.

targeting audiences in allied nations with messages aligned with Iranian national interests, will continue to use online influence operations to try to advance its interests.

In Africa, Nigeria has faced contentious presidential elections as evidenced by the February 2019 elections and has also been faced with constant attacks from Boko Haram and Isis West Africa.[7] Additionally, violence and a critical humanitarian crisis still persists in South Sudan and Al-Qa'ida-affiliated Jama'at Nusrat al-Islam wal-Muslimin (JNIM) and its extremist allies present a growing threat in countries such as Chad, Mali and Niger. Big data and analytics serve as a way to allow the exploration of the decision cycle to identify and tackle such instances through the orientation and decision phases of the decision cycle. This means that action can be taken more quickly when opportunities present themselves, and problems can be anticipated ahead of time and measure put in place ahead of time. Furthermore, big data analytics capabilities can act as an enabler to collaboration. In an age of complexity, the next operation may require cooperation across multiple nations, non-governmental organizations, and both civilian and military authorities. Big data can facilitate this, even despite the fact that each nation operates within its sovereign process.[8]

Kenya can be described as a big data generation engine with the new wave of internet users and mobile technology in the country. According to the report by the Communications Authority of Kenya (2018), active mobile subscriptions stood at 46.6 million with an SMS traffic of around 14.5 billion and data and internet subscription of 42 million. This has thus created a vast amount of data that constitute big data which contain information that show the participation, conversation,

[7] E Nwanga et al., "Leveraging Big Data in Enhancing National Security in Nigeria," *International Journal of Knowledge, Innovation and Entrepreneurship* 2, no. 2 (2014): 66–80.
[8] Kevjn Lim, "Big Data and Strategic Intelligence," *Intelligence and National Security* 31, no. 4 (July 3, 2015): 619–35, https://doi.org/10.1080/02684527.2015.1062321.

opinions and collaboration among the population and this can be a valuable resource for the intelligence community in its effort to uphold the nation's security by protecting against terror attacks and cyber-crimes. For instance, various social media tools have become common in the everyday lives of many people and are the main methods of social connection and interaction around the world, be it between individuals, or within business entities as well as governments. Globalization has shifted the focus to the effects of non-state actors such as terrorists as opposed to the previous focus on the actions of states only. The non-state actors, including terrorists, criminals, protestors, hate mongers and rioters have benefited from the advancement in social media technology which has increased their ability to impact national security.

The issue of national security is at the forefront of every sovereign nation's policy, and every nation therefore has guidelines on how the nation is to be secured.[9] In Kenya, national security has evolved over the years but the security environment remains uncertain and full of many challenges that are transnational in nature and have the ability to get worse if not acted upon in the right way. Some of these transnational threats include terrorism, cybercrimes, drug trafficking, weapons proliferation among others. In the face of globalization and the dynamic changing landscape of technology, the scale of the threats that face the country are increasing exponentially making it very difficult for the intelligence community to effectively defend the country from all these threats.[10]

In this regard, there is a need to understand these threats, their nature and what effects they have on the national security of Kenya. Most of the threats facing the nation rely on clandestine

[9] Anneli Botha, "Assessing the Vulnerability of Kenyan Youths to Radicalisation and Extremism" (Institute for Security Studies, 2013).
[10] Anneli Botha, "Assessing the Vulnerability of Kenyan Youths to Radicalisation and Extremism," *Institute for Security Studies Papers* 2013, no. 245 (2013): 28–28.

networks to pass information and conduct their activities and in the age of new technologies, most of these activities have a digital print through avenues such as social media interactions. However, given how vast the data is and the lack of the necessary infrastructure or know how, these usually go unnoticed and eventually lead to attacks that threaten and weaken the national security of Kenya.

The inclusion of Big Data in the intelligence cycle of Kenya will provide a great advancement in the quest to improve national security since it will introduce objective and quantitative methods in a discipline highly characterized by its subjectivity. Additionally, it will help to reduce intelligence uncertainty through the collection of a huge volume of data and the identification of hidden correlations that would otherwise go unnoticed.

1.2 Statement of the research problem

The use of Big Data tools in various security policies has an effect on the freedom and security of people in a society and therefore affects the foundation of a constitutional state. Freedom and security are fundamental rights in accordance with the Universal Declaration of Human Rights. It is the duty of the government to protect its citizens and in doing so, the government may have to gather information on its citizens and this is where the balance of distance and this duty has to be balanced well.

Big Data, however, constitutes an assault on the protective function of distance. The amount of information that is now available or can be accessed for surveillance, investigation and prosecution has risen sharply. Combined with cheaper and more flexible forms of data storage and computers that can carry out ever more complicated data processing tasks, this results in government bodies increasingly encroaching on the lives of citizens. It is not so easy to analyze how and to what extent

Big Data applications already manifest themselves in the field of security. This is due to the secrecy that often shrouds security policy operations as well as the experimental nature of some applications. One such case has been the argument about the National Security Agency (NSA) spying on American citizens in the interest of National Security.

Recently, Nigeria has faced contentious 2019 presidential elections and has also been faced with constant attacks from Boko Haram and Isis West Africa. This has negatively affected the growth and development of the country alongside its socio-political landscape. It is, therefore, about time the nation seeks new techniques to tracking and curbing crimes. In this age of big data, as this data is generated by people in real time, it can be analyzed in real time by high performance computing networks, thus creating a potential for improved decision making and insight

According to the Article 238 (I) of the Kenyan constitution, national security is described as the protection of Kenya's territorial integrity, laws, its people, values and national interests against internal and external threats of any kind. Internally, tribal tensions that sometimes have erupted into open clashes, perennial violence associated with political elections, cattle rustling, conflicts among pastoralists, urban crimes and land clashes has been a constant security challenge to Kenya. Externally, Kenya has been the target of a number of international attacks and the one that is most common recently being the Westgate Mall terrorist attack in 2013 and the Dusit attack in 2019. Since then to date there have been numerous small-scale attacks on the national security not limited to terrorism but also in cybercrimes, drug trafficking among others.

These occurrences have negatively affected the socio-political and economic landscape of the country. One of the sectors that has suffered from this is the tourism sector in Kenya. The recent developments in technology such as mobile technology in Kenya and the internet wave have made

it easier for these threats to emanate and threaten the national security of the country. Individuals use social media to send alarming messages, hate messages and false information to the public regarding state of national security affairs. Most of the social media users remain unanimous and cannot be easily traced by law enforcement agencies and subsequent prosecution. Inherently, social media and vast electronic tools for generating huge amount of data have convoluted the government's appetite in Big Data management. This study aims to investigate the benefits of using big data in the intelligence cycle for national security of the country and highlight the possible ways that it can be used in reducing the threats to national security in the country.

1.3 Research Objectives

1.3.1 Overall Objective

The general objective of this study is to investigate the role Big data and analytics can play in improving the National security in Kenya.

1.3.2 Specific Objectives

The specific objectives of this study will be:

1. To examine the big data and analytics approaches the government can undertake to improve the intelligence cycle.
2. To look into the challenges faced by the government in implementing big data and analytics in the national security environment and their possible solutions.
3. To examine the unique value proposition big data and analytics presents to address issues on national security and the intelligence cycle.

1.4 Hypothesis

The study will test the following hypothesis:

1. Big data and analytics can help improve the intelligence aspects for National security of Kenya if policies are set in place to implement it.

1.5 Justification of the study

In a world of startling change, the first duty of the Government remains: that of providing security to the people of the country. Kenya has faced several security challenges since independence. Despite existing security policies, these challenges appear to be on the increase and are emerging in new forms. The new era of digitization has resulted in a vast amount of data and through big data and analytics, studies have shown a lot of actionable insight can be obtained by harnessing the benefits of big data and analytics. Big data has been important in the area of Aerial reconnaissance as demonstrated by the use of the Gorgon stare to collect, process, analyze and disseminate information through the use of the reaper and predator sensors.

Developed nations like the United States have made strides in using big data in their security agencies such as homeland security and the CIA to improve their national security. This clearly shows there exists an opportunity for Kenya to adopt some of the policies and frameworks and build a competitive advantage for the country to fight a number of transnational threats. However there has been challenges that have been highlighted in the implementation of big data in national security such as the lack of explicitly defined or instituted framework for assessing the big data phenomenon. Big data analytics does come with privacy issues as it entails collection of information that at times can encroach on the privacy of the concerned individual, institution or nation. Furthermore, there exist challenges in the implementation of big data into policy through the government budget process, data compartmentalization between the different government agencies in implementation of a policy if effected among other and government privacy obligations which may cause the process to take longer to be effectively implemented.

Nonetheless, the importance of big data cannot be understated and it plays a pivotal role in enhancing the national security of a country. This study will therefore be of importance in the policy making process of the relevant government bodies to formulate a framework that will enable the implementation of big data and analytics in the pursuit of National security. Secondly, this study will contribute to the literature on this topic, in the Kenyan context, as there is a paucity of studies done on this area.

Earlier research points to the fact that there is little study that has been conducted in Kenya and Africa as a whole, on the impact of big data on the national security. This study therefore aims to fill the literature gap as the discussion on big data and national security is a topic that is gaining traction and is in its infancy. With regards to academics, students and various scholars, especially those undertaking International Relations and Diplomacy courses as well as security studies will also benefit from the findings of this research as a source of information. Furthermore, in the study of security there still exists ideological conflicts and there exists various school of thoughts such as constructionists, liberalists and realists programs of research. All these school of thoughts try to approach security from various angles and this study will contribute to this scholarly discourse.

1.6 Literature Review

1.6.1 Introduction

Big data and analytics offers the intelligence community tremendous benefits but implementation challenges incorporating it are likely. For the purposes of ensuring that the ensuring research project is both robust and thorough, the ensuing literature seeks to offer insights on the subject matter and is categorized according to the theoretical frameworks for big data, its relationship with national security and the empirical evidence of the two.

1.6.2 Theoretical Framework

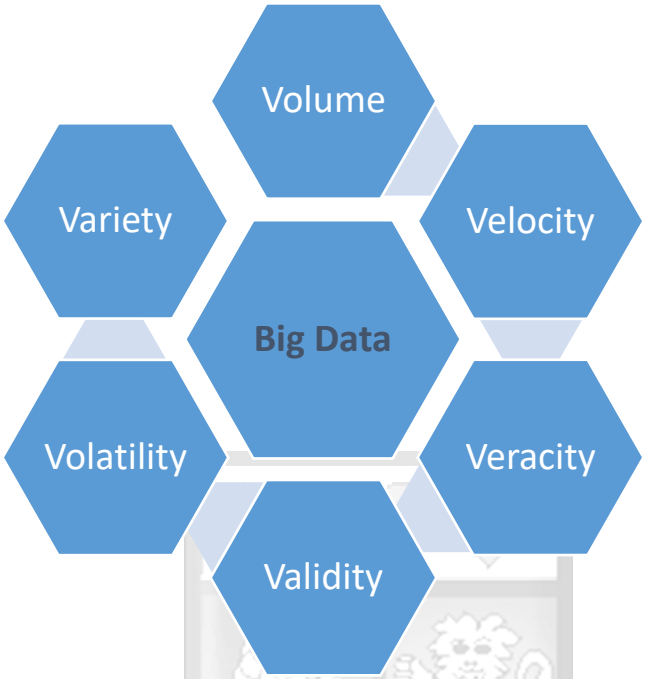
This section explores the theories that help define what big data is and additionally, the theories of national security. It explores how the two concepts have evolved and grown over time with the views of different scholars being examined to better understand their fundamentals.

1.6.2.1 Big Data Theoretical Framework

Big Data concern large-volume, complex, growing data sets with multiple, autonomous sources. With the fast development of networking, data storage, and the data collection capacity, Big Data are now rapidly expanding in all fields. Big data has six main characteristics that differentiates it from ordinary data and they include: volume, velocity, variety, validity, veracity and volatility.[11]

[11] Wei Tan et al., "Social-Network-Sourced Big Data Analytics," *IEEE Internet Computing* 17, no. 5 (2013): 62–69, https://doi.org/10.1109/mic.2013.100.

Figure 1: Big Data differentiators



Source: Author (2020)

The top three differentiators on Figure 1 represent the most fundamental differentiators and can be explained as follows based on Tan[12]

Volume - This represents the vast amount of data that is created every second that are larger than what normal relational database infrastructure can handle

Velocity – This is the frequency at which new data is created, captured and shared

Variety – This represents the various types of data that do not fit to consume structures.

The bottom three differentiators are the additional characteristics described as follows:

[12]Wei Tan et al., "Social-Network-Sourced Big Data Analytics," *IEEE Internet Computing* 17, no. 5 (2013): 62–69, https://doi.org/10.1109/mic.2013.100.

Validity- This is the accuracy of the data for the intended use.

Veracity- This is how meaningful the results are for the given problem at hand.

Volatility- This is representative of how long you need to store the data

Big Data involves delving into the various datasets, identifying and relating different unknown relationships within these datasets and coming up with insights that can enables good decision making It is anticipated that by the year 2020, worldwide production of data will get to around 35 zettabytes [13] and the need and capacity to process such huge volumes of data will be integral in providing a definition of what constitutes big data. The ability to use data is not only challenged by the increase in large data volumes, but also on data velocity. The speed at which new data is created and changed is increasing, which creates more challenges in its storage and processing. For instant, Twitter users generate an average of 6,000 tweets every second.[14]

Other than data, Big Data is also about the means by which data are stored, processed and analyzed, technologically and methodologically. To process enormous amounts of data hardware and software solutions are with robust capabilities are required, as well as ways of extracting information from the data since it is only a series of symbols when not processed to generate meaning. The potential in leveraging big data analytics exists but supporting the application is lacking theory. Chen et al. found that big data analytics is progressing technically, but a theoretical framework is required for design and integration.[15] Similarly, Seddon and Currie also opined that

[13] J Gantz and D Reinsel, "The Digital Universe in 2020: Big Data, Bigger Digital Shadows, and Biggest Growth in the Far East," in *IView* (IDC Analyze the Future, 2012), 1–16.
[14] K Morabia et al., "SEDTWik: Segmentation-Based Event Detection from Tweets Using Wikipedia," *Conference of the North American Chapter of the Association for Computational Linguistics: Student Research Workshop*, June 2019.
[15] Chen, Kun, Xin Li, and Huaiqing Wang. "On the Model Design of Integrated Intelligent Big Data Analytics Systems." *Industrial Management & Data Systems* 115, no. 9 (2015): 1666–1682.

big data analytics is under-theorized and empirically underrepresented in business research and also found the vast majority of related research articles were analytical and more research using case studies in different organizations is needed.[16]

Chen et al. applied Schut's CI model to test the integration of big data analytics in an ecommerce company. Chen, Li, and Wang concluded that the CI framework was suitable for designing and implementing big data analytics applications.[17] However, Al Nuaimi et al. contend that despite the advantages of implementing a big data framework, there does exist ethical and private issues that need to be addressed if the framework should be effective and this is dependent on what policy will be created to support these frameworks.[18] Similarly, Btihaj studied big data analytics regarding the ethical concerns of immigration. [19] Btihaj revealed that governments are increasingly applying big data analytics technologies to regulate the flows of immigration. His study was centered on the Australian Department of Immigration and Citizenship's big data analytics system developed by IBM to identify risky travelers and improve boarder security. Btihaj presents the argument that big data analytics can be used as means of digital discrimination and prevents movements of people based on a preconceived risky pattern.

Data processing is also another field big data could be useful according to Kitchin however the study Waterman and Bruening addressing risks pertaining to big data analytics related to the Enron financial scandal argue that this could be problematic. In their study, the initial emails released by

[16] Seddon, J. J., & Currie, W. L. (2017). A model for unpacking big data analytics in high-frequency trading. *Journal of Business Research*, *70*, 300-307.
[17] Chen, K., Li, X., & Wang, H. (2015). On the model design of integrated intelligent big data analytics systems. *Industrial Management & Data Systems*, *115*(9), 1666-1682.
[18] Al Nuaimi, E., Al Neyadi, H., Mohamed, N., & Al-Jaroodi, J. (2015). Applications of big data to smart cities. *Journal of Internet Services and Applications*, *6*(1), 25.
[19] Ajana, B. (2015). Augmented borders: Big Data and the ethics of immigration control. *Journal of information, Communication and Ethics in Society*, *13*(1), 58-78.

the Federal Energy Regulatory Commission totaled over one million. After the emails were scrubbed, removing duplicates and blanks, the total number of emails was reduced to fewer than 600,000 and the total number of users was reduced by 7.5%. These types of discrepancies could skew the larger set of analyses and possibly result in misguided decisions. The integrity of the data and how the data is processed create risks, as the analytical models depend on uncompromised data[20].

Big data analytics has the potential to enhance Homeland Security Investigations, but the technology does present risks. Merging high volumes of data from disparate datasets has potential risks. To maximize the potential of big data analytics and avoid pitfalls, oversight of data integrity and processing is necessary. Most importantly, verifying the results of findings produced through big data analytics is imperative to protecting individuals, particularly related to potential algorithmic discrimination, and the reputation of the supported organization.

1.6.2.2 National Security Theoretical Framework

The concept of national security to this day remains an ambiguous area that has evolved from simpler definitions which made emphasis on freedom from military threats and political coercion. The theories that will be explored in this section will be the strategies that integrate military, political and economic pursuits of a nation's ultimate objective as a grand strategy in the international framework.[21] The study by Moore and Turner explores six broad theories that address national security namely which will be discussed below:[22]

[20] Krasnow Waterman, K., & Bruening, P. J. (2014). Big Data analytics: risks and responsibilities. *International Data Privacy Law*, *4*(2), 89-95.
[21] Lindell Hart, *Strategy*, 2nd Revised Edition (Plume, 1954).
[22] John Moore, *Newer Theories in Understanding War: From the Democratic Peace to Incentive Theory* (Moore & Turner National Security Law. Durham, NC: Carolina Academic Press, 2005).

1. Balance of Power approach
2. Collective security approach
3. World Federalist approach
4. Functionalist approach
5. Democratic peace approach
6. Incentive approach.

1.6.2.2.1 Balance of Power approach

This theory suggests that an approach that intends to prevent any country from becoming too strong such that it will enforce its will upon other countries.[23] It sees the world to be made up of rational actor who act in such a natural way by uniting in various coalitions with each other to conter any threats against them. The balancing of this strategy can either occur internally or externally where internal balancing is when a country strengthens itself through the mobilization of resources within the confines of its own borders while external balancing is when the coalitions formed pool their resources to be used against a common enemy. History has shown that eventually a power comes up to challenges such arrangements and this sometimes leads to the criticism of this theory that this can be considered an invitation for by them balancing the power in this way.[24] Ideally, wars should not happen in this theory because each country is constantly on the lookout and attentive to each other's' alliances but it has been identified that a concept called power transition theory

[23] T Lansford and J Pauly Jr, *To Protect and Defend: US Homeland Security Policy* (London: Routledge, 2016).
[24] Dina A. Zinnes, "An Analytical Study of the Balance of Power Theories," *Journal of Peace Research* 4, no. 3 (1967): 270–87, https://doi.org/10.1177/002234336700400304.

has made the realization that wars do often come as a result of small shifts in the distribution of power.[25]

1.6.2.2.2 Collective security approach

This theory came to be during the World War 1 as an alternative to the balance of power theory and it can be attributed to Immanuel Kant who first came up with it as alternative to just war theory which was founded on ethical duty towards renunciation of any form of aggression.[26] Collective security is an approach which sees national security as a side benefit of world order to be managed by some transcendent authority from above. It is the theoretical foundation upon which United Nations was built. It sees national security as a byproduct of world order and it to be managed by an authority figure and it's formed the foundations for how the United Nations was formed.

It does not agree with the idea that alliances can work and they make a substitution of that idea the concept that if you attack one person, its equivalent to attacking everyone. This then relives various countries with the load of having to provide security for themselves, especially weaker nations that have no way of defending themselves against stronger ones. UN and NATO can be considered to be collective security and defense institutions respectively. According to Claude (2005), violence should always be a measure of last resort.[27]

1.6.2.2.3 World Federalist approach

This theory advocates for a system that is democratic of coequal regions that has citizens that can be considered as global and this is to replace nation states as the major form of government in the

[25] Joseph S. Nye Jr, *The Paradox of American Power: Why the World's Only Superpower Can't Go It Alone* (Oxford University Press, 2003).
[26] J Moore, *Newer Theories in Understanding War: From the Democratic Peace to Incentive Theory*, 2005.
[27] Inis Claude, "Theoretical Approaches to National Security and World Order," *Moore & Turner National Security Law. Durham*, 2005, 3–14.

world. It believes in the spirit as opposed to the letter of the law and according to the Federalist Paper No. 20, the principles of federalism have been laid out and key among them is that decisions for the common good of the society should be made on a level that is high and that every individual has the right to exercise their maximum influence on all matters that are of concern to them.[28] There have existed many other theories before this and the paper by Baratta tries to establish the history and it traces it as far as Dante (1265-1321) all the way to the times of Winston Churchill (1874-1965).[29]

Some of the concepts are also drawn from the cosmopolitan ideas more so those by Jacques Derrida who pushes for the idea of asylum for immigrants.[30] Most federalists usually are in favor of ideas such as appeasement with terrorist groups specifically those that do have nationalist motivations.[31] On the other hand, cosmopolitanism views achievement of global security through peaceful coexistence[32]

1.6.2.2.4 The functionalist approach

This approach was brought about by Karl Deutsch and its key idea is that an establishment of a predictable growth pattern in the world is necessary and it is also sufficient to have international organizations that address the needs that are in support of human welfare.[33] Its key principle is based on synergy where organizations work in tandem to complete important functions that the

[28] Alexander Hamilton, James Madison, and John Jay, *The Federalist: A Commentary on the Constitution of the United States: Being a Collection of Essays Written in Support of the Constitution Agreed Upon September 17, 1787 by the Federal Convention: Reprinted from the Original Text of Alexander Hamilton, John Jay, and James Madison* (GP Putnam's sons, 1888).
[29] Joseph Preston Baratta, *The Politics of World Federation: United Nations, UN Reform, Atomic Control*, vol. 2 (Greenwood Publishing Group, 2004).
[30] Jacques Derrida, *Writing and Difference* (Routledge, 2001).

[32] David Held, *Democracy and the Global Order: From the Modern State to Cosmopolitan Governance* (Stanford: Stanford University Press, 1996).
[33] Karl W. Deutsch, *Politische Kybernetik: Modelle Und Perspektiven* (Rombach, 1969).

state is required to undertake responsibility for. The works Mitrany [34] and Sewell[35] elaborate functionalism as global peace theory.

They do share a certain commonality with federalists in the sense that there is no need of nation states and that the needs of the people can be well taken care of by international organizations and the cooperation between state and these organizations will eventually lead to a cooperation that is political in nature.

1.6.2.2.5 The democratic peace approach

This idea was put forward by Immanuel Kant (1795) and it opines that that responsible democracies shouldn't go to war with each other since they are in principle good at achieving several goals associated with peace and things such as human rights, avoiding corruption, handling terrorism among many other responsible actions.[36] Jack Levy has called this idea the closest thing we have to a law in the field of international relations.[37] Moore[38] goes further and says that democracies are not only good at preventing war, but they are good at achieving various peacetime goals too - things such as human rights, economic development, environmental protection, famine avoidance, control of terrorism, corruption avoidance, and ending mass refugee flows.

Contemporary spokespersons for the democratic peace approach include Doyle [39], Russet[40], and Rummel[41] who all view democracy as a non-violent method and the key principle behind it leads

[34] David Mitrany, *A Working Peace System: Introd. by Hans J. Morgenthau* (Quadrangle Books, 1966).
[35] James Patrick Sewell, "Policy Processes and International Organisation Tasks," in *International Organisation: World Politics* (Springer, 1969), 98–112.
[36] Immanual Kant, "To Perpetual Peace: A Philosophical Sketch Trs," *Ted Humphrey, Indianapolis: Hacket Pub*, 1795.
[37] Jack S. Levy, "Domestic Politics and War," *The Journal of Interdisciplinary History* 18, no. 4 (1988): 653–673.
[38] Moore, *Newer Theories in Understanding War*.
[39] Michael W. Doyle, "Ways of War and Peace: Realism," *Liberalism, and Socialism*, 1997, 24–25.
[40] Bruce Russett, *Grasping the Democratic Peace: Principles for a Post-Cold War World* (Princeton university press, 1994).
[41] Rudolph J. Rummel, *Power Kills: Democracy as a Method of Nonviolence* (Routledge, 2017).

to rule of law and encourages a certain culture that adheres to these laws. It is the most commonly debated theory across various international relations fields and is key in studies that involve peace. A lot of research is being done into it involving a number of case studies of other theories in agreement with it such as equifinality.

1.6.2.2.6 The incentive approach

The incentive approach is the name Moore[42] gives is variant of the democratic governance theory that opines that the best foreign policy framework is one that focuses on the rule of law and the wealth of nations through trade agreements. He argues that not all states that are non-democratic are a threat to the peace and the only reason democracies go to war is more of a defense reason than an aggressor reason against the various actions of leaders in non-democratic regimes. These dictators capitalize on the incentive that democracies do not have a penchant of going to war but this theories does advocate for war as a defensive strategy should sanctions, diplomatic actions or otherwise fail to deter potential aggressors.

1.6.3 Big data and National Security

In the intelligence process, huge datasets allow trends and patterns to be deduced with a very high level of confidence and this ability can then be used to provide key information for national security professionals which can be used in decision making in the intelligence cycle. The Intelligence process or cycle can generally be categorized into 6 core processes namely: requirements, collection, collation, analysis, dissemination and security.[43]

[42] Moore, *Newer Theories in Understanding War*.
[43] Peter Gill and Mark Phythian, "From Intelligence Cycle to Web of Intelligence: Complexity and the Conceptualisation of Intelligence," in *Understanding the Intelligence Cycle* (Routledge, 2013), 35–56.

In requirements, the intelligence, surveillance and reconnaissance needs that are related to the fight against threats in various areas usually have an inclusion of an ability to tackle high data volume and variety in limited periods of time. The collection aspect aims to look at new data and information to bridge the knowledge gap and the big data capabilities play a key role in this area of facilitating collection of large data sets through certain indexing algorithms that can identify and summarize the relevant data.

The main contribution of this in the intelligence cycle can be in the area of processing and exploitation. This is where data is transformed into a format that can be used in drawing meaningful conclusions. For example, big data can be used to bring structure to data that is unstructured from various sources such as websites and this can then be used to draw conclusions.[44] The intelligence analysis can be considered as the thinking part of the intelligence cycle[45] since it involves the application of the knowledge and reasoning methods necessary to transform raw data from various sources into useful information that can be used in the decision making process.

Various tools can be used in the dissemination of intelligence from the creators to the consumers besides the traditional tools that have previously been used in dissemination of information such as periodic reports. The use of technology can help in conveying the information in a much more efficient way and in a much faster delivery mode. One such example of this is applications that rely on the insights from big data to give recommendations of what consumers may like on various ecommerce sites, as in the case on e-commerce platforms such as Jumia and additionally, the

[44] Damien Van Puyvelde, Stephen Coulthart, and M. Shahriar Hossain, "Beyond the Buzzword: Big Data and National Security Decision-Making," *International Affairs* 93, no. 6 (November 1, 2017): 1397–1416, https://doi.org/10.1093/ia/iix184.

[45] Roger Z. George and James B. Bruce, *Analyzing Intelligence: National Security Practitioners' Perspectives* (Georgetown University Press, 2014).

visualization aspect can also present data that would have otherwise have looked so complex in a much simpler and understandable way for a much wider audience.

1.6.4 Empirical literature

This section looks at the various studies examining the relationship between big data and national security and how they relate with each other. It explores studies that have been done internationally, then relates them to those done on the African region and finally the studies that are done in the Kenyan context.

Big data analytics as a research area is new and has the interest of several industries, including the government. Although it presents value and opportunity for the government, academic literature supporting big data and analytics in national security is lacking. Kitchin acknowledges that big data analytics is popular within business but the governments are also beginning to recognize the potential. [46] In his study he assessed the extent to which big data is involved in the paradigm shifts across multiple disciplines including sciences, social sciences and humanities. The study was a critical review of emerging epistemological positions and it concluded that big data can be a fruitful approach for governments in the field of national security. Where business is leveraging data analytics to increase revenue and market share, governments can apply it to improve national security. The potential in leveraging big data analytics exists but supporting the application is lacking theory.

Wang et al. found that big data analytics is progressing technically, but just like Kitchin, they found that a theoretical framework is required for design and integration. [47] Their study was trying

[46] Rob Kitchin, "Big Data, New Epistemologies and Paradigm Shifts," *Big Data & Society* 1, no. 1 (2014): 2053951714528481.

[47] Xiaojun Wang et al., "On the Model Design of Integrated Intelligent Big Data Analytics Systems," *Industrial Management & Data Systems*, 2015.

to resolve the problems associated with the collective intelligence (CI) model suggested by Schut.[48] In the model instantiation, they used a multi-agent paradigm as the specific model and then the general model was the hierarchical colored petri net. By doing do, their study found that it will be suitable for a dynamic data analyst environment very much so how the national security of a nation is dynamic. Their findings helped in the integration problem of the big data environment and made suggestions of how institutions and governments can employ big data technologies.

Tankard in his study explains that, of the many advantages of big data analytics, the most compelling is operational efficiency. [49] This includes the timely detection of cyber or terrorism attacks by harnessing the power of big data to provide insights. He goes on to explain that big data analytics can be useful to governments for the detection of threats from foreign countries, terrorists, hacktivists and criminal elements in the real world and in cyberspace. His study however focused more on the controls that need to be placed around the data rather than the applications and systems that are responsible for storing the data. Cardenas et al. [50] extols the use of big data analytics, but they focus more on its uses for cyber security. They explain that the idea of data analysis for cyber-attack detection is not new in that the information security community have been monitoring network traffic, and analyzing system logs and other sources of data in order to detect threats and malicious activities for more than a decade, but the use of big data analytics is better and has overcome the many challenges that faced the traditional data analysis (for security) of monitoring network traffic and security logs. One of these challenges is the inability to perform long term and large scale analytics because it was not

[48] Martijn C. Schut, "On Model Design for Simulation of Collective Intelligence," *Information Sciences* 180, no. 1 (2010): 132–155.
[49] Colin Tankard, "Big Data Security," *Network Security* 2012, no. 7 (2012): 5–8.
[50] Alvaro A. Cárdenas, Pratyusa K. Manadhata, and Sreeranga P. Rajan, "Big Data Analytics for Security," *IEEE Security & Privacy* 11, no. 6 (2013): 74–76.

economically feasible to keep large volumes of data for a long period. They explain that one of the main impacts of big data technologies is the facilitation of the development of affordable infrastructures – such as storage and maintenance - for security monitoring by various industries, thus making it possible for large scale analytics to be carried out. However, they argue that despite the significant promise of big data analytics for security, there are several challenges, such as privacy laws, that can prevent this development from realizing its true potential if not addressed.

Although big data analytics offers numerous benefits and value, it is not without its challenges and risk. One such challenge is the ethical concerns that surround the access of data used in intelligence. Ajana studied big data analytics regarding the ethical concerns of immigration and found that governments were increasingly applying it on the Australian Department of Immigration and Citizenship's big data analytics system made by IBM. [51] The data was then complied and algorithms were built to provide profiles on the "riskiest" travelers. This could be dangerous to immigration because governments can unfairly target undocumented immigrants and prevent migrant population flow through an unchallenged algorithm.

The other challenge Ajana found was that of finding patterns where none exist. This type of pattern analysis can be hazardous, and this concept of pattern analysis through big data analytics in the identification high-risk persons based on analytic profiles is supported by homeland security in the United States. [52] Bottles et al. also highlight some other challenges of big data such as its susceptibility to the bias that could inherently be present in the data collected for analysis. [53] Flawed data can lead to faulty discoveries, which can result in consequential decisions. This problem is

[51] Btihaj Ajana, "Augmented Borders: Big Data and the Ethics of Immigration Control," *Journal of Information, Communication and Ethics in Society* 13, no. 1 (2015): 58–78.
[52] Ajana.
[53] Kent Bottles, Edmon Begoli, and Brian Worley, "Understanding the Pros and Cons of Big Data Analytics," *Physician Executive* 40, no. 4 (2014): 6–12.

common when applying the new big data analytics technologies to old government data systems. The way in which the data is stored differs by each system or application that has been introduced over time. The problem is multiplied when data is merged from different government agencies or private sector organizations.

Nwanga et al. looks into how big data can be used to create new leads in investigations in an electronic fashion as opposed to the traditional field gathering exercise as a way of tackling terrorism in Nigeria and they do this through an analysis of the dark web.[54] Their study looks into how big data can be leveraged to improve the awareness that will help agencies be proactive rather than reactive to national threats. They suggested the creation of a big data center as an active collaboration strategy to combating terrorism in Nigeria.

Ezumah and Adekunle focused on cybersecurity in four countries in Africa namely Nigeria, Kenya, Egypt and South Africa. Their work presents detailed information on the legislative framework proposed and implemented by these countries to combat and control cybercrimes.[55] Notable among them are the Egypt's e-Signature Law 15, Kenya's e-Transaction Bill, Nigeria's Computer Security and Critical Information Infrastructure Protection Bill, and South Africa's Electronic Communications and Transaction Act. They come to the realization that even though cybercrime can never be abolished, efforts that are aimed at curbing it can go a long way at trying to reduce it and big data can play a role at making the cyberspace safer.

[54] MATHEW E. Nwanga et al., "Leveraging Big Data in Enhancing National Security in Nigeria," *International Journal of Knowledge, Innovation and Entrepreneurship* 2, no. 2 (2014): 66–80.
[55] Bellarmine Ezumah and Suraj Olunifesi Adekunle, "A Review of Privacy, Internet Security Threat, and Legislation in Africa: A Case Study of Nigeria, South Africa, Egypt, and Kenya," in *Internet and Distributed Computing Advancements: Theoretical Frameworks and Practical Applications* (IGI Global, 2012), 115–136.

Van Vuuren et al. opines that the government does have a key role in providing, regulating and maintaining national security which is a right deserved by the citizens of a country. [56] They acknowledge that despite the South African government approving the draft of the National Cyber Security Policy framework that was done in March 2012, the nation's still needs a good structure to efficiently control its cyber infrastructure. They suggest that structures need to be in place to set the security controls and policies and also to govern their implementation arguing that partnerships between business, government and civil society needs to be put in place to achieve this goal. Although they do not mention big data as one of the ways of setting up the infrastructure, they do suggest that the government set up organizational structures to control cybersecurity and follow the policy implementation, something which big data has been shown to assist in doing.

In Kenya, Kimutai looked at the effects of social media on the National security of Kenya. The target population were employees from the DCI, NIS, Communications Authority of Kenya National Cohesion and Integration Commission (NCIC) and a few members of the public. [57] Her study established that terrorist groups do indeed take advantage of social media to compromise national security in Kenya and made a suggestion that the government put in place measure to ensure this does not happen such as the cybercrime act. An additional suggestion of close monitoring of social media groups was suggested but no framework was provided as to how this will be done. Big data has proved to be a valuable asset in the intelligence cycle when it comes the suggestion about monitoring that is suggested by the study.

[56] Joey Jansen van Vuuren et al., "An Approach to Governance of Cybersecurity in South Africa," in *Cyber Behavior: Concepts, Methodologies, Tools, and Applications* (IGI Global, 2014), 1583–1597.
[57] JULIUS KIPKORIR Kimutai, "Social Media and National Security Threats: A Case Study of Kenya," *Unpublished MA Thesis: University of Nairobi*, 2014.

Kones looks at transnational threats to the national security in Kenya and identifies the most common of these threats in Kenya are terrorism, drug smuggling and small arms and light weapon proliferation.[58] The study also found that global trends also played a part in the transnational threats especially terrorism with the rise of extremist groups serving to incite other local groups such as al-Shabaab and other militants. The solutions provided by the study are however very vague and no specific reform has been suggested to curb this. Big data has proven to be valuable especially in issues such as surveillance of borders as shown in the study by Fahey[59] and additionally in the monitoring of communication channels that may be used to pass information and recognizing patterns in advance before events happen that threaten national security. If the government as a policy framework that incorporates big data then the issue of transnational threats will be greatly reduced.

1.7 Research gap

From the literature review, it is clear that although big data analytics is seen as a very powerful and relevant technology for the detection and prevention of cyber-attacks, there is limited documentation of exactly big data and analytics can be used to enhance national security in Kenya. Additionally, there is no meaningful framework about how effective it is in the Kenyan context. Such a framework can help provide guidance on the policy initiation and implementation by the Kenyan government to achieve the success that is being achieved by countries that have a working policy framework incorporating big data in their national security.

[58] Kefa Kones Kibet, "Transnational Threats to the National Security in Kenya - Google Search" (University of Nairobi, 2015).
[59] Sean Fahey, "Big Data and Analytics for National Security," *Stanford University. Pristupljeno* 5 (2012): 2017.

1.8 Conceptual Framework

The conceptual framework maps out the actions required during the course of the study to help achieve the objectives of the study. This study seeks to investigate how big data affects the intelligence cycle under the 5 key processes of the intelligence cycle. Figure 1.1 illustrates what aspects of the cycle will be examined in relation to big data.

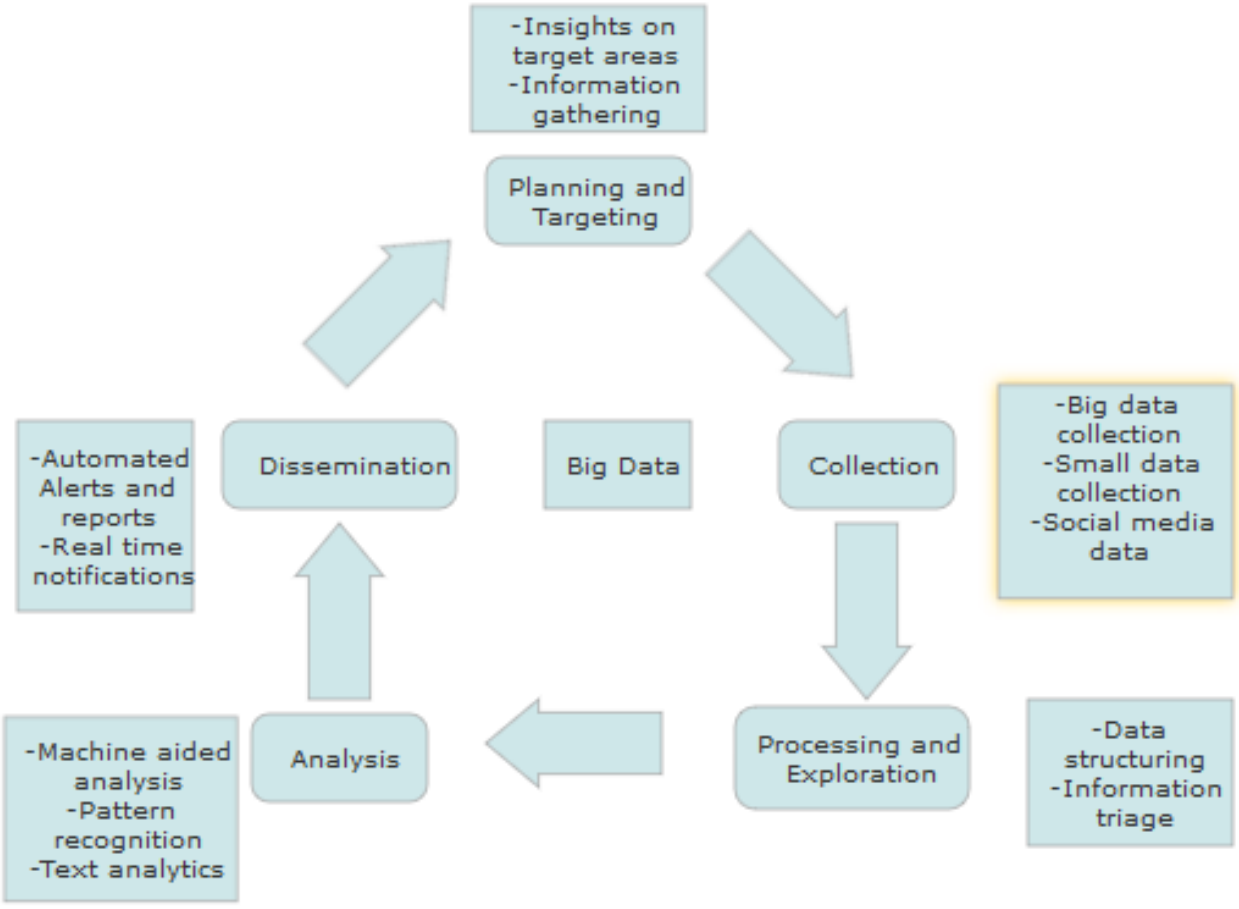


Figure 1.1: Conceptual framework of study

CHAPTER TWO

INTELLECTUAL HISTORY AND CONCEPTUAL ANALYSIS OF INTELLIGENCE ASPECTS OF BIG DATA

2.1 Introduction

This chapter seeks to offer insights on the subject matter in five key areas: evolution of big data analytics and big data trends, big data and the intelligence cycle, applications of big data, and the ethical issues in the use of big data analytics.

2.2 Evolution of Big Data Analytics

Ohlhorst traces back the conception of big data to the pre-personal computer era where unstructured data in the form of paper records was the norm with any attempt to review data involved tedious human actions of first seeking the data from the storage archives. Central to this, the author notes that the first recognizable emergence of big data was the 1880 census in the USA whereby user data of an estimated 50 million US residents had to be collected, classified, and analysed.[60] Key to note is that the mere collection of the census numbers was not enough; the government needed to derive more insights from the information such as the level of education, age, sex, occupation, and the number of insane people in every household. Suffice to say that this information held intrinsic value to the government if only it could be analysed.

Ohlhorst noted that the above desire led to the need to correlate the various categories of the information such as one's level of education to their geographical area of residence and European country of origin.[61] However, due to the limited data analytical technologies available during the time, meant that the vast insights from the collected data could only come to fore in seven years.[62]

[60] Frank J. Ohlhorst, *Big Data Analytics: Turning Big Data into Big Money*, vol. 65 (John Wiley & Sons, 2012).
[61] Ohlhorst.
[62] Ohlhorst.

From the above, a case for the coining of the phrase big data is made to note that the phrase implied, from the pre-definition era of the 1950s, to encompass the process towards the collection and analysis of the data and not merely the data in itself.

Nair further advances the notion on the evolution of big data by stating that organizations proceeded to develop advanced technologies that would in turn enable them to derive more insights from the collected data. Central to the above exercise was the need to make sure that such insights were derived from data in the shortest amount of time possible.[63] For the purposes of enhancing the above exercise, it then becomes apparent that organizations in need of big data have increasingly turned towards external sources such as social media, clickstreams, and the internet, as a hub of such data. The decision to gain vast amount of data from the external organization environment implied that such data would differ, in volume, from the internal organizational data (small data) hence given meaning to the phrase, big data. Firms with the ability to provide insights from the big data stood to profit more from their enterprise.[64]

Nair pointed out that whereas this phenomenon presented a lucrative enterprise for organizations, who responded by developing fast processing engines, another challenge emerged in the form of the Open Source Community. The community played a central role in the evolution of the big data analytics. For example, the community was essential in the developing of new computer frameworks tasked with receiving and analysing big data in the form of Data Lakes as well as ensuring that such data is predictive.[65]

[63] Deepesh Nair, "The Evolution of Analytics with Data," *Medium*, October 21, 2018, https://towardsdatascience.com/the-evolution-of-analytics-with-data-8b9908deadd7.
[64] Nair.
[65] Nair.

2.3 Emerging Big Data Trends

Big data is creating new opportunities for the intelligence cycle and its leading to an improvement in the investigative capabilities of national security agencies when dealing with the various transnational threats. The Homeland Research Security Corp. (HRSC) predicts that the big data market will witness a rapid growth by 2022[66] due to the increased use of smartphones and pother smart devices such Internet of Things (IoT) which are expected to have grown to 212 billion U.S. dollars by the end of the year 2019 according to the report by Liu.[67] Furthermore, the increased advancement in data collection tools will allow organizations to increase the amount of data generated such as the use of augmented analytics.[68] The use of cloud technology by the agencies in the government will act as an incentive for the implementation of analytics tools and given that most agencies don't have the infrastructure to support big data analytics, some of them will have to turn to cloud solutions which will make it easier to adopt data analytics solutions.[69]

Transnational threats have changed the way countries are dealing with national security issues and new security threats that pose as a threat will make countries to make use of technologies that are well equipped to deal with this advanced threats. These systems will then generate a big amount of data that will then lead to an increase in the demand for big data solutions.

Machine learning provides the opportunity for researchers to analyse data in ways that were not possible before and the computing power that is currently in existence can allow for it to do correct

[66] Homeland Security Research Corp. (HSRC), "Homeland Security Research Corp. (HSRC): Big Data and Data Analytics in National Security Is Forecast to Grow at a 2015-2022 CAGR of 17.5%," accessed March 20, 2020, https://www.prnewswire.com/news-releases/homeland-security-research-corp-hsrc-big-data-and-data-analytics-in-national-security-is-forecast-to-grow-at-a-2015-2022-cagr-of-17-5-300869799.html.
[67] Statista Research Department, "Global IoT End-User Spending Worldwide 2017-2025," *Statista*, February 19, 2020, https://www.statista.com/statistics/976313/global-iot-market-size/.
[68] Gerard George, Martine R. Haas, and Alex Pentland, Big Data and Management (Academy of Management Briarcliff Manor, NY, 2014).
[69] Statista Research Department, "Global IoT End-User Spending Worldwide 2017-2025."

classification and recognize very complex patterns that would otherwise have not been discernible by humans.[70] Another similar concept is deep learning which simulates the workings of biological neurons through artificial neural networks and this is then used to understand complex features in data by linking it with historical recognized patterns. Around the mid 1980's, AI required some human input in the classification of data but this has changed over time with the AI learning by itself.[71] The study by Neal Jean et al. goes on to explain how this can be applied in economic development where he used a combination of survey and satellite data from countries such as Nigeria, Uganda, Rwanda, Tanzania and Malawi to train algorithms to identify visual patterns that then helped them predict certain socioeconomic distributions.

Luvembe and Mutai further examine how big data can be adopted into the county governments in Kenya in an aim to strengthen citizen engagement and participation in the security process of Kenya. They argue that a majority of the current county digital systems exist in isolation from one another and that the sharing of data becomes very difficult. The benefits of big data analytics, according to them, will only be realised once the different counties merge their systems and amass the advantages of the collective data that will be available from the integration of the systems to provide solutions in crime prevention, improve devolution service, assist in natural disaster management among other benefits.[72]

[70] Robert D. Hof, "Deep Learning: With Massive Amounts of Computational Power, Machines Can Now Recognize Objects and Translate Speech in Real Time," *Artificial Intelligence Is Finally Getting Smart. MIT Technology Review* 116, no. 2 (2013): 78–86.
[71] Neal Jean et al., "Combining Satellite Imagery and Machine Learning to Predict Poverty," *Science* 353, no. 6301 (2016): 790–794.
[72] Alex Luvembe and Hillary Mutai, "Big Data Framework for Kenya's County Governments," *Journal of Computer and Communications* 07 (January 1, 2019): 1–9, https://doi.org/10.4236/jcc.2019.71001.

2.3.1 Generalized Partial Directed Coherence (GPDC)

National security is an extremely complicated system in which different parties play various roles with asymmetric access to information at different intelligence levels. The success or failure of a government in securing its national security has big ramifications. With the help of complex network theory, it is possible to model and extract the network topological structures to reveal hidden information and relationships among various fields that affect national security.[73] Due to the lack of the ability to reveal the information of mutual influences among different global and local events that affect national security, it becomes challenging to identify how one event can affect another event or which event leads or lags another event when it comes to threats to national security.[74] Correlation among events does not imply causation, hence other methods are needed to construct directed networks to catch the embedded causal relationships among the inter-influences of events related to national security.

The casual inferences driven by data in these complex dynamic systems can be challenging because the data are usually very high dimensional and nonlinear and also tends to be limited in terms of sample sizes. In big data analytics the analysis tools used employ the use of Generalized Partial Directed Coherence as a methodology that is used in the investigation, usually in the frequency domain, of the concept of Granger causality when dealing with multivariate time series. Baccala et al. introduced this methodology when trying to find a frequency-domain quantifier for the multivariate relationship and their paper was mainly to look at its application in functional connectivity inference in the field of neuroscience.[75]Their paper built upon that of Saito and

[73] Yong Tang et al., "How Do the Global Stock Markets Influence One Another? Evidence from Finance Big Data and Granger Causality Directed Network," *International Journal of Electronic Commerce* 23, no. 1 (2019): 85–109.
[74] Tang et al.
[75] Luiz A. Baccala, K. Sameshima, and D. Y. Takahashi, "Generalized Partial Directed Coherence," in *2007 15th International Conference on Digital Signal Processing* (IEEE, 2007), 163–166.

Harashima[76] who looked at Directed Coherence and its applications in analysing neural data to identify the direction of information flow following the principles of Granger Causality. It also was an extension of their paper that was earlier done in Baccala that focused on Partial Directed Coherence (PDC).[77]

Identifying causal relationships and quantifying their strength from observational time series data are key problems in disciplines dealing with complex dynamical systems and the concept of Granger causality helps in trying to explain these relationships. The concept of causality, introduced by Wiener[78] and Granger,[79] constitutes a basic notion for analyzing dynamic relationships between time series. It determines whether one time series is useful in giving a forecast on another time series. However, due to the issues of confounding effects that do not capture the instantaneous non-linear relationships, several extensions have been proposed to tackle the issue and GPDC is one of this extensions.[80]

GPDC offers the tool to identify such causal relationships and has made a lot of strides in the neuroscience field and has also been used by Tang et al.[81] in investigating the influence global stocks have on each other using the evidence from financial big data, which is one of its application outside the field of medicine. With the dynamic nature of technology, these techniques can have

[76] Y. Saito and H. Harashima, *Tracking of Information within Multichannel EEG Record-Casual Analysis in EEG. Recent Advances in EEG and EMG Data Processing* (Elsevier/North-Holland, Amsterdam, 1981).
[77] Koichi Sameshima and Luiz Antonio Baccalá, "Using Partial Directed Coherence to Describe Neuronal Ensemble Interactions," *Journal of Neuroscience Methods* 94, no. 1 (1999): 93–103.
[78] Norbert Wiener, "The Theory of Prediction. Modern Mathematics for Engineers," *New York*, 1956, 165–190.
[79] C. W. J. Granger, "Investigating Causal Relations by Econometric Models and Cross-Spectral Methods," in *Essays in Econometrics: Collected Papers of Clive WJ Granger*, 2001, 31–47.
[80] Xiaojun Song and Abderrahim Taamouti, "A Better Understanding of Granger Causality Analysis: A Big Data Environment," *Oxford Bulletin of Economics and Statistics* 81, no. 4 (2019): 911–936.
[81] Tang et al., "How Do the Global Stock Markets Influence One Another?"

extensions to the national security field to aid in understanding the complex systems that govern different events in the wake of big data and technology.

2.3.2 Social Media intelligence

Social Media Intelligence (SOCMINT) can be considered to be the tools that can allow a certain agency to respond to the various social signals possible and then make analysis of these data points to have a meaningful trend that can be used for specific inferences.[82] Social media has made it possible for people globally to plan revolutions and even make recruitments into terrorists groups in addition to inciting violence while to most, it can conjure up thoughts of friends connecting with each other on the various platforms such as Facebook or Instagram. However, several activists and even individuals have started using it to connect with people to make their voices heard even more by a broader audience in addition to coordinating actions against the government and explain their side of the story.

Platforms such as Facebook can be considered to pose a threat to the national security of countries and foreign enemies can use it and other similar platforms to recruit and pass messages to their members in forms of propaganda and this can lead to radicalization.[83] For instance, Al-Shabaab runs a twitter account and use blogs to recruit and coordinate activities with different factions internationally.[84] SOCMINT technologies can be used to crowd-source information ensuring a better flow of information between citizens and their governments, especially in times where there are emergencies. Access to it can cause passive bystanders to become active citizen journalists

[82] David Omand, Jamie Bartlett, and Carl Miller, "Introducing Social Media Intelligence (SOCMINT)," *Intelligence and National Security* 27, no. 6 (2012): 801–823.
[83] Hammaad Salik and Zaheema Iqbal, "Social Media and National Security," *The Geopolitics*, September 10, 2019, https://thegeopolitics.com/social-media-and-national-security/.
[84] David Omand, Jamie Bartlett, and Carl Miller, "Introducing Social Media Intelligence (SOCMINT)," *Intelligence and National Security* 27 (December 1, 2012), https://doi.org/10.1080/02684527.2012.716965.

who relay information straight from the ground in areas where government agencies would otherwise have been unable to reach.

Social media has helped people globally organize revolutions and riots, recruit terrorists, encourage attacks, glorify gangs and spread violence. To most, social media conjures up thoughts of long-forgotten friends connecting with each other on Facebook, and aspiring performers posting videos of their antics on YouTube. However, activists and individuals globally have begun using social media to connect with each other, amplify their voices, coordinate actions against government and law enforcement, and publicize their side of the story.

Additionally, SOCMINT can be used in research to create understanding on various issues pertinent to the national security of a nation such that the law enforcement can generate operational intelligence that could be used to identify criminal activities or even give early warnings on incidences of disorderly conduct as well as understand the concerns of the public. Some of this information can be obtained on the public net as it is open source however, there are instances where the government may have to break ethical boundaries and invade the privacy of certain individuals for the safety of the general public and this can be a controversial area as the boundaries are not clear cut.[85]

Mutahi and Kimari look at how social media and digital technology influenced the electoral violence in Kenya during the 2007 general elections. They opine that social media technology was used to propagate hate speech and mobilize for violence and this was used to help the intelligence community map out violence hotspots around the country. Their research sought to present opportunities that exist for possible partnerships between the state and non-state actors to help in

[85] Maxim Pinkovskiy and Xavier Sala-i-Martin, "Lights, Camera… Income! Illuminating the National Accounts-Household Surveys Debate," *The Quarterly Journal of Economics* 131, no. 2 (2016): 579–631.

the prevention of political violence in Kenya. Their findings showed that the effectiveness of national legislative and judicial efforts to combat online hate speech, incitement and mobilization to violence remains limited, with little success. They suggested that the government should consider other measures that may complement whatever limited legislative and judicial mechanisms exist such as self- regulation and the exploring the use of big data while maintaining a balance between protecting freedom of expression and combating hate speech.[86]

Such a potential by these tools then suggest that SOCMINT will have a useful place in the national intelligence cycle of a country. However, where new forms of technology come into existence, it usually takes a bit of time for there to be a rigorous infrastructure that can allow the intelligence cycle to benefit from these tools. There are various challenges that do need to be looked into before SOCMINT can be fully utilized in the interest of national security.[87]

2.4 Big Data and the Intelligence Cycle

Puyvelde et al. notes that big data analytics plays a key role in the discovery and analysis of both trends and threats to national security. Key to note is that one emerging issue in the use of big data in the intelligence community occur when the voluminous big data collected is compressed. In this regard, the authors proposed a review of the impact of big data on the intelligence cycles through the lens of six core areas: *requirements*, *collection*, *collation*, *analysis*, *dissemination*, and *security*.[88] The reason behind the above focus by the scholar is informed by the fact that the six core areas lie at the core of the decision-making process of the national intelligence security framework. With regards to the *requirements functionality*, policy makers direct the intelligence

[86] P Mutahi and B Kimari, "The Impact of Social Media and Digital Technology on Electoral Violence in Kenya," 2017.
[87] Omand, Bartlett, and Miller, "Introducing Social Media Intelligence (SOCMINT)," December 1, 2012.
[88] Van Puyvelde, Coulthart, and Hossain, "Beyond the Buzzword."

requirements to intelligence managers who in turn direct intelligence analysts to direct big data analytics to differentiate general trends and anomalies with the detection of anomalies and associative algorithms. Consequently, the requirements and collection process may become inverted.[89] In other words, where a specific target would be required, it then holds that more voluminous data would be required to be collected so as to enable the machine sort out data for potential matches. *Collation* through the use of big data analytics within the intelligence community became apparent after the release of classified information by Edward Snowden who stole around 1.7 million classified documents from the NSA which exposed many secret illegal programs that it was conducting against the citizens of its country.[90] Big data analytical programs such as natural language processing (NLP) were able to implement structure to the unstructured 1.7 million data sets. In this regard for example, software programs could automatically transcribe audio data and collate it in such a manner that allows the data to be searchable through the use of key words.[91]

Verble further notes that other example of use of the big data within the intelligence collation process would be in analysing the sentiments expressed by Twitter users by region through the Geofeedia program software. Despite the vast potential, he cautions that big data does not necessarily imply all data as for example, social media sentiment analysis, is only as useful where the targets use the tools. Moreover, even with the availability of the above tool, the intelligence analyst is still required to use discretion in deciding how to contextualize the data output.

[89] Van Puyvelde, Coulthart, and Hossain.
[90] Joseph Verble, "The NSA and Edward Snowden: Surveillance in the 21st Century," *ACM SIGCAS Computers and Society* 44, no. 3 (2014): 14–20.
[91] Verble.

George and Bruce made contribution to the *analysis* component of big data analytics on the intelligence cycle. In this regard, the process is based on correlation that in turn facilitates in answering the who, what, where and when questions.[92] This allows the intelligence analyst to map out potential trends that adds values to current opinions and facts within the intelligence community. For example, the authors mention the anomaly detection function of big data analytics programs that enable one to anticipate threats. Such anticipatory output was evident during the 2011 Egyptian revolution where the CIA was able to forecast social instability and unrest to within a measure of three to five days.[93] However, despite the immense role played by big data analytics in analysing intelligence information, its ability cannot replace the human analysts who must be retained as an essential component as the human analyst is able to add context and make judgements on the data analysed.[94] Absent of context, software may present data that is biased that if used on face value may result in inappropriate intelligence action taken.

Treverton makes contribution on the role of big data analytics on the intelligence cycle. The scholar points out that the intelligence community utilizes big data analytics tools to *disseminate* analysis, in the form of recommendation, to different intelligence operatives depending on one's designation. For example, visualization tools such as Palantir may be used to display a map of the evolving trends of conflict zones in terms of people displacement and position of armed groups over a period of time.[95] Such then implies that such a tool is critical in aiding the intelligence analyst to shape their briefing report especially where the focus is trends over time. Suffice to say that the above software, or like-functioned ones, may play a critical role in the visualization of

[92] George and Bruce, *Analyzing Intelligence*.
[93] Robert E. Wilson, Samuel D. Gosling, and Lindsay T. Graham, "A Review of Facebook Research in the Social Sciences," *Perspectives on Psychological Science* 7, no. 3 (2012): 203–220.
[94] George and Bruce, *Analyzing Intelligence*.
[95] Gregory F. Treverton, *New Tools for Collaboration* (Rowman & Littlefield, 2016).

terrorism data for groups such as Al Shabab that may in turn be useful in the disruption of the same.

With regards to the *security* component of the intelligence cycle, Guira and Wang noted that big data analytics plays a key role in the deterrence of security threats. For example, the scholars take note of the NLP capabilities that enable the analyst to detect malware in the cyberspace. But then again, even such noble ability of big data analytics cannot deny the fact that the human component of this stage of the intelligence cycles is essential in detecting social engineering scams that programs may be unable to detect.[96]

One overriding conclusion from the above is the fact that whereas big data analytics aids in expanding the intelligence capabilities, it may not alter the human aspects of intelligence.[97] Such a point is particularly essential in unpredictable environments where dramatic changes are the norm that in turn make automated analysis unable to factor such in real-time.

2.5 Role of Intelligence in National Security Decision Making

Decision making as in policy making, faces the issue of whether there can be an intelligence free national security decisions. Decision makers can and often do formulate polices and plans without seeking insights and information from the intelligence analysts. However, it would be exceedingly difficult for decision makers to implement them without knowing what intelligence has to say regarding their efficacy and unintended consequences.[98] Therefore, the role of intelligence is to provide intelligence derived information and insights to inform the decisions and not to dictate

[96] Erik Gartzke and Jon R. Lindsay, "Weaving Tangled Webs: Offense, Defense, and Deception in Cyberspace," *Security Studies* 24, no. 2 (2015): 316–348.
[97] Van Puyvelde, Coulthart, and Hossain, "Beyond the Buzzword."
[98] Richard H. Immerman, "Intelligence and Strategy: Historicizing Psychology, Policy, and Politics," *Diplomatic History* 32, no. 1 (2008): 1–23.

them. Intelligence does not make or define strategies, but they are likely to be more successful if they incorporate the information and insights from intelligence.[99]

Strategic intelligence is concerned with the long-term strategic issues of the state and its supporting functions. It operates together with intelligence, visionary management and strategic foresight[100]. The mission of strategic intelligence is to understand past events, accurately analyze and interpret current conditions and predict the future. The product of this intellectual function is used by senior policy makers who make strategic national security[101]. Defective national security decisions result from use of wrong information, lack of sufficient information and disregard of available information by policy makers. Strategic intelligence analysts provide policy makers with intelligence that will help them in making different views on an issue or change the view they had before.

Intelligence contributes to the making of national security decisions by providing support, objectivity and information. Intelligence community exists to support policymakers with its primary missions being to reduce uncertainty, provide warning, and identify opportunities, making it an ideal instrument for collecting and assessing information applicable to the formulation of security strategies. As a support function, intelligence is guided by the needs and priorities of the individuals and institutions it serves. Policy makers rely on the IC for support to ensure that they have all the information needed to understand the dynamics of the challenges they want to address and the effectiveness of the strategies they develop to meet those challenges.[102] Objectivity is

[99] Thomas Fingar, *Reducing Uncertainty: Intelligence Analysis and National Security* (Stanford University Press, 2011).
[100] Tuomo Kuosa, *Towards Strategic Intelligence: Foresight, Intelligence, and Policy-Making* (Dynamic Futures, 2014).
[101] Harry Howe Ransom, "Strategic Intelligence and Foreign Policy," *World Politics* 27, no. 1 (1974): 131–46.
[102] Richard K. Betts, *Enemies of Intelligence: Knowledge and Power in American National Security* (Columbia University Press, 2009).

critical to the way issues are assessed by intelligence analysts. Policy makers have ideas about what policies and strategies they want to accomplish. These predetermined ideas shape the requirements they give the IC and specify the issues they want assessed by the analysts but they do not and should not specify how they will be assessed. This is important so as to avoid politicization of the intelligence process. Politicization has been described as the willful distortion of information to serve interests which it was not originally aimed to serve, which can render such information useless.[103] The work of the intelligence analysts is therefore to give objective analytic judgements regardless of how the policy makers will receive the intelligence findings.[104]

Intelligence is therefore crucial for the formulation and implementation of the national security decisions. Proper implementation needs proper feedback on the implementation process, since things can go wrong from the many unanticipated consequences arising from the VUCA environment. Intelligence analyses and assesses these consequences and gives the insight to the decision makers, helping them adjust the policies with regards to these consequences. The work of the IC is to ensure that the policy makers have as much information as possible from both classified and open sources. They are both crucial for protecting the national interests of the state and in helping decision makers do their job. Intelligence does not make or define its efficacy, but it can make strategies better.

[103] Richard K. Betts and Thomas Mahnken, "Politicization of Intelligence Costs and Benefits," in *Paradoxes of Strategic Intelligence* (Routledge, 2004), 70–89.
[104] Fingar, "Intelligence and Grand Strategy."

2.5 Application of Big Data in National Security

This section aims to explore the practical applications of big data analytics in addressing threats to national security. The focus is on application of big data and analytics in the areas of cyber security, counter terrorism and human and drug trafficking.

2.5.1 Cyber Security

Tankard opined that big data analytics plays a critical role in the cyber security due to its operational efficiency in the detection of cyber-attacks. For this to work, the scholar proposes the separation of security access controls and networks and instead have the said controls closer to the data set that requires the protection.[105] Key to note here is that, organizations holding volumes of big data are often prone to cyber-attacks as intruders attempt to access such information for their own motives.

Cardenas et al. concur with the above notion on the role of big data analytics on cyber security. The scholars argue that the analysis on network traffic and system logs is critical in the detection of potential attacks with ease over a short period of time unlike traditional means of analysing event logs for threat detection.[106] With the traditional methods, analysing large volumes over a long period of time was not economical as the storage of vast data over prolonged period required vast resources. However, with big data, the scholars aver that, organizations are better placed to develop infrastructure that stores and maintains such data for both present and future analysis cost effectively. Despite the above developments, Cardenas et al. make it clear that big data analytics cannot be considered as the ultimate solution to cyber security as cyber aggressors would often adapt to every new measure put in place including the above.

[105] Tankard, "Big Data Security."
[106] Cárdenas, Manadhata, and Rajan, "Big Data Analytics for Security."

Curry et al. further opined that big analytics had the potential to alter the nature of the vast majority of information security products (such as IDS, fraud detection, network monitoring, and authentication) that may in the fullness of time evolve to encompass advanced real-time and predictive features. Ultimately, the scholars foresee a scenario where big data analytics would improve cyber security mechanism especially with the increased frequency in sophisticated cyber-attacks.[107] Interestingly, the scholars noted that one unwanted outcome in the growth of big data analytics is the dissolution of network boundaries to permit the flow of data across mobile devices and cloud services that in turn increased the ease in which attack surfaces increase in occurrence and sophistication. In this regard, Curry et al. noted that for systems to retain their integrity, information systems needed to rely more on agile dynamic risk assessments.

Brewer however, notes that the current information security infrastructure lacks a network perimeter thus any attempt at stopping a network intrusion at such a perimeter is likely to fail.[108] Moreover, preventing intrusions have been complicated by the speed and stealth at which the intrusion occurs. With the above challenges, the scholar suggests the need for a fundamental shift in operations from prevention to detection (as it occurs) through the use of big data analytics. Mahmood and Afzal reinforce the calls for the use of big data analytics being central to the deterrence of cyber-attacks. For the big data analytics to be effective in cyber-attack deterrence, the data ought to emanate from diverse sources with the user interface being interactive and the analytics engine being sophisticated.[109]

[107] Sam Curry et al., "Big Data Fuels Intelligence-Driven Security," *RSA Security Brief*, 2013.
[108] Ross Brewer, "Cyber Threats: Reducing the Time to Detection and Response," *Network Security* 2015, no. 5 (2015): 5–8.
[109] Tariq Mahmood and Uzma Afzal, "Security Analytics: Big Data Analytics for Cybersecurity: A Review of Trends, Techniques and Tools," in *2013 2nd National Conference on Information Assurance (Ncia)* (IEEE, 2013), 129–134.

With the above calls for a radical change from the traditional cyber security solutions, Ahn et al. point out why there is need to change from a systems review point of view. Central to the argument is that traditional cyber security solutions were based on signatures and features of known malwares.[110] Consequently, in the event of new malware such as advanced persistent threats, (APTs) that lacked known signatures, detection and deterrence would be impossible. As a solution, the scholars suggest that a deterrence model fashioned around a three-stage big data analytics model: data collection, data processing, and an alert system. In the first stage, data collection would encompass collation of data from traditional sources that is mined using trending data mining methods for the purpose of determining abnormal information security behaviours.

2.5.2 Counter Terrorism

The September 11, 2001 terrorist attack in the United States marked the dawn of the era of global terrorism anchored on the new trends of technological advancement mainly in communication through the internet, social media and a voluminous open source information. States are now alive to the fact that in order to respond to this threat, counterterrorism must be intelligence led and collaborative using both covert and overt means by the intelligence and law enforcement agencies. Akhgar et al. opines that big data analytics can be used in identification of terrorist networks and their associations by using social media analysis of open source intelligence (OSINT) to identify radical roots in online communities. This can concurrently be applied in the intelligence and

[110] Sung-Hwan Ahn, Nam-Uk Kim, and Tai-Myoung Chung, "Big Data Analysis System Concept for Detecting Unknown Attacks," in *16th International Conference on Advanced Communication Technology* (IEEE, 2014), 269–272.

knowledge requirement for the prevent, pursue, protect and prepare pillars of the counterterrorism strategy.[111]

Dolenko and Lobach undertook a study of terror activities between 2000 and 2010 conducted in Algeria, Afghanistan, and Iraq factoring in key attributes such as geo-political zones, region, date and city. Through the use of exponential distribution model in studying the intervals between a series of terror activities, the scholars were able to predict the next attack. The scholars argued that the insights they derived from the analysis could be used to enhance counter terrorism efforts[112]. Whereas the method used by the scholars was admirable, that applied earlier by Kyung et al. resulted in a finer precision in the prediction of the next potential terror attack as it relied on the Dirichlet process which was introduced by Ferguson in 1973[113]. Their study recognizes the need to use sophisticated modelling to draw reasonable inferences from data describing terrorist events which usually have poor measurement levels coming from the observation of political actors who do not seek to provide reliable data on their activities.[114]Their model had a random-effects specification that used the Dirichlet process which they found to be better than the normal random-effects model since it could remove the underlying variability from the data hence providing latent information which would otherwise not have been revealed. Kolajo and Daramola propose Social Media Analysis for Combating Terrorism (SMACT) model that uses data from multiple social media platforms to detect terrorist activities by using Apache Spark technology for implementation

[111] Babak Akhgar et al., *Application of Big Data for National Security: A Practitioner's Guide to Emerging Technologies* (Butterworth-Heinemann, 2015).
[112] G. Dolenko, and S. Lobach, "System and statistical approach of analysing and forecasting terrorist activity," Model Assisted Statistics and Applications, vol. 9(3), pp.267-275, 2014.
[113] M. Kyung, J. Gill, and G. Casella, "New findings from terrorism data: Dirichlet process process random-effects models for latent groups," Journal of the Royal Statistical Society, vol. 60(5), pp.701-721, 2011.
[114] Ibid

as a useful approach that leverages Big Data analytics to address the terrorism threat in developing nations.[115]

Nusratullah et al. focused on how to use big data in the detection of changes in the use of social media network usages through time series analysis. The focus of the scholars was in analysing emails and identification of changes in the frequency of communication over time. For example, the scholars were keen to monitor communication from a person to another in Libya especially where such ties were previously non-existent; the individual would then be tracked for acquisition of further information[116]. The scholars further stated that the data could be enhanced through integrating it with a facial tracking system that would track the movement of the target continuously. Suffice to say that the facial recognition tracking would only work in an environment where camera surveillance is available.

In addition to the above, Piu et al. pursued a study of mining relationship amongst online users of online discussions through the use of sentiment analysis to define subgroups from the online forum. The above was achieved by developing discussant attitude profiles that were later clustered into correlative subgroups. Once the above was achieved, the scholars proceeded to subject the data to probabilistic matrix factorization in order to obtain a "low rank representation from the opinion scores".[117] The scholars claim that such information then allows potential terrorists and their sympathizers to be detected and singled out from the population for further security action. Gerber undertook a similar work in topical modelling over twitter posts with a view of signalling

[115] Taiwo Kolajo and Olawande Daramola, "Leveraging Big Data to Combat Terrorism in Developing Countries," in *2017 Conference on Information Communication Technology and Society (ICTAS)* (IEEE, 2017), 1–6.
[116] K. Nusratullah, A. Shah,, S.A. Khan, and W. H. Butt, "Detecting changes in context using time series analysis on social network," SAI Intelligent System Conference, November, 10-11, 2015, London, UK.
[117] M. Piu, L. Yang, and J. Jiang, "Mining user relations from online discussions using sentiment analysis and probabilistic matrix factorization," Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics – Human Language Technology, Atlanta, pp.401-410, 9-14 June, 2013.

out discussion topics that would later be integrated with crime prediction model. Key to note is that this model was compared with kernel density estimation that in turn revealed that the topic modelling was more efficient[118].

Vajjhala et al. undertook a study of terrorist activities between 1970 and 2013 with a view of preparing a summary of visual trends in terror activities that would in turn be used to predict terror activities.[119] Their study identified a gap that there were few examples that applied open big data analytics in terrorism and they used correspondence analysis to find hidden factor relationships that could exist in decision making. Similarly, Baig and Jabeen opined that big data analytics could be used to monitor students with a view of predicting whether a student's deviant ideologies may lead to one becoming a potential terrorist. To achieve this, the model determined a threshold upon which a student who surpassed it would call for closer monitoring[120].

In Kenya, Alfano and Gorlach looked into how al-Shabaab attacks in Kenya are related to education and perceptions. While undertaking their research, they used geo-coded data to examine how the attacks that the distance from children's way to school or also their other side of international borders affects their school enrolment. They then used this data alongside other behavioural responses from their study to model earnings loss that arises from the reduction in schooling cause by terror attacks in a country.[121]

[118] M. S. Gerber, "Predicting crime using twitter and kernel density estimation," Decision Support System, vol. 61, pp.115-125, 2014

[119] N. R. Vajjhala, K. D. Strang, and Z. Sun, "Statistical modeling and visualizing open big data using a terrorism case study," IEEE Computer Society, pp.489-495, 2015.

[120] A. R. Baig, and H. Jabeen, "Big data analytics for behaviour monitoring of students," Procedia Computer Science, vol. 82, pp.43- 48. 2016.

[121] Marco Alfano and Joseph-Simon Görlach, "Terrorism, Education, and the Role of Expectations: Evidence from al-Shabaab Attacks in Kenya," 2019.

Nyabira et al further looked into the effect of police training programme on the counter terrorism capability of Kenya. The study interviewed a sample of 85 respondents and in their results they found that the programme delivery method had a positive impact in the counter terrorism capabilities of the police officers. Part of the content of the programme was on the use of IT in combating terrorism though it wasn't specified to be on big data analytics. The study however, made a recommendation that the programme should consider aligning itself with the current international trends in IT to be more effective in combating terrorism in Kenya.[122]

From the above subsection, it becomes evident that the literature presented focused on a single type of social media with no effort made to undertake a study of a variety of social media platforms. Moreover, the studies have been largely implemented within the developed nations and thus leaving the need to undertake a similar study in the developing nations.

2.5.3 Human and Drug Trafficking

According to a study by the International Labour Organisation, human trafficking victimizes around 40 million people globally with most of these victims being women and children. These victims are forced into industrial labour, prostitution and various other forms of exploitation and governments and NGOs have engaged in various activities to try thwart these cases. One of the approaches that is taken in tackling this menace is the 'follow the money' approach where governments have worked with financial institutions to develop anti-money laundering systems that identify and expose these activities and thus makes it easier to facilitate information sharing.[123]For instance, IBM has developed a cloud hosted data centre that enables banks and other financial institutions to access enhanced information with the help of augmented intelligence and

[122] Ben Christopher Nyabira and Zemelak Ayitenew Ayele, "The State of Political Inclusion of Ethnic Communities under Kenya's Devolved System," *Law, Democracy & Development* 20, no. 1 (2016): 131–153.
[123] Lopez, Rolando R. "Battling Human Trafficking with Big Data." (2014).

machine learning which will detect certain human and drug trafficking incidences.[124]Additionally, the augmented intelligence can use a lot of open source data from various social platforms to identify the characteristics of human and drug trafficking incidences such as recruitment or transportation methods.[125]

Grothaus notes that between the years 2007 and 2015, Google and Palantir created a database and an analytics platform that utilizes big data for the National Human Trafficking Resource Center (NHTRC), a non-profit initiative in the US, to gather and analyse trafficking data through pattern recognition and enable coordinated responses. [126] During that period, it received over one hundred thousand calls on top of thousands of email and web submission regarding human trafficking.[127] As the database has grown, it has been able to identify patterns across state lines and trace cases of recruitment by fake organizations involved with trafficking.

The drug trade has become a high-tech enterprise, as traffickers turn to technology to track shipments, communicate, sell their product and more. Governments have naturally responded in kind with technology to investigate and apprehend drug traffickers. However, the biggest challenge has not been the ability to gather information but rather sorting through the vast amount of data and analyzing it for potential patterns for actionable decisions. [128]According to a report by Thompson Reuters,[129] law enforcement in the US have partnered with Reuter to get a platform

[124] Joh, Elizabeth E. "The new surveillance discretion: Automated suspicion, big data, and policing." *Harv. L. & Pol'y Rev.* 10 (2016): 15.
[125] Ibid
[126] Michael Grothaus, "How Google is Fighting Sex Trafficking with Big Data", Fast Company, March 14, 2013. Accessed 02/04/2020. https://www.fastcompany.com/3009686/how-google-is-fighting-sex-trafficking-with-big-data
[127] Ibid
[128] Connell, Elizabeth, Steven Jones, and Javonda Williams. "Human Trafficking and the Transportation Profession: How Can We Be Part of the Solution?." *Institute of Transportation Engineers. ITE Journal* 88, no. 7 (2018): 45-49.
[129] Thompson Reuters, "Haystacks of Needles: Law Enforcement Fights Organized Crime with Smart Analytics". Accessed 02/04/2020. clear.thomsonreuters.com/analytics.

called TETRA which allows them to normalize data and identify the most relevant information more efficiently through big data analytics on smartphone data.

Hong Kong customs authorities are using AI to tackle drug trafficking by scouring social media websites to help officers find drugs and fake goods being sold online and according to Leung, this has tripled the number of cases that are being solved over a span of one year.[130]The system helps in reducing the cases that the officers have to look through by shortlisting the cases it considers high risk and then the officers can spend time looking into them to prevent smuggling. A similar software, called DataWalk, is being used in the US by the state of Philadelphia to reveal patterns and anomalies for large scale multi-source intelligence operations. The software provides a singular data view using intuitive visualizations including histograms, link charts, maps, and timelines for faster intelligence-led decision-making using secure workflows for capturing organizational knowledge and delivering accurate, complete, and consistent results.[131]

2.6 Ethical Issues in Big Data

Newell and Marabelli opined that one needs to understand the underlying moral premise for making a case for understanding how ethical issues affect big data analytics. In this regard, the scholars suggested the need for an appreciation to the utilitarian and Kantian viewpoints on the principles of truth and moral guidelines[132]. Here, the Kantian point of view is that any ethical action is based on principles and moral values that in turn differs with the utilitarian point of view which

[130] Leung, Christy, "Customs looks to AI to combat more cases of smuggling of drugs and cigarettes in Hong Kong, using it to sift through thousands of online posts". August 7 2019. Accesed 06/02/2020.
[131] Newell, S., and Marabelli, M. 2015. "Strategic opportunities (and challenges) of algorithmic decisionmaking: A call for action on the long-term societal effects of 'datification,'" The Journal of Strategic Information Systems, Elsevier B.V., pp. 1–12 (doi: 10.1016/j.jsis.2015.02.001).
[132] Ibid

focuses on outcomes and consequences. In a nutshell, an action can only be considered ethical where the outcome of such an exercise is beneficial to the majority of that community.

With the above in mind, it becomes apparent that as big data analytics increases in utility for the community, it tends to encompass more responsibilities and thus calls for greater ethical considerations to take into account the dynamics. Such are the dynamics that presents the case for the need for a literature review on ethical issues revolving big data analytics. Naturally, any such ethical consideration leads to the full realization that there are social consequences when the ethical considerations are not considered.

Lv et al. opined that a central issue on the ethical use of big data revolves around is data security and privacy. The underlying reasoning behind the above concern is the nature of the content of big data that contains personal user details such as health and financial information that in turn calls for a special form of data transmission. In this regard, the scholar calls for the installation of secure data transmission protocols that would safeguard the data from potential data leaks that would in turn place the privacy of the users in jeopardy. Whereas the solution to the above issue requires the data handler to implement secured certification mechanisms, such protocols are often complex and costly to implement. Moreover, where the data handler considers anonymization, the data confidence would reduce proportionally.[133]

Clarke noted that the increased use of algorithmic decision making based on the overall premise that the selected relationships are not only meaningful but also an objective representation of cause and effect. Key to note here is that algorithms function by collation of collected data from various

[133] Lv, Z., Song, H., Basanta-Val, P., Steed, A. and Jo, M., 2017. IEEE Transactions on Industrial Informatics journal, Volume 13, pp. 1891-1899.

platform.[134] The scholar argues that users may decide to present a false narrative on their online social media presence thus implying that the where the algorithm analyse such information, the information it presents would be factually incorrect as its based its analyses on an incorrect data set. In other words, algorithmic decision making often fails to take cognisance of possible bias in data, measure, and analysis as well as human design error.

Martin further suggested that big data analytics could lead to the production of profiles that would in turn be used in the discrimination of users in the community based on certain metrics. For example, a target could be incorrectly placed in a subgroup such as sex, sexual orientation, economic status, and race that would in turn be used by the data handlers in discriminating an individual thus resulting in an unethical social impact especially where such a classification would make the target locked out from certain benefits based on the classification.[135] In other words, the algorithms result in outcomes that disregard key outliers such as the need to revert back to mean.

Zuboff opined that surveillance and target monitoring is an essential component in the generation of big data that has the undesirable effect of eliminating free choice from an individual. The reason behind the above assertion is based on the operationalization of big data analytics; organizations analyse surveillance data of a target user and make comments and predictions on what they perceive as to be the taste and preference of the user.[136] In this regard, commercial firms would result in only advertising products and services based of the decisions made by the algorithm thus making the user miss out on other options that might have been of interest to them. In simpler terms, where the user results into purchasing products and services based on the advertisements

[134] Clarke, R. 2016. "Big data, big risks," Information Systems Journal (26:1), pp. 77–90 (doi: 10.1111/isj.12088).
[135] Martin, K. E. 2015. "Ethical Issues in the Big Data Industry," MIS Quarterly Executive (14:2), pp. 67– 85 (doi: 1540-1960).
[136] Zuboff, S. 2015. "Big other: surveillance capitalism and the prospects of an information civilization," Journal of Information Technology (30:1), Nature Publishing Group, pp. 75–89 (doi: 10.1057/jit.2015.5).

made from the algorithm, then it can be argued that the user has lost their freedom of choice as the algorithms is now responsible for determining their choices.

Someh et al. further opined that another ethical issue emerges in the use of the data collected by organizations. Whereas data collected by one party may be given consent by the owner of the information, the said person might not be fully aware of the fact that the data might exchange parties, absent of their permission. Ordinarily, such data collected often moves through the information value chain from the companies undertaking the data tracking through the data aggressors and finally the retailers.[137] Key to note is that data aggressors often combine data sets of an individual from various sources to produce a profile on an individual. With the data handling and purpose of the data changing along the value chain, it becomes apparent that the final use of the data might be markedly different from what the owner of the data accented to. Therefore, at different stages of the value chain, risks might emerge for the party that was responsible for first collecting the data from the owner.

[137] Ida Someh et al., "Ethical Issues in Big Data Analytics: A Stakeholder Perspective," *Communications of the Association for Information Systems*, 2019, 718–47, https://doi.org/10.17705/1cais.04434.

CHAPTER THREE

RESEARCH METHODOLOGY

3.1 Introduction

This chapter outlines a comprehensive discourse on the research methodology that is used in collection and analysis of data and presentation of the research findings. The chapter is divided into the following parts; research design, target population, data collection methods, sampling techniques and data collection instruments.

3.2 Research design

Research design refers to the structure or strategy of obtaining data with the aim of answering the research questions of the study. It is the blueprint for collection, measurement and analysis of data.[138] This study largely depended on an exploratory design method. This type of research design is used when a problem is not well researched and it focuses on explaining the aspects of the study in a more detailed manner. Its aim is not to give conclusive evidence but rather provide an understanding of the problem and obtain insights that can be used to provide suggestions for implementation of solutions to the problem.

3.3 Target Population

For the primary data, the population of interest was drawn from the staff working in the government agencies whose operations are critical to national security. These agencies of interest include the Ministry of Interior and Coordination of National Government, National Police

[138] Olive M. Mugenda and Abel G. Mugenda, *Research Methods: Quantitative and Qualitative Approaches* (Acts press, 1999).

Service, Kenya Defense Forces and the National Intelligence Service as well as several big data experts. Well-designed questionnaires were used to obtain the data required for the research.

3.4 Sampling techniques

Sampling technique refers to the method used to select representative elements form the target population so as to generalize the findings of the study.[139] Sampling techniques can be classified as either probability or non-probability sampling. Due to the secretive nature of the study non-probability sampling method was used where the respondents were selected based on the personal judgement of the researcher. Purposive sampling is one of the non-probability sampling techniques in which the researcher uses their own judgement to select the population sample. Given the nonprobability nature of the sampling method, it may have a bit of bias within it but it can be very useful when the researcher has limited resources or workforce[140]. Purposive sampling was used to help in choosing the people that would provide the best information on the topic. Due to the sensitive nature of the information being sought, the information was limited in scope and this was supplemented by the information obtained from secondary sources.

3.5 Data Collection

Questionnaires were used as the method of data collection. The questionnaire employed the use of descriptive open ended questions as a way to obtain the respondents insight in the areas of interest for the research. Questionnaires are preferred because they are practical, inexpensive and are a simple way of getting valid results with the respondent anonymity maintained. However, they do fall prone to dishonest answers, differences in understanding and interpretation and if not designed

[139] Mark NK Saunders and Keith Townsend, "Reporting and Justifying the Number of Interview Participants in Organization and Workplace Research," *British Journal of Management* 27, no. 4 (2016): 836–852.
[140] Etikan, I., Musa, S. A., & Alkassim, R. S. (2016). Comparison of convenience sampling and purposive sampling. *American journal of theoretical and applied statistics*, *5*(1), 1-4.

correctly can be difficult to analyze. To prevent this, the researcher was available during the data collection process to provide clarity on any unclear question and reliability tests were conducted on the questionnaire to ensure that the questions were set correctly to address objectives of the study. Two methods were used to administer the questionnaire; one is where the researcher administered them personally and the other was by use of an online survey platform, where it was deployed. To ensure every option of obtaining feedback from the respondents was explored, emails were also sent to the respondents.

With regards to deciding what questions to ask, Leung[141] explains that there are two main types of information to be obtained:

1. The first type is the main information one is seeking to obtain from the chosen target audience. This is known as the dependent parameters which in the study was the 5 stages of the intelligence cycle.
2. The second type is the information that might bring more meaning to the main information (dependent variables). The second type is known as the independent parameters which in the study was the big data knowledge.

3.6 Reliability of Data collection instruments

To ascertain the reliability of the questionnaire and identify any weaknesses in the design and instrumentation, a pilot test was conducted on a few of the individuals from the sample of study. In addition to this, Cronbach's alpha test was conducted to also test the internal consistency of the questions in the questionnaire. Cronbach's alpha is a measure of internal consistency, that is, how

[141] Wai-Ching Leung, "How to Design a Questionnaire," *BMJ* 322, no. Suppl S6 (2001): 0106187.

closely related a set of items are as a group and it's considered to be a measure of scale reliability.[142] Using the Cronbach's alpha indicates whether the score is reliable, depending on the result of the test, which should lie between 0 and 1, whereby a score above 0.7 is the best.[143] This test has been used in various studies that have to work on data from questionnaires and a good example of such a study is the study by Taber which looked at how Cronbach's alpha is used when developing and reporting the research instruments used in education.[144]

3.7 Data Analysis

With respect to the nature of the information to be obtained in this research, the data analysis was conducted by means of descriptive statistical techniques such as frequency tables; to facilitate this process, the statistical application SPSS was used. In addition to this, inferential statistics was also conducted for causality to be established from the variables used for the study. To test the reliability of the questionnaire, a Cronbach's alphas test was conducted. The model that was used was a Multivariate Generalized Linear Model (GLM), the MANOVA variant. The MANOVA in multivariate GLM extends the ANOVA by taking into account multiple continuous dependent variables, and bundles them together into a weighted linear combination or composite variable. The independent variable for the data is big data knowledge and the dependent variables are the five aspects of the intelligence cycle. The MANOVA compares whether or not the newly created combination differs by the different groups, or levels, of the independent variable. In this way, the MANOVA essentially tests whether or not the independent grouping variable simultaneously

[142] Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E., & Tatham, R. L. (2006). others. Multivariate data analysis. *Pearson Prentice Hall Upper Saddle River*, *NJ*, *44623*(4.2), 2.
[143] Moonseong Heo, Namhee Kim, and Myles S. Faith, "Statistical Power as a Function of Cronbach Alpha of Instrument Questionnaire Items," *BMC Medical Research Methodology* 15, no. 1 (October 14, 2015): 86, https://doi.org/10.1186/s12874-015-0070-6.
[144] Taber, K. S. (2018). The use of Cronbach's alpha when developing and reporting research instruments in science education. *Research in Science Education*, *48*(6), 1273-1296.

explains a statistically significant amount of variance in the dependent variable. The method of estimation for the MANOVA is the Maximum Likelihood Estimation (MLE) and it is a twostep methodology which will involve diagnostic tests like the Hotelling's T square test, the Pillai's trace test and the most commonly used, Wilk's Lambda U test. The commonly used Wilk's lambda measures how well the methodology separates the data into groups by measuring the proportion of total variance in the scores that are not explained by the differences among the groups.[145]

This kind of analysis is appropriate when there are multiple dependent variables and these variables should be moderately correlated for MANOVA to be applicable. Unlike the ANOVA which looks at the differences across two groups or more on one dependent variable, the MANOVA looks at the same difference but over two or more dependent variables.[146] It does have its limitations one being that it is a more complex methodology compared to the simpler ANOVA methodology. In addition, its results can be a bit ambiguous especially if the assumptions of the model are not met. However, given the nature of the study where we have multiple dependent variables, this was the most appropriate methodology and the key assumption of some correlation between the dependent variables was met which meant that the methodology would be suited to giving better results.

3.8 Ethical Considerations

The process of data collection was guided by the ethical considerations of confidentiality, anonymity, responsibility, respect, competence, consent, security and understanding.

[145] Liu, C., Bathke, A. C., & Harrar, S. W. (2011). A nonparametric version of Wilks' lambda—Asymptotic results and small sample approximations. *Statistics & probability letters*, *81*(10), 1502-1506.
[146] Huberty, C. J., & Olejnik, S. (2006). Applied MANOVA and discriminant analysis (Vol. 498). John Wiley & Sons.

CHAPTER FOUR

ANALYSIS AND PRESENTATION OF FINDINGS

4.1 Introduction

This chapter delves into the data analysis and the interpretation of the findings. The analysis was structured into several themes that capture the research objectives. The data collected was analyzed for its descriptive statistics and their implications in line with the research objectives were explored. Additionally, a MANOVA analysis was run to further investigate the relationship between the variables of the study and how the findings relate to similar studies done on the same topic.

4.2 Rate of responses

Seventy (70) questionnaires were administered to the sample population of the study and 46 questionnaires were correctly filled and returned, which represented a 65.7% response rate which was adequate for the analysis.

4.3 Reliability test

The study made use of a questionnaire for data collection and to measure the internal consistency of the questions, Cronbach's alpha test was conducted to determine the scale reliability of the questions capturing five processes of the intelligence cycle. Table 5.1 below gives a summary of the results:

Variable	Cronbach's Alpha	Number of items
Planning and targeting	0.712	5
Collection	0.838	5

Processing and Evaluation	0.802	5
Analysis	0.781	5
Dissemination	0.689	5

Table 5.1: Cronbach's Alpha reliability statistics for the intelligence cycle variables

The findings indicates that most of the study variables had an alpha value greater than 0.7 with only dissemination having a value slightly lower than 0.7. This implies that the measures for the same constructs do exhibit high internal consistency and the scale can be considered to be reliable for use in further analysis.

4.4 Demographic Characteristics

This section describes the basic characteristics of the 46 respondents which include characteristics such as level of education, gender, years worked in the organization among others.

4.4.1 Education level of the respondents

The education level statistics of the respondents showed that 43.48% of the respondents had Degrees, 54.35% had Masters while 2.17% had PHDs as shown in Figure 3, implying that most of the employees in the organization have master's level education.

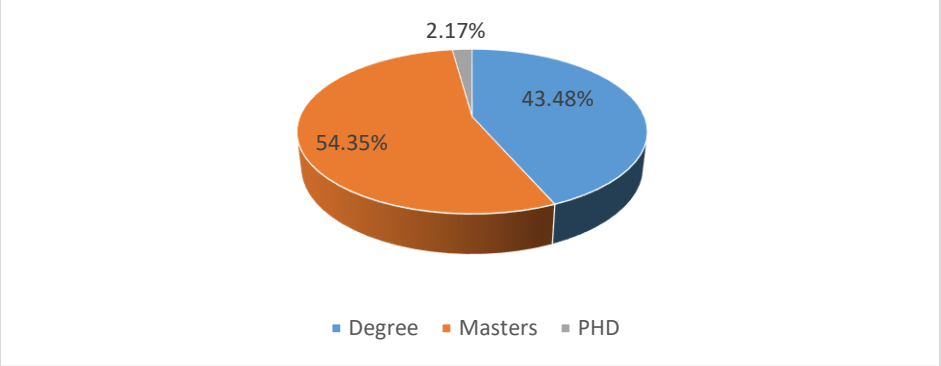


Figure 1: Education level summary

4.4.2 Years worked in the organization

A majority of the respondents had worked in their organization for more than 10 years while only a few of them had worked for less than two years. More than half of the respondents had worked in the company for more than 5 years which meant that most of the respondents were in an informed position to fill in the questionnaire. Figure 4 gives a summary of these statistics.

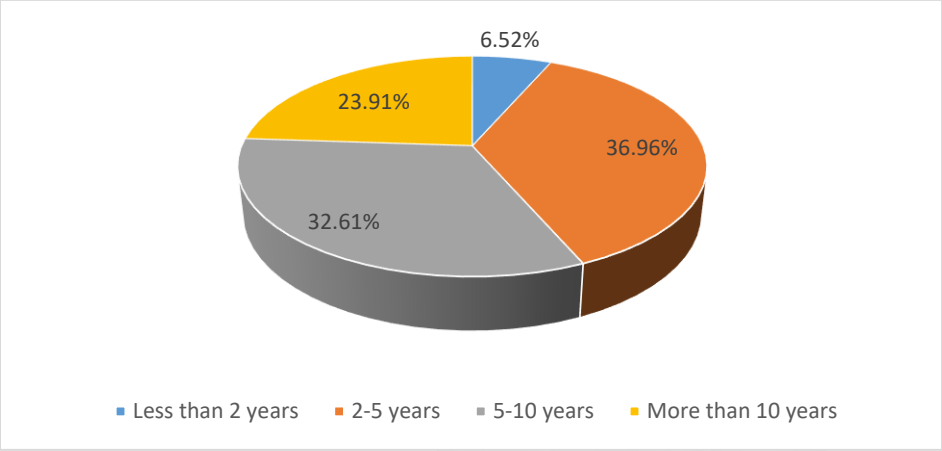


Figure 2: Number of years worked in the organization

4.4.3 Areas of National Security tackled by respondent

Respondents were asked to choose from a list of possible national security areas that they encounter in the execution of their duties and they could choose more than one choice. Terrorism was the most common area of national security tackled by the respondents followed by cybercrime then financial fraud and money laundering. Figure 5 gives a summary of the areas of national security that were chosen by the respondents. Terrorism was the most common since over the last few years, Kenya has been having terrorism incidences and according to Nyabira et al, the

government has had an appreciation on how big data can help combat terrorism and the data collected seems to indicate this.[147]

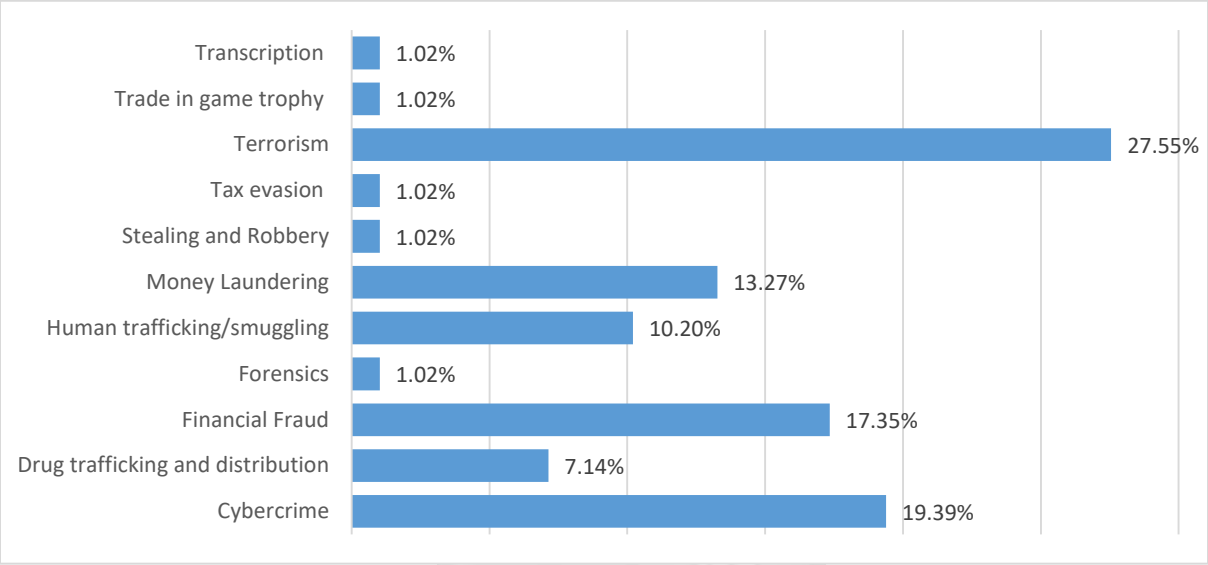


Figure 3: Areas of national security encountered by respondent in the execution of their duties

4.4.4 Emerging National Threats

Included in the questionnaire was a question on what the respondents felt were emerging national threats and cybercrime and terrorism were among the threats that were chosen. Figure 6 gives a summary on the various national threats. These threats are in line with trends in other countries as noted in the study by Dolenko and Lobach[148] and Alfano and Gorlach.[149] The scholars argued that the advent of technology had enabled most of the perpetrators of these acts to thrive in the new age of technology and as such has led to the rise of such threats and the responses reflected this as cybersecurity and terrorism were among the most chosen in regards to emerging threats.

[147] Nyabira and Ayele, "The State of Political Inclusion of Ethnic Communities under Kenya's Devolved System."
[148] Galyna Dolenko and Sviatoslav Lobach, "System and Statistical Approach of Analysis and Forecasting Terrorist Activity," *Model Assisted Statistics and Applications* 9, no. 3 (2014): 267–275.
[149] Alfano and Görlach, "Terrorism, Education, and the Role of Expectations."

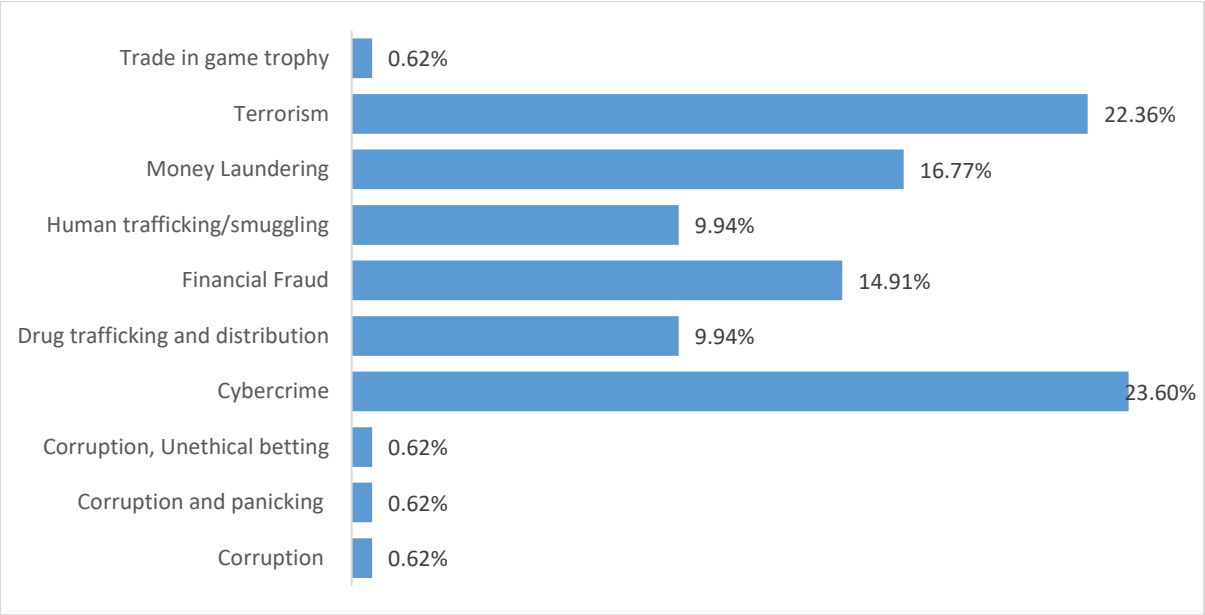


Figure 4: Emerging national threats

4.4.5 Big data knowledge

The knowledge of big data of the respondents was captured and in the summary, a break down in terms of the level of education was done to have a better understanding if the level of education had any relationship with the knowledge of big data. Generally, most of the respondents, around 52.17%, had moderate knowledge of big data with around 6.52% not having any knowledge at all on big data. An interesting observation is that the respondents who did not have any knowledge of big data had a Masters level of education and also those that had very good knowledge of big data had their Masters. The only respondent with a PHD had moderate knowledge of big data and there was one respondent who did not fill in this part of the questionnaire. Table 5.2 provides further analysis of big data knowledge and the level of education. This is in line with what other studies such as Dolenko and Lobach[150] or Leung[151] regarding the level of education level where they found

[150] Dolenko and Lobach, "System and Statistical Approach of Analysis and Forecasting Terrorist Activity."
[151] Christy Leung, "Customs Looks to AI to Combat More Cases of Smuggling of Drugs and Cigarettes in Hong Kong, Using It to Sift through Thousands of Online Posts," August 7, 2019, https://www.thestar.com.my/authors?q=Christy%20Leung.

that that having a university degree was in most cases enough to make people aware of big data but they found that increasingly over the years, departments such as Homeland Security are increasingly hiring employees that have a technology background.

Big data Knowledge	Degree	Masters	PHD	Total
Not at all	0.00%	6.52%	0.00%	**6.52%**
To a small extent	13.04%	2.17%	0.00%	**15.22%**
To a moderate extent	23.91%	26.09%	2.17%	**52.17%**
To a large extent	4.35%	8.70%	0.00%	**13.04%**
To a very large extent	2.17%	8.70%	0.00%	**10.87%**
(blank)	0.00%	2.17%	0.00%	**2.17%**
Total	**43.48%**	**54.35%**	**2.17%**	**100.00%**

Table 5.2: Summary of Big data knowledge

4.4.6 Big data Technological solutions

Respondents were given a choice of more than one technology to choose from that their organization uses to collect information and Open Source Intelligence and Human Intelligence were the most common methods of obtaining data. Figure 6 gives a summary of the responses:

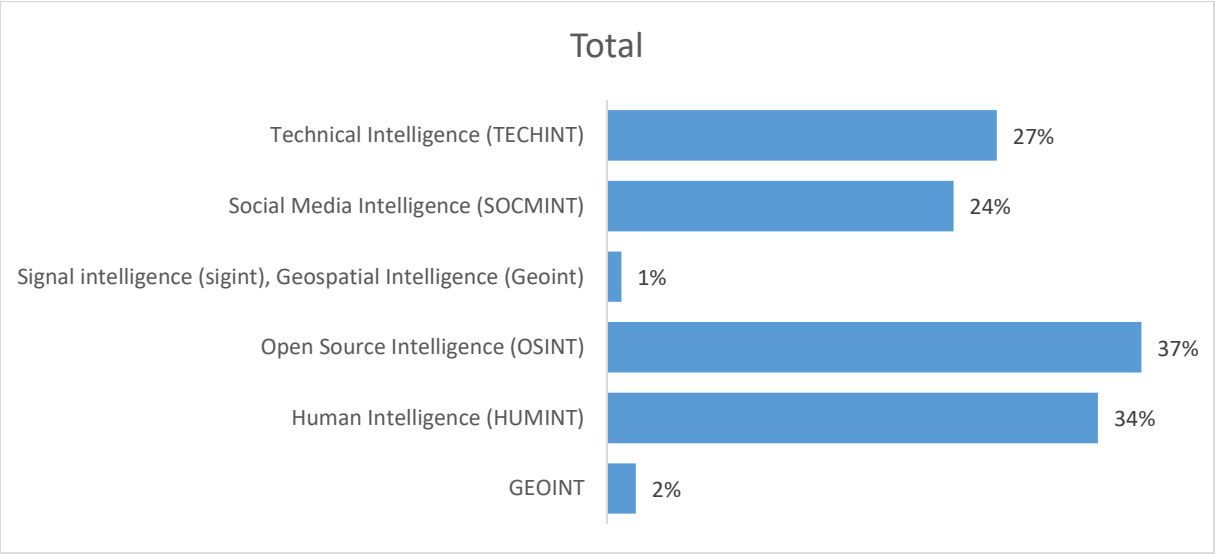


Figure 5: Summary of the big data tools used for information

Most respondents seemed to indicate that the tools mostly used were SOCMINT and TECHINT and this is what most developed nations use in their analysis as was noted in Mutahi and Kimari when looking at how the government tried to find people who were spreading hate speech using social media during the post-election violence.[152] Kolajo and Daramola also reviewed how many countries are increasingly adopting big data tools to combat threats such as terrorism and cybercrimes and they even went ahead to suggest the use of another advanced technology called Social Media Analysis for Combating terrorism (SMACT) which uses data from multiple social media platforms to detect terrorism activities using the Apache Spark technology.[153] This technology is however too advanced for most developed nations but still goes to underscore the importance of the use of these big data tools.

[152] Mutahi and Kimari, "The Impact of Social Media and Digital Technology on Electoral Violence in Kenya."
[153] Kolajo and Daramola, "Leveraging Big Data to Combat Terrorism in Developing Countries."

4.4.7 Challenges faced in the implementation of big data and analytics

As part of the questionnaire, the respondents were asked to answer open ended questions as a substitute to informant interviews to give any challenges they felt impeded the application of big data and analytics in their organization. The biggest challenge faced by most of the respondents is the lack of enough technological knowledge to adapt to the changing landscape. Table 5.3 gives a summary of their responses.

Challenges	Percentage
Data privacy and ethical considerations of the sources of information	12.23%
Difficulty having access to at risk individuals or community	5.85%
Human being remains indispensable in the analysis.	0.53%
Insufficient cooperation with other organizations or agencies	10.11%
Insufficient funding and resources	13.83%
Insufficient staff	8.51%
Lack of enough technological knowledge to adapt to the changing landscape	19.68%
Lack of enough time to implement programs and ideas in an effective way	8.51%
Management have not given it priority	0.53%
Overload of responsibilities	9.57%
Reluctance of decision makers to accept the results of automated analysis	10.64%

Table 5.3: Summary of challenges faced in the implementation of big data and analytics

Based on the open-ended questions, a number of issues came to light regarding the challenges faced in the implementation of big data analytics in countering national threats. One issue that constantly came up from the responses was the lack of the technical knowhow of the ever-changing landscape which involved knowledge of technology. Most respondents felt that their institution had not taken the right steps towards positioning the institution and its employees to understand how to adopt and make full use of the advancements in the field. In addition to this, most respondents felt that the amount of resources was not enough to enable them to make this transition. This makes the implementation of some of the necessary technologies or training to take a lot of time or even be non-existent at certain times making it very difficult for the institution to conduct its duties.

Luvembe and Mutahi in their study made note of similar issues when they examined how big data can be adopted into the county governments in Kenya in an aim to strengthen citizen engagement and participation in the security process of Kenya. They argue that a majority of the current county digital systems exist in isolation from one another and that the sharing of data becomes very difficult.[154] The benefits of big data analytics, according to them, will only be realised once the different counties merge their systems and amass the advantages of the collective data that will be available from the integration of the systems to provide solutions in crime prevention, improve devolution service, assist in natural disaster management among other benefits. This lack of coordination among different agency mentioned in their study was also reflected by the responses from the study alongside data privacy issues which were also explored in the study by Mutahi and Kimari.[155]

[154] Luvembe and Mutai, "Big Data Framework for Kenya's County Governments."
[155] Mutahi and Kimari, "The Impact of Social Media and Digital Technology on Electoral Violence in Kenya."

4.5 Descriptive statistics

This section provides the descriptive statistics of the responses on each process of the intelligence cycle which was measured using the Likert scale. The scale used is described in Table 5.4 below:

Scale
1= Not at all
2= To a small extent
3= To a moderate extent
4= To a large extent
5= To a very large extent

Table 5.4: Likert Scale

4.5.1 Planning and targeting

The Table 5.5 below gives a summary of the average score given for every question under planning and targeting. Most of the employees were aware of their role in the organization and agreed that their organizations gave attention equally to all national security threats. However, the respondents did not feel like they had the funding to execute their organization's policies and additionally, they did not feel that they had received adequate training by their organization on the role technology can play in the intelligence cycle. This was a sentiment that was captured in the study by Luvembe and Mutai[156] and the data seems to capture what they raised in their study regarding the amount of

[156] Luvembe and Mutai, "Big Data Framework for Kenya's County Governments."

funding committed by the government to help in the planning process to tackle national threats. The overall mean of the questions under planning and targeting had a mean of 3.43

	Planning and Targeting	**Mean**	**Standard Deviation**
1	To what extent do you feel you have adequate funding to execute your organization's policies?	2.98	1.022
2	How well aware are your of your role in the organization?	4.13	0.859
3	Do you feel you have received adequate training from your organization on the role technology can play in the intelligence cycle?	2.85	1.032
4	To what extent do you agree that your organization's policies effectively acknowledge the role that big data can play in helping to prepare against threats to national security	3.28	1.004
5	Do you feel your organization gives attention equally on all national security threats	3.91	0.865
Mean		**3.43**	**0.96**

Table 5.5: Descriptive statistics on the responses on Planning and Targeting

4.5.2 Collection

Most of the respondents felt like their organization did not have enough staff to collect intelligence nor the right speed to act upon such information. Additionally, the responses indicated that there could be room for improvement in terms of the collection tools required to access intelligence and

most felt that their organizations had good systems to protect the privacy and ensure safety of individuals offering intelligence. Most of the responses under this area had lower scores from the respondent compared to planning and targeting with the overall mean coming to 3.05. This is a very key aspect in the intelligence cycle and the study by George and Bruce[157] goes to show how important it is for a government through their study which goes to underscore the importance of how the data should be collected and collated to make it easier for different agencies to evaluate it and draw insights from it much quicker.

	Collection	**Mean**	**Standard Deviation**
1	Does your organization have adequate digital capabilities to collect intelligence and conduct surveillance and reconnaissance	3.17	0.973
2	Are there enough collection tools to be able to adequately access the required intelligence for instance in cybercrimes or terrorism among other national threats	2.89	0.875
3	Do you feel your organization has enough staff to collect intelligence	2.78	0.951
4	How would you describe your speed of execution on acting upon information	2.80	0.934
5	Would you say your organization has systems to protect privacy, ensure safety and surety of individuals offering intelligence	3.59	1.066
Mean		**3.05**	**0.96**

[157] George and Bruce, *Analyzing Intelligence*.

Table 5.6: Descriptive statistics on the responses on Collection

4.5.3 Processing and Evaluation

The respondents felt that big data is important in the processing and evaluation of information with a mean response of 4.28. They also felt that incorporating big data and analytics speeds up the time taken to process intelligence but their responses showed that most of them felt that their organizations did not use big data analytics enough to process information. The responses under this category had higher scores underscoring the importance of big data in the processing and evaluation of intelligence and Table 5.7 below provides further statistics. Leung[158] notes how the use of various big data and machine learning technologies can be used to process vast amounts of data and lead to expedited results as was done when big data analytical programs such as natural language processing (NLP) were able to implement structure to the unstructured 1.7 million data sets from the Edward Snowden leak.

	Processing and Evaluation	**Mean**	**Standard Deviation**
1	To what extent do you feel big data important in the processing and evaluation of information	4.28	0.861
2	To what extent does your organization use any big data analytics tools in the processing of collected information	3.26	0.905
3	To what extent do you feel incorporation of big data and analytics speed up the time it takes to process collected intelligence	3.91	0.812

[158] Leung, "Customs Looks to AI to Combat More Cases of Smuggling of Drugs and Cigarettes in Hong Kong, Using It to Sift through Thousands of Online Posts."

4	Do you feel incorporating big data and analytics in your organization will provide information superiority and be better than the current systems in place	3.85	0.942
5	To what extent do big data programs bring structure to unstructured data	3.57	0.860
Mean		**3.77**	**0.88**

Table 5.7: Descriptive statistics on the responses on Processing and Evaluation

4.5.4 Analysis

The respondents felt big data was useful in analysis to discern long term development, generate intelligence hypothesis, identify patterns between various threats and adduce refuting facts. However their responses showed that their organizations use big data and analytics in analysis to a moderate extent. They also felt, to a moderate extent, that the human element in the intelligence analysis can be replaced by big data analytics. George and Bruce give a good example in their study on how these tools can be used in analysis and show how this allows the intelligence analyst to map out potential trends that adds values to current opinions and facts within the intelligence community.[159] For example, the authors mention the anomaly detection function of big data analytics programs that enable one to anticipate threats. Such anticipatory output was evident during the 2011 Egyptian revolution where the CIA was able to forecast social instability and unrest to within a measure of three to five days. Table 5.8 below provides the mean values of the responses.

[159] George and Bruce, *Analyzing Intelligence*.

	Analysis	Mean	Standard Deviation
1	To what extent does your organization use big data and analytics in its analysis of intelligence	3.20	0.851
2	How important do you feel big data would be in the speed of analysis of collected information	3.89	0.859
3	To what extent do you feel big data may help in identifying patterns between various threats	3.98	0.917
4	How useful is big data in helping analysis to discern long term development, generate intelligence hypothesis and adduce refuting facts	4.09	0.793
5	To what extent can the human element in the intelligence analysis be replaced by big data analytics	3.13	0.919
Mean		**3.66**	**0.87**

Table 5.8: Descriptive statistics on the responses on Analysis

4.5.5 Dissemination

The Table 5.9 below gives a summary of the scores given to the various questions under dissemination. The responses in this section had the most deviation from the mean compared to the other responses from the other sections and additionally, it had the lowest scoring average from the responses. Most of the respondents expressed that they utilize automatic alerts and reports to a moderate extent and also that there wasn't enough cooperation between law enforcement and other

relevant agencies in regard to national security. However, one of the lower scores was a positive indication as incidences of disclosures of sensitive information occurred to a very small extent which is in agreement with the responses given in regard to protecting the privacy of individuals providing intelligence. Symon and Tarapore in their study point out that the intelligence community utilizes big data analytics tools to disseminate analysis, in the form of recommendation, to different intelligence operatives depending on one's designation. They give an example of how visualization tools such as Palantir may be used to display a map of the evolving trends of conflict zones in terms of people displacement and position of armed groups over a period of time.[160]

	Dissemination	**Mean**	**Standard Deviation**
1	To what extent do you utilize automated alerts and reports within the organization	2.93	1.009
2	Are there mechanisms to share intelligence findings to other related agencies and institutions	3.58	1.270
3	To what extent do your share the results of the intelligence process with the wider public	3.42	1.118
4	How frequent have there been incidences of disclosures of sensitive information in your organization	2.31	0.973

[160] Symon and Tarapore, "Defense Intelligence Analysis in the Age of Big Data."

5	To what extent do you agree that there is sufficient transnational cooperation between law enforcement, civil security agencies and other relevant actors to prevent and respond to national security threats	2.89	0.885
Mean		**3.03**	**1.05**

Table 5.9: Descriptive statistics on the responses on Dissemination

4.6 MANOVA Analysis

A MANOVA analysis was used to investigate the effect of big data and analytics on the various areas of the intelligence cycle. This is because we had more than one dependent variable affected by one independent variable which is big data analytics and this makes MANOVA the appropriate statistical tool to investigate this relationship. The independent variable was big data knowledge and the dependent variables were planning and targeting, collection, processing and evaluation, analysis and finally dissemination. From the reliability tests, the Cronbach's alpha value indicated that the questionnaire was reliable to provide data to conduct analysis and hence all the six intelligence cycle variables were used for analysis.

An analysis on the how the knowledge of big data and analytics by the respondents affected each of the dependent variable under the intelligence cycle was conducted. This was done to see if there was any difference in how respondents with different knowledge of big data responded to the various stages of the intelligence cycle. The mean response on all the questions under each of the variable representing the intelligence cycle was used as the response score of that variable when conducting the MANOVA analysis. There were two respondents who did not provide complete

data on all the variables under the intelligence cycle hence they were omitted from analysis and this gave us 44 data points.

The Wilks' lambda test, which is used to identify if there are differences between the group means of a particular combination of dependent variables, was conducted and it was found that there did not exist a statistically significant difference for the different stages of the intelligence cycle at the 5% significance level with the *p*-value of 0.871. A detailed analysis of this can be found in Appendix I. Since the results did not achieve a statistically significant result, there is no need to do further follow up tests.

However, it was decided to group the independent variable into 2 groups where those with no knowledge at all and to a small extent will form one group and the others will form the second group to see if this will provide a statistically significant difference in how this two groups will affect the depend variables.

The Table 5.10 below provides the descriptive statistics for the dependent variables, split by the independent variable, moreover, the table gives a row labeled as Total which gives means and the standard deviation for groups that are only split by the dependent variable to be known.

Big data knowledge		**Mean**	**Std. Deviation**	**N**
Planning and Targeting	1	3.4200	0.62858	10
	2	3.4529	0.66114	34
	Total	3.4455	0.64680	44
Collection	1	2.9400	0.86436	10
	2	3.0765	0.73363	34
	Total	3.0455	0.75682	44

Processing and Evaluation	1	3.9000	0.34319	10
	2	3.7588	0.67650	34
	Total	3.7909	0.61600	44
Analysis	1	3.7400	0.52536	10
	2	3.6279	0.67904	34
	Total	3.6534	0.64334	44
Dissemination	1	3.0800	0.58271	10
	2	2.9824	0.72089	34
	Total	3.0045	0.68674	44

Table 5.10: Descriptive statistics for dependent variables, split by independent variable

In this case when the analysis was done, the Wilks' lambda had a significance value of .000 which meant that p<.0005 and therefore, it can be concluded that there exist a statistically significant difference in the intelligence cycle variables based on the respondents knowledge of big data based on the two groups created. To find out how the dependent variables differ for the independent variable, the Test of Between-Subjects Effects was conducted and the Table 5.11 below gives a summary of the results:

Big data knowledge		**Type III Sum of Squares**	**df**	**Mean Square**	**F**	**Sig.**	**Partial Eta Sqrd.**	**Noncent. Parameter**	**Obs. Power**
	Planning and Targeting	522.339	2	261.170	610.050	0.000	0.967	1220.100	1.000

	Collection	408.235	2	204.117	350.127	0.000	0.943	700.255	1.000
	Processing and Evaluation	632.478	2	316.239	821.788	0.000	0.975	1643.576	1.000
	Analysis	587.383	2	293.691	696.896	0.000	0.971	1393.793	1.000
	Dissemination	397.275	2	198.637	412.898	0.000	0.952	825.795	1.000

a. R Squared = .967 (Adjusted R Squared = .965)

b. R Squared = .943 (Adjusted R Squared = .941)

c. R Squared = .975 (Adjusted R Squared = .974)

d. R Squared = .971 (Adjusted R Squared = .969)

e. R Squared = .952 (Adjusted R Squared = .949)

f. Computed using alpha = .05

Table 5.11: Summary statistics of the Test of Between-Subjects Effects

It can be observed from Table 5.11 that knowledge of big data based on the two groups we created has a statistically significant effect on Planning and Targeting (F (2,42)=610.05:p<.0005; partial $\eta^2 = .967$), Collection (F (2,42)=350.127:p<.0005; partial $\eta^2 = .943$, Processing and Evaluation (F (2,42)=6821.788:p<.0005; partial $\eta^2 = .975$, Analysis (F (2,42)=696.896:p<.0005; partial $\eta^2 = .971$ and Dissemination (F (2,42)=412.898:p<.0005; partial $\eta^2 = .952$. An alpha correction was important to take into account the fact that we have run multiple ANOVAs. A Benferroni

correction was conducted, which is done by dividing the original alpha level by the number of tests being performed, and a statistical significance of $p<.0125$ is what will be deemed acceptable and all our dependent variables meet this level.

Additionally, a MANOVA analysis was conducted to examine whether the big data technologies used by organizations has any effect on the intelligence cycle and the detailed result can be found in Appendix II. From the Wilks' Lambda test result, it was found that there was no significant effect, at the 5% significance level, on the intelligence cycle for national security organizations that used any of the big data technologies namely; Technical Intelligence, Social Media Intelligence, Signal Intelligence, Open Source Intelligence, Human Intelligence and Geospatial Intelligence. Furthermore, an analysis on the effect of the training provided by national security organizations did not have a significant effect, at the 5% significance level, on the intelligence cycle as measured by the Wilks' Lambda test. As the results were not significant, there was no need to carry a test of between subjects' effect and the detailed results of this can be found in Appendix III. This result was different from what Gartzke and Lindsay found in their study where they found training of employees was significant at the 5% level in improving the knowledge of big data and how it can be used to tackle cybercrime.[161] Furthermore, Nyabira et al. looked into the effect of police training programme on the counter terrorism capability of Kenya. The study interviewed a sample of 85 respondents and in their results they found that the programme delivery method had a positive impact in the counter terrorism capabilities of the police officers.[162]

[161] Gartzke and Lindsay, "Weaving Tangled Webs."
[162] Nyabira and Ayele, "The State of Political Inclusion of Ethnic Communities under Kenya's Devolved System."

4.6 Chapter Summary

This chapter examined the demographic characteristics of the respondents and it was found that most of the respondents had a Master's degree and knowledge on big data analytics and a majority of them had knowledge on big data and analytics. A MANOVA analysis was conducted to establish whether the knowledge of big data and analytics had an effect on the intelligence cycle and it was observed that there was a significant effect of the knowledge of big data on the intelligence cycle. Additionally, it was observed that the big data technologies being used by organizations did not have a significant effect on the intelligence cycle.

CHAPTER FIVE:

DISCUSSION, CONCLUSION, AND RECOMMENDATIONS

5.1 Introduction

This chapter discusses the implication of the results obtained from the previous chapter and draws conclusions from them in relation to the objectives of the study.

5.2 Discussion of the Findings

On the relationship between big data and analytics and the intelligence cycle, the study found that the knowledge of big data and analytics did have a significant effect on the intelligence cycle. This had a significant effect on all the six stages of the intelligence cycle at $p<.0125$ and meant that the knowledge of big data did have an impact at each stage of the intelligence cycle. From the responses, 93.48% of the respondents had some knowledge of big data and a majority of them had worked in their organizations for more than 2 years which meant that they understood their roles in the company and how they could contribute to better their organization. This was reflected in their responses when asked how well they understood their role in the organization with the mean response coming to 4.13 on the Likert scale. Despite their knowledge of big data and analytics, the respondents felt that their organizations could do a better job of formulating policies to effectively acknowledge the role big data could play in the organization. Nonetheless, they felt that big data played an important role in the processing of information and the whole intelligence cycle as a whole and the analysis provided the statistical results to ascertain this. These results therefore address the first objective of the study which aimed to examine the role that knowledge of big data analytics had in the intelligence cycle and the findings show that there exist a statistically significant effect of big data knowledge on the intelligence cycle. Van Puyvelde et al., opined to this fact in their study where they looked at how big data knowledge affected the intelligence cycle

and how big data can be used to bring structure to data that is unstructured from various sources such as websites and this can then be used to draw conclusions for various agencies.[163]

The study also found that the big data technologies being used by national security organizations at the moment do not have a statistically significant effect on the intelligence cycle, in line with addressing the second objective of the study. The technologies that were highlighted by respondents as being used in their organization included Technical Intelligence, Social Media Intelligence, Signal Intelligence, Open Source Intelligence, Human Intelligence and Geospatial Intelligence. Their responses on other questions in the questionnaire did however, indicate that the respondents generally felt that their organizations did not have the necessary policies to ensure the effective use of these tools. One of the responses to a question examining the extent to which their organizations used big data in its analysis of intelligence had a mean response of 3.2 which indicated that the organizations used these tools to a moderate extent in their intelligence cycle process.

Based on the open-ended questions that replaced the key informant interviews, a number of issues came to light regarding the challenges faced in the implementation of big data analytics in countering national threats. One issue that constantly came up from the responses was the lack of the technical knowhow of the ever-changing landscape which involved knowledge of technology. Most respondents felt that their institution had not taken the right steps towards positioning the institution and its employees to understand how to adopt and make full use of the advancements in the field. In addition to this, most respondents felt that the amount of resources was not enough to enable them to make this transition. This makes the implementation of some of the necessary technologies or training to take a lot of time or even be non-existent at certain times making it very

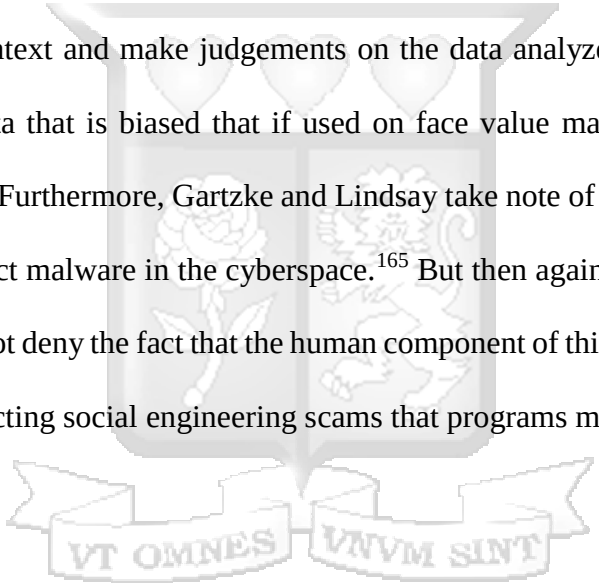
[163] Van Puyvelde, Coulthart, and Hossain, "Beyond the Buzzword."

difficult for the institution to conduct its duties. These issues did show up in the analysis of the respondents questions on the questionnaire when conducting analysis on the questionnaire results where respondents responses did show how these various issues affected the intelligence cycle, For instance, it was noted that the lack of adequate funding did affect the planning and targeting phase of the intelligence cycle and this ties down to the issues of necessary funding to necessary to build up the big data knowledge that will help make this part of the intelligence cycle more effective. These findings also in line with the theories on national security such as the functionalist approach whose principle is based on synergy where organizations work in tandem to complete important functions that the state is required to undertake responsibility for.

Training of the employees to understand the role big data can play to an organization and in the use of various big data technology could play an effect in how effective big data can be in the intelligence cycle and this was the third objective the study sought to examine. Information was collected on the current training programs conducted by the organizations and whether the program were effective in helping improve the intelligence cycle. The findings did not find a statistically significant effect of the current training program undertaken by the national security organizations in helping improve how big data can impact the intelligence cycle. The responses from the respondents also showed that they felt that they did not receive adequate training from their organizations on the role big data technologies can play in the intelligence cycle with the mean response to this question coming to 2.85 which indicated that they felt the training was adequate to a low extent.

As part of the questionnaire, the respondents were asked to list any challenges they felt impeded the application of big data and analytics in their organization and the majority of the respondents cited the lack of knowledge to adapt to the changing technological landscape as the major

challenge. In addition to this, they also felt that the lack of adequate funding, reluctance of the decision makers to accept the results of automated analysis since they do not understand it well, data privacy and ethical considerations of the sources of information and the lack of adequate cooperation among agencies to also be major challenges faced. They further opined that even though big data is important to the intelligence cycle, it could never completely replace the human aspect of the job. George and Bruce in their study also alluded to this where they noted that despite the immense role played by big data analytics in analyzing intelligence information, its ability cannot replace the human analysts who must be retained as an essential component as the human analyst is able to add context and make judgements on the data analyzed.[164] Absent of context, software may present data that is biased that if used on face value may result in inappropriate intelligence action taken. Furthermore, Gartzke and Lindsay take note of the NLP capabilities that enable the analyst to detect malware in the cyberspace.[165] But then again, even such noble ability of big data analytics cannot deny the fact that the human component of this stage of the intelligence cycles is essential in detecting social engineering scams that programs may be unable to detect.



[164] George and Bruce, *Analyzing Intelligence*.
[165] Gartzke and Lindsay, "Weaving Tangled Webs."

5.3 Conclusion of the study

The study established that big data and analytics did have a statistically significant effect on the intelligence cycle. What this means is that national security organizations should focus on ensuring that that their employees receive adequate training on the role that technology and specifically, big data analytics, can play on making the intelligence cycle effective. The new era of digitization has resulted in a vast amount of data that, if used correctly, can have a lot of actionable insights and big data and analytics plays a key role in unlocking this potential. The study findings agree with Tankard where his study explains that, of the many advantages of big data analytics, the most compelling is operational efficiency. This includes the timely detection of cyber or terrorism attacks by harnessing the power of big data to provide insights. He goes on to explain that big data analytics can be useful to governments for the detection of threats from foreign countries, terrorists, hacktivists and criminal elements in the real world and in cyberspace.[166]

The study showed that emerging national threats in Kenya such as cybercrime, terrorism, money laundering, drug and human trafficking and financial fraud require more effective preventative policies and big data plays a key role in this. The findings of the study show that current technologies being used and the training conducted do not have a significant effect on the intelligence cycle but this could be due to the fact that there does not exist good policies and technological infrastructure to fully utilize the power of these tools. The responses from the questionnaire does in fact point to this and majority of respondents indicated that their organization did not fully utilize the power of big data nor did they feel adequately trained to effectively utilize the role of big data and the findings from the analysis reflect this. Although big data and analytics

[166] Tankard, "Big Data Security."

offers numerous benefits, it is not without its challenges and key among them is the ethical and privacy concerns. Legislative frameworks have to be implemented to ensure that this issue is combated and Kenya has the e-Transaction bill for instance, to combat and control cybercrimes which is a step in the right direction.

5.4 Limitations of the study

The findings of the study should be interpreted and comprehended within the confines of some limitations. Firstly, the study was limited in terms of the data that could be collected from the respondents. The research was conducted during a period where the Covid-19 pandemic affected the data collection process from the respondents as some of the respondents could not be available to fill in the questionnaire and thus, the data obtained for analysis could have been much better than it was but the study had to work with the available data. This further posed a challenge to the researcher as the nature of the data being collected was of a confidential nature and physical presence by the researcher was necessary to assuage their fears. However, some of the respondents had to provide their responses online but did not answer some of the questions due to fear and despite the researcher's best efforts to assure the respondents, it was a challenge doing this over email.

Secondly, the intelligence world is very secretive. Staff within a certain department in an organization may not even be in a position to know what the other departments do and the tools they use. Therefore, this may have influenced the adequacy of the data collected and consequently the validity of the conclusions that could be derived from the data. The respondents were however requested to answer the questions to the extent of their interactions with other departments, however limited that may have been.

5.5 Recommendations

The study recommends that the government enact more legislation to help develop and efficient and effective policy infrastructure for the various stakeholders in different agencies involved in the intelligence cycle. Developed nations like the United States have made strides in using big data in their security agencies such as Homeland Security and the CIA to improve their national security. Vuuren et al. opines that the government does have a key role in providing, regulating and maintaining national security which is a right deserved by the citizens of a country. They acknowledge that despite the South African government approving the draft of the National Cyber Security Policy framework that was done in March 2012, the nation's still needs a good structure to efficiently control its cyber infrastructure. They suggest that structures need to be in place to set the security controls and policies and also to govern their implementation arguing that partnerships between business, government and civil society needs to be put in place to achieve this goal.[167]

The study findings have also shown that knowledge of big data does improve the intelligence cycle and as such, this clearly shows there exists an opportunity for Kenya to adopt some of the policies and frameworks to build a competitive advantage for the country to fight several transnational threats. Chatfield et al. examined how the Cape Verde government started setting up structures across the country that made this transition to big data analytics seamless and they look at the National Identification System (NIS)[168], similar to the Huduma number initiative that the Kenyan government is currently implementing and the eCitizen platform that is currently operational, and how it was used by the government as a tool to bring ICT resources closer to the people. They opined that such an infrastructure then makes the transition to the digital age of use for national

[167] van Vuuren et al., "An Approach to Governance of Cybersecurity in South Africa."
[168] Akemi Takeoka Chatfield and Christopher G. Reddick, "Collaborative Network Governance Framework for Aligning Open Justice and E-Justice Ecosystems for Greater Public Value," *Social Science Computer Review* 38, no. 3 (2020): 252–273.

security easier as the government has already established a good ICT infrastructure base within its agencies and the transition becomes quite easy. The Kenyan government can therefore borrow from such an example into how this can be implemented at a county level to enable the intelligence cycle to be more effective.

Additionally, the government should provide more funding to enable the implementation of the various big data technologies that will aid in making the intelligence cycle more effective. Part of the concerns raised by some of the respondents was that they felt their departments did not have enough funding to conduct adequate training and provide the necessary tools that will enable them to expense their duties in the most efficient way in light of big data analytics. A full integration of the various systems within and across departments had not been well implemented and lack of adequate funding was cited as the main reason. Furthermore, those who lacked or had limited knowledge on big data suggested they would appreciate further training sessions to fully understand how to integrate big data into their duties.

The big data and analytics field is a growing field in Kenya and Africa in general and further research would serve to shed more light into further possibilities for the use of this technology. This can be achieved by partnerships between various academic institutions, governmental and non-governmental agencies to look into the possible approaches the government can delve into to combat the transnational threats that the nation currently faces. This clearly shows there exists an opportunity for Kenya to adopt some of the policies and frameworks and build a competitive advantage for the country to fight a number of transnational threats. However there has been challenges that have been highlighted in the implementation of big data in national security such as the lack of explicitly defined or instituted framework for assessing the big data phenomenon. Big data analytics does come with privacy issues as it entails collection of information that at times

can encroach on the privacy of the concerned individual, institution or nation. Furthermore, there exist challenges in the implementation of big data into policy through the government budget process, data compartmentalization between the different government agencies in implementation of a policy if effected among other and government privacy obligations which may cause the process to take longer to be effectively implemented.

APPENDIX

Appendix I

Multivariate Tests[a]

Effect		Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[d]
Intercept	Pillai's Trace	0.958	159.267[b]	5.000	35.000	0.000	0.958	796.337	1.000
	Wilks' Lambda	0.042	159.267[b]	5.000	35.000	0.000	0.958	796.337	1.000
	Hotelling's Trace	22.752	159.267[b]	5.000	35.000	0.000	0.958	796.337	1.000
	Roy's Largest Root	22.752	159.267[b]	5.000	35.000	0.000	0.958	796.337	1.000
Bigdataandanalyticsknowledge	Pillai's Trace	0.405	0.687	25.000	195.000	0.866	0.081	17.187	0.583
	Wilks' Lambda	0.638	0.677	25.000	131.521	0.871	0.086	12.378	0.394
	Hotelling's Trace	0.503	0.672	25.000	167.000	0.878	0.091	16.809	0.562
	Roy's Largest Root	0.341	2.663[c]	5.000	39.000	0.036	0.255	13.315	0.751

a. Design: Intercept + Bigdataandanalyticsknowledge

b. Exact statistic

c. The statistic is an upper bound on F that yields a lower bound on the significance level.

d. Computed using alpha = .05

Appendix II

Multivariate Tests[a]

Effect		Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[d]
Intercept	Pillai's Trace	0.977	1082.117[b]	5.000	127.000	0.000	0.977	5410.584	1.000
	Wilks' Lambda	0.023	1082.117[b]	5.000	127.000	0.000	0.977	5410.584	1.000
	Hotelling's Trace	42.603	1082.117[b]	5.000	127.000	0.000	0.977	5410.584	1.000
	Roy's Largest Root	42.603	1082.117[b]	5.000	127.000	0.000	0.977	5410.584	1.000
Technologicalsolutionsusedbycompany1	Pillai's Trace	0.107	0.716	20.000	520.000	0.812	0.027	14.312	0.568
	Wilks' Lambda	0.897	0.707	20.000	422.161	0.820	0.027	11.687	0.457
	Hotelling's Trace	0.111	0.699	20.000	502.000	0.829	0.027	13.978	0.554
	Roy's Largest Root	0.049	1.269[c]	5.000	130.000	0.281	0.047	6.347	0.439

a. Design: Intercept + Technologicalsolutionsusedbycompany1

b. Exact statistic

c. The statistic is an upper bound on F that yields a lower bound on the significance level.

d. Computed using alpha = .05

Appendix III

Multivariate Tests[a]

Effect		Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[d]
Intercept	Pillai's Trace	0.880	181.216[b]	5.000	124.000	0.000	0.880	906.080	1.000
	Wilks' Lambda	0.120	181.216[b]	5.000	124.000	0.000	0.880	906.080	1.000
	Hotelling's Trace	7.307	181.216[b]	5.000	124.000	0.000	0.880	906.080	1.000
	Roy's Largest Root	7.307	181.216[b]	5.000	124.000	0.000	0.880	906.080	1.000
Formaltrainingcategories	Pillai's Trace	0.175	0.663	35.000	640.000	0.933	0.035	23.216	0.710
	Wilks' Lambda	0.835	0.655	35.000	524.051	0.938	0.035	19.212	0.591
	Hotelling's Trace	0.185	0.648	35.000	612.000	0.943	0.036	22.674	0.695
	Roy's Largest Root	0.080	1.454[c]	7.000	128.000	0.189	0.074	10.181	0.595

Appendix IV: Questionnaire

Introduction

The general objective of this study is to investigate the role big data and analytics can play in improving the National security in Kenya.

The specific objectives of this study will be:

4. To examine the big data and analytics approaches the government can undertake to improve the intelligence cycle.
5. To examine the unique value proposition big data and analytics presents to address issues on national security and the intelligence cycle.
6. To look into the challenges faced by the government in implementing big data and analytics in the national security environment and their possible solutions.

SECTION A

Kindly respond to the questions as honestly as possible.

Tick inside the box to indicate your choice of answer

1. Please specify your gender

   Male [ ]

   Female [ ]

   Other [ ]

2. How long have your worked at your organization

   Less than 2 years [ ]

   2-5 years [ ]

   5-10 years [ ]

   More than 10 years [ ]

3. What is your highest level of education

   Diploma/Certificate [ ]

   Bachelor's Degree [ ]

   Master's Degree [ ]

   PhD [ ]

4. In relation to your profession, which of the following broad areas of crime are you confronted with?

   Cybercrime [ ]

   Terrorism [ ]

   Human trafficking/smuggling [ ]

   Drug trafficking and distribution [ ]

   Money laundering [ ]

   Other (specify)_____.

5. In your opinion, what emerging national threats need more effective preventative or response policies or strategies? Please tick as many from the following list that apply.

   Human trafficking/smuggling [ ]

   Drug trafficking and distribution [ ]

   Money laundering [ ]

   Cybercrime [ ]

   Terrorism [ ]

   Other (specify)_____.

6. How conversant are you with big data and analytics

   Not at all [ ]

   To a small extent [ ]

   To a moderate extent [ ]

   To a large extent [ ]

   To a very large extent [ ]

7. Have you received any formal training/education in the following

   Information retrieval/online searching [ ]

   Information seeking [ ]

   Other relevant training (specify)_____.

8. How you gather information or intelligence

   Human Intelligence (HUMINT) [ ]

   Open Source Intelligence (OSINT) [ ]

   Technical Intelligence (TECHINT) [ ]

   Social Media Intelligence (SOCMINT) [ ]

   Other (specify)_____.

9. Is your organization working with any one of the following technological solutions?

   Social media surveillance [ ]

   Information database creation [ ]

   Cyber-intelligence hacking software [ ]

   Big data software solutions [ ]

   Other (specify)_____.

10. Which of the following big data tools and techniques do you use in your organization?

    Hadoop [ ]

    Apache Spark [ ]

    Apache Storm [ ]

    Cassandra [ ]

    R Programming Tool [ ]

    Other (specify)_____.

SECTION B

This section focuses on the application of big data and analytics to the intelligence cycle. Please indicate the extent to which the following statements describe your organization.

Use the scale:

1= Not at all

2= To a small extent

3= To a moderate extent

4= To a large extent

5= To a very large extent

(Tick the appropriate scale)

	Planning and Targeting	1	2	3	4	5
1	To what extent do you feel you have adequate funding to execute your organization's policies?					
2	How well aware are your of your role in the organization?					
3	Do you feel you have received adequate training from your organization on the role technology can play in the intelligence cycle?					
4	To what extent do you agree that your organization's policies effectively acknowledge the role that big data can play in helping to prepare against threats to national security					
5	Do you feel your organization gives attention equally on all national security threats					
	Collection					
1	Does your organization have adequate digital capabilities to collect intelligence and conduct surveillance and reconnaissance					
2	Are there enough collection tools to be able to adequately access the required intelligence for instance in cybercrimes or terrorism among other national threats					
3	Do you feel your organization has enough staff to collect intelligence					
4	How would you describe your speed of execution on acting upon information					
5	Would you say your organization has systems to protect privacy, ensure safety and surety of individuals offering intelligence					
	Processing and Evaluation					
1	To what extent do you feel big data is important in the processing and evaluation of information					

2	To what extent does your organization use any big data analytics tools in the processing of collected information					
3	To what extent do you feel incorporation of big data and analytics speed up the time it takes to process collected intelligence					
4	Do you feel incorporating big data and analytics in your organization will provide information superiority and be better than the current systems in place					
	Analysis					
1	To what extent does your organization use big data and analytics in its analysis of intelligence					
2	How important do you feel big data would be in the speed of analysis of collected information					
3	To what extent do you feel big data may help in identifying patterns between various threats					
4	Do you feel big data and analytics will make the explain ability of complicated issues much easier within the organization					
	Dissemination					
1	To what extent do you utilize automated alerts and reports within the organization					
2	Are there mechanisms to share intelligence findings to other related agencies and institutions					
3	To what extent do your share the results of the intelligence process with the wider public					
4	How frequent have there been incidences of disclosures of sensitive information in your organization					
5	To what extent do you agree that there is sufficient transnational cooperation between law enforcement, civil security agencies and other relevant actors to prevent and respond to national security threats					

SECTION C

1. To what extent does your organization use the following mechanisms to counter threats to national security?

	Very great extent	Great extent	Neutral	Low extent	Not at all
Counter propaganda					
Diplomacy					
Open source intelligence					
News					
Social Media					
Any other Specify:					

2. What are the main challenges, in your opinion, that impede the application of big data and analytics in your organization? Please tick any that applies

   Insufficient staff [ ]

   Insufficient cooperation with other organizations or agencies [ ]

   Overload of responsibilities [ ]

   Insufficient funding and resources [ ]

   Lack of enough time to implement programs and ideas in an effective way [ ]

   Difficulty having access to at risk individuals or community [ ]

   Lack of enough technological knowledge to adapt to the changing landscape [ ]

   Other_____.

3. Overall, has using big data analytics made your operations more efficient?
   Yes [ ]
   No [ ]

4. Please indicate here any issues you would like to mention not covered by the questionnaire

5. Please indicate any concerns or recommendations you would like to mention regarding the questionnaire itself

BIBLIOGRAPHY

Ahn, Sung-Hwan, Nam-Uk Kim, and Tai-Myoung Chung. "Big Data Analysis System Concept for Detecting Unknown Attacks." In *16th International Conference on Advanced Communication Technology*, 269–272. IEEE, 2014.

Ajana, Btihaj. "Augmented Borders: Big Data and the Ethics of Immigration Control." *Journal of Information, Communication and Ethics in Society* 13, no. 1 (2015): 58–78.

Akhgar, Babak, Gregory B. Saathoff, Hamid R. Arabnia, Richard Hill, Andrew Staniforth, and Petra Saskia Bayerl. *Application of Big Data for National Security: A Practitioner's Guide to Emerging Technologies*. Butterworth-Heinemann, 2015.

Alfano, Marco, and Joseph-Simon Görlach. "Terrorism, Education, and the Role of Expectations: Evidence from al-Shabaab Attacks in Kenya," 2019.

Baccala, Luiz A., K. Sameshima, and D. Y. Takahashi. "Generalized Partial Directed Coherence." In *2007 15th International Conference on Digital Signal Processing*, 163–166. IEEE, 2007.

Baratta, Joseph Preston. *The Politics of World Federation: United Nations, UN Reform, Atomic Control*. Vol. 2. Greenwood Publishing Group, 2004.

Botha, Anneli. "Assessing the Vulnerability of Kenyan Youths to Radicalisation and Extremism." *Institute for Security Studies Papers* 2013, no. 245 (2013): 28–28.

Bottles, Kent, Edmon Begoli, and Brian Worley. "Understanding the Pros and Cons of Big Data Analytics." *Physician Executive* 40, no. 4 (2014): 6–12.

Brewer, Ross. "Cyber Threats: Reducing the Time to Detection and Response." *Network Security* 2015, no. 5 (2015): 5–8.

Cárdenas, Alvaro A., Pratyusa K. Manadhata, and Sreeranga P. Rajan. "Big Data Analytics for Security." *IEEE Security & Privacy* 11, no. 6 (2013): 74–76.

Chatfield, Akemi Takeoka, and Christopher G. Reddick. “Collaborative Network Governance Framework for Aligning Open Justice and E-Justice Ecosystems for Greater Public Value.” *Social Science Computer Review* 38, no. 3 (2020): 252–273.

Claude, Inis. “Theoretical Approaches to National Security and World Order.” *Moore & Turner National Security Law. Durham*, 2005, 3–14.

Corp. (HSRC), Homeland Security Research. “Homeland Security Research Corp. (HSRC): Big Data and Data Analytics in National Security Is Forecast to Grow at a 2015-2022 CAGR of 17.5%.” Accessed March 20, 2020. https://www.prnewswire.com/news-releases/homeland-security-research-corp-hsrc-big-data-and-data-analytics-in-national-security-is-forecast-to-grow-at-a-2015-2022-cagr-of-17-5-300869799.html.

Curry, Sam, Engin Kirda, Eddie Schwartz, William H. Stewart, and Amit Yoran. “Big Data Fuels Intelligence-Driven Security.” *RSA Security Brief*, 2013.

Derrida, Jacques. *Writing and Difference*. Routledge, 2001.

Deutsch, Karl W. *Politische Kybernetik: Modelle Und Perspektiven*. Rombach, 1969.

Dolenko, Galyna, and Sviatoslav Lobach. “System and Statistical Approach of Analysis and Forecasting Terrorist Activity.” *Model Assisted Statistics and Applications* 9, no. 3 (2014): 267–275.

Doyle, Michael W. “Ways of War and Peace: Realism.” *Liberalism, and Socialism*, 1997, 24–25.

Ezumah, Bellarmine, and Suraj Olunifesi Adekunle. “A Review of Privacy, Internet Security Threat, and Legislation in Africa: A Case Study of Nigeria, South Africa, Egypt, and Kenya.” In *Internet and Distributed Computing Advancements: Theoretical Frameworks and Practical Applications*, 115–136. IGI Global, 2012.

Fahey, Sean. “Big Data and Analytics for National Security.” *Stanford University. Pristupljeno* 5 (2012): 2017.

Gartzke, Erik, and Jon R. Lindsay. “Weaving Tangled Webs: Offense, Defense, and Deception in Cyberspace.” *Security Studies* 24, no. 2 (2015): 316–348.

George, Roger Z., and James B. Bruce. *Analyzing Intelligence: National Security Practitioners' Perspectives*. Georgetown University Press, 2014.

Gill, Peter, and Mark Phythian. “From Intelligence Cycle to Web of Intelligence: Complexity and the Conceptualisation of Intelligence.” In *Understanding the Intelligence Cycle*, 35–56. Routledge, 2013.

Granger, C. W. J. “Investigating Causal Relations by Econometric Models and Cross-Spectral Methods.” In *Essays in Econometrics: Collected Papers of Clive WJ Granger*, 31–47, 2001.

Hamilton, Alexander, James Madison, and John Jay. *The Federalist: A Commentary on the Constitution of the United States: Being a Collection of Essays Written in Support of the Constitution Agreed Upon September 17, 1787 by the Federal Convention: Reprinted from the Original Text of Alexander Hamilton, John Jay, and James Madison*. GP Putnam’s sons, 1888.

Heo, Moonseong, Namhee Kim, and Myles S. Faith. “Statistical Power as a Function of Cronbach Alpha of Instrument Questionnaire Items.” *BMC Medical Research Methodology* 15, no. 1 (October 14, 2015): 86. https://doi.org/10.1186/s12874-015-0070-6.

Hof, Robert D. “Deep Learning: With Massive Amounts of Computational Power, Machines Can Now Recognize Objects and Translate Speech in Real Time.” *Artificial Intelligence Is Finally Getting Smart. MIT Technology Review* 116, no. 2 (2013): 78–86.

IBM. “Big Data for the Intelligence Community,” 2019, 20.

Jani, Karan. “The Promise and Prejudice of Big Data in Intelligence Community.” *ArXiv Preprint ArXiv:1610.08629*, 2016.

Jean, Neal, Marshall Burke, Michael Xie, W. Matthew Davis, David B. Lobell, and Stefano Ermon. "Combining Satellite Imagery and Machine Learning to Predict Poverty." *Science* 353, no. 6301 (2016): 790–794.

Kant, Immanual. "To Perpetual Peace: A Philosophical Sketch Trs." *Ted Humphrey, Indianapolis: Hacket Pub*, 1795.

Kimutai, JULIUS KIPKORIR. "Social Media and National Security Threats: A Case Study of Kenya." *Unpublished MA Thesis: University of Nairobi*, 2014.

Kitchin, Rob. "Big Data, New Epistemologies and Paradigm Shifts." *Big Data & Society* 1, no. 1 (2014): 2053951714528481.

Kolajo, Taiwo, and Olawande Daramola. "Leveraging Big Data to Combat Terrorism in Developing Countries." In *2017 Conference on Information Communication Technology and Society (ICTAS)*, 1–6. IEEE, 2017.

Kones, Kefa, Kibet. "Transnational Threats to the National Security in Kenya - Google Search." University of Nairobi, 2015.

Leung, Christy. "Customs Looks to AI to Combat More Cases of Smuggling of Drugs and Cigarettes in Hong Kong, Using It to Sift through Thousands of Online Posts," August 7, 2019. https://www.thestar.com.my/authors?q=Christy%20Leung.

Leung, Wai-Ching. "How to Design a Questionnaire." *BMJ* 322, no. Suppl S6 (2001): 0106187.

Levy, Jack S. "Domestic Politics and War." *The Journal of Interdisciplinary History* 18, no. 4 (1988): 653–673.

Luvembe, Alex, and Hillary Mutai. "Big Data Framework for Kenya's County Governments." *Journal of Computer and Communications* 07 (January 1, 2019): 1–9. https://doi.org/10.4236/jcc.2019.71001.

Mahmood, Tariq, and Uzma Afzal. “Security Analytics: Big Data Analytics for Cybersecurity: A Review of Trends, Techniques and Tools.” In *2013 2nd National Conference on Information Assurance (Ncia)*, 129–134. IEEE, 2013.

Mitrany, David. *A Working Peace System: Introd. by Hans J. Morgenthau*. Quadrangle Books, 1966.

Moore, John. *Newer Theories in Understanding War: From the Democratic Peace to Incentive Theory*. Moore & Turner National Security Law. Durham, NC: Carolina Academic Press, 2005.

Mugenda, Olive M., and Abel G. Mugenda. *Research Methods: Quantitative and Qualitative Approaches*. Acts press, 1999.

Mutahi, P, and B Kimari. “The Impact of Social Media and Digital Technology on Electoral Violence in Kenya,” 2017.

Nair, Deepesh. “The Evolution of Analytics with Data.” *Medium*, October 21, 2018. https://towardsdatascience.com/the-evolution-of-analytics-with-data-8b9908deadd7.

Nwanga, MATHEW E., ELIZABETH N. Onwuka, A. M. Albinu, and O. C. Ubadike. “Leveraging Big Data in Enhancing National Security in Nigeria.” *International Journal of Knowledge, Innovation and Entrepreneurship* 2, no. 2 (2014): 66–80.

Nyabira, Ben Christopher, and Zemelak Ayitenew Ayele. “The State of Political Inclusion of Ethnic Communities under Kenya’s Devolved System.” *Law, Democracy & Development* 20, no. 1 (2016): 131–153.

Nye Jr, Joseph S. *The Paradox of American Power: Why the World’s Only Superpower Can’t Go It Alone*. Oxford University Press, 2003.

Ohlhorst, Frank J. *Big Data Analytics: Turning Big Data into Big Money*. Vol. 65. John Wiley & Sons, 2012.

Omand, David, Jamie Bartlett, and Carl Miller. “Introducing Social Media Intelligence (SOCMINT).” *Intelligence and National Security* 27, no. 6 (2012): 801–823.

———. “Introducing Social Media Intelligence (SOCMINT).” *Intelligence and National Security* 27 (December 1, 2012). https://doi.org/10.1080/02684527.2012.716965.

Pinkovskiy, Maxim, and Xavier Sala-i-Martin. “Lights, Camera… Income! Illuminating the National Accounts-Household Surveys Debate.” *The Quarterly Journal of Economics* 131, no. 2 (2016): 579–631.

Rummel, Rudolph J. *Power Kills: Democracy as a Method of Nonviolence*. Routledge, 2017.

Russett, Bruce. *Grasping the Democratic Peace: Principles for a Post-Cold War World*. Princeton university press, 1994.

Saito, Y., and H. Harashima. *Tracking of Information within Multichannel EEG Record-Casual Analysis in EEG. Recent Advances in EEG and EMG Data Processing*. Elsevier/North-Holland, Amsterdam, 1981.

Salik, Hammaad, and Zaheema Iqbal. “Social Media and National Security.” *The Geopolitics,* September 10, 2019. https://thegeopolitics.com/social-media-and-national-security/.

Sameshima, Koichi, and Luiz Antonio Baccalá. “Using Partial Directed Coherence to Describe Neuronal Ensemble Interactions.” *Journal of Neuroscience Methods* 94, no. 1 (1999): 93–103.

Saunders, Mark NK, and Keith Townsend. “Reporting and Justifying the Number of Interview Participants in Organization and Workplace Research.” *British Journal of Management* 27, no. 4 (2016): 836–852.

Schut, Martijn C. “On Model Design for Simulation of Collective Intelligence.” *Information Sciences* 180, no. 1 (2010): 132–155.

Sewell, James Patrick. “Policy Processes and International Organisation Tasks.” In *International Organisation: World Politics*, 98–112. Springer, 1969.

Song, Xiaojun, and Abderrahim Taamouti. “A Better Understanding of Granger Causality Analysis: A Big Data Environment.” *Oxford Bulletin of Economics and Statistics* 81, no. 4 (2019): 911–936.

Statista Research Department. “Global IoT End-User Spending Worldwide 2017-2025.” *Statista*, February 19, 2020. https://www.statista.com/statistics/976313/global-iot-market-size/.

Symon, Paul B., and Arzan Tarapore. “Defense Intelligence Analysis in the Age of Big Data.” *Joint Forces Quarterly—JFQ* 79 (2015): 4–11.

Tang, Yong, Jason Jie Xiong, Yong Luo, and Yi-Cheng Zhang. “How Do the Global Stock Markets Influence One Another? Evidence from Finance Big Data and Granger Causality Directed Network.” *International Journal of Electronic Commerce* 23, no. 1 (2019): 85–109.

Tankard, Colin. “Big Data Security.” *Network Security* 2012, no. 7 (2012): 5–8.

Treverton, Gregory F. *New Tools for Collaboration*. Rowman & Littlefield, 2016.

Van Puyvelde, Damien, Stephen Coulthart, and M. Shahriar Hossain. “Beyond the Buzzword: Big Data and National Security Decision-Making.” *International Affairs* 93, no. 6 (November 1, 2017): 1397–1416. https://doi.org/10.1093/ia/iix184.

Verble, Joseph. “The NSA and Edward Snowden: Surveillance in the 21st Century.” *ACM SIGCAS Computers and Society* 44, no. 3 (2014): 14–20.

Vuuren, Joey Jansen van, Louise Leenen, Jackie Phahlamohlaka, and Jannie Zaaiman. “An Approach to Governance of Cybersecurity in South Africa.” In *Cyber Behavior: Concepts, Methodologies, Tools, and Applications*, 1583–1597. IGI Global, 2014.

Wang, Xiaojun, Leroy White, Xu Chen, Kun Chen, Xin Li, and Huaiqing Wang. “On the Model Design of Integrated Intelligent Big Data Analytics Systems.” *Industrial Management & Data Systems*, 2015.

Wiener, Norbert. “The Theory of Prediction. Modern Mathematics for Engineers.” *New York*, 1956, 165–190.

Wilson, Robert E., Samuel D. Gosling, and Lindsay T. Graham. “A Review of Facebook Research in the Social Sciences.” *Perspectives on Psychological Science* 7, no. 3 (2012): 203–220.

VT OMNES
VNVM SINT