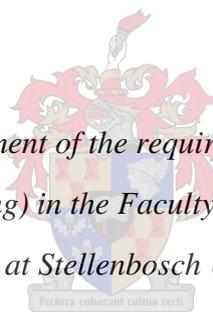


**MITIGATING SIGNIFICANT RISKS PERTAINING TO THE IMPLEMENTATION  
OF COGNITIVE COMPUTING**

by

Jana van Wyk

*Thesis presented in partial fulfilment of the requirements for the degree of Master of  
Commerce (Computer Auditing) in the Faculty of Economic and Management  
Sciences at Stellenbosch University*



Supervisor: Professor Riaan Rudman

March 2017

## Declaration

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

Jana van Wyk

March 2017

Copyright © 2017 Stellenbosch University

All rights reserved

## **ACKNOWLEDGEMENTS**

My sincere gratitude to:

- my Heavenly Father,
- my parents and friends, and
- my supervisor

for your support, patience and guidance.

## ABSTRACT

Cognitive computing is recognised as the third era in the evolution of computing. This era is driven by the exponential growth in data, advances in enabling technologies and enterprise's need to realise significant business value from data resources. The capabilities of cognitive computing creates significant and immediate opportunities for enterprises. The problem is that management are implementing cognitive computing systems without understanding the technology or the risks the enterprise are exposed. The aim of the research is therefore to identify and mitigate significant risks pertaining to the implementation of cognitive computing. The research aims to investigate cognitive computing, identify significant risks and recommend safeguards to mitigate these risks.

A literature review was performed to provide a theoretical foundation for the research and focussed on cognitive computing, corporate governance, IT governance, data protection and the use of control frameworks to achieve effective governance. COBIT 5 was selected as the most appropriate control framework to identify significant risk. In order to identify the risks the core components of a cognitive computing system were identified and classified into specific phases based on their function within the cognitive computing system. The research found that a cognitive computing system consists of consist of twelve core components and four phases. The core components include: unstructured, semi-structured and structured data; data access, metadata, feature extraction, natural language processing and deep learning; corpus and advances analytics; and hypothesis generation and scoring, and machine learning.

Based on the understanding of the core components, COBIT 5 was used to identify significant risks. Significant risks were identified at a strategic and operational or technological level. Risks at a strategic level involved inadequate governance and management, as well as insufficient human skills and resource management. Significant risks at an operational or technological level comprised of cost, privacy, security, scalability, integration, interoperability, veracity, ownership and life cycle risks. The research proceeded to formulate appropriate internal control techniques to mitigate the significant risks identified. The internal control techniques include establishing a cognitive computing strategies and policies, implementing human skills and resource controls, data controls, infrastructure controls, supplier controls and life cycle controls. The final product of the findings is a risk matrix, which maps the relevant core components with the significant risk which they introduce and a risk-control matrix which maps the risk to the control technique which mitigates the risk.

## UITTREKSEL

Kognitiewe verwerking (cognitive computing) word erken as die derde era in die evolusie van rekenaar verwerking (computing). Die era word gedryf deur die eksponensiële groei van data, die verbetering van bemagtigende tegnologieë en ondernemings se behoefte om beduidende besigheidswaarde uit data bronne te realiseer. Kognitiewe verwerking is in staat om beduidende en onmiddellike geleenthede vir ondernemings te skep. Die probleem is egter dat bestuur kognitiewe verwerking stelsels (cognitive computing systems) implementeer sonder dat hulle die tegnologie of die risiko's waaraan die onderneming blootgestel word verstaan. Die doel van die navorsing is dus om wesenlike risiko's in verband met die implementering van kognitiewe verwerking te identifiseer en aan te spreek. Die navorsing beoog om 'n dieper begrip te ontwikkel van kognitiewe verwerking, om wesenlike risiko's te identifiseer en om kontroles aan te beveel wat die risiko's aanspreek.

'n Literatuur studie was uitgevoer om 'n teoretiese basis vir die navorsing te bied en het gefokus op kognitiewe verwerking, korporatiewe beheer, IT beheer, data beskerming en die gebruik van kontrole raamwerke om effektiewe beheer te bewerkstellig. COBIT 5 was geselekteer as die beste kontrole raamwerk om wesenlike risiko's te identifiseer. Om die risiko's te identifiseer is die onderliggende komponente van 'n kognitiewe verwerking stelsel geïdentifiseer en geklassifiseer gebaseer op hul funksies in die kognitiewe verwerking stelsel. Die navorsing het gevind dat 'n kognitiewe verwerking stelsel uit twaalf onderliggende komponente en vier fases bestaan. Die onderliggende komponente sluit in: ongestruktureerde, semigestruktureerde en gestruktureerde data; data toegang, metadata, kenmerk ontrekking (feature extraction), natuurliketaalverwerking (natural language processing) en dieper leer (deep learning); korpus (corpus) en gevorderde ontleding (advanced analytics); en hipotese generering en meting (hypothesis generation and scoring), en masjien leer (machine learning).

COBIT 5 is gebruik om wesenlike risiko's te identifiseer, gebaseer op kennis van die onderliggende komponente. Wesenlike risiko's is op 'n strategiese en operasionele of tegnologiese geïdentifiseer. Risiko's op 'n strategiese vlak sluit in onvoldoende beheer en bestuur, sowel as onvoldoende menslike hulpbron bestuur. Wesenlike risiko's op 'n operasionele of tegnologiese vlak bestaan uit koste, privaatheid, sekuriteit, skaalbaarheid (scalability), integrasie, interoperasionaliteit (interoperability), geldigheid (veracity), data eienaarskap en lewensiklus risiko's. Die navorsing het voortgegaan om interne beheermaatreëls te formuleer om die wesenlike risiko's aan te spreek. Die interne

beheermaatreëls sluit in die vestiging van kognitiewe verwerking strategieë en beleide, die implementering van menslike hulpbron kontroles, data kontroles, infrastruktuur kontroles, verskaffer kontroles en lewensiklus kontroles. Die finale produk van die bevindinge is 'n risiko matriks, wat die relevante onderliggende komponente verbind met die wesenlike risiko's wat hulle skep en 'n risiko-kontrole matriks wat die risiko's verbind met die beermaatreëls wat die risiko's aan spreek.

**TABLE OF CONTENTS**

<b>CHAPTER 1: INTRODUCTION AND RESEARCH OBJECTIVE</b>	<b>1</b>
1.1 Introduction and background	1
1.2 Problem statement and research objective	2
1.3 Scope limitations	2
1.4 Organisational structure of the research	3
<b>CHAPTER 2: RESEARCH DESIGN AND METHODOLOGY</b>	<b>4</b>
<b>CHAPTER 3: LITERATURE REVIEW</b>	<b>7</b>
3.1 Introduction	7
3.2 Historic review of prior research	7
3.3 Cognitive computing	8
3.3.1 Defining cognitive computing	9
3.3.2 Application of cognitive computing	10
3.4 Corporate governance and Information Technology governance	13
3.4.1 Corporate governance	13
3.4.2 Information Technology governance	14
3.4.3 Data protection and cognitive computing	15
3.4.4 IT gap and alignment	17
3.5 Control frameworks	17
3.5.1 Control objectives for information and related technologies (COBIT)	18
3.5.2 Information technology infrastructure library (ITIL)	20
3.5.3 Committee of sponsoring organizations (COSO)	22
3.5.4 Framework selected for the purpose of this research	23
3.6 Summary and conclusion	24
<b>CHAPTER 4: CORE COMPONENTS OF A COGNITIVE COMPUTING SYSTEM</b>	<b>26</b>
4.1 Data ingestion	28
4.2 Read phase	28

4.3	Resolve phase	30
4.4	Reason phase	32
4.5	Infrastructure	35
4.6	Enabling technologies	37
	4.6.1 Hadoop	37
	4.6.2 Cloud technologies	39
	4.6.3 Big data	40
4.7	Summary and conclusion	41
<b>CHAPTER 5: RISKS PERTAINING TO THE IMPLEMENTATION OF A COGNITIVE COMPUTING SYSTEM</b>		<b>42</b>
5.1	Significant risks at a strategic level	42
	5.1.1 Inadequate governance and management of cognitive computing systems	42
	5.1.2 Inadequate human skills and resource management	44
5.2	Significant risks at an operational or technological level	45
	5.2.1 Cost	45
	5.2.2 Privacy	46
	5.2.3 Security	49
	5.2.4 Ownership	51
	5.2.5 Scalability	51
	5.2.6 Integration	53
	5.2.7 Interoperability	54
	5.2.8 Veracity (Quality)	55
	5.2.9 Cognitive computing life cycle risks	56
5.3	Summary and conclusion	58
<b>CHAPTER 6: SAFEGUARDS AND CONTROLS IN A COGNITIVE COMPUTING SYSTEM</b>		<b>60</b>
6.1	Governance and management at a strategic level	60
	6.1.1 Cognitive computing governance	60
	6.1.2 Cognitive computing strategy and policies	61

6.1.3	Human skills and resources controls	63
6.2	Controls at an operational or technological level	64
6.2.1	Data controls	64
6.2.2	Infrastructure controls	68
6.2.3	Service provider controls	70
6.2.4	Cognitive computing life cycle controls	72
6.3	Summary and conclusion	73
<b>CHAPTER 7: CONCLUSION</b>		<b>85</b>
<b>REFERENCES</b>		<b>88</b>

**LIST OF FIGURES, TABLES AND APPENDICES****Figures**

Figure 1: The fundamental capabilities of cognitive computing	9
Figure 2: Core components of a cognitive computing system	27
Figure 3: Hypothesis generation and scoring process	34

**Tables**

Table 1: The benefits and limitations of COBIT, ITIL and COSO	24
Table 2: A risk matrix: linking the cognitive computing components to the significant risks it gives rise to	58
Table 3: A risk-control matrix: linking the significant cognitive computing risks to the relevant mitigating internal controls	74

**Appendices**

Appendix A: Identifying risk by means of COBIT 5's detailed processes	101
---	-----

## **CHAPTER 1: INTRODUCTION AND RESEARCH OBJECTIVE**

### **1.1 Introduction and background**

Cognitive computing is recognised as the third era in the evolution of computing (Hurwitz, Kaufman & Bowles, 2015; Kelly III & Hamm, 2013). An era where computers learn from experience, instead of depending on programmed algorithms to respond to a set of predefined rules and questions, and are capable of formulating new ideas, hypotheses and knowledge based on an understanding of natural language (Willis Towers Watson, 2016). The advent of the cognitive computing era is driven by the exponential growth in data, specifically unstructured data which accounts for 80-90% of the digital universe, the advances in and increasing sophistication of enabling technologies, and the enterprise's need to realise business and economic value from data resources (Willis Towers Watson, 2016; Bataller & Harris, 2015; Hurwitz, Kaufman & Bowles, 2015; Sarkar & Zaharchuk, 2015).

An increasing number of industries are utilising the innovative capabilities of cognitive computing, including health care, customer services, insurance, IT and telecommunications, and financial services. Deloitte estimates that, due to the increase in enterprises and industries using cognitive computing, more than 80 of the world's largest enterprise software companies (by revenues) will have incorporated cognitive technologies into their products by the end of 2016. This is expected to rise to 95 out of 100 by 2020 (Willis Towers Watson, 2016).

The capabilities of cognitive computing creates significant and immediate opportunities for enterprises. Management requires knowledge about cognitive computing and the risks enterprises are exposed to pertaining to the implementation of a cognitive computing system.

### **1.2 Problem statement and research objective**

Pioneering enterprises across industries are implementing cognitive computing systems to utilise cognitive computing capabilities. The problem is that management is implementing the cognitive computing system without understanding the technology or the risks enterprises are exposed to pertaining to the implementation of the cognitive computing system.

The objective of this research is to identify and mitigate significant risks pertaining to the implementation of cognitive computing and a cognitive computing system.

The research proposes to provide stakeholders with insight into (i) the core components of a cognitive computing system; (ii) the significant risks pertaining to the implementation of a cognitive computing system and (iii) to recommend safeguards to mitigate these risks to an acceptable level.

### **1.3 Scope limitations**

The aim of the research is to mitigate and investigate *significant* risks pertaining to the *implementation* of a cognitive computing system and its core components. The purpose is not to identify and mitigate all the risks associated with cognitive computing in general or to identify all the risks associated with the interaction between cognitive computing and its surrounding environment.

Cognitive computing systems can be developed in-house or by cognitive computing developers. In-house development will introduce significant development risks and require appropriate internal control techniques to address these risks. The research will only highlight some of the significant development risks and appropriate control techniques that can be directly linked to cognitive computing systems and does not intend to identify and mitigate all risks relating to system or software development.

The focus of the research is on the four phases of the cognitive computing system and the core components included in each phase, as well as the infrastructure and enabling technologies which support the four phases. There are two components (virtualisation and application) which form part of the cognitive computing system but do not fall within the four phases based on their function. These components will be excluded from the research.

### **1.4 Organisational structure of research**

The structure of the remainder of the research is illustrated in the chapter outline below:

#### **CHAPTER 2: Research design and methodology**

Chapter 2 contains the research design and methodology used in the research.

#### **CHAPTER 3: Literature review**

The literature review provides the theoretical foundation for the research and commences with an overview of prior research. The review establishes the definition, capabilities and

applications of cognitive computing. Thereafter the review focusses on corporate governance, IT governance and data protection, as well as the use of control frameworks to achieve effective governance. In order to select the most appropriate control framework to identify risks pertaining to the implementation of cognitive computing three control frameworks are reviewed. Chapter 3 forms the basis for the findings in Chapters 4, 5 and 6.

#### **CHAPTER 4: Core components of a cognitive computing system**

In Chapter 4 the core components of a cognitive computing system are identified and classified into specific phases based on their function within the cognitive computing system. Each component is defined and the purpose of each component describe within the context of the cognitive computing system.

#### **CHAPTER 5: Risks pertaining to the implementation of a cognitive computing system**

In this chapter, the control framework selected in Chapter 3, is used to identify and investigate the significant risks pertaining to the implementation of a cognitive computing system, based on the knowledge obtained in Chapter 4. Chapter 5 concludes with a risk matrix which maps the relevant core components of each cognitive computing phase with the significant risk which they introduce.

#### **CHAPTER 6: Safeguards and controls in a cognitive computing system**

In Chapter 6, the appropriate internal control techniques to mitigate the significant risks identified in Chapter 5 are formulated. Chapter 6 concludes with a risk-control matrix which maps the significant risk to the relevant control techniques formulated to mitigate the specific risk to an acceptable level.

#### **CHAPTER 7: Conclusion**

Chapter 7 presents a summary of the findings of the research and identifies areas for future research.

## CHAPTER 2: RESEARCH DESIGN AND METHODOLOGY

In this non-empirical study, literature from peer reviewed and non-peer reviewed sources were considered in order to provide a foundation for the research. According to Levy and Ellis (2006) an effective literature review should provide a theoretical foundation for the proposed study and establish that the proposed research will enhance current knowledge or contribute something new to the overall body of knowledge. An effective literature review is enabled by a process which consists of a structured approach to collect, comprehend, synthesise and evaluate quality literature (Levy & Ellis, 2006; Webster & Watson, 2002). Sylvester, Tate and Johnson (2010) suggest that a structured literature review should be performed in five stages. In the beginning stages relevant literature with a broad scope is accumulated and as the stages progress the selection is reduced to literature focussed on a specific area. Only four of the five stages were considered relevant to this study and were used:

1. **The searching stage:** The initial search criteria to identify and select relevant literature was deliberately diverse and with a broad scope. The terms used in the initial search included: “cognitive computing”, “cognitive technologies”, “cognitive analytics”, “cognitive computing and big data analytics”, “artificial intelligence”, “business value of cognitive computing”, “IT governance”, “corporate governance”, “control frameworks” and “risks related to cognitive computing, big data analytics and big data”. The sources used in the search include printed books and e-books, organisational articles and white papers, theses, scholarly articles published in local and international academic journals, electronic databases (such as IEEE, Elsevier, Emerald, Scopus) and web articles. Cognitive computing as a new generation technology has limited research available, therefore the literature reviewed was not evaluated or discarded based on the quality, academic focus or the reputation of the sources. Seeing as big data, big data analytics and cognitive computing were included in the search, it yielded 323 articles, books and white papers.
2. **The mapping stage:** The mapping stage focuses on narrowing the scope by identifying recurring themes, keywords and phrases. The themes, keywords and phrases identified during this stage included “data and information governance”, “COBIT”, “ITIL”, “COSO”, “cognitive computing systems”, “IBM’s Watson”, “machine learning”, “big data analytics”, “big data analytics and related risks” and “data privacy and security”. In order to refine the collection of literature the abstracts, introductions and conclusions were studied. This reduced the collection of literature to 131 articles, books and white papers.

3. **The appraisal stage:** During this stage the refined selection of articles, books and white papers were read, analysed and the contributions to the identified concepts were linked. Important concepts regarding cognitive computing and cognitive computing systems, the significant risks enterprises are exposed to due to the implementation of a cognitive computing system and the related mitigating control techniques, were identified, categorised and grouped.
4. **The synthesis stage:** During this stage the available literature from the previous stages are synthesised in order to enable a consistent approach in reaching conclusions and assists in creating a clear and structured final document.

The literature review, as described above, provided a theoretical foundation for a deeper understanding of cognitive computing, cognitive computing systems, corporate and IT governance, and control frameworks.

The objective of the research is to identify and mitigate significant risks pertaining to the implementation of cognitive computing. In order to achieve this objective the research was structured in the following manner:

- a) **Defining cognitive computing and the core components of a cognitive computing system:** The aim was to define cognitive computing, as well as the fundamental capabilities which differentiate cognitive computing from other computing, based on the knowledge obtained from available literature. Thereafter the core components of a cognitive computing system were identified and classified into specific phases based on their function within the cognitive computing system.
- b) **Selecting the most appropriate control framework to achieve the objective of this research:** The content, scope, benefits and limitation of a selection of control frameworks were compared, in order to select the most appropriate control framework to identify significant risks pertaining to the implementation of a cognitive computing system. The control frameworks reviewed included Control Objectives for Information and Related Technology (COBIT 5), IT Information Library (ITIL 4) and Committee of Sponsoring Organisation (COSO 2013). Based on the comparison COBIT 5 was selected as the most appropriate control framework.
- c) **Performing a study of COBIT 5 and use the detailed processes of COBIT 5 to identify significant risks associated with the implementation of a cognitive computing system:** The detailed processes of COBIT 5 were studied and relevant processes pertaining to the

governance and management of cognitive computing were identified. These relevant processes were used to identify significant risks associated with the implementation of a cognitive computing system. The significant risks were mapped to the relevant core components of each cognitive computing phase which introduces the risk (Risk matrix).

- d) **Formulating controls to mitigate the significant risk pertaining to the implementation of a cognitive computing system:** After identifying the significant risks appropriate controls were formulated to mitigate the risks to an acceptable level. The controls were mapped to the relevant risks that they will mitigate in a risk control matrix.

The methodology used in the research provided the foundation to obtain insight into cognitive computing and the core components of a cognitive computing system, identify and investigate the significant risks pertaining to the implementation of cognitive computing system and formulate safeguards to mitigate these risks.

## CHAPTER 3: LITERATURE REVIEW

### 3.1 Introduction

The first section (3.2) of the literature review provides a review of prior cognitive computing studies. The second section (3.3) of the literature review briefly establishes the definition, capabilities and application of cognitive computing. The primary focus of the third (3.4) and fourth (3.5) sections of the literature review is corporate governance, IT governance, data protection and the use of control frameworks to achieve effective IT governance. Three control frameworks are reviewed and compared in order to select the most appropriate control framework to employ in the identification of significant risks pertaining to the implementation of a cognitive computing system (Chapter 5).

### 3.2 Historic review of prior research

The majority of research on cognitive computing has been conducted by independent organisations such as IBM Corporation (Drury, Harper, Marshall & Sarkar, 2015; Fox, Lala & Coelho, 2015; Sarkar & Zaharchuk, 2015; Ballard, Compert, Jesionowski, Milman, Plants, Rosen & Smith, 2014; Jewell, Barros, Diederichs, Duijvestijn, Hammersley, Hazra, Holban, Li, Osaigbovo, Plach, Portilla, Saptarshi, Seera, Stahl & Zolotow, 2014; Sudarsan, 2013; IBM Corporation, 2014a; IBM Corporation, 2014b; IBM Corporation, 2014c; High, 2012); Accenture (Bataller & Harris, 2015); SAS Institute (n.d.); and Deloitte (Danson, Pierce & Shilling, 2015; Schatsky, Muraskin & Gurumurthy, 2014; Ronanki & Steier, 2014a; Ronanki & Steier, 2014b). The focus of these articles and white papers are to define cognitive computing, describe the underlying technologies and the capabilities of the technologies, present a perspective on how the technology will impact business, and the opportunities that exist. IBM has focused much of their research on IBM Watson, one of the first cognitive computing platforms, while Accenture offers a perspective on cognitive computing challenges and presents a framework for understanding in what ways cognitive computing can deliver value.

Recent research concentrated on providing an overview of cognitive computing. These include the book by Hurwitz *et al.* (2015) on cognitive analytics and big data analysis. The book focuses on imparting both theoretical and practical guidance to technologists. On a theoretical level, the book defines cognitive computing, the elements within a cognitive system and the underlying technologies. On a practical level, case studies from the financial, healthcare, and

manufacturing industries address the design and testing of cognitive systems. In the book '*Smart Machines*', Kelly III and Hamm (2013) introduce the world of cognitive systems to general audiences and investigate the future of computing. Kelly III and Hamm's (2013) comprehensive perspective describes the technology and explains how it will assist in the utilising and understanding of big data.

Academic research has also been conducted. Wang (2011 & 2009) explored the theoretical foundations of cognitive computing in terms of cognitive informatics, denotational mathematics, and neural informatics. A survey by Wang (2011) focused on a theoretical framework, architectural techniques, and conceptual models of cognitive computing.

The literature review established that prior research discussed the underlying technologies which forms part of the cognitive computing system, as well as the general challenges relating to cognitive computing. The prior research did not present an approach to identify specific risks pertaining to the different components of a cognitive computing system. This research proposes to address this gap by providing a structured approach to identify significant risks pertaining to the implementation of a cognitive computing system, with a specific focus on linking the underlying core components to the significant risks the component generates.

### **3.3 Cognitive computing**

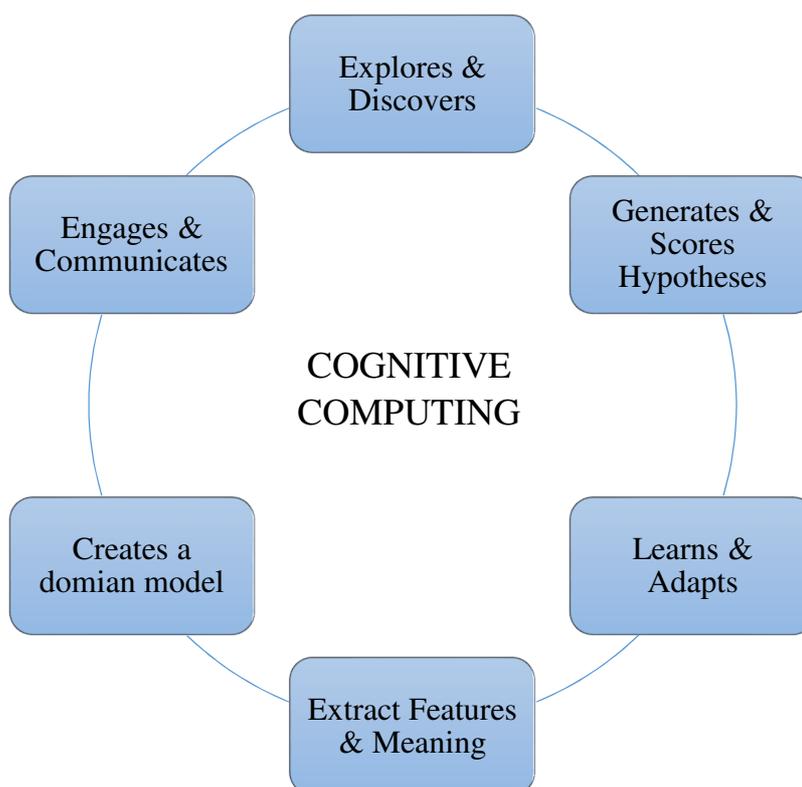
The evolution of information technology and computing consists of three eras. The first and second era encompassed instruction-driven computing, while the third era focuses on data-driven computing. The first era, known as the *tabulating era*, commenced in the 19th century and comprised of computers which automated the process of logging numbers and performing calculations. The second era, known as the *programmable computing era*, developed in the 1940s and is based on the Von Neumann architectural principles. These computers perform tasks, such as calculations and storing of information, based on a set of instructions embedded in software. Programmable computers are still used today, but do not support the enormous amounts of data generated daily by digital technologies (Kelly III & Hamm, 2013; Wladawsky-Berger, 2013).

*Cognitive computing* represents the third era in the evolution. An era which will be characterised by a collaboration between humans and machines (Kelly III & Hamm, 2013). Cognitive computing systems extract meaning from data, gain insight and solve problems in the same manner as the human brain does (Wladawsky-Berger, 2013).

### 3.3.1 Defining cognitive computing

In a 2014 cognitive computing survey Steve Adire, a contributor to the survey, defined cognitive computing as: “*Natural language processing of structured, unstructured, streaming in Big Data or Smart Data layers with machine learning for reasoning and learning to generate contextual patterns and associations that enables humans to connect the dots faster and smarter for more informed decisions to drive better outcomes*” (Zaino, 2014:2).

Adire’s definition highlights that cognitive computing consists of several components, which enables different capabilities when combined. Some of these components are fundamental to cognitive computing, while others may differ depending on the objective of the cognitive computing system, as well as the approach used to design the cognitive computing system.



**Figure 1: The fundamental capabilities of cognitive computing**

Figure 1 presents the fundamental capabilities of a cognitive computing system which differentiate cognitive computing from other computing (Hurwitz *et al.*, 2015; Noor, 2015; Bellissimo, 2015; Sarkar & Zaharchuk, 2015; Zaino, 2014; Oberlin, 2012):

- **Explores and discovers:** Cognitive computing systems uses context driven dynamic algorithms to discover patterns and insight in vast amounts of data that otherwise would have remained obscured knowledge.
- **Generates and scores hypotheses:** Cognitive computing systems generate, evaluate and score contradictory hypotheses based on its corpus of knowledge. The cognitive computing system is bias free and probabilistic, therefore it presumes that there are multiple correct answers for a hypothesis and selects the most appropriate answer based on the applicable data.
- **Learns and adapts:** Cognitive computing systems learn from experience with data, evidence, and hypotheses; and based on this experience, the system is able to improve its knowledge and its performance without direct programming.
- **Extracts features and meaning:** Cognitive computing systems extract meaning and make sense of unstructured text data through Natural Language Processing and extract features from non-text data (images, videos, voice, and sensors) through deep learning tools.
- **Creates domain models:** Cognitive computing systems constructs a model of a domain, which includes internal and external data, in the corpus and create assumptions to determine what learning algorithms are required to enable the system to learn.
- **Engages and communicates:** Cognitive computing systems are highly interactive, facilitating advanced communication between human and computer. These systems offer expert assistance by gaining deep domain-specific insights and providing this information to people in a timely, natural and usable format.

### 3.3.2 Application of cognitive computing

According to Bataller and Harris (2015), cognitive computing capabilities can be divided into four types of activity models. Each model utilises the cognitive computing system in a different manner to create value for the enterprise. The following activity models can be implemented:

1. **Efficiency model:** The efficiency model provides consistent, low cost performance for routine, predictable, rule-based activities. In this model the cognitive computing system senses, comprehends and acts, while humans monitor the accuracy of the results, and determine how the rules need to evolve as conditions change (Bataller & Harris, 2015). The following examples illustrate how cognitive computing systems are used within industries to facilitate efficiency:

- **Communication industry:** Call centres implement cognitive computing systems to provide relevant and accurate automated responses to inquiries posed in natural language. This assists call center employees in quickly retrieving the correct responses, thereby improving call center productivity as well as customer satisfaction (Fox *et al.*, 2015).
  - **Banking industry:** Banks are employing cognitive computing systems to assist customers in making better investment decisions based on their individual preferences. The cognitive computing system searches large quantities of data in order to answer specific questions, enable dialogue regarding investment options and deliver evidence-based recommendations (Drury *et al.*, 2015).
2. **Effectiveness model:** An effectiveness model provides seamless integration and collaboration for routine, predictable, rule-based activities. However, the data is more complex in comparison with the efficiency model due to an increase in volume and unstructured data. In this model, the cognitive computing system acts as a personal assistant which assist humans in scheduling, communicating, monitoring and executing activities (Bataller & Harris, 2015). Cognitive computing systems can be used within consumer products to facilitate effectiveness. Virtual agents receive requests in a textual or verbal form, process them in the cognitive computing system using natural language processing or speech recognition, search the knowledge repositories (corpus), formulate hypotheses and consequently provide answers to the users in text or speech. For example, consumers are utilising agents such as Siri, Cortana and Google Now on their smartphones, while in a corporate situation, virtual agents will answer routine questions that come into a customer service center (Bataller & Harris, 2015). In the banking industry cognitive computing systems assist in account management, security management and identity management.
3. **Expert model:** An expert model provides specialised expertise for ad-hoc, unpredictable, judgment-based activities. In this model, the cognitive computing system explores vast data stores, and subsequently makes inferences and recommendations based on the knowledge obtained during the exploration. The humans will make the ultimate decision based on recommendations (Bataller & Harris, 2015). The following examples illustrate how cognitive computing systems are used to provide expert advice:
- **Healthcare industry:** A medical diagnostic system is an example of an advisory cognitive computing system that enhances human understanding and judgement. The system analyses patient data, medical literature and guidelines from world-class experts

in order to provide data-driven recommendation. The medical doctors interpret the results of the expert system's analysis and present a diagnosis to the patient in person (Bataller & Harris, 2015).

- **Banking industry:** The cognitive computing system supports wealth management by improving the advice and experience provided to customers. The system analyses research reports, product information and customer profiles in order to identify connections between customers' needs and the investment knowledge in the corpus, and to weigh the available financial options. Based on the recommendations provided by the cognitive computing system, the bank's relationship managers will be able to provide services in accordance with the client's needs in a timely manner and create a personalised client experience (Drury *et al.*, 2015).
4. **Innovation model:** The innovation model enhances ideas and creativity by identifying alternatives and optimising recommendations. However, the data is unstructured and more complex in comparison with the expert model due to an increase in volume. The cognitive computing system enhances creativity and ideas of biomedical researchers, fashion designers, chefs, musicians and entrepreneurs (Bataller & Harris, 2015). The following examples illustrate how cognitive computing systems are used within industries to augment innovation:
- **Music industry:** The cognitive computing system assists artists and producers by using spectral sound-wave analysis to analyse songs and provide recommendations to increase the likelihood that the song will be a hit (Bataller & Harris, 2015).
  - **Healthcare industry:** The cognitive computing system assists biomedical researchers by extracting information from scientific literature and automatically identifying direct and indirect patterns, thereby accelerating research and discovering new insights. The time required for researchers to test hypotheses and formulate conclusions can be extensive and cognitive computing is a new and innovative method to advance and accelerate medical research (Sarkar & Zaharchuk, 2015).

The implementation of cognitive computing systems creates value in enterprises; however, it also exposes the enterprises to new and additional risks. In order to mitigate the exposure to these risks, enterprises should consider corporate governance, IT governance and control frameworks.

### **3.4 Corporate governance and Information Technology governance**

The importance of corporate governance significantly increased over the past decade in response to corporate fraud, managerial misconduct and negligence that lead to the loss of shareholder wealth (Krechovská & Procházková, 2014; ISACA, 2012a). Information technology has become a core asset for most enterprises and in order to enhance corporate governance, it must be supported by the effective management of these IT assets (IT governance). The effective governance of cognitive computing (IT asset) requires investigation in order to strengthen and support corporate governance.

#### **3.4.1 Corporate governance**

Corporate governance is a structure of policies, practices and procedures by which enterprises are controlled, directed and organised (Grose, Kargidis & Vasilios, 2014; Krechovská & Procházková, 2014; Zalewska, 2014). Governance ensures that stakeholder needs are assessed to determine appropriate enterprise objectives that the decision making and procedures of the enterprise aligns with the objectives, and performance and compliance are monitored in accordance with the objectives (Grose *et al.*, 2014; Gartner, 2013; ISACA, 2012a).

Effective corporate governance principles promote fairness, accountability, responsibility and transparency which is fundamental in managing an enterprise successfully (IODSA, 2009). The third King Code of Corporate Governance for South Africa (King III) was specifically written to assist South African enterprises in achieving good corporate governance principles. However, in order to maintain good corporate governance, the principles must be adjusted and grow with changes in business environment, new trends and new concepts (Krechovská & Procházková, 2014). The King report also had to adapt to these changes and as a result the King III specifically includes IT governance principles to address new developments in IT environments (IODSA, 2009).

On 15 March 2016, The Institute of Directors in Southern Africa (IODSA) and the King Committee made the draft version of King IV available for public comment. King IV addresses a number of new developments in corporate governance, including executive remuneration and the role of both social and ethics committees (IODSA, 2009).

Information systems have become an integral part of all businesses, revealing and creating new risks. Directors must take prudent and reasonable steps to generate business value and mitigate risks by incorporating international guidelines and frameworks (Juiz & Toomey, 2015; IODSA, 2009).

### 3.4.2 Information Technology governance

Information technology is a key resource for all enterprises and as a result IT governance forms an integral part of corporate governance. The core principle of IT governance is that it is a structured mechanism (framework) that enables the effective and efficient management of IT resources and assists a company in achieving its strategies and objectives (Gartner, 2013; Goosen & Rudman, 2013a; ISACA, 2012a; Marks, 2010; IODSA, 2009).

According to King, the responsibility of implementing good IT governance principles lies with the directors and senior management and recommends utilising international guidelines set by ITGI, ISACA and ISO authorities to assess and assist in IT governance (Gartner, 2013; Goosen & Rudman, 2013a; Marks, 2010; IODSA, 2009). Regardless of the guideline applied by those charged with this responsibility, King III contains seven principles that South African organisations should adhere to (IODSA, 2009; Liell-Cock, Graham & Hill, 2009):

- 1. The board should be responsible for information technology governance:** The board of a company is responsible for IT governance and should ensure that appropriate IT policies are established and implemented.
- 2. IT should be aligned with the performance and sustainability objectives of the entity:** The board should provide direction to facilitate the integration of the IT strategy and IT processes, with the company's overall strategic and business processes.
- 3. The board should delegate the responsibility for the implementation of an IT governance framework to management:** Directors should assign the responsibility for the implementation of the structures, processes and mechanisms for the IT governance framework to management.
- 4. The board should monitor and evaluate significant IT investments and expenditure:** The board should ensure that IT resources and projects deliver the promised returns and benefits, with specific focus on adding value by optimising cost efficiency.
- 5. IT should form an integral part of the entity's risk management process:** Management should ensure that IT risks are included in the company's risk management process. An adequate business resilience plan should be in place to recover from business disruptions and disasters, and the board must ensure that the company complies with IT laws and regulations.
- 6. The board should ensure that information assets are managed effectively:** Processes should be implemented to enhance the performance and sustainability of the

company through the effective use of IT resources. The board has the responsibility to ensure that there are systems in place for the effective management of information, including information security, information privacy and the identification and processing of personal information in accordance with applicable laws and regulations (Refer to section 3.4.3 Data protection and cognitive computing).

- 7. A risk committee and audit committee should assist the board in carrying out its IT duties:** The risk committee should ensure that all IT risks are identified and properly addressed, while the audit committee should assess IT in relation to the company's financial reporting and going concern.

In order to effectively apply the IT governance principles listed above, an integrated framework should be implemented that provides structures and processes which align business and IT, and attend to all relevant risk areas pertaining to the implementation of the cognitive computing system (Goosen & Rudman, 2013a; Liell-Cock *et al.*, 2009). An essential element of the IT governance principles is that the board and management should obtain an understanding of the laws and regulations applicable to the cognitive computing system. King III specifically addresses privacy of information, therefore it is important to obtain an understanding of data protection laws and regulation in South Africa in the context of a cognitive computing system.

### **3.4.3 Data protection and cognitive computing**

Cognitive computing stimulates innovation and discovery. However, it is fundamental to eliminate the possibility of individuals, within or outside the organisation, misusing private data in unjustifiable and intrusive ways, which would incur both regulatory action, as well as financial losses and reputational damage (Wang, 2009). In accordance with Principle 7 of King III, the board should ensure that all personal information is identified and that there are systems in place to manage information security and information privacy (IODSA, 2009). As such, organisations must adhere to data protection legislation.

The Protection of Personal Information Act No. 4 of 2013 (POPI Act) was enacted on 26 November 2013 and the commencement date is expected to be in the second half of 2016, after the local government elections in August 2016 (de Bruyn, 2014). The POPI Act arises out of global developments with regard to data protection regulation and the purpose of this legislation is to enhance local privacy regulation and to prescribe data protection practices that aligns South Africa's data privacy and protection legislation with global best practice (Jangara & Bezuidenhout, 2015; PWC, 2011).

The POPI Act presents eight conditions for the lawful processing of personal information (de Bruyn, 2014; POPI, 2013; PWC, 2011):

1. **Accountability:** The responsible party must ensure that the conditions for lawful processing with regard to the POPI Act are implemented and must also monitor adherence to the conditions.
2. **Processing limitation:** The responsible party must ensure that personal data is processed in a fair, lawful and reasonable manner, which does not infringe the right to privacy of the data subject. Processing limitations ensure that consent is provided by the 'data subject' and that objections from the data subject is adhere to, if justified in accordance with the provisions of the Act.
3. **Purpose specification:** Private data must be collected for a specific purpose, and the 'data subject' must be notified of the collection of their personal data as well as for what specific purpose it is collected.
4. **Further processing limitation:** Further processing of personal data must be consistent with the original purpose for which collection took place. Therefore prohibiting excessive processing.
5. **Information quality:** The private data collected must be accurate and relevant, to ensure the quality of information.
6. **Openness:** The data subject must be notified of the collection of their personal data and documentation must be maintained in terms of Section 14 to 51 of the Promotion of Access to Information Act.
7. **Security safeguards:** Policies and practices must be implemented to ensure the security, confidentiality and integrity of personal data, and that data subjects are notified of security breaches.
8. **Data subject participation:** The data subjects must have access to their personal information and have the right to request the correction or deletion of the personal information.

In order to identify and address confidentiality and privacy risks in a cognitive computing environment, these conditions must be taken into account. Although the POPI Act does not have a commencement date, de Bruyn (2014) determined that the core principles of the POPI Act and the UK's Data Protection Act are significantly similar and therefore it can be concluded that it is reasonable to expect that the implementation of both these acts will produce similar outcomes.

### 3.4.4 IT gap and alignment

Within an organisation, the board is responsible for the implementation of an IT governance control framework, while IT specialists are responsible for the implementation of control techniques as indicated by the control framework. This is problematic, seeing as the board and top management have insufficient knowledge with regards to the control techniques and technology, and the IT specialist lacks understanding of the framework. This is known as the 'IT gap'. The IT gap causes a misalignment between IT strategies / processes and business strategies / processes, which in turn creates risks and weaknesses in an IT system. In order to bridge the gap, the board must focus on integrating business and IT strategies by using a framework which facilitates alignment (Goosen & Rudman, 2013b; Rudman, 2010; Rudman, 2008).

### 3.5 Control frameworks

Effective governance of IT requires the implementation of a control framework which aligns business and IT, and addresses all IT related risks and relevant control areas (Juiz & Toomey, 2015; Goosen & Rudman, 2013a; Rudman, 2010).

There are numerous established standards, frameworks and best practices available to govern IT, which include Control Objectives for Information and Related Technology version 5 (COBIT 5), Information Technology Information Library (ITIL), ISO38500:2008 and ISO/IEC 27002. These frameworks can be applied individually or combined in order to establish a comprehensive IT framework which eliminates the weaknesses and combines the strengths of the various frameworks (Rubino & Vitolla, 2014). However, combining the frameworks may result in an ineffective control structure where work may be duplicated, controls could overlap and management may find it difficult to obtain a comprehensive understanding of the organisation's risk exposure and control processes (Anisingaraju, 2013).

According to Zhang and Le Fever (2013), control frameworks can be separated into three main categories:

1. Business oriented controls:
  - (i) Committee of Sponsoring Organisation (COSO);
  - (ii) Statement of Auditing Standards (SAS);
2. IT focused controls:
  - (i) Information Technology Infrastructure Library (ITIL);

(ii) ISO/IEC 17799:2000, ISO/IEC 27000;

3. Business-IT alignment focused controls:

(i) Control Objectives for Information and Related Technology (COBIT).

One framework from each category was selected for review. COSO, ITIL and COBIT 5 were reviewed in order to choose the most suitable framework for the identification of cognitive computing risks.

### 3.5.1 Control Objectives for Information and Related Technology

Control Objectives for Information and Related Technology version 5 (COBIT 5) is a globally accepted, comprehensive framework that enables enterprises to create value through the effective governance and management of Information Technology (Rubino & Vitolla, 2014; Huang, Hung, Yen, Chang, & Jiang, 2011). The objective of this framework is to find a balance between the benefits and risks of IT, while considering the interests of all stakeholders (ISACA, 2012a).

COBIT was developed by ISACA (Information Systems Audit and Control Association) and was initially used as a framework for executing IT audits. The COBIT framework was developed further and in April 2012, the newest version, COBIT 5, was released. COBIT 5 consolidates and incorporates other frameworks such as Val IT and Risk IT and was updated to be in accordance with ITIL practices (Rubino & Vitolla, 2014). Val IT focuses on the attainment of business value through investment in IT and Risk IT addresses risk management (Sahibudin, Sharifi & Ayat, 2008). ISACA (2012) indicates that COBIT 5 is focused on five key principles:

- **Principle 1:** Meeting stakeholder needs by creating business value through the use of information technology, specifically by transforming business objectives into information technology related objectives.
- **Principle 2:** Covering the enterprise end-to-end by incorporating governance and management of enterprise information and related information technology into enterprise wide governance.
- **Principle 3:** Applying a single, integrated framework by aligning appropriate standards and frameworks to function as an overarching framework for governance and management of information technology.

- **Principle 4:** Enabling a holistic approach through the use of categorised enablers (for example, policies, processes etc.), in order to create efficient and effective governance and management of information technology.
- **Principle 5:** Separating governance from management since these two disciplines involve different types of activities it necessitates different organisational structures and achieve different goals.

COBIT groups 34 IT processes into five domains (Kusumah, Sutikno & Rosmansyah, 2014; Rubino & Vitolla, 2014; Goosen & Rudman, 2013a; ISACA, 2012a; Sahibudin *et al.*, 2008).

The domains are defined as follows:

- **Evaluate, direct and monitor (EDM):** The five processes included in the EDM domain provide guidance on the successful governance of IT-enabled business investments; through structures, principles, processes and practices; in order to achieve the company's objectives.
- **Align, plan and organize (APO):** The APO domain comprises of thirteen processes which provide guidance on the effective utilisation of internal and external IT resources in order to achieve business objectives and optimal IT results.
- **Build, acquire and implement (BAI):** The ten processes included in the BAI domain provide guidance on the processes required to implement the IT strategy, specifically how to identify, develop / acquire, implement and integrate IT solutions.
- **Deliver, service and support (DSS):** The six processes of the DSS domain focus on the delivery and support of services required by end users. The domain also covers the management of security; continuity; training; and data and operational facilities.
- **Monitor, evaluate and assess (MEA):** According to the three processes included in the MEA domain, all IT processes must be continuously assessed to ensure quality and compliance with control requirements. The processes include performance management, monitoring of internal control, regulatory compliance and governance.

The benefits of applying COBIT 5 are the following (Bartens, de Haes, Lamoen, Schulte, & Voss, 2015; Crespo, 2015; Anisingaraju, 2013; Kneller, 2010; ISACA, 2012a):

- COBIT 5 offers an end-to-end business approach which integrates IT governance and enterprise governance, taking into account the interests of both business and IT stakeholders in the process;

- COBIT 5 integrates (COBIT 4.1, Val IT 2.0, Risk IT, BMIS) and aligns (ITIL, TOGAF and ISO standards) with other universally established standards, frameworks and practices, in order to create an overall, comprehensive governance and management framework;
- The framework achieves optimal value from IT investment by creating a balance between realising benefits, optimising risk and effective and innovative utilisation of IT resource;
- COBIT 5 improves user satisfaction through IT engagement and services by creating an increase in the contributions of users to the investment and use of IT;
- The framework supports compliance with relevant local and international laws, regulations, and internal policies;
- The framework focuses on aligning business needs and IT objectives;
- The framework is generic and flexible, therefore it can be adapted to suit any enterprise's specific situation. However the enterprise (user) is responsible for selective implementation of COBIT 5 processes; and
- COBIT 5 and the 2013 COSO Internal Control – Integrated Framework are complimentary and compatible.

COBIT 5 has the following inherent limitations (Bartens *et al.*, 2015; Anisingaraju, 2013; Zhang & Le Fever, 2013; Kneller, 2010):

- There is a lack of implementation guidance, specifically for selective implementation and customisation;
- COBIT requires detailed understanding and significant resources for its implementation and this could exclude small- and medium-sized companies from applying the framework; and
- The complicated concepts and structure of COBIT 5 guidance may discourage new users and prevent its adoption.

### **3.5.2 Information Technology Infrastructure Library**

Information Technology Infrastructure Library (ITIL) is a comprehensible framework of best practices in IT service management and supports the governance, management and control of IT services (ITIL, 2012). ITIL was developed by the UK's Office of Government Commerce (OGC) to provide value to users in the form of services (Peña, Vicente & Ocaña, 2013). It is currently managed by the Information Technology Service Management Forum (ITSMF) and is the most widely established approach to IT Service Management. ITIL is structured in five

lifecycle phases (ITIL, 2012). The phases are defined as follows (Kusumah, Sutikno & Rosmansyah, 2014; Peña *et al.*, 2013; ITIL, 2012):

- **Service strategy:** This phase provides guidance on how to transform service management into a strategic asset and using it to achieve strategic objectives. The processes included in this phase are Financial Management, Service Portfolio Management and Demand Management.
- **Service design:** This phase provide guidance for the design and development of information technology services, including their architectures, processes and policies, to ensure quality service delivery, customer satisfaction and cost-effective service provision. The processes are Service Catalogue Management, Service Level Management, Capacity Management, Availability Management, Information Security Management, Supplier Management and IT Service Continuity Management
- **Service Transition:** This phase provides guidance for transitioning new and modified services into operational use, ensuring that the service strategy requirements are followed through to the service design and effectively implemented in the operation phase. The processes are Change Management, Service Asset and Configuration Management, Release and Deployment Management, Knowledge Management, Service Validation and Testing, Evaluation.
- **Service operation:** This phase provides guidance to coordinate and perform processes necessary to deliver and manage services effectiveness and efficiency at agreed levels to business users / service provider and customers. During this phase of the lifecycle, the services deliver value to the business by realising the strategic objective. The processes are Incidence Management, Event Management, Problem Management, Access Management and Request Fulfilment.
- **Continual service improvement:** This phase provides guidance on maintaining and creating value for customers through the assessing and advancing the quality of services and overall maturity of each phase and its underlying processes.

The benefits of applying ITIL are the following (ITIL, 2012; Kneller, 2010; Fry, 2005):

- ITIL enables organisations to align IT services and business objectives in order to increase benefits and create a return on investment;
- Costs are reduced as a result of increased business productivity and effective resource management;

- ITIL has a philosophy of continuous improvement, which is aided by well-defined, consistent processes;
- ITIL provides a standard set of terminology which facilitates better communication between all internal stakeholders; and
- ITIL consists of standard processes which create reliable, consistent and available IT Services. These IT services enhances user satisfaction.

ITIL has the following inherent limitations (Küller, Grabowski, PetrSameš & Vogt, 2010; Fry, 2005):

- The complexity of ITIL discourages implementation, specifically for Small and Medium-sized Enterprises;
- ITIL does not contain sufficient work instructions and practical guidance in order to implement and maintain processes; and
- Processes which affect multiple departments may result in interdepartmental conflicts, especially where the performance of each department is evaluated independently.

### 3.5.3 Committee of Sponsoring Organizations

Committee of Sponsoring Organizations (COSO) is the most widely applied internal control framework for designing, implementing, and managing internal controls, as well as evaluating the effectiveness of these controls (Rubino & Vitolla, 2014; D'Aquila, 2013). The Committee of Sponsoring Organizations issued the Internal Control–Integrated Framework in 1992 and revised it in 2013. The 2013 COSO framework version groups 17 principles over five integrated components of internal control (Rubino & Vitolla, 2014; D'Aquila, 2013; Rittenberg, 2013).

- **Control environment:** The control environment consists of five principles and establishes a set of standards, processes, and structures that provide the foundation for applying internal control across the entire organisation.
- **Risk assessment:** Risk assessment provides a process to identify and assess the risk of an organisation not achieving its objectives, not producing reliable financial reporting and not considering business and technological changes which could significantly impact the organisation's internal control system. The component consists of four principles which also create a basis for deciding how to manage the identified risks.

- **Control activities:** These are actions developed and established by policies and procedures to assist management in mitigating risks in order to achieve the organisation's objectives. These activities are implemented at all levels of an organisation, at various stages within business processes, and throughout the information technology environment. This component consists of three principles.
- **Information and communication:** This component, which consist of three principles, emphasises the importance of relevant, quality information and efficient communication processes. The principles cover both internal and external communicate of information.
- **Monitoring activities:** This component, which consist of two principles, requires management to perform evaluations to ascertain whether all five components of internal control, including controls relating to the principles within each component, are in place and operating efficiently. In accordance with this component, any deficiencies identified must be communicated in a timely manner and corrective actions taken.

The benefits of applying COSO are the following (McNally, 2012; Küller *et al.*, 2010):

- The COSO framework is flexible, which allows the framework to be applied to various business and operation models;
- The framework offers agility to adapt the internal controls to changing business needs.
- COSO produces an effective system of internal control through its cohesive approach to all controls within an organisation;
- COSO 2013 provides coverage for financial, operational and compliance reports; and
- COSO 2013 provides additional guidance for implementation, which enables more effective internal controls at lower costs.

COSO has the following inherent limitations (Rubino & Vitolla, 2014):

- The COSO framework only provides high-level guidance for internal controls and does not stipulate detailed control objectives which auditors require in the design of audit tests; and
- While COSO 2013 increased the focus on technology, it still does not provide detailed guidance on the evaluation of specific controls relating to technology or addressing risks and complexities of IT.

### 3.5.4 Framework selected for the purposes of this research

The scope, content, benefits and limitations of the COSO, ITIL and COBIT frameworks were reviewed in order to select the most suitable framework for the identification of cognitive

computing risks. Table 1 summarises the benefits and limitations of the COBIT, ITIL and COSO control frameworks.

**Table 1: The benefits and limitations of COBIT, ITIL and COSO**

	COBIT 5	ITIL v 3	COSO 2013
<b>BENEFITS</b>			
Improves alignment between IT and business strategy	X	X	
Comprehensive framework	X		
Single integrated framework	X		
Flexibility to adapt to enterprise size, business and operations models and changing needs	X	X	X
Optimal value creation (cost saving)	X	X	X
Detailed processes		X	
Standard terminology and processes (cohesive approach)		X	X
Improves user satisfaction	X		
Promotes continuous improvement of IT processes		X	X
<b>LIMITATIONS</b>			
Requires detailed understanding (complex model)	X	X	X
Significant resources required	X	X	
Lack of implementation guidance	X	X	X
Lack of detailed processes and controls	X		X
IT security not addressed	X		
Insufficient focus on IT (lacks detailed guidance)			X
Creates interdepartmental conflict		X	

(Bartens *et al.*, 2015; Crespo, 2015; Sahd, 2015; Rubino *et al.*, 2014; Anisingaraju, 2013; Zhang & Le Fever, 2013; ITIL, 2012; ISACA, 2012a; Kneller, 2010; Küller *et al.*, 2010; McNally, 2012; Fry, 2005)

COBIT 5 was chosen as the most suitable framework, given that COBIT 5 is a comprehensive framework which seamlessly integrates IT governance into enterprise governance. The framework covers IT functions and processes, as well as other business functions and processes affected by IT. COSO was not selected due to insufficient focus on IT, and ITIL because it only focused on IT service management.

### 3.6 Summary and conclusion

The second section of the literature review established the definition, capabilities and application of cognitive computing. The knowledge obtained in this section will form the foundation for the identification, classification and definition of the core components of the

cognitive computing system in Chapter 4. The third and fourth sections of the literature review provided an understanding of corporate governance, IT governance, data protection and the implementation of control frameworks to achieve effective IT governance, identify significant risks pertaining to the implementation of a cognitive computing system, as well as mitigating control techniques. Three control frameworks were reviewed and COBIT was selected as the most suitable control framework. COBIT and its detailed processes form the foundation for the identification of significant risks pertaining to the implementation of the cognitive computing system in Chapter 5, as well as the formulation of mitigating control techniques to address the significant risks in Chapter 6.

## CHAPTER 4: CORE COMPONENTS OF A COGNITIVE COMPUTING SYSTEM

The COBIT 5 framework was selected as the most appropriate framework to identify significant risks pertaining to the deployment of a cognitive computing system. In order to utilise the framework effectively, the core components of cognitive computing system need to be identified, defined and classified. Based on studies performed by Bowles *et al.* (2015), Digital Reasoning (2015) and Kelly III & Hamm (2013), the core components of a cognitive computing system were identified. The core components were further classified into specific phases based on their functions within the cognitive computing system. The core components per phase are the following:

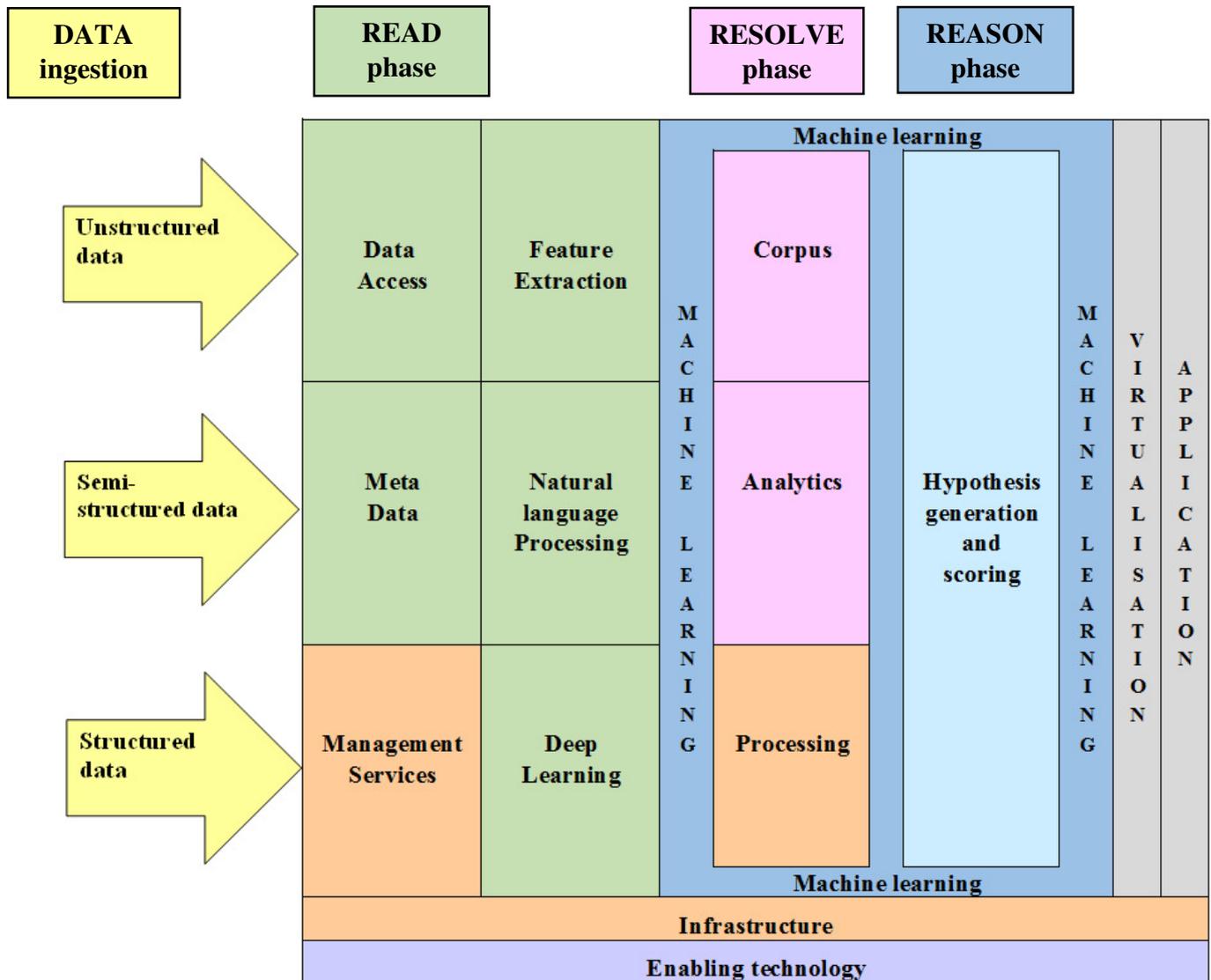
- **Data ingestion:** Unstructured data, semi-structured data and structured data.
- **Read:** Metadata, feature extraction, natural language processing (NLP) and deep learning.
- **Resolve:** Corpus and advances analytics.
- **Reason:** Hypothesis generation and scoring, and machine learning.

The four phases in the cognitive computing system are supported by:

- **Infrastructure:** Storage, processing and management.
- **Enabling technologies:** Hadoop, Cloud and Big Data.

Unstructured, semi-structured and structured data from various internal and external data sources are “ingested” into the cognitive computing system. The “Read” phase provides access to the ingested data and metadata (data regarding the origin, structure and meaning of data). In this phase Natural Language Processing (NLP) and Deep Learning are utilised to extract data elements and meaning from the ingested data and metadata in order to prepare the data and produce machine-readable data for the “Resolve” phase. In the “Resolve” phase, data from the “Read” phase is assembled, organised, and analysed in the corpus of the cognitive computing system to create a knowledge base. From the body of knowledge, hypotheses are generated and scored to uncover relationships in order to resolve problems. This is known as the “Reason” phase. The machine learning allows the cognitive computing system to continuously learn from the ingested data as well as hypotheses’ results in order to become a more effective system. These four phases are supported by an IT infrastructure and enabled by other technologies, which is also recognised as components of the cognitive computing system. Figure 2 illustrates how the core components and phases fit together in the cognitive computing system. The figure shows a further virtualisation and application phase, however these phases fall outside the

scope of this research and is only presented for completeness' sake (Refer to 1.3 Scope limitations).



**Figure 2: Core components of a cognitive computing system**

In order to gain a comprehensive understanding of each phase and its core components, the core cognitive computing components will be defined and its purpose explained within the context of the cognitive computing system. Thereafter the IT infrastructure supporting the cognitive computing system will be defined. Lastly the enabling technologies will be defined and its benefits considered. Significant risks relating to each component, IT infrastructure and enabling technologies will be considered in the following chapter (Chapter 5).

## 4.1 Data ingestion

A cognitive computing system requires a combination of a variety of data to discover patterns and anomalies, and to gain insight from the data. It also often requires a large data set in order to substantiate that the results of analyses and hypothesis are trustworthy and consistent. The following data formats should be ingested:

- **Structured data:** Structured data refers to data that has a formal structure with a distinct length and format. The semantics of such data are explicitly defined in metadata, schemas and glossaries. Sources of structured data include computer processed transactional data, as well as machine generated data from devices such as sensors (Hurwitz *et al.*, 2015; Chan, 2013).
- **Semi-structured:** Semi-structured data does not have a formal structure, however it contains tags that separate semantic elements (Chan, 2013). These user-defined tags allow the data to be machine readable. An example of semi-structured data includes extensible mark-up language (XML), a textual language for exchanging data on the Web (SAS Institute, n.d.).
- **Unstructured data:** Unstructured data in the form of written material, audio, video, and images are data with no identifiable formal structure (SAS Institute, n.d.; Chan, 2013). The semantics of the data is discovered and extracted through techniques such as natural language processing and analytics (Hurwitz *et al.*, 2015).

Structured data combined with insight gained from unstructured data forms the foundation of a cognitive computing system (Zikopoulos, deRoos, Bienko, Buglio & Andrews, 2015; Hurwitz *et al.*, 2015).

## 4.2 Read phase

A cognitive computing system does not require all ingested data to be scrubbed, interpreted, and translated into a common format. Instead, the system incorporates metadata and uses natural language processing and deep learning techniques to create linkages in data elements. These linkages are used to interpret unstructured data within its specific context and gain insight into a particular domain area of knowledge (Hurwitz *et al.*, 2015; Ronanki & Steier, 2014a; Ronanki & Steier, 2014b). This phase consists of five core components:

- **Data access:** The data access component provides the interface between the cognitive computing system and the external data sources. The applicable IT infrastructure will manage the ingested data and metadata (Hurwitz *et al.*, 2015).

- **Metadata:** Metadata provides information about data and can address the data content or the whole dataset. If the metadata refers to content, it will include the names and descriptions of specific fields as well as data definitions. This assists the cognitive computing system in understanding the composition of the data and how the data should be used and interpreted. If the metadata refers to a dataset as a whole, it will include descriptive metadata (for example, title, author, publisher, subject, and description), which assists in identification and discovery, and administrative metadata which addresses provenance of data (when and how the dataset was created), ownership of data (who owns and can use the data) and technical aspects of the data (file format). The metadata is kept within a metadata repository in the cognitive computing system (Hurwitz *et al.*, 2015; Kitchin, 2014).
- **Feature extraction services:** The feature extraction services component uses statistical algorithms to identify relevant ingested data which requires refinement, through the use of NLP, deep learning and analysis, before it can be ingested in the corpus. After the identification process is completed, the data is extracted and introduced to the NLP, deep learning or analytical (“Resolve” phase) components (Hurwitz *et al.*, 2015).
- **Natural language processing:** NLP is a set of techniques that establishes the meaning of unstructured text by utilising dictionaries, identifying recurring patterns of co-occurring words and recognising other contextual clues (Hurwitz *et al.*, 2015). Unstructured text includes raw text, handwritten content, emails, blog posts, mobile and sensor data, and voice transcriptions (Ronanki & Steier, 2014a; Ronanki & Steier, 2014b). NLP enables a cognitive computing system to (Hurwitz *et al.*, 2015; Schatsky *et al.*, 2014; Ronanki & Steier, 2014a; Ronanki & Steier, 2014b):
  - process text, written or recorded in a language used for human communication, and extract meaning from it;
  - identify linkages in data elements in order to interpret the meaning of unstructured text in the right context;
  - interact with humans by interpreting the meaning of spoken natural language and generating a natural language response; and
  - identify and extract names, location, actions and events in or across documents in order to find relationships.

NLP distinguishes cognitive computing from other data-driven analytical techniques.

- **Deep learning:** Deep learning is a variant of neural networks with multiple processing layers to allow for higher-level abstractions (features). The objective of deep learning is to identify objects and extract features in non-text based data such as videos and sensor data. Therefore deep learning must be applied in a cognitive computing system in order to transform images to interpret and capture the meaning and allow for further processing (Hurwitz *et al.*, 2015; Harper, 2015; SAS Institute, n.d).

### 4.3 Resolve phase

The “Resolve” phase is linked with the “Read” phase given that the “Read” phase provides the data which is incorporated into the corpus and analysed by advanced analytical techniques in preparation for the “Reason” phase. This phase consists of two core components:

- **The Corpus:** The corpus is the body of knowledge within a cognitive computing system that consists of a complete record of machine-readable, searchable, and comprehensible data. The base corpus focuses on a specific domain and combines validated structured, semi-structured and unstructured data relating to the domain. The content of the corpus enables the system to answer questions, discover new patterns or relationships, and deliver new insights (Hurwitz *et al.*, 2015).

The corpus determines the types of questions and hypotheses the system can solve. As a result the corpus must perform the following tasks:

- *Source acquisition:* The corpus determines and acquires the external, internal and dark data essential for the specific domain and objective of the cognitive computing system. This will include text or non-text based data, domain specific databases, ontologies, taxonomies and catalogues.
- *Source transformation and integration:* The corpus establishes if the acquired data (from previous phases) requires further advanced analytics or if it can be directly incorporated into the corpus.
- *Source expansion and updates:* The corpus identifies data sources which must be updated continuously to update and expand the corpus.

Through initial design and continuous machine learning, the cognitive computing system will learn to perform the acquisition, transformation and expansion without additional training (Hurwitz *et al.*, 2015).

The structure of the corpus is created by knowledge models, specifically ontologies and taxonomies, that provide mechanisms for determining context and meaning of concepts

(objects) within the domain, by clarifying and defining terminology, and by creating accurate mappings and a common vocabulary (Enterrasolutions, 2016; Bradbury, 2015; Hurwitz *et al.*, 2015). A taxonomy provides formal structures of the types of objects within a domain and consists of rules which are used to classify objects. An ontology, on the other hand, provides a more comprehensive approach than a taxonomy and includes vocabulary, definitions and rules. The knowledge model used within the corpus will depend on the types of queries which must be solved (Hurwitz *et al.*, 2015).

- **Advanced Analytics:** Advanced analytics refers to a collection of techniques and algorithms for identifying patterns or relationships in large, complex, or high-velocity (speed of data creation, streaming, and aggregation) data sets with varying degrees of structure. The analytics process allows the cognitive computing system to identify and understand the relationships that exist amongst data elements and puts it into context. Advanced analytics may include a combination of predictive analytics, prescriptive analytics, text analytics, image analytics and speech analytics (Hurwitz *et al.*, 2015).

In a cognitive system, machine learning is applied to the analytics to improve accuracy, reduce errors and enhance future predictions (Hurwitz *et al.*, 2015). The main categories of advanced analytics are (Gartner, 2015; Hurwitz *et al.*, 2015; Siegel, 2013; Fluss, 2011; Bailor, 2006):

- *Predictive Analytics:* Predictive analytics applies algorithms and techniques, including data mining and statistical models, to predict future outcomes. The analysis finds hidden patterns in all data types and the cognitive system uses these patterns to form the basis of the answers and predictions it makes. The difference between a cognitive system and other applications of predictive analytics is that it is not only used to predict future behaviour, but also to predict if an answer is correct. In a cognitive system the unknown factor being predicted is already known, rather than becoming known when it occurs in the future.
- *Prescriptive Analytics:* Prescriptive analytics creates a framework that supports decisions about what should and should not be done, with specific focus on consequences of actions.
- *Text Analytics:* Text analytics is the process of isolating critical information from text-based unstructured sources. Relevant information is extracted from the text, transformed into machine-readable information, and analysed to identify patterns and determine relationships and trends. It can also be used to gain insight into masked

sentiment in text, in order to improve predictive analytics. The analysis and extraction process uses techniques which was derived from computational linguistics, natural language processing, and statistics.

- *Image Analytics*: Image analytics analyses images in order to extract meaning from it. The sources used in the development of the corpus in the cognitive system will include many different types of images, such as videos, photos, or medical images. As a result, image analytics are important in cognitive computing system to identify clusters and patterns in these images. Image analytics can index and search video, photos and images by classifying objects into different categories, or to look for anomalies in a digital images.
- *Speech Analytics*: Speech analytics analyses recorded speech to extract information about either the person speaking or the content of the words. It transcribes unstructured spoken words, using a variety of techniques, into structured output. Identifying the patterns of words and phrases provides more clues of emotion and intent behind the speech and can lead to improved accuracy of predictive analytics.

The corpus and advanced analytics components are supported by IT infrastructure processing capabilities.

#### 4.4 Reason phase

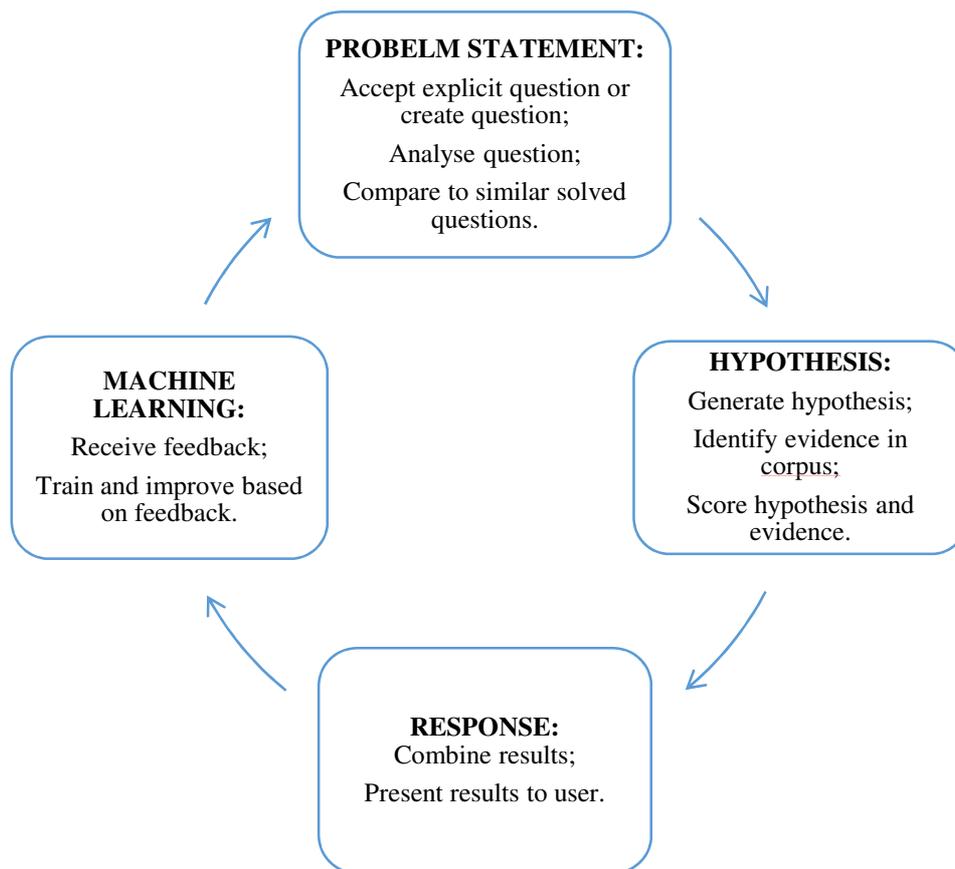
In the “Reason” phase hypotheses are generated and scored to uncover relationships, provide recommendations and resolve problems. The hypothesis is based on data / knowledge obtained from the corpus. The machine learning algorithms enable the cognitive computing system to continuously learn from the data and knowledge in the corpus, as well as hypothesis results in order to become a more effective system. This phase consist of two core components namely Hypothesis generation and hypothesis scoring and Machine learning:

- **Hypothesis generation and hypothesis scoring**: Hypothesis generation is a fundamental cognitive ability of the human brain (Lange, Thomas & Davelaar, 2013). A hypothesis has some supporting evidence or knowledge that is used to formulate plausible explanations regarding an occurrence or relationship (Hurwitz *et al.*, 2015; Lange *et al.*, 2013). A cognitive computing system explores numerous combinations of potential relationships for evidence to corroborate or refuse a hypothesis (Hurwitz *et al.*, 2015; Ronanki & Steier, 2014a). In a cognitive system, a hypothesis can be generated with or without an explicit question:

- *With explicit question:* The objective with an explicit question is to detect a relationship within a domain to generate the best potential response to the question. The system will search for a relationship between a cause and effect in the domain where a known set of causes with effects exists (Hurwitz *et al.*, 2015; Sudarsan, 2013).
- *Without explicit question:* The cognitive system continuously scrutinises the corpus to discover unusual data patterns and relationships that may reveal threats or opportunities. The nature of the new pattern or relationship identified form the basis for the creation of a hypothesis (Hurwitz *et al.*, 2015).

The hypothesis generated by the system must be scored and assigned a confidence level in order to identify the answer with the highest level of confidence. The hypothesis is compared to the data in the corpus to ascertain what evidence exists to corroborate or refute it (Hurwitz *et al.*, 2015; Sudarsan, 2013). Scores are assigned based on the relevance of the evidence and are weighted against statistical models to produce a percent confidence. The scores can be adjusted based on experience with the system and feedback through machine learning (Hurwitz *et al.*, 2015; Sudarsan, 2013). The scores algorithms, also known as reasoning algorithms, are used to create a score (High, 2012).

The hypothesis generation and scoring is a continues process which initiates when a problem is presented to the cognitive computing system with or without an explicit question. The hypothesis component analyses the question and compares it to similar questions, solved by the cognitive computing system in the past, before generating a hypothesis. While generating the hypothesis, the cognitive computing system scans through the corpus to identify knowledge that will present useful insight and valuable responses to the hypothesis. Reasoning algorithms are applied to produce scores; the resulting scores are weighted against a statistical model and a summary with recommendations are presented to the user (High, 2012). Throughout the process, machine learning takes place based on feedback from the system itself or from the users. Figure 3 illustrates the flow of the hypothesis process:



**Figure 3 Hypothesis generation and scoring process**

- **Machine learning:** Machine learning is an artificial intelligence technique, modelled after characteristics of the human brain (Ronanki & Steier, 2014a; Ronanki & Steier, 2014b). It allows computer systems to learn continuously (Hurwitz *et al.*, 2015) and improve performance by exposure to data, without the need to follow explicitly programmed instructions (Schatsky *et al.*, 2014).

The cognitive computing system uses machine learning algorithms to explore divergent concepts (found in structured and unstructured data sets) for possible connections and patterns, express potential new ideas with relative confidence through hypothesis generation and scoring, and adjust the strength of decision frameworks and future hypothesis based on direct feedback to those ideas (Ronanki & Steier, 2014a; Ronanki & Steier, 2014b; Kelly III & Hamm, 2013; Oberlin, 2012).

Three classes of machine learning algorithms are used (Hurwitz *et al.*, 2015; Ronanki & Steier, 2014b; Oberlin, 2012; SAS Institute, n.d.):

- *Supervised learning:* Supervised learning refers to an approach where the cognitive computer is trained by humans, using sample data, to detect or match patterns in a data

set. The training data set for a supervised learning classifier algorithm will include examples of patterns or question-answer pairs which the system will encounter. The supervised learning algorithm receives a set of inputs with the corresponding correct outputs, and learns and improves its performance by comparing its actual output with the correct outputs. Supervised learning is used where large data sets with known patterns are available, and regression or classification problems must be solved.

- *Reinforcement learning*: Reinforcement learning is a special type of supervised learning in which the cognitive computing system improves its “thought process” and refines future hypotheses based on feedback received on its performance. The system learns and discovers through trial and error which actions yields the greatest rewards and uses this as the basis for its next actions. Reinforcement learning is used when it is too complicated to create a representative training set, because the number of variables and series of tasks are too high.
- *Unsupervised learning*: Unsupervised learning refers to an approach that uses inferential statistical modelling algorithms to discover rather than detect patterns, associations or relationships in data. It learns through experience by identifying new patterns and not by matching patterns which it learned through human training. Unsupervised learning is used when representative relationships or question-answer pairs are not available to train the cognitive system. The lack of availability may be due to the complexity of the data, a substantial amount of variables, or when the data structure is unspecified. Therefore the objective is to explore the domain instead of detecting something known (Hurwitz *et al.*, 2015; SAS Institute, n.d.).

The use of the different classes of machine learning is not exclusive. A hybrid approach, which includes both supervised and unsupervised learning, will be more effective in some domains. The unsupervised learning approach will be implemented to discover a new pattern or relationship. Based on this discovery, a training set will be designed for supervised learning purposes (Hurwitz *et al.*, 2015).

#### **4.5 Infrastructure**

A cognitive computing system requires an agile and flexible infrastructure to support a distributed environment. The following data storage, processing and management solutions are available:

- **Distributed file system and distributed processing framework:** Distributes data files and computation over large server clusters. The individual servers work independently, while data is consistent throughout the decentralised environment (Miller, 2012). This provides quick access to large data sets, as well as the opportunity for scalability by increasing the number of servers (Chan, 2013). Cognitive computing systems require the management of various types and large quantities of data. In order to enable this, the system requires infrastructure that facilitates flexibility and scalability. The best way to achieve this is through distributed storage and management through the cloud (Hurwitz *et al.*, 2015).
- **Parallelism:** In parallel processing tasks can be performed simultaneously across multiple computers (Oracle, 1997), and therefore it is essential to support high volume and high speed data (Chan, 2013). The hypothesis generation and scoring in the cognitive system requires a software that supports parallel generation and scoring of multiple hypotheses. Each independent hypothesis must also be performed on a separate hardware thread or core in order to enable the corpus to scale up as the number of hypotheses increases. Therefore a hardware architecture is needed that supports seamless expansion through additional processors (Hurwitz *et al.*, 2015).
- **NoSQL (Not Only SQL) databases:** NoSQL databases use a non-relational data model that supports flexible schemas (structures), horizontal scalability, unstructured data, distributed processing and storage (Dix, 2014; Yuhanna, 2013). NoSQL structures, include:
  - *Key value pairs:* an identifier (Key) is matched with an associated data set (Value) and access is provided through the key. A Key-value database provides quick access to distributed data (Hurwitz *et al.*, 2015; Chen & Zhang, 2014; Zikopoulos *et al.*, 2015; DigitalOcean, 2014; Yuhanna, 2013).
  - *Columnar database:* a collection of one or more key value pairs combined with a record that consists of one or more columns containing information (Hurwitz *et al.*, 2015; Zikopoulos *et al.*, 2015).
  - *Document databases:* a collection of one or more key value pairs combined with a document. The structure of the document is flexible (for example, text documents, web pages, complete books) and can evolve rapidly (Hurwitz *et al.*, 2015; Zikopoulos *et al.*, 2015; DigitalOcean, 2014; Yuhanna, 2013).
  - *Graph databases:* use a graph (tree-like structure) with nodes (things), edges (relationships) and properties (key values). Graph databases create speed access to

connected data containing complex relationships (Hurwitz *et al.*, 2015; DigitalOcean, 2014; Yuhanna, 2013).

These simple structures enable horizontal scalability and flexibility (Dix, 2014; Dijcks, 2011). NoSQL allows for less strict rules regarding a database schema (schema-free), which supports the modification of the structure of data without rewriting the structure (Chen & Zhang, 2014). This is referred to as a “schema-on-read” approach (Markham, Kowolenko & Michaelis, 2015).

- **SQL database:** SQL databases are based on a relational model that organises data into interrelated tables (also known as relations) of rows and columns (Tech Target, n.d.). The columns contain data categories, and the rows a unique instance of data relating to the category (Tech Target, n.d.). The unique instance is assigned a unique key which is used to link rows in different tables. A structured query language (SQL) is used in a relational database to enable efficient interaction with data tables for both interactive queries and gathering of data (Tech Target, n.d.). SQL also allows for a broad set of questions to be asked to a single database (Dix, 2014). A traditional relational model enforces a set of rules to address scalability by reducing the information in a data set. The set of rules restricts the type of data and data structures leading to the loss of valuable information, relationships and patterns.
- **Stream processing:** Cognitive computing incorporates all types of data, including a new type of data known as streaming data. Streaming data (data in motion) is a continuous sequence of data that moves at a fast speed and is often generated by sensors (Hurwitz *et al.*, 2015; Chen & Zhang, 2014; IBM Corporation, 2014c). This high-velocity, high-volume data requires real-time processing due to the fact that there is not sufficient time available to store it before it changes. Real-time platforms include, Hadoop 2, SQLstream, Storm and StreamCloud (Chen & Zhang, 2014; IBM Corporation, 2014c).

## 4.6 Enabling technologies

Three essential enabling technologies that support the cognitive computing system were identified: (i) a standards-based, open-source software framework (Hadoop), (ii) cloud technology and (iii) big data.

### 4.6.1 Hadoop

Hadoop is a standards-based, open-source software framework which enables distributed processing and storage of large data sets, as well as massive parallel processing (IBM

Corporation, 2014a; Schneider, 2012). Unstructured data, such as text documents, raw data, social media data and sensor data are managed efficiently due to the fact that Hadoop uses distributed computing techniques. As a result, Hadoop is critical to the development of corpus for cognitive computing systems (Hurwitz *et al.*, 2015). Hadoop consists of two key elements: (i) a distributed, fault-tolerant file system for storing large files (Hadoop Distributed File System), and (ii) a distributed parallel processing framework (MapReduce) (IBM Corporation, 2014a).

- **Hadoop Distributed File System (HDFS):** A data storage cluster, both highly reliable and low in cost, used to make it easy to manage related files across different machines (Hurwitz *et al.*, 2015).
- **MapReduce engine:** MapReduce parallel programming model provides the capability to break down elements of a task into small pieces and process them in parallel, thereby facilitating the distributed processing of analytical algorithms across a large number of systems (Hurwitz *et al.*, 2015; Vaughan & Loshin, 2014; Schneider, 2012). After the distributed processing is complete, all the elements are combined to either produce a result or for additional processing (Hurwitz *et al.*, 2015; IBM Corporation, 2014a).

Utilising Hadoop in a cognitive computing system provides the following benefits (Hurwitz *et al.*, 2015; IBM Corporation, 2014a; Chan, 2013; Schneider, 2012):

- **High volume:** HDFS is well-suited to support large data volumes, high-velocity data and processing of large sequential operations.
- **Increased processing speed:** In HDFS, data is written once and then read many times thereafter, instead of multiple read-writes which occur in other file systems. This speeds up the processing of large data volumes which facilitates faster problem solving.
- **Increased transformation speed:** Hadoop can quickly transform massive amounts of non-traditional data from unstructured data to structured data in order for the cognitive system to search for patterns and find answers.
- **Scalability and Flexibility:** Hadoop is useful in cognitive computing because it is easy to dynamically scale and make changes quickly.
- **Fault-tolerance:** MapReduce offers fault-tolerant distributed processing across Hadoop clusters.

Hadoop 1 combines HDFS with the MapReduce, and therefore has a batch oriented format which exclude real-time processing (stream processing). In Hadoop 2, Yet Another Resource

Negotiator (YARN) has been added. YARN is a rebuilt cluster resource manager which splits the resource management and job scheduling capabilities, controlled by MapReduce. Therefore, MapReduce becomes a processing engine that can sit on top of YARN in Hadoop clusters. YARN facilitates other programming frameworks and new types of applications, and as such real-time processing (Vaughan & Loshin, 2014).

#### 4.6.2 Cloud technologies

Cognitive computing systems require an environment that integrates various data sources, software, hardware and networks. Therefore it necessitates the use of cloud computing services, as well as distributed architectures (Hurwitz *et al.*, 2015). Cloud computing is an information and communication technology, which provides shared computing resources to organisation as a service via the internet. These services include software as a service, infrastructure as a service (Amazon EC2), and platform as a service (Google AppEngine and Microsoft Azure) and has flexible specification of details such as number of processors and servers; memory size and storage; operating system, applications and networks (Chen & Zhang, 2014). There are various deployment models of cloud services. According to Hurwitz *et al.* (2015), there are four cloud computing models:

- **Public clouds:** are utilised by multiple users (enterprises), and owned and managed by external third-party service providers.
- **Private clouds:** are utilised exclusively by one user and managed within the company's private data centre.
- **Managed service providers:** are utilised by a specific user or multiple associated users with shared interests and managed by external third-party service providers.
- **Hybrid clouds:** are utilised by one user and offers the user the ability to connect services across public cloud, private cloud and managed services.

Cloud services are provided in the form of three delivery models (Hurwitz *et al.*, 2015):

- **Software as a Service (SaaS)** is a model that provides applications to users on a public cloud service.
- **Platform as a Service (PaaS)** is a model that provides a full infrastructure package (platform), including the databases, middleware and development tools, required to design and deploy applications on a public or private cloud.

- **Infrastructure as a Service (IaaS)** is a model that provides computing, storage and network services directly on a computer or virtually.

Utilising Cloud computing in a cognitive computing system provides the following benefits (Hurwitz *et al.*, 2015; Chen & Zhang, 2014):

- **Affordability:** Cloud computing allows self-service provisioning making it cost effective.
- **Flexibility and scalability:** Cloud computing provides shared computing resources (for example, processors, operating system, *etc.*) which can be utilised as the services and resources are required. It also allows the enterprise to scale services to address high data volume and variable data types.
- **Distributed processing:** Cloud design is built on distributed computing models which supports distributed processing which is a requirement for cognitive computing.

#### 4.6.3 Big data

According to Gartner (2015), big data refers to high-volume, high-velocity and high-variety information assets which requires cost-effective and innovative forms of processing in order to enhance insight, decision making, and process automation. Cognitive computing provides the computing power required to gain value from big data. A cognitive computing system requires high-volume and high-variety data to support hypotheses, and answer questions (Kelly III & Hamm, 2013). Big data has the following characteristics (Heller, & Piziak, 2015; Hurwitz *et al.*, 2015; Hagen, Khan, Evans, Thota, Wall & Seshadri, 2014; Jewell *et al.*, 2014; Kitchin, 2014; Kshetri, 2014):

- **Volume:** Volume refers to the quantity of data available. It is estimated that about 2.5 quintillion bytes have been created daily in the last couple of years and it is expected that the volume of business data worldwide will double every 1.2 years. The volume of data increased drastically due to the introduction of new data sources such as social media, sensors, machine to machine data and the internet of things.
- **Velocity:** The velocity of data is the speed at which data is received and processed and perhaps acted upon. Data can be ingested and processed in periodic batches, or in real time, depending on the time-sensitivity of the data.
- **Variety:** Variety refers to the format of data, which includes structured (numeric transaction data), unstructured (unstructured text, audio, and video) and semi-structured data. The variety of data also increased due to the introduction of new data sources.

- **Veracity:** Veracity is a requirement for data accuracy and integrity.
- **Variability:** Data variability refers to the flow of data which can vary greatly with periodic peaks and troughs. The peaks and troughs relates to social media trends, daily, seasonal and event-triggered peak data.
- **Value:** Data has intrinsic value which must be discovered. Cognitive computing systems are applied in order to find connection and insight in data in order to create value.

The benefits of big data in a cognitive computing environment are the following:

- **Value delivery:** Big data allows enterprises to develop a sound and insightful understanding of their business, which the enterprise can use to improve decisions, enhance productivity and create a strong competitive advantage (Dijcks, 2011).
- **Innovation:** Big data enables enterprises to enhance old products and services, develop new products and services, and create new business models (Manyika, Chui, Brown, Bughin, Dobbs, Roxburgh & Byers, 2011).

#### **4.7 Summary and conclusion**

This chapter identified and explained the core components of a cognitive computing system. These components, divided into the Data, Read, Resolve, Reason, Infrastructure and Enabling technologies phases, will be integrated with the detailed processes of COBIT in order to facilitate the identification of the significant risks that the enterprise is exposed to by employing a cognitive computing system.

## CHAPTER 5: RISKS PERTAINING TO THE IMPLEMENTATION OF A COGNITIVE COMPUTING SYSTEM

A cognitive computing system is a strategic platform that continuously interacts with its surrounding environment which includes humans, business processes and other information technology systems. Therefore it cannot be governed and managed in isolation. In Chapter 3, COBIT was selected as the most appropriate control framework to identify significant risks pertaining to the implementation of a cognitive computing systems. In order to ensure that all significant risks are identified, the detailed processes of COBIT were used to identify risks based on the understanding of the components of cognitive computing developed in Chapter 4. The risks were identified and rated as High (H); Medium (M) or Low (L). Annexure A presents the significant risks identified at a strategic level (with regards to inadequate governance and management) and at an operational or technological level.

### 5.1 Significant risks at a strategic level

Significant risks at a strategic level were identified using the detailed COBIT processes and divided into: inadequate governance and management of cognitive computing systems and inadequate human skills and resource management.

#### 5.1.1 Inadequate governance and management of cognitive computing systems.

Cognitive computing systems fundamentally change the way humans and systems interact. The success of this relationship depends on the trust the stakeholders (company and customers) have in the ability of the cognitive computing solution. In order to promote trust, a mature implementation of IT Governance principles, is required. The absence of an effective information governance framework and comprehensive governance strategies and policies in the cognitive computing environment will expose the enterprise to the following risks:

- **A lack of board involvement and inadequate information governance policies:** The adoption of cognitive computing will require the advancement of existing policies (for example, data security and privacy), as well as the development of entirely new policies in response to advances in cognitive capabilities (for example, corpus management and traceability of the decision-making process) (Sarkar & Zaharchuk, 2015). Often when new technologies are introduced there is a lack of board involvement, implementation of appropriate policies and documentation of the policies (Zikopoulos *et al.*, 2015). This may

prohibit the efficient deployment of the cognitive computing system and will hamper an ethical IT governance culture and awareness within the organisation (IODSA, 2009).

- **The cognitive computing strategy does not align with the objectives of the entity:** The organisation may not achieve alignment between the cognitive computing strategy and processes, and the enterprise's strategies and processes. If the organisation only achieves strategic alignment, it may not be enough to ensure effective IT governance. It also requires alignment between business objectives and IT objectives at an operational level (daily activities) (Rubino & Vitolla, 2014). If strategic and operational alignment is not achieved, IT governance will be ineffective.
- **Poor stewardship and ownership for the implementation of IT governance structures and processes:** Management should be responsible and accountable for the implementation of structures and processes to enable the governance of the cognitive computing system. However, the board and senior management lack sufficient confidence and understanding of cognitive computing strategies and processes. Consequently, they hesitate to employ the necessary structures and processes (Court, 2015). A lack of stewardship and ownership over cognitive computing governance strategies exists, due to the fact that few companies have chief data officers directly responsible for issues associated with it. This may lead to inadequate management of the cognitive management system as well as miscommunication between stakeholders.
- **The investment in cognitive computing exceeds the return in the investment:** According to Court (2015), early investment in data analytics did not yield significant return as a result of efforts being open-ended without sufficient focus and monitoring. In a data intensive environment such as a cognitive computing system, the board may not place sufficient value on their data assets. Therefore they may not create value statements around data sources which stipulate how the data ingested and created in the cognitive system must be used, trusted, curated, and invested in (Zikopoulos *et al.*, 2015).
- **Inadequate risk assessment and management processes:** The risk management system may not identify all the risks the enterprise is exposed to due to the deployment of the cognitive computing system or address the risks effectively to reduced risk to an acceptable risk tolerance level (ISACA, 2016a). For instance: governance requirements and IT laws which place usage restrictions on ingested and generated data in the corpus, may not be adhered to. This may lead to lost trust in the enterprise and as a result financial losses. Other examples are copyright images which may require a licence, social media data which may

violate privacy rules and sensitive data regarding competitive best practices (Hurwitz *et al.*, 2015; Hodges & Creese, 2013). The enterprise may also not have adequate business resilience arrangements in place for disaster recovery (IODSA, 2009).

- **Ineffective management of information assets and resources:** The cognitive computing system may not be managed or used effectively and as a result potential business opportunities may be overlooked and resources misallocated. Resources may also not be sufficient (Suer & Nolan, 2015). In addition the enterprises may not have sufficient resources available to meet the cognitive computing objective (Suer & Nolan, 2015).
- **Insufficient assistance from the risk committee:** The risk committee may not ensure that IT risks are adequately addressed and controls are in place and effective in addressing IT risks, which will result in risk exposure which exceeds the organisation's risk tolerance levels (IODSA, 2009).

### 5.1.2 Inadequate human skills and resource management

A key challenge for the advancement of cognitive computing capabilities and implementation of cognitive systems is the unique skill sets required. Some of the most essential skills include those of machine learning experts, natural language processing scientists and data scientist. However, a significant concern for companies is the scarcity of these technical talents (Sarkar & Zaharchuk, 2015; Kitchin, 2014; McAfee & Brynjolfsson, 2012).

McKinsey predicts, based on sources such as the US Bureau of Labor Statistics; US Census; Dun & Bradstreet; company interviews and McKinsey Global Institute analysis, that the United States alone will face a shortage of 50 to 60 percent of people with deep analytical talent by 2018 (Manyika *et al.*, 2011). According to ATKearney (Hagen *et al.*, 2014), 80% of data scientist jobs in the United States are already unoccupied and the small group of people with these unique skills are predominantly utilised by industry heavyweights (Hagen *et al.*, 2014).

Education is currently lacking in the preparation of such skilled scientists. According to Dr Bruce Porter, Professor and Chair of Computer Science at The University of Texas in Austin, the focus of current Computer Science curricula needs to be modified. Focus must shift from inward looking programmability to outward looking application, with specific focus on Artificial Intelligence and Cognitive Computing. Universities are now starting to create new programs, however the technology is developing faster than the curricula is changing.

Cognitive computing development and management may be outsourced to external service providers or vendors due to a shortage of internal skills and resources. The risk exists that the service providers and suppliers may not provide the required skill set, resources or services and that the cognitive computing requirements are not met.

If the design and development of the cognitive system is successful and effective, the self-service nature of machine learning algorithms will reduce the need for further assistance from the scientists (Harper, 2015). However, people still remain a key component in building the system (Kitchin, 2014).

## 5.2 Significant risks at an operational or technological level

Significant risks at an operational level or technological were identified using the detailed COBIT processes and can be divided into three groups:

1. Risks that affect the objective of the cognitive computing system. These include cost, privacy, security and ownership.
2. Risks that affect the ability of the cognitive computing system to function effectively. These include scalability, integration, interoperability and veracity.
3. Risks that affect both the objective of the cognitive computing system, as well as the ability of the cognitive computing system to function effectively. These include cognitive computing life cycle risks.

### 5.2.1 Cost

Deploying a cognitive computing system may lead to significant cost implications for organisations. According to McKinsey & Company (2015) senior management sees large investments in analytical technologies as a major challenge. The cost implications involved in the implementation of a cognitive computing system include (Hurwitz *et al.*, 2015; Chen & Zhang, 2014; Géczy, 2014; IBM Corporation, 2014b; Jewell *et al.*, 2014):

- **Development cost:** The initial investment in a cognitive computing system is extensive since the majority of cognitive computing systems must still be built from scratch by vendors in collaboration with the user.
- **Infrastructure and management cost:** A cognitive computing system requires a significant investment in infrastructure additions, modifications and upgrades. The costs involved include the cost of scalable data storage, processing capacity and transmission capabilities. The changes in infrastructure and the implementation of the new technology will also require investment in new management approaches.

- **Human skills cost:** Designing and developing a cognitive computing system; specifically selecting, accessing, acquiring, and preparing data for the corpus; is time-intensive and requires the involvement of domain experts and end users. The investment in experts, personnel changes and retraining may be substantial.
- **Security and privacy infrastructure and monitoring cost:** A cognitive computing system requires an effective security infrastructure to address security and privacy risks. Continuous monitoring of data access and protection against data breaches will involve significant investment.

### 5.2.2 Privacy

Privacy refers to the right of individuals to control or influence what personal and sensitive information related to them may be accessed and by whom and to whom that information may be disclosed (Kitchin, 2014). The deployment of cognitive computing introduces the following significant privacy risks:

- **Re-identification risk:** Re-identification risk occurs when data is aggregated and as a result of the aggregation process semi-anonymous information or personally non-identifiable information become non-anonymous or identifiable. In a cognitive computing system, data from various data sources are combined and new connections are discovered which increases the risks of the re-identification of individuals. In addition, a large component of the data sources ingested consists of unstructured data which is more probable to contain sensitive data, personally identifiable information (PII) and intellectual property (IP) (ISACA, 2014; Kitchin, 2014; Kshetri, 2014; Jensen, 2013; Manyika *et al.*, 2011).
- **Transparency risk:** Transparency risk occurs when personal data is used, processed or disclosed without consent from the affected individual or for purposes they do not expect or understand (Kshetri, 2014). A cognitive computing system introduces transparency risk because:
  - data aggregated from various data sources and generated by the cognitive computing system may not have enough direct or indirect identifiers to trace the data back to the individual in order to obtain consent for the use of the personal data (Nelson, 2015; Kshetri, 2014);
  - consent must be obtained before processing. Given the numerous types of algorithms applied in cognitive computing, informed consent entails that the individual must be

provided with an explanation of all of these algorithms in order to understand what happens to their personal data. This is a significant challenge due to the complex nature of cognitive computing algorithms (Jensen, 2013);

- permission may be obtained for the use of personal data for a specific purpose (original or primary). However after the information was collected, analysed and new data generated from the personal data, it might be used for a different (secondary) purpose without obtaining additional consent (de Bruyn, 2014; Kitchin, 2014). The data collected may also be stored in the corpus and reused indefinitely, increasing the risk that the data may be used for different purposes than originally intended; and
- the data sources ingested into the cognitive computing system may include clickstream data (consisting of the route taken by a user when they navigate through an internet site), which can be manipulated by tracking tools to build a detailed database of personal profiles without notifying the data subject or specifying the use (Hurwitz *et al.*, 2015; Kshetri, 2014).
- **Violation of individual participation rights:** Individuals are entitled to refuse usage, revoke consent and request corrections to their personal data and these rights may be disregarded (de Bruyn, 2014; Ekbia, Mattioli, Kouper, Arave, Ghazinejad, Bowman, Suri, Tsou, Weingart & Sugimoto, 2014; Jensen, 2013). In a cognitive computing system, the corpus is constantly updated with newly generated information. The risk exists that individuals may not be able to correct personal information, enforce the deletion of data or revoke their consent with regards to their personal data included in the newly generated information in the corpus (Hurwitz *et al.*, 2015). The risk increases when service providers are used for transferring, storing and processing purposes. This limits the access and control that the enterprise has over personal information, which in turn limits the ability of the individual to access and correct of his/her personal data (Hurwitz *et al.*, 2015; Jangara & Bezuidenhout, 2015).
- **Information quality risk:** Large quantities of data obtained from a variety of sources and spread across various systems are not always precise or faithful. Specifically, data collected from social media may be inaccurate, manipulated, falsified and often outdated, which will result in incorrect correlations and statements (Kitchin, 2014; Jensen, 2013; PWC, 2011).
- **Unauthorised access:** This refers to unauthorised access to personal data by an entity without authentication. Unauthorised access to sensitive personal data may lead to reputational damage, legal liability and ethical harm (Hurwitz *et al.*, 2015; ISACA, 2014;

Kitchin, 2014). A cognitive computing system is vulnerable to unauthorised access due to the (Hurwitz *et al.*, 2015; Jangara & Bezuidenhout, 2015; Kshetri, 2014):

- centralised aggregation of high volume of data and information in the corpus, exposing the entirety of the data to unauthorised access rather than just a subset of the data (known as amplified technical impact);
  - high variety and variability of data from multiple sources increasing the risk that unauthorised access may go undetected and / or adequate responses being delayed;
  - high volumes, variety and variability of data ingested into the cognitive computing system, increasing the risk of attracting the attention of cybercriminals;
  - extensive nature of data and the speed that data sources are deployed at, which means that the security infrastructure supporting the cognitive computing system might not be able to protect sensitive data in motion or distributed data against unauthorised access. Data from real-time devices such as sensors and medical devices are specifically exposed to a greater risk; and
  - service providers transferring information inside or outside South Africa, exposing enterprises to additional risks that arise when they surrender control over the data and the right to respond directly to unauthorised access events which may affect the personal information residing at the services provider.
- **Compliance risk:** Cognitive computing exists in a data rich environment which is becoming highly regulated, particularly personally identifiable information. The manner in which personal data must be secured may differ in various industries, markets, and countries. As such, a risk of non-compliance exists (Hurwitz *et al.*, 2015). The company is responsible to ensure that data and meta-data imported and generated is in compliance with the applicable regulations, and remains compliant on an ongoing basis (Hurwitz *et al.*, 2015). In addition, when international cloud solutions are used for storing and processing purposes, jurisdictional conflicts may arise if data centres are situated across geopolitical boundaries in locations with inadequate or incompatible data protection and privacy laws. This will increase the risk of non-compliance with the POPI Act (Jangara & Bezuidenhout, 2015; Hurwitz *et al.*, 2015).
  - **Gaps in privacy management and policies:** The management of personal data and enforcement of privacy policies are a challenge for an enterprise using a cognitive computing system due to the following factors (Hurwitz *et al.*, 2015; Jangara & Bezuidenhout, 2015; Salido, 2010):

- inadequate notice of data collection, use, disclosure and restoration policies as well as a lack of controls to enforce the policies;
- insufficient documentation of privacy plans, policies, controls, and system configurations;
- a lack of configuration controls;
- lack of a breach notification plan in order to inform individual of unauthorised access to personal data; and
- service providers are often used to support the cognitive computing system however, most service providers do not have adequate privacy policies, protection procedures and controls in place. The enterprise will remain responsible for protecting sensitive data and the POPI Act stipulates that loss of privacy due to attacks and unauthorised access may result in severe penalties and jail time.

### 5.2.3 Security

Information security ensures that data is protected against disclosure to unauthorised users (confidentiality), unauthorised modification (integrity) and inaccessibility when required (availability) (ISACA, 2016a). In a cognitive computing system the ability to secure content and results are essential. The organisation must trust the system and its hypothesis and as such the information cannot be compromised. Therefore, security has to be incorporated at every level of the cognitive computing environment (Hurwitz *et al.*, 2015).

Cognitive computing environments are exposed to various security challenges such as:

- **Unauthorised access:** This refers to unauthorised access to sensitive and confidential data by an entity without authentication. Refer to Unauthorised access under section 5.2.2 for detailed discussion. In addition, the extensive number of users of a cognitive computing system complicates the decision of which users should be granted access to the different components within the cognitive computing system (Paryasto, Alamsyah & Kuspriyanto, 2014).
- **Intentional security breaches:** The cognitive computing system is exposed to malicious attacks such as hacking, malware, viruses, phishing and denial of service. Due to the high variety of data from multiple sources, it becomes more challenging to detect security breaches and respond appropriately to these invasions (ISACA, 2016b; Ballard *et al.*, 2014; Kshetri, 2014).

- **Distribution risk:** Distributed infrastructures used to support the volume and velocity of data in a cognitive computing system will increase the following security risks:
  - leakage of confidential data due to malfunctioning computing nodes (ISO, 2014);
  - eavesdropping on confidential data by adding rogue nodes in the distributed system (ISO, 2014);
  - interference, modification or destruction of a significant fraction of the system or the entire system by a partial infrastructure breach due to high levels of connectivity and dependency (ISO, 2014);
  - challenges in establishing access control across the distributed environments, as well as physical security of data infrastructure, data networks, data applications, and data (Hurwitz *et al.*, 2015; Chen & Zhang, 2014; IBM Corporation, 2014b); and
  - operational inefficiency due to the fact that implementing several security controls across a diverse enterprise IT infrastructure may be complex, time-consuming and cost inefficient (CA Technologies, 2015).
- **Non-Compliance risk:** Enterprises are must comply with numerous data security regulatory requirements from governments and industry organisations. These regulations contain specific mandates around management, control and monitoring of financial, personal, intellectual property and sensitive data. There is a risk that the security infrastructure of the cognitive computing system will not identify all the regulatory requirements and that the enterprise will consequently fail to comply (CA Technologies, 2015; IBM Corporation, 2014b). Refer to Non-compliance risk under section 5.2.2, for additional detailed discussion.
- **Insider breaches:** The risk of lost, stolen or unauthorised sharing of privileged credentials by privileged users and administrative accounts (CA Technologies, 2015).
- **Insecure computation:** An insecure program which has access to confidential data in the cognitive system, can corrupt the data leading to incorrect results as well as denial of service to the cognitive computing system (Paryasto *et al.*, 2014).
- **Validation risk:** The acquisition of high volume, variety and velocity data for the cognitive computing system makes it difficult to validate and ensure data integrity (Paryasto *et al.*, 2014).
- **Database risk:** Security features embedded in traditional databases are not always present in new databases such as NoSQL databases. For example, NoSQL databases do not make use of encryption for data at rest (Smitha, Suma & Sunitha, 2015).

- **Outsourcing risks:** Due to the variability of data, the enterprise may not have the capacity to collect and store data securely during peak data traffic. Inevitably the organisation will require outsourcing services from service providers, thereby limiting the enterprise's control over confidential information (Jangara & Bezuidenhout, 2015; Hurwitz *et al.*, 2015; Kshetri, 2014).
- **Gaps in security management and policies:** The management of sensitive data and enforcement of security policies are a challenge for an enterprise using a cognitive computing system due to the following factors (CA Technologies, 2015; Jangara & Bezuidenhout, 2015; Hurwitz *et al.*, 2015; Salido, 2010):
  - inadequate access control, data sharing, data quality and data integration policies, as well as a lack of controls to enforce the policies;
  - insufficient documentation of security plans, policies, controls, and system configurations; and
  - service providers are often used to support the cognitive computing system, however most service providers do not have adequate security policies, protection procedures and controls in place. This exposes the enterprise to loss of proprietary secrets and confidential information through unauthorised access and security breaches.

#### 5.2.4 Ownership

Ownership can be defined as a right that associates data with one or more entities, who own and control what can be done with the data (ISO, 2014). In a cognitive computing environment, the new knowledge produced by the system creates uncertainty about data ownership and intellectual property rights (Hurwitz *et al.*, 2015). Challenges include who owns a piece of data or controls it, what are the rights attached to the data and whether the data can be sold or shared (Jagadish, Gehrke, Labrinidis, Papakonstantinou, Patel, Ramakrishnan & Shahabi, 2014; Manyika *et al.*, 2011). This in the end also threatens the privacy of the individuals who shared their information (Hodges & Creese, 2013).

#### 5.2.5 Scalability

Scalability refers to the ability of a system to increase or decrease its performance and related cost in response to changes in data, application and system processing demands (Gartner, 2015). The rapid growth in data volume and the increase in data velocity are forcing cognitive computing to evolve at an extreme pace. The risks exist that variety and scalability capabilities of the cognitive computing system will not advance rapidly enough to cope with this

information supply (Sarkar & Zaharchuk, 2015). A critical issue is whether or not the algorithms used in the cognitive system are able to scale as the data volumes and aggregation increase by orders of magnitude. Algorithms contain an inherent limitation known as a “*knee*”. Kaisler (2013) defines this as the point at which the algorithm’s performance ceases to increase with increasing computational resources and a new algorithm is needed (Kaisler, Armour, Espinosa & Money, 2013). In context with cognitive computing there are some scale machine learning algorithms, however natural language processing specifically, still face the scalability problems (Chen & Zhang, 2014).

In addition, the data infrastructure (storage capabilities, data transmission and computing power) supporting the cognitive computing system contain inherent limitations that restrict performance levels of the cognitive computing system. The following limitations exist:

- **Storage capacity:** The volume of data integrated into the cognitive computing system poses a great challenge for information technology structures and their capacity to store information. (Chen & Zhang, 2014; ISO, 2014; Jewell, 2014). The performance level of the cognitive computing system will be limited if data storage is not scalable and does not provide the required data density on disks.
- **Data access:** The risk is that data cannot be accessed easily and promptly for further processing and analysis. According to Chen & Zhang (2014), current storage devices, architecture and technologies do not provide the same high performance for both sequential and random Input/Output (I/O) simultaneously. They indicate that new storage technologies, such as solid-state drive and phase-change memory will alleviate the access challenges, but will not eliminate them.
- **Data management:** Database technologies are used in a cognitive system for effective management and processing of data. Traditionally, databases entailed software systems running on specialised single-rack high-performance hardware. These traditional database approaches have been unable to match the rapid growth of data and management demands (Chen & Zhang, 2014; Géczy, 2014).
- **Data processing:** A fundamental concern is whether the processing power of the infrastructure can keep up with the processing demands of scaling data volume, or if it can be suitably expanded to meet the demands. Linear scalability will be required, which entails infrastructures that deliver linear increases in processing throughput with linear increases in software and hardware resources (IBM Corporation, 2014a). However, the problem is

that the data volumes are growing faster than the computational power of processing (CPU speeds) (Chen & Zhang, 2014; Géczy, 2014).

- **Real-time / Stream processing:** A big challenge for stream processing is to provide a timely response when large volume of data must be processed (Chen & Zhang, 2014).
- **Data transmission:** The transmission of large volumes of data is creating bottlenecks in communication networks (cloud and distributed systems) due to the fact that the network bandwidth is insufficient (Chen & Zhang, 2014; Géczy, 2014). This includes transmission from data sources, as well as between components.

The solution for many of the challenges listed above is the deployment of both SQL and NoSQL database systems; cloud solutions; and parallel and distributed processing (Chen & Zhang, 2014; Géczy, 2014). However, none of the solutions are sufficient to address all of the challenges and all of the solutions introduces new risks such as security risk, integration risk, and privacy risk.

### 5.2.6 Integration

Integration risk encompasses both data integration, and system and infrastructure integration.

- **Data integration:** IBM (n.d) defines data integration as technical and business processes which combine data from a variety of sources into valuable information. In the context of a cognitive computing environment, data integration can be narrowed down to utilising software to link data in order to create reliable, trustworthy and consistent information for the corpus. The challenge is to combine data which is highly heterogeneous, unstructured and variable in quality to obtain a common representation (Knoblock & Szekely, 2015; Kitchin, 2014).

Diverse data sources produce data with various formats, structures, timescales and semantics which may be incompatible (ISO, 2014; Kaisler *et al.*, 2013). In order to combine data, the cognitive system must employ both traditional data integration mechanisms such as extraction, transformation, and load (ETL), as well as new integration mechanisms such as extract, load, transform (ELT) (Heller & Piziak, 2015; Hagen *et al.*, 2014; Jewell, 2014).

Cognitive computing has the ability to integrate data from various heterogeneous sources and generate answers from them (Noor, 2015). However, the risk currently lies in integrating sensor data, specifically static and moving images; languages, spoken or

written; and music as sound and music in its written form (ISO, 2014; Wolff, 2014). Sensing technologies and scientists involved with the development thereof, still tend to focus on each sensory field in isolation, which creates a major challenge in creating a meaningful common representation of sensor data (Wolff, 2014; Kelly III & Hamm, 2013). The cognitive computing system, and specifically the corpus, needs access to a variety of frequently updated data sources to keep current about the domain it operates in and provide accurate results. As a result, the integration of data will continue indefinitely and there must be a fusion of the newly acquired data with the original data in the corpus. The risks remain the same as in the data access layer (Hurwitz *et al.*, 2015).

- **Systems and infrastructure integration:** Gartner (2015) defines system integration as creating a multifaceted information system by integrating a customised architecture with new or existing hardware, software, and communications. The infrastructure supporting a cognitive computing system faces integration challenges resulting from the variety, uncertainty, and complexity of the data environment (Chen, Li & Wang, 2015). The heterogeneous nature of data in a cognitive system requires diverse storage capacity, varied processing power, different management mechanisms and network technologies. If these components (data, software, hardware and technologies) do not integrate seamlessly, the risk exists that the system will not deliver the desired results or function as intended (Géczy, 2014).

The cloud is often used as a solution for these diverse requirements in a cognitive computing system. However, the availability of data in the cloud provides both potential and complexities. The complexities include the need to provide links and techniques for integrating cloud data sources (Hurwitz *et al.*, 2015).

### 5.2.7 Interoperability

Interoperability refers to the ability of different systems or components within an infrastructure to exchange and use information or functionality by adhering to common standards (Janssen, Estevez & Janowski, 2014; Nielsen, 2013). ETSI (2008) and Janssen *et al.* (2014) identified the four different levels of interoperability as technical, syntactical, semantics and pragmatic interoperability. The two main risks associated with cognitive computing are:

- **Technical interoperability:** Technical interoperability ensures that the hardware and software components, systems and networks within the cognitive system infrastructure can

communicate. This kind of interoperability is often established through standardised communication protocols. If a standardised communication protocol is absent, there is a risk of incompatibility which will create a barrier for machine-to-machine communication (Nielsen, 2013).

- **Semantics interoperability:** Semantic interoperability ensures that the cognitive computing system interprets data in the same way. The misinterpretation of data creates the risk that correlation between data may not be recognised or incorrect correlation may be made prohibiting the discovery of new patterns, valuable analysis and decision making. Ontologies and taxonomies enable semantic interoperability (Haav & Kungas, 2013).

### 5.2.8 Veracity (Quality)

Data veracity refers to the trustworthiness, applicability, accuracy, consistency, bias and other quality properties in data (ISO, 2014; Kitchin, 2014). The risk exists that low levels of data quality in individual databases and resources will result in lower levels of data quality within the cognitive computing system (Wigan & Clarke, 2013). Consequently, the degree of confidence and trust that can be placed in the analyses and hypotheses rendered from the data is effected (Kitchin, 2014).

Kitchin (2014) established that the quality and veracity of data within a system may be weakened due to: (i) instrument error (for example, sensors); (ii) working parameters of applied technologies changing the nature of the data; (iii) faked data from false accounts and; (iv) hacked accounts. A cognitive computing environment is data rich and may be influenced by all these challenges.

Within a cognitive computing system, data must be cleaned and corrected in the data access layer in order to standardise data. If data is not standardised before it is ingested in the corpus, it may lead to misinterpretation or misapplication of the data, as well as inconsistencies in the decision making process (Trites, 2013).

Bias in training, with regard to supervised learning, is a significant risk. Specifically, unstructured data and domains with no standards available to understand the domain data requires experts to make judgements based on their own experiences. Their judgements may be biased because most individual will not be exposed to all possible interpretations available. Bias may result in tainted analyses and hypotheses with weakened validity (Hurwitz *et al.*, 2015; Kitchin, 2014).

Errors may also occur in the hypothesis generation, as well as the recommendations provided by the cognitive computing system, due to insufficient data and cognitive computing models that cannot capture correlations and nuances between similar data sources.

Even though there are risks associated with dirty data there is also a significant risk that valuable insights may be missed due to data cleaning in a cognitive computing system. A core element of a cognitive computing system remains the identification of abnormalities and identifying new patterns. For example, when a cognitive computing system is used for detecting fraud or for security purposes, data must be ingested in its original state to enable the cognitive computing system to deliver the required results (Hurwitz *et al.*, 2015).

### 5.2.9 Cognitive computing life cycle risks

In terms of COBIT 5, the planning phase includes the identification of objectives, information architecture, standards and definitions. The design phase involves the implementation of the plan obtained from the planning phase, and the build phase entails the creation of the system. Lastly the use phase involves how the system is operated, specifically how information is stored, shared and used (Suer & Nolan, 2015). Significant risks must be identified in every life cycle phase.

Enterprises deploying cognitive computing systems may experience the following planning, design and building challenges:

- **Inadequate high-level cognitive computing road map:** If the cognitive computing road map does not address the objective of the cognitive computing system and user requirements; the cognitive computing components required to support the objective, and the development plan for the cognitive computing system, the cognitive computing system will be ineffective.
- **Ineffective cognitive computing component development because of problem with logic:**
  - The contents of the corpus limit the types of user questions (problems) that the cognitive computing system can solve. Therefore, if the corpus is too narrowly defined it will not be comprehensive enough and new insights will be missed.
  - The corpus must include internal and external data. If the corpus does not include the right combination of relevant data resources, it may not be able to deliver accurate responses.

- If data from the external sources are trimmed or cleaned before they are imported into the corpus, they will be excluded in the generation and scoring of hypotheses. This will limit discovery and create incorrect correlations.
- It is essential that the appropriate machine learning algorithms, as well as suitable techniques of analysis, are employed which are best suited for the specific domain and specific problem which must be solved. Experts must possess great skill to combine the best algorithms for the best results as well as a good understanding of the domain. Failing to do so may lead to mistakes in the interpretation the hypothesis.
- If a taxonomy or ontology is not available for the specific domain a new taxonomy or ontology must be developed. The main risk associated with the development of an ontology or taxonomy is inconsistencies. Inconsistent assumptions, beliefs and practices may be identified during the planning, design and development (building) phases. If these inconsistencies are not identified and addressed appropriately, the body of knowledge (corpus) cannot be trusted and will be ineffective.
- A cognitive computing system must be able to identify and request additional data from internal and external sources when that new data will enable the system to make better decisions. If the cognitive system is not properly trained by the right experts and does not have the correct combination of algorithms, the system will not be able to update and maintain the corpus as required and knowledge gaps might arise leading to weaker recommendations.
- **Inadequate change plan:** with regards to business processes; infrastructure, operating systems and networks; people, *etc.* (ISACA, 2012b)
- **Inadequate software development process which is not appropriate and not followed** (Hurwitz *et al.*, 2015; Kitchin, 2014):
  - The development procedures do not adhere to the enterprise development standards.
  - Third parties involved in the development do not adhere to contractual obligations and enterprise development standards.
  - The changes during the development process are not authorised and monitored.
  - The different stages of the development process are not controlled and monitored for effectiveness and performance.

Insufficient monitoring of the core cognitive computing components during the use/operate phase will leave additional requirements for the cognitive computing system unidentified, and as such will hamper continual improvement (ISACA, 2012b).

### 5.3 Summary and conclusion

The processes of COBIT 5 were used in Annexure A to identify all significant risks relating to the deployment of a cognitive computing system. A summary of the significant risks identified, at a strategic and operational or technological level, in relation to the components of a cognitive computing system (Chapter 4) which give rise to the risks, is documented in Table 2.

**Table 2: A risk matrix: linking cognitive computing components to the significant risks it generates**

	Information governance	Human skills & resources	Cost	Privacy	Security	Ownership	Scalability	Integration	Interoperability	Veracity	Life cycle risks
<b>DATA</b>											
Unstructured data	X	X		X	X	X		X		X	X
Semi-structured data	X	X		X	X	X		X		X	X
Structured data	X	X		X	X	X		X		X	X
<b>READ</b>											
Data access	X	X		X	X						X
Metadata	X	X		X	X						X
Feature extraction services	X	X					X				X
NLP	X	X					X				X
Deep learning	X	X					X				X
<b>RESOLVE</b>											
Corpus	X	X	X	X	X	X	X	X			X
Advanced analytics	X	X		X	X	X	X				X
<b>REASON</b>											
Hypothesis generation and scoring	X	X		X	X	X	X				X
Machine learning	X	X		X	X		X			X	X
<b>DATA INFRASTRUCTURE</b>											
Storage	X	X	X	X	X		X	X	X		X
Processing and management	X	X	X	X	X		X	X	X		X
<b>ENABLING TECHNOLOGIES</b>											
Hadoop	X	X		X	X				X		
Cloud computing	X			X	X				X		
Big data	X	X		X	X	X		X		X	X

The risks identified in this chapter must be addressed in order to avoid negative consequences for the enterprise. The objective of Chapter 6 is to identify mitigating control techniques which can be implemented to mitigate these risks to an acceptable level.

## CHAPTER 6: SAFEGUARDS AND CONTROLS IN A COGNITIVE COMPUTING SYSTEM

According to ISACA (2016a), internal controls are policies, processes and practices developed to provide reasonable assurance that undesirable events (risks) are prevented, detected and corrected. Enterprises implementing cognitive computing systems are exposed to significant risks. In order to mitigate these risks comprehensive controls must be implemented in accordance with COBIT 5 to govern and manage the cognitive computing system.

In this chapter, controls are formulated on both a governance and management level, and an operational or technological level.

### 6.1. Governance and management at a strategic level

At a strategic level, the enterprise should develop a cognitive computing governance strategy accompanied by a list of comprehensive policies to provide practical guidance for implementation, and a human resources strategy.

#### 6.1.1 Cognitive computing governance

According to ISACA (2016a), IT extends an enterprise's business strategies. To ensure that the cognitive computing strategies and plans align with the enterprise's business strategies, an effective governance framework should be established (ISACA, 2016a; Manyika *et al.*, 2011). The framework for the governance of the cognitive computing system should identify and engage with the enterprise's stakeholders, and document their requirements; direct the structures, processes and practices enabling the governance; and monitor, report and improve the effectiveness and performance of the governance (ISACA, 2012b). A governance framework should be developed and implemented, taking the following considerations into account (Cavoukian, Chibba, Williamson & Ferguson, 2015; Rubino & Vitolla, 2014; PWC, n.d.):

- **Responsibility for governance:** The board should establish a framework for the governance of the cognitive computing system to ensure that the enterprise's objectives are achieved. The enablers of governance, specifically policies and procedures, should be reviewed, monitored and improved by the board in order to facilitate effective cognitive computing governance.
- **Risk committee:** The risk committee must address cognitive computing risks through risk management, monitoring and assurance processes.

- **Risk management:** The cognitive computing governance system should include a risk management system which establishes processes for risk identification, risk assessment and risk response. This will allow management to prioritise risks, according to the likelihood of occurrence and potential impact, address the risks with effective mitigating controls and assign responsibility. All legal and regulatory requirements, pertaining to the cognitive computing system should be addressed by the risk management system.
- **Alignment:** The cognitive computing strategy must be integrated with the enterprise's strategic and business processes; and evaluated, directed, monitored and remediated in order to ensure alignment.
- **Delegation of responsibility:** The board must ensure that roles, responsibilities and accountability are established and communicated to all stakeholders, including management. This can include formal reporting lines for the cognitive computing system in the enterprise. The enterprise should also appoint a Chief Information Officer, with appropriate experience, to lead the cognitive computing team and create an effective environment.
- **Management of IT resources:** Formal processes must be established to manage information and data in the cognitive computing system, and must consist of information security management, protection of personal information and both planned changes and incident management.
- **Monitoring of IT investment:** The investment in and return from the cognitive computing system must be measured and managed regularly (for example, by means of cost-benefit analyses and budgets) and should be overseen by the enterprise's board. A resource gap analysis process should be established to ensure that the resources available for the governance of cognitive computing is sufficient. The allocation of resources should be evaluated and monitored to ensure that the allocation is optimal.

### 6.1.2 Cognitive computing strategy and policies

In order to implement a cognitive computing system the enterprise must develop a strategy. The cognitive computing strategy will allow management to consider all the cognitive computing elements, as well as enabling technologies, holistically. A holistic approach will assist in identifying gaps in the enterprises' current information technology system with regards to policies, infrastructure, risk management, IT resources, security and compliance, and assist management in establishing a comprehensive set of policies (Manyika *et al.*, 2011). Policies

translate desired behaviour into practical guidance (ISACA, 2012b). According to ISACA (2010), effective policies are: simple and easy to implement; flexible in order to adapt to changes in requirements; auditable; reliable under abnormal circumstances and reflect the risk appetite of the enterprise. The policies must be communicated to all of the enterprise's stakeholders, including customers. Specifically customers must be informed that policies comply with privacy regulations in order to create trust (ISACA, 2012b; Manyika *et al.*, 2011).

The enterprise should consider the following when developing a cognitive computing strategy and policies (Khan, Chan & Chua, 2016; Hayee, 2015; Hurwitz *et al.*, 2015; Smitha *et al.*, 2015; Zikopoulos *et al.*, 2015; Ballard *et al.*, 2014; Gartner, 2013; Crowe Horwath LLP, 2012; ISACA, 2012b):

- determine whether the cognitive computing system will be developed in-house, outsourced or by means of a cognitive platform (for example, IBM Watson Ecosystem);
- establish a usage policy, which identifies which of the cognitive computing system components should be supported by services from service providers (for example, cloud service providers);
- determine the resources and investment necessary to create the appropriate IT infrastructure to support the cognitive computing system and prepare a budget;
- establish ownership and accountability for cognitive computing system resource investment;
- establish performance measures to evaluate and monitor the optimisation of resources allocated to the cognitive computing system;
- leverage the cognitive computing investments by integrating the cognitive computing system with the existing IT environment and extending current controls and processes into the system;
- complete a compliance risk assessment to prioritise compliance requirements, mapping the regulations to the policies to identify gaps and redesigning policies and controls to address the identified gaps;
- establish risk management policies and procedures for the continuous identification, monitoring and evaluation of new and emerging risk relating to the cognitive computing system;
- establish data classification policies which define the purpose, ownership, and sensitivity of data types. The data classification policy must ensure that sensitive information is managed according to the risks it poses to the enterprise and that the sensitive information

is labelled with an appropriate risk level which controls encryption level, storage and transmission requirements, *etc.*

- establish privacy policies which define sensitive data and personally identifiable information; and address securing the data, transparency of usage, receiving and revocation of consent;
- establish compliance policies to ensure that imported data, derived data and metadata remain in compliance with privacy laws and other applicable regulations;
- establish policies to control inbound and outbound data traffic by implementing network filtering mechanisms such as firewalls, anti-malware and intrusion detection software;
- develop policies for the assessment, training and development of staff; and
- develop procedures for identifying of areas of innovation.

### **6.1.3 Human skills and resources controls**

The human skills and resources risks can be mitigated by ensuring that (Heller & Piziak, 2015; Hagen *et al.*, 2014; Sudarsan, 2013; McAfee & Brynjolfsson, 2012):

- Considerations and decisions regarding human skills and resources required for the cognitive computing system are included in the IT governance program.
- The process of managing resource costs and best leveraging resources is standardised.
- The skills requirements are assessed early and potential skill gaps are proactively identified.
- Existing skills of internal employees are cultivated through targeted training.
- The skills gaps are addressed by hiring new talent or experts and by leveraging consulting firms.
- Partnerships are formed with vendors and service providers involved in cognitive computing. For example, by participating in the Watson Ecosystem access can be gained to a number of resources such as tools, content, hosting services and talent.

In addition, enterprises should consider establishing a Centre of Excellence (COE). Oracle (Heller & Piziak, 2015) advises that a CEO is a new organizational best practice which consists of a cross-functional team which will include analysts, domain specialists, data engineers and data scientists. The CEO has the responsibility to plan and prioritise cognitive computing initiatives; manage, develop and support the cognitive computing initiatives, and promote the use of cognitive computing best practices throughout the enterprise. The implementation of a COE will facilitate (Heller & Piziak, 2015; Hagen *et al.*, 2014):

- cross-training of specialists from diverse data science disciplines;
- communication between the various experts, as well as aligning their goals to produce an effective cognitive computing system;
- mobilising resources for the cognitive computing initiatives; and
- the promotion of a culture within the enterprise to trust and appreciate the value of cognitive computing decisions.

## 6.2 Controls at an operational or technological level

At an operational or technological level, the enterprise should implement techniques to detect and mitigate risk exposure. These include data controls, infrastructure controls, supplier controls and life cycle controls.

### 6.2.1 Data controls

The core characteristic of cognitive computing is to deliver insight and value from data. In order to enable this characteristic, the data ingested into and created within the cognitive computing system must be managed and monitored. A significant challenge for enterprises using cognitive computing systems is the protection of sensitive data within the cognitive computing system. Data security and privacy can be controlled using the following techniques:

- **Anonymisation:** This process de-identifies all data which can be uniquely tied to an individual by removing or obscuring any personally identifiable information (Nelson, 2015; Terzi, Terzi & Sagioglu, 2015). In a survey on security and privacy approaches for big data, the following techniques were identified which will also support data anonymisation in a cognitive computing system (Panackala & Pillaib, 2015; Terzi *et al.*, 2015):
  - *AES symmetric key encryption:* In order to anonymise sensitive fields in log data, an AES symmetric key encryption is applied. When de-anonymisation is required, the masking areas are decrypted by employing the same key.
  - *Adaptive Utility based Anonymisation:* This approach is based on association mining. The technique anonymises quasi-identifiers (indirect identifiers) and consists of two steps. The first step focuses on quasi-sensitive associations amid the total data set. The second step focuses on quasi-quasi associations in cases of non-frequent data set.
  - *Sub-Tree Anonymisation:* This hybrid scheme combines two classical methods for anonymisation. By combining the Top-Down and Bottom-up methods, it increases scalability capabilities for anonymisation using MapReduce.

- *Two-phase clustering algorithm*: This technique provides scalable privacy preservation and is executed in two steps. The first step utilises tancestor clustering to split a dataset and the second step records the data with a proximity-aware agglomerative algorithm.
- **Privacy preferences**: This approach enables individuals to tag their data or information with privacy preferences. Software is then employed to track the usage of these individual parcels of data. In this way metadata is used to protect data, while placing the onus of privacy protection on individuals (Ekbia *et al.*, 2014).
- **Masking**: Masking techniques are used to disguise sensitive data by substituting real data with realistic looking fictitious data (Terzi *et al.*, 2015; Ballard *et al.*, 2014). ISACA distinguishes between Static data masking (SDM) and Dynamic data masking (DDM). However, SDM will not deliver the best results for cognitive computing because it replaces sensitive data permanently. Semantic masking is a newer masking technique. It supports cognitive computing analytics by retaining the utility (usefulness) of the data (IBM, 2014).
- **Tokenisation (data scrubbing)**: This technique replaces sensitive data with tokens or alias values obtained from a token table. The token table pair's blocks of the original data with random values and only authorised users who has access to the token table will be able to restore the data to its original form (Mattsson, 2016).

In addition, the enterprise should employ the following data management and data monitoring controls to ensure data security and privacy:

- **Proactive management**: To ensure the security of personal information and sensitive information, as well as compliance with legislative requirements (for example, the POPI Act), risks should be proactively identified, understood and management protocols developed before processing occurs (Jangara & Bezuidenhout, 2015).
- **Data life cycle control**: This is a framework for managing data from collection to retirement. The controls include documenting policies for data retention and disposition, which specifically address the manner in which collected data is preserved in its original format, and how the data is destroyed in a manner that creates a verifiable data disposition audit trail (Ballard *et al.*, 2014).
- **Monitoring system model**: This model will include security during data collection, integration, analysis, and interpretation. For example, the data collection phase will include security and network logs and the data integration process will include data filtering and classifying. The data analysis phase will include correlations and association rules to catch

events and breaches, and the data interpretation will provide visual and statistical outputs to predict network behaviour and respond to events and breaches (Terzi *et al.*, 2015).

- **Data activity monitoring:** Data activity monitoring (DAM) ensures that data access is secure by continuously monitoring activities in real time, using pattern-based policies to identify unauthorised, suspicious, and/or malicious activity (internal and external), which terminates the request and subsequently alert key personnel. The DAM solution will also produce forensic data in order to facilitate the investigation of data breaches (Zikopoulos *et al.*, 2015; Ballard *et al.*, 2014).

The following security tools and products from suppliers can be implemented to reduce data security and privacy risks in the cognitive computing system (Terzi *et al.*, 2015):

- **Trust mechanism in Hadoop:** This approach implements a trust mechanism between user and name node, which is a component of HDFS. In order to access the name node the user has to authenticate himself. MapReduce is used to encrypt data in this approach.
- **Bull eye algorithm in HDFS:** This algorithm is used to monitor all sensitive information by managing relations between original data and replicated data, and only allowing authorised users to read or write critical data.
- **Name node approach in HDFS:** This approach uses a two name node, which consists of one master and one slave node. If an incident occurs which negatively affects the master node, the administrator will provide data from the slave name node. Therefore this approach will ensure data availability in secure manner.
- **Security based data structures:** This involves security developed for a data node, in Hadoop, which consists of different types of data as well as security services for each data type. The approach consist of two stages: in the first data analytics phase, data is filtered and classified based on a data sensitivity level (sensitive, confidential, public). In the second phase, the data node of the database is created and a scheduling algorithm is applied to identify the appropriate service (identification, confidentiality, integrity, authentication, non-repudiation) and sensitivity level.

Another significant challenge in a cognitive computing system is controlling data quality. Enterprises must determine whether data sources require quality checking before integrating it into the cognitive computing system. This will differ for each cognitive computing system based on the problem which must be solved (Zikopoulos *et al.*, 2015).

For example, in cognitive security applications unclean data with anomalies and outliers will assist in identifying threats (Hurwitz *et al.*, 2015).

If the decision is made that quality checks are required, the following safeguards should be incorporated into data pre-processing to remove noise, biases, and inconsistencies, and ensure validity (Ritter, 2016; Hurwitz *et al.*, 2015; Zikopoulos *et al.*, 2015; Ballard *et al.*, 2014; Kitchin, 2014):

- **Data cleaning software:** Data cleaning software facilitates the identification of potential data-quality issues and the correction thereof in standardising data. For example, if a customer's name is listed several times due to variations in the spelling of the name, the software will make necessary corrections to assist in standardisation.
- **Data quality software:** Data quality software ensures that data and meta-data elements are represented in the same way in the entire cognitive system, thereby increasing trustworthiness of the data.
- **Standardisation:** Standardisation is used to normalise data into defined standards. It creates a consistent representation of data by parsing free-form data into single-domain data elements. An example of a defined standard is data provenance standards and rules. Provenance data validates that stored data has not been altered.
- **Data profiling:** Data profiling is a technique or process which allows the system to validate data against technical rules.
- **Metadata management:** This provides a metadata definition and a glossary to facilitate data quality, data provenance and data governance.

Other safeguards such as Mapping, Linking, Matching and filtering will occur in the corpus with the use of taxonomies, analytics and NLP.

Standardisation and/or translation into a universal form or presentation also facilitates effective data integration in a cognitive computing system. This can be done by either leveraging mature data integration tools or third-party products such as Hadoop, or by utilising the interpretation level in the cognitive computing system (NLP, text analytics, *etc.*) (Smitha *et al.*, 2015; Zikopoulos *et al.*, 2015; Wolff, 2014).

In terms of data ownership, enterprises should maintain data provenance throughout the data lifecycle by utilising data provenance standards and rules. Data provenance will allow traceability of data with regards to who owns the data, the rights attached to the data, *etc.*

The enterprise should obtain legal advice for guidance on specific issues regarding the data ownership and liability.

It is essential to continuously evaluate and monitor the results of these methods, with regards to known risks, and to identify new risks (Jangara & Bezuidenhout, 2015).

### **6.2.2 Infrastructure controls**

The cognitive computing system requires an infrastructure with sufficient storage, processing, transmission and management capacity. Enterprises should consider the following to ensure appropriate supportive infrastructure and reliable service delivery (Heller & Piziak, 2015; Hagen *et al.*, 2014; ISACA, 2012a; ISACA, 2012b):

- Identify the IT infrastructure (hardware, software, network, information systems, applications, services, assets and resources) needed to support the cognitive computing strategy.
- Assess the ability of the current IT infrastructure to support the cognitive computing system, performing a gap analysis, maturity analysis and technology dependencies analysis.
- Develop a solution to address the IT infrastructure gap through third party providers, in-house development or open-source technologies (Hadoop).
- Provide a reliable, agile, and cost-effective infrastructure, which enables innovation,
- Define and implement procedures and controls to manage and monitor the IT infrastructure and related services.
- Implement a change management process, disaster recovery plan, business continuity plan, and an infrastructure migration plan.
- Implement physical security controls.
- Establish and maintain a logical model for the configuration of infrastructure items as well as regular software updates.
- Availability of skilled personnel and teams to manage the infrastructure.

A significant challenge for enterprises utilising cognitive computing systems is to develop a secure system and networks, and secure transmission to maintain data confidentiality and integrity. The following security techniques can assist the enterprise with this challenge:

- **Authentication:** Authentication of authorised users with email, passwords, digital signatures and two-factor authentication (Danson *et al.*, 2015; Terzi *et al.*, 2015; Zikopoulos *et al.*, 2015).
- **Encryption:** Encryption provides secure transmission of data by scrambling sensitive data (Danson *et al.*, 2015). This only allows authorised users to see the clear text information while unauthorised users see a string of numbers and letters that obscure the original source (Terzi *et al.*, 2015). For example, block layer encryption will improve security while enabling clusters to scale (Smitha *et al.*, 2015; Zikopoulos *et al.*, 2015).
- **Anti-malware software:** Anti-malware software, including anti-spy and anti-virus software, eliminates the threat of malicious infections of both inbound and outbound data transmissions (Rudman, 2010).
- **Access control:** To facilitate access control the enterprise should make use of an Access Control List (ACL). ACL limits the access rights of system users and assigns the proper access rights. The enterprise should also utilise access control models, such as Role-Based Access Control (RBAC) or Attribute-Based Access Control (ABAC). RBAC introduces role-based access controls which grants permission to users based on their roles within the enterprise, while ABAC makes a context-aware decision to grant access to the system and resources based on multiple attributes. ABAC supports collaboration and data sharing across and outside the enterprise, because it provides fine-granularity, high flexibility, rich semantics and partial authentication (Cavoukian *et al.*, 2015; Rezaeibagha, Win & Susilo, 2015).
- **Key establishment scheme:** This scheme secures data by using cryptographic virtual mapping to create separate data paths which are located at different storage providers, with information encryption. It also maintains availability by holding multiple copies of each part of the data (Terzi *et al.*, 2015).
- **Secure group key transfer:** This involves group key transfer protocols for secure communications between multiple groups through key freshness, key authentication and key confidentiality. This protocol will include an online key generation centre (based on Diffie-Hellman key agreement) and linear secret sharing scheme (Terzi *et al.*, 2015).
- **Secure group data sharing:** Conditional proxy re-encryption is employed to provide secure group data sharing. In a cloud environment, the key generation and decryption processes will be executed on an outsourcing server. A condition value changing key will

be calculated and sent to the cloud, where the cloud storage will use it to transform existing data (Terzi *et al.*, 2015).

- **Secure communication channel:** A secure communication channel can be established before data transmission with mechanisms such as firewalls, VPN, network segregation, Secure Socket Layer and Transport Layer Security (Rezaeibagha *et al.*, 2015).
- **Self-assuring system:** A self-assuring system prohibits the user from proceeding with a task or accessing data if the system classifies the user as suspicious (Terzi *et al.*, 2015).

In addition, the enterprise should also employ the following data management and data monitoring controls to ensure a secure network and communication:

- **Privileged access management:** Privileged users, such as system and network administrators, vendors, and business partners, should be managed by a privileged access management solution. The solution will establish privileged user authentication (providing stronger or multi-factor authentication), privileged user credential management (providing a credential safe where sensitive passwords and key pairs can be stored and encrypted at rest, in transit or at use), privileged user session management (establishing privileged sessions, with a single sign-on, and monitoring and recording privileged user session activity for future investigation) (CA Technologies, 2015).
- **Intrusion detection and prevention architecture:** This entails the use of a security monitoring architecture which stores and processes data in distributed sources through data correlation schemes. A maliciousness likelihood matrix is then used to identify whether a domain name, packet or data flow is malicious. According to the score obtained, an alert will either arise in the detection system, or the process will be stopped by the prevention system (Danson *et al.*, 2015; Terzi *et al.*, 2015).
- **Data encryption security server:** The server administers, manages and controls encryption policies and keys, as well as access to unencrypted data (Terzi *et al.*, 2015).

### 6.2.3 Service provider controls

Cognitive computing platforms (for example, Watson Ecosystem), cloud computing platforms and data platforms (for example, Hadoop) provide solutions to address resourcing, scalability and integration risks with a service provider or supplier. However, due to the vulnerability of these dynamic open environments, it is critical to establish mitigating controls in order to ensure appropriate and reliable service delivery. Enterprises should consider the following with

regards to management and monitoring of suppliers (Khan *et al.*, 2016; Crowe Horwath LLP, 2012; ISACA, 2012b):

- The service level agreement must define, formalise and assign roles and responsibilities to the enterprise and supplier; establish and communicate procedures to review practices and internal controls applied by the supplier; define, communicate and agree on ways to identify and implement required improvements; and establish procedures to address disputes (COBIT 5) (framework).
- Supplier risks must be identified, monitored and managed in order to ensure that the supplier has the ability to continually provide secure, efficient, effective and reliable service delivery.
- The service delivery agreement must clearly define service requirements, such as security and protection of IP; and any legal or regulatory requirements.
- Service delivery must be monitored and reviewed to ensure alignment with the service level agreement.
- Management must identify which controls are relinquished to the provider.
- Management must determine the specific monitoring controls which must be implemented due to this relinquishment.
- The control activities of the provider must be validated to ensure that they align with the enterprises risk appetite.
- The controls maintained by the provider must be periodically verified for effectiveness and request independent reviews of the controls if deemed necessary.
- The ability of the supplier to provide adequate incident responses and procedures, to address system disruption and security breaches, should be assessed and monitored.
- The ability of the supplier to restore operations in the event of a disaster should be assessed and monitored.
- Management must establish an incident response plan and business continuity plan to support the suppliers' plan.
- The enterprise must integrate key internal IT management processes with those of suppliers, specifically change management, configuration management, incident management, security management and business continuity management.

#### 6.2.4 Cognitive computing life cycle controls

Controls must be established for each life cycle phase. An enterprise deploying cognitive computing systems should take the following into account during the plan, design, build and operate phases (Hurwitz *et al.*, 2015; Kitchin, 2014):

- Establish a cognitive computing strategy road map.
- Determine what the objective is of the cognitive computing system and the type of question it will have to solve.
- Define the domain or subject area for the cognitive computing system.
- Based on the domain definition, determine the domain experts needed to train and test the system.
- Establish the user requirements for the cognitive computing system.
- Evaluate the data resources the enterprise owns and which additional data resources are required to create new opportunities for insight.
- Determine the right combination of relevant data resources (internal and external) needed, to deliver the most accurate response to the questions.
- Determine the life cycle for each data source to establish which sources must be updated regularly, and to create a process to ensure that the updates are made on a timely basis.
- Determine if data from external sources should be cleaned or transformed before it is imported into the corpus.
- Validate the ingested data to ensure that data is readable, comprehensible and searchable.
- Monitor the data ingestion process to ensure that the deletion of records for security purposes was done.
- Identify which machine learning algorithms and analysis techniques are best suited to the specific domain and specific question which must be solved.
- Determine if a taxonomy or ontology is available for the domain or if a new taxonomy or ontology must be developed.
- Monitor the development of the ontology or taxonomy to identify any inconsistent assumptions, beliefs and practices which may affect the corpus.
- Determine the correct combination of algorithms which will enable the corpus to update and maintain the corpus itself.
- Develop a training and testing strategy.

- Test sample data.
- Improve system errors by adding glossaries and ontologies.
- Manage the development process by assigning ownership of the project, establish a development methodology which aligns with enterprise development standards, implement quality assurance processes throughout the lifecycles, implement project risk management processes to identify, analyse, respond to, mitigate, monitor and control risks, and implement change management processes.

### **6.3 Summary and conclusion**

A cognitive computing system exposes enterprises to significant risks which are not always sufficiently mitigated by appropriate controls. Exposure can be limited by applying a structured approach to implementing controls at a strategic and operational or technological level. Table 2 summarises the specific control which can be employed for each identified risk.

The columns of the table shows the significant risks and the related detailed risks (Chapter 5). These columns are linked to the internal control techniques that will mitigate the risks in the rows of the table. These are linked by the “X”.

**Table 3: A risk-control matrix: linking the significant cognitive computing risks to the relevant mitigating internal controls**

	<b>Governance</b>	<b>Human skills</b>	<b>Cost</b>				<b>Privacy</b>					<b>Security</b>							<b>Ownership</b>	<b>Scalability</b>	<b>Integration</b>		<b>Inter-operability</b>		<b>Life Cycle risk</b>			<b>Veracity</b>				
	Inadequate governance	Shortage of human skills and resources	Development cost	Infrastructure & management cost	Human skills cost	Security & privacy infrastructure and monitoring cost	Re-identification risk	Transparency risk	Violation of rights	Information quality risk	Unauthorised access	Regulatory compliance risk	Unauthorised access	Intentional breaches	Distributed infrastructure risk	Regulatory compliance risk	Insider breaches	Insecure computation	Data integrity risk	Outsourcing risk (Limited control)	Ownership and intellectual property rights	Infrastructure scalability risk	Algorithm scalability risk	Incompatible data from diverse sources	Incompatible hardware & software	Technical interoperability risk	Semantics interoperability risk	Inadequate high-level road map	Ineffective component development	Inadequate change management	Inadequate software development	Errors in data, training and hypothesis
<b>GOVERNANCE AND MANAGEMENT</b>																																
Implement a governance framework and develop a governance system	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Develop and implement a cognitive computing strategy	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Develop and implement cognitive computing policies	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
<b>HUMAN SKILLS AND RESOURCES CONTROLS</b>																																
Include human skills and resource requirements in the IT governance program	X	X																														
Standardise resource management	X	X																														
Perform skills gap analysis	X	X																														
Cultivate existing talent with targeted training		X			X																											

**Table 3: A risk-control matrix: linking the significant cognitive computing risks to the relevant mitigating internal controls**

	Governance	Human skills	Cost				Privacy				Security						Ownership	Scalability	Integration	Inter-operability	Life Cycle risk		Veracity									
	Inadequate governance	Shortage of human skills and resources	Development cost	Infrastructure & management cost	Human skills cost	Security & privacy infrastructure and monitoring cost	Re-identification risk	Transparency risk	Violation of rights	Information quality risk	Unauthorised access	Regulatory compliance risk	Unauthorised access	Intentional breaches	Distributed infrastructure risk	Regulatory compliance risk	Insider breaches	Insecure computation	Data integrity risk	Outsourcing risk (Limited control)	Ownership and intellectual property rights	Infrastructure scalability risk	Algorithm scalability risk	Incompatible data from diverse sources	Incompatible hardware & software	Technical interoperability risk	Semantics interoperability risk	Inadequate high-level road map	Ineffective component development	Inadequate change management	Inadequate software development	Errors in data, training and hypothesis
<b>HUMAN SKILLS AND RESOURCES CONTROLS</b>																																
Address skills gap by employing new talent and leveraging consulting firms		X			X																											
Enter into partnerships to gain access to skills resources		X			X																											
Establish a Centre of Excellence		X			X																											
<b>DATA COTROLS</b>																																
De-identify personally identifiable information (PII) with <b>anonymisation</b> techniques						X					X																					
Allow individuals to tag personal data with <b>privacy preferences</b>							X	X			X																					
Implement <b>masking</b> techniques						X					X		X	X																		
Implement <b>tokenization</b> techniques						X					X		X	X																		

**Table 3: A risk-control matrix: linking the significant cognitive computing risks to the relevant mitigating internal controls**

	<b>Governance</b>	<b>Human skills</b>	<b>Cost</b>				<b>Privacy</b>				<b>Security</b>						<b>Ownership</b>	<b>Scalability</b>		<b>Integration</b>		<b>Inter-operability</b>		<b>Life Cycle risk</b>		<b>Veracity</b>							
	Inadequate governance	Shortage of human skills and resources	Development cost	Infrastructure & management cost	Human skills cost	Security & privacy infrastructure and monitoring cost	Re-identification risk	Transparency risk	Violation of rights	Information quality risk	Unauthorised access	Regulatory compliance risk	Unauthorised access	Intentional breaches	Distributed infrastructure risk	Regulatory compliance risk	Insider breaches	Insecure computation	Data integrity risk	Outsourcing risk (Limited control)	Ownership and intellectual property rights	Infrastructure scalability risk	Algorithm scalability risk	Incompatible data from diverse sources	Incompatible hardware & software	Technical interoperability risk	Semantics interoperability risk	Inadequate high-level road map	Ineffective component development	Inadequate change management	Inadequate software development	Errors in data, training and hypothesis	
<b>DATA COTROLS</b>																																	
Perform <b>proactive management</b> to identify risks and develop protocols	X					X	X	X	X	X	X	X	X	X	X	X	X	X	X	X													
Manage data throughout its lifecycle with <b>data lifecycle controls</b>								X	X	X		X			X	X			X	X													X
Implement a <b>monitoring system model</b> to detect breaches										X	X		X	X	X		X			X													
Implement <b>data activity monitoring</b> to ensure secure data access										X		X	X			X																	
Utilise <b>Hadoop security controls</b>						X				X	X		X	X	X																		X
Utilise <b>data cleaning software</b> to ensure data quality									X																								X
Utilise <b>data quality software</b> to ensure data quality									X										X														X

**Table 3: A risk-control matrix: linking the significant cognitive computing risks to the relevant mitigating internal controls**

	<b>Governance</b>	<b>Human skills</b>	<b>Cost</b>				<b>Privacy</b>				<b>Security</b>						<b>Ownership</b>	<b>Scalability</b>		<b>Integration</b>		<b>Inter-operability</b>		<b>Life Cycle risk</b>		<b>Veracity</b>							
	Inadequate governance	Shortage of human skills and resources	Development cost	Infrastructure & management cost	Human skills cost	Security & privacy infrastructure and monitoring cost	Re-identification risk	Transparency risk	Violation of rights	Information quality risk	Unauthorised access	Regulatory compliance risk	Unauthorised access	Intentional breaches	Distributed infrastructure risk	Regulatory compliance risk	Insider breaches	Insecure computation	Data integrity risk	Outsourcing risk (Limited control)	Ownership and intellectual property rights	Infrastructure scalability risk	Algorithm scalability risk	Incompatible data from diverse sources	Incompatible hardware & software	Technical interoperability risk	Semantics interoperability risk	Inadequate high-level road map	Ineffective component development	Inadequate change management	Inadequate software development	Errors in data, training and hypothesis	
<b>DATA COTROLS</b>																																	
Standardise data to defined standards									X												X			X	X	X	X						
Utilise <b>data profiling</b> to validate data									X										X													X	
Implement <b>metadata management</b> to ensure data quality and provenance	X																				X											X	
Obtain <b>legal advice</b>																					X												
Implement <b>data provenance</b> standards to allow traceability									X												X												X
<b>INFRASTRUCTURE COTROLS</b>																																	
Perform a <b>need assessment</b> to identify IT Infrastructure required for a cognitive computing system	X																																
Perform a <b>gap analysis &amp; maturity analysis</b> to access current IT infrastructure	X			X																													

**Table 3: A risk-control matrix: linking the significant cognitive computing risks to the relevant mitigating internal controls**

	<b>Governance</b>	<b>Human skills</b>	<b>Cost</b>				<b>Privacy</b>				<b>Security</b>						<b>Ownership</b>	<b>Scalability</b>		<b>Integration</b>		<b>Inter-operability</b>		<b>Life Cycle risk</b>			<b>Veracity</b>						
	Inadequate governance	Shortage of human skills and resources	Development cost	Infrastructure & management cost	Human skills cost	Security & privacy infrastructure and monitoring cost	Re-identification risk	Transparency risk	Violation of rights	Information quality risk	Unauthorised access	Regulatory compliance risk	Unauthorised access	Intentional breaches	Distributed infrastructure risk	Regulatory compliance risk	Insider breaches	Insecure computation	Data integrity risk	Outsourcing risk (Limited control)	Ownership and intellectual property rights	Infrastructure scalability risk	Algorithm scalability risk	Incompatible data from diverse sources	Incompatible hardware & software	Technical interoperability risk	Semantics interoperability risk	Inadequate high-level road map	Ineffective component development	Inadequate change management	Inadequate software development	Errors in data, training and hypothesis	
<b>INFRASTRUCTURE COTROLS</b>																																	
Develop an <b>IT infrastructure solution</b> to address the 'gap'	X			X																													
<b>Manage and monitor</b> the IT infrastructure	X			X										X										X	X								
Implement <b>change management, disaster recovery &amp; business continuity plans</b>	X																												X				
Implement <b>physical security</b>										X		X		X																			
Implement and maintain <b>configuration &amp; software updates</b>	X									X		X	X																				
Implement <b>authentication techniques</b>										X		X																					
Implement <b>encryption techniques</b>						X				X	X	X	X	X	X					X													
Utilise <b>anti-malware software</b>													X																				

**Table 3: A risk-control matrix: linking the significant cognitive computing risks to the relevant mitigating internal controls**

	<b>Governance</b>	<b>Human skills</b>	<b>Cost</b>				<b>Privacy</b>				<b>Security</b>						<b>Ownership</b>	<b>Scalability</b>	<b>Integration</b>	<b>Inter-operability</b>		<b>Life Cycle risk</b>		<b>Veracity</b>								
	Inadequate governance	Shortage of human skills and resources	Development cost	Infrastructure & management cost	Human skills cost	Security & privacy infrastructure and monitoring cost	Re-identification risk	Transparency risk	Violation of rights	Information quality risk	Unauthorised access	Regulatory compliance risk	Unauthorised access	Intentional breaches	Distributed infrastructure risk	Regulatory compliance risk	Insider breaches	Insecure computation	Data integrity risk	Outsourcing risk (Limited control)	Ownership and intellectual property rights	Infrastructure scalability risk	Algorithm scalability risk	Incompatible data from diverse sources	Incompatible hardware & software	Technical interoperability risk	Semantics interoperability risk	Inadequate high-level road map	Ineffective component development	Inadequate change management	Inadequate software development	Errors in data, training and hypothesis
<b>INFRASTRUCTURE COTROLS</b>																																
Implement <b>access control</b>										X	X	X				X			X													
Secure data by implementing a <b>key establishment scheme</b>									X	X	X					X																
Implement <b>secure group key transfer</b>										X	X																					
Implement <b>secure group data sharing</b>										X	X																					
Establish a <b>secure communication channel</b> for data transmission										X	X		X							X												
Implement a <b>self-assuring system</b> to prohibit suspicious tasks and users						X				X	X	X	X		X	X				X												
Establish <b>privileged access management</b>																X																
Monitor the system through <b>intrusion detection software</b>										X	X	X	X						X													

**Table 3: A risk-control matrix: linking the significant cognitive computing risks to the relevant mitigating internal controls**

	<b>Governance</b>	<b>Human skills</b>	<b>Cost</b>				<b>Privacy</b>				<b>Security</b>						<b>Ownership</b>	<b>Scalability</b>	<b>Integration</b>	<b>Inter-operability</b>	<b>Life Cycle risk</b>		<b>Veracity</b>									
	Inadequate governance	Shortage of human skills and resources	Development cost	Infrastructure & management cost	Human skills cost	Security & privacy infrastructure and monitoring cost	Re-identification risk	Transparency risk	Violation of rights	Information quality risk	Unauthorised access	Regulatory compliance risk	Unauthorised access	Intentional breaches	Distributed infrastructure risk	Regulatory compliance risk	Insider breaches	Insecure computation	Data integrity risk	Outsourcing risk (Limited control)	Ownership and intellectual property rights	Infrastructure scalability risk	Algorithm scalability risk	Incompatible data from diverse sources	Incompatible hardware & software	Technical interoperability risk	Semantics interoperability risk	Inadequate high-level road map	Ineffective component development	Inadequate change management	Inadequate software development	Errors in data, training and hypothesis
<b>SERVICE PROVIDER CONTROLS</b>																																
Establish <b>service level agreements</b>	X																															
Manage and monitor <b>service provider (SP) risks</b>	X																															
Define service requirements in <b>service delivery agreements</b>	X																															
<b>Monitor and review service delivery</b>	X																															
Implement <b>monitoring controls</b> (SP control management)	X																			X												
Request <b>independent reviews</b> to evaluate controls implemented by SP	X																			X												
Assess and monitor the service providers' <b>incident response</b>	X																			X												

**Table 3: A risk-control matrix: linking the significant cognitive computing risks to the relevant mitigating internal controls**

	<b>Governance</b>	<b>Human skills</b>	<b>Cost</b>				<b>Privacy</b>				<b>Security</b>						<b>Ownership</b>	<b>Scalability</b>		<b>Integration</b>		<b>Inter-operability</b>		<b>Life Cycle risk</b>		<b>Veracity</b>						
	Inadequate governance	Shortage of human skills and resources	Development cost	Infrastructure & management cost	Human skills cost	Security & privacy infrastructure and monitoring cost	Re-identification risk	Transparency risk	Violation of rights	Information quality risk	Unauthorised access	Regulatory compliance risk	Unauthorised access	Intentional breaches	Distributed infrastructure risk	Regulatory compliance risk	Insider breaches	Insecure computation	Data integrity risk	Outsourcing risk (Limited control)	Ownership and intellectual property rights	Infrastructure scalability risk	Algorithm scalability risk	Incompatible data from diverse sources	Incompatible hardware & software	Technical interoperability risk	Semantics interoperability risk	Inadequate high-level road map	Ineffective component development	Inadequate change management	Inadequate software development	Errors in data, training and hypothesis
<b>SERVICE PROVIDER CONTROLS</b>																																
Assess and monitor the service providers' <b>disaster recovery plans</b>	X																			X												
Establish <b>incident response and business continuity plans to support SP plans.</b>	X																			X												
<b>Integrate key IT management processes</b> with SP processes																				X												
Determine if taxonomies and ontologies are available and monitor the development process																														X		
Develop a <b>training and testing strategy</b>																													X			X
<b>Test sample data</b>																													X			X
<b>Add glossaries and ontologies</b> to improve the CCS																													X			X



**Table 3: A risk-control matrix: linking the significant cognitive computing risks to the relevant mitigating internal controls**

	<b>Governance</b>	<b>Human skills</b>	<b>Cost</b>				<b>Privacy</b>				<b>Security</b>						<b>Ownership</b>	<b>Scalability</b>	<b>Integration</b>		<b>Inter-operability</b>		<b>Life Cycle risk</b>		<b>Veracity</b>								
	Inadequate governance	Shortage of human skills and resources	Development cost	Infrastructure & management cost	Human skills cost	Security & privacy infrastructure and monitoring cost	Re-identification risk	Transparency risk	Violation of rights	Information quality risk	Unauthorised access	Regulatory compliance risk	Unauthorised access	Intentional breaches	Distributed infrastructure risk	Regulatory compliance risk	Insider breaches	Insecure computation	Data integrity risk	Outsourcing risk (Limited control)	Ownership and intellectual property rights	Infrastructure scalability risk	Algorithm scalability risk	Incompatible data from diverse sources	Incompatible hardware & software	Technical interoperability risk	Semantics interoperability risk	Inadequate high-level road map	Ineffective component development	Inadequate change management	Inadequate software development	Errors in data, training and hypothesis	
<b>LIFE CYCLE CONTROLS</b>																																	
Identify the best algorithms for the domain and the right combination																							X							X			
Determine if taxonomies and ontologies are available and monitor the development process																														X			
Develop a <b>training and testing strategy</b>																													X				X
<b>Test sample data</b>																													X				X
<b>Add glossaries and ontologies</b> to improve the CCS																													X				X
<b>Manage the system and software development process</b>																												X	X	X			

## CHAPTER 7: CONCLUSION

The exponential growth of data, advances in enabling technology and the ability of cognitive computing to realise significant business value from data sources have accelerated the growth and application of cognitive computing. However, enterprises are implementing cognitive computing systems without understanding the technology or the risks the enterprise are exposed to pertaining to the implementation of the cognitive computing system. Given that enterprises are unaware of these risks, they are not implementing a system of internal control in a comprehensive manner. The objective of the research was to identify and mitigate significant risks pertaining to the implementation of a cognitive computing system. The research aimed to provide stakeholders with an understanding of the core components of a cognitive computing system, to identify the significant risks pertaining to the implementation of cognitive computing and to recommend safeguards to mitigate these risks to an acceptable level.

The research found that a cognitive computing system consist of twelve core components which can be classified into four phases based on their function within the cognitive computing system:

- **Structured, semi-structured and unstructured data (Data ingestion phase):**  
A cognitive computing system requires large quantities of structured, semi-structured and unstructured data to discover patterns and gain new insight from the data.
- **Data access, feature extraction, natural language processing, deep learning and metadata (Read phase):** The function of these components within a cognitive computing system are to extract features, meaning and context from ingested data and metadata in preparation for the corpus, in essence making it machine-readable.
- **Corpus and advanced analytics (Resolve phase):** The corpus is the body of knowledge of the cognitive computing system and consists of comprehensible data (obtained from the read phase) about a specific domain. The content of the corpus enables the cognitive computing system to answer questions, discover new patterns and deliver new insight. The advanced analytic algorithms assist in the identification of new patterns and relationships to increase insight and to generate new data for hypotheses.
- **Hypothesis generation and scoring, and machine learning (Reason phase):** Hypotheses are generated and scored to uncover relationships, provide recommendations and resolve problems. The hypothesis is based on data / knowledge obtained from the

corpus. The machine learning algorithms enable the cognitive computing system to continuously learn from the data and knowledge in the corpus, as well as the hypothesis results in order to become a more effective system.

The core components of the cognitive computing system depend on a distributed environment supported by an agile and flexible infrastructure. Moreover, the functioning of cognitive computing systems are dependent on enabling technologies such as Hadoop, cloud computing and big data.

The detailed processes of COBIT were used to identify significant risks based on the understanding of the cognitive computing system and its core components. Significant risks were identified at a strategic and operational or technological level:

- **Significant risks at a strategic level:** The risks identified included inadequate governance and management of cognitive computing systems and inadequate human skills and resource management.
- **Significant risks at an operational or technological level:** The risks identified were divided into three groups: (i) risks that affect the objective of the cognitive computing system, including cost, privacy, security and ownership, (ii) risks that affect the ability of the cognitive computing system to function effectively, including scalability, integration, interoperability and veracity, and (iii) risks that affect both the objective of the cognitive computing system, as well as the ability of the cognitive computing system to function effectively, including cognitive computing life cycle risks.

Safeguards and controls were formulated in order to ensure that the significant risks pertaining to the implementation of a cognitive computing system are mitigated to an acceptable level. These include:

- Governance and management at a strategic level by establishing a cognitive computing governance framework, which include the development and implementation of cognitive computing strategies and policies, as well as the implementation of human skills and resources controls.
- Internal control techniques to detect and mitigate risks at an operational or technological level, which include data controls, infrastructure controls, supplier controls level and life cycle controls.

A risk matrix (Table 2) and risk-control matrix (Table 3) were compiled from the research. The risk matrix maps the relevant core components of each cognitive computing phase with the significant risk which they introduce, while the risk-control matrix maps the significant risk to the relevant control techniques formulated to mitigate the specific risk to an acceptable level.

There are two areas of potential future research. The first research area involves the identification of significant risks and mitigating controls with regard to the virtualisation and application components of the cognitive computing system. These two components did not fall within the scope of this study. The second research area would involve a study of the process to develop a cognitive computing application. The research would include how cognitive computing can be embedded in a business application, as well as, how to integrate the cognitive computing application with other systems.

**BIBLIOGRAPHY**

- Anisingaraju, S. 2013. What does COBIT 5 mean for your business? *COBIT Focus*, 4, October 2013 [Online]. [http://www.isaca.org/knowledge-center/cobit/documents/cf-vol-4-what-does-cobit-5-mean-for-your-business-anisingaraju\\_nlt\\_1013.pdf](http://www.isaca.org/knowledge-center/cobit/documents/cf-vol-4-what-does-cobit-5-mean-for-your-business-anisingaraju_nlt_1013.pdf) [2015, November 11].
- Bailor, C. 2006. The why factor in speech analytics. *CRM Magazine* [Electronic], 10(8):32-37. Available: <http://web.b.ebscohost.com.ez.sun.ac.za/ehost/pdfviewer/pdfviewer?vid=4&sid=01d29368-97a4-4991-9555-479e61ba1d2d%40sessionmgr114&hid=101> [2015, July 7].
- Ballard, C., Compert, C., Jesionowski, T., Milman, I., Plants, B., Rosen, B. & Smith, H. 2014. *Information governance principles and practices for a big data landscape*. [New York]: IBM Redbooks [Online]. Available: <http://www.redbooks.ibm.com/Redbooks.nsf/RedbookAbstracts/sg248165.html?Open> [2016, January 20].
- Bartens, Y., de Haes, S., Lamoen, Y., Schulte, F. & Voss, S. 2015. On the Way to a Minimum Baseline in IT Governance: Using Expert Views for Selective Implementation of COBIT 5. *2015 48th Hawaii International Conference on System Sciences*. 5-8 January, 4554-4563. Kauai, Hawaii [Electronic]. Available: <http://ieeexplore.ieee.org.ez.sun.ac.za/xpl/articleDetails.jsp?arnumber=7070363> [2015, October 26].
- Bataller, C. & Harris, J. 2015. *Turning Cognitive Computing into Business Value Today*. Accenture [Online]. Available: [https://www.accenture.com/t20150521T005731\\_\\_w\\_/us-en/\\_acnmedia/Accenture/Conversion-Assets/DotCom/Documents/Global/PDF/Dualpub\\_8/Accenture-Turning-Cognitive-Computing-Business-Value-Today.pdf](https://www.accenture.com/t20150521T005731__w_/us-en/_acnmedia/Accenture/Conversion-Assets/DotCom/Documents/Global/PDF/Dualpub_8/Accenture-Turning-Cognitive-Computing-Business-Value-Today.pdf) [2015, June 15].
- Bellisimo, J. 2015. *What's The Future Of Cognitive Computing? IBM Watson*. [Online]. Available: <http://www.forbes.com/sites/ibm/2015/02/23/whatsthefutureofcognitivecomputingibmwatson/> [Accessed: 5 April 2015].
- Bradbury, D. 2015. *Cognitive computing: What can and can't we do, and should lipreading be banned?* [Online]. Available: [http://www.theregister.co.uk/2015/09/04/cognitive\\_computing\\_humans\\_hal/](http://www.theregister.co.uk/2015/09/04/cognitive_computing_humans_hal/) [2016, 4 May].
- CA Technologies. 2015. *How Can I Defend my Hybrid Enterprise From Data Breaches and Insider Threats?* [Online]. Available: [http://docs.media.bitpipe.com/io\\_12x/io\\_128619/item\\_1283253/EC-solutionbrief-privilegedaccessmanagement-Final.pdf](http://docs.media.bitpipe.com/io_12x/io_128619/item_1283253/EC-solutionbrief-privilegedaccessmanagement-Final.pdf) [2016, February 9].
- Cavoukian, A., Chibba, M., Williamson, G. & Ferguson, A. 2015. *The Importance of ABAC: Attribute-Based Access Control to Big Data: Privacy and Context*. Toronto: Ryerson University [Online]. Available: <http://www.ryerson.ca/content/dam/pbdi/Resources/The%20Importance%20of%20ABAC%20to%20Big%20Data%2005-2015.pdf> [2016, June 28].

- Chan, J.O. 2013. An Architecture for Big Data Analytics. *Communications of the IIMA* [Electronic], 13(2):1-13. Available: <http://search.proquest.com.ez.sun.ac.za/docview/1518604853/fulltextPDF/D8BDC37A0BAC4EA9PQ/1?accountid=14049> [2015, June 9].
- Chen, K., Li, X. & Wang, H. 2015. On the model design of integrated intelligent big data analytics systems. *Management & Data Systems* [Electronic], 115(9):1666-682. Available: <http://www.emeraldinsight.com.ez.sun.ac.za/doi/pdfplus/10.1108/IMDS-03-2015-0086> [2016, March 22].
- Court, D. 2015. *Getting big impact from big data*. McKinsey & Company [Online]. [http://www.mckinsey.com/insights/business\\_technology/getting\\_big\\_impact\\_from\\_big\\_data](http://www.mckinsey.com/insights/business_technology/getting_big_impact_from_big_data) [2015, June 4].
- Crespo, O. 2015. COBIT 5 Adoption: Understand and Be Understood. *COBIT Focus*, 2 November 2015 [Online]. Available: <http://www.isaca.org/COBIT/focus/Pages/cobit-5-adoption-understand-and-be-understood.aspx> [2015, November 25].
- Crowe Horwath LLP. 2012. *Enterprise risk management for cloud computing* [Online]. Available: <http://www.coso.org/documents/Cloud%20Computing%20Thought%20Paper.pdf> [2016, March 10].
- D'Aquila, J. 2013. COSO's Internal Control – Integrated Framework. *CPA Journal* [Electronic], 83(10):22-29. Available: <http://search.proquest.com.ez.sun.ac.za/docview/1468440142/AACB6BBAF46746CBPQ/8?accountid=14049> [2015, November 24].
- Danson, F., Pierce, D. & Shilling, M. 2015. *Tech Trend 2015*. Deloitte University Press: 96-109 [Online]. Available: <http://dupress.com/articles/tech-trends-2015-amplified-intelligence/?id=gx:2el:3dc:dup1009:eng:cons:tt15:dcpromo> [2015, April 5].
- de Bruyn, M. 2014. The Protection Of Personal Information (POPI) Act - Impact On South Africa. *International Business & Economics Research Journal* [Electronic], 13(6):1315-1340 Available: <http://www.cluteinstitute.com/ojs/index.php/IBER/article/view/8922> [2016, January 28].
- DigitalOcean. 2014. *A comparison of NoSQL Database Management Systems and Models* [Online]. Available: <https://www.digitalocean.com/community/tutorials/a-comparison-of-nosql-database-management-systems-and-models> [2015, June 4].
- Digital Reasoning Systems. 2015. *Synthesys: Technology overview*. [Franklin]: Digital Reasoning Systems Incorporated [Online]. Available: [http://www.digitalreasoning.com/resources/Synthesys\\_v3.9\\_Technology\\_Overview\\_FINAL\\_Jan\\_2015.pdf](http://www.digitalreasoning.com/resources/Synthesys_v3.9_Technology_Overview_FINAL_Jan_2015.pdf) [2015, December 21].
- Dijcks, J. 2011. *Oracle: Big Data for the Enterprise*. [Redwood Shores]: Oracle Corporation [Online]. Available: <http://www.oracle.com/us/products/database/big-data-for-enterprise-519135.pdf> [2015, July 31].
- Dix, J. 2014. *What's better for your big data application, SQL or NoSQL?* [Online]. Available: <http://www.networkworld.com/article/2226514/tech-debates/what-s-better-for-your-big-data-application--sql-or-nosql-.html> [2015, July 7].

- Drury, N., Harper, A., Marshall, A. & Sarkar, S. 2015. *Breakthrough banking*. [Somers, New York]: IBM Corporation [Online]. Available: <http://www-01.ibm.com/common/ssi/cgi-bin/ssialias?subtype=XB&infotype=PM&htmlfid=GBE03713USEN&attachment=GBE03713USEN.PDF> [2016, July 22].
- Ekbia, H., Mattioli, M., Kouper, I., Arave, G., Ghazinejad, A., Bowman, T., Suri, V.R., Tsou, A., Weingart, S. & Sugimoto, C.R. 2014. Big Data, Bigger Dilemmas: A Critical Review. *Journal of the Association for Information Science and Technology* [Electronic], 66(8):1523-1545. Available: <http://onlinelibrary.wiley.com.ez.sun.ac.za/doi/10.1002/asi.23294/full> [2015, December 22].
- Enterra Solutions. 2016. *The power of ontologies* [Online]. Available: <http://www.enterrasolutions.com/products/ontologies> [2016, May 10].
- Fluss, D. 2011. Realizing the Benefits of Speech Analytics. *CRM Magazine* [Electronic], 15(3):40-41. Available: <http://web.b.ebscohost.com.ez.sun.ac.za/ehost/pdfviewer/pdfviewer?vid=2&sid=01d29368-97a4-4991-9555-479e61ba1d2d%40sessionmgr114&hid=101> [2015, July 7].
- Fox, B., Lala, R. & Coelho, O.C. 2015. *Dialing in a new frequency: Your cognitive future in the communications industry*. [Somers, New York]: IBM Corporation [Online]. Available: <http://www-01.ibm.com/common/ssi/cgi-bin/ssialias?subtype=XB&infotype=PM&htmlfid=GBE03722USEN&attachment=GBE03722USEN.PDF> [2016, July 22].
- Fry, M. 2005. *Top ten reasons organizations are unsuccessful implementing ITIL* [Online]. Available: <http://costkiller.net/tribune/Tribu-PDF/Top-Ten-Reasons-Organizations-are-Unsuccessful-Implementing-ITIL.htm> [2015, November 26].
- Gartner. 2013. *Information Governance Best Practice: Adopt a Use Case Approach* [Online]. [http://www.bitpipe.com/detail/RES/1392921757\\_401.html](http://www.bitpipe.com/detail/RES/1392921757_401.html) [2015, June 5].
- Gartner. 2015. *Gartner's IT glossary* [Online]. Available: <http://www.gartner.com/it-glossary/> [2015, July 8].
- Géczy, P. 2014. Big data characteristics. *The Macrotheme Review* [Electronic], 3(6):94-104. Available: [http://macrotheme.com/yahoo\\_site\\_admin/assets/docs/8MR36Pe.97110828.pdf](http://macrotheme.com/yahoo_site_admin/assets/docs/8MR36Pe.97110828.pdf) [2016, January 19].
- Goosen, R. & Rudman, R. 2013a. An Integrated Framework To Implement It Governance Principles At A Strategic And Operational Level For Medium-To Large-Sized South African Businesses. *International Business & Economics Research Journal* [Electronic], 12(7):835-854. Available: <http://www.cluteinstitute.com/ojs/index.php/IBER/article/view/7972/8026> [2015, September 11].
- Goosen, R. & Rudman, R. 2013b. The development of an integrated framework in order to address King III's IT governance principles at a strategic level. *South African Journal of Business Management* [Electronic], 44(4):91-101. Available: <http://web.a.ebscohost.com.ez.sun.ac.za/ehost/pdfviewer/pdfviewer?vid=2&sid=f4174fa7-73a6-470a-a527-b0cb6d73294e%40sessionmgr4002&hid=4207> [2014, October 24].

- Grose, C., Kargidis, T. & Vasilios, C. 2014. Corporate Governance in practice. The Greek case. *Procedia Economics and Finance* [Electronic], 9:369-379. Available: [http://ac.els-cdn.com/S2212567114000380/1-s2.0-S2212567114000380-main.pdf?\\_tid=07e50188-718c-11e4-86f9-00000aab0f6c&acdnat=1416580877\\_a8085273ad48001706bbd6d2008997e1](http://ac.els-cdn.com/S2212567114000380/1-s2.0-S2212567114000380-main.pdf?_tid=07e50188-718c-11e4-86f9-00000aab0f6c&acdnat=1416580877_a8085273ad48001706bbd6d2008997e1) [2015, August 5].
- Haav, H. & Kungas, P. 2013. *Big data computing*. New York: CRC Press. 245-269 [Online]. Available: [https://www.researchgate.net/profile/Peep\\_Kungas/publication/260706654\\_Semantic\\_data\\_in\\_teroperability\\_the\\_key\\_problem\\_of\\_big\\_data/links/53dfc6190cf2aede4b493d2b.pdf/download?version=vtp](https://www.researchgate.net/profile/Peep_Kungas/publication/260706654_Semantic_data_in_teroperability_the_key_problem_of_big_data/links/53dfc6190cf2aede4b493d2b.pdf/download?version=vtp) [2016, May 17].
- Hagen, C., Khan, K., Evans, H., Thota, B., Wall, D. & Seshadri, A. 2014. *IT's Challenge: Bringing Structure to the Unstructured World of Big Data*. A.T.Kearney [Online]. Available: <https://www.atkearney.com/documents/10192/5148172/ITs+Challenge-+Bringing+Structure+to+the+Unstructured+World+of+Big+Data.pdf/86dd0149-6abb-4252-8e6f-3cddde3df247> [2015, April 15].
- Harper, J. 2015. *2015 Trends in Data Modeling: The Machine Learning Effect* [Online]. Available: <http://www.dataversity.net/2015-trends-data-modeling-machine-learning-effect> [2015, March 7].
- Hayee, B. 2015. *Tips for creating a data classification policy* [Online]. Available: <http://searchsecurity.techtarget.com/feature/Tips-for-creating-a-data-classification-policy> [2016, June 24].
- Heller, P. & Piziak, D. 2015. *An Enterprise Architect's Guide to Big Data*. [Redwood Shores]: Oracle Corporation [Online]. Available: <http://www.oracle.com/technetwork/topics/entarch/articles/oea-big-data-guide-1522052.pdf> [2015, June 9].
- High, R. 2012. *The Era of Cognitive Systems: An Inside Look at IBM Watson and How it Works*. [New York]: IBM Redbooks [Online]. Available: <http://www.redbooks.ibm.com/redpapers/pdfs/redp4955.pdf> [2015, June 17].
- Hodges, D. & Creese, S. 2013. Breaking the Arc: Risk Control for Big Data. *2013 IEEE International Conference on Big Data*. 6-9 October, 613-621. Santa Clara, CA, USA [Electronic]. Available: [http://ieeexplore.ieee.org.ez.sun.ac.za/xpl/articleDetails.jsp?arnumber=6691630&filter%3DAND%28p\\_IS\\_Number%3A6690588%29%26pageNumber%3D4](http://ieeexplore.ieee.org.ez.sun.ac.za/xpl/articleDetails.jsp?arnumber=6691630&filter%3DAND%28p_IS_Number%3A6690588%29%26pageNumber%3D4) [2015, June 15].
- Huang, S.M., Hung, W.H., Yen, D.C., Chang, I.C. & Jiang, D. 2011. Building the evaluation model of the IT general control for CPAs under enterprise risk management, *Decision Support Systems* [Electronic], 50(4):692-701. Available: [http://ac.els-cdn.com.ez.sun.ac.za/S0167923610001399/1-s2.0-S0167923610001399-main.pdf?\\_tid=928b207e-9f08-11e5-b735-00000aab0f6c&acdnat=1449729667\\_a58a06f939d2b0385b0d46520c5a5d3e](http://ac.els-cdn.com.ez.sun.ac.za/S0167923610001399/1-s2.0-S0167923610001399-main.pdf?_tid=928b207e-9f08-11e5-b735-00000aab0f6c&acdnat=1449729667_a58a06f939d2b0385b0d46520c5a5d3e) [2015, November 11].
- Hurwitz, J.S., Kaufman, M. & Bowles, A. 2015. *Cognitive Analytics and Big Data Analysis*. Indianapolis: Jodn Wiley & Sons, Inc.

IBM Corporation. 2014a. *Big data integration and Hadoop* [Online]. Available: <http://static.ziftsolutions.com/files/8a28785f4b557204014b5f54d78315ff.PDF> [2015, June 5].

IBM Corporation. 2014b. *Data protection for big data environments* [Online]. Available: <https://tdwi.org/~media/3192291DBEC446129F464922D782A09C.PDF> [2015, June 23].

IBM Corporation. 2014c. *The MDM advantage: Creating insight from big data* [Online]. Available: [http://www-01.ibm.com/common/ssi/cgi-bin/ssialias?subtype=BK&infotype=PM&appname=SWGE\\_IM\\_EZ\\_USEN&htmlfid=IMM14124USEN&attachment=IMM14124USEN.PDF&ce=ISM0056&ct=swg&cmp=ibmsocial&m=h&cr=crossbrand&ccy=us#loaded](http://www-01.ibm.com/common/ssi/cgi-bin/ssialias?subtype=BK&infotype=PM&appname=SWGE_IM_EZ_USEN&htmlfid=IMM14124USEN&attachment=IMM14124USEN.PDF&ce=ISM0056&ct=swg&cmp=ibmsocial&m=h&cr=crossbrand&ccy=us#loaded) [2015, June 5].

IBM Corporation. n.d. *What is data integration?* [Online]: Available: <http://www.ibm.com/analytics/us/en/technology/data-integration/> [2016, January 20].

Institute of Directors Southern Africa (IODSA). 2009. *King Code of Governance for South Africa 2009* [Online]. Available: [http://c.yimcdn.com/sites/www.iodsa.co.za/resource/collection/94445006-4F18-4335-B7FB-7F5A8B23FB3F/King\\_III\\_Code\\_for\\_Governance\\_Principles\\_.pdf](http://c.yimcdn.com/sites/www.iodsa.co.za/resource/collection/94445006-4F18-4335-B7FB-7F5A8B23FB3F/King_III_Code_for_Governance_Principles_.pdf) [2014, July 7].

ISACA. 2012a. *COBIT 5. A Business Framework for the Governance and Management of Enterprise IT* [Online]. Available: <http://www.isaca.org/cobit/pages/cobitLiteRegistrationdownload.aspx?RegID=72492e8e-70a1-4ee6-91a4-fcb5e3f37539> [2015, August 11].

ISACA. 2012b. *COBIT 5: Process Reference Guide*. United States of America.

ISACA. 2014. *Generating value from big data analytics* [Online]. Available: <http://f6ce14d4647f05e937f4-4d6abce208e5e17c2085b466b98c2083.r3.cf1.rackcdn.com/generating-value-from-big-data-analytics-pdf-2-w-907.pdf> [2015, December 24].

ISACA. 2016a. *Internal control using Cobit 5* [Online]. Available: <http://www.isaca.org/knowledge-center/research/researchdeliverables/pages/internal-control-using-cobit-5.aspx> [2016, March 17].

ISACA. 2016b. *State of cybersecurity implications for 2016*. An ISACA and RSA Conference Survey [Online]. Available: [http://www.isaca.org/cyber/Documents/state-of-cybersecurity\\_res\\_eng\\_0316.pdf](http://www.isaca.org/cyber/Documents/state-of-cybersecurity_res_eng_0316.pdf) [2016, March 3].

ISO. 2014. *Big Data*. ISO/IEC JTC 1 Information technology [Online]. Available: [http://www.iso.org/iso/big\\_data\\_report-jtc1.pdf](http://www.iso.org/iso/big_data_report-jtc1.pdf) [2016, January 19].

ITIL. 2012. *An Introductory Overview of ITIL® 2011* [Online] Available: [http://www.doc-developpement-durable.org/file/Projets-informatiques/cours-&-manuels-informatiques/ITIL/An\\_Introductory\\_Overview\\_of\\_ITIL\\_V3.pdf](http://www.doc-developpement-durable.org/file/Projets-informatiques/cours-&-manuels-informatiques/ITIL/An_Introductory_Overview_of_ITIL_V3.pdf) [2015, August 6].

Jagadish, H.V., Gehrke, J., Labrinidis, A., Papakonstantinou, Y., Patel, J.M., Ramakrishnan, R. & Shahabi, C. 2014. Big data and its technical challenges. *Communications of the ACM* [Electronic], 57(7): 86-94. Available: <http://web.a.ebscohost.com.ez.sun.ac.za/ehost/pdfviewer/pdfviewer?sid=86fe4d13-5f46-42de-bd9e-29b58f96dd1e%40sessionmgr4007&vid=2&hid=4112> [2015, June 1].

- Jangara, T.M. & Bezuidenhout, H. 2015. Addressing emerging risks in transborder cloud computing and the protection of personal information: The role of internal auditors. *Southern African Journal of Accountability and Auditing Research* [Electronic], 17(1): 11-24 Available:  
[http://reference.sabinet.co.za.ez.sun.ac.za/webx/access/electronic\\_journals/sajaar/sajaar\\_v17\\_n1\\_a2.pdf](http://reference.sabinet.co.za.ez.sun.ac.za/webx/access/electronic_journals/sajaar/sajaar_v17_n1_a2.pdf). [2016, March 10].
- Janssen, M., Estevez, E. & Janowski, T. 2014. Interoperability in Big, Open and Linked data - Organizational Maturity, Capabilities, and Data Portfolios. *Computer* [Electronic], 47(10):44-49. Available:  
<http://ieeexplore.ieee.org.ez.sun.ac.za/stamp/stamp.jsp?tp=&arnumber=6926683> [2016, May 17].
- Jensen, M. 2013. Challenges of Privacy Protection in Big Data Analytics. *2013 IEEE International Congress on Big Data*. 27 June - 2 July, 235-238. Santa Clara, California, USA [Electronic]. Available:  
<http://ieeexplore.ieee.org.ez.sun.ac.za/stamp/stamp.jsp?tp=&arnumber=6597142> [Accessed: 5 January 2016].
- Jewell, D., Barros, R.D., Diederichs, S., Duijvestijn, L.M., Hammersley, M., Hazra, A., Holban, C., Li, Osaigbovo, O., Plach, A., Portilla, I., Saptarshi, M., Seera, H.P., Stahl, H. & Zolotow, C. 2014. *Performance and Capacity Implications for Big Data*. IBM Corporation [Online]. Available: <http://www.redbooks.ibm.com/redpapers/pdfs/redp5070.pdf> [2015, April 8].
- Juiz, C. & Toomey, M. 2015. To Govern IT, or Not to Govern IT? *Communications of the ACM* [Electronic], 58(2):58-64 Available:  
<http://web.a.ebscohost.com.ez.sun.ac.za/ehost/pdfviewer/pdfviewer?vid=2&sid=baa207b7-83b9-4427-922a-3a2a7f320fbd%40sessionmgr4001&hid=4207> [2015, August 14].
- Kaisler, S., Armour, F., Espinosa, J. & Money, W. 2013. Big Data: Issues and Challenges Moving Forward. *46th Hawaii International Conference on System Sciences (HICSS)*. 7-10 January, 995–1004. Wailea, Maui, Hawaii [Electronic]. Available:  
<http://ieeexplore.ieee.org.ez.sun.ac.za/stamp/stamp.jsp?tp=&arnumber=6479953> [2015, December 22].
- Kelly III, J.E. & Hamm, S. 2013. *Smart Machines: IBM's Watson and the Era of Cognitive Computing*. New York: Columbia University Press [Online]. Available:  
<http://sun.eblib.com.ez.sun.ac.za/patron/FullRecord.aspx?p=1319720&echo=1&userid=rbNnV9bOROM%3d&tstamp=1470211737&id=EDF9C6D450C63A1783E59B221A4594EA47F20AFF> [2015, June 11].
- Khan, H.M., Chan, G. & Chua, F. 2016. An adaptive monitoring framework for ensuring accountability and quality of services in cloud computing. *The 30th International Conference on Information Networking (ICOIN)* 13-15 January 2016. Kota Kinabalu, Malaysia [Electronic]. Available:  
<http://ieeexplore.ieee.org.ez.sun.ac.za/stamp/stamp.jsp?tp=&arnumber=7427071> [2016, June 19].

- Kitchin, R. 2014. *The data revolution: big data, open data, data infrastructure & their consequences*. London: SAGE Publications Ltd [Online]. Available: <http://srmo.sagepub.com.ez.sun.ac.za/view/the-data-revolution/n9.xml> [2015, December 22].
- Kneller, M. 2010. *Executive Briefing: The Benefits of ITIL*. The Stationery Office [Online]. Available: [http://addingvalue.se/wp-content/uploads/Executive\\_Briefing\\_Benefits\\_of\\_ITIL.pdf](http://addingvalue.se/wp-content/uploads/Executive_Briefing_Benefits_of_ITIL.pdf) [2015, November 25].
- Knoblock, C.A. & Szekely, P. 2015. Exploiting Semantics for Big Data Integration. *AI magazine* [Electronic], 36(1):25-38. Available: <http://search.proquest.com.ez.sun.ac.za/docview/1667668090?OpenUrlRefId=info:xri/sid:wcdiscovery&accountid=14049> [2016, March 12].
- KPMG International Cooperative. 2013. *The Road to Transition: COSO's Internal Control 2013 – Integrated Framework* [Online]. Available: <http://www.kpmg.com/sg/en/issuesandinsights/articlespublications/pages/advisory-rc-the-road-to-transition.aspx> [2015, November 26].
- Krechovská, M. & Procházková, P.T. 2014. Sustainability and its Integration into Corporate Governance Focusing on Corporate Performance Management and Reporting. *Procedia Engineering* [Electronic], 69:1144-1151. Available: [http://ac.els-cdn.com/S187770581400349X/1-s2.0-S187770581400349X-main.pdf?\\_tid=f684d5de-718c-11e4-9db3-00000aab0f01&acdnat=1416581277\\_1d7ed0e3af4214419aa42383c171ea1f](http://ac.els-cdn.com/S187770581400349X/1-s2.0-S187770581400349X-main.pdf?_tid=f684d5de-718c-11e4-9db3-00000aab0f01&acdnat=1416581277_1d7ed0e3af4214419aa42383c171ea1f) [2015, August 5].
- Kshetri, N. 2014. Big data's impact on privacy, security and consumer welfare. *Telecommunications Policy* [Electronic], 38(11):1134–1145. Available: <http://www.sciencedirect.com.ez.sun.ac.za/science/article/pii/S0308596114001542> [2015, June 2].
- Küller, P., Grabowski, M., PetrSameš & Vogt, M. 2010. *IT Service Management Methods and Frameworks Systematization*. Innotrain IT [Online]. Available: [http://www.central2013.eu/fileadmin/user\\_upload/Downloads/outputlib/Innotrain\\_Systematization\\_2011\\_04\\_05\\_FINAL.PDF](http://www.central2013.eu/fileadmin/user_upload/Downloads/outputlib/Innotrain_Systematization_2011_04_05_FINAL.PDF) [2015, March 27].
- Kusumah, P., Sutikno, S. & Rosmansyah, Y. 2014. Model Design of Information Security Governance Assessment with Collaborative Integration of COBIT 5 and ITIL. *2014 International Conference on ICT for Smart Society*, 23-24 September, 1-6 [Electronic]. Available: <http://ieeexplore.ieee.org.ez.sun.ac.za/stamp/stamp.jsp?tp=&arnumber=7013193> [2015, October 26].
- Lange, N.D., Thomas, R.P & Davelaar, E.J. 2013. Data Acquisition Dynamics and Hypothesis Generation. *Cognitive Systems Research* [Electronic], 24, September:9-17. Available: <http://www.sciencedirect.com.ez.sun.ac.za/science/article/pii/S1389041712000599> [2015, June 19].
- Levy, Y. & Ellis, T.J. 2006. A Systems Approach to Conduct an Effective Literature Review in Support of Information Systems Research. *Informing Science* [Electronic], 9. Available: <http://go.galegroup.com.ez.sun.ac.za/ps/i.do?issn=1547-9684&v=2.1&u=27uos&it=Jlourl&p=AONE&sw=w&authCount=1> [2016, October 12].

- Liell-Cock, S., Graham, J. & Hill, P. 2009. *IT Governance aligned to KING III* [Online]. Available: <http://lgict.org.za/document/it-governance-aligned-king-iii> [2016, April 7].
- Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C. & Byers, A.H. 2011. *Big data: The next frontier for innovation, competition and productivity*. McKinsey & Company [Online]. Available: <http://www.mckinsey.com/business-functions/business-technology/our-insights/big-data-the-next-frontier-for-innovation> [2015, December 23].
- Markham, S.K., Kowolenko, M. & Michaelis, T.L. 2015. Unstructures text analytics to support new product development decisions. *Research Technology Management* [Electronic], 58(2):30-38. Available: <http://web.a.ebscohost.com.ez.sun.ac.za/ehost/pdfviewer/pdfviewer?vid=2&sid=8dfef89a-d081-4e01-a048-797b1616cd57%40sessionmgr4001&hid=4207> [2015, July 7].
- Marks, N. 2010. The pulse of IT Governance. *Internal Auditor* [Electronic], 67(4):32-37. Available: <http://web.b.ebscohost.com.ez.sun.ac.za/ehost/pdfviewer/pdfviewer?vid=2&sid=ad3a892e-612d-492f-b69b-fb19b77cc1ab%40sessionmgr111&hid=101> [2015, August 6].
- Mattsson, U.T. 2014. Bridging the gap between access and security in big data. *ISACA Journal*, 6, 2014. [Online]. Available: <http://www.isaca.org/Journal/archives/2014/Volume-6/Pages/Bridging-the-Gap-Between-Access-and-Security-in-Big-Data.aspx> [2016, January 15].
- McAfee, A. & Brynjolfsson, E. 2012. Big Data: The management revolution. *Harvard Business Review* [Electronic], 90(10):60-68. Available: <http://web.a.ebscohost.com.ez.sun.ac.za/ehost/pdfviewer/pdfviewer?sid=a61b58b6-b187-4ccb-9b0f-bbcd36c55490%40sessionmgr4006&vid=4&hid=4114> [2015, July 7].
- McNally, J.S. 2012. COSO Framework Holding Strong and Getting a Polish. *Pennsylvania CPA Journal* [Electronic], 83(2):1-6. Available: <http://web.a.ebscohost.com.ez.sun.ac.za/ehost/pdfviewer/pdfviewer?vid=2&sid=705bcb86-bd9b-4909-b2f5-dcd0ab72a573%40sessionmgr4009&hid=4214> [2015, November 26].
- Miller, M.J. 2012. *Storing Massive data: Distributed data and the noSQL movement* [Online]. Available: <http://forwardthinking.pcmag.com/pc-hardware/297512-storing-massive-data-distribution> [2015, June 4].
- Nelson, G.S. 2015. *Practical Implications of Sharing Data: A Primer on Data Privacy, Anonymization, and De-Identification* [Online]. Available: <http://support.sas.com/resources/papers/proceedings15/1884-2015.pdf> [2016, June 29].
- Nielsen, P. 2013. *Integrated Health Information Architecture*. Department of Informatics, UiO, 58-82 [Online]. Available: <https://www.mn.uio.no/ifi/english/research/networks/hisp/integrated-health-information-architecture/ch-03.pdf> [2016, May 17].
- Noor, A.K. 2015. Potential of Cognitive Computing and Cognitive Systems. *Open Engineering* [Electronic], 5(1):75-88. Available: <http://www.degruyter.com.ez.sun.ac.za/view/j/eng.2015.5.issue-1/eng-2015-0008/eng-2015-0008.xml> [2015, October 26].

- Oberlin, S. 2012. *Machine Learning, Cognition, and Big Data*. [Online]. Available: <http://www.ca.com/us/~media/files/articles/ca-technology-exchange/machine-learning-cognition-and-big-data-oberlin.aspx> [2015, June 03].
- Oracle. 1997. *Oracle8 Parallel Server Concepts & Administration* [Online]. Available: [http://docs.oracle.com/cd/A58617\\_01/server.804/a58238/ch1\\_unde.htm](http://docs.oracle.com/cd/A58617_01/server.804/a58238/ch1_unde.htm) [2015, June 11].
- Panackala, J.J & Pillaib., A.S. 2015. Adaptive Utility-based Anonymization Model: Performance Evaluation on Big Data Sets. *Procedia Computer Science* [Electronic], 50:347-352. Available: <http://www.sciencedirect.com/science/article/pii/S1877050915005384> [2016, June 2].
- Paryasto, M., Alamsyah, A. & Kuspriyanto, B.R. 2014. Big-Data Security Management Issues. *2014 2nd International Conference on Information and Communication Technology (ICoICT)*. 28-30 May, 59-63. Bandung, Indonesia [Electronic]. Available: <http://ieeexplore.ieee.org.ez.sun.ac.za/stamp/stamp.jsp?tp=&arnumber=6914040> [2016, April 8].
- Peña, J.J.S., Vicente, E.F. & Ocaña, A.M. 2013. ITIL, COBIT and EFQM: Can They Work Together? *International Journal of Combinatorial Optimization Problems and Informatics* [Electronic], 4(1):54-64. Available: <http://ijcopi.org/ojs/index.php?journal=ijcopi&page=article&op=view&path%5B%5D=114&path%5B%5D=172> [2015, August 6].
- Philip Chen, C.L. & Zhang, C. 2014. Data-intensive applications, challenges, techniques and technologies: A survey on Big Data. *Information Sciences* [Electronic], 275 (2014):314-347. Available: [http://ac.els-cdn.com.ez.sun.ac.za/S0020025514000346/1-s2.0-S0020025514000346-main.pdf?\\_tid=db91a82a-9f11-11e5-81a0-00000aab0f6b&acdnat=1449733655\\_fcdbd1569f29ce60d82ab06432d4ea9d](http://ac.els-cdn.com.ez.sun.ac.za/S0020025514000346/1-s2.0-S0020025514000346-main.pdf?_tid=db91a82a-9f11-11e5-81a0-00000aab0f6b&acdnat=1449733655_fcdbd1569f29ce60d82ab06432d4ea9d) [2015, June 1].
- Protection of Personal Information Act No 4 of 2013 (RSA).
- PWC. 2011. *The protection of personal information bill: The journey to implementation* [Online]. Available: [http://reference.sabinet.co.za.ez.sun.ac.za/webx/access/electronic\\_journals/sajaar/sajaar\\_v17\\_n1\\_a2.pdf](http://reference.sabinet.co.za.ez.sun.ac.za/webx/access/electronic_journals/sajaar/sajaar_v17_n1_a2.pdf)<http://www.pwc.co.za/en/assets/pdf/popii-white-paper-2011.pdf>. [2016, March 10].
- PWC. n.d. *King III, IT governance and your organisation* [Online]. Available: <https://www.pwc.co.za/en/assets/pdf/steeringpoint-kingiii-it-governance-and-kingiii-15.pdf> [2015, August 6].
- Rezaeibagha, F., Win, K.T. & Susilo, W. 2015. A systematic literature review on security and privacy of electronic health record systems: technical perspectives. *Health Information Management Journal* [Electronic], 44(3):23-38. Available: <http://web.b.ebscohost.com.ez.sun.ac.za/ehost/pdfviewer/pdfviewer?vid=28&sid=116c450f-ea11-481d-8559-1537752786ad%40sessionmgr102&hid=105> [2016, June 17].
- Rittenberg, L.E. 2013. COSO 2013. *Internal Auditor* [Electronic], 70(4):60-65. Available: <http://web.a.ebscohost.com.ez.sun.ac.za/ehost/pdfviewer/pdfviewer?sid=eb097a8b-29cb-4feb-84f6-9023d674bb69%40sessionmgr4009&vid=3&hid=4214> [2015, November 24].

- Ritter, J. 2016. *Regulating big data: Monitoring systems to create new wealth* [Online]. Available: [http://searchcompliance.techtarget.com/tip/Regulating-big-data-Monitoring-systems-to-create-new-wealth?utm\\_content=recipe4&utm\\_medium=EM&asrc=EM\\_ERU\\_59403853&utm\\_campaign=20160621\\_ERU%20Transmission%20for%2006/21/2016%20\(UserUniverse:%202091933\)\\_myka-reports@techtarget.com&utm\\_source=ERU&src=5524211](http://searchcompliance.techtarget.com/tip/Regulating-big-data-Monitoring-systems-to-create-new-wealth?utm_content=recipe4&utm_medium=EM&asrc=EM_ERU_59403853&utm_campaign=20160621_ERU%20Transmission%20for%2006/21/2016%20(UserUniverse:%202091933)_myka-reports@techtarget.com&utm_source=ERU&src=5524211) [2016, June 21].
- Ronanki, R. & Steier, D. 2014a. *Human Brain inspires new cognitive analytics*. [Online]. Available: <http://deloitte.wsj.com/cio/2014/05/13/human-brain-inspires-new-cognitive-analytics> [2015, April 5].
- Ronanki, R. & Steier, D. 2014b. *Tech Trend 2014*. Deloitte University Press: 20-29. [Online]. Available: [http://dupress.com/wp-content/uploads/2014/02/Tech-Trends-2014\\_FINAL-ELECTRONIC\\_single.2.24.pdf](http://dupress.com/wp-content/uploads/2014/02/Tech-Trends-2014_FINAL-ELECTRONIC_single.2.24.pdf) [2015, April 5].
- Rubino, M. & Vitolla, F. 2014a. Coporate governance and the information system: how a framework for IT governance supports ERM. *Corporate Governance* [Electronic], 14(3):320-338. Available: <http://search.proquest.com.ez.sun.ac.za/docview/1658479811/fulltextPDF/C290062E04824BAAPQ/3?accountid=14049> [2014, July 2].
- Rubino, M. & Vitolla, F. 2014b. Internal control over financial reporting: oppotunities using the COBIT framework. *Managerial Auditing Journal* [Electronic], 29(8):736-771. Available: <http://search.proquest.com.ez.sun.ac.za/docview/1660950575/fulltextPDF/6FDE82448DD14643PQ/4?accountid=14049> [2015, October 26].
- Rudman, R.J. 2008. IT Governance: A New Era. *Accountancy SA*, March 2008:12-14.
- Rudman, R. 2010. Framework to identify and manage risks in Web 2.0 applications. *African Journal of Business Management* [Electronic], 4(13):3251-3264. Available: <http://search.proquest.com.ez.sun.ac.za/docview/1663919262/A47743610F67408FPQ/4?accountid=14049> [2015, November 23].
- Sahd, L. 2015. A structured approach to the identification of significant risks related to enterprise mobile solutions at a mobile technology component level. Unpublished master's thesis. Stellenbosch: Stellenbosch University.
- Sahibudin, S., Sharifi, M. & Ayat, M. 2008. Combining ITIL, COBIT and ISO/IEC 27002 in order to design a comprehensive IT framework in organizations. *AMS 2008 Second Asia International Conference on Modelling & Simulation*. 13-15 May, 749-753. Kuala Lumpur: Malaysia [Electronic]. Available: <http://ieeexplore.ieee.org.ez.sun.ac.za/stamp/stamp.jsp?arnumber=4530569> [2015, October 26].
- Salido, J. 2010. Data Governance for Privacy, Confidentiality and Compliance: A Holistic Approach. *ISACA Journal*, 6, 2010. [Online]. Available: <http://www.isaca.org/Journal/archives/2010/Volume-6/Documents/jpdf1006-data-governance-for.pdf> [2016, July 2].
- Sarkar, S. & Zaharchuk, D. 2015. *Your cognitive future*. IBM Corporation [Online]. Available: <http://www-01.ibm.com/common/ssi/cgi->

bin/ssialias?subtype=XB&infotype=PM&appname=CB\_BU\_B\_CBUE\_GB\_TI\_USEN&htmlfid=GBE03641USEN&attachment=GBE03641USEN.PDF [2015, June 15].

Schatsky, D., Muraskin, C. & Gurusurthy, R. 2014. *Demystifying artificial intelligence*. Deloitte University Press: 1-13 [Online]. Available: [http://dupress.com/wp-content/uploads/2014/02/Tech-Trends-2014\\_FINAL-ELECTRONIC\\_single.2.24.pdf](http://dupress.com/wp-content/uploads/2014/02/Tech-Trends-2014_FINAL-ELECTRONIC_single.2.24.pdf) [2015; April 5].

SAS Institute. n.d *Machine Learning: What it is and why it matters*. [Online]. Available: [http://www.sas.com/en\\_us/insight/analytics/machine-learning.html](http://www.sas.com/en_us/insight/analytics/machine-learning.html) [2015, July 2].

Schneider, R.D. 2012. *Hadoop For Dummies, Special Edition*. Canada: John Wiley & Sons, Inc.

Siegel, E. 2013. *Predictive Analytics: The Power to Predict Who Will Click, Buy, Lie or Die*. Wiley [Online] Available: <http://web.a.ebscohost.com.ez.sun.ac.za/ehost/detail/detail?sid=bda49be2-64a4-44ba-a6ab-202e6845c8f7%40sessionmgr4005&vid=0&hid=4207&bdata=JnNpdGU9ZWhvc3QtOGl2ZS5zY29wZT1zaXRl#AN=535996&db=nlebk> [2015, July 7].

Smitha, R., Suma, S.N. & Sunitha, M. 2015. Security Solutions For Big Data Analytics In Healthcare. *2015 Second International Conference on Advances in Computing and Communication Engineering*. 1-2 May, 510-514. Dehradun, India [Electronic]. Available: <http://ieeexplore.ieee.org.ez.sun.ac.za/stamp/stamp.jsp?tp=&arnumber=7306738> [2016, April 18].

Sudarsan, S. 2013. *An Ecosystem of Innovation: Creating Cognitive Applications Powered by Watson*. [Somers, New York]: IBM Corporation [Online]. Available: <https://developer.ibm.com/watson/wp-content/uploads/sites/19/2013/11/An+Ecosystem+Of+Innovation+-+Creating+Cognitive+Applications+PoweredByWatson.pdf> [2015, March 7].

Suer, M. & Nolan, R. 2015. Using Cobit to deliver information and data governance. *COBIT Focus*, 12 January 2015 [Online]. Available: <http://www.isaca.org/cobit/focus/pages/using-cobit-5-to-deliver-information-and-data-governance.aspx> [2015, August 6].

Sylvester, A., Tate, M. & Johnstone, D. 2013. Beyond Synthesis: re-presenting heterogeneous research literature. *Behaviour & Information Technology* [Electronic], 32(12):1199-1215. Available: <http://web.b.ebscohost.com.ez.sun.ac.za/ehost/pdfviewer/pdfviewer?sid=9803fdb3-867f-493a-befb-243b59cfd35e%40sessionmgr102&vid=1&hid=128> [2015, April 7].

Tech Target. n.d. *Top Data Management Terms to Know* [Online]. Available: [http://cdn.ttgtmedia.com/CascadingTargetedDownloads/downloads/DataManagement\\_TopTerms\\_Eguide\\_updated.pdf](http://cdn.ttgtmedia.com/CascadingTargetedDownloads/downloads/DataManagement_TopTerms_Eguide_updated.pdf) [2015, July 7].

Terzi, D.S., Terzi, R. & Sagiroglu, S. 2015. A Survey on Security and Privacy Issues in Big Data. *The 10th International Conference for Internet Technology and Secured Transactions (ICITST-2015)*. 14-16 December. London, UK [Electronic]. Available: <http://ieeexplore.ieee.org.ez.sun.ac.za/stamp/stamp.jsp?tp=&arnumber=7412089> [2016, June 1].

Trites, G. 2013. *Information integrity*. [New York]: American Institute of CPAs [Online]. <https://www.aicpa.org/interestareas/frc/assuranceadvisoryservices/downloadabledocuments/a-sec-information-integrity-white-paper.pdf> [2016, March 4].

Van der Veer, H. & Wiles, A. 2008. *Achieving Technical Interoperability – the ETSI approach*. ETSI [Online]. Available: <http://www.etsi.org/WebSite/document/whitepapers/IOP%20whitepaper%20Edition%203%20final.pdf> [2015, December 17].

Vaughan, J. & Loshin, D. 2014. *The Hitchhiker's Guide to Hadoop 2*. TechTarget [Online]. Available: [https://docs.google.com/viewer?url=http://tngconsultores.com/kw/pluginfile.php/53/mod\\_glossary/attachment/880/Hitchhikers%20Guide%20to%20Hadoop%202.pdf](https://docs.google.com/viewer?url=http://tngconsultores.com/kw/pluginfile.php/53/mod_glossary/attachment/880/Hitchhikers%20Guide%20to%20Hadoop%202.pdf) [2015, July 23].

Wang, Y. 2009. On Cognitive Computing. *International Journal of Software Science and Computational Intelligence*, 1(3):1-15.

Wang, Y. 2011. Towards the synergy of Cognitive Informatics, Neural Informatics, Brain Informatics and Cognitive Computing. *International Journal of Cognitive Informatics and Natural Intelligence* [Electronic], 5(1):75-93. Available: <http://go.galegroup.com.ez.sun.ac.za/ps/publicationSearch.do?lm=&inPS=true&prodId=AONE&userGroupName=27uos&method=doLinkDirectedSearch&searchType=AdvancedSearchForm&qt=PU%7E%22International+Journal+of+Cognitive+Informatics+and+Natural+Intelligence%22%7E%7EDA%7E120110101%7E%7EIU%7E%221%22%7E%7EVO%7E5> [2015, June 1]

Webster, J. & Watson, R.T. 2002. Analyzing the past to prepare for the future: Writing a literature review. *MIS Quarterly* [Electronic], 26(2):xiii-xxiii. Available: [http://www.jstor.org.ez.sun.ac.za/stable/4132319?seq=1#page\\_scan\\_tab\\_contents](http://www.jstor.org.ez.sun.ac.za/stable/4132319?seq=1#page_scan_tab_contents) [2016, July 7].

Wigan, M.R. & Clarke, R. 2013. Big data's unintended consequences. *Computer* [Electronic], 46(6):46-53. Available: <http://ieeexplore.ieee.org.ez.sun.ac.za/stamp/stamp.jsp?tp=&arnumber=6527249> [2016, February 16].

Willis Towers Watson. 2016. *TechTalk: Cognitive Computing* [Online]. Available: [http://www.willis.com/documents/publications/Industries/Technology\\_and\\_Telecomm/15644%20PUBLICATION\\_TMT%20Cognitive%20Computing.pdf](http://www.willis.com/documents/publications/Industries/Technology_and_Telecomm/15644%20PUBLICATION_TMT%20Cognitive%20Computing.pdf) [2016, July 16].

Wolff, G.J. 2014. Big Data and the SP Theory of Intelligence. *IEEE Access* [Electronic], 2: 301-315. Available: <http://ieeexplore.ieee.org.ez.sun.ac.za/stamp/stamp.jsp?tp=&arnumber=6782396> [2015, June 29].

Wladawsky-Berger, I. 2013, July 1. The Era of Cognitive Computing. *Irving Wladawsky-Berger* [Web log post]. Available: <http://blog.irvingwb.com/blog/2013/07/the-dawn-of-a-new-era-in-computing.html> [2015, August 23].

Yuhanna, N. 2013. *The Steadily Growing Database Market Is Increasing Enterprises' Choices*. [Cambridge]: Forrester Research [Online]. Available:

[http://docs.media.bitpipe.com/io\\_11x/io\\_115609/item\\_926011/Forrester%20Research%20-%20New%20Database%20Choices%20for%20Enterprise%20IT.pdf](http://docs.media.bitpipe.com/io_11x/io_115609/item_926011/Forrester%20Research%20-%20New%20Database%20Choices%20for%20Enterprise%20IT.pdf) [2015, August 2].

Zaino, J. 2014. *Bringing Clarity to the Topic of Cognitive Computing* [Online]. Available: <http://www.datavercity.net/bringing-clarity-topic-cognitive-computing> [2015, June 17 June].

Zalewska, A. 2014. Challenges of Corporate Governance: Twenty years after Cadbury, ten years after Sarbanes-Oxley. *Journal of Empirical Finance* [Electronic], 27:1-9. Available: [http://ac.els-cdn.com.ez.sun.ac.za/S092753981300100X/1-s2.0-S092753981300100X-main.pdf?\\_tid=feb99ff8-9f0f-11e5-811c-00000aab0f6b&acdnat=1449732855\\_7851aa67201dba7982a99cfaf0e9a8d5](http://ac.els-cdn.com.ez.sun.ac.za/S092753981300100X/1-s2.0-S092753981300100X-main.pdf?_tid=feb99ff8-9f0f-11e5-811c-00000aab0f6b&acdnat=1449732855_7851aa67201dba7982a99cfaf0e9a8d5) [2015, August 5].

Zhang, S. & Le Fever, H. 2013. An Examination of the Practicability of COBIT Framework and the Proposal of a COBIT-BSC Model. *Journal of Economics, Business and Management* [Electronic], 1(4):391-395. Available: <http://www.joebm.com/papers/84-M021.pdf> [2015, August 6].

Zikopoulos, P., deRoos, D., Bienko, C., Buglio, R. & Andrews, M. 2015. *Big data beyond the hype: A guide to Conversations for Today's Data Center*. McGraw-Hill Education. [Online]. Available: <http://www.ibmbluhub.com/big-data-ebook/> [2015, July 10].

**APPENDIX A: Identifying risk by means of COBIT 5's detailed processes**

COBIT 5 was used to identify the risks pertaining to the implementation of a cognitive computing system and formulate mitigating internal control techniques to address these risks. The 37 processes of COBIT 5, summarised in column two and three, were applied to identify specific risks the enterprise is exposed to and to evaluate the impact of these risks on the enterprise. The risks are summarised in column four, and the risks were rated as High (H); Medium (M) or Low (L) in column five. The relevant internal controls formulated to address each risk are summarised in column six.

Domain	Relevant COBIT process	Detail processes' requirements	Risk(s) identified	Impact of the risk	Control(s) to mitigate the risk(s)
Evaluate, Direct and Monitor	EDM01	Evaluate, direct and monitor the governance system	<ul style="list-style-type: none"> <li>• An absence of effective governance strategies and policies in the cognitive computing environment.</li> <li>• Cognitive computing strategies and policies are not comprehensive and efficient; and not well documented.</li> <li>• Poor implementation of policies and procedures, as well as a lack of stewardship and ownership of cognitive computing strategies and policies.</li> <li>• Inadequate involvement from the Board.</li> <li>• Cognitive computing governance is not monitored for effectiveness and performance.</li> </ul>	High	<ul style="list-style-type: none"> <li>• The Board must design and implement a governance system which ensures that: <ul style="list-style-type: none"> <li>○ all stakeholders are identified and their requirements are obtained and documented;</li> <li>○ comprehensive cognitive computing governance policies are established;</li> <li>○ governance policies and procedures are reviewed and monitored to enable improvement;</li> <li>○ authority, responsibility and decision-making regarding investment in and use of cognitive computing is established and communicated.</li> </ul> </li> </ul>

Domain	Relevant COBIT process	Detail processes' requirements	Risk(s) identified	Impact of the risk	Control(s) to mitigate the risk(s)
<b>Evaluate, Direct and Monitor</b>	EDM02	Evaluate, direct and monitor the value contribution from the cognitive computing system to the enterprise.	<ul style="list-style-type: none"> <li>The cost of the investments in cognitive computing and use there off exceeds its benefits.</li> <li>Insufficient focus and monitoring of value delivery.</li> </ul>	High	<ul style="list-style-type: none"> <li>Compile a comprehensive cost-benefit analysis and budget to measure and manage the investment in and return from cognitive computing</li> </ul>
	EDM03	Evaluate, direct and monitor the enterprise's risk appetite, risk tolerance and risk exposure with regard to the cognitive computing system.	<ul style="list-style-type: none"> <li>All risks relating to the utilisation of cognitive computing are not identified and identified risks are not mitigated appropriately to reduce risk to an acceptable risk tolerance level. Thus the risk management process is ineffective.</li> <li>All governance and legal requirement are not identified and mitigated.</li> <li>Inadequate business resilience arrangements.</li> </ul>	High	Develop and implement a risk management system that: <ul style="list-style-type: none"> <li>Establishes and documents processes for risk identification, risk assessment and risk response;</li> <li>assigns the responsibility;</li> <li>addresses legal and regulatory compliance risks; and</li> <li>addresses the management of changes in risks and disaster recovery.</li> </ul>
	EDM04	Evaluate, direct and monitor if the cognitive computing system is capable to effectively support enterprise objectives at an optimal cost.	<ul style="list-style-type: none"> <li>The cognitive computing system and the resources utilised are not effectively managed and used.</li> <li>Misallocation of resources occur and are not identified.</li> </ul>	High	<ul style="list-style-type: none"> <li>Perform a resource gap analysis to establish if sufficient resources are available.</li> <li>Establish ownership and accountability for resource investment.</li> <li>Establish performance measures to evaluate and monitor the optimisation of allocated resources.</li> </ul>

Domain	Relevant COBIT process	Detail processes' requirements	Risk(s) identified	Impact of the risk	Control(s) to mitigate the risk(s)
Evaluate, Direct and Monitor	EDM05	Evaluate, direct and monitor stakeholder requirements and communication.	<ul style="list-style-type: none"> <li>Miscommunication between stakeholders involved in the investment in and use of cognitive computing exist.</li> </ul>	High	<ul style="list-style-type: none"> <li>Refer EDM01</li> </ul>
Align, Plan and Organise	APO01	Align the cognitive computing investments and use with the enterprise strategies and objectives. Define and establish the right mechanisms and authorities to achieve alignment.	<ul style="list-style-type: none"> <li>Cognitive computing objectives are not aligned with enterprise objectives.</li> <li>The necessary organisational structures are not established.</li> <li>Ownership of cognitive computing policies and procedures are not assigned.</li> <li>Cognitive computing policies are insufficient.</li> <li>Cognitive computing investments do not create value is.</li> <li>Data ownership of new data and information produced by the cognitive computing system are not established or controlled.</li> <li>Continual improvement of cognitive systems and procedures are hampered by insufficient monitoring and management.</li> </ul>	High	<ul style="list-style-type: none"> <li>Define enterprise and cognitive computing strategies and objectives; and ensure alignment between the strategies and objectives.</li> <li>Establish cognitive computing strategy, policies and procedures to support alignment.</li> <li>Communicate the cognitive computing strategy and policies to the stakeholders and establish their roles and responsibilities.</li> <li>Establish and implement data provenance standards and rules throughout the data lifecycle in the cognitive computing system.</li> <li>Implement a system of continuous monitoring and improvement of the strategy, policies and procedures.</li> </ul>

Domain	Relevant COBIT process	Detail processes' requirements	Risk(s) identified	Impact of the risk	Control(s) to mitigate the risk(s)
Align, Plan and Organise	APO02	Establish a cognitive computing strategy and leverage current IT infrastructure.	<ul style="list-style-type: none"> <li>The cognitive computing road map is inadequate.</li> <li>The infrastructure supporting the cognitive computing system is not sufficient, scalable or compatible.</li> </ul>	High	<ul style="list-style-type: none"> <li>Design a cognitive computing road map which defines the objective of the cognitive computing system, establishes user requirements, identify the required cognitive computing components and establishes a development plan</li> <li>Establish the IT infrastructure required to support the cognitive computing strategy.</li> <li>Perform a maturity analysis to assess the ability of the current infrastructure.</li> <li>Perform a gap analysis to identify shortcomings between the required infrastructure and the current infrastructure.</li> </ul>
	APO03	Identify and define cognitive computing component requirements and establish policies, procedures and standards for the implementation of the cognitive computing components.	<ul style="list-style-type: none"> <li>The cognitive computing infrastructure and components are not sufficient for the achievement of the cognitive computing objective and strategy.</li> <li>New investments in cognitive computing components are not managed effectively leading to excessive costs.</li> </ul>	High	<ul style="list-style-type: none"> <li>Define and implementing procedures and controls to manage and monitor the IT infrastructure and related services.</li> <li>Implement a change management process and an infrastructure migration plan.</li> <li>Use cognitive computing platforms, cloud computing platforms and data platforms to increase scalability and integration.</li> </ul>

Domain	Relevant COBIT process	Detail processes' requirements	Risk(s) identified	Impact of the risk	Control(s) to mitigate the risk(s)
Align, Plan and Organise	APO03		<ul style="list-style-type: none"> <li>• Cognitive computing infrastructure (including storage, access, processing, management and transmission) is not scalable and cannot integrate.</li> <li>• Opportunities to advance enterprise operations are missed due to ineffective management.</li> <li>• The algorithms used in the cognitive system is not scalable.</li> <li>• Insufficient technical and semantics interoperability.</li> </ul>		<ul style="list-style-type: none"> <li>• Leverage mature integration tools or the cognitive computing components to Standardise and /or translate data into a universal form / presentation for effective data integration and to improve interoperability.</li> <li>• Refer APO02, APO04</li> </ul>
	APO04	Manage cognitive computing innovations.	<ul style="list-style-type: none"> <li>• New solutions and opportunities are missed due to ineffective management.</li> </ul>	Medium	<ul style="list-style-type: none"> <li>• Develop procedures for the identification of new areas of innovation.</li> <li>• Establishing a Centre of Excellence (COE) to facilitate the mobilisation of resources for the cognitive computing initiatives.</li> <li>• Evaluate the data resources the enterprise owns and which additional data resources are required to create new opportunities for insight.</li> </ul>

Domain	Relevant COBIT process	Detail processes' requirements	Risk(s) identified	Impact of the risk	Control(s) to mitigate the risk(s)
Align, Plan and Organise	APO05	Identify required investments and manage, monitor and prioritise the portfolio of investments based on the availability of resources, as well as risks and rewards.	<ul style="list-style-type: none"> <li>Refer EDM04, APO01 and APO03</li> </ul>	Medium	<ul style="list-style-type: none"> <li>Refer EDM04, APO01 and APO03</li> </ul>
	APO06	Identify and manage budgets, costs and benefits related to the investment in and use of cognitive computing.	<ul style="list-style-type: none"> <li>The costs related to the investment in infrastructure and management of the cognitive computing components are substantial and exceeds budgets.</li> <li>The investment required to develop the cognitive computing system is extensive.</li> <li>The investment in experts, personnel changes and retraining is substantial.</li> <li>Costs required to ensure security and privacy of data is considerable and deviates from the budget.</li> </ul>	High	<ul style="list-style-type: none"> <li>Use cognitive computing platforms, cloud computing platforms and data platforms to reduce and control development cost; infrastructure cost; human skills cost; and security and privacy costs.</li> <li>Determining the resources and investment necessary to create the appropriate IT infrastructure to support the cognitive computing system and prepare a budget.</li> <li>Establish performance measures to evaluate and monitor the optimisation of resources.</li> <li>Reduce cost by integrating the cognitive computing system with the existing IT environment and extending current controls and processes into system.</li> </ul>
	APO07	Identify and manage key IT employees, maintain required skills and competencies and evaluate employee performance.	<ul style="list-style-type: none"> <li>Shortage of experts, scientists and other IT personnel with the required technical skill sets and experience.</li> </ul>	High	<ul style="list-style-type: none"> <li>Include human skills and resource requirements in the cognitive computing strategy and governance program.</li> </ul>

Domain	Relevant COBIT process	Detail processes' requirements	Risk(s) identified	Impact of the risk	Control(s) to mitigate the risk(s)
Align, Plan and Organise	APO07				<ul style="list-style-type: none"> <li>• Perform a gap analysis to identify potential skill and resource gaps.</li> <li>• Provide targeted training for existing employees.</li> <li>• Hire new talent and leverage consulting firms.</li> <li>• Form partnerships with vendors and service providers involved in cognitive computing.</li> <li>• Establish a Centre of excellence to facilitate: <ul style="list-style-type: none"> <li>○ cross training</li> <li>○ communication between experts and IT teams</li> <li>○ a culture of trust</li> </ul> </li> <li>• Developing policies for the assessment, training and development of staff.</li> </ul>
	APO08	Manage the relationships between business and IT.	<ul style="list-style-type: none"> <li>• The cognitive computing strategies do not align with the enterprise strategies.</li> </ul>	High	<ul style="list-style-type: none"> <li>• Refer APO01 and APO02</li> </ul>
	APO09	Manage and maintain IT service delivery.	<ul style="list-style-type: none"> <li>• Service providers and suppliers do not provide the required skill set or services in accordance with enterprise requirements.</li> <li>• The IT services provided do not meet user requirement or are insufficient.</li> </ul>	High	<ul style="list-style-type: none"> <li>• Determining whether the cognitive computing system will be developed in-house, outsourced or by means of a cognitive platform. Establishing a usage policy, which identifies which components of the cognitive computing system should be supported by services from service providers.</li> </ul>

Domain	Relevant COBIT process	Detail processes' requirements	Risk(s) identified	Impact of the risk	Control(s) to mitigate the risk(s)
Align, Plan and Organise	APO09				<ul style="list-style-type: none"> <li>Establishing performance measures for outsourced services, compiling and reviewing service level agreements; assigning responsibility within the enterprise to monitor compliance with the service level agreement and establish controls to address security, change management, and access rights relating to the service providers.</li> <li>Refer APO02</li> </ul>
	APO10	Manage external service providers' relationships and monitor supplier performance.	<ul style="list-style-type: none"> <li>Service providers do not have adequate privacy and security policies, procedures and controls.</li> <li>Outsourcing services limits the control enterprises have over data.</li> </ul>	High	Refer APO09
	APO11	Establish standards and manage the quality of information in the cognitive computing system.	<ul style="list-style-type: none"> <li>Insufficient management of data quality.</li> <li>The data ingested into a cognitive computing system are not precise or faithful.</li> <li>Insecure programs corrupt data resulting in incorrect results.</li> <li>Ingested data cannot be validated creating data integrity issues.</li> </ul>	High	<ul style="list-style-type: none"> <li>Control data quality through: <ul style="list-style-type: none"> <li>Data cleaning software</li> <li>Data quality software</li> <li>Standardisation</li> <li>Data profiling</li> <li>Meta data management</li> </ul> </li> </ul>

Domain	Relevant COBIT process	Detail processes' requirements	Risk(s) identified	Impact of the risk	Control(s) to mitigate the risk(s)
Align, Plan and Organise	APO12	Identify, manage and reduce all risks relating to the cognitive computing system.	<ul style="list-style-type: none"> <li>Refer to EDM03</li> </ul>	High	<ul style="list-style-type: none"> <li>Establishing risk management policies and procedures for the continuous identification, monitoring and evaluation of new and emerging risk relating to the cognitive computing system</li> <li>Identify, monitor and manage supplier risks in order to ensure that the supplier has the ability to continually provide secure, efficient, effective and reliable service delivery.</li> </ul>
	APO13	Establish, maintain and monitor an information security and privacy management system.	<ul style="list-style-type: none"> <li>Inadequate policies and management of sensitive data.</li> <li>Refer DSS05</li> </ul>	High	<ul style="list-style-type: none"> <li>Establish data classification policies which define the purpose, ownership, and sensitivity of data types to ensure that sensitive information is managed according to the risks it poses to the enterprise.</li> <li>Establishing privacy policies which defines sensitive data and personally identifiable information; and addresses securing the data, transparency of usage, receiving and revocation of consent</li> <li>Refer DSS05</li> </ul>

Domain	Relevant COBIT process	Detail processes' requirements	Risk(s) identified	Impact of the risk	Control(s) to mitigate the risk(s)
Build, Acquire and Implement	BAI01	Manage and coordinate all the cognitive computing investments, programmes and projects.	<ul style="list-style-type: none"> <li>• Inadequate software development process:               <ul style="list-style-type: none"> <li>○ The development procedures are not adhere to the enterprise development standards.</li> <li>○ Third parties involved in the development do not adhere to contractual obligations and enterprise development standards.</li> <li>○ The changes during the development process are not authorised and monitored.</li> <li>○ The different stages of the development process are not and controlled and monitored for effectiveness and performance.</li> </ul> </li> </ul> <p>Refer EDM04</p>	High	<ul style="list-style-type: none"> <li>• Manage the development process by:               <ul style="list-style-type: none"> <li>○ assigning ownership of the project;</li> <li>○ establishing a development methodology which aligns with enterprise development standards;</li> <li>○ implementing quality assurance processes;</li> <li>○ implementing project risk management processes; and</li> <li>○ implementing change management processes.</li> </ul> </li> <li>• Develop a training and testing strategy Refer APO09</li> </ul>
	BAI02	Analyse, define and manage enterprise and user requirements for cognitive computing infrastructure and components.	<ul style="list-style-type: none"> <li>• The cognitive computing infrastructure and components do not meet the enterprise and user requirement.</li> </ul>	High	<ul style="list-style-type: none"> <li>• Refer APO02</li> </ul>

Domain	Relevant COBIT process	Detail processes' requirements	Risk(s) identified	Impact of the risk	Control(s) to mitigate the risk(s)
<b>Build, Acquire and Implement</b>	BAI03	Establish, build and maintain the cognitive computing system in line with the enterprise requirements and cognitive computing strategy.	<ul style="list-style-type: none"> <li>• The corpus is inadequate seeing as is too narrowly defined and do not include the right combination of relevant data resources.</li> <li>• External sources are trimmed or cleaned before they are imported into the corpus limiting discovery.</li> <li>• The incorrect algorithms are utilised.</li> <li>• Inconsistencies occur in the ontology or taxonomy development.</li> <li>• The cognitive system is not trained correctly.</li> <li>• Lack of integration between the different cognitive computing components.</li> <li>• Insufficient technical and semantic interoperability.</li> <li>• Refer APO02, APO03, APO08 and BAI02</li> </ul>	High	<ul style="list-style-type: none"> <li>• Determine the objective of the cognitive computing system and the type of question it will have to solve.</li> <li>• Define the domain or subject area for the cognitive computing system.</li> <li>• Based on the domain definition determine the domain experts needed to train and test the system.</li> <li>• Establish the user requirements.</li> <li>• Evaluate the data resources the enterprise owns and which additional data resources are required to create new opportunities for insight.</li> <li>• Determine the right combination of relevant data resources (internal and external) needed.</li> <li>• Determine the life cycle for each data source in order to establish which sources must be updated regularly and create a process to ensure that the updates are made on a timely basis.</li> <li>• Determine if data from the external sources should be cleaned or transformed before they are imported into the corpus.</li> <li>• Validate the ingested data to ensure that the data is readable, comprehensible and searchable.</li> </ul>

Domain	Relevant COBIT process	Detail processes' requirements	Risk(s) identified	Impact of the risk	Control(s) to mitigate the risk(s)
<b>Build, Acquire and Implement</b>	BAI03				<ul style="list-style-type: none"> <li>• Monitor the data ingestion process to ensure that the deletion of records for security purposes have been done.</li> <li>• Identify which machine learning algorithms and analysis techniques are best suited for the specific domain and specific question which must be solved.</li> <li>• Determine if a taxonomy or ontology is available for the domain or if a new taxonomy / ontology must be developed.</li> <li>• Monitor the development of the ontology or taxonomy in order to identify any inconsistent assumptions, beliefs and practices which may affect the corpus.</li> <li>• Determine the correct combination of algorithms which will enable the corpus to update and maintain the corpus itself.</li> </ul>
	BAI04	Assess and manage current availability and performance as well as future requirements.	<ul style="list-style-type: none"> <li>• The enterprise has insufficient resources to support the cognitive computing strategy.</li> </ul>	High	<ul style="list-style-type: none"> <li>• Refer EDM04</li> </ul>
	BAI05	Manage organisational change enablement.	<ul style="list-style-type: none"> <li>• Changes during the development process are not authorised and monitored.</li> </ul>	Medium	<ul style="list-style-type: none"> <li>• Implement change management processes.</li> </ul>

Domain	Relevant COBIT process	Detail processes' requirements	Risk(s) identified	Impact of the risk	Control(s) to mitigate the risk(s)
<b>Build, Acquire and Implement</b>	BAI06	Manage all changes in a controlled manner.	<ul style="list-style-type: none"> <li>Refer BAI05</li> </ul>	High	Refer BAI05
	BAI07	Establish an implementation plan and manage the implementation of new cognitive computing solutions.	<ul style="list-style-type: none"> <li>Refer BAI01 and BAI5</li> </ul>	High	Refer BAI01 and BAI5
	BAI08	Manage information and knowledge in the cognitive computing system to ensure that it is available, current, validated and reliable.	<ul style="list-style-type: none"> <li>Refer APO11</li> </ul>	High	Refer APO11
	BAI09	Manage all cognitive computing assets.	<ul style="list-style-type: none"> <li>Interference, modification or destruction of the cognitive computing system by a partial infrastructure breaches.</li> <li>Challenges in establishing access control across the distributed environments, including physical security of cognitive computing components including infrastructure, data networks, data applications, and data.</li> </ul>	High	<ul style="list-style-type: none"> <li>Establish and maintaining a logical model for the configuration of infrastructure items as well as regular software updates.</li> <li>Establish an Access Control List (ACL) to limit the access rights of system users and assigns the proper access rights.</li> <li>Utilise access control models, such as Role-Based Access Control (RBAC) or Attribute-Based Access Control (ABAC).</li> <li>Implementing physical security controls.</li> </ul>

Domain	Relevant COBIT process	Detail processes' requirements	Risk(s) identified	Impact of the risk	Control(s) to mitigate the risk(s)
Build, Acquire and Implement	BAI09		<ul style="list-style-type: none"> <li>• Software are not updated regularly.</li> <li>• EDM04</li> </ul>	High	<ul style="list-style-type: none"> <li>• Establishing policies to control inbound and outbound data traffic by implementing network filtering mechanisms such as firewalls, anti-malware and intrusion detection software.</li> </ul>
	BAI10	Manage configuration.	<ul style="list-style-type: none"> <li>• Insufficient documentation of system configurations.</li> <li>• Poor configuration controls.</li> </ul>	High	<ul style="list-style-type: none"> <li>• Establish and maintaining a logical model for the configuration of infrastructure items as well as regular software updates.</li> </ul>
Deliver, Service and Support	DSS01	Manage and coordinate cognitive computing operations.	<ul style="list-style-type: none"> <li>• Refer APO09 and APO10</li> <li>• Insufficient monitoring of cognitive computing components will leave additional requirements for cognitive computing system unidentified, and as such hamper continual improvement.</li> </ul>	High	<ul style="list-style-type: none"> <li>• Refer APO09</li> <li>• Service level agreement (SLA) must clearly define service requirements,</li> <li>• Monitor and review service to ensure it aligns with the SLA.</li> <li>• Identify which controls is relinquished to the provider and determine the specific monitoring controls which must be implemented due to this relinquishment.</li> <li>• Validate the control activities of the provider to ensure that they align with the enterprises risk appetite.</li> <li>• Periodically verify whether the controls maintained by the provider is effective and request independent reviews of the controls.</li> </ul>

Domain	Relevant COBIT process	Detail processes' requirements	Risk(s) identified	Impact of the risk	Control(s) to mitigate the risk(s)
<b>Deliver, Service and Support</b>	DSS01				<ul style="list-style-type: none"> <li>• Assess and monitor the ability of the supplier to provide adequate incident response and procedures to address system disruption and security breaches.</li> <li>• Assess and monitor the ability of the supplier to restore operations in the event of a disaster.</li> <li>• Establish an incident response plan and business continuity plan to support the suppliers plan.</li> <li>• Integrate key internal IT management processes with those of suppliers, specifically change management, configuration management, incident management, security management and business continuity management.</li> </ul>
	DSS02	Manage user request and resolve incidents.	<ul style="list-style-type: none"> <li>• Refer APO02 and DSS01</li> </ul>	High	<ul style="list-style-type: none"> <li>• Refer APO02 and DSS01</li> </ul>
	DSS03	Identify and manage the source of problems.	<ul style="list-style-type: none"> <li>• Incorrect answers and hypotheses due to:               <ul style="list-style-type: none"> <li>○ quality of data;</li> <li>○ lack of sufficient data;</li> <li>○ bias in training</li> </ul> </li> </ul>	High	<ul style="list-style-type: none"> <li>• Improve answers and hypotheses by:               <ul style="list-style-type: none"> <li>○ adding glossaries and ontologies to the corpus;</li> <li>○ testing sample data; and</li> <li>○ continually acquiring new data to update the corpus.</li> </ul> </li> </ul>

Domain	Relevant COBIT process	Detail processes' requirements	Risk(s) identified	Impact of the risk	Control(s) to mitigate the risk(s)
<b>Deliver, Service and Support</b>	DSS03		<ul style="list-style-type: none"> <li>○ cognitive computing models that do not capture relationships between similar data sources</li> </ul>		
	DSS04	Establish, implement and maintain business continuity plans.	<ul style="list-style-type: none"> <li>● Interference, modification or destruction of the cognitive computing system resulting in significant disruptions.</li> <li>○ Refer EDM03</li> </ul>	Medium	<ul style="list-style-type: none"> <li>● Establish an incident response plan and business continuity plan to support the supplier's plans.</li> <li>● Refer EDM03</li> </ul>
	DSS05	Establish and maintain security procedures and controls to ensure security, privacy and integrity of data.	<ul style="list-style-type: none"> <li>● Unauthorised access to sensitive, confidential and personal data.</li> <li>● Intentional security breaches through hacking, malware and phishing.</li> <li>● Distributed infrastructure and environment leads to security vulnerabilities.</li> <li>● Non-compliance with IT laws.</li> <li>● Insider breaches by privileged users.</li> <li>● Inadequate management of security.</li> <li>● Inadequate validation of data affecting the integrity of data.</li> </ul>	High	<ul style="list-style-type: none"> <li>● Implement privacy controls: <ul style="list-style-type: none"> <li>○ Anonymisation</li> <li>○ Tokenization</li> <li>○ Making</li> <li>○ Privacy preferences</li> </ul> </li> <li>● Implement privacy and security management: <ul style="list-style-type: none"> <li>○ Proactive management</li> <li>○ Data lifecycle control</li> <li>○ Monitoring system model</li> <li>○ Data activity monitoring</li> </ul> </li> <li>● Leverage third party security controls: <ul style="list-style-type: none"> <li>○ Trust mechanism in Hadoop</li> <li>○ Bull Eye algorithm in HFDS</li> <li>○ Name node approach in HDFS</li> <li>○ Security based on data structures</li> </ul> </li> </ul>

Domain	Relevant COBIT process	Detail processes' requirements	Risk(s) identified	Impact of the risk	Control(s) to mitigate the risk(s)
Deliver, Service and Support	DSS05		<ul style="list-style-type: none"> <li>• Re-identification of individuals.</li> <li>• Using, processing or disclosing personal data without consent.</li> <li>• Using data for a secondary purpose without obtaining consent.</li> </ul> Violation of individual participation rights.	High	<ul style="list-style-type: none"> <li>• Implement security controls:               <ul style="list-style-type: none"> <li>○ Authentication</li> <li>○ Encryption</li> <li>○ Anti-malware</li> <li>○ Access control</li> <li>○ Key establishment scheme</li> <li>○ Secure group key transfer</li> <li>○ Secure group data sharing</li> <li>○ Secure communication channel</li> <li>○ Self-assuring system</li> </ul> </li> <li>• Employ the following data management and data monitoring controls to ensure secure network / communication:               <ul style="list-style-type: none"> <li>○ Privileged access management</li> <li>○ Intrusion detection architecture</li> <li>○ Data Encryption Security Server</li> </ul> </li> </ul>
	DDS06	Manage business process controls with regards to cognitive computing information (including roles, responsibilities, data security and integrity).	<ul style="list-style-type: none"> <li>• Refer APO01, APO11 and DSS05</li> </ul>	High	Refer APO01, APO11 and DSS05
Monitor, Evaluate and Assess	MEA01	Monitor and evaluate cognitive computing performance.	<ul style="list-style-type: none"> <li>• Refer APO09, APO10</li> </ul>	High	Refer APO09, APO10
	MEA02	Monitor and evaluate the internal control system.	<ul style="list-style-type: none"> <li>• Refer EDM01</li> </ul>	High	<ul style="list-style-type: none"> <li>• Refer EDM01</li> </ul>
	MEA03	Monitor and evaluate compliance with IT laws and regulations.	<ul style="list-style-type: none"> <li>• Refer EDM03 and DSS05</li> </ul>	High	<ul style="list-style-type: none"> <li>• Refer EDM03 and DSS05</li> </ul>