Mitigating Ethical Risks of Using Artificial Intelligence in the Insurance Industry

Student Name: Mark Mkorongo

Student number: 2702380

Supervisor Name: Professor Patrick Ndayizigamiye

A research project submitted to the Faculty of Commerce, Law and Management, University of the Witwatersrand, in partial fulfilment of the requirements for the degree of Master of Management in the field of Digital Business

Johannesburg, 2025

ABSTRACT

The insurance industry, like other financial services sectors, has been disrupted by artificial intelligence (AI), transforming decision-making processes such as underwriting, pricing, and claim management. While AI can improve productivity, accuracy, and efficiency, it can also raise ethical concerns such as bias, discrimination, data quality, and data privacy concerns. These ethical risks can potentially have a negative impact on fairness, transparency, and trust in AI-driven decision-making.

This study explored the ethical implications of AI-driven decision-making in the insurance industry through the ethical lens of utilitarianism (consequentialist) and deontology (rule-based).

A qualitative research approach was employed within an interpretivist paradigm, with data collected through semi-structured interviews. A purposive and snowball sampling strategy was used to identify professionals within the insurance industry with knowledge and/or experience in AI. Thematic analysis, guided by Braun and Clarke's six-phase framework, was conducted to identify recurring patterns and insights within the data.

The study's findings highlight the ethical risks associated with the integration of AI in insurance processes. Current strategies were found to be somewhat effective however there remains a need for continuous improvement and adaptability to advances in AI. It also emerged that balancing innovation with ethical considerations was important to ensure that AI-driven insurance systems achieve optimal efficiency while maintaining ethical integrity and consumer trust.

A conceptual framework that incorporates utilitarian and deontological ethical principles, was therefore proposed as a guide to balancing accuracy, efficiency, continuous improvement and maximising overall good while ensuring adherence to moral rules, transparency and explainability, responsibility and accountability. The proposed conceptual framework serves as a guide for insurance industry stakeholders to infuse both utilitarian and deontological ethical principles into AI-driven decision-making.

KEYWORDS

AI, Artificial intelligence, Decision-making, Deontology, Ethics, Financial services, Insurance, Insurtech, Utilitarianism

DECLARATION

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

Name: Mark Mkorongo Signature: *M. Mkorongo*

Signed at **Kyalami**

On the **28th** day of **February 2025**

DEDICATION

To my family, my son, Walinase, my wife, Chido, my mum and dad, and my siblings, Phillip, Rachel, and Sarah, you are my source of strength and motivation. Your unwavering love and support inspire me to pursue my goals with determination and perseverance.

To Walinase, my greatest inspiration, your curiosity, joy, and boundless energy remind me every day why perseverance matters. This is for you, so that you may always chase your dreams fearlessly.

Family, this is for all of you!

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to my supervisor, Professor Patrick Ndayizigamiye, for his unwavering support, patience, and encouragement. His guidance and persistence pushed me forward even when I struggled to keep going. I truly would not have been able to persevere without him.

To my wife, Chido, thank you for holding everything together at home and giving me the time and space to complete this research. Your love and support made this journey possible. And to Walinase, my inspiration, thank you for simply being you, the coolest kid on this earth.

I am also immensely grateful to the participants, many of whom were strangers before I approached them, yet they willingly gave me their time without expecting anything in return. Their valuable insights were the cornerstone of this research, and this study would not have been possible without them.

And lastly, but most importantly, I give thanks to the Almighty for granting me the strength, perseverance, and dedication to complete not only this research report but my entire master's qualification.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ACRONYMS AND ABBREVIATIONS

ADM	Algorithm decision making
AI	Artificial intelligence
AMAs	Artificial moral agents
ARB	Algorithm review boards
AWS	Amazon Web Services
EIOPA	European Insurance and Occupational Pensions Authority
EU AI Act	European Union Artificial Intelligence Act
Fair-ML	Fairness-aware machine learning
FAT	Fairness accountability and transparency
FIDO	Framework for inhibiting data overcollection
Fintech	Financial services and technology
GDPR	General Data Protection Regulation
HITL	Human-in-the-loop
Insurtech	Insurance and technology
JSE	Johannesburg Stock Exchange
LLM	Large language model
ML	Machine learning
NLP	Natural language processing
POPIA	Protection of Personal Information Act
PPML	Privacy-preserving machine learning
RAI	Responsible AI
RPA	Robotic process automation
UBI	Usage-based insurance
XAI	Explainable artificial intelligence

CHAPTER 1. INTRODUCTION

1.1 Introduction

This chapter serves as the foundation of the study, outlining its purpose, background, and rationale. The research problem and the guiding research questions are clearly defined. The role of this chapter is to provide clarity and direction for the study, setting the stage for a more in-depth exploration in the subsequent chapters.

1.2 Statement of purpose

This qualitative study explored the ethical implications of artificial intelligence (AI)-driven decision-making in the insurance industry. The study's findings are discussed through the lens of the conceptual framework that guides this study, which encompasses the utilitarianism and deontological ethics constructs.

1.3 Background

The financial services industry provides financial services to individuals and organisations and is made up of banks, investment houses, lenders and insurance companies, to name a few (Perez, 2023). Fintech, which is a blend of financial services and technology, has disrupted the financial services industry by introducing innovative solutions driven by technology (Gomber et al., 2018). Insurtech refers to the fintech companies that utilise technology to innovate within the insurance industry (Catlin et al., 2017). AI is one of the many technologies being utilised by financial services companies that can assist in automating decision-making processes using machine learning (ML) algorithms and enhance the efficiency, consistency, and accuracy of decision-making (Guan et al., 2022). Other potential benefits of leveraging AI include increased productivity, lower costs due to automation, and improved customer experience (Balasubramanian et al., 2021). This study focused on the insurance segment of the financial services industry. Furthermore, it focused on the Insurtech segment of the insurance industry, which leverages technological advancements, specifically AI, to enhance and optimise insurance-related processes (Hargrave, 2024).

Although AI has the potential to deliver considerable advantages, it may also present numerous challenges. According to Guan et al. (2022), AI can struggle with complex decision-making scenarios due to tacit knowledge including customs, emotions and beliefs which are difficult to fully digitise and structure. The study by Guan et al. (2022) further identifies several ethical concerns with AI. These include algorithmic discrimination, data quality and integrity, privacy violations, ethical governance and regulation, as well as a lack of transparency in the AI decision-making process.

1.4 Research problem

AI technologies are being deployed in the insurance industry for various applications such as behavioural insurance, parametric products, novel pricing and risk assessment algorithms, e-services, and claims management (Holland et al., 2021). These applications can potentially lead to more efficient and accurate decision-making processes in the industry. According to Balasubramanian et al. (2021), AI is significantly transforming the following key processes within the insurance industry:

Underwriting: Advanced data analytics powered by AI enhance the precision of risk assessments by integrating large volumes of structured and unstructured data from diverse sources (Balasubramanian et al., 2021).

Claims processing: AI-driven automation accelerates claims adjudication, mitigates fraud, and improves customer experience by ensuring prompt and accurate claim settlements (Balasubramanian et al., 2021).

Pricing: ML models utilise historical and real-time data to refine pricing strategies, resulting in competitive and equitable premiums (Balasubramanian et al., 2021).

However, the use of AI in decision-making in the insurance industry can have negative ethical implications for insurance customers. According to Guan et al. (2022), the main sources of these ethical risks are as follows:

1.4.1 Technological factors

1.4.1.1 Technological uncertainty

AI systems inherently possess a degree of unpredictability, particularly in complex and dynamic environments. This uncertainty can lead to unforeseen consequences and a loss of control over AI actions, potentially resulting in harmful decisions (Guan et al., 2022).

1.4.1.2 Misuse and abuse of technology:

There exists a significant risk of AI technologies being misused or abused, either intentionally or unintentionally. Such misuse can lead to ethical dilemmas, including the deployment of AI for malicious purposes or beyond its intended scope, with potentially detrimental outcomes (Guan et al., 2022).

1.4.1.3 Algorithmic and program design issues:

Flaws in the design of algorithms and programs underpinning AI decision-making systems can embed biases and errors, leading to unfair or discriminatory results. This includes biases present in the data used for training AI models and the opacity of algorithmic processes (Guan et al., 2022).

1.4.1.4 Algorithmic bias

AI decision-making processes are susceptible to biases inherent in their algorithms. These biases may result from flawed design, inadequate testing, or intentional manipulation, leading to decisions that are unfair or discriminatory (Guan et al., 2022).

1.4.1.5 Data quality and integrity

The quality and integrity of data used in AI systems are paramount. Poor data quality, encompassing incomplete, biased, or incorrect data, poses significant ethical risks, as decisions based on such data can be inaccurate or unjust, adversely affecting individuals and groups (Guan et al., 2022).

1.4.1.6 Lack of transparency and accountability

The opacity of AI decision-making processes, often referred to as the “black box problem”, complicates the understanding of how decisions are made (Guan et al., 2022). This lack of transparency can hinder accountability, making it challenging to ensure that AI systems operate ethically and justly.

1.4.2 Human factors

1.4.2.1 Limited human rationality:

The limitations of human rationality can introduce biases and errors into AI systems, as these systems are both designed and managed by humans (Guan et al., 2022). This includes biases in data selection and problem framing, which can adversely affect AI decision-making processes.

1.4.2.2 Loss of human control

There is a concern that AI decision-making may surpass human control, leading to situations where AI systems make critical decisions without adequate human oversight (Guan et al., 2022). This lack of control can result in unpredictable and potentially harmful outcomes.

1.4.2.3 Impact on human dignity and autonomy

AI decision-making might undermine human dignity and autonomy by making decisions that significantly affect people’s lives without their input or understanding (Guan et al., 2022). This concern is particularly pertinent in sensitive contexts such as healthcare, legal judgements, and employment.

1.4.2.4 Incomplete and inadequate data:

The data utilised for AI decision-making could be incomplete, biased, or not representative of the entire population (Guan et al., 2022). Such deficiencies can result in decisions that are not only inaccurate but also unethical, as they may perpetuate existing biases or create new forms of discrimination.

1.4.2.5 Privacy violations

AI systems often necessitate vast amounts of data, raising concerns about individual privacy. The misuse of personal data, such as through unauthorised access or exploitation, represents a major ethical issue, potentially leading to privacy infringements (Guan et al., 2022).

1.4.2.6 Ethical governance and regulation

The absence of comprehensive ethical governance frameworks and regulations for AI technologies poses a risk. Ensuring that AI systems are developed and deployed in an ethically responsible manner is crucial to mitigating these risks. These ethical concerns highlight the necessity for ethical guidelines and governance frameworks to effectively address the risks associated with AI decision-making, ensuring that these systems are employed in a manner that promotes fairness, accountability, and respect for human rights (Guan et al., 2022).

1.4.2.7 Social and economic inequities

AI has the potential to exacerbate social and economic inequities by perpetuating existing biases and creating new forms of discrimination. This can lead to increased inequality and social injustice, negatively impacting marginalised communities.

1.4.2.8 Complex interactions among technology, humans, and society:

Ethical risks in AI decision-making emerge from the intricate interactions between technological systems, human users, and societal norms (Guan et al., 2022). These interactions can lead to unintended consequences, such as the erosion of social justice or the infringement of individual privacy.

Guan et al. (2022) study suggests that these risks can lead to social risks such as discrimination, inequality, erosion of privacy, loss of autonomy and human dignity, and a lack of trust and acceptance of AI systems. Therefore, this study aimed to gain insights into the ethical risks posed by AI and determine how they can be mitigated.

1.4.3 Problem statement

This study aimed to identify the ethical risks associated with the use of AI in the insurance industry and to propose strategies for mitigating these ethical concerns. Additionally, the study suggests strategies that need to be put in place to minimise the ethical risks that can emanate from the implementation of AI in the insurance industry. Therefore, this study primarily benefits insurance companies, regulators, policymakers, and insurance customers. By understanding and addressing the ethical ramifications, insurance companies can ensure their AI systems operate with fairness, transparency, and regulatory compliance, thereby maintaining customer trust and safeguarding their reputation. Simultaneously, the insights derived from the study are instrumental for regulators and policymakers in formulating guidelines and regulations that govern the ethical deployment of AI, ensuring consumer protection and promoting a fair market environment. Ultimately, consumers benefit from more equitable and transparent AI-driven decision-making processes, which reduce occurrences of discrimination and bias, ensuring fair treatment in various insurance procedures, including claims management, underwriting, and premium pricing.

To guide the reader's journey, this study's conceptual framework which is developed from utilitarian and deontological ethical principles will be introduced in Chapter 2 and revisited throughout the analysis to demonstrate how each ethical lens highlights specific AI-risk and mitigation themes in insurance.

1.5 Research questions

The study's research questions were:

a) What ethical risks arise from using AI in decision-making within the insurance industry?
b) How are these ethical risks currently addressed in the insurance industry?
c) How effective are the current strategies used to address ethical risks in the insurance industry?
d) What strategies can be implemented to ensure AI technologies are used ethically in the insurance industry?

1.6 Rationale

The deployment of AI technologies in the insurance industry is becoming increasingly prevalent, enhancing a range of processes including risk assessment, pricing, and claims management (Balasubramanian et al., 2021). As the integration of AI continues to expand, it is imperative to thoroughly understand its impact to ensure that it is leveraged responsibly. While AI promises significant improvements in efficiency and accuracy, it also raises profound ethical concerns. Issues such as algorithmic bias, discrimination, lack of transparency, and accountability must be critically examined to prevent potential harm to consumers and to uphold the integrity of the insurance industry. Addressing these concerns is essential for maintaining public trust (Guan et al., 2022).

1.7 Contribution to academic and practical knowledge

This study makes three key contributions:

- Theoretical contribution, as it extends current AI-ethics literature by integrating utilitarian and deontological perspectives into a unified ethical-risk framework tailored to the insurance context.
- Managerial contribution as it offers insurance executives a practical roadmap for embedding ethical-by-design principles into AI-driven decision processes, including concrete checks and balances at data-collection, model-development and deployment stages.
- Policy contribution as it informs South African and broader regulatory debates by outlining targeted guidelines that policy-makers can adopt to ensure fair, transparent and accountable AI use in insurance.

1.8 Delimitations

This study focused on the ethical implications of AI-driven decision-making within the insurance industry. The ethical implications of AI-driven decision-making outside of the insurance industry were thus not within the scope of the study.

1.9 Definition of terms

- **AI-driven decision-making**: refers to the process whereby AI systems analyse data, identify patterns, and provide recommendations or make decisions autonomously or with minimal human intervention (Colson, 2019). This approach leverages various AI technologies, such as ML, Natural Language Processing (NLP), and neural networks, to enhance the decision-making process.
- **Insurtech:** Insurtech is defined as the application of technology advancements to enhance efficiency and reduce costs within the insurance industry (Hargrave, 2024). This concept is derived from the combination of the terms “insurance” and “technology”, similar to fintech.
- **AI algorithms:** AI algorithms allow machines to analyse data and make decisions based on instructions (Baluja, 2024). These algorithms fall under the broader category of ML, which enables computers to learn and function autonomously.
- **Ethical implications:** Ethical implications refer to the consequences and considerations that arise from the development, deployment, and use of technology, particularly digital technologies such as AI (Martin et al., 2019).
- **Usage-based insurance (UBI):** UBI is an auto insurance model where premiums are determined based on individual driving behaviour (Ben-Shahar, 2023).
- **Responsible AI (RAI):** RAI, also known as ethical AI, is the practice of using AI with good intentions to empower those who implement it whilst positively impacting the recipients such as customers (Eitel-Porter, 2021).
- **Explainable AI (XAI):** XAI refers to AI systems designed to provide clear and understandable explanations of their decision-making processes to stakeholders (Owens et al., 2022).
- **Fairness-aware ML (fair-ML):** Fair-ML is characterised by the attempt to design and implement ML algorithms that account for fairness considerations (Selbst et al., 2019).

- **Black box:** Many modern algorithms, especially those based on deep learning, are often black boxes, meaning their decision-making processes are not transparent or easily explainable (Dexe et al., 2021).

1.10 Assumptions

This study was conducted through the lens of a combined conceptual framework that leverages both the utilitarian and deontological ethical theories. The combination of these frameworks helped to understand the ethical implications of AI in the insurance industry. This included examining how AI can maximise benefits for most people (utilitarianism), such as clients, insurance companies, and regulators, while also considering the morality of using AI in decision-making in the insurance sector (deontology). The study's guiding conceptual framework is discussed in section 2.6.2.

1.11 Chapter outline

The structure of the research report is as follows:

Chapter 1: Introduction

Chapter 1 set the stage for the research by outlining the background and rationale of the study. It defined the research problem and presented the research questions that guided the study.

Chapter 2: Literature review and theoretical framework

Chapter 2 provides a comprehensive literature review applicable to the study. The chapter presents the literature based on the four research questions that focus on the ethical risks of AI within the insurance industry, current strategies employed to mitigate the ethical risks, the effectiveness of the current strategies, and possible future strategies to mitigate ethical risks. Lastly, the chapter introduces the conceptual framework that integrates both utilitarian and deontological perspectives.

Chapter 3: Research methodology

Chapter 3 explores the methodology employed in the study. It begins by outlining the research paradigm, followed by the research approach that guided the investigation. The chapter then examines the research design, clarifying the rationale behind it. This is followed by a discussion of the sampling techniques, along with the methods used for data collection and analysis. The chapter offers a detailed account of the data collection process and the subsequent procedures for data analysis. Lastly, the chapter highlights the ethical considerations that were central throughout the research journey.

Chapter 4: Presentation of findings

Chapter 4 presents the findings from the thematic analysis of expert interviews. The chapter begins by detailing the demographics of the participants and providing context for the findings. It then systematically presents the findings, providing interpretations in relation to the posed research questions.

Chapter 5: Discussion of findings

Chapter 5 analyses the findings against the conceptual framework and the literature review. Using the themes from the findings and insights from the literature review, along with the initial conceptual framework, this chapter constructs a revised conceptual framework. The framework maps the themes to the tenets of the ethical theories of utilitarianism and deontology to ensure ethical outcomes in the use of AI within the insurance industry.

Chapter 6: Conclusions and recommendations

Chapter 6 concludes the study by offering a summary of the study. It emphasises the study's contributions and acknowledges the limitations of the study. The chapter details the theoretical contributions, along with the broader implications and practical applications arising from the research. It also presents recommendations and identifies potential directions for future research.

1.12 Conclusion

This chapter established the research foundation by outlining the study's purpose, background, and rationale. The chapter also presented the research problem and subsequent research questions that guide the study. Ultimately, the introduction sets the stage for a structured investigation, ensuring a thorough exploration in the following chapters.

Chapter 2 follows and comprises the literature review and theoretical framework.

CHAPTER 2. LITERATURE REVIEW AND THEORETICAL FRAMEWORK

2.1 Introduction

This chapter provides a comprehensive review of the literature on the ethical implications of AI in the insurance industry. The review is based on the four research questions underpinning the study. The chapter also provides the theoretical foundation of the study, integrating utilitarian and deontological ethics to ensure a structured approach to assessing AI ethics. Finally, the chapter presents a conceptual framework that combines utilitarian and deontological viewpoints.

AI is transforming insurance by automating claims processing, enhancing risk assessment, and personalising pricing through real-time data analysis from connected devices (Balasubramanian et al., 2021). This shift enables more efficient operations, informed customer decision-making, and advanced digital capabilities, fundamentally reshaping the industry's landscape.

ML typically enables systems to autonomously learn and improve through experience without requiring explicit programming. It is widely recognised as one of the most prominent and advanced technologies within the Fourth Industrial Revolution (4IR or Industry 4.0) (Sarker, 2021). One of the main benefits of ML is its learning and intelligent decision-making capabilities. The data-driven predictive analytics employed by ML can explore data from past events to predict the outcome of future events (Sarker, 2021).

Neural networks and decision trees are AI techniques that can potentially be utilised for predictive underwriting in the insurance industry. Neural networks consist of interconnected nodes organised into layers, which can be trained on data to learn complex patterns and relationships. Once trained, these networks can analyse large datasets and generate predictions, such as forecasting future outcomes to support underwriting decisions (Zanke & Sontakke, 2021).

Supervised learning algorithms, such as decision trees, are used to segment data into subgroups based on defined decision-making rules. In a decision tree, each internal node represents a decision or test on a specific feature, while each leaf node represents a predicted outcome or class label. Due to the need for explainability and versatility in handling categorical and numerical data, decision trees are suitable for risk management and segmentation (Zanke & Sontakke, 2021).

AI techniques can also help improve pricing efficiency by using predictive capabilities to determine potential losses and set competitive premiums for each policy (Zanke & Sontakke, 2021). This involves analysing data such as loss history, market conditions, and competitor pricing.

Additionally, claims management for insured assets such as motor vehicles can be automated using a combination of NLP, image analysis, and pattern anomaly detection (Eling et al., 2022). These technologies can streamline the claims management workflows, including key processes such as damage assessment using ML models trained in image recognition to assess the damage to motor vehicles based on photos submitted by the client.

The potential for AI to make decisions that impact individual lives raises significant ethical questions. AI technologies can lead to ethical issues such as algorithm bias towards clients, privacy concerns, discrimination, lack of transparency, and accountability (Dhirani et al., 2023).

Ethical risks within the African context may differ from Western societies. Hassan (2023) argues that the literature around AI ethics is biased towards the Euro-American perspectives, thus the implications from an African perspective are understudied.

AI systems rely heavily on the variety, quality, and diversity of data for the models to learn, make decisions and provide solutions. However, due to the minimal nature or lack of African-specific data that accurately represents African social, economic, and environmental contexts, AI systems risk being ineffective, biased, or even harmful when applied in the African context (Hassan, 2023).

Dancy and Saucier (2022) argue that AI bias, such as “antiblackness” or “structural racism”, is not a surface issue but is deeply entrenched in the structures and concepts that influence the design, development and deployment of AI systems. Antiblack outcomes are a symptom of these structural forms. As noted by Hassan (2023), AI systems can be influenced by historical and ongoing bias, as these systems rely on large datasets that may reflect the racial inequalities of the societies from which they are drawn.

The development of technology such as AI decision-making systems can be influenced by coloniality, which is rooted in systems, institutions and values from the past that remain unresolved in the present (Mohamed et al., 2020). These AI decision-making systems can, therefore, reinforce historical inequalities that benefit developed nations and large corporations, while, in turn, marginalising underrepresented groups. This phenomenon is referred to as algorithmic coloniality.

2.2 Ethical risks of using AI in decision-making within the insurance industry

The first research question focuses on the potential ethical risks that AI decision-making can have on the insurance industry. The following are some of the risks that were identified in the literature:

2.2.1 Algorithm bias

According to Fabris et al. (2021), insurance quotes from various popular comparison websites could be generated by biased algorithms. Their study, which was conducted on the pricing algorithms that Italian car insurance companies employ, shows that both gender and birthplace directly affected the quoted insurance premiums. While there is no clear conclusion that one gender is favoured over the other, birthplace shows systematic discrimination. Foreign-born drivers and natives of Naples (Italy) are charged higher premiums than those born in Milan, even when other factors are constant. In addition, riskier drivers, on average, received fewer quotes when querying comparison websites.

These findings point to significant and potentially unlawful biases in the way car insurance premiums are quoted based on gender and birthplace, with both the comparison websites and individual insurers contributing to these disparities. This bias can be attributed to the following factors:

- The use of protected attributes such as gender and birthplace either explicitly or through correlation.
- Training data with historical bias that is now been perpetuated by the algorithm.
- Decisions made during the design and implementation of the AI systems.
- Lack of sufficient regulatory oversight to ensure fair algorithm practices.

Although this bias pertains to the Italian market, it highlights the potential for discrimination by the algorithms within the insurance market context.

2.2.2 Data quality and integrity

Breidbach (2024) presents a comprehensive analysis of the challenges and ethical considerations associated with algorithmic decision-making across various industries and sectors. This qualitative study argues that the quality of data used by algorithms can be poor, which can lead to inaccurate or biased decision-making. The integration of diverse data formats from multiple sources can compromise data accuracy, timeliness, and comprehensiveness. The algorithms that are developed can be limited in their decision-making abilities. For example, fitness trackers, which sometimes produce inaccurate data about users' physical activities, highlight the challenges in ensuring high-quality data inputs for algorithms. Breidbach (2024) further suggests that unsupervised learning algorithms often identify correlations in data and make decisions based on these probabilities. However, correlation does not imply causation. Therefore, decisions based solely on correlations can be flawed because they might not reflect actual causal relationships.

Additionally, Cai and Zhu (2015) highlight the significant challenges posed by the volume, velocity, and quality assessment of data in the big data era. The exponential growth in data volume, ranging from terabytes to zettabytes, creates substantial difficulties in ensuring data accuracy and integrity due to the complexity and resource demands of validation, cleaning, and integration processes. Additionally, the rapid generation and flow of data require real-time processing to maintain relevance, yet existing systems often struggle to handle the velocity, leading to potential delays and errors. These challenges are further compounded during quality assessment, as the diverse sources and inconsistent formats of data increase the difficulty of extracting high-quality information. The authors propose a dynamic data quality assessment framework to address these issues, emphasising tailored solutions to uphold the timeliness, reliability, and accuracy of big data.

2.2.3 Privacy concerns

Liu et al. (2020) conducted a comprehensive review of various scenarios related to privacy and ML. The findings revealed that the sensitive nature of the data that can be utilised by ML algorithms to make decisions raises concerns regarding possible data breaches (Liu et al., 2020). In addition, there is the potential for reidentification, in which individuals can be identified by linking various anonymised datasets. Liu et al. (2020) offer a distinct perspective on privacy, differing from other authors who primarily concentrate on the misuse of data by companies rather than external attacks on client information.

Ben-Shahar (2023) conducted a study on UBI in the United States with an emphasis on how the benefits, such as improved driving behaviour, cost savings for policyholders, and enhanced risk assessment, are being limited by privacy protection. The author argues that enrolling in UBI is a voluntary decision that the client makes and can enhance the overall safety and freedom of driving; therefore, the choice to join, despite the collection of personal data and surveillance devices infiltrating the client's personal space, suggests a willingness to trade certain freedoms, in this case, privacy, for perceived benefits.

Privacy issues thus highlight the complex balance between leveraging AI for improved insurance processes and protecting individual privacy rights, to which insurance companies need to adhere.

Curzon et al. (2021) identify several privacy concerns related to AI, including extensive data collection and unauthorised secondary use, which can lead to privacy violations and misuse of personal information. The risks of data misrepresentation and poor decision-making arise from incorrect or narrow data representations. Data aggregation poses additional risks by potentially exposing sensitive information through the combination of disparate datasets. Ambiguities in the use of public information, particularly in social media and facial recognition, further complicate privacy issues. Moreover, ML systems introduce new privacy risks, necessitating continuous reassessment and rigorous privacy-preserving mechanisms to safeguard against vulnerabilities.

2.2.4 Transparency and accountability

Aysolmaz et al. (2023) investigated the impact of transparency on perceived fairness, accountability, and privacy and how these factors influence trust, usefulness, and the intention to adopt algorithmic decision-making (ADM) systems. The study found that transparency is essential for customers to evaluate the fairness, accountability, and privacy of ADM systems. Customers' perceptions of these elements influence how useful they find ADM systems in enhancing customer-facing processes, which in turn affects their willingness to adopt such systems. The study found that transparency is a prerequisite for people to make judgements about ADM systems when it comes to fairness, accountability, and privacy (Aysolmaz et al., 2023). How the individual views these elements can impact how they perceive usefulness in improving customer-facing processes, thereby facilitating the adoption of ADM systems.

Akinrinola et al. (2024) on the other hand, conducted a qualitative study that examined ethical dilemmas, real-world examples, and case studies to propose strategies that could be used to ensure transparency, fairness, and accountability in AI development of algorithms and models. According to Akinrinola et al. (2024), the lack of transparency regarding how ADM systems make decisions, which is commonly referred to in various literature as the black box problem, exacerbates concerns around trust and accountability in AI. AI models function as intricate, ambiguous systems, making it difficult for stakeholders and users to comprehend the decision-making process.

2.2.5 Ethical governance and regulation

De Almeida et al. (2021) conducted a systematic literature review of AI regulations. The study describes the current global regulatory framework on AI as fragmented and reactive, lagging behind the rapid advancements in AI. The challenge is managing these AI technologies to adhere to ethical standards such as transparency, fairness, respect for privacy, and accountability. The absence of appropriate regulations in some countries and regions constitutes a risk, due to the lack of a regulatory body that can ensure that AI advancements align with ethical guidelines.

2.2.6 Overreliance on AI and hallucination

According to Roychowdhury et al. (2023), “hallucination” refers to the phenomenon in which large language models (LLMs) produce responses that contain false or misleading information presented as factual. These inaccuracies often stem from inherent biases and gaps within the training data, which can skew the model’s outputs. In the context of insurance, such hallucinations can have significant repercussions, leading to clients receiving incorrect information about their policies, premiums, or claims processes. This not only risks eroding the trust that clients place in their insurers but can also inflict lasting reputational damage on the company as misinformation spreads. The potential consequences highlight the critical need for vigilance and accuracy in the deployment of AI technologies within client-facing services.

According to Buçinca et al. (2021), the idea of merging human capabilities with AI to form what are known as sociotechnical systems is grounded in the belief that this integration can produce results that surpass the individual performance of either humans or AI alone. However, a significant concern arises when users develop an overreliance on AI systems. When individuals place complete trust in AI recommendations, accepting them without question, even in cases where those suggestions may be flawed or inaccurate, can lead to poor decision-making. In such scenarios, choices made with the assistance of AI may ultimately turn out to be less effective than those that could have been made independently by users, free from the influence of potentially misguided AI guidance.

2.2.7 Proposition 1

Based on the literature review on the ethical risks that arise from using AI in decision-making within the insurance industry, it can be proposed that ethical risks such as privacy violations, bias, and lack of transparency can arise from the use of AI-driven decision-making within the industry.

2.3 Strategies currently used to address ethical risks of AI used in the insurance industry

The second research question focuses on how ethical risks are currently being addressed within the insurance industry.

2.3.1 Key principles of responsible AI (RAI)

Eitel-Porter (2021) discusses the general principles and frameworks for implementing AI across various organisations. The study suggests that many organisations have embraced a principle-based approach. This approach involves implementing and applying fundamental ethical principles to steer the growth, implementation, and management of AI technologies. The aim is to minimise the impact of ethical risks by integrating five crucial pillars: fairness, accountability, transparency, explainability, and privacy. However, principles alone are not sufficient to ensure RAI, they need to be put into practice.

2.3.2 Regulatory compliance

In the context of Europe, the General Data Protection Regulation (GDPR) provides a framework for the privacy and protection of personal data. It ensures that the processing of personal data is lawful, fair, and transparent and that the rights of the data subject are protected (Van Den Boom, 2021). These aspects are relevant to addressing ethical risks in the insurance industry and any industry incorporating AI decision-making within its processes. The GDPR established the governance of “algorithmic accountability” and insurance customers’ “right to explanation” of the decision-making algorithms, which encourages XAI (Owens et al., 2022). Therefore, it is evident that privacy regulations such as the GDPR in Europe can provide guidelines to mitigate ethical risks of AI decision-making in the insurance industry, as key aspects include ensuring fairness, transparency, and accountability of the ADM systems.

In South Africa, the Protection of Personal Information Act (POPIA) serves as the cornerstone of the nation’s data protection framework, meticulously overseeing the processing of personal information belonging to individuals, referred to as data subjects (Sharma & Sharma, 2024). This pivotal legislation preserves principles of fairness, accountability, and transparency, drawing parallels to the GDPR implemented in the European Union. While the POPIA provides a regulatory structure aimed at safeguarding personal information, it only indirectly addresses some privacy concerns associated with AI. As of now, South Africa lacks dedicated legislation specifically aimed at mitigating the potential privacy risks posed by AI technologies, leaving a critical gap in the regulatory landscape (Sharma & Sharma, 2024).

Additionally, in the European landscape, the GDPR serves as a comprehensive framework dedicated to ensuring the privacy and safeguarding of personal data. This regulation establishes essential principles that govern the lawful, fair, and transparent processing of personal information. It empowers individuals by protecting their rights regarding their data, ensuring they can access, rectify, and request the deletion of their personal information. By enforcing strict guidelines on data handling, the GDPR aims to foster trust and accountability among organisations and uphold the fundamental rights of data subjects across Europe (Van Den Boom, 2021).

In addition to data protection frameworks like the GDPR and POPIA, the European Union Artificial Intelligence Act (EU AI Act) introduces targeted regulatory obligations for insurance applications, particularly those deemed “high-risk” such as life and health underwriting (Mahajan et al., 2025). According to Mahajan et al. (2025), the Act mandates rigorous fairness testing, technical documentation and post-deployment monitoring for AI systems used in pricing and underwriting, and imposes substantial penalties for non-compliance.

While these measures aim to reduce algorithmic bias and enhance transparency, Mahajan et al. (2025) notes that such compliance requirements can introduce capital strain and distort premium structures, unless insurers carefully calibrate their models and governance frameworks. For insurers, this implies that aligning AI governance with GDPR and POPIA alone may be insufficient. Instead, it is advisable to adopt a compliance architecture that integrates EU AI Act-specific mechanisms such as bias audits, conformity assessments and documentation protocols alongside existing data protection practices, to effectively manage risk while maintaining operational viability.

2.3.3 Data integrity

Deekshith (2021) emphasises the vital role that high-quality data plays in the success of ML models. By ensuring the accuracy and reliability of this data, one can significantly reduce the risks related to biases, inaccuracies, and errors that may compromise the performance of these models. Deekshith (2021) also highlights the paramount importance of utilising diverse and representative datasets. Such diversity is essential not only for mitigating biases but also for promoting fairness in outcomes, ultimately leading to more equitable and effective solutions in ML applications.

Shanmugam et al. (2022) emphasise the importance of data minimisation in AI systems as a means to reduce risks such as privacy violations and data misuse. They propose FIDO (Framework for Inhibiting Data Overcollection), which uses performance-based criteria to identify the minimum amount of data necessary for achieving acceptable model performance, aligning with the GDPR's principle of processing only relevant and necessary data. Additionally, the authors stress the importance of safeguarding data by controlling its usage and preventing unauthorised reuse, particularly in contexts where data could be exploited for unintended applications. These insights reinforce the necessity of limiting data collection and ensuring its responsible use to enhance compliance with privacy regulations and mitigate ethical risks (Shanmugam et al., 2022).

2.3.4 Fairness-aware machine learning (fair-ML)

Selbst et al. (2019) explain that fair-ML focuses on embedding fairness into the ML algorithms. However, this could be an oversimplification of fair-ML, as fairness in ML is complex and requires more than just adjusting the inputs or tweaking the algorithms.

2.3.5 Transparency and explainability

The study by Dexe et al. (2021) focuses on how Swedish insurance companies address the dilemma of implementing transparency within the insurance industry, which lies in balancing the need for openness, consumer trust, and the risks of revealing sensitive information, overwhelming consumers, and compromising competitive advantage. Their findings reveal that a small number of companies strategically use transparency to gain customer trust and, therefore, a competitive advantage. This is in line with companies acknowledging the importance of transparency in aligning with ethical practices. Insurance companies that use transparency to their advantage gain trust by ensuring that consumers are well-informed, understand how their data is used, and feel confident that the company operates in an ethical and responsible manner.

Owens et al. (2022) underscore the critical role of explainability in AI systems within the insurance industry. Their findings highlight the importance of transparency and explainability in fostering trust and accountability, particularly in high-impact contexts such as insurance, where decisions significantly affect individuals. They emphasise the challenges posed by black-box models, noting that their opacity can undermine fairness and accountability, thereby hindering stakeholder confidence. Owens et al. (2022) identify the adoption of XAI techniques as essential for embedding trust, transparency, and moral values into AI systems, enabling stakeholders to understand and justify complex decision-making processes.

2.3.6 Governance committees and frameworks

Hadley et al. (2024) highlight the growing importance of governance committees, such as algorithm review boards (ARBs), as key mechanisms to manage risks and ensure ethical oversight in AI and data practices. These committees play a crucial role in reviewing AI systems from both technical and societal perspectives, addressing considerations such as ethics, accountability, and operational impacts. To be effective, Hadley et al. (2024) argue that governance committees must include a balanced mix of technical and non-technical expertise, ensuring that RAI practices are guided by diverse insights. This multidisciplinary composition enables these committees to holistically evaluate AI systems and contribute to the development of ethical, transparent, and accountable AI processes.

Similarly, Birkstedt et al. (2023) underscore the necessity of multi-stakeholder involvement in AI governance, asserting that effective governance structures must incorporate diverse expertise to ensure comprehensive oversight and accountability. Their study highlights that legal experts play a critical role in ensuring compliance with evolving regulatory frameworks, while risk managers assess potential ethical, operational, and reputational risks associated with AI deployment. Furthermore, business leaders contribute strategic oversight, aligning AI governance with broader organisational objectives, whereas IT professionals provide the technical expertise necessary to implement governance mechanisms effectively. The authors argue that without such interdisciplinary collaboration, AI governance frameworks risk being fragmented and ineffective, particularly in addressing complex ethical dilemmas and aligning AI initiatives with corporate social responsibility (Birkstedt et al., 2023). Consequently, the study advocates for a structured, multi-disciplinary governance model that integrates these diverse perspectives to enhance the transparency, accountability, and ethical integrity of AI systems.

Complementing the role of governance committees, Ferrell et al. (2024) emphasise the need for AI ethical frameworks that are thoughtfully aligned with cultural, industry-specific, organisational, and legal standards. These frameworks must be designed to address the unique challenges posed by AI technologies while incorporating diverse perspectives from various stakeholders. Ferrell et al. (2024) suggest that engaging a wide range of contributors during the development process enhances the comprehensiveness and applicability of these frameworks, ultimately strengthening their capacity to guide ethical AI practices effectively.

Fahmideh et al. (2021) underscore the importance of integrating ethical considerations into every phase of the AI lifecycle, rather than treating them as isolated or supplementary concerns. The authors argue that ethical principles should be embedded from the initial stages of data collection and model training through to deployment and ongoing monitoring. This perspective aligns with the broader notion that ethical frameworks should not only inform internal AI development processes but also extend to procurement decisions, ensuring that organisations uphold ethical standards even when sourcing AI technologies from external vendors. By adopting this comprehensive approach, businesses can foster greater accountability and mitigate the potential ethical risks associated with AI-driven decision-making.

By integrating governance structures such as ARBs with well-developed ethical frameworks, organisations can establish a robust foundation for RAI. These approaches work in tandem to ensure that ethical considerations are systematically embedded in both decision-making and operational processes, fostering trust and accountability in AI deployments.

2.3.7 Building ethical awareness and culture

Santoni de Sio (2024) highlights the critical importance of educating employees about the ethical implications of AI throughout its entire development lifecycle, from the initial design phase to full deployment. This education is essential to equipping employees with the knowledge needed to recognise and address potential ethical challenges that may arise at various stages of AI implementation. Furthermore, Santoni de Sio (2024) emphasises that embedding ethical, legal, and societal considerations into employee education is vital for mitigating significant risks, including bias, discrimination, and a lack of accountability in AI systems. By fostering a deeper understanding of these issues, organisations can promote ethical awareness and proactive risk management. This perspective aligns with the growing emphasis on enhancing AI literacy within organisations, ensuring that employees are well-equipped to use AI responsibly and ethically in their respective roles.

Figaredo and Stoyanovich (2023) substantiate this argument by emphasising the critical role of AI literacy in providing individuals with the cognitive competencies necessary to identify and address the ethical and technical risks inherent in algorithmic decision-making systems. The authors contend that insufficient AI literacy may hinder individuals, including professionals working in sectors such as the insurance industry, from fully comprehending and mitigating the ethical and privacy challenges associated with AI applications. These challenges include, but are not limited to, data breaches, algorithmic bias, and the potential misuse of sensitive personal information. Without a foundational understanding of AI and its implications, employees may inadvertently contribute to or fail to detect these risks, thereby exacerbating their impact on both businesses and consumers.

2.3.8 Balance innovation with ethics

Konda et al. (2024) contend that the adoption of AI technologies can deliver transformative benefits, including enhanced efficiency, improved accuracy, and enriched customer experiences. These advancements have the potential to drive significant innovation and create substantial value for organisations. However, Konda et al. (2024) also caution that the integration of AI systems is not without challenges, as it can give rise to complex ethical dilemmas. To address these challenges, they stress the importance of balancing technological progress with robust ethical considerations and adherence to regulatory frameworks. This balance is crucial for fostering RAI, ensuring that its deployment aligns with societal values, protects stakeholders, and mitigates potential risks associated with its adoption.

Baker and Rajab (2024) argue that the challenge of balancing competitiveness with ethical integrity reflects the tension businesses face in navigating market pressures while maintaining ethical standards in AI innovation. Organisations must continuously weigh the benefits of adopting cutting-edge AI technologies against the risks of compromising ethical principles to retain market share. Baker and Rajab (2024) reinforce this dilemma by emphasising that the pursuit of rapid innovation often leads companies to deploy AI systems without fully assessing their fairness, transparency, and long-term societal impact. Concurrently, regulatory evolution and global cooperation play a crucial role in mitigating these risks by fostering harmonised ethical standards and compliance frameworks. Baker and Rajab (2024) advocate for cross-border collaboration to ensure AI governance aligns with international best practices, preventing regulatory fragmentation and strengthening accountability. By integrating competitive imperatives with evolving ethical governance structures, organisations can pursue responsible innovation that balances market leadership with long-term societal trust and regulatory compliance.

2.3.9 AI model selection, training, and validation

Smith and Bochanski (2022) explored the trade-offs between tailored (in-house) AI solutions and vendor-provided models in mitigating ethical risks. Tailored solutions allow organisations greater control over data, functionality, and ethical compliance, making them well-suited for complex or sensitive applications. However, they require significant financial and technical resources, as well as continuous monitoring and maintenance. Conversely, vendor-provided AI models offer cost efficiency and faster deployment, enabling organisations to leverage external expertise. However, these models often lack transparency and flexibility, posing risks related to bias, regulatory non-compliance, and ethical misalignment. To mitigate these challenges, Smith and Bochanski (2022) advocate for rigorous vendor assessments, validation frameworks, and contractual agreements that define ethical usage and compliance measures. They emphasise a context-sensitive approach in determining whether to build or buy AI systems, ensuring that AI solutions, whether developed in-house or procured externally, support both innovation and ethical governance while aligning with organisational values and evolving regulatory standards.

2.3.10 Human-in-the-loop decision-making

Mosqueira-Rey et al. (2023) emphasise that ML models that operate without continuous human oversight are susceptible to becoming static, difficult to evaluate, and prone to performance degradation when exposed to evolving real-world scenarios. This finding reinforces the argument that AI should not be granted full autonomy in making final decisions, particularly in sensitive, client-facing processes, where human validation serves as a crucial safeguard. Human oversight ensures that AI systems remain adaptable to new contexts, maintain transparency in their decision-making processes, and uphold accuracy by mitigating potential biases or errors that may arise from shifts in data patterns or operational conditions.

Sele and Chugunova (2024), however, challenge the assumption that human involvement in AI decision-making inherently leads to better outcomes. Their study reveals that human interventions frequently reduce decision accuracy rather than enhancing it, as human reviewers often fail to detect or appropriately correct errors within AI-generated outputs. This finding contradicts the widely held belief that human oversight always improves decision quality, suggesting instead that reviewers may inadvertently introduce biases, overlook significant errors, or over-rely on algorithmic recommendations without critically assessing their validity. This raises concerns about the effectiveness of human-in-the-loop (HITL) systems as a universal safeguard, particularly in contexts where decision accuracy is paramount.

2.3.11 Proposition 2

There are current strategies that are used to address ethical risks in the insurance industry. These must be investigated to establish their effectiveness and ways to enhance them.

2.4 The effectiveness of current strategies used to address ethical risks in the insurance industry

The third research question focuses on how effective current strategies are at addressing ethical risks within the insurance industry.

2.4.1 Effectiveness of regulatory compliance

Stahl et al. (2022) provide a comprehensive analysis of how organisations are addressing ethical issues related to AI from the European perspective. Their findings reveal that compliance with regulations such as the GDPR in Europe ensures that insurance companies adhere to strict privacy and data protection standards. While regulatory compliance is a significant step towards addressing ethical risks, it is also often seen as a minimum requirement rather than a comprehensive solution. Regulations can sometimes lag behind technological advancements, leaving gaps in addressing new ethical challenges. Therefore, specific regulations are needed to manage ethical risks, ensure ethical standards, and maintain transparency and accountability in AI development and deployment (De Almeida et al., 2021).

2.4.2 Effectiveness and evaluation of current strategies

Stahl et al. (2022) argue that insurance companies can adopt ethical frameworks to govern the use and development of AI models. These frameworks often include principles such as fairness, accountability, transparency, and privacy. The effectiveness of these frameworks depends on how rigorously they are implemented and enforced. In many cases, there is a gap between the prescribed principles and their practical application.

Stahl et al. (2022) further argue that while regulatory frameworks such as the GDPR establish strict privacy and data protection standards in the insurance industry, they often serve as a baseline rather than a comprehensive ethical safeguard. Compliance with such regulations ensures adherence to fundamental legal requirements; however, the rapid pace of technological advancement frequently outstrips regulatory evolution, creating gaps in addressing emerging ethical challenges. As a result, organisations that rely solely on compliance as a measure of ethical responsibility may fail to anticipate and mitigate novel risks associated with AI-driven decision-making. This perspective underscores the limitations of static regulatory mechanisms, suggesting that organisations must proactively adapt their governance strategies beyond legal mandates to uphold ethical integrity in AI applications.

Thierer (2023) critiques the assumption that existing governance structures are sufficient for ensuring ethical AI deployment, arguing that a reactive approach based solely on compliance can lead to a false sense of security. While regulatory compliance plays a crucial role in risk mitigation, Thierer (2023) contends that rigid, one-size-fits-all regulatory frameworks may be ineffective in addressing the nuanced challenges posed by AI. Instead, Thierer (2023) advocates for flexible, decentralised governance models that emphasise voluntary compliance, iterative learning, and stakeholder collaboration. This approach challenges the traditional compliance-centric perspective, proposing that soft-law mechanisms, such as industry-led best practices, self-regulation, and adaptive oversight, offer a more dynamic and responsive means of managing AI ethics. By prioritising proactive governance over regulatory rigidity, Thierer (2023) highlights the need for AI governance frameworks that evolve alongside technological advancements, ensuring both ethical accountability and innovation.

Vazquez-Zapien et al. (2022) emphasise the necessity of rigorous validation protocols before deploying AI models, underscoring the importance of establishing high accuracy thresholds to ensure reliability. The authors highlight that accuracy, sensitivity, and specificity obtained through validation techniques serve as critical indicators of a model’s readiness for real-world application. This reinforces the perspective that AI systems must meet stringent performance benchmarks before being operationalised, ensuring their effectiveness and reducing the risk of inaccurate or biased decision-making. By advocating for comprehensive validation processes, the study underscores the need for robust evaluation frameworks that go beyond theoretical performance metrics to guarantee real-world applicability.

Similarly, Poenaru-Olaru et al. (2023) stress the importance of continuous monitoring and adaptation of AI models to maintain their long-term accuracy and relevance. The authors identify concept drift, where changes in data distribution over time lead to model degradation, as a significant challenge that must be proactively managed. Without regular assessment and adjustment, AI models risk becoming obsolete or unreliable, leading to erroneous predictions and ethical concerns. By advocating for adaptive strategies such as real-time drift detection and strategic model retraining, the study highlights the necessity of ongoing performance evaluation to ensure AI models remain both effective and fair in dynamic environments.

2.4.3 Ensuring algorithm fairness

Mehrabi et al. (2021) investigated biases in AI systems arising from both the data and the algorithms themselves when applied to various use cases and regions. They argue that tools and techniques that can detect and mitigate bias in datasets and models are being integrated into the AI development process. This includes pre-processing, in-processing, and post-processing methods to identify and correct potential biases. Although these tools and techniques could potentially reduce bias, they may be ineffective at detecting subtle or unidentified biases that are not easily recognised. Tampering with the model may also risk model overfitting, which can potentially identify false correlations as biases.

2.4.4 Proposition 3

The effectiveness of current strategies in addressing ethical risks in the insurance industry is limited, especially when implemented individually. Therefore, a multifaceted approach may be more effective in bridging the gaps in individual strategies.

2.5 Strategies that can be implemented to ensure AI technologies are used ethically in the insurance industry

The fourth research question focuses on what strategies can be implemented to ensure AI technologies are used ethically in the insurance industry.

2.5.1 Transparency and explainability

Mullins et al. (2021) discuss the development of AI ethical frameworks by the European Insurance and Occupational Pensions Authority (EIOPA), for the European insurance market. There are two types of transparency: model transparency and process transparency. Model transparency refers to the model's code and how input data and output data relate to decision-making, while process transparency involves making the steps, logic, and criteria used by AI systems to make decisions visible and understandable to stakeholders. Additionally, transparency within the insurance industry ensures that policyholders have visibility into how decisions about their policy were made whilst explainability ensures policyholders understand why those decisions were made. The challenge, however, is that AI and other digital technologies complicate the explainability due to the black-box nature of most AI systems. Therefore, to address transparency concerns, it is crucial to ensure both model and process transparency to build trust and ensure customers understand how AI makes decisions.

2.5.2 Fairness

Ntoutsi et al. (2020) provide a holistic view of AI bias within the European market, its challenges, solutions, and suggested research directions. They argued that bias can be mitigated at different stages of AI decision-making, namely, pre-processing, in-processing, and post-processing. Pre-processing mitigation requires modifying the training data before using it to train AI models. In-processing mitigation involves modifying the model to intentionally minimise bias during training. Post-processing mitigation adjusts the model's predictions after training to ensure fairness.

Akinrinola et al. (2024) argue that three key strategies can ensure accountability in AI development: regulatory measures, ethical AI governance, and HITL approaches. Existing regulations, such as the GDPR and the POPIA, are not tailored for AI and are evolving too slowly to keep pace with the rapid advancements in the field. Therefore, to address these limitations, it is imperative that regulatory frameworks be developed in collaboration with policymakers, technologists, and ethicists, providing clear guidelines for AI developers to follow in mitigating ethical risks. Ethical AI governance requires the creation of diverse boards that oversee ethics within AI development and their role is to enforce ethical guidelines to ensure the development of RAI. HITL strategies ensure that there is a balance between AI automation and human intervention, enhancing accountability by minimising the reliance on fully autonomous AI systems, and ensuring alignment with ethical guidelines.

2.5.3 Privacy

According to Shahriar et al. (2023), privacy risks can be categorised into the risk of identification, the risk of inaccurate decisions, the risk of non-transparent AI, and the risk of non-compliance with regulations. The integration of data minimisation practices, which involve collecting and retaining only the necessary data, along with the application of privacy-preserving machine learning (PPML) techniques, which enhance model accuracy while safeguarding privacy, can significantly reduce the risks associated with risk identification and the potential for inaccurate decisions.

Dhirani et al. (2023) contend that addressing privacy issues in AI requires a multifaceted approach, including obtaining user consent, implementing robust security measures, developing comprehensive regulations and standards, and ensuring compliance with data protection laws across different jurisdictions. Obtaining explicit user consent for data usage and storage is fundamental to respecting user privacy, and this must be coupled with stringent security measures to prevent unauthorised access and protect sensitive information.

Finally, Dhirani et al. (2023) argue that developing and aligning comprehensive regulations and standards, such as those provided by international policymakers and professional bodies, are critical for mitigating privacy risks. However, ensuring these frameworks are interoperable and consistently applied across different regions and industries, such as the insurance industry, is essential for their effectiveness. Aligning AI deployments with existing data protection regulations, like the GDPR in Europe and POPIA in South Africa, can provide a robust framework for ethical data handling.

2.5.4 Regulations and frameworks

De Almeida et al. (2021) state that the rapid integration of AI into various industries, including insurance, has necessitated the establishment of standardised frameworks and industry-specific guidelines to ensure ethical, efficient, and transparent operations. Standardisation and industry-specific approaches are critical to addressing the unique challenges posed by AI technologies within the insurance industry while fostering consistency across global and local contexts.

Moreover, to ensure the responsible use of AI, ethical frameworks must be specifically tailored to the needs of individual organisations. The study by Pant et al. (2024) emphasises the importance of developing customised ethical guidelines for organisations adopting AI technologies. Their study highlights that ethical frameworks should align with an organisation's unique operational, cultural, and regulatory contexts to enhance their practical applicability and overall effectiveness.

Li and Goel (2024) emphasise the necessity of automated and continuous auditing in AI systems to ensure ethical compliance, scalability, and efficiency. They argue that real-time auditing mechanisms enhance transparency and accountability, enabling AI governance frameworks to identify and mitigate potential risks proactively. By leveraging automation, auditing processes can be seamlessly integrated into AI development and deployment, ensuring that ethical considerations are maintained throughout the model lifecycle. This perspective supports the view that effective governance requires dynamic, real-time monitoring rather than periodic manual audits, thus enhancing the adaptability and responsiveness of AI regulatory mechanisms.

Bharadhwaj et al. (2021), however, caution that the lack of interpretable AI structures presents significant challenges to implementing meaningful audits. They argue that many contemporary AI models, particularly deep learning architectures, operate as black boxes, making it difficult to assess how decisions are made and whether they adhere to ethical principles. This perspective challenges the assumption that auditing can be seamlessly integrated into AI governance frameworks, highlighting the need for improved model interpretability and explainability before large-scale audit mechanisms can be effectively deployed. The study suggests that without significant advancements in AI transparency, automated audits may struggle to provide meaningful insights, thus complicating their role in ensuring RAI deployment.

2.5.5 Adaptability and continuous improvements

Birkstedt et al. (2023) emphasise that AI governance should be approached as an ongoing and iterative process that evolves alongside technological advancements and the dynamic needs of organisations. They highlight the critical role of feedback loops in evaluating the effectiveness of governance structures, suggesting that continuous refinement and improvement are essential for maintaining robust AI governance frameworks. By regularly assessing and updating these structures, organisations can better address emerging challenges and align governance with evolving ethical, regulatory, and operational requirements.

Building on the notion of continuous improvement, Zarifis et al. (2019) argue that AI serves as a powerful driver of innovation, offering organisations the potential to optimise processes, enhance decision-making, and secure a competitive advantage. However, leveraging these opportunities requires more than just adopting advanced technologies – organisations must cultivate a skilled and knowledgeable workforce capable of navigating and applying AI-driven solutions effectively. This underscores the necessity of fostering a culture of continuous learning and development, with an emphasis on equipping employees with AI-related skills to support innovation and adaptability.

By integrating continuous monitoring with a focus on adaptability, organisations can establish resilient governance and operational frameworks that not only address current challenges but also position them to harness future opportunities in the rapidly evolving landscape of AI.

2.5.6 Aim for efficiency and accuracy

Eling et al. (2022) provide a detailed exploration of how AI facilitates process automation within the insurance sector. The authors highlight the significant advantages of AI-driven process automation, which greatly boosts operational efficiency across various business functions. By minimising the likelihood of human error, AI streamlines workflows and enhances accuracy. Additionally, it plays a crucial role in expediting essential business processes, particularly in areas such as claims management, where swift and precise handling is vital; underwriting, which benefits from improved data analysis; and customer service, where automated responses ensure timely and effective communication. Overall, these advancements contribute to a more agile and responsive organisation.

Zarifis et al. (2019) highlight the transformative potential of AI in enhancing efficiency by automating routine processes, allowing insurers to shift from reactive to proactive operations. AI-driven automation has been shown to streamline workflows, reduce inefficiencies, and optimise insurance processes, enabling organisations to focus on higher-value tasks. However, the study also identifies significant organisational, regulatory, and technological barriers that hinder the seamless integration of AI within the insurance sector. Many insurers struggle with legacy systems, fragmented data, and resistance to process change, limiting the effectiveness of AI-driven optimisations. This challenges the assumption that AI inherently improves efficiency, emphasising instead that organisational readiness, robust data infrastructure, and adaptability are critical prerequisites for successful AI implementation. Thus, while AI holds considerable promise for operational improvements, its impact is contingent on insurers' ability to overcome integration challenges and align AI adoption with their existing structures.

Lastly, Eling et al. (2022) highlight that AI in the insurance industry primarily functions to automate processes and decision-making, thereby reducing dependence on human intervention across traditional operational roles. This perspective challenges the notion that AI serves solely as a tool for workforce empowerment, instead suggesting that its integration is inherently geared toward replacing many manual and decision-based tasks. Consequently, the assumption that human adaptability alone can safeguard against job displacement is called into question, as the widespread adoption of AI may lead to a fundamental reduction in workforce demand rather than merely necessitating reskilling efforts.

2.5.7 Building ethical awareness and culture

Migdadi et al. (2024) highlight the positive correlation between AI ethical awareness and the intention to use AI, reinforcing the notion that understanding AI technology is a foundational step in identifying and addressing ethical challenges. Their findings suggest that individuals who possess a greater awareness of AI ethics are more likely to engage with AI responsibly, emphasising the need for structured education and training programmes.

Similarly, Durant et al. (2022) underscore the critical role of ongoing AI ethics education in mitigating ethical concerns, particularly within industries like insurance, where AI adoption is accelerating. However, despite increasing AI implementation, gaps in AI literacy persist, necessitating continuous professional development initiatives to ensure that employees comprehend both the risks and opportunities associated with AI technologies.

While ethical awareness is crucial, the studies also reveal that awareness alone does not necessarily lead to ethical compliance. Migdadi et al. (2024) caution that while individuals with high AI ethical awareness are more inclined to adopt AI, this does not directly translate into responsible or ethical AI decision-making. This finding challenges the assumption that education alone is sufficient to ensure ethical AI use, highlighting the importance of strong compliance mechanisms, regulatory oversight, and accountability frameworks to reinforce ethical standards in AI deployment. Therefore, while education and awareness serve as essential building blocks, they must be complemented by formal governance structures to ensure ethical adherence in AI-driven environments.

2.5.8 Developing AI ethical frameworks

Pant et al. (2024) emphasise the importance of customised ethical guidelines for organisations implementing AI technologies, arguing that ethical frameworks must be tailored to the specific operational, cultural, and regulatory contexts in which they are applied. This perspective underscores the need for practical and context-sensitive approaches to AI governance, ensuring that ethical principles are not merely theoretical but integrated into business operations in a meaningful way. By advocating for adaptable governance models, the study highlights the necessity of aligning AI ethics with industry-specific requirements, enabling organisations to effectively navigate both ethical and compliance challenges.

While Filabi and Duffy (2021) acknowledge the value of extending existing governance frameworks, they also contend that new AI-specific governance mechanisms are necessary to address the unique risks posed by ML and big data. The opacity and complexity of AI-driven decision-making, particularly in areas such as insurance underwriting, require more than just modifications to current structures. Filabi and Duffy (2021) caution that existing governance frameworks may not be sufficient to manage the ethical and regulatory challenges associated with AI deployment and suggest that organisations must develop novel oversight mechanisms to ensure fairness, transparency, and accountability. This challenges the assumption that extending traditional governance models is enough, reinforcing the argument that AI governance must evolve to accommodate the distinct ethical dilemmas introduced by emerging technologies.

2.5.9 Collaboration and expertise

Keller (2020) underscores the value of leveraging external expertise to develop RAI frameworks, supporting the argument that organisations benefit from consulting specialists to understand best practices in AI ethics. This reinforces the idea that external insights contribute to shaping AI governance strategies and mitigating ethical risks. However, while collaboration with industry bodies and regulators is often advocated for standardising AI ethics, Keller (2020) highlights the significant challenges in achieving universal AI governance frameworks across the insurance sector. Variations in data sources, regulatory requirements, and business models create obstacles to establishing a single standardised approach, suggesting that AI governance must remain flexible and adaptable rather than adopting rigid, uniform policies that may not account for industry-specific complexities.

2.5.10 Balancing innovation with ethics

Wang and Wu (2024) emphasise the critical challenge of deploying AI technologies while maintaining public trust, underscoring the necessity for organisations to integrate AI responsibly to preserve customer confidence. They highlight the delicate balance between AI accuracy and user sentiment, asserting that trust in AI-driven decisions depends not only on technical precision but also on how these technologies align with consumer expectations and ethical standards. This reinforces the argument that AI adoption must be approached with a dual focus on innovation and transparency, ensuring that AI systems are both effective and socially acceptable.

While advocating for AI innovation, Wang and Wu's (2024) study cautions against an unregulated AI landscape, warning that the lack of oversight can result in ethical blind spots and governance failures. This challenges the view that reducing AI regulations inherently fosters progress, suggesting instead that a measured approach to regulation is necessary to prevent unintended consequences such as bias, privacy violations, and accountability gaps. The authors argue that effective AI governance must strike a balance between enabling technological advancement and safeguarding ethical integrity, ensuring that AI serves both business interests and broader societal needs.

2.5.11 Ensure data integrity

Van Bekkum et al. (2024) emphasise the critical role of high-quality data in mitigating bias and discrimination within AI-driven insurance models. They underscore that data-intensive underwriting, if reliant on poor-quality or unrepresentative datasets, can inadvertently reinforce biases and lead to unfair differentiation in insurance pricing. Ensuring that datasets are comprehensive, unbiased, and representative is essential for fostering fair and equitable risk assessments. Van Bekkum et al. (2024) assert that insurers must implement rigorous data validation and quality control measures to prevent discriminatory pricing structures and uphold ethical AI deployment in underwriting processes.

2.5.12 Proposition 4

Implementing a multifaced strategy that addresses fairness, transparency, explainability, privacy, and accountability is essential to ensure the ethical use of AI technologies in the insurance industry.

2.6 Analytical framework

2.6.1 Theoretical framework

A theoretical framework presented by Fabris et al. (2021) primarily examines ethical risks from the perspective of fairness, discrimination, and transparency, focusing on the process of pricing and quoting for insurance policies. This theory is limited in its scope of ethical risks as important concepts such as privacy and data quality are not considered.

Mullins et al. (2021) examine ethics from a different perspective, as the theoretical framework they present is based on relational ethics. The framework emphasises the ethical considerations arising from the relationships and interactions between various stakeholders involved in the implementation and use of AI and big data in the insurance industry. The goal is to maximise the benefits derived from AI in the insurance industry and for its stakeholders. As Mullins et al. (2021) examines various stakeholders, it was limited from a customer-centric perspective.

Owens et al. (2022) integrate several viewpoints when it comes to a theoretical framework. These include XAI, which can increase stakeholder trust inversely, and regulatory considerations, which provide a theoretical guide to transparency, fairness, and accountability in AI systems. Again, there is a gap in the framework when it comes to a customer-centric perspective and addressing specific ethical risks pertaining to the customer. The absence of a customer-centric perspective in Owens et al.'s (2022) framework raises ethical concerns, particularly regarding how AI-driven decisions impact policyholders. The following theories seek to address this.

2.6.1.1 Ethical theory: Utilitarianism

Utilitarianism focuses on the consequences of an action rather than the action itself, and therefore this theory looks to maximise good outcomes while simultaneously minimising bad ones (Driver, 2009). The potential benefits of examining ethics from a utilitarian lens in the insurance industry can include productivity improvements, increased accuracy of risk assessments, and favourable pricing. However, doing so could also increase the risk of bias and a lack of transparency for specific segments of the population. This could be considered acceptable if most of the population is unaffected by it.

2.6.1.2 Ethical theory: Deontology

Deontology emphasises the importance of rules and intentions over consequences, therefore actions are morally right if they are done in accordance with a moral rule or principle, regardless of the outcome (Alexander & Moore, 2021). This study also adopted a deontological approach to ethics to examine AI systems' adherence to moral principles such as fairness, transparency, and privacy. However, this approach could stifle innovation, which can improve efficiency, the accuracy of risk assessments, and pricing.

The selection of utilitarianism and deontology as the primary ethical frameworks for this study is justified by their complementary perspectives in assessing the ethical implications of AI in the insurance industry. Utilitarianism, which prioritises maximising positive outcomes and minimising harm, is particularly relevant in evaluating AI's potential to improve efficiency, accuracy, and affordability in insurance decision-making. However, a strictly utilitarian approach may justify ethically questionable practices such as biased risk assessments or opaque decision-making if the overall benefits to the majority outweigh the disadvantages to a marginalised group. This limitation necessitates the incorporation of deontological ethics, which places inherent value on fairness, transparency, and privacy, ensuring that AI-driven decision-making aligns with fundamental moral duties rather than merely its consequential benefits.

The combination of utilitarianism and deontology provides a balanced ethical lens that mitigates the shortcomings of each theory when considered in isolation. While utilitarianism allows for a pragmatic evaluation of AI's impact on efficiency and accessibility, deontology safeguards against ethical trade-offs that may disproportionately disadvantage vulnerable populations. This dual approach ensures that AI implementation in the insurance industry is both beneficial and morally justifiable, promoting innovation while maintaining accountability to ethical standards. By integrating these perspectives, the study offers a comprehensive ethical assessment that acknowledges both the practical advantages of AI-driven decision-making and the moral imperatives that must guide its application.

2.6.2 Conceptual framework

The conceptual framework for this study was based on utilitarianism and deontological ethical theories.

2.6.2.1 Utilitarian ethics

Utilitarian principles, as mentioned earlier, focus on maximising overall benefits such as efficiency, cost reduction, customer satisfaction, and favourable pricing. This theory ensures that AI decision-making solutions maximise positive outcomes while minimising harm.

Maximising overall good

Insurance companies should strive to achieve the greatest amount of good for the greatest amount of people while minimising harm (Zoshak & Dew, 2021). This includes fair pricing, efficient claims processing, and fair underwriting for clients.

Cost-benefit analysis

Insurance companies need to continuously assess the impact of AI decisions on clients to balance the interests of the insurance company, such as profitability, with ethical considerations and ensure decisions contribute positively to social welfare (Singh & Mishra, 2018).

Accuracy and efficiency

Efficiency can be improved by the ADM systems through the implementation of advanced optimisation algorithms to streamline processes such as pricing, risk assessment, policy underwriting, and claims processing (Zoshak & Dew, 2021). Accurate information and predictions about the consequences of actions are crucial for making decisions that truly maximise utility. This includes decisions on pricing, underwriting, and claim processing (Cervantes et al., 2016). Decisions based on inaccurate data or faulty assumptions can lead to suboptimal outcomes, which would be contrary to the utilitarian goal of maximising overall well-being

Continuous improvement

As ethical risks are constantly evolving, AI systems need to be regularly reviewed and updated to address emerging ethical issues and improve their mitigation of client well-being (Singh & Mishra, 2018). Therefore, insurance companies would benefit from encouraging a culture of continuous improvement, where feedback from clients and regulators is used to refine AI-driven processes.

2.6.2.2 Deontological ethics

Deontological principles, as mentioned earlier, focus on fairness, transparency, and privacy. This theory ensures that AI decision-making solutions adhere to ethical guidelines such as fairness and transparency.

Adherence to moral rules

AI decision-making systems need to be developed to ensure that the established moral rules are followed such as fairness, honesty, and respect for individual rights (Kumar et al., 2022). Insurance companies, therefore, need to ensure their algorithms do not unfairly discriminate against individuals based on race, gender, or other protected attributes.

Universalisability of actions

Insurance companies must take diligent actions to guarantee that all clients and potential clients are treated fairly and equally. This includes implementing policies and practices that foster transparency, inclusivity, and consistency in their interactions, and ensuring that every individual receives the same level of respect and service regardless of their background or circumstances (Micewski & Troy, 2007). Therefore, AI decision-making processes should be consistent and applied to all similar cases.

Transparency and explainability

ADM systems in insurance processes should be transparent in their decision-making processes, how they come to those outcomes, and be explainable in how the ADM reached that decision for clients and potential clients (Kumar et al., 2022). This helps build trust and enables clients to understand how decisions affecting their coverage or claims are made.

Responsibility and accountability

Insurance companies should establish clear accountability mechanisms for AI decisions to ensure that ADM systems adhere to ethical standards (Zoshak & Dew, 2021). This includes having human oversight of AI outputs and ensuring that there are processes in place to address any ethical breaches or errors in decision-making.

2.6.2.3 Breakdown of the conceptual framework

The conceptual framework integrates utilitarian and deontological ethical theories to guide AI decision-making processes in the insurance industry. The goal is to achieve ethical outcomes while leveraging AI technologies in key insurance processes such as underwriting, claim management, and pricing.

The framework begins with the foundational ethical theories of utilitarianism and deontology. Utilitarian principles focus on maximising overall good through cost-benefit analysis, achieving high accuracy and efficiency, and ensuring continuous improvement. On the other hand, deontological principles emphasise adherence to moral rules, the universalisability of actions, transparency, explainability, responsibility, and accountability. These principles ensure that AI decision-making systems in the insurance industry are designed and implemented in a way that adheres to ethical norms and moral duties. These principles aim to optimise the benefits of AI applications while balancing efficiency with ethical considerations.

AI decision-making solutions are central to the framework and are applied across underwriting, claim management, and pricing processes. Underwriting involves evaluating risks and determining the terms and pricing of insurance policies, claim management handles and processes insurance claims, and pricing sets the cost for insurance policies based on risk assessment and other factors. By incorporating ethical principles into these AI-driven processes, the framework ensures that AI decision-making systems enhance decision-making while adhering to ethical standards.

The outcomes of applying AI in these processes, guided by ethical theories, are measured in terms of accuracy, efficiency, fairness, transparency, unbiased decisions, and privacy protection. Accuracy and efficiency ensure precise and effective processing in underwriting, claims, and pricing. Fairness and transparency maintain equitable treatment and clear decision-making processes. Unbiased decisions and privacy protection mitigate biases in AI models and safeguard customer privacy, respectively.

In summary, the conceptual framework ensures that AI decision-making in the insurance industry is ethically guided, integrating both utilitarian and deontological principles to maximise benefits while adhering to moral principles. This approach helps mitigate ethical risks and achieve positive outcomes for insurers, customers, regulators, and society.

Figure 2.1 presents the conceptual framework as explained above.

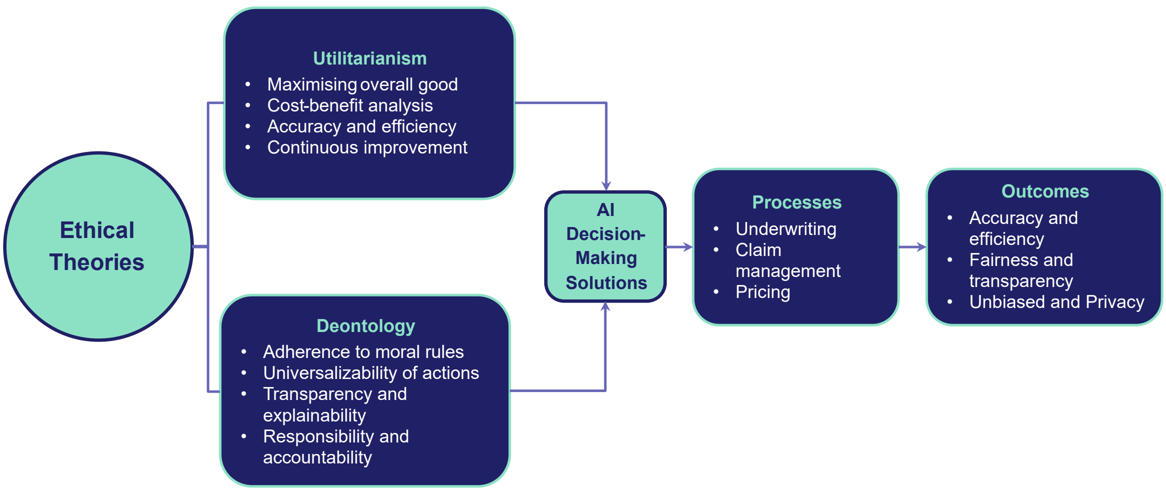


Figure 2.1: Conceptual framework

2.7 Conclusion

The literature review highlights the profound ethical implications of AI-driven decision-making within the insurance industry. AI's transformative potential in automating claims processing, enhancing risk assessment, and personalising pricing comes with significant ethical risks, including algorithm bias, data quality and integrity issues, privacy concerns, and challenges in transparency and accountability.

Algorithm bias in AI systems, as evidenced by the study on Italian car insurance premiums, highlights the risk of discrimination based on attributes like gender and birthplace. The quality of data used by AI algorithms also poses ethical challenges, as poor data quality can lead to inaccurate or biased decisions. Privacy concerns are prominent, with risks of data breaches and reidentification of individuals from anonymised datasets being critical issues. Additionally, the black-box nature of AI decision-making processes exacerbates concerns about transparency and accountability.

Current strategies to address these ethical risks include the implementation of principles of RAI, regulatory compliance (such as the GDPR), and fair-ML. However, these strategies often fall short in practice. Regulatory frameworks like the GDPR provide a foundation for ethical AI use, but they are not always comprehensive or rapidly evolving enough to keep pace with technological advancements. Ethical frameworks adopted by insurance companies, while theoretically sound, frequently suffer from gaps between prescribed principles and their practical application.

To ensure the ethical use of AI technologies in the insurance industry, a multifaceted approach is essential. This includes enhancing transparency and explainability, ensuring fairness through various stages of AI decision-making, strengthening accountability mechanisms, and rigorously protecting privacy. Strategies such as regulatory measures tailored for AI, ethical AI governance, and HITL approaches are crucial for developing robust ethical guidelines and maintaining human oversight.

The integration of utilitarian and deontological ethical theories provides a comprehensive framework for guiding AI decision-making processes in insurance. Utilitarian principles emphasise maximising overall benefits, such as efficiency and customer satisfaction while minimising harm. Deontological principles focus on adherence to moral rules, transparency, and accountability. This combined approach aims to balance innovation with ethical considerations, ensuring that AI technologies are used responsibly and ethically in the insurance industry.

The next chapter, Chapter 3, presents the methodology used to conduct the research.

CHAPTER 3. RESEARCH METHODOLOGY

3.1 Introduction

This chapter outlines the research design and methodology used to explore the ethical implications of AI in the insurance industry. To begin with, it discusses the research paradigm adopted. This is followed by the research approach and design, and the data collection method. The sampling techniques used to identify participants with relevant expertise in insurance and knowledge or experience with AI are described. The research instrument used is then described. The steps in the analysis of the data are presented and the possible limitations and challenges of the study are outlined. The issue of quality assurance is discussed and the chapter ends with the ethical considerations of the study.

3.2 Research paradigm

In this study, the interpretivism paradigm was selected to explore the ethical implications of AI in decision-making within the insurance industry. Interpretivist researchers aim to understand complex social experiences by collecting and interpreting data that reflects participants' subjective experiences (Saunders et al., 2019). This paradigm focuses on understanding the subjective meanings and experiences of individuals, which aligns well with exploring the nuanced ethical concerns and perceptions of stakeholders in the insurance industry. The paradigm's focus on subjective experiences aligns with the ontological view that reality is complex and open to multiple interpretations (Saunders et al., 2019).

Ontology

The nature of the reality (AI and the insurance industry) being studied was complex, and therefore multiple interpretations could be derived based on the individual (Saunders et al., 2019). In this study, the ethical implications of AI were described through the experiences and interpretations of industry experts.

Epistemology

What constitutes acceptable knowledge can include narratives, stories, perceptions, and interpretations (Saunders et al., 2019). By interviewing industry experts, the study sought to gather insights based on their experiences and interpretations, which are essential for understanding the ethical implications of AI in the insurance industry.

Axiology

Interpretivism acknowledges that research is value-bound and, therefore, the researcher's values influence the research process and outcomes (Saunders et al., 2019). Hence, ethical considerations such as ensuring the respectful treatment of interview participants and considering the implications of the findings, were central to the research process.

3.3 Research approach

This study adopted a qualitative research approach to explore the ethical implications of AI in decision-making within the insurance industry, focusing on customers/policyholders and interviewing industry experts. The qualitative approach was chosen for the following reasons:

- The qualitative research approach aims to investigate complex and context-specific issues while retaining deep meaning when interpreting data (Bhandari, 2020b). Ethical implications of AI in the insurance industry are multifaceted and require an in-depth exploration to understand the unique experiences and perspectives of various stakeholders.
- Qualitative research allows researchers to collect detailed data using methods such as interviews and focus groups, allowing for the capturing of rich and descriptive information about the experiences of those interviewed (Bhandari, 2020b).

3.4 Research design

The qualitative case study design was the most suitable research design for this study, as it allowed for a comprehensive, context-specific, and flexible exploration of the ethical implications of AI in decision-making within the insurance industry (McCombes, 2019).

3.5 Data collection method

Data collection is a structured method of obtaining observations and measurements in a systematic manner (Bhandari, 2020a). This study collected data via semi-structured interviews. Conducting semi-structured interviews with industry experts allowed for flexibility in exploring various themes while ensuring that key topics related to the ethical implications of AI within the insurance industry were covered. This method enabled the collection of rich, detailed data that captures the participants' perspectives, experiences, and insights on the ethical implications of AI in decision-making.

Data collection continued until thematic saturation was achieved. This was evidenced after the ninth interview, whereby there were no new themes emerging from the participants' responses. This saturation criterion ensured that all perspectives relevant to the research questions were captured.

3.6 Population and sample

3.6.1 Population

The population for this study consisted of experts in the insurance industry who have knowledge and/or experience related to AI and the applicable processes. This included professionals, technology specialists, and legal and compliance officers working in the insurance industry. The inclusion criteria were individuals in the insurance industry with knowledge and/or experience in AI. There was no specific organisation selected, as the participants came from different insurance organisations of varying AI adoption maturity.

3.6.2 Sample and sampling method

Purposive sampling

This non-probability sampling method involves selecting participants based on specific criteria and the research objectives (Nikolopoulou, 2022a). Purposive sampling was used in this study to ensure that the selected participants had the relevant experience and knowledge required to provide meaningful insights into the ethical implications of AI in the insurance industry.

Snowball sampling

This non-probability sampling method involves asking initial participants to refer other potential participants who meet the study criteria (Nikolopoulou, 2022b). Snowball sampling is useful for reaching a broader network of experts, especially in specialised fields such as AI in the insurance industry. It helps in identifying additional knowledgeable participants who may not be immediately accessible through other means. In the case of this study, the initial respondents who were recruited through the purposive sampling method were asked to refer other potential participants who had the same expertise and who could provide rich insights into the implications of AI-driven decision-making and were well-versed in ethical risks related to the insurance industry.

Therefore, this study utilised the purposive and snowball sampling methods in selecting suitable participants. Nine participants took part in the study. This allowed for in-depth analysis while ensuring a diverse set of perspectives.

3.7 Research instrument

The selected research instrument was a semi-structured interview guide (Please see Appendix C for the data collection instrument). The guide was aligned with the research questions and the conceptual framework.

3.7.1 Sequential development of themes from research questions

The first research question, “What ethical risks arise from using AI in decision-making within the insurance industry?”, served as the foundation of the study by identifying and categorising the key ethical risks inherent in AI-driven decision-making processes. The themes emerging from this question provided a basis for understanding the challenges that necessitate mitigation efforts. These themes fed into research question 2, as they established the risks that existing mitigation strategies sought to address. They also informed the themes in research question 3, which evaluated the effectiveness of those mitigation strategies. Without a clear articulation of the risks, it would have been impossible to assess how well current approaches are mitigating them. Additionally, the insights gained from identifying ethical risks contributed to research question 4, as they highlighted gaps that future strategies must aim to address.

Building upon the ethical risks identified, the second research question, “How are these ethical risks currently addressed in the insurance industry?”, examined the existing strategies and mechanisms used to mitigate AI-related ethical concerns. The themes emerging from this question provided a structured overview of current interventions, including regulatory compliance, ethical AI frameworks, and governance structures. These themes directly fed into those of research question 3, as they established the baseline against which effectiveness is assessed. If certain risks remain inadequately addressed, the effectiveness of existing strategies is called into question. Additionally, these themes influenced research question 4 by highlighting areas where strategies require improvement or innovation, ensuring that future mitigation efforts build upon the strengths and limitations of current approaches.

The third research question, “How effective are the current strategies used to address ethical risks in the insurance industry?”, critically evaluated the impact and adequacy of existing mitigation strategies. This question built on the second by moving beyond the identification of strategies to assess their practical implementation and measurable success. The themes emerging from this question not only reflected the successes and limitations of current strategies but also fed directly into research question 4, which sought to propose improved mitigation measures. By identifying gaps and shortcomings in existing efforts, these themes helped shape the development of future strategies. Additionally, the themes from research question 3 indirectly linked back to those in research question 1, as any persistent ethical risks, despite mitigation efforts, suggest that certain risks require more targeted intervention.

Finally, the fourth research question, “What strategies can be implemented to ensure AI technologies are used ethically in the insurance industry?”, was forward-looking and aimed to identify improvements and innovations for ethical AI governance. The themes emerging from this question built directly on the findings from research questions 2 and 3, ensuring that the proposed strategies addressed identified gaps and enhanced the effectiveness of existing approaches. These themes were also informed by those from research question 1, as any newly proposed strategies must account for the full spectrum of ethical risks to ensure comprehensive mitigation. Thus, the themes from research question 4 served as a culmination of the thematic insights derived from the preceding questions, integrating risk identification, assessment of current strategies, and evaluation of their effectiveness into a coherent set of recommendations.

3.8 Procedure for data collection

Participants were approached and recruited based on the sampling strategy defined above. Online interviews were then scheduled with the participants, as this was the preference of all the participants. The interviews were recorded (consent to do so was obtained from the participants), and detailed key points and non-verbal cues were captured. After the interviews, the recordings were transcribed.

3.9 Data analysis and interpretation of the results

Thematic analysis was used to analyse the data collected to uncover themes that emerged from the participants’ responses to the questions posed (Jansen, 2021). The six-phase process of thematic analysis, as described by Braun & Clarke (2006), was utilised for the thematic data analysis:

Step 1: Get familiar with the data

The first step in thematic analysis involves the researcher immersing themselves in the interview data by reading through it multiple times to thoroughly understand its content (Braun & Clarke, 2006). This includes transcribing the audio recording into text. Doing so assisted with developing preliminary thoughts about what to code and how these codes relate to the research topic, research questions, and conceptual framework. The researcher maintained the initial coding thoughts and processes on the qualitative data analysis software, ATLAS.ti, thus enhancing reliability and consistency (Please see Appendix D for a sample of the codes captured on ATLAS.ti).

Step 2: Search for patterns or themes in the codes

After initial coding, the next step was to search for patterns or themes within these codes (Braun & Clarke, 2006). Similar codes were grouped together to identify broader themes that captured significant aspects of the data related to the research questions. During this stage, subthemes, which are more specific aspects within a broader theme, might also emerge. The researcher then recorded the process in ATLAS.ti, noting how codes were combined to form themes.

Step 3: Review themes

In this step, the researcher reviewed all the identified themes to ensure they accurately reflected the data (Braun & Clarke, 2006). The researcher checked for coherence within themes and distinctness between them and verified that they encompassed all relevant data points. The themes were refined as necessary by merging or splitting them to better represent the data. ATLAS.ti was used to note how themes were supported by the data and how they aligned with the research questions.

Step 4: Finalise themes

Once themes were reviewed and refined, the next step was to finalise them. This involved defining and describing each theme in detail, ensuring that they were accurately named and reflected their essence (Braun & Clarke, 2006). The researcher ensured that the finalised themes aligned with the research questions. The final descriptions of the themes and their relevance to the research were documented in ATLAS.ti.

Step 5: Produce the report

The final step involved writing the findings in a structured report, which included an introduction, methodology, results, and conclusion (Braun & Clarke, 2006). The report detailed the process of the analysis, providing enough information for readers to evaluate its rigour. The researcher supported the findings with quotations from the data to demonstrate the grounding of themes in the data. Additionally, all findings were ensured to be relevant to the research questions, maintaining the report's focus and coherence.

3.10 Possible limitations and challenges of the study

- Gaining access to busy industry experts was challenging. This was mitigated by building relationships with key stakeholders within the industry. The researcher was also flexible with scheduling and provided assurances of confidentiality and anonymity to encourage participation.
- The researcher's own values and beliefs can influence the interpretation of data, especially in a study focused on ethical implications, if not controlled. This was mitigated by regularly reflecting on and documenting any biases that might affect the research.
- Ethical implications such as fairness, transparency, and accountability can be interpreted differently by different participants, making it challenging to develop consistent themes. This was mitigated by using the conceptual framework based on utilitarianism and deontology to guide the analysis and interpretation of ethical implications. The researcher also ensured clarity and consistency in definitions and interpretations.

3.11 Quality assurance

3.11.1 Transferability

This study on the ethical implications of AI in decision-making within the insurance industry meets the transferability criteria by providing detailed and context-rich descriptions of the industry, the specific AI applications examined, and the ethical implications observed (Tracy, 2010). These descriptions allow others to determine the applicability of the findings to their own settings.

Utilising purposive and snowball sampling ensured that the selected participants had the required knowledge and experience to provide meaningful data for the study. The diversity within the sample, covering a broad spectrum of perspectives from industry experts, enhanced the applicability of the findings across different scenarios within the industry.

To maximise transferability, the clear description of the research design and how data was collected and analysed will assist other researchers in replicating or adapting the research within their contexts.

This study on the ethical implications of AI in decision-making within the insurance industry meets the internal credibility criteria by employing strategies that ensure the accuracy and trustworthiness of the findings. Through triangulation, the researcher collected data from interviews and cross-verified it with relevant literature to enhance the consistency of the findings.

Detailed documentation of the research process, including decisions made and justifications for methodological choices, created an audit trail that enhances transparency and allows others to follow the research steps.

Finally, to maximise internal credibility, the researcher maintained reflexivity by regularly reflecting on and documenting their biases and how they could have affected the research. Engaging in peer debriefing with colleagues and the supervisor provided alternative perspectives and helped identify any potential biases or errors in analysis. These strategies ensured that the study offers a credible and trustworthy exploration of the ethical implications of AI in decision-making within the insurance industry.

3.11.2 Dependability

This study on the ethical implications of AI in decision-making within the insurance industry meets the dependability criteria by ensuring that the research process was logical, traceable, and documented. This was achieved through the use of an audit trail, where the researcher meticulously recorded the research design, data collection methods, and data analysis procedures (Tracy, 2010). Such detailed documentation ensured transparency and provided a clear pathway for others to understand and potentially replicate the study.

To further maximise dependability, the researcher engaged in review and debriefing sessions, where colleagues and the supervisor critiqued the research process and findings, offering insights and identifying any inconsistencies (Tracy, 2010). This collaborative approach helped ensure that the research findings were consistent and the methods were applied correctly. Additionally, maintaining a reflexive journal throughout the research process allowed the researcher to document reflections, decisions, and adjustments, providing a comprehensive record of the research journey (Tracy, 2010). These strategies enhanced the dependability of the study, ensuring it was reliable, and that the findings were consistent over time.

3.12 Ethical considerations

The study adhered to the research ethical requirements set by the University of the Witwatersrand. Ethical clearance was applied for and received prior to the commencement of data collection (Please see Appendix E for the ethical clearance certificate). All participants were fully informed of their involvement in the research and provided their consent by completing and signing the applicable consent form (Please see Appendix A for the participation information sheet and Appendix B for the consent form). The interview process and data collection were conducted in accordance with the ethical requirements of the POPIA. Participants were provided the opportunity to either meet in person or online via video conferencing. Consent to record and transcribe the interviews was also obtained from the participants. To ensure confidentiality, participants’ identities were protected when quotes from the interviews were included in the research report.

3.13 Alignment to Framework

The study’s dual-ethical theoretical framework, grounded in utilitarian and deontological principles served as a foundational guide for the overall research design, data collection and thematic analysis. This alignment ensured conceptual consistency between what the study sought to explore (ethical risks and mitigation strategies) and how the data were collected and interpreted.

In particular, utilitarian ethics, concerned with the balance of harms and benefits to maximise collective good, guided the exploration of perceived outcomes of existing mitigation strategies and the effectiveness of AI implementations. Deontological ethics, which emphasise adherence to moral duties and the intrinsic rights of individuals, informed the formulation of questions related to fairness, transparency, and respect for policyholders in AI-driven decision-making.

While the study employed an inductive thematic analysis approach to generate themes directly from expert insights, the interpretation of those themes was informed by the underlying ethical framework. Themes were therefore not only understood in relation to participant narratives but also evaluated in terms of their alignment with outcome-based (utilitarian) and principle-based (deontological) ethical reasoning.

3.14 Conclusion

The research methodology chapter provided a detailed account of the research design, sampling methods, data collection techniques, and analytical approach. Ethical research practices, including compliance with data protection regulations, were strictly followed. This methodological framework ensures the reliability and validity of the study, allowing for meaningful insights into the ethical implications of AI in insurance.

The next chapter, Chapter 4, presents the findings and their interpretation.

CHAPTER 4. PRESENTATION OF FINDINGS

4.1 Introduction

This chapter presents the results of the thematic analysis conducted on expert interviews. The findings are structured around the four research questions, focusing on identifying ethical risks, evaluating current mitigation strategies, assessing their effectiveness, and proposing future improvements. The chapter provides insights into the impacts of ethical risks, setting the stage for the discussion of mitigation strategies.

4.2 Demographics of participants

The selection of the participants was based on purposive and snowball sampling. This was to ensure that the participants met the specific criteria stipulated. A total of nine participants agreed to take part in the study. All the participants had extensive professional experience in the insurance industry. In addition, most participants had managerial positions with work experience ranging from eight to 30 years as depicted in Table 4.1. Their insights provided a strategic high-level overview of the use of AI in meeting strategic objectives. Most importantly, their collective expertise ensured that the study uncovered the ethical risks related to the use of AI in the insurance industry and the potential strategies that could be used to mitigate them. To ensure anonymity, all the participants are referred to by their assigned codes (from P1 to P9).

Table 4.1: Participants' demographics

Code Name	Title	Industry	Experience	Highest Qualification
P1	Manager: AI	Short term insurance	15 years	Master of Business Administration (MBA)
P2	Data Analyst	Long term insurance	9 years	Master of Applied Data Science
P3	Manager: Information Management	Short term insurance	20 years	Master of Business Administration (MBA)
P4	Chief Data Officer	Health insurance	30 years	Post Graduate Diploma in Business Administration
P5	Head: Digital Strategy	Insurance (General)	8 years	Post Graduate Diploma in Accounting
P6	Chief Information Officer (CIO)	Long term insurance	15 years	Bachelor of Business Science Honours in Computer Science
P7	Chief Risk Officer	Insurance (General)	21 years	Doctor of Philosophy (PhD), Informatics
P8	Manager: IT Operations	Insurance (General)	18 years	Doctor of Philosophy (PhD), Information Technology
P9	Head: Machine Learning	Long term insurance	24 years	Master of Big Data Science

4.3 Participants’ perspectives on the use of AI in the insurance industry

As observed during the interviews and the analysis of the data, all the participants agreed that AI has the potential to drive innovation within the insurance industry. This innovation can lead to significant benefits for both organisations and customers, such as enabling more efficient processes and enhancing accuracy in risk assessments.

Participant 4 (P4) highlighted that NLP enables efficiency in customer query responses by generating answers without human intervention.

> *“So it’s a good use for using large language foundation models that have been trained on additional content, so it can be stitched into any one of your channels and potentially answer some questions before there’s a need to refer to a human so it’s just really an efficiency play in our space.” (P4)*

P6 emphasised that robotic process automation (RPA) powered by AI is particularly useful in businesses with older, well-defined processes that have remained unchanged for years. Instead of redesigning or modernising these processes, RPA offers a way to automate repetitive tasks without major system overhauls.

> *“Robotic process automation, and in a legacy business where there’s lots of call it older, there’s been business process defined many years ago, and it just works in a certain way, and using RPA is a nice and easy way to automate some of that thing.” (P6)*

P6 further outlined the benefits of introducing an AI-generated affordability model in the claims underwriting process. The affordability predictive model maximises benefits by improving efficiency and reducing risks. It ensures resources are allocated to customers likely to afford premiums, reducing financial losses for the organisation while protecting individuals from over-commitment.

> *“So we built what we call the affordability predictive model, which basically looks at all of our previous sales information and looks at the propensity for that client to be able to pay their first premium. And we use that to build this model that kind of now works in real-time.” (P6)*

The findings show that AI's application in the insurance industry enhances operational efficiency and decision-making while promoting fairness and customer protection. Thus, the use of AI in the insurance industry aligns with broader insurance organisations' goals of maximising value while mitigating risks.

4.4 Presentation of the findings based on the research questions

The findings of the study are presented in accordance with the themes that emerged from each research question. An inductive approach was used to identify the emerging themes from the participants' responses.

4.4.1 Research question 1: What ethical risks arise from using AI in decision-making within the insurance industry?

The insurance industry can reap numerous benefits from leveraging AI. However, like any technology, there are potential risks, particularly ethical risks, that may arise from its use. Through an inductive approach, the following themes were identified to unpack the potential ethical risks of AI in the insurance industry.

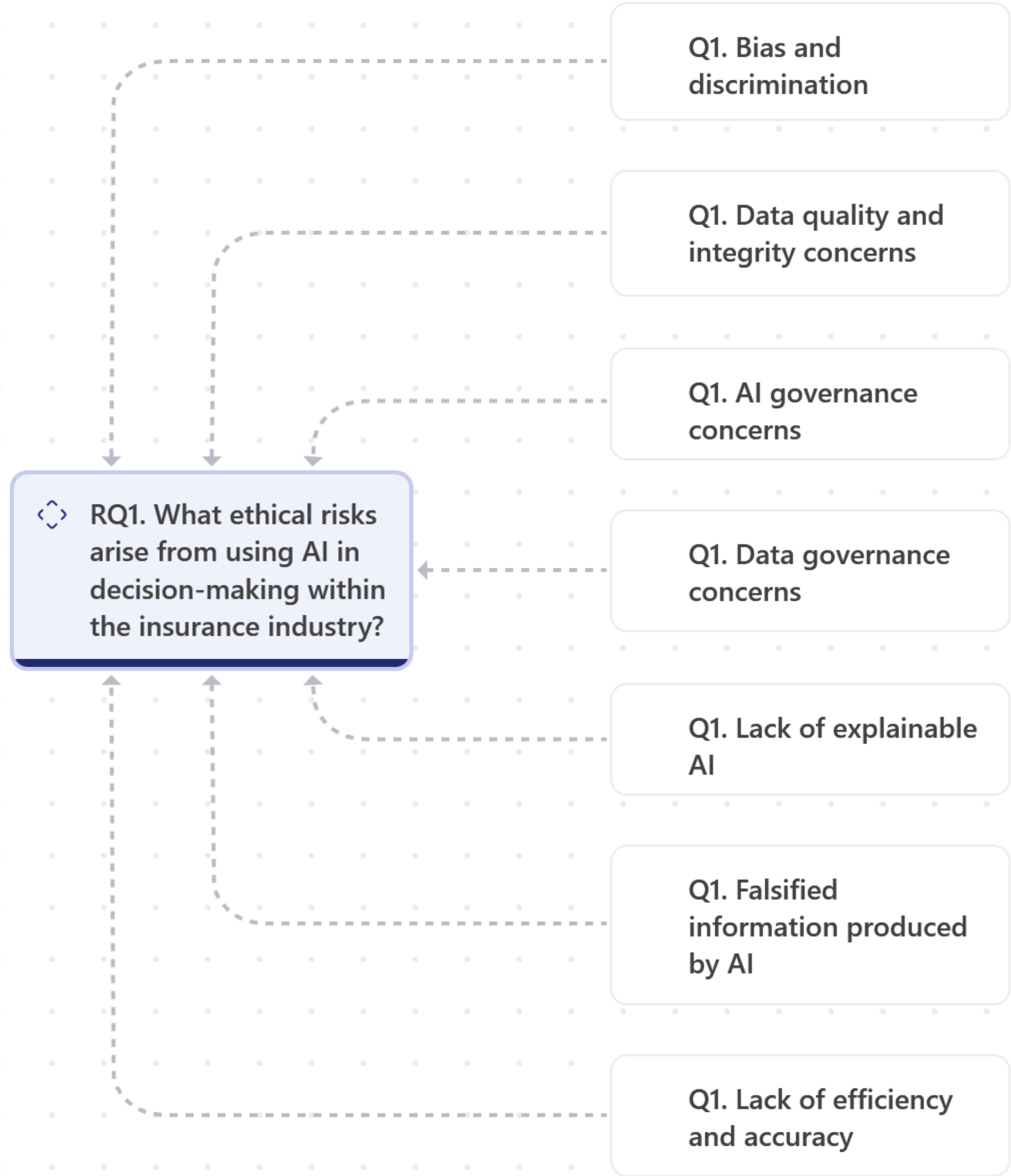


Figure 4.1: Emerging themes for research question 1 (RQ1)

4.4.1.1 Bias and discrimination

Bias and discrimination are critical ethical concerns in the application of AI within the insurance industry. This theme emerged prominently in participant discussions, highlighting how biases can stem from both the design of AI models and the datasets used for training. Participants pointed to the potential of AI to perpetuate historical inequalities, particularly in a context such as South Africa, where socioeconomic disparities are deeply rooted. The discussion below explores the dual sources of bias in AI model design and training data while emphasising the importance of mitigating these risks to ensure fairness, trust, and ethical decision-making.

Bias and discrimination within the AI models

P1 highlighted the risk of reinforcing existing inequalities through AI systems, particularly in a country with a history of socioeconomic disparity such as South Africa:

> *"I think with that, it means that our lower-income individuals and our previously discriminated against individuals will become more discriminated against in these models." (P1)*

Guan et al. (2022) note that AI decision-making is prone to biases in algorithms. These biases may arise from poor design, inadequate testing, or intentional manipulation, leading to unjust or discriminatory outcomes. This is an important consideration, especially in a country like South Africa, which has a history of socioeconomic disparities. It is crucial to avoid reinforcing systemic inequalities that AI models can perpetuate. Failure to mitigate bias can result in unethical outcomes, eroding client trust and fairness. Similarly, P7 elaborated that bias can be introduced based on the way these models make decisions:

> *"The other one is discrimination, discrimination of clients through the way in which the models make decisions or operate or based on the output it produces that allows the results to perpetuate unfair treatment." (P7)*

AI models may introduce biases based on gender, age, culture, and socioeconomic status. P8 alluded to this stating:

> *"Certain things can be a bit discriminatory, as an example, based on either the race or socioeconomic status or age or what have you, and all of that, I think that would have to be taken into consideration to ensure that those challenges are properly mitigated." (P8)*

Bias and discrimination emanating from datasets

Bias can be inherited from the data that is used to train the model, which then drives discrimination against customers. Therefore, if the data is not clean, accurate, and representative of the population, the model may inherit and amplify those biases.

Fabris et al. (2021) point out that if the training data contains historical bias, it can be perpetuated by the algorithm. To this end, P6 emphasised the low quality of data as the potential driver of bias in AI models:

> *"There's certainly a potential flaw that is pretty much based on your data. Is your dataset broad enough? Because if there is a bias, then certainly AI is going to almost over exaggerate that bias that's in the data training sets." (P6)*

P9 echoed this by providing an example of how a biased training dataset can lead to erroneous results:

> *"I mean by biases, just refer to the model. For example, a pricing model, you know, using a biased training set, probably to train the model, which then results in bias that treats a subpopulation unfairly." (P9)*

Similarly, Fabris et al. (2021) studied the pricing algorithms used by Italian automobile insurance companies and found that both gender and place of birth significantly influenced the quoted insurance premiums. Therefore, this calls for a thorough examination of the quality of data used to train ML models to mitigate risks such as noise and inadequate representation, which can impact model performance (Pothuganti, 2018). Additionally, careful monitoring is necessary to prevent overfitting, where the model learns patterns too specifically from the training data or underfitting, where it fails to generalise effectively to unseen data (Pothuganti, 2018).

P1 stated that unclean data is a potential threat that can lead to bias within the data. The participant also emphasised the need to pay particular attention to bias within the data (especially in the context of South Africa, which has a history of racial and gender discrimination):

> *"So the biggest ethical risk that I perceive in AI decision-making is 100% bias. Like I said, if there is, if the data that we feed a model is not clean and not correct and not proper, that model will obviously learn from that data. And if that data has bias in it, which in a country like ours, it probably does, then that model will learn that bias, and that model will then make decisions based on that bias." (P1)*

P1's response emphasised that bias and discrimination are significant ethical risks in AI applications within the insurance industry. These risks can originate from both the design and operation of AI models as well as the data used to train them. Left unaddressed, biases can perpetuate or amplify historical inequalities, leading to unjust outcomes, unfair treatment, and an erosion of trust among customers. Addressing these challenges requires organisations to prioritise robust data quality practices, ensure representative datasets, and implement rigorous oversight of AI model design and decision-making processes. By tackling these issues proactively, organisations can ensure that AI technologies are not only effective but also equitable and ethically sound.

4.4.1.2 Data quality and integrity concerns

The importance of ensuring data quality and integrity emerged as a critical theme among participants, particularly in the context of AI-driven decision-making. The reliability of AI systems hinges on the accuracy, relevance, and comprehensiveness of the data they are trained on and utilise. Poor-quality or outdated data not only compromises model outputs but also introduces significant ethical risks, such as unfair treatment of clients and biased decisions. The discussion below explores participants' perspectives on data quality, the challenges posed by outdated and inaccurate data, and the complexities introduced by the increasing volume and velocity of data in the digital era.

As alluded to by P3, high-quality data is critical for the reliability and effectiveness of AI-driven decision-making. Inaccurate or poor-quality data compromises model outputs, leading to incorrect decisions that can impact customers, operations and compliance.

> *"I think data is a big thing for any organisation, and that's becoming the fuel for any organisation now. You know, the more data you have the better. But data that's accurate and quality data is even better because that's driving decision-making." (P3)*

In a similar vein, Breidbach (2024) argues that if the quality of data used by algorithms is poor, this can lead to inaccurate or biased decision-making. Furthermore, P1 emphasised the importance of insurance companies utilising recent and relevant data to train the models. The use of outdated data can potentially introduce ethical risks as decisions are based on irrelevant or incorrect information, which can result in the unfair treatment of clients. For example, underwriting or claims decisions may not consider recent changes in a customer's circumstances, leading to inaccurate decision-making.

> *"The other part of it is that if insurance companies don't keep up with and update the data and update the rules and update all these things. It could also be using historical issues, and then in using historical data and historical rules, it's now outdated and is stale, and so it's not a true reflection of where we are today." (P1)*

P4 introduced big data characteristics, namely, volume and velocity, as important impediments to data integrity. Due to the vast amount of data that is now being collected and the speed at which both structured and unstructured data are collected, it becomes increasingly difficult to manage and ensure that such data is accurate and of the right quality.

> *"Volume of data. So, the volume and scale of data, I mean, we have millions of policyholders. That's a lot of data. That's a lot of transactional data and the data that we collect about individuals keep increasing. So, I think just the ability to always understand what data is flowing is becoming increasingly difficult because it's not just structured." (P4)*

Equally, Cai and Zhu (2015) highlight the challenges in managing data, emphasising volume, velocity, and quality assessment. The vast scale complicates validation and integration, while rapid data generation necessitates real-time processing, often exceeding system capabilities. Diverse sources and inconsistent formats hinder high-quality information extraction.

These insights from participants emphasise the paramount importance of data quality and integrity in AI-driven decision-making. Accurate, relevant, and up-to-date data are foundational to ethical and effective AI systems, while poor-quality or outdated data introduces significant ethical risks, such as unfair treatment and biased outcomes. Additionally, the growing volume and velocity of data in the digital era present unique challenges in maintaining data integrity. Organisations must adopt effective strategies to address these complexities, including real-time processing and comprehensive quality assessments, to ensure the reliability and fairness of AI systems.

4.4.1.3 AI governance and data governance concerns

Data privacy and security are fundamental ethical concerns in the use of AI within the insurance industry. As AI systems depend heavily on large datasets to function effectively, the need to safeguard client data and comply with data protection regulations becomes critical. Participants highlighted the risks of non-compliance, data breaches, and inadequate anonymisation, all of which can lead to privacy violations and erode trust. Moreover, the absence of formalised AI regulations, particularly in contexts such as South Africa, further complicates the ethical management of data. This theme (AI governance and data governance concerns) emerging from the data analysis, explores the ethical and operational challenges associated with data privacy and security in AI.

P7 highlighted the necessity of safeguarding client privacy as a basic right, reinforcing the ethical responsibility of organisations:

> *"If you are using models to collect or to process client data, the sensitivity around the privacy of client data is always important, and it's the right of a client to have such privacy." (P7)*

In light of this, P8 questioned how organisations can ensure AI models adhere to data protection frameworks such as the POPIA. Non-compliance with data protection laws could result in hefty legal penalties, loss of trust from clients, and reputational damage to the organisation.

> *"So, the question is, what guarantees that the information is properly protected, alright? I mean, we use POPIA over here in South Africa. How can we ensure that the models comply with those regulations? Alright? The NCA [National Credit Act], the POPIA act and the likes, alright, so, how, how, what? What is the degree or percentage of compliance?" (P8)*

Liu et al. (2020) note the sensitive nature of the data that can be utilised by ML algorithms to make decisions, raising concerns regarding possible data breaches. Therefore, insufficient data security measures are a critical concern in the use of AI in the insurance industry. P2 echoed Liu et al. (2020), expressing that insurers need to prioritise the protection of sensitive customer information:

> *"Well, there's always an issue with data security. You know, whether the client's data is safe, I think that's one of the main issues that you want to make sure that nothing exposes your client information." (P2)*

P8 also emphasised the importance of ensuring data security, pointing out that as the adoption of AI increases, so too will the data security risks. The participant added that inadequate anonymisation of client data may allow for the re-identification of individuals, leading to unintended privacy violations. Liu et al. (2020) also highlighted the risk of reidentification should there be a data breach from an external attack on client information. Similarly, using data without explicit and informed consent from the data owners violates ethical obligations to respect client autonomy and trust. In this regard, P8 stated:

> *"I think security as well, alright, yeah, I think security. Security is something that also needs to be looked into because that relates a little bit with what I mentioned just now. So, data protection, but specifically about reducing privacy risk. Alright, the data that we use, how best are they anonymised to ensure that you cannot trace them? I mean, when we started this conversation, we mentioned the consent form. I mean, my name and whatnot will not be explicitly made available. So, how best can we ensure that information is now anonymised? Because, as things [AI adoption], you know grow, it is something that needs to be solved sooner rather than later, right?" (P8)*

The ethical risks associated with data security and privacy in AI decision-making are overarching, encompassing privacy violations, regulation non-compliance, and security breaches.

However, De Almeida et al. (2021) highlight that the absence of appropriate AI regulations in some countries is a major concern. The lack of established ethical standards for AI creates uncertainty and increases the risk of unethical practices, intentionally or unintentionally, as organisations may not have clear principles or frameworks to guide the ethical use of AI technologies.P8 emphasised the absence of formalised ethical standards or guidelines in many South African organisations regarding AI may be due to the lack of AI regulations in the country that these organisations can follow:

"As I said, there are more questions than answers at this time because, for many organisations, it's still like a niche, alright, but this will have to be solved. For example, the ethical standards, there is nothing in place." (P8)

The participant's insights into the ethical risks associated with data security and privacy reveal critical ethical and operational challenges in AI-driven decision-making. From an ethical standpoint, ensuring privacy and security reflects a commitment to fairness, accountability, and respect for clients' rights. From a practical perspective, compliance with regulations like the POPIA, coupled with proactive risk mitigation strategies such as data anonymisation and secure handling of client data, is necessary to address these concerns. These measures ultimately strengthen client trust and align AI practices with both legal and ethical standards.

4.4.1.4 Lack of explainable AI

The lack of transparency in AI decision-making, often referred to as the black box effect, is a significant challenge, particularly in industries such as insurance, where accountability and trust are of paramount importance. The inability to understand or explain how AI models arrive at specific decisions reduces clients' confidence in these models, complicates compliance with regulatory requirements, and raises ethical concerns, such as the risk of discrimination. Participants' insights on the implications of the black box effect and its impact on trust, accountability, and ethical decision-making in the insurance sector are discussed below.

P1 pointed out that autonomous learning in AI models can result in decisions that are not transparent, leading to the black box effect. Similarly, Akinrinola et al. (2024) identify the black box problem in AI models, which are complex and ambiguous systems, making it difficult for stakeholders to understand the decision-making process.

The inability to understand how AI systems arrive at a specific outcome makes it difficult to justify and validate its outcomes. This obviously can erode trust among stakeholders and more importantly clients, particularly in an industry such as insurance where transparency in claims or pricing is expected. P1 continued:

"As AI goes, we will train models, but there it is an autonomous learning model, and therefore, we will not know exactly what the model will come out with and how

it will learn. And that leads to what they call the black box effect, where sometimes the decision will not be transparent. We won't be able to understand how that tool or that system got to that answer." (P1)

Moreover, P9 highlighted that the inability to explain the decisions made by an AI solution can lead to ethical risks, including discrimination. When the outcomes cannot be clarified, complying with regulations that require accountability and transparency in the decision-making process becomes challenging. Thus, addressing the black box problem is imperative not only for fostering trust and ensuring fairness but also for aligning AI practices with ethical and regulatory standards in the sector.

"So if you've got a black box model, you simply can't explain why the model rates one life worse than another, you know, etc. So, you have to be able to explain what the drivers of that is." (P9)

4.4.1.5 Falsified information produced by AI

AI models are inherently designed to generate responses even when they lack certainty, which presents considerable risks to decision-making processes. This tendency to produce answers without reliable fact-checking raises serious concerns about the reliability and trustworthiness of AI, especially in situations where accuracy or precision is of utmost importance. This phenomenon is commonly referred to as "hallucination", that is, AI's propensity to fabricate information rather than acknowledge uncertainty. Roychowdhury et al. (2023) demonstrated that hallucination happens when LLMs provide false information as facts, primarily due to biases in the training data. P4 highlighted this issue of AI "manufacturing" answers even if it lacks the data or information to provide an answer, raising concerns about reliability and trustworthiness in contexts of AI, especially in cases where accuracy is critical, such as in the insurance industry:

"… but the precarity of AI to want to give an answer, even when it doesn't have an answer, it will manufacture an answer. So, hallucinations are real." (P4)

P7 pointed out that the lack of adequate validation of the training of the AI model in terms of the information it produces can amplify the hallucination problem:

"The one where I spoke about integrity of data, where the model is trained in such a way that there's no adequate validation of information, then it produces non-valid

information because it goes into a process of hallucination, and it falsifies information." (P7)

In conclusion, the issue of AI hallucinations presents a critical ethical risk in decision-making processes. Ensuring the validation of AI models' outputs and designing systems with safeguards against hallucinations are essential steps for mitigating this risk. Ethical considerations demand that organisations prioritise accuracy, transparency, and accountability in the development and deployment of AI systems to prevent harm and maintain trust.

4.4.1.6 Lack of efficiency and accuracy

While AI offers significant benefits in enhancing decision-making efficiency, the risk of overreliance on AI presents critical ethical and operational challenges, particularly in the insurance industry. Overreliance on AI can diminish human oversight, leading to accepting non-verified AI outputs, even when they are inaccurate or biased. This can result in suboptimal decisions, reputational harm, and a loss of trust among stakeholders, including employees and clients. The participants' perspectives on the risks of overreliance on AI emphasise the importance of maintaining a balance between AI capabilities and human judgment.

P4 argued that if AI systems fail to deliver accurate results consistently, this can lead to suboptimal or even harmful decisions. In the insurance industry, this might translate to incorrect pricing, incorrect claim assessments, or incorrect underwriting decisions.

> *"So I still perceive the greatest challenge for AI, particularly in some of the instances I've encountered, is the consistency of precise answers; however, they are unlikely to be identical every time." (P4)*

P1 highlighted the issue of overreliance on AI from an employee point of view, as accepting AI results before verifying them can pose a risk, even if the model has high accuracy. Buçinca et al. (2021) argue that combining humans and AI in sociotechnical systems yields better results than using either alone. However, over-reliance on AI, where users trust and follow its incorrect recommendations, can result in poorer decisions than if AI were absent.

> *"I think also, what I'm starting to see is like that reliance on AI. So, like I just said to you, now we know that the model is 95% correct, so it's about making sure that our investigator still does that job properly and makes that decision at the end. And he's not like I know that the model's 95% correct, so I'm not even going to check this, just I am going to let it go through." (P1)*

Additionally, P3 looked at the issue from the point of view of the client, highlighting the risk of clients excessively relying on AI-generated information, assuming it to be completely accurate and relevant. This overreliance risks eroding the trust of clients who make use of AI if it provides inconsistent or unfair outcomes, raising concerns about transparency and fairness.

> *"I think over time, as an end user, as a client or someone that is taking out insurance, you know, the over-reliance on AI as well. And I think that might also be a risk for them because you're going to over-rely on the fact that, listen, I would prefer this machine to give me the information. And that might not be the most accurate thing initially, and the most relevant thing for that individual." (P3)*

While AI can potentially enhance efficiency and decision-making in the insurance industry, overreliance on AI poses significant ethical risks. Lack of or reduced human oversight, blind trust in AI outputs, and client dependency on AI-generated information can lead to inaccuracies, biases, and eroded trust. To address these challenges, organisations must implement reliable human supervision to validate AI outputs, ensure transparency to manage client expectations, and continuously refine AI systems to deliver consistent and accurate results. By balancing AI capabilities with human judgment, organisations can safeguard trust, accountability, and fairness in decision-making.

4.4.2 Research question 2: How are these ethical risks currently addressed in the insurance industry?

This research question aimed to determine how the insurance industry currently addresses ethical risks associated with AI, assess the effectiveness of existing measures, and identify potential gaps in ethical practices.

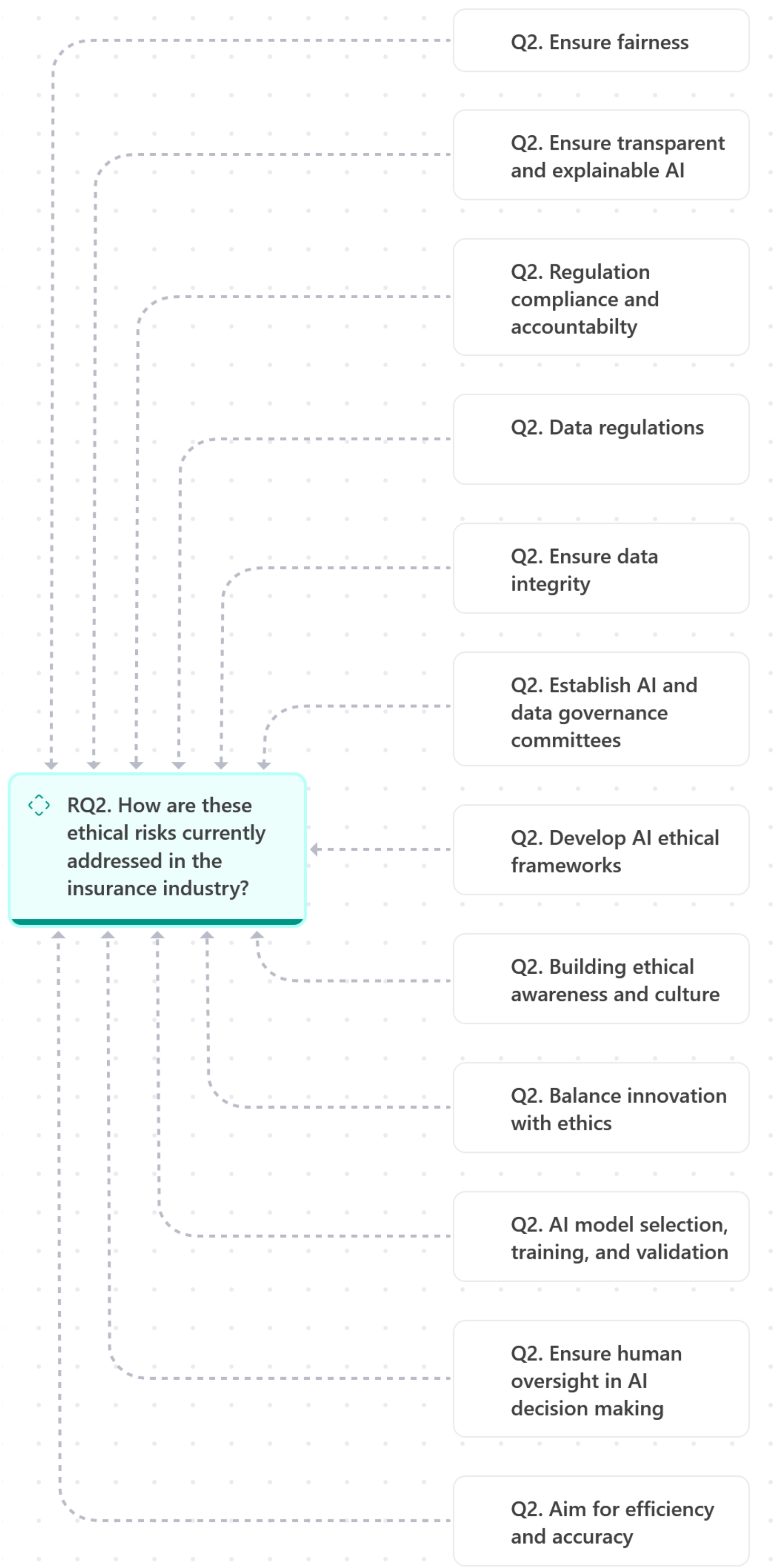


Figure 4.2: Emerging themes for research question 2 (RQ2)

4.4.2.1 Fairness, transparency and explainable AI

The ethical use of AI in the insurance industry requires alignment with organisational values, adherence to principles such as fairness, accountability, and transparency (FAT), and a strong emphasis on explainability. These principles ensure that AI-driven decisions are responsible, equitable, and comprehensible to both employees and clients. Participants emphasised the importance of trust, transparency, and ethical decision-making as the foundations for the successful integration of AI technologies in client-focused processes.

P7 stated that AI models should represent the values of the organisation, ensuring alignment with ethical standards and customer expectations. This builds trust with clients and drives the responsible use of AI.

> *"I think the biggest thing is, from an ethical point of view, one you want to make sure that the models that you have represent who you are as a company, that it speaks to your values as a company, in terms of what you do, how you do it, and what you want to express as an organisation to your customers." (P7)*

P1 highlighted the significance of fair, accountable, and transparent AI models (referring to them as FAT models). Adhering to these principles, the FAT framework ensures that customers are treated equitably, the organisation is held responsible and accountable for its AI decisions, and the decision-making process is clear to all clients, irrespective of whether they are positively or negatively affected by those decisions.

> *"The reason that we do that is that we really are looking at what we call FAT models, where we try and make sure that there's fairness, accountability and transparency within those solutions." (P1)*

Similarly, Aysolmaz et al. (2023) argue that transparency is essential for customers to assess the fairness, accountability, and privacy of ADM systems. These perceptions influence the perceived usefulness of the systems in enhancing customer-centric processes, affecting their willingness to adopt them.

Expanding on the point of transparency, P5 opined that it enhances customers' awareness of their interactions with AI technology in client-centric processes, which, in turn, helps manage their expectations and builds trust with clients:

> *"So by transparency, we want to make sure that our customers are aware of the tools that we're using so that it manages expectations in relation to the technology that they interact with." (P5)*

Some participants stressed the importance of explainability of the decisions and outcomes made by AI solutions. P4 argued that explainability is particularly significant in contexts like health insurance, where decision outcomes can have profound impacts on individuals:

> *"So, the ability to explain why a decision was made, why a claim was paid or not paid, and how you arrived at that answer is very important in health insurance specifically, but other forms of insurance as well." (P4)*

P9 echoed P4's sentiment regarding the importance of explainability within the insurance industry and added that the risk of having a black box model is that it makes decisions that cannot be explained:

> *"Explainability is, is quite important in the insurance industry, because we must be able to justify the premiums that we're charging. So, if you've got a black box model, you simply can't explain why the model rates one life worse than another, you know, etc. So, you must be able to explain," (P9)*

Similarly, Owens et al. (2022) highlight the importance of explainability in AI systems, particularly in sensitive contexts like insurance, where decisions profoundly impact individuals. They emphasise that the opacity of "black-box models" undermines fairness and accountability, reducing stakeholder trust.

By aligning AI models with organisational values, adhering to the FAT framework, and prioritising explainability, insurance companies can build trust and manage customer expectations effectively. Explainability ensures that critical decisions, such as claims assessments and premium calculations, are justifiable and transparent, reducing risks associated with black box models. These principles provide a roadmap for RAI adoption, fostering fairness, accountability, and trust in a heavily regulated industry.

4.4.2.2 Regulation compliance, accountability and data regulations

Regulation compliance

The absence of AI-specific regulations in South Africa presents a significant challenge for organisations navigating the ethical and operational risks of AI adoption. Participants emphasised the importance of involving legal experts, adhering to governance frameworks, ensuring compliance with data protection laws, and promoting fairness and accountability. These measures collectively provide a foundation for ethical AI governance and data security in an evolving regulatory landscape.

In this context, P5 recommended involving legal and compliance experts to provide guidelines that can mitigate ethical risks. As highlighted by De Almeida et al. (2021), the absence of appropriate regulations in some countries and regions constitutes a risk, due to the lack of a regulatory body that can ensure that AI advancements align with ethical guidelines. Therefore, to mitigate this risk, P5 suggested that these experts can ensure alignment with current legal frameworks and help establish governance for AI technologies:

> *"We still need someone to take a look, preferably an expert in legal matters, to determine if this aligns with our understanding of the regulatory environment. Are we including all the necessary elements when we communicate with our stakeholders and customers?" (P5)*

As mentioned earlier, South Africa currently does not have AI regulations that govern the use of AI within its borders. However, discussions are currently underway regarding specific regulations for AI, and it is essential that organisations contribute to the development of these regulations. In this regard, P5 stressed the importance of actively engaging in regulatory development to ensure that AI-related regulations are practical, enforceable, and aligned with industry needs. Noordhoek (2023) supports P5's assertion that organisations must actively contribute to the development of AI regulations to ensure that they are practical and aligned with industry needs. The author (Noordhoek, (2023) goes on to emphasise that regulators should engage with insurers to create balanced frameworks that protect consumers while fostering innovation.

"I know the EU has one [regulation], and there are some that have been developed, so almost contributing and making sure we are a voice in that since we are already using it [AI], so actively getting involved in what regulation should look like around AI in general in the country." (P5)

In addition to contributing to regulatory development, P1 highlighted the importance of adopting global regulations and best practices as a stopgap measure. This approach allows organisations to leverage internationally recognised standards to mitigate risks such as bias, discrimination, and privacy violations. Similar to P1's view on leveraging global standards, Noordhoek (2023) advocates for the use of existing technology-neutral regulations as a stopgap measure, arguing that many AI-related risks in insurance, such as bias and data governance, are already addressed by sector-specific regulations, such as the GDPR. However, P1 cautioned that although global standards may provide guidelines, the global standards often require adaptation to address the unique socioeconomic and legal context in South Africa:

"With there not being any real regulations in South Africa, we do try and follow, I don't want to say industry best practice more like global best practices and understand the global best practices and try and incorporate that as best as possible into our organisation." (P1)

Additionally, P6 noted that other existing governance frameworks, such as the King IV Code, encourage a holistic approach to compliance in the financial services industry that not only looks at financial compliance but also ethical and societal dimensions as well. Applying these principles to AI governance helps insurers align their AI practices with broader organisational values and stakeholder expectations.

"So all, all JSC [Johannesburg Stock Exchange] listed companies have to adhere to King Four Code, which is like integrating reporting that doesn't just talk about financial performance, it speaks about financial performance, good governance and all the seven areas of capital, which speaks about financial, people, etc." (P6)

Finally, P9 raised the important point that regulatory oversight within the financial services industry is rigorous, thereby enforcing fairness and accountability in insurers' practices, which helps reduce risks such as discrimination and bias. However, AI-specific regulations will still be required to ensure that fairness and accountability are enforced in AI-integrated processes.

"I think the regulators are quite active in ensuring that insurers are treating clients fairly and that the actuaries, for example, are bound by professional standards in terms of how it's used for underwriting." (P9)

Data regulations

Insurers can mitigate ethical risks by leveraging established legal standards, contributing to regulatory development, adopting global best practices, and adhering to existing data regulations such as the POPIA and the GDPR.

From a data regulation point of view, P8 believed that regulatory frameworks like the POPIA and the GDPR provide critical guidelines for organisations to protect personal information, prevent misuse of personal information, and maintain customer trust. Likewise, Sharma and Sharma (2024) point out that the POPIA governs the processing of personal information in South Africa, while the GDPR provides a framework for data privacy in Europe. Both ensure that the processing of personal data is lawful, fair, and transparent and that the rights of the data subject are protected (Van Den Boom, 2021). The guidelines also apply to AI as personal information can be used in AI models. This ensures legal and ethical accountability of insurers, where sensitive client data is heavily used in several of their processes.

"So, the question is, what guarantees that the information is properly protected, alright? I mean, we use POPIA over here in South Africa. How can we ensure that the models comply with those regulations? Alright? The NCA, the POPIA Act and the likes, alright, so, how, how, what? What is the degree or percentage of compliance?" (P8)

P4 noted that regulation compliance is not new and organisations are accustomed to being compliant. However, what is new is the emerging complexity of the ethical use of data for AI, which introduces new challenges that require new regulations or extensions of the current regulatory frameworks.

"This is not new. We've been having POPIA, GDPR, and data security, we get audited on an annual basis. So, this is not new, it's just that the ethical use of data for AI is an additional layer to those regulations." (P4)

Lastly, according to P1, insurers need to be accountable for the privacy of the clients that they serve by adhering to regulatory requirements, as this is their ethical obligation:

> *"So we need to make sure that we are accountable for how we use our customer's data and the fact that we keep it private and that we make sure that we are compliant with all the regulations." (P1)*

Although South Africa lacks AI-specific regulations, organisations can adopt a proactive and holistic approach by combining global best practices, existing governance frameworks, data regulations, and active engagement in shaping future AI policies. Leveraging principles like fairness, accountability, societal responsibility, and data privacy ensures that AI aligns with organisational values and stakeholder expectations. Compliance with data regulations such as the POPIA and the GDPR reinforces ethical practices, while rigorous oversight in the financial sector helps insurers build trust, mitigate ethical risks, and prepare for the eventual introduction of AI-specific regulations.

4.4.2.3 Ensure data integrity

The quality and integrity of the data are crucial for ensuring the ethical and effective use of AI in the insurance industry. High-quality and well-managed data can reduce risks such as bias, discrimination, inaccuracies, and unintended consequences in decision-making. According to the findings, data integrity elements such as data quality and representative data are key to ensuring the ethical use of AI.

Data quality is key to ensuring that AI models are trained on accurate and relevant information, reducing the above-mentioned risks. P4 emphasised that data quality can be improved by enhancing the labelling, cataloguing, and understanding of the data while also defining the purpose of the data. Deekshith (2021) noted that high-quality data is crucial for effective ML models and for mitigating risks from biases and inaccuracies.

> *"So, what we're doing is trying to improve how well our data is labelled and catalogued and understood, and also that we know exactly what data is allowed to be used for what purposes. So, we just strengthened our general data governance, which will help us to select the right data when we do train algorithms." (P4)*

P3 underscored the importance of minimising the amount of data collected by AI models, which can aid in reducing the risks associated with data misuse, such as breaches and privacy violations. Shanmugam et al. (2022) also highlight the importance of limiting data collection in AI models to mitigate risks such as privacy violations and data misuse. This aligns with the principles of regulations such as the POPIA and the GDPR.

> *"I think things like, like data minimisation is where, you know, AI, because it's feeding off, you know, I guess models, you know, LLMs, large language models, action models. It's basically feeding off information, historical information, that's being fed into the ether, if you want to call it that, and it's learning, and it's evolving from that. So, I think, from an insurance perspective, if you're going to use the data, you know, we must be well aware that we need to minimise the data to a point that we only use what we need." (P3)*

Furthermore, P1 underlined the need for safeguarding data, which involves controlling where and how it is stored and processed, as well as preventing unauthorised reuse for training other models, especially in countries with weak data privacy regulations. Likewise, Shanmugam et al. (2022) point out the importance of monitoring and restricting the reuse of collected data to prevent unintended or unauthorised applications. This aligns with P1's concern about not misusing data for training other models:

> *"So if we using a tool or technology, even on a hyperscalers platform like your Microsoft or Amazon Web Services (AWS) we still need to ask the question of, what's going to happen to that data. Is that data that we're using going to be used for other training models? Where is that data going to be prepared? Where is it going to be stored? Where is it going to be processed as well, and because of the POPIA and the GDPR, are we okay with anything that goes into any of the European areas to be processed and come back, as long as it's not being used to train other models?" (P1)*

As mentioned earlier, bias is a major concern according to the participants and one way to combat this is to ensure the dataset is representative of the target population. Likewise, Deekshith (2021) notes the importance of diverse and representative datasets in mitigating bias and ensuring fairness. This reduces the risk of bias in AI outputs and upholds fairness and equity, aligning with ethical frameworks such as deontological ethics (respect for all individuals) and utilitarian ethics (maximising fairness for the greater good). P8 remarked:

> *"Also, for the biases, I think that will have to take, deliberate steps to ensure that the data used is from a diverse population. So don't just use, you know, a certain age range. Use as many diverse data as possible, not just from certain races. Take it from all over the place, and not just for people with certain economic status" (P8)*

In addition, P8 pointed out that organisations can make use of their proprietary data to ensure better control over data quality, which directly impacts the accuracy and reliability of AI outputs. This practice enhances data integrity by reducing dependence on external, potentially less reliable data sources, thereby mitigating risks.

> *"Luckily, many organisations are now using their own data to train models, which I believe will greatly help in achieving accuracy and reducing potential biases as well." (P8)*

Ensuring the quality, integrity, and ethical use of data should be at the forefront of mitigating risks and maximising the benefits of AI in the insurance industry. By improving data governance, adopting data minimisation practices, safeguarding storage and processing, and using diverse and proprietary datasets, organisations can enhance accuracy, reduce bias, and promote fairness. These actions not only align with regulatory and ethical standards but also reinforce trust in AI systems, enabling organisations to leverage AI responsibly and effectively.

4.4.2.4 Establish AI and data governance committees

Governance committees play a pivotal role in ensuring the ethical use of AI within organisations, particularly in the insurance industry. Participants highlighted various aspects of these committees, from championing discussions on ethical implications to ensuring diverse stakeholder representation and a client-centric approach. These committees serve as collaborative platforms for addressing ethical risks, aligning AI strategies with organisational values, and improving services for clients and staff, ultimately promoting responsible and inclusive AI governance. Hadley et al. (2024) argue that governance committees, such as ARBs, have emerged as significant mechanisms to address risks and ensure ethical oversight in AI and data practices. These committees ensure that AI models align with ethical and operational standards.

Some participants stated that their organisation has a governance committee responsible for overseeing AI implementation. This includes the ethical use of AI and mitigating strategies. For instance, P6 pointed out that governance committees foster discussions around the use of AI within the organisation, which includes the ethical implications of AI. These committees provide a collaborative environment for committee members to explore how best to mitigate ethical risk, creating a foundation for RAI development.

> *"I do know that there are governance forums, and one of these is a community of practice focused on AI. One of the topics discussed in this community revolves around ethics in AI." (P6)*

Echoing P6, P9 pointed to the importance of dedicated governance committees in addressing ethical risks associated with the use of AI in the insurance industry:

> *"And then we've got a very dedicated machine learning forum that supports the product management committees, that specifically looks at the implementation of machine learning and AI models, particularly in the product space and that then looks at the explainability of the models, any potential bias, fairness, you know, those types of things." (P9)*

Additionally, P1 emphasised the importance of ensuring that the committee has diverse stakeholders from different backgrounds such as IT, business, risk, and security, and also different demographics, such as race and gender. Likewise, Birkstedt et al. (2023) emphasise that effective governance structures require the participation of diverse stakeholders, including legal experts, risk managers, business leaders, and IT professionals. A multidisciplinary AI governance committee ensures that AI strategies are developed and implemented with input from all relevant areas of the organisation and population groups.

> *"The next part is that we have an AI governance committee, which consists of individuals from across the organisation, not just within our division. It's a company-wide initiative, rather than limited to the short-term insurer division. The commission comprises people from various areas of the business, including risk and compliance, as well as cyber security and security." (P1)*

On the other hand, P8 recommended a client-centric approach to AI governance, emphasising the leveraging of AI to improve the services provided to clients and advisors, while also considering the staff benefits. This is in line with the utilitarian principle of maximising overall benefits.

> *"I've also participated or currently participating in what we call the AI community of practice within the organisation, where we looking at how best to utilise AI within insurance, but specifically within the organisation to be able to provide better service for our clients and advisers alike, and, of course, even for our staff also, but major focus is on our clients and advisers." (P8)*

By fostering ethical discussions, incorporating diverse perspectives, and prioritising client-centric strategies, governance committees provide a structured approach to mitigating ethical risks in AI applications. They ensure AI models align with fairness, transparency, and accountability principles, balancing organisational goals with societal and client needs. These committees are essential for building trust, enhancing accountability, and promoting the responsible development and deployment of AI within the insurance industry.

4.4.2.5 Develop AI ethical frameworks

As several participants expressed, one of the most important outputs that the governance committee must produce and oversee is the ethical framework that will guide the development of AI solutions within the organisation. These frameworks ensure responsible and ethical AI use by aligning AI practices with global standards, organisational values, and stakeholder expectations.

Ethical frameworks can be based on global standards, which can provide organisations with a starting point in framework development. This ensures that the implementation of AI solutions is informed by established standards, fostering responsible and ethical AI use. However, organisations may need to adapt the framework to ensure that it is suitable for their environment and target populations. P5 remarked:

> *"Obviously, business is aware of what is happening in the world and what best practices are emerging in specific use cases that we wish to implement within the organisation." (P5)*

According to P7, the establishment of AI principles that are aligned with organisational values is crucial and should be applied throughout the AI development lifecycle, from initiation to deployment, including those procured from third-party vendors. This alignment with organisational values ensures consistency in ethical decision-making. Fahmideh et al. (2021) also emphasise that AI ethics should be integrated at all stages of the AI lifecycle, including data collection, model training, deployment, and monitoring. These principles provide a foundation for evaluating AI systems, ensuring that they reflect the organisation's commitment to fairness, transparency, and accountability.

> *"We've defined AI principles, and these principles speak to what we deem as things we need to consider throughout a life cycle, or when we talk about AI and those principles are to align to our values, and those things are to be brought in. When you develop a system, a model, you need to think of that. If you're buying a model, you need to think of that if you're engaging with a third party on a model, those principles and those values are to be thought of, and how we drive it is to make sure that everybody understands that those things are there." (P7)*

Although aligning with organisational values is crucial, P5 stressed the importance of developing AI ethical frameworks through collaboration across diverse disciplines such as legal, IT, and cybersecurity teams. Additionally, Ferrell et al. (2024) note the need to align AI ethical frameworks with cultural, industry, organisational, and legal standards. AI development should, therefore, incorporate diverse stakeholder inputs for a comprehensive ethical framework. This approach will help ensure that the framework is comprehensive and addresses various dimensions of ethical risk.

> *"So again, speaking to making sure that legal IT, cyber security, everyone is in the know of what is required of them to develop and contribute to this ethical framework and what that looks like." (P5)*

Lastly, P6 explained that ethical frameworks provide guidelines for stakeholders to consider in the development and use of AI solutions. They provide a comprehensive approach to identifying and mitigating risks and ensure that no critical aspects are overlooked.

> *"So, frameworks are meant to typically help at least frame your thoughts and make sure you holistically thought about everything." (P6)*

The development of AI ethical frameworks, informed by global best practices and adapted to organisational needs, is critical for ensuring RAI use. By aligning these frameworks with organisational values and engaging diverse stakeholders, organisations can address ethical risks comprehensively, ensuring fairness and transparency in AI-driven decision-making. These frameworks not only provide clear guidelines for mitigating risks but also establish a strong foundation for ethical governance, fostering trust and accountability in AI solutions.

4.4.2.6 Building ethical awareness and culture

The rapid advancement of AI in the insurance industry necessitates that organisations prioritise AI literacy and ethical awareness among employees. Educating staff ensures RAI usage, mitigates ethical risks, and fosters trust and accountability. Participants highlighted the importance of equipping employees with a comprehensive understanding of AI's impact, from data privacy risks to ethical considerations throughout the AI lifecycle. Additionally, they stressed the role of organisational culture, led by senior leadership, in embedding ethical principles into AI practices and decision-making.

AI has made significant advancements in the insurance industry recently. Therefore, insurers must ensure that their staff are AI literate, equipping them with the knowledge needed to use AI responsibly and minimise the risks of ethical violations. This education builds a foundation of trust and accountability within the organisation.

To illustrate, P4 pointed to the need to educate employees about AI literacy, focusing on risks like data privacy and the potential public exposure of sensitive information, such as through tools like ChatGPT. Figaredo and Stoyanovich (2023) support this argument, stating that AI literacy is essential for equipping individuals with cognitive tools to recognise and mitigate the risks associated with ADM systems.

> *"Literacy, we have to make sure that all our employees understand that we're in the business of trust. We did do that quite well with POPIA and privacy of personal information, etc. But the reality is, whenever you go on ChatGPT, those prompts and information that you upload are actually now in the public domain. So, I do think we need to go into education drives around AI and the use of AI to be more efficient in the workplace." (P4)*

An aspect highlighted by P3 is to ensure all employees understand AI's impact, regulatory requirements, and implications for business and end users. Therefore, it is essential to bridge the gap between technical and non-technical stakeholders, fostering a collaborative environment for the effective use and implementation of AI.

> *"… and we're going to need individuals that are skilled, that they might not be technically savvy, but understand the concept of AI, understand the impact, understand how regulation works, how it's going to impact the end user, how they don't have to speak ones and zeros, but they should be able to understand a business." (P3)*

P7 provided a critical perspective, stating that like any other technology, AI has a development lifecycle from initiation to deployment; therefore, it is important to ensure employees are aware of ethical risks at all stages of the model's lifecycle. Similarly, Santoni de Sio (2024) highlights the need for educating employees on the ethical implications of AI across its development lifecycle from initiation to deployment. Raising awareness ensures that ethical considerations are integrated into every phase of AI development and eventual use.

> *"Because these risks play out in various stages of a model's life cycle, it can play out in the development in the conceptual phase of a model right up until when it's deployed. So I think it's just making sure that the various people in the various stages of a model are aware of the risk associated around discrimination, fair treatment, etc."(P7)*

On the other hand, P6 connected ethics with organisational culture and postulated that the organisation as a whole must embody ethical principles that can guide AI ethical frameworks to effectively foster a strong AI ethical culture:

> *"Yeah, because I don't know for me, ethics is it like, in a way, it almost feels like when we talk a bit about the culture, right? Do you have an ethical organisation? Like, what's the culture of your organisation?" (P6)*

P9 argued that an organisation lacking an ethical culture driven by leadership is unlikely to implement RAI that ensures the fair treatment of clients. When ethical behaviour is championed by senior leaders, it cascades throughout the organisation.

> *"I think the company culture is really what is important in ensuring that things are done correctly and that the clients are treated fairly, um, and I think that's something that's driven from the top, you know, from the CEO to make sure that you know that the company's culture and approach to clients, etc, is appropriate, is fair, you know, etc." (P9)*

Building ethical awareness and a strong ethical culture are essential strategies for mitigating the ethical risks of AI use in the insurance industry. By prioritising education, integrating ethics across the AI lifecycle, developing AI-aware professionals, and driving ethical behaviour from the top, organisations can ensure RAI use. These efforts not only mitigate risks but also enhance trust, accountability, and long-term sustainability in AI-driven operations.

4.4.2.7 Balance innovation with ethics

In a competitive, for-profit-driven environment such as the insurance industry, organisations face the dual challenge of fostering innovation while maintaining ethical integrity. This dilemma becomes particularly pronounced with the adoption of AI, especially in the absence of AI-specific regulations. Participants highlighted the need to balance innovation with ethics, focusing on governance, customer-centricity, and measurable ethical outcomes to ensure RAI deployment while maintaining competitiveness and client trust.

P6 pointed out the challenges that businesses face when trying to be competitive in a high-pressure environment while also being ethical:

> *"And I think similarly, the challenge of business is balancing trying to be an ethical organisation, but also trying to survive in a very competitive industry." (P6)*

P6 further pointed out that the pressure to remain competitive may tempt organisations to prioritise innovation at the expense of ethics. This creates a dilemma: should an insurer compromise ethics to maintain market share or adhere to ethical principles even if it risks losing competitive ground? On the same note, Baker and Rajab (2024) emphasise that rapid AI development creates tension between technological progress and ethical responsibility. The authors argue that while businesses feel pressure to innovate for competitive advantage, prioritising speed and profit over ethics may lead to unintended social and regulatory consequences. Compromising ethics in favour of innovation may bring short-term rewards; however, navigating this balance is crucial for long-term trust from clients and sustainability.

> *"There is always this balance of striving to remain relevant as a leading insurer while offering innovative products that leverage data to provide clients with the best options. However, it is also crucial to act ethically and not misuse that data, particularly when competitors might wield it to their advantage and gain a competitive edge. If such a scenario unfolds, would you be comfortable losing market share? If that occurs, do you then conform and use that data in a manner that perhaps you shouldn't? I believe this will always be the dilemma, wouldn't you agree?" (P6)*

P9, however, emphasised the need to ensure the balance of innovation and ethics is guided by proper governance without stifling creativity. While Konda et al. (2024) argue that AI adoption provides transformative benefits, including increased efficiency, accuracy, and improved customer experiences, they also point out that these advancements present ethical dilemmas, necessitating a balance between ethical considerations and regulatory compliance for RAI. Effective governance ensures that innovation is pursued within ethical and regulatory boundaries. Therefore, a tailored approach to governance based on the complexity and risks of the AI model is necessary to avoid hindering innovation.

> *"Where you've got a balance between innovation and exploring new solutions to problems, versus ensuring that governance is appropriate, commensurate with the importance, complexity, and challenges of the specific model being implemented. So, you don't want to stifle innovation on one hand, but we want to ensure that it is done responsibly." (P9)*

An interesting view was proffered by P3, namely, that the innovation needs to be customer-centric, balancing efficiency gains and convenience with ethical practices such as privacy and fair treatment:

> *"I think at some point customers are going to be like, listen, as an organisation, as much as you are speeding up processes. You're speeding up turnaround times. You're getting to me a lot quicker. You're answering questions a lot quicker. I'm not waiting for longer. You know, things like that. You know whether it's paying claims or whether it's underwriting a policy, everything is a lot quicker now. But I think when it comes to the impact on the client, they also need to understand that their*

data is safe, the quality is consistent, and that the information is not going to be used for other aspects, you know." (P3)

Lastly, P1 suggested the need for governance committees to include ethically measurable metrics for innovation benefits when evaluating AI initiatives. This could help ensure that ethical outcomes from the proposed innovations align with organisational culture and values. In contrast, Baker and Rajab (2024) advocate for global regulatory alignment to tackle ethical AI challenges. While localised governance is necessary, aligning with global ethical frameworks enhances compliance and best practices.

"And then the second part of that is, how will you measure those benefits in there? And so at that stage of actually going to the Governance Committee, we need to show that we will be able to measure these initiatives. And then part of measuring the initiatives is making sure that we also include some kind of an ethical measurement." (P1)

Balancing innovation with ethics in the insurance industry requires navigating competitive pressures without compromising long-term client trust. A balance between innovation and ethics is necessary, focusing on governance tailored to the complexity of AI models, customer-centric innovation that prioritises ethics, and measurable ethical outcomes to ensure RAI deployment. By embedding these principles into organisational culture, insurers can drive innovation responsibly while maintaining ethical integrity and client trust.

4.4.2.8 AI model selection, training, and validation

The process of selecting, training, and validating AI models plays a critical role in mitigating ethical risks, particularly when guided by an AI ethical framework. Participants emphasised the importance of tailoring AI solutions to organisational needs, balancing in-house development with external procurement, and implementing rigorous validation processes. These strategies ensure alignment with ethical standards, minimise risks such as bias and discrimination, and uphold client trust in the insurance industry.

Insurers generally collect a large amount of client data and use it to make data-driven decisions, which is not new. Therefore, P4 argued that using smaller, contextually trained models makes sense, as it allows insurers to control the data used for training and narrow the focus to specific use cases, reducing the likelihood of bias, irrelevant outputs, or misuse of data.

> *"And so to counter that, what I have seen is a lot of businesses are actually opting for small language models, contextually trained, rather than very big, large language models because they're trying to control the data that the model is trained on, and boundary it more to one use case or two use cases, rather than being trying to be ultra-wide." (P4)*

Moreover, P5 noted that opting for internally bespoke AI solutions, rather than relying on a third-party vendor, provides insurers with greater control over data and functionality, thereby reducing risks associated with vendor reliance, such as data exposure or non-compliance with AI ethical frameworks. It also ensures that the AI system aligns with the organisation's ethical and operational needs. Similarly, Smith and Bochanski (2022) argue that tailored solutions provide better control over data, functionality, and ethical alignment, making them ideal for complex applications. However, they demand substantial resources and ongoing maintenance.

> *"Another precaution that we take, I think I mentioned it earlier, is our decision for most of these technologies that are changing all the time, like Gen AI, is to have a more bespoke build, so not exposing ourselves as an organisation and our data directly with a vendor, unless we are happy that the solution delivers on what we want it to and for other business reasons." (P5)*

However, P9 pointed out that there are cases where it makes more sense for AI solutions to be completely outsourced, especially for universal use cases. The decision to build or buy models depends on the complexity, specificity, and ethical considerations of the application. Smith and Bochanski (2022) also highlight that, while outsourced models may be cost-effective and faster to deploy, they can lack transparency and flexibility, posing risks of bias and ethical misalignment. Therefore, purchasing pre-built models for straightforward use cases can save time and resources, while building bespoke solutions is better suited for sensitive or highly specific use cases to ensure accurate decision-making and ethical alignment.

> *"So if something you know, certain models are appropriate, they are so unique to the company that you have to develop it in-house versus some others, it's just*

appropriate for something like an underwriting model, for instance, that measures blood pressure or something similar. That seems to make sense to invest in something like that and rather than to attempt developing it in-house." (P9)

P8 explained that when opting to outsource the model, a rigorous validation process can minimise the ethical risk. Rigorous validation processes mitigate ethical risks by ensuring that data privacy and usage agreements are strictly enforced. This approach prevents data from being shared or misused by anyone external to the organisation, upholding client trust and compliance with data protection regulations.

"So we went through an over a month process, alright, with risk and legal back and forth, just validating and triple checking that the model that will be built for us, will strictly use only our organisation's data and any other available public data. I will never share our organisation's data into the pool, for example, our organisation's data should never be used for anything other than for our organisation, so the agreement that we'll eventually have to make with the vendor was eventually we have our own dedicated sandbox within the platform and so and the data that we have within the namespace will not leave, will not go anywhere else, will only be used by us." (P8)

AI model selection, training, and validation are critical steps in mitigating ethical risks in the insurance industry. By choosing context-specific models, developing bespoke solutions, balancing build-versus-buy decisions, and enforcing rigorous validation processes, organisations can ensure that AI systems are ethical, secure, and aligned with regulatory and business needs. These practices build trust and reduce risks associated with bias, discrimination and non-compliance.

4.4.2.9 Ensure human oversight in AI decision-making

The role of human oversight in AI-driven processes is a crucial consideration in ensuring ethical and practical outcomes, particularly in sensitive and client-facing operations. Participants emphasised the need to integrate HITL practices, where human judgment is retained for critical decisions while allowing AI to enhance efficiency. According to Akinrinola et al. (2024), HITL strategies ensure that there is a balance between AI automation and human intervention, which enhances accountability by minimising fully autonomous AI systems and ensuring alignment with ethical guidelines. P1 elaborated on the role of human oversight by explaining that AI systems are used to generate outputs, which are subsequently reviewed and validated by a human who then makes the final decision. This approach underscores the importance of leveraging AI for efficiency while retaining human judgment to maintain transparency, explainability, and accountability.

> *“So in that way, what it does is it gives us a list, and then we still keep what we call human-in-the-loop, so a final person to take that and then analyse it and then make a decision on that.” (P1)*

P1 emphasised the importance of not allowing AI to make final decisions in some sensitive processes. Mosqueira-Rey et al. (2023) warn that ML models without continuous human oversight may become static, difficult to evaluate, and suffer performance degradation in dynamic real-world conditions. This supports P1’s argument that AI should not make final decisions in sensitive processes without human validation to ensure adaptability, transparency, and accuracy. This practice ensures that ethical and practical considerations are applied, reducing the likelihood of errors or biases in decision-making.

> *“And then one of the things that we really make sure that we do is we don’t allow at this point for our AI tools or technologies or models to make final decisions.” (P1)*

P5 reinforced the importance of the HITL approach, noting that it shifts employees’ focus toward reviewing AI-generated outputs rather than analysing information. This review process serves as a critical checkpoint to ensure that the final outputs align with organisational values, ethical standards, and customer expectations.

> *“And again, the human-in-the-loop is so important for us, because we have to make sure that before it even goes out, it’s gone through that process. So, the job sort of shifts into a review process for those that work with our customers, and less about pulling the information together unless there are updates and what those look like.” (P5)*

However, Sele and Chugunova (2024) argue that human interventions often reduce decision accuracy rather than improve it. This contradicts the assumption that human involvement inherently enhances decision quality, suggesting that human reviewers might fail to detect or correct major errors in AI outputs.

P4 echoed the importance of maintaining human oversight. However, the participant was of the view that for low-risk processes that have minimal impact on clients, AI should be afforded more autonomy. This perspective illustrates how organisations can balance efficiency and control by restricting AI autonomy to areas with limited ethical implications.

> *"But again, I've got to stress at the moment, it all got human-in-the-loop, okay, except for the low empathy, low risk, not crucial to a claim's outcome, kind of like space, where we are using generative AI or have projects undergo on that." (P4)*

P4's insights collectively underscore the critical role of human oversight in ensuring ethical, transparent, and accountable AI-driven decision-making. Retaining human judgement in sensitive or high-stakes processes helps mitigate errors and biases while preserving customer trust. At the same time, delegating more autonomy to AI in low-risk scenarios enables organisations to strike a balance between efficiency and control. This balanced, measured approach ensures that AI technologies are implemented responsibly, aligning innovation with ethical and organisational values.

4.4.2.10 Aim for efficiency and accuracy

The pursuit of efficiency and accuracy has become a central focus for organisations leveraging AI technologies. Tools such as LLMs and RPA are being increasingly adopted to streamline workflows, modernise legacy systems, and enhance productivity. Participants emphasised the transformative potential of these technologies in optimising resource utilisation, automating routine tasks, and driving operational excellence. As highlighted by Eling et al. (2022), AI-driven process automation enhances operational efficiency, reduces human errors, and accelerates business processes, particularly in claims management, underwriting, and customer service.

P4 highlighted the role of LLMs in enhancing efficiency by automating initial customer interactions. These models, trained on extensive datasets, can be integrated across communication channels to address routine queries before escalating to human agents. This approach not only optimises resource utilisation but also ensures timely and accurate responses in customer service contexts.

"So it's a good use for using large language foundation models that have been trained on additional content, so it can be stitched into any one of your channels and potentially answer some questions before there's a need to refer to a human so it's just really an efficiency play in our space." (P4)

P4 also highlighted the impact of RPA in automating processes that previously required human intervention. By introducing automation, organisations achieve significant efficiency gains while maintaining accuracy in routine operations.

"It's brought a lot of, and now I'm referring to the original one, so deterministic models. It's brought a lot of automation into the process even where we have got human processes, robotics process automation is another source of AI that we are using, and obviously, there's huge efficiencies there as well." (P4)

P6 described how RPA can be used to modernise legacy business processes. These systems are often outdated but still functional and benefit significantly from automation as a cost-effective and efficient solution. This automation is particularly valuable in industries with well-defined processes where consistency is crucial.

"Robotics process automation, and in a legacy business where there's lots of, call it older processes, there's been business process defined many years ago, and it just works in a certain way, and using RPA is a nice and easy way to automate some of those processes." (P6)

While P4 and P6 emphasised the efficiency gains of AI-driven automation, Zarifis et al. (2019) argue that AI implementation in insurance often faces significant organisational, regulatory, and technological challenges. This questions the participants' assertion that RPA and LLMs automatically enhance operational efficiency, suggesting instead that organisational readiness and process adaptability are crucial for successful AI implementation.

Lastly, P8 observed a broader organisational trend toward process optimisation using AI. The focus on speeding up operations and reducing inefficiencies aligns with the overarching theme of aiming for both accuracy and efficiency. This reflects a shift in strategic priorities, where AI becomes a key enabler of operational excellence.

"I think that for me, the main observation that I have had in recent times and in the organisation is that many of us are now looking to speed up their processes and optimise them using the power of AI." (P8)

The participants' insights collectively underscore the critical role of AI technologies such as LLMs and RPA in achieving efficiency within organisational processes. These tools not only automate routine tasks but also enable organisations to modernise legacy systems, optimise workflows, and focus on higher-value activities. The strategic adoption of AI reflects a broader shift toward operational excellence, where efficiency and accuracy are prioritised to drive productivity and consistency. These advancements highlight the transformative potential of AI in reshaping organisational capabilities and delivering sustainable improvements.

4.4.3 Research question 3: How effective are the current strategies used to address ethical risks in the insurance industry?

The purpose of this research question was to evaluate the effectiveness of current strategies used in the insurance industry to address ethical risks associated with AI and to identify strengths, weaknesses, and areas for improvement.

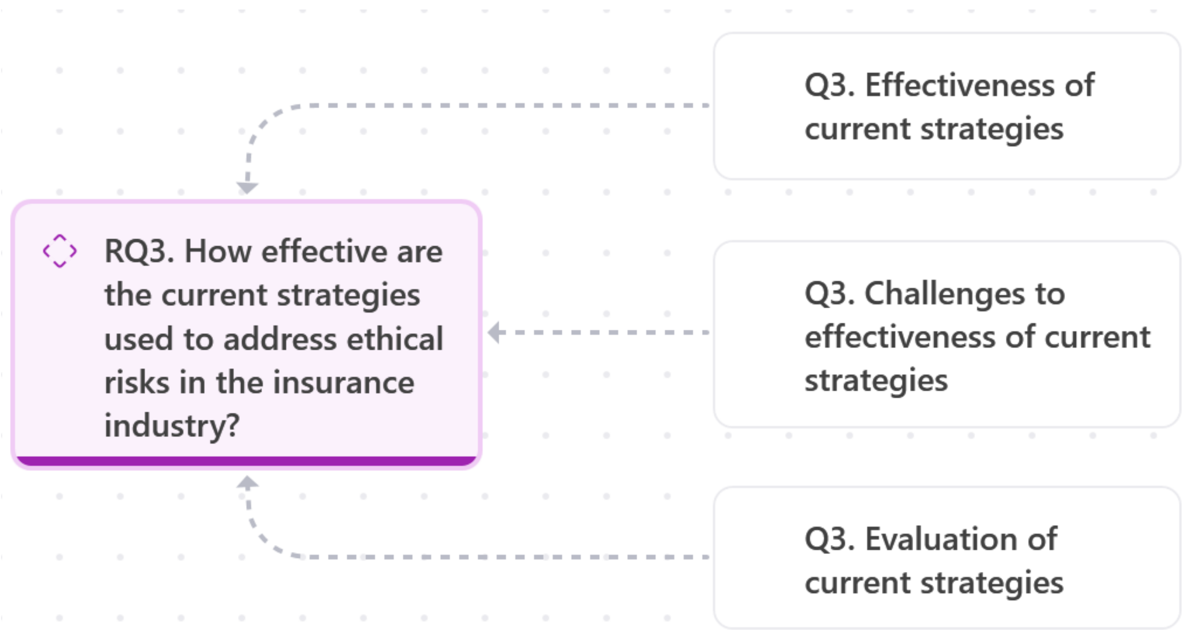


Figure 4.3: Emerging themes for research question 3 (RQ3)

4.4.3.1 Participants understanding of the effectiveness of current strategies

Understanding the effectiveness of current strategies for mitigating ethical risks associated with the use of AI is critical in the insurance industry, where regulatory compliance, governance, and customer trust are paramount. This section explores participants' perceptions regarding the adequacy and impact of these strategies, drawing from their insights to identify key ethical considerations. The participants' responses indicate a general consensus that the strategies employed are at least somewhat effective, with some viewing them as highly effective. Their reflections reveal the importance of adherence, governance structures, continuous improvement, and dual responsibility in ensuring ethical AI practices.

P7 argued that while strategies such as governance frameworks, controls, and policies are fundamental, their effectiveness is dependent upon strict adherence to these strategies. P7 further argued that merely establishing governance structures is insufficient without ensuring compliance at various stages:

> *"It is extremely effective if people adhere to them. Look, things can be put out there. However, if there is no adherence, it raises another issue. So, I think it's good to have these things, all these things that I've spoken about, your governance, your controls, to define these things, to be aware of them, to know them. But at the end of the day, you need to make sure that there's adherence to all of it at various stages." (P7)*

P9 also noted that governance plays a key role in the effectiveness of mitigation strategies. The participant alluded to the regulated nature of the insurance industry, which has governance structures in place to mitigate ethical risks, including those associated with the use of AI. These mechanisms, such as product approvals and management processes, ensure ethical decision-making.

> *"Yeah, so it's the governance of it. Ja, so look, on the one hand, I'm going to tell you that I think it works quite well in terms of identifying the risks and making sure that they are addressed because it's a regulated industry, I think there's fairly strong governance around product approvals, product management etc." (P9)*

However, Stahl et al. (2022) argue that although compliance with regulations such as the GDPR in Europe ensures that insurance companies adhere to strict privacy and data protection standards, it is often seen as a minimum requirement rather than a comprehensive solution. Regulations can sometimes lag behind technological advancements, leaving gaps in addressing new ethical challenges.

P1 described the current strategies as sufficient for maintaining ethical standards while also identifying gaps within them. P1 emphasised the importance of continuously adapting these strategies to keep pace with changes in technology and address existing shortcomings. This perspective reflects a dynamic approach, where strategies evolve to meet emerging challenges.

> *"I think it is sufficient for now to keep us in a good, ethical place, and to assist us in also understanding what more needs to be done, and for us to continuously learn and improve upon that." (P1)*

P4 perceived the current strategies as adequate, citing the absence of ethical infringements or incidents as evidence. However, they acknowledged the potential for increased regulatory demands and stressed the dual responsibility insurers have towards policyholders and shareholders to ensure compliance.

> *"Well, we haven't had an incident, so we haven't had a finding. So I suppose right now it is adequate, but I'm sure that the regulator will raise the bar and force us to comply, but also, as I said, we've got a responsibility towards our policyholders to do the right thing, and we have got a responsibility to our shareholder to not harm the reputation of the company." (P4)*

However, Thierer (2023) contradicts P1 and P4's view that current governance structures are adequate. While P4 argued that the adequacy of governance structures is evident by the absence of regulatory infringements, Thierer (2023) warns against complacency, stressing that governance must be proactive, not reactive.

In summary, participants' perspectives on the effectiveness of current strategies highlight both strengths and areas for improvement. Governance frameworks and regulatory compliance are widely regarded as fundamental to mitigating ethical risks, but their success is contingent on consistent adherence and proactive engagement. However, the dynamic nature of AI necessitates strategies that evolve to address emerging challenges and gaps. Overall, the current strategies appear adequate for addressing existing risks, while ongoing adaptation and alignment with ethical principles remain essential for long-term effectiveness in mitigating ethical risks associated with AI.

4.4.3.2 Evaluation of current strategies utilised to mitigate AI risks within the insurance industry

The evaluation of AI models in terms of their accuracy, fairness, and alignment with business goals is critical in mitigating ethical risks in their deployment. Insights from participants revealed a comprehensive approach to assessing AI effectiveness, focusing on high accuracy thresholds, real-time monitoring, fairness across diverse demographics, and the fulfilment of business objectives. The participants' contributions highlight the importance of technical rigour and ethical responsibility in ensuring that AI systems operate effectively and equitably.

P1 highlighted the importance of high accuracy in model output before deployment. This demonstrates the organisation's commitment to ensuring that AI systems meet stringent benchmarks, including accuracy, before being operationalised. Likewise, Vazquez-Zapien et al. (2022) emphasise that AI models should undergo rigorous validation before being deployed, aligning with P1's assertion that models must meet high accuracy thresholds to ensure reliability.

> *"So like for one of our initiatives, which is our fraud detection one, we trained the model, and the model was coming out on the training data with an accuracy of 75% right, and it sounds like it's really good. And our execs were like, No, that's not really good. We need a 95% accuracy on this thing before we even think about taking it to a deployment level." (P1)*

P4 explained how they use model drift detection to determine if the model requires retraining based on specified tolerance levels. This approach ensures that AI models remain effective, relevant, and accurate over time, addressing the ethical risk of inaccuracies arising from outdated or biased data.

> *"In our machine learning platform, we have automated drift detection and retraining. Okay, so, basically for any machine learning algorithm that we do have in production, we set the tolerances for data drift, and if any of the answers start going outside of those data drift, like limits that we set, then it triggers an automated process to retrain the algorithm on the data that we now have available." (P4)*

P9 reinforced the importance of real-time monitoring of the model for detecting model drift. Similarly, Poenaru-Olaru et al. (2023) pointed to the necessity of continuous model monitoring and adaptation to maintain AI effectiveness and accuracy over time to address model drift. This further emphasises the importance of ensuring that AI models remain effective, relevant, and accurate over time.

> *"… and not only when the models are implemented, but also, over time, to assess models for drift, the approach, the continued appropriateness of the models, etc." (P9)*

P8 highlighted the importance of ensuring that the model is fair by having it tested against different demographics representing the target audience. This approach directly addresses ethical risks such as discrimination and unfair treatment of clients, ensuring that AI models are inclusive and equitable.

> *"Now, one thing that can be done is, if the model is tested against different demographics, for example, we should be able to get a fair feedback assessment, not something that tilts towards a specific economic status or towards a certain age range or towards certain events." (P8)*

Lastly, P7 mentioned the importance of evaluating the AI model against business expectations to ensure business value is met. This perspective highlights a practical, outcome-oriented approach to assessing AI effectiveness.

> *"Remember, you also create these models to ensure that you achieve a specific business outcome. Thus, you can establish business metrics related to return on investment." (P7)*

The participants' insights show the multifaceted nature of evaluating AI models, combining technical accuracy, fairness, and business value to address ethical risks. High accuracy standards, proactive monitoring and retraining mechanisms, demographic fairness, and alignment with business outcomes collectively ensure that AI models remain effective, relevant, and ethically sound. These guidelines provide a framework to tackle the ethical challenges of using AI in ever-changing and diverse settings.

4.4.4 Research question 4: What strategies can be implemented to ensure AI technologies are used ethically in the insurance industry?

This research question aimed to identify actionable strategies for ensuring the ethical use of AI technologies in the insurance industry, addressing identified risks and enhancing ethical decision-making practices.

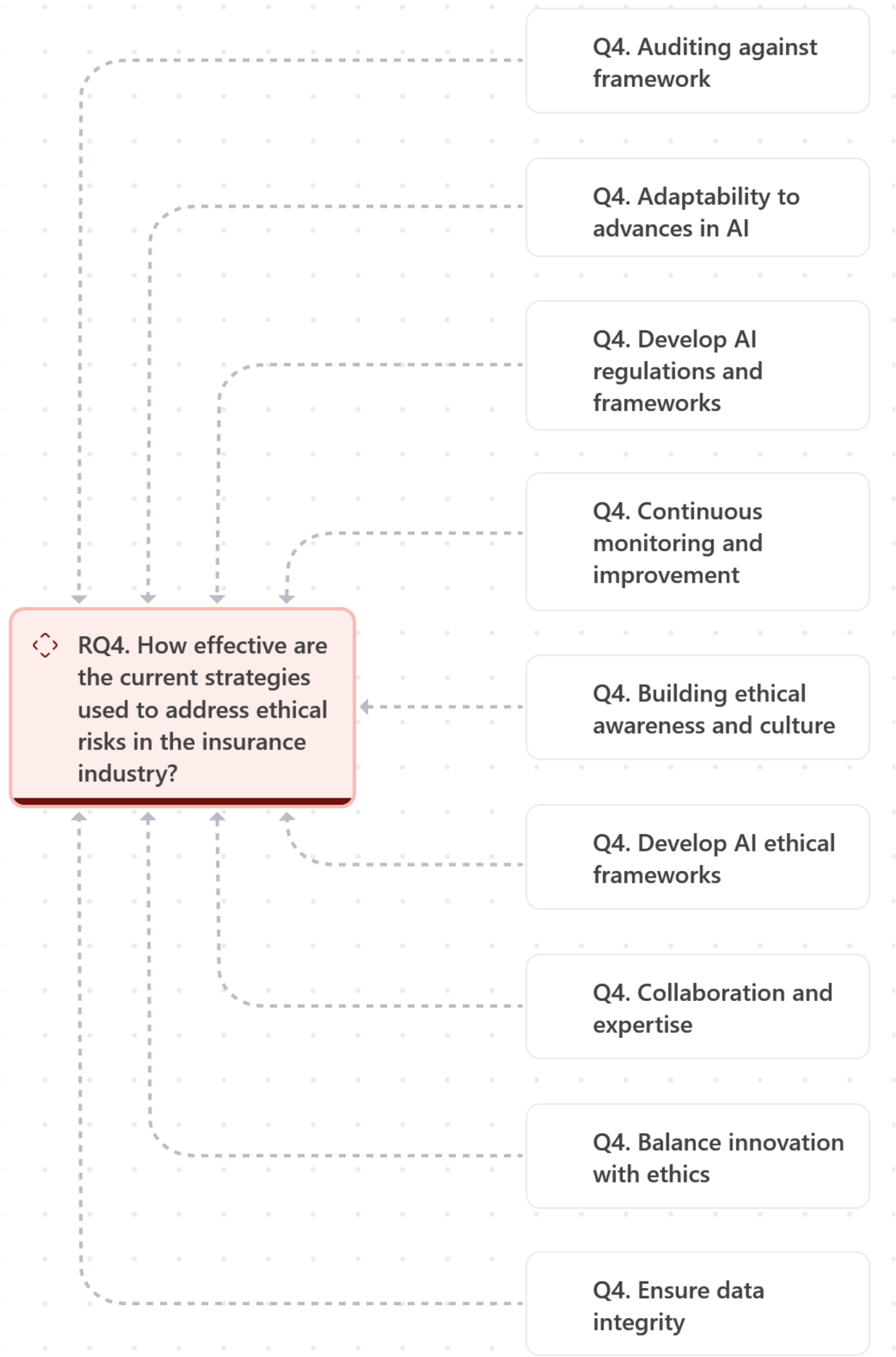


Figure 4.4: Emerging themes for research question 4 (RQ4)

4.4.4.1 Auditing against framework

The rapid advancement of AI has raised significant concerns about the adequacy of existing governance mechanisms to ensure its ethical and responsible use. Participants highlighted the need for robust auditing systems to address these challenges, emphasising the importance of both proactive and reactive measures. The insights illustrate the growing recognition that traditional human governance processes may be insufficient to keep pace with AI's evolution. Automated and scalable auditing mechanisms are proposed as critical tools for maintaining ethical standards and ensuring accountability throughout the AI development lifecycle.

P4 highlighted the possibility that AI could be used to govern itself, pointing to human governance not being sufficient and potentially lagging behind the rapid advancement in AI. This will necessitate automated auditing mechanisms to ensure effective governance.

> *"And Gartner brought out a prediction about two years ago where they said, by 2030, half of AI will be used to govern the other half of AI, which I find, quite an interesting stat [fact], because this is the reality our human governance processes will not be able to keep pace with how quickly AI is going to embed itself in our decision-making processes." (P4)*

Equally, Li and Goel (2024) advocate for automated and continuous auditing processes as they are essential for maintaining ethical standards in AI systems, ensuring scalability and efficiency of the auditing.

P8 noted the absence of an existing structure that would allow for regular audits, suggesting that such mechanisms must be built and automated to ensure scalability and feasibility over time. This aligns with P4's view.

> *"We currently don't have anything in place at this time, so it's something that we have to build as part of what we are doing, I believe, to be able to perform regular and frequent audits. Alright, of course, they have to be done in an automated fashion, otherwise, it becomes almost impossible to maintain." (P8)*

P6 elaborated on the comprehensive lifecycle of AI models, emphasising the critical importance of integrating ethical considerations at every stage of the model's development. This includes not only the initial design and deployment phases but also the necessity of conducting thorough audits after implementation. By embedding ethics throughout the lifecycle, the aim is to ensure RAI, with mechanisms in place to address any potential issues that may arise during their use. However, Bharadhwaj et al. (2021) highlight that current AI models lack interpretable structures, making it difficult to conduct meaningful audits. Therefore, it is important to keep this in mind when conducting audits after implementation.

> *"I suppose we'd have to perhaps it ought to be in the audit as well, wouldn't it? So, we discussed it upfront. There's a body that regulates and formulates policy. We established the practice whereby individuals, through the design and build, remain conscious of the policy. We've created awareness through training, and then perhaps after the fact, there's an audit. So, an audit, similar to an audit practice, that measures whether your AI models are behaving ethically, conducted by an independent body." (P6)*

The collective perspectives of the participants underscore the urgency of establishing comprehensive auditing frameworks for AI governance. The emphasis on automated solutions reflects the need for scalability and efficiency in managing ethical risks while integrating ethics throughout the AI lifecycle ensures responsible development and use. By addressing these challenges, the industry can better align with ethical principles, fostering trust and accountability in the deployment of AI technologies. This forward-looking approach is essential for navigating the complexities of AI's integration into decision-making processes and safeguarding its societal impact.

4.4.4.2 Adaptability to advances in AI

Adaptability to advances in AI, like other 4IR technologies, is increasingly recognised as a critical factor for organisations and individuals to remain competitive and relevant in the rapidly evolving insurance industry. Participants emphasised the need for businesses to embrace change, reconfigure processes, and proactively seize AI-driven opportunities. They also highlighted the importance of fostering a culture of continuous learning and equipping the workforce with the necessary skills to leverage AI effectively. These insights illustrate the multifaceted nature of adaptability, encompassing strategic, technological, and human dimensions.

P7 pointed to the importance of an organisation being adaptable, advocating for openness to change and the ability to reconfigure processes and structures. This underscores a broader strategic need for organisations to embed flexibility into their operations to adapt to AI-driven transformations.

> *"And the only thing honestly, that organisations should be open to is being responsive. They need to be responsive to changes, and they need to be open to adapting and reconfiguring themselves." (P7)*

P6 linked adaptability to competitiveness, viewing AI as an opportunity to level the playing field. This suggests that organisations must be forward-thinking and proactive in identifying and exploiting AI-driven opportunities. Zarifis et al. (2019) attest that AI drives innovation, giving organisations a competitive advantage by optimising processes and creating opportunities. Moreover, a skilled workforce with AI-related skills is essential to sustain this innovation.

> *"I think it's picking up massively, and there will be lots more opportunities around the usage of AI in business. So, and again, you know, I think, as someone working in the business, we always looking for opportunities to use AI to give you that competitive edge." (P6)*

Given the above, P1 emphasised the importance of organisations and employees staying updated with technological advancements and changes. This underscores the necessity of continuous learning and awareness to remain relevant in a rapidly evolving AI landscape.

> *"We need to be updated and keep up to date with what the technology is, how the technology is advancing and how it's changing." (P1)*

P4 contended that while AI will not replace people, those who leverage AI might. Therefore, adaptability requires equipping individuals with the necessary skills to effectively utilise AI in achieving business goals, emphasising the importance of education and training. This reflects the shift from a purely technological focus to empowering the human workforce to work alongside AI.

Eling et al. (2022) emphasise that AI's role in insurance is primarily in process automation and decision-making, significantly reducing reliance on human intervention in many traditional insurance functions. This suggests that, contrary to P4's assertion that AI will not replace individuals but will rather empower skilled employees, AI is designed to replace many manual and decision-making roles in the industry. This contradicts the claim that human adaptability alone will suffice to mitigate job displacement, as AI adoption may significantly reduce workforce demand rather than simply requiring reskilling.

> *"In many industries, including our own, it won't be AI replacing people, it will be people that can leverage AI replacing people who can't. Okay, so I think we've got a responsibility to ensure that our people are trained in the use of AI." (P4)*

The perspectives shared by the participants illustrate that adaptability to AI is not merely a technical challenge but a strategic imperative for organisations and individuals. By fostering openness to change, embracing continuous learning, and empowering employees with AI skills, businesses can navigate AI-driven transformations effectively. This adaptability is essential for maintaining competitiveness, leveraging AI opportunities, and ensuring that both organisations and their workforce can thrive in an AI-enhanced future. Building resilience and flexibility into operational and strategic frameworks will be key to succeeding in the age of AI.

4.4.4.3 Develop AI regulations and frameworks

The development of strong AI governance frameworks and regulations has emerged as a critical need in addressing the ethical challenges posed by AI technologies in the insurance industry. Participants highlighted various approaches to achieving this, emphasising the importance of balancing global standards with local contextual relevance, industry-specific needs, and ethical adoption. Their perspectives shed light on the interconnected nature of AI governance, where a multi-layered approach ensures both global consistency and flexibility to address unique industry and regional challenges. The participants' views focus on the necessity of global standards, local regulations, organisational guidelines, and tailored industry frameworks to establish effective AI governance.

P3 pointed out the need for global standards as a foundational step for AI governance, followed by the development of industry-specific guidelines. This perspective highlights the interconnectedness of industries in the global economy, where consistent standards can reduce regulatory fragmentation and promote cross-border collaboration. Similarly, De Almeida et al. (2021) highlight that the rapid integration of

AI into various industries, including insurance, has necessitated the establishment of standardised frameworks and industry-specific guidelines to ensure ethical, efficient, and transparent operations.

> *"I'm just looking at this thing from a global perspective. So, I think maybe global standards, I think that might be the first step is to ensure that there's a global standard for AI use across the board, across industries, and then from there you can kind of break it down into sub-categories of industry and all that." (P3)*

P1, however, recommended that there should be regulations that govern the use of AI that are country-specific, therefore aligning with local ethical, legal, and social norms. Adapting global frameworks to local industries ensures flexibility while maintaining core ethical principles.

> *"I think we need to actually get proper guidelines and governance for South Africa, for ourselves, so that we have something to follow and that we know we're doing the right thing." (P1)*

P9 advocated for detailed industry standards that address the governance of ML models, including best practices for model classification. This echoes P3's views on having industry standards that are tailored to that industry's needs.

> *"I think it would be very useful to develop industry guidelines on what a governance framework should look like, and what best practices are around the governance of machine learning models as well as the classification of machine learning models." (P9)*

Lastly, P8 emphasised the need for frameworks and guidelines that can guide organisations when integrating AI into their processes. The emphasis on ethical adoption and compliance suggests a proactive approach to mitigate risks such as bias, discrimination, privacy violations, and operational inefficiencies.

> *"… but we need to have frameworks and guidelines that organisations will need to comply with in the usage of AI to improve their processes, to ensure that proper ethics is put into place in the adoption of AI while rendering their insurance services to their various clients. So, that's what I meant by ethical standard so that the governing body for insurance, or let me say, for Financial Services, will have to help come up with something that the organisation can then try to comply with." (P8)*

The participants' insights highlight the necessity of a comprehensive approach to AI governance. While global standards provide a unified foundation, their effectiveness depends on adaptability to local legal, ethical, and social contexts. Industry-specific guidelines further enhance this framework by addressing each sector's unique challenges and requirements. By integrating these perspectives, the insurance industry can create governance systems that ensure RAI usage, cultivate trust, and promote sustainable innovation.

4.4.4.4 Continuous monitoring and improvement

Continuous monitoring and improvement have emerged as a critical theme for ensuring effective and ethical governance of AI technologies within the insurance industry. Participants pointed to the need for organisations to adopt iterative processes that align with evolving regulatory requirements, address ethical risks, and refine operational processes. This approach is particularly crucial given the dynamic nature of AI systems and the challenges associated with their implementation. The insights shared by the participants highlight the importance of adaptability, organisational commitment, and iterative learning in establishing robust governance structures that evolve alongside AI advancements.

P8 noted the need for organisations to constantly review their processes to ensure alignment with regulatory standards and improve overall effectiveness. Similarly, Birkstedt et al. (2023) highlight that AI governance must be treated as an ongoing process capable of adapting to technological advancements and organisational needs, emphasising the importance of feedback loops to assess the effectiveness of governance in refining these structures. This reinforces the notion that AI governance is a continuous cycle of monitoring compliance and refining operational frameworks to address evolving regulatory landscapes.

> *"Well, it's constantly reviewing our process to make them better, and also constantly check our compliance level with what the regulatory body will eventually come up with." (P8)*

P5 pointed out that as AI is forever evolving, there is a need for continuous improvement in how AI is managed, particularly in terms of ethical risk. Singh and Mishra (2018) also point to ethical risks that are constantly evolving. Therefore, AI systems need to be regularly reviewed and updated to address emerging ethical issues and improve their mitigation of client well-being. P5 further stated that the governance committees should frequently meet to discuss findings, processes, and frameworks to be on top of any required improvements, demonstrating an organisational commitment to adaptive AI governance:

> *"So, it will always involve continuous improvement. I believe there will never be a point, at least with the types of technologies we utilise, where we can confidently say we are effectively managing these. I think it is simply an ongoing process, which is why we have the forum that meets monthly to share what is happening and determine if we need to revisit certain processes or frameworks." (P5)*

Additionally, P7 pointed out the importance of integrating continuous improvement into organisational strategies for AI governance. This approach requires insurers to adopt vigorous strategies that are sufficiently flexible to accommodate changes. The focus on maturity and adaptability in implementing governance frameworks reflects a proactive stance toward the ethical challenges posed by emerging technologies.

> *"So I think right now, it's just for people to bed down what strategies they have for AI to put in the right sort of governance in place, and be open to the fact that it is part of the continuous improvement insurance companies or financial institutions would have to do when it comes to emerging technologies, and also be open to the fact that one has to be robust in approach and implementation." (P7)*

P6 highlighted the importance of continuous innovation in terms of AI use cases, which will, in turn, allow the organisation to mature and evolve AI governance. Such iterative learning helps businesses refine their AI processes while ensuring ethical boundaries are respected.

> *"I think as we develop more and more use cases for AI and the application of data in decision-making, we will start to mature. However, if you can at least provide feedback to your governance committee, I believe it should assist your organisation in evolving, hopefully… "I believe that we may see more of this occurring in the future as businesses seek to, in a sense, maximise the benefits of AI technology. As we develop new use cases, I think we will encounter more fallacies and issues that will need to be addressed through a governance committee, ensuring we do not breach or overstep ethical boundaries regarding the use of AI in business." (P6)*

The participants' insights emphasise the importance of continuous monitoring and improvement as a cornerstone of effective AI governance. The dynamic nature of AI technologies necessitates an iterative approach that combines regular process reviews, organisational adaptability, and innovative use-case development. By integrating these practices into their governance structures, organisations can address emerging ethical risks, maintain compliance with evolving regulatory standards, and enhance their operational maturity. This proactive stance not only strengthens trust with stakeholders but also ensures that the ethical boundaries of AI usage are respected, thereby fostering sustainable innovation in the insurance industry.

4.4.4.5 Building ethical awareness and culture

As identified in research question 2, building ethical awareness and fostering a culture of ethical decision-making is critical for organisations leveraging AI technologies. The integration of AI brings about ethical challenges, which require deliberate efforts to educate stakeholders, ensure accountability, and embed shared ethical values. Participants referred to various strategies to build and sustain ethical awareness within organisations, with a strong focus on understanding AI capabilities, prioritising education and training, and employing practical measures to reinforce ethical considerations in daily processes.

P1 emphasised the need to understand AI tools and technologies as an initial step in building ethical awareness. Equally, Migdadi et al. (2024) argue that AI ethical awareness is positively correlated with the intention to use AI. This awareness enables stakeholders to better identify ethical concerns and challenges associated with AI, highlighting that awareness begins with clarity on what the technology can and cannot do.

> *"I think we also need to make sure that we have a better understanding, first of all, of what these tools and technologies can actually do, because if we can understand what they can actually do, we can understand more what the issues, the challenges and the ethical concerns are related to them." (P1)*

P7 noted the importance of ongoing education and training to build and maintain ethical awareness, suggesting that continuous education is critical to addressing risks associated with AI misuse and ensuring that employees are informed about ethical principles. Durant et al. (2022) also emphasise the role of education and awareness in mitigating AI-related ethical concerns, reinforcing P7's view that ongoing education is critical for addressing AI risks and ensuring compliance with ethical principles.

> *"Training and development also encompasses consistently training your various audiences or groups within your organisation about the risks associated with the use and the development or misuse of AI. But what exactly does that entail?" (P7)*

P6 recommended practical measures, such as discussing past mistakes, sharing examples, and producing educational materials, which are effective tools for promoting awareness and fostering ethical behaviour. These measures contribute to a culture where ethical considerations are part of everyday decision-making, thus reinforcing an ethical culture. However, Migdadi et al. (2024) argue that while increased AI ethical awareness correlates with AI adoption, it does not necessarily translate into better ethical decision-making. This challenges the assumption that simply educating employees on AI ethics ensures ethical AI usage, suggesting that compliance mechanisms and accountability structures are also necessary.

> *"I spoke about awareness and how you create awareness. I guess, often through discussions, making materials or sharing examples about the way we've made mistakes in the past, and then, so maybe there's, like, a training element, right?" (P6)*

Lastly, P4 argued that systems have three elements, namely, process, technology, and people; therefore, it is imperative that organisations better equip their employees to integrate ethical considerations into their implementation of AI. This perspective aligns with the view that education and awareness among employees are key drivers of an ethical organisational culture.

> *"So with any solution, there are people, process technology and content, okay? So, what I'm seeing is not a lot of focus on people in it, okay, by and large, in many businesses, it will still be people in your business using AI to service clients better. So, the education of your employees is key." (P4)*

The insights shared by participants emphasise the critical role of understanding, education, and practical engagement in building ethical awareness and culture within organisations using AI. Establishing an ethical foundation begins with understanding

AI capabilities and is sustained through ongoing training, stakeholder involvement, and reflective practices such as learning from past mistakes. These strategies not only enhance individual and organisational accountability but also embed ethical considerations into everyday processes, ensuring a more responsible use of AI technologies. The shared perspectives highlight the importance of proactive efforts to cultivate a robust ethical culture that can effectively navigate the complexities of AI implementation.

4.4.4.6 Developing AI ethical frameworks

The development of AI ethical frameworks emerged as a critical theme across both research question 2 (focusing on current strategies) and research question 4 (exploring future strategies for mitigating ethical risks in AI). This overlap reflects the diverse levels of AI maturity among participants' organisations, ranging from those adapting existing governance structures to those prioritising the creation of foundational frameworks. The participants' perspectives on extending current governance processes, creating new frameworks, addressing high-risk areas, and standardising assessment methodologies, underscore the importance of comprehensive and adaptable approaches to ethical AI governance.

P4 argued that developing AI ethical frameworks is not about reinventing new AI-specific governance strategies but rather about extending existing governance processes to incorporate AI-specific ethics and the management of unstructured data. This perspective aligns with the theme of adaptability to advances in AI and the integration of AI ethics into well-established systems, which can provide a practical and scalable approach to ethical governance.

> *"A lot of these things are not new. So, I'm saying it's not a case of brand-new strategies. It's adding AI ethics and governing unstructured data to your existing governance processes. That's what you need. So, it's extending existing governance processes. It's not brand-new strategies." (P4)*

P7 contended that although governance committees have been established, the ethical frameworks are still in progress. This suggests that some organisations are actively working towards institutionalising ethical practices through formalised governance mechanisms, signalling a commitment to addressing the ethical challenges of AI. Pant et al. (2024) established the need for tailored ethical guidelines for AI in organisations. They stress that these frameworks must align with the specific operational, cultural, and regulatory contexts to ensure effectiveness.

> *"So those are things that are there in terms of maturity when it comes to AI, specifically from a governance perspective as we have moved we've formed an AI governance structure. We're now working on policies. We're working on guidance notes." (P7)*

P8 echoed P7, adding that the absence of foundational frameworks is a significant challenge, and, in contrast to P4, emphasised that the focus should be on creating rather than merely improving existing structures. While Filabi and Duffy (2021) acknowledge the value of extending governance frameworks, they also argue that new AI-specific governance mechanisms are necessary to address the unique risks posed by ML and big data. This challenges P4's assumption that existing frameworks are sufficient and supports P8's assertion that governance structures must be created rather than just improved.

> *"The problem is we don't even have a lot of things in place now. That is actually the challenge. So, so it's more of creation than improving, alright?" (P8)*

Significantly, P9 underlined the importance of governance in high-risk processes, such as underwriting decisions and the use of LLMs, where decisions can significantly impact individuals. This perspective stresses the need for rigorous oversight and governance to mitigate risks and ensure fairness, particularly in scenarios where AI-driven decisions could lead to exclusion or discrimination.

> *"I think the crucial point is where decisions are being made about people. It's like underwriting decisions or people are included or excluded from some processes or opportunities based on underwriting decisions or decisions of large language models. It's, you know, I think those are the very, very high-risk areas where it's very important to ensure there's a lot of governance and I, and I think the use of models will remain quite limited in those cases." (P9)*

Lastly, P9 advocated for the standardisation of frameworks that can be used to assess models. This approach would enhance the consistency, reliability, and transparency of model evaluations, contributing to the responsible use of AI.

"That would be the frameworks that are used to assess the models, by developing and standardising those frameworks." (P9)

The participants' responses reflect diverse perspectives on the development of AI ethical frameworks, shaped by varying levels of organisational maturity and AI adoption. While some participants emphasised extending existing governance structures, others stressed the need to create foundational frameworks to address current gaps. High-risk processes, such as underwriting and LLM use, were identified as critical areas requiring rigorous governance. Moreover, the standardisation of assessment frameworks was highlighted as a key strategy for ensuring transparency and accountability in AI implementation. Together, these insights describe the importance of flexible, proactive, and vigorous governance strategies to navigate the ethical complexities of AI and foster its responsible use in organisations.

4.4.4.7 Collaboration and expertise

Collaboration and expertise are critical for addressing the ethical challenges posed by AI. The participants highlighted the importance of leveraging external expertise, forming strategic partnerships, and fostering collaboration across industries to develop comprehensive ethical frameworks. These approaches enable organisations to access specialised knowledge, manage resource constraints, and establish standards that promote accountability and fairness in AI use. The findings and discussion below explore participants' perspectives on consulting experts, building partnerships, and engaging in industry-wide collaborations.

P1 highlighted how beneficial consulting external experts on best practices for mitigating ethical risk can be. This approach enables organisations to identify and address the specific ethical considerations that can potentially impact their industry, stakeholders, and operations. Organisations can then utilise these expert insights to develop robust ethical frameworks. Equally, Keller (2020) emphasises the importance of leveraging external expertise to develop RAI frameworks, reinforcing P1's assertion that organisations benefit from engaging consultants to understand best practices in AI ethics. This supports the notion that external expertise plays a crucial role in shaping AI governance strategies and mitigating ethical risks.

> *"What you will then need to do is also probably get in consultants to assist you with best practices, or if you can't do that, literally, do the research on your own to understand what are all the considerations or all the ethical implications that are out there, and which of those ethical implications are going to affect you in your organisation, your industry, and your stakeholders, and how are those stakeholders going to be affected by that and react to that?" (P1)*

P4 emphasised the importance of partnerships to access specialised AI solutions that would be challenging and costly to build in-house. By collaborating with external providers, organisations can leverage advanced technologies and expertise, such as pre-trained LLMs, to remain competitive and resource-efficient.

> *"I think we are going to have to get better at evaluating and deciding on partners because we will not be able to build everything in-house. We will have to integrate and leverage solutions developed outside of our space... I mean, training Falcon costs $14 million for a week of training. Now, we're not going to have that kind of money to train large language models ourselves, which means we will partner with others who have done the training." (P4)*

P5 stressed the need for collaboration across the industry to address ethical issues, suggesting that competitive advantage should not come at the expense of ethical considerations. This reflects the importance of shared dialogue and collective decision-making in defining the boundaries of ethical AI use.

> *"So I think there needs to be collaboration across the industry, and though the participants in the industry in developing what that looks like, because some things may be in a competitive advantage, but maybe are not ethical, and conversations need to be had about how far can we take it, which I think is not an insurance-specific issue." (P5)*

Similarly, P6 emphasised the value of co-creating standards with industry bodies and regulators. Collaborative efforts to develop compliance regulations and ethical standards are viewed as beneficial for ensuring consistency and accountability across the industry. While P6 advocated for co-creating AI ethical standards with industry bodies and regulators, Keller (2020) points to significant challenges in standardising AI governance across the industry due to variations in data sources, regulatory requirements, and business models. This suggests that while standardisation is ideal, achieving it across different insurers remains a complex challenge, requiring flexible and adaptable regulatory approaches rather than rigid, one-size-fits-all frameworks.

> *"But I'm pretty sure there's also some industry, industry compliance regulations around AI and AI ethics. And I think if companies and industry bodies that regulate these things can work together to co-create these standards, I think it will be beneficial for the industry as a whole." (P6)*

The participants' responses emphasise the critical role of collaboration and expertise in navigating the ethical challenges of AI implementation. Consulting external experts equips organisations with best practices while forming partnerships with specialised providers helps address resource constraints and foster technological innovation. Industry-wide collaboration, including the co-creation of standards with regulators, ensures consistency, accountability, and fairness in AI governance. Together, these approaches show the importance of shared efforts in fostering responsible and ethical AI practices that balance innovation with stakeholder and societal interests. The interaction between individual organisational efforts and broader industry collaboration highlights a pathway toward robust and ethical AI ecosystems.

4.4.4.8 Balancing innovation with ethics

Balancing innovation with ethics emerged as a key theme in response to both research question 2 and research question 4. The interaction between technological advancement and ethical responsibility is complex, requiring organisations to innovate while maintaining customer trust and adhering to ethical and governance principles. Participants' perspectives focused on navigating this balance through careful integration, regulatory guidance, and mature governance frameworks.

P4 sheds light on the fact that balancing the provision of an innovative AI solution while maintaining customer trust can be challenging, and it is important to ensure that organisations get this balance right. Wang and Wu (2024) highlight the challenge of deploying AI technologies while preserving public trust, reinforcing P4's assertion that organisations must integrate AI responsibly to maintain credibility with customers. This supports the notion that AI deployment must balance accuracy and customer sentiment to ensure that users trust AI-driven decisions and outputs.

"So, sentiment and accuracy are not always the same thing, so we've got to figure out a way to embed this in our business without it eroding trust and with the client being satisfied with the answers that they get." (P4)

P5 advocated for guidance on how AI should be used ethically but warned against imposing constraints that may limit its innovative potential. This perspective emphasises the need for a regulatory balance that empowers organisations to utilise AI effectively while simultaneously safeguarding ethical considerations. P5 highlighted the importance of allowing AI to drive both organisational and customer benefits without being overly restrictive:

"I believe some guidance is necessary, but there should not be any constraints on how and why organisations can use this tool to their advantage or for the benefit of their customers." (P5)

While P5 warned against AI constraints, Wang and Wu (2024) caution that an unregulated AI environment could lead to ethical blind spots and governance failures. This challenges the assumption that minimising AI regulations is always beneficial, suggesting that some level of oversight is necessary to prevent ethical lapses in AI use.

Lastly, P6 argued that the value of regulatory frameworks in supporting organisations to utilise AI responsibly is both effective and ethical. The participant further argued that mature guidelines from AI regulatory bodies can help organisations address their challenges using AI, ensuring that innovation aligns with governance principles and ethical policies. This perspective reinforces the role of external standards in striking a balance between innovation and ethics.

"I believe that having a mature framework from an AI regulation industry body probably aids organisations in using AI effectively. It allows them to address their challenges while ensuring that their approach remains ethical and adheres to good governance and ethical policies." (P6)

The participants' insights highlight the critical interaction between innovation and ethics in AI implementation. Achieving this balance involves addressing challenges such as maintaining customer trust while delivering innovative solutions, providing regulatory guidance that empowers rather than restricts organisations, and adopting mature governance frameworks to ensure ethical alignment. By integrating these approaches, organisations can drive innovation responsibly, fostering trust and aligning technological advancements with ethical principles.

4.4.4.9 Ensuring data integrity

The theme of ensuring data integrity emerged prominently in research question 2, highlighting its critical role in the ethical and responsible use of AI. High-quality, validated, and diversified datasets form the foundation for reducing bias, enhancing fairness, and ensuring trustworthiness in AI systems. Participants emphasised that effective data management practices, including quality assurance, vetting, and transparency, are essential for mitigating risks and ensuring ethical AI implementation. Participants' perspectives on these aspects highlight the importance of RAI use.

P3 emphasised the importance of ensuring that the data utilised for training models is of the highest quality. High-quality data reduces the likelihood of bias and discrimination, ensuring the responsible use of AI.

> *"I think for me, the biggest thing that needs to be looked at from an improvement perspective, and it's something that's happening now, is data quality, that is ensuring that the data we're feeding our systems, our platforms, is accurate, is correct, as there's no duplication, there's no bias," (P3)*

Similarly, P8 stressed the importance of thoroughly vetting datasets used to train AI models. P8 advocated for careful validation to ensure data quality and emphasised the need for diversification to enhance fairness and reduce bias. This perspective highlights that RAI begins with the datasets used in training and development processes.

Similarly, Van Bekkum et al. (2024) strongly support the argument that data quality is crucial for preventing bias and discrimination in AI-driven insurance. They highlight how data-intensive underwriting, when based on poor-quality data, can lead to unintended biases and unfair differentiation in insurance pricing. This aligns with P3 and P8's perspective that AI models should be trained on accurate, complete, and bias-free data to ensure fairness in risk assessments. P8 stated:

> *"But if I can just speak generally, actually what we can do in the absence of that is ensure that the dataset used for training AI models are properly, properly vetted, alright? I think that will definitely help a lot. And special concentration should be put in place to ensure that there is a proper diversification in there which will improve fairness and also reduce bias." (P8)*

Another aspect highlighted by P4 focused on the need for organisations to manage client data effectively to mitigate risks when personalised decisions are being made that may impact the client negatively or positively. An emphasis on transparency, explainability, and accountability is necessary when the organisation relies on AI more to make decisions.

> *"Okay, the best way that we can mitigate the risk is to have a fantastic handle on what data we have about that client and what data about that client we use to either augment or ground some of these models when we start doing personalisation." (P4)*

The participants' insights collectively highlight the critical role of data integrity in mitigating ethical risks associated with AI. Ensuring high-quality, validated, and diversified datasets is foundational to building responsible and fair AI systems. Additionally, effective management of client data and prioritising transparency, explainability, and accountability are essential for mitigating risks in personalised decision-making. Together, these strategies highlight that robust data integrity practices are indispensable for fostering trust, promoting fairness, and ensuring the ethical use of AI technologies.

4.5 Conclusion

This chapter presented the findings from expert interviews, offering valuable insights into the ethical risks associated with AI-driven decision-making in the insurance industry, as well as the strategies currently employed to mitigate these risks. The participants highlighted various ethical concerns, including bias and discrimination, data governance and integrity issues, lack of transparency, and the potential for AI to produce unreliable or misleading outputs. Additionally, the findings accentuated the tension between efficiency-driven AI innovation and ethical considerations, with participants emphasising the need for a balanced approach that prioritises fairness, accountability, and customer trust. The study also revealed that while existing strategies such as regulatory compliance, AI governance frameworks, data quality assurance, and human oversight are considered somewhat effective, there remains a need for continuous improvement and refinement in response to evolving AI capabilities and regulatory landscapes.

Furthermore, the findings identified key future strategies that could enhance the ethical deployment of AI in the insurance sector. These include the development of robust AI governance frameworks, increased collaboration with industry stakeholders and regulators, continuous monitoring and auditing of AI models, and fostering an organisational culture of ethical awareness. The study suggests that adaptability to AI advancements, along with a commitment to RAI use, is essential for ensuring that AI-driven decision-making aligns with ethical and legal expectations. Ultimately, these insights contribute to a growing discourse on AI ethics in insurance, reinforcing the importance of proactive governance, transparency, and accountability in mitigating ethical risks while harnessing AI's potential for innovation and efficiency.

The next chapter, Chapter 5, discusses the findings in light of the conceptual framework.

CHAPTER 5. DISCUSSION OF FINDINGS

5.1 Introduction

From the findings presented in Chapter 4, it is evident that ensuring the responsible use of AI in the insurance industry requires a multifaceted approach. This chapter discusses the findings presented in Chapter 4 in light of the conceptual framework presented in Chapter 2. The chapter proposes a revised conceptual framework that depicts strategies that can be used to adhere to utilitarian and deontological ethics, processes that may benefit from those strategies, and the expected outcomes.

The findings are derived from nine experts within the South African insurance sector and reflect prevailing organisational and regulatory norms. Consequently, generalisation beyond this context should be undertaken cautiously, and further research is recommended to validate these insights in other jurisdictions and lines of business.

5.2 Ethical theories and themes

In this section, the researcher discusses the themes that emerged and how they map to the tenets of the utilitarian and deontological ethical theories.

5.2.1 Ethical risks

The following themes related to ethical risks associated with the use of AI within the insurance industry emerged from the thematic analysis:

- ***Bias and discrimination (P1, P6, P7, P9)***
  This theme highlights how bias and discrimination can emanate from both the data used to train AI models and the design of the models.
- ***Data quality and integrity concerns (P1, P3, P4)***
  Poor quality data can unearth ethical risks such as bias and discrimination. This means that AI decision-making can be compromised due to data quality and integrity concerns.

- ***AI and data governance concerns (P2, P7, P8)***

  The governance of AI and data is crucial to mitigating ethical risks; therefore, if there are no adequate frameworks and policies to regulate AI and data, then it can lead to ethical issues such as regulatory non-compliance and a lack of accountability.

- ***Lack of explainable AI (P1, P9)***

  The lack of transparency and explainability in AI decisions complicates the interpretation and justification of outcomes, raising concerns about accountability, fairness, and trust.
- ***Falsified information produced by AI (P4, P7)***

  AI hallucinations, whereby the models produce inaccurate, misleading or fabricated results as factual, raise ethical concerns regarding trust, accuracy, and reliability.
- ***Lack of efficiency and accuracy (P1, P3, P4)***

  The emergence of suboptimal and error-prone outputs from AI systems raises significant ethical concerns regarding their efficiency and accuracy.

5.2.2 Utilitarianism ethics

Utilitarianism is a consequentialist ethical theory that prioritises the outcomes of actions rather than the actions themselves (Driver, 2009). Its central aim is to maximise positive outcomes while minimising negative ones, thereby promoting the greatest overall good. The main tenets of utilitarianism are:

- Maximising overall good
- Cost-benefit analysis
- Accuracy and efficiency
- Continuous improvement.

The discussion below depicts how the themes that emerged from the participants' responses align with the main tenets of utilitarianism.

Maximising overall good

Maximising overall good refers to how organisations, industries, as well as countries, should strive to achieve the greatest benefits for the greatest number of people while minimising harm (Zoshak & Dew, 2021). The following theme that emerged from the thematic analysis align with this perspective:

- ***Collaboration and expertise (P1, P4, P5, P6)***
  This theme encourages diverse input from various stakeholders to develop robust ethical frameworks. This aligns with maximising overall good by encouraging collaborative efforts to enhance the societal benefits of AI systems.

Cost-benefit analysis

It is essential for insurance companies to consistently evaluate how AI-driven decisions affect their clients. This ongoing assessment is crucial to strike a balance between the insurance companies' objectives, including profitability, and important ethical considerations (Singh & Mishra, 2018). By doing so, these companies can ensure that their decision-making processes not only enhance their financial performance but also contribute positively to the overall well-being of society. The following theme aligns with this perspective:

- ***Balance innovation with ethics (P1, P4, P5, P6, P9)***
  Balancing innovation with ethics aligns with cost-benefit analysis as it evaluates the trade-offs between innovation and potential societal harm, striving for solutions that maximise overall benefit and minimise harm.

Accuracy and efficiency

Increased efficiency can be achieved through the integration of AI in insurance processes to streamline operations (Zoshak & Dew, 2021). Furthermore, having accurate information and predictions about the results of proposed actions is key to making decisions that truly maximise benefits (Cervantes et al., 2016). Therefore, accuracy and efficiency contribute to the maximisation of overall benefits, which is a core tenet of utilitarian ethics. The following themes align with this perspective:

- ***AI model selection, training, and validation (P5, P8, P9)***
  AI model selection, training, and validation align with the elements of accuracy and efficiency, as rigorous selection and validation processes ensure AI decisions are precise and optimised for maximum societal benefit.
- ***Aim for efficiency and accuracy (P4, P6, P8)***
  Organisations should prioritise accuracy and efficiency by optimising processes and outcomes through AI integration, ultimately maximising benefits for all users.

Continuous improvement

Continuous improvement of AI systems ensures that organisations remain competitive whilst leveraging AI opportunities. Additionally, ethical risks are evolving in parallel with advances in AI, therefore, AI systems need to be regularly reviewed and updated to address emerging ethical issues and mitigate their impact on client well-being (Singh & Mishra, 2018). The following themes align with this perspective:

- ***Adaptability to advances in AI (P1, P4, P6, P7)***
  Adaptability to advances in AI aligns with continuous improvement because adapting to technological advances ensures that AI remains beneficial and effective for future needs and use cases.
- ***Ongoing monitoring and development (P5, P6, P7, P8)***
  Ongoing monitoring and development are essential for effective continuous improvement. Given the dynamic nature of AI, organisations must address emerging ethical risks and maintain compliance with evolving regulatory standards. This requires an iterative approach that incorporates regular reviews, organisational adaptability, and the development of innovative use cases.

5.2.3 Deontological ethics

Deontology is an ethical framework that emphasises adherence to moral rules or principles rather than the consequences of an action (Alexander & Moore, 2021). According to this theory, an action is morally right if it aligns with a moral rule or principle, regardless of the resulting outcomes. The main tenets of deontology are:

- Adherence to moral rules
- Universalisability of actions
- Transparency and explainability
- Responsibility and accountability.

The discussion below depicts how the themes that emerged from the participants' responses align with the main tenets of deontology.

Adherence to moral rules

AI solutions, like any other technology, must be developed in such a way that they uphold established moral principles such as fairness, honesty, and respect for individual rights (Kumar et al., 2022). Therefore, organisations need to ensure their models do not discriminate. The following themes align with this perspective:

- ***Building ethical awareness and culture (P1, P3, P4, P6, P7)***
  This theme reflects adherence to moral rules as it emphasises a workplace committed to core ethical principles. By prioritising education, integrating ethics throughout the AI lifecycle, developing professionals who are knowledgeable about AI, and promoting ethical behaviour from leadership, organisations can ensure RAI use.
- ***Develop AI ethical frameworks (P4, P7, P8, P9)***
  This theme aligns with adherence to moral rules because ethical frameworks embody principles of fairness, justice, and respect for all parties involved.

Universalisability of actions

The actions taken by insurance companies must ensure equal treatment for all clients and potential clients (Micewski & Troy, 2007). Therefore, AI decision-making processes should always remain consistent and applicable to all similar cases. The following theme aligns with this perspective:

- ***Develop AI regulations and frameworks (P1, P3, P8, P9)***
  Developing AI regulations and frameworks aligns with the universalisability of actions as ethical frameworks establish principles applicable across contexts, ensuring consistent and morally grounded AI use. By integrating regulations

into AI implementation, the insurance industry can create governance systems that ensure RAI usage, foster trust, and promote sustainable innovation.

Transparency and explainability

AI solutions in insurance need to be transparent about how outcomes are determined and be explainable in terms of how AI reached its conclusions, for both clients and potential clients (Kumar et al., 2022). The following themes align with this perspective:

- ***Ensure data integrity (P1, P3, P8)***
  Ensuring data integrity provides a foundation for an organisation to be transparent and ensure explainability because maintaining accurate, high-quality, and truthful data supports clear and reliable decision-making.
- ***Ensure transparent and explainable AI (P1, P4, P9)***
  Ensuring fairness, transparency, and explainability in AI aligns strongly with the ethical principle of being open and accountable, reducing the harm caused by opaque or misunderstood AI processes. This fosters trust, respect, and equity in AI-driven insurance decisions.

Responsibility and accountability

Insurance companies should implement clear accountability mechanisms for AI decision-making to ensure that ADM processes align with ethical standards (Zoshak & Dew, 2021). This involves maintaining human oversight of AI outputs and establishing robust procedures to address any ethical breaches or errors that may arise during decision-making. The following themes align with this perspective:

- ***Auditing against frameworks (P4, P6, P8)***
  While it is important to have frameworks in place to mitigate ethical risk, they need to be enforced to ensure the responsibility and accountability of the organisation. This is why auditing is important to ensure adherence to these frameworks.
- ***Ensure human oversight in AI decision-making (P1, P4, P5)***
  Ensuring HITL decision-making aligns with responsibility and accountability because it ensures that humans retain oversight and ultimate responsibility for AI-driven decisions. By involving humans in critical decision points,

organisations preserve ethical judgment, address AI limitations, and ensure compliance with legal and moral obligations.

- ***Establish AI and data governance committees (P1, P8, P9)***
  The theme embodies responsibility and accountability by creating structured governance bodies to oversee and enforce ethical AI usage within organisations.
- ***Data regulations (P1, P4, P8)***
  Data regulations such as the POPIA and the GDPR closely align with responsibility and accountability as they impose ethical and legal duties on organisations to ensure fair, transparent, and lawful handling of personal data. While data regulations may indirectly support transparency and explainability by promoting clear data handling practices, their primary focus on ethical stewardship and compliance positions them more strongly within the realm of responsibility and accountability.

5.2.4 Themes that intersect with utilitarian and deontological ethics

- ***Ensuring fairness (P1, P7)***
  Fairness aligns with **maximising overall good** (utilitarian perspective) by promoting equitable treatment, reducing harm, and fostering trust in AI systems, thereby benefiting society as a whole. Simultaneously, it reflects the **universalisability of actions** (deontological perspective) by upholding the moral obligation to treat all individuals consistently and justly, regardless of personal characteristics. This dual alignment ensures that fairness in AI decision-making balances societal well-being with adherence to universal ethical principles.
- ***Regulation compliance and accountability (P1, P5, P6, P9)***
  Regulation compliance and accountability align with **cost-benefit analysis** (utilitarian perspective) by ensuring that societal and organisational benefits, such as fairness, data protection, and reputational preservation, outweigh the costs of compliance. From a utilitarian perspective, adhering to regulations like privacy and anti-discrimination laws maximises overall good by safeguarding customer trust and preventing harm. Simultaneously, it reflects **adherence to moral rules** (deontological perspective) by upholding a principled duty to act

ethically and within legal boundaries, ensuring justice and respect for individual rights irrespective of outcomes. Together, these perspectives integrate practical utility with moral responsibility.

Table 5.1 illustrates the alignment between the themes that represent the mitigation strategies and the tenets of the utilitarian and deontological ethical theories.

Table 5.1: Theme to theory mapping

Theme	Utilitarianism tenets	Deontology tenets
Adaptability to advances in AI	Continuous improvement	
AI model selection, training, and validation	Accuracy and efficiency	
Aim for efficiency and accuracy	Accuracy and efficiency	
Auditing against frameworks		Responsibility and accountability
Balance innovation with ethics	Cost-benefit analysis	
Building ethical awareness and culture		Adherence to moral rules
Collaboration and expertise	Maximising overall good	
Continuous monitoring and improvement	Continuous improvement	
Develop AI ethical frameworks		Adherence to moral rules
Develop AI regulations and frameworks		Universalisability of actions
Ensure data governance		Responsibility and accountability
Ensure data integrity		Transparency and explainability

Theme	Utilitarianism tenets	Deontology tenets
Ensure fairness	Maximising overall good	Universalisability of actions
Ensure human oversight in AI decision-making		Responsibility and accountability
Ensure transparent and explainable AI		Transparency and explainability
Establish AI and data governance committees		Responsibility and accountability
Regulation compliance and accountability	Cost-benefit analysis	Adherence to moral rules

5.3 Proposed conceptual framework

5.3.1 Breakdown of the conceptual framework

The proposed revised conceptual framework integrates utilitarian and deontological ethical theories to guide AI-driven decision-making processes in the insurance industry. Its primary objective is to achieve ethical outcomes while leveraging AI technologies in critical insurance operations, including underwriting, claims management, and pricing.

The framework is grounded in the foundational ethical theories of utilitarianism and deontology. Utilitarianism focuses on maximising overall good by employing cost-benefit analyses, enhancing accuracy and efficiency, and fostering continuous improvement in decision-making processes. Conversely, deontological principles emphasise strict adherence to moral rules, the universalisability of actions, and key ethical attributes such as transparency, explainability, responsibility, and accountability. Together, these principles ensure that AI systems in the insurance industry are designed and deployed in a manner that aligns with established ethical norms and moral duties. The framework should be interpreted as a guide toward optimising the benefits of AI applications while carefully balancing operational efficiency with ethical considerations.

The themes that emerged from the participants' responses help understand the ethical considerations that should guide AI in decision-making within the insurance industry. As explained in the previous section, these themes are linked to specific ethical elements derived from the ethical theories of utilitarianism and deontology. The themes presented in the framework were developed from participants' responses and the researcher's interpretation of those responses. Research question 1 revealed themes related to ethical risks associated with the use of AI in the insurance industry. Research questions 2 and 4 identified themes specifically focused on strategies to mitigate these ethical risks. Research question 3 influenced the themes that emerged in research questions 2 and 4.

As depicted in Chapter 3, section 3.7.1, the research questions in this study were designed to provide a structured exploration of the ethical risks associated with AI-driven decision-making in the insurance industry. They were sequentially linked to facilitate a comprehensive examination of the issue, progressing from risk identification to the evaluation and improvement of mitigation strategies. This logical progression ensured a holistic understanding of the ethical implications of AI deployment in insurance while also enabling the formulation of recommendations for ethical AI governance.

Although the themes from research questions 1 and 3 are not explicitly represented in the conceptual framework, they inform and contribute to the themes aligned with strategies to mitigate ethical risks.

5.3.1.1 Sequential development of themes from research questions

The first research question, "What ethical risks arise from using AI in decision-making within the insurance industry?" served as the foundation of the study by identifying and categorising the key ethical risks inherent in AI-driven decision-making processes. The themes emerging from this question provide a basis for understanding the challenges that necessitate mitigation efforts. These themes feed into research question 2, as they establish the risks that existing mitigation strategies seek to address. Furthermore, they also inform the themes in research question 3, which evaluated the effectiveness of those mitigation strategies. Without a clear articulation of the risks, it would be impossible to assess how well current approaches are mitigating them.

Additionally, the insights gained from identifying ethical risks contribute to research question 4, as they highlight gaps that future strategies must aim to address.

Building upon the ethical risks identified, the second research question, “How are these ethical risks currently addressed in the insurance industry?”, examined the existing strategies and mechanisms used to mitigate AI-related ethical concerns. The themes emerging from this question provide a structured overview of current interventions, including regulatory compliance, ethical AI frameworks, and governance structures. These themes directly feed into those of research question 3, as they establish the baseline against which effectiveness is assessed. If certain risks remain inadequately addressed, the effectiveness of existing strategies is called into question. Additionally, these themes influence research question 4 by highlighting areas where strategies require improvement or innovation, ensuring that future mitigation efforts build upon the strengths and limitations of current approaches.

The third research question, “How effective are the current strategies used to address ethical risks in the insurance industry?”, critically evaluated the impact and adequacy of existing mitigation strategies. This question builds on the second by moving beyond the identification of strategies to assess their practical implementation and measurable success. The themes emerging from this question not only reflect the successes and limitations of current strategies but also feed directly into research question 4, which sought to propose improved mitigation measures. By identifying gaps and shortcomings in existing efforts, these themes help shape the development of future strategies. Additionally, the themes from research question 3 indirectly link back to those in research question 1, as any persistent ethical risks, despite mitigation efforts, suggest that certain risks require more targeted intervention.

Finally, the fourth research question, “What strategies can be implemented to ensure AI technologies are used ethically in the insurance industry?”, is forward-looking and aimed at identifying improvements and innovations for ethical AI governance. The themes emerging from this question build directly on the findings from research questions 2 and 3, ensuring that the proposed strategies address identified gaps and enhance the effectiveness of existing approaches. These themes are also informed by those from research question 1, as any newly proposed strategies must account for the full spectrum of ethical risks to ensure comprehensive mitigation. Thus, the themes from research question 4 serve as a culmination of the thematic insights derived from the preceding questions, integrating risk identification, assessment of current strategies, and evaluation of their effectiveness into a coherent set of recommendations.

5.3.1.2 AI decision-making solutions and processes

AI decision-making solutions are central to the framework and are applied across underwriting, pricing, and claim management. Underwriting and pricing involves assessing risks to determine insurance policy terms and pricing. Claim management focuses on processing and handling insurance claims. By incorporating ethical principles into these AI-driven processes, the framework ensures that AI decision-making systems enhance decision-making while adhering to ethical standards.

The implementation of strategies to mitigate ethical risks associated with AI in decision-making within the insurance industry yields a range of critical outcomes. These outcomes reflect the extent to which organisations can address ethical concerns while leveraging the benefits of AI technology. They are interconnected and collectively contribute to fostering trust, ensuring fairness, and promoting accountability in AI-driven processes. The following section explores these outcomes in detail, emphasising their importance in achieving RAI ethical innovation and aligning AI systems with both regulatory standards and societal expectations.

5.3.1.3 Outcomes

Ethical innovation

Ethical innovation refers to the development and implementation of novel technologies, processes, or systems that prioritise ethical considerations. By embedding ethical principles into AI-driven processes within the insurance industry, organisations can ensure that innovation promotes not only operational efficiency and profitability but also upholds societal and moral values. This outcome enhances stakeholder trust and positions organisations as leaders in RAI practices.

Adaptability to ethical risk

Adaptability to ethical risk signifies an organisation's ability to dynamically respond to new and emerging ethical challenges posed by advancements in AI technology. As AI systems evolve, unforeseen risks such as novel biases or unintended consequences may arise. This outcome ensures that mitigation strategies remain relevant, enabling organisations to address these challenges proactively while safeguarding ethical integrity.

Accuracy and efficiency

The implementation of effective strategies improves the accuracy of AI-driven decision-making processes while maintaining operational efficiency. This outcome ensures that AI systems generate precise and reliable outputs, minimising errors that could result in unethical practices, such as biased claims processing or inequitable pricing. Enhancing both accuracy and efficiency reinforces trust in AI systems and supports resource optimisation within organisations.

Fairness

Fairness is a fundamental ethical principle that ensures AI systems treat all customers equitably, irrespective of their demographics or personal circumstances. By addressing biases in data and algorithms, effective mitigation strategies foster fairness in outcomes, particularly in areas such as underwriting, claims approval, and customer engagement. This outcome aligns with both deontological ethics, which emphasise duty and principles, and utilitarian ethics, which focus on maximising equitable outcomes for all stakeholders.

Transparency and explainability

Transparency and explainability are critical outcomes that enable organisations to provide clear and comprehensible justifications for AI-driven decisions. These qualities enhance trust and accountability by allowing customers, regulators, and other stakeholders to scrutinise the reasoning behind AI outputs. By promoting transparency, organisations demonstrate their commitment to ethical decision-making and RAI use.

Client privacy

Ensuring client privacy involves the robust governance and protection of sensitive customer information. This outcome is achieved through effective data governance and cybersecurity measures that prevent the misuse or breach of personal data. By safeguarding privacy, organisations not only comply with regulatory requirements but also encourage customer confidence and loyalty, reinforcing the ethical use of AI systems.

Regulation compliance

Compliance with relevant regulatory frameworks and industry standards represents a critical outcome of ethical risk mitigation strategies. This ensures organisations align with legal and societal expectations for RAI use. By adhering to these requirements, organisations minimise the risk of legal sanctions, maintain their reputation, and demonstrate their commitment to upholding ethical standards.

Accountability

Accountability entails the clear assignment of responsibility for decisions and actions associated with AI systems. Effective mitigation strategies establish governance structures and define roles to ensure individuals and organisations are held accountable for the outcomes of AI-driven processes. This outcome promotes ethical behaviour, enhances credibility, and ensures alignment with both organisational and societal ethical principles.

The proposed conceptual framework is depicted in Figure 5.1 below.

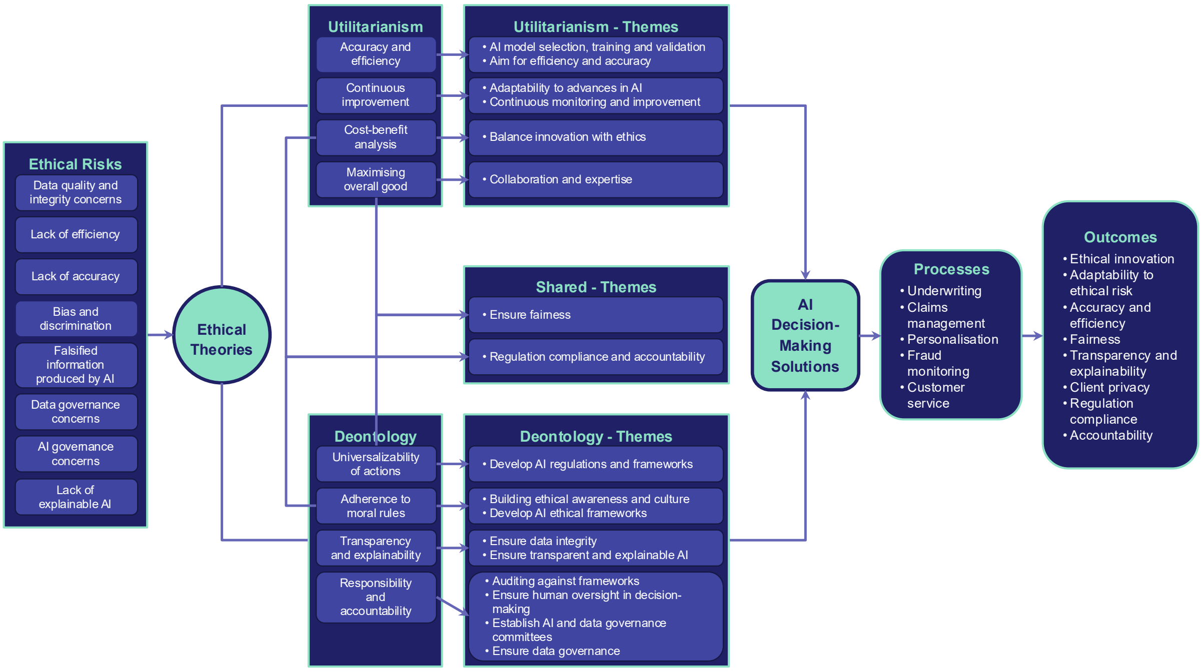


Figure 5.1: Proposed conceptual framework

5.3.2 Practical implementation of the conceptual framework

As Eling et al. (2022) note, AI enables advanced risk assessment in pricing and underwriting, allowing insurers to differentiate premiums based on increasingly detailed data points. From a utilitarian perspective, this can be justified as it promotes efficiency and ensures that premiums more accurately reflect individual risk, thereby benefiting the majority of policyholders through lower average costs. However, a deontological perspective highlights ethical concerns where this practice could disadvantage vulnerable groups or indirectly reinforce bias. To balance these perspectives, insurers could adopt fairness constraints in pricing algorithms, ensuring that while risk-based differentiation enhances overall affordability, protections are introduced to prevent discrimination against specific groups. In this way, both aggregate benefits (utilitarianism) and individual rights to fair treatment (deontology) are upheld.

In addition, Eling et al. (2022) also emphasise the growing role of AI in claims management, particularly in detecting fraud and expediting settlement. A utilitarian approach supports this application, as it reduces overall costs and accelerates processing times, producing benefits for both insurers and the broader customer base. In contrast, a deontological perspective underscores the duty of insurers to ensure fairness and transparency in decision-making. A balanced approach would be to use AI for initial fraud detection and claims triage, while introducing a human-in-the-loop review for contested or high-stakes cases. This preserves efficiency gains (utilitarianism) while ensuring fairness and accountability in protecting claimants' rights (deontology).

These examples demonstrate how utilitarian and deontological reasoning illuminate tensions in AI-driven decision-making across the insurance value chain. By showing how both perspectives can be simultaneously addressed in practice, the framework provides not only a lens for interpreting ethical risks but also a roadmap for developing actionable mitigation strategies in the insurance sector.

Barriers to adoption and mitigation strategies

While the proposed conceptual framework offers a structured approach for embedding ethical principles into AI use in insurance, several barriers to adoption may limit its effectiveness. A key challenge lies in organisational resistance, where insurers may prioritise cost efficiency and speed-to-market over ethical safeguards. Additionally, resource constraints, such as the availability of skilled professionals capable of conducting ethical audits, may hinder implementation. Regulatory uncertainty in South Africa, where AI governance guidelines are still evolving, also poses a risk by creating ambiguity around compliance expectations.

Overcoming these barriers requires deliberate strategies. Organisational resistance can be mitigated through strong leadership endorsement, where executive teams champion ethical AI as both a compliance necessity and a source of long-term customer trust. Resource limitations may be addressed by investing in cross-functional training to upskill existing staff, while partnerships with academic or industry bodies could supplement internal expertise. To counter regulatory ambiguity, insurers can align their practices with emerging global guidelines while actively engaging with local regulators to shape context-appropriate standards.

5.4 Conclusion

The findings of this study underline the complexity of ethical risks associated with AI-driven decision-making in the insurance industry and the necessity of a structured, ethical approach to mitigating these challenges. Through a thematic analysis aligned with utilitarian and deontological ethical theories, key ethical risks such as bias, lack of transparency, regulatory non-compliance, and data integrity concerns were identified. The study further examined existing strategies employed within the industry to mitigate these risks, including AI governance frameworks, human oversight mechanisms, and regulatory compliance initiatives. While these strategies address ethical concerns to a certain extent, their effectiveness varies, necessitating continuous improvement and adaptive governance approaches.

The proposed revised conceptual framework developed in this study integrates ethical theories into AI-driven insurance processes, ensuring that decision-making aligns with both utilitarian principles of accuracy and efficiency, continuous improvement, cost-benefit analysis, and maximising overall benefit, and deontological principles of universalisability of actions, adhering to moral rules, transparency and explainability, and responsibility and accountability. The framework highlights the importance of accountability, fairness, transparency, and adaptability to emerging AI advancements. By fostering ethical innovation and regulatory compliance, it provides a pathway for organisations to balance operational efficiency with RAI deployment.

Importantly, the study also highlights how the conceptual framework can be practically implemented within insurance operations. For example, the framework can be operationalised through targeted interventions such as fairness constraints in underwriting algorithms and human-in-the-loop oversight in claims processing. These examples illustrate how the framework moves beyond theory, providing insurers with actionable strategies for embedding ethical principles, balancing efficiency, transparency, and fairness into core AI-enabled processes.

Ultimately, this study reaffirms the critical role of ethical oversight in AI applications within the insurance industry and advocates for a continuous, multi-stakeholder approach to safeguarding policyholder interests while leveraging the benefits of AI technologies.

The next chapter, Chapter 6, concludes the study and provides recommendations.

CHAPTER 6. CONCLUSIONS AND RECOMMENDATIONS

6.1 Introduction

This chapter concludes the study. To begin with, a summary of the study, by chapter, is presented. This is followed by the contributions and implications of the study, its limitations, and recommendations for future research. The chapter (and study) ends with some concluding remarks.

6.2 Summary of the study

The primary aim of this qualitative study was to explore the ethical implications of AI-driven decision-making in the insurance industry. The study comprised six chapters. Each chapter played an important role, from explaining the rationale for the research to exploring the applicable existing literature, developing a conceptual framework, data collection, thematic analysis and discussing the findings. This section summarises the study by reviewing the main points of each chapter:

Chapter 1

The first chapter provided a foundation for the study by outlining its background, rationale, and the research problem. It introduced the research problem, which examines the ethical implications of AI-driven decision-making in the insurance industry. AI is increasingly being used in underwriting, claims management, and pricing to enhance efficiency and accuracy. However, its implementation raises ethical concerns such as algorithmic bias, lack of transparency, regulatory non-compliance, and data privacy issues. The chapter framed these ethical risks within the theoretical perspectives of utilitarian and deontological ethics, ensuring a balanced evaluation of AI's benefits and moral obligations. The study aimed to identify these ethical risks, assess current mitigation strategies, and propose additional measures to foster RAI deployment. Through a qualitative approach, the research sought to contribute to ethical AI governance in the insurance sector.

Chapter 2

The second chapter provided a comprehensive literature review. It explored the integration of AI, ethics, and insurance, providing insights into how AI technologies are reshaping decision-making processes. AI-driven automation enhances efficiency in risk assessment, customer service, and fraud detection, yet ethical concerns persist. Algorithmic bias may result in unfair insurance pricing, while data integrity issues undermine trust in AI-generated decisions. Privacy concerns arise due to the vast data requirements of AI models and the opacity of AI decision-making, often referred to as the black box problem, challenges accountability. Existing mitigation strategies include regulatory frameworks such as the POPIA and the GDPR, fair-ML techniques, governance committees, and transparency measures such as XAI. This chapter also brought together the theoretical framework and gave rise to the conceptual framework that integrated both utilitarian and deontological perspectives to ensure RAI deployment.

Chapter 3

The third chapter detailed the qualitative research approach used to explore the ethical risks of AI integration into insurance processes. An interpretivism research philosophy and a qualitative research methodology were adopted. Purposive sampling was used, with participants being chosen based on their experience in the insurance industry and/or experience and knowledge of AI. The study also used the snowball sampling strategy, by which the initial group of participants referred other potential participants who met the sampling criteria. Thematic analysis, following Braun and Clarke's six-phase approach, was employed to identify key themes emerging from the data. Nine participants were interviewed online using an interview guide comprising open-ended questions. Ethical considerations, including informed consent, data confidentiality, and compliance with ethical research guidelines, were central to the research process.

Chapter 4

The fourth chapter presented the findings of the thematic analysis of the transcripts of the interviews held with the nine participants. The focus was on the ethical risks of AI-driven decision-making in the insurance industry, the current strategies used to mitigate these risks, the effectiveness of those strategies, and recommendations for future improvements. The findings were organised according to the study's four research questions, highlighting key themes that emerged from the data.

Research question 1's findings confirmed that AI-driven decision-making in the insurance industry introduced significant ethical risks. Participants identified several ethical risks including bias, discrimination, lack of transparency, and data integrity concerns.

Research question 2 explored current strategies being implemented to mitigate the ethical risk of the integration of AI within insurance processes. Participants identified current strategies such as regulatory compliance, AI governance frameworks, transparency and explainability, and HITL decision-making.

Research question 3 explored the effectiveness of current strategies in addressing ethical risks. Participants expressed mixed views on their effectiveness. While current strategies contributed to addressing AI ethics, their effectiveness depended on continuous monitoring, industry collaboration, and regulatory evolution.

Research question 4 explored future strategies for ensuring ethical AI use in the insurance industry. Participants recommended several strategies to improve the ethical governance of AI in insurance, including auditing, adaptability to AI advances, and continuous monitoring and development of AI-specific regulations and frameworks.

Chapter 5

The fifth chapter provided an in-depth analysis of the findings in relation to the study's conceptual framework, as well as the broader literature on AI ethics in the insurance industry. The discussion synthesised the themes that emerged from the research, focusing on how the ethical risks identified aligned with or diverged from existing knowledge and the extent to which current and proposed mitigation strategies addressed these concerns. The findings reinforced that AI-driven decision-making in the insurance sector presented significant ethical risks, particularly in the areas of algorithmic bias, lack of transparency, data privacy, and regulatory compliance. These risks were consistent with concerns raised in prior research, further validating the necessity for vigorous AI governance and ethical frameworks.

The study reaffirmed the importance of integrating utilitarian and deontological ethics into AI governance, ensuring that AI systems maximised efficiency and accuracy while upholding moral obligations such as fairness, accountability, and transparency. Therefore, a revised framework was proposed that aligns with these utilitarian and deontology principles. The framework is informed by themes derived from participant responses that describe ethical risks (research question 1), current mitigation strategies (research question 2), assessing their effectiveness (research question 3), and future mitigation strategies (research question 4) that address gaps in current strategies. The chapter concluded that an adaptive and multi-layered approach to AI ethics was essential for mitigating risks while leveraging AI's potential to improve decision-making and operational efficiency.

6.3 Contributions of the study

This study makes significant contributions to both academic literature and industry practice by addressing key gaps in AI ethics within the insurance industry in the South African context. It explores the ethical risks associated with AI-driven decision-making, evaluates current mitigation strategies, and proposes recommendations for ensuring RAI utilisation. By integrating theoretical and practical insights, the study advances the understanding of AI ethics while providing actionable strategies for industry stakeholders.

6.3.1 Theoretical contributions

This research makes a theoretical contribution by applying utilitarian and deontological ethics as a dual framework for evaluating AI decision-making in the insurance industry. While previous studies have predominantly focused on individual ethical theories, this study integrated the two theories to provide a comprehensive understanding of the ethical risks that need to be addressed in the insurance industry concerning AI-driven decision-making.

6.3.2 Theoretical implications

The ethical implications of AI adoption in the insurance industry are multifaceted, requiring a nuanced and comprehensive understanding. This study highlights that addressing AI ethics cannot be effectively achieved through a singular ethical perspective, whether utilitarianism (which focuses on consequences) or deontology (which emphasises duty and moral principles). Instead, a balanced integration of both frameworks is necessary.

By combining utilitarianism and deontological ethics, this study provides a structured approach to navigating the ethical complexities of AI in insurance. It emphasises the need to maximise overall societal benefits while simultaneously upholding fundamental moral principles. This theoretical contribution offers a framework for balancing efficiency with ethical responsibility.

6.3.3 Practical implications

From an industry perspective, the study provides valuable recommendations for insurance companies, regulators, AI developers, and policymakers seeking to ensure ethical AI implementation. It highlights the importance of transparency and accountability in AI-driven decision-making, recommending that insurance firms adopt XAI practices to enhance consumer trust. The study suggests that AI ethics training should be embedded within organisational policies to improve employees' understanding of AI's impact on customer outcomes and regulatory compliance.

The findings also point to the need for continuous monitoring and auditing of AI systems. Given that AI models evolve, periodic audits should be conducted to assess fairness, accuracy, and bias in decision-making processes. Additionally, the study advocates for the establishment of AI ethics governance committees within insurance firms to oversee AI deployment and ensure adherence to ethical standards. Policymakers are encouraged to develop sector-specific AI regulations that go beyond general data protection laws, providing clear guidelines on fairness, accountability, and explainability in AI-driven insurance decisions.

6.3.4 Policy implications

In addition to its theoretical and practical contributions, the study offers valuable insights for policy development within the South African insurance industry. The identification of ethical risks such as bias, lack of transparency, and regulatory non-compliance highlights areas where policymakers can strengthen oversight frameworks. For example, the findings suggest the need for clearer guidance on explainability standards in AI-driven decision-making and more robust requirements for fairness audits of pricing and claims algorithms. By aligning regulatory frameworks with both utilitarian objectives (ensuring overall benefits such as affordability and efficiency) and deontological duties (safeguarding fairness, rights, and accountability), policymakers can provide insurers with actionable standards that promote responsible AI adoption.

The study therefore contributes to the broader discourse on AI governance in South Africa by offering evidence-based recommendations that can inform regulatory updates, industry codes of conduct, and ethics-by-design principles in AI development pipelines. These contributions have the potential to not only improve customer outcomes in insurance but also to serve as starting points for other financial services sectors adopting AI.

6.4 Limitations of the study

This study has some limitations that future research in this area can address. While the sample size was relatively small, it was composed of highly knowledgeable and experienced professionals who met the study's criteria. Their expertise helped compensate for the limited number of participants, ensuring valuable, in-depth insights. As AI adoption in the insurance industry is still in its early stages, the experiences of some participants with AI will naturally expand over time. However, their insights remain highly valuable, as they highlight key considerations that must be addressed as AI integration progresses in the industry. While early-stage adoption may limit the depth of practical experience, the perspectives shared by participants still provide crucial foresight into emerging ethical risks and necessary mitigation strategies.

Another limitation concerns the evolving nature of AI technologies and regulatory frameworks. AI-driven decision-making is a rapidly developing field, with continuous advancements in ML models, regulatory oversight, and governance mechanisms. As a result, the study's findings represent a snapshot in time and may require ongoing validation as new AI policies, ethical considerations, and insurance applications emerge. Lastly, although it investigated the ethical implications of AI within the insurance industry primarily from a customer or policyholder perspective, it did not focus on these implications from an employee or other stakeholder point of view.

6.5 Recommendations for future research

The study identified several areas for future research that could enhance the understanding and implementation of ethical AI in the insurance industry. First, there is a need for studies that assess the long-term impact of AI governance frameworks and ethical guidelines. As AI technologies continue to evolve, research should examine how regulatory measures and industry standards adapt over time and whether they effectively mitigate emerging ethical risks.

Future research could also explore comparative studies across different territories to assess how different regulatory environments shape AI ethics in the insurance sector. Given the varying regulatory landscapes across countries, studies comparing AI governance in regions with strict data protection laws, such as Europe's GDPR, with less regulated markets could provide valuable insights into best practices for global AI governance.

Furthermore, future research could consider creating a comprehensive ethical risk framework specifically designed for insurance companies. This framework should take into account the unique challenges and responsibilities these organisations face, including data privacy, algorithmic bias, and transparency. Moreover, it would be beneficial to explore the principles of ethics by design, which advocate for the integration of ethical considerations at every stage of the algorithm development process. This includes early-stage planning, data collection, model training, and deployment, ensuring that ethical safeguards are embedded in the technology from the outset. By implementing these principles, insurance companies can foster a culture of accountability and fairness, ultimately leading to more trustworthy and responsible algorithmic practices.

Another crucial area for future research involves customers' perspectives on AI-driven decision-making. While much of the existing research focuses on industry and regulatory viewpoints, there is limited understanding of how customers perceive AI-driven insurance decisions, particularly in relation to fairness, transparency, and trust. Investigating consumer responses to AI-driven underwriting claims, assessments, and policy pricing could help refine AI governance strategies to align with customer expectations.

Additionally, research could focus on the intersection of AI and human oversight in insurance decision-making. The effectiveness of HITL systems remains a contentious issue, with some studies suggesting that human oversight can either mitigate or introduce biases. Future studies could investigate how human oversight can be optimally integrated into AI decision-making without undermining efficiency or ethical integrity.

6.6 Concluding remarks

This study has provided a comprehensive exploration of the ethical implications of AI-driven decision-making in the insurance industry, offering valuable insights into the risks, current mitigation strategies, and future governance considerations. Across its chapters, the research has highlighted the transformative potential of AI in enhancing efficiency, accuracy, and automation in insurance processes. However, it has also highlighted the pressing ethical challenges associated with algorithmic bias, transparency deficits, data privacy concerns, and regulatory non-compliance. By examining these risks through the lens of utilitarian and deontological ethics, the study has contributed to the growing discourse on RAI governance in the insurance sector.

A key contribution of this research lies in its identification of gaps within existing mitigation strategies and the recommendation of a more structured, proactive approach to ethical AI deployment. While regulatory frameworks such as the POPIA and the GDPR provide a foundation for AI governance, the study has revealed that these measures are often reactive rather than adaptive to the fast-evolving nature of

AI technologies. The findings suggest that the integration of comprehensive AI governance frameworks, continuous monitoring and auditing mechanisms, and enhanced transparency initiatives is essential for ensuring that AI-driven insurance decisions remain fair, accountable, and ethically sound.

Despite its contributions, the study has limitations. The reliance on a small sample of industry experts may have restricted the breadth of perspectives captured, and the rapidly evolving nature of AI presents challenges in ensuring that the findings remain applicable over time. These limitations highlight the need for further research to validate and expand upon the study's findings, particularly in areas such as consumer perspectives on AI ethics, cross-regional regulatory comparisons, and the role of human oversight in AI decision-making.

Looking ahead, the responsible adoption of AI in the insurance industry will require ongoing collaboration between insurers, regulators, technology developers, and policymakers. The study's recommendations provide a foundation for industry-wide efforts to establish ethical AI practices that balance innovation with accountability. By fostering greater transparency, implementing adaptive regulatory frameworks, and prioritising consumer protection, the insurance sector can ensure that AI is leveraged as a force for positive transformation rather than a source of ethical risk.

Finally, this study has reinforced the importance of a multi-layered and continuously evolving approach to AI ethics in insurance. The proposed revised conceptual framework serves as a strategic model for integrating ethical principles into AI-driven decision-making, promoting an ecosystem where AI can maximise overall good while safeguarding moral rules. As AI technologies continue to advance, the industry must remain committed to ethical decision-making that prioritises fairness, transparency, and public trust. The insights presented in this research serve as a guiding framework for future initiatives aimed at ensuring that AI-driven decision-making aligns with societal and ethical expectations, paving the way for a more responsible and equitable insurance landscape.

REFERENCES

Akinrinola, O., Okoye, C. C., Ofodile, O. C., Ugochukwu, C. E., Akinrinola, O., Okoye, C. C., Ofodile, O. C., & Ugochukwu, C. E. (2024). Navigating and Reviewing Ethical Dilemmas in AI Development: Strategies for Transparency, Fairness, and Accountability. *GSC Advanced Research and Reviews*, *18*(3), Article 3. https://doi.org/10.30574/gscarr.2024.18.3.0088

Alexander, L., & Moore, M. (2021). Deontological Ethics. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2021). Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/win2021/entries/ethics-deontological/

Aysolmaz, B., Müller, R., & Meacham, D. (2023). The Public Perceptions of Algorithmic Decision-Making Systems: Results from a Large-Scale Survey. *Telematics and Informatics*, *79*, 101954. https://doi.org/10.1016/j.tele.2023.101954

Baker, M., & Rajab, H. (2024). *Balancing Innovation and Ethics in AI*.

Balasubramanian, R., Libarikian, A., & McElhaney, D. (2021, March 12). *Insurance 2030—The impact of AI on the future of insurance | McKinsey*. https://www.mckinsey.com/industries/financial-services/our-insights/insurance-2030-the-impact-of-ai-on-the-future-of-insurance

Baluja, H. (2024, February 7). *The Complete Guide to AI Algorithms*. https://blog.hubspot.com/marketing/ai-algorithms

Ben-Shahar, O. (2023). Privacy Protection, At What Cost? Exploring the Regulatory Resistance to Data Technology in Auto Insurance. *Journal of Legal Analysis*, *15*(1), 129–157. https://doi.org/10.1093/jla/laad008

Bhandari, P. (2020a, June 5). *Data Collection | Definition, Methods & Examples*. Scribbr. https://www.scribbr.com/methodology/data-collection/

Bhandari, P. (2020b, June 19). *What Is Qualitative Research? | Methods & Examples*. Scribbr. https://www.scribbr.com/methodology/qualitative-research/

Bharadhwaj, H., Huang, D.-A., Xiao, C., Anandkumar, A., & Garg, A. (2021). *Auditing AI Models for Verified Deployment Under Semantic Specifications*. https://doi.org/10.48550/arXiv.2109.12456

Birkstedt, T., Minkkinen, M., Tandon, A., & Mäntymäki, M. (2023). AI Governance: Themes, Knowledge Gaps and Future Agendas. *Internet Research*, *33*(7), 133–167. https://doi.org/10.1108/INTR-01-2022-0042

Braun, V., & Clarke, V. (2006). Using Thematic Analysis in Psychology. *Qualitative Research in Psychology*, *3*, 77–101. https://doi.org/10.1191/1478088706qp063oa

Breidbach, C. F. (2024). Responsible Algorithmic Decision-Making. *Organizational Dynamics*, *53*(2), 101031. https://doi.org/10.1016/j.orgdyn.2024.101031

Buçinca, Z., Malaya, M. B., & Gajos, K. Z. (2021). To Trust or to Think: Cognitive Forcing Functions Can Reduce Overreliance on AI in AI-assisted Decision-making. *Proceedings of the ACM on Human-Computer Interaction*, *5*(CSCW1), 1–21. https://doi.org/10.1145/3449287

Cai, L., & Zhu, Y. (2015). The Challenges of Data Quality and Data Quality Assessment in the Big Data Era. *Data Science Journal*, *14*. https://doi.org/10.5334/dsj-2015-002

Catlin, T., Lorenz, J.-T., Münstermann, B., & Ricciardi, V. (2017, March 1). *Insurtech: The Threat That Inspires*. Mckinsey. https://www.mckinsey.com/industries/financial-services/our-insights/insurtech-the-threat-that-inspires

Cervantes, J.-A., Rodríguez, L.-F., López, S., Ramos, F., & Robles, F. (2016). Autonomous Agents and Ethical Decision-Making. *Cognitive Computation*, *8*(2), 278–296. https://doi.org/10.1007/s12559-015-9362-8

Colson, E. (2019, July 8). *What AI-Driven Decision Making Looks Like*. https://hbr.org/2019/07/what-ai-driven-decision-making-looks-like

Curzon, J., Kosa, T. A., Akalu, R., & El-Khatib, K. (2021). Privacy and Artificial Intelligence. *IEEE Transactions on Artificial Intelligence*, *2*(2), 96–108. IEEE Transactions on Artificial Intelligence. https://doi.org/10.1109/TAI.2021.3088084

Dancy, C. L., & Saucier, P. K. (2022). AI and Blackness: Toward Moving Beyond Bias and Representation. *IEEE Transactions on Technology and Society*, *3*(1), 31–40. IEEE Transactions on Technology and Society. https://doi.org/10.1109/TTS.2021.3125998

De Almeida, P. G. R., Dos Santos, C. D., & Farias, J. S. (2021). Artificial Intelligence Regulation: A framework for governance. *Ethics and Information Technology*, *23*(3), 505–525. https://doi.org/10.1007/s10676-021-09593-z

Deekshith, A. (2021). Data Engineering for AI: Optimizing Data Quality and Accessibility for Machine Learning Models. *International Journal of Management Education for Sustainable Development*, *4*(4), 1–33.

Dexe, J., Franke, U., & Rad, A. (2021). Transparency and Insurance Professionals: A Study of Swedish Insurance Practice Attitudes and Future Development. *The Geneva Papers on Risk and Insurance. Issues and Practice*, *46*(4), 547–572. https://doi.org/10.1057/s41288-021-00207-9

Dhirani, L. L., Mukhtiar, N., Chowdhry, B. S., & Newe, T. (2023). Ethical Dilemmas and Privacy Issues in Emerging Technologies: A Review. *Sensors*, *23*(3), Article 3. https://doi.org/10.3390/s23031151

Driver, J. (2009). *The History of Utilitarianism*. https://seop.illc.uva.nl/entries/utilitarianism-history/

Durant, A., Mcclure, F., Karunakaran, M., Anderson, L., & Nakato, G. (2022). *Artificial Intelligence is Transforming the Insurance Industry, Introducing Innovative Methods that Revolutionize the Buying Process for Customers*. https://www.researchgate.net/publication/386050902_artificial_intelligence_is_transforming_the_insurance_industry_introducing_innovative_methods_that_revolutionize_the_buying_process_for_customers

Eitel-Porter, R. (2021). Beyond the Promise: Implementing Ethical AI. *AI and Ethics*, *1*(1), 73–80. https://doi.org/10.1007/s43681-020-00011-6

Eling, M., Nuessle, D., & Staubli, J. (2022). The Impact of Artificial Intelligence Along the Insurance Value Chain and on the Insurability of Risks. *The Geneva Papers on*

Risk and Insurance - Issues and Practice, *47*(2), 205–241. https://doi.org/10.1057/s41288-020-00201-7

Fabris, A., Mishler, A., Gottardi, S., Carletti, M., Daicampi, M., Susto, G. A., & Silvello, G. (2021). Algorithmic Audit of Italian Car Insurance: Evidence of Unfairness in Access and Pricing. *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society*, 458–468. https://doi.org/10.1145/3461702.3462569

Fahmideh, M., Ahmad, A., Khan, A., Badshah, S., Liang, P., Waseem, M., Niazi, M., & Azeem Akbar, M. (2021). *Ethics of AI: A Systematic Literature Review of Principles and Challenges*. https://doi.org/10.48550/arXiv.2109.07906

Ferrell, O. C., Harrison, D. E., Ferrell, L. K., Ajjan, H., & Hochstein, B. W. (2024). A Theoretical Framework to Guide AI Ethical Decision Making. *AMS Review*, *14*(1), 53–67. https://doi.org/10.1007/s13162-024-00275-9

Figaredo, D. D., & Stoyanovich, J. (2023). Responsible AI Literacy: A Stakeholder-First Approach. *Big Data & Society*, *10*(2), 20539517231219958. https://doi.org/10.1177/20539517231219958

Filabi, A., & Duffy, S. (2021). *AI, Ethics and Life Insurance: Balancing Innovation With Access*. https://www.researchgate.net/publication/386082923_Balancing_Innovation_and_Ethics_in_AI

Gomber, P., Kauffman, R. J., Parker, C., & Weber, B. W. (2018). On the Fintech Revolution: Interpreting the Forces of Innovation, Disruption, and Transformation in Financial Services. *Journal of Management Information Systems*, *35*(1), 220–265. https://doi.org/10.1080/07421222.2018.1440766

Guan, H., Dong, L., & Zhao, A. (2022). Ethical Risk Factors and Mechanisms in Artificial Intelligence Decision Making. *Behavioral Sciences*, *12*(9), 343. https://doi.org/10.3390/bs12090343

Hadley, E., Blatecky, A., & Comfort, M. (2024). Investigating Algorithm Review Boards for Organizational Responsible Artificial Intelligence Governance. *AI and Ethics*. https://doi.org/10.1007/s43681-024-00574-8

Hargrave, M. (2024, April 30). *Overview of Insurtech & Its Impact on the Insurance Industry*. Investopedia. https://www.investopedia.com/terms/i/insurtech.asp

Hassan, Y. (2023). Governing Algorithms from the South: A Case Study of AI Development in Africa. *Ai & Society*, *38*(4), 1429–1442. https://doi.org/10.1007/s00146-022-01527-7

Holland, C. P., Mullins, M., & Cunneen, M. (2021). *Creating Ethics Guidelines for Artificial Intelligence (AI) and Big Data Analytics: The Case of the European Consumer Insurance Market* (SSRN Scholarly Paper 3808207). https://doi.org/10.2139/ssrn.3808207

Jansen, D. (2021, April 16). What Is Thematic Analysis? *Grad Coach*. https://gradcoach.com/what-is-thematic-analysis/

Keller, B. (2020). Promoting Responsible Artificial Intelligence in Insurance. *Artificial Intelligence*. https://www.genevaassociation.org/sites/default/files/ai_in_insurance_web_0.pdf

Konda, K. R., Koyya, V. S. R. U. R., & Voddi, V. K. R. (2024). Ethical Considerations in AI for Financial Services: Balancing Innovation with Regulatory Compliance. *International Journal of Communication Networks and Information Security (IJCNIS)*, *16*(2), 169–175.

Kumar, S., Arora, R., Rani, N., Mishra, D., & Ramkumar, M. (2022). AI Led Ethical Digital Transformation: Framework, Research and Managerial Implications. *Journal of Information, Communication & Ethics in Society*, *20*(2), 229–256. https://doi.org/10.1108/JICES-02-2021-0020

Li, Y., & Goel, S. (2024). Making It Possible for the Auditing of AI: A Systematic Review of AI Audits and AI Auditability. *Information Systems Frontiers*, 1–31. https://doi.org/10.1007/s10796-024-10508-8

Liu, B., Ding, M., Shaham, S., Rahayu, W., Farokhi, F., & Lin, Z. (2020). *When Machine Learning Meets Privacy: A Survey and Outlook* (arXiv:2011.11819). arXiv. https://doi.org/10.48550/arXiv.2011.11819

Mahajan, S., Agarwal, R., & Gupta, M. (2025). Algorithmic Bias Under the EU AI Act: Compliance Risk, Capital Strain, and Pricing Distortions in Life and Health Insurance Underwriting. *Risks*, *13*(9), 160. https://doi.org/10.3390/risks13090160

Martin, K., Shilton, K., & Smith, J. (2019). Business and the Ethical Implications of Technology: Introduction to the Symposium. *Journal of Business Ethics*, *160*(2), 307–317. https://doi.org/10.1007/s10551-019-04213-9

McCombes, S. (2019, May 8). *What Is a Case Study?* Scribbr. https://www.scribbr.com/methodology/case-study/

Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2021). A Survey on Bias and Fairness in Machine Learning. *ACM Computing Survey*, *54*(6), 115:1-115:35. https://doi.org/10.1145/3457607

Micewski, E. R., & Troy, C. (2007). Business Ethics – Deontologically Revisited. *Journal of Business Ethics*, *72*(1), 17–25. https://doi.org/10.1007/s10551-006-9152-z

Migdadi, M. K., Oweidat, I. A., Alosta, M. R., Al-Mugheed, K., Saeed Alabdullah, A. A., & Farghaly Abdelaliem, S. M. (2024). The Association of Artificial Intelligence Ethical Awareness, Attitudes, Anxiety, and Intention-to-Use Artificial Intelligence Technology Among Nursing Students. *Digital Health*, *10*, 20552076241301958. https://doi.org/10.1177/20552076241301958

Mohamed, S., Png, M.-T., & Isaac, W. (2020). Decolonial AI: Decolonial Theory as Sociotechnical Foresight in Artificial Intelligence. *Philosophy & Technology*, *33*(4), 659–684. https://doi.org/10.1007/s13347-020-00405-8

Mosqueira-Rey, E., Hernández-Pereira, E., Alonso-Ríos, D., Bobes-Bascarán, J., & Fernández-Leal, Á. (2023). Human-in-the-Loop Machine Learning: A State of the Art. *Artificial Intelligence Review*, *56*(4), 3005–3054. https://doi.org/10.1007/s10462-022-10246-w

Mullins, M., Holland, C. P., & Cunneen, M. (2021). Creating Ethics Guidelines for Artificial Intelligence and Big Data Analytics Customers: The Case of the Consumer European Insurance Market. *Patterns*, *2*(10), 100362. https://doi.org/10.1016/j.patter.2021.100362

Nikolopoulou, K. (2022a, August 11). *What Is Purposive Sampling?* Scribbr. https://www.scribbr.com/methodology/purposive-sampling/

Nikolopoulou, K. (2022b, August 17). *What Is Snowball Sampling?* Scribbr. https://www.scribbr.com/methodology/snowball-sampling/

Noordhoek, D. (2023). *Regulation of Artificial Intelligence in Insurance: Balancing Consumer Protection and Innovation*. https://www.genevaassociation.org/publication/public-policy-and-regulation/regulation-artificial-intelligence-insurance-balancing

Ntoutsi, E., Fafalios, P., Gadiraju, U., Iosifidis, V., Nejdl, W., Vidal, M.-E., Ruggieri, S., Turini, F., Papadopoulos, S., Krasanakis, E., Kompatsiaris, I., Kinder-Kurlanda, K., Wagner, C., Karimi, F., Fernandez, M., Alani, H., Berendt, B., Kruegel, T., Heinze, C., … Staab, S. (2020). Bias in Data-Driven Artificial Intelligence Systems—An Introductory Survey. *WIREs Data Mining and Knowledge Discovery*, *10*(3), e1356. https://doi.org/10.1002/widm.1356

Owens, E., Sheehan, B., Mullins, M., Cunneen, M., Ressel, J., & Castignani, G. (2022). Explainable Artificial Intelligence (XAI) in Insurance. *Risks*, *10*(12), Article 12. https://doi.org/10.3390/risks10120230

Pant, A., Hoda, R., Tantithamthavorn, C., & Turhan, B. (2024). *Ethics in AI through the Practitioner's View: A Grounded Theory Literature Review* (arXiv:2206.09514). arXiv. https://doi.org/10.48550/arXiv.2206.09514

Perez, L. (2023, March 19). *An Overview of the Financial Services Industry—SmartAsset | SmartAsset*. https://smartasset.com/financial-advisor/the-financial-services-industry-what-you-need-to-know

Poenaru-Olaru, L., Sallou, J., Cruz, L., Rellermeyer, J. S., & Van Deursen, A. (2023). Retrain AI Systems Responsibly! Use Sustainable Concept Drift Adaptation Techniques. *2023 IEEE/ACM 7th International Workshop on Green And Sustainable Software (GREENS)*, 17–18. https://doi.org/10.1109/GREENS59328.2023.00009

Pothuganti, S. (2018). Review on Over-Fitting and Under-Fitting Problems in Machine Learning and Solutions. *International Journal of Advanced Research in Electrical*

Electronics and Instrumentation Engineering, *7*, 3692–3695. https://doi.org/10.15662/IJAREEIE.2018.0709015

Roychowdhury, S., Alvarez, A., Moore, B., Krema, M., Gelpi, M. P., Rodriguez, F. M., Rodriguez, A., Cabrejas, J. R., Serrano, P. M., Agrawal, P., & Mukherjee, A. (2023). *Hallucination-Minimized Data-to-Answer Framework for Financial Decision-Makers* (arXiv:2311.07592). arXiv. https://doi.org/10.48550/arXiv.2311.07592

Santoni de Sio, F. (2024). Artificial Intelligence and the Future of Work: Mapping the Ethical Issues. *The Journal of Ethics*, *28*(3), 407–427. https://doi.org/10.1007/s10892-024-09493-6

Sarker, I. H. (2021). Machine Learning: Algorithms, Real-World Applications and Research Directions. *SN Computer Science*, *2*(3), 160. https://doi.org/10.1007/s42979-021-00592-x

Saunders, M., Lewis, P., Thornhill, A., & Bristow, A. (2019). Understanding Research Philosophy and Approaches to Theory Development. In *Research Methods for Business Students* (8th ed., pp. 128–171). Pearson Education.

Selbst, A. D., Boyd, D., Friedler, S. A., Venkatasubramanian, S., & Vertesi, J. (2019). Fairness and Abstraction in Sociotechnical Systems. *Proceedings of the Conference on Fairness, Accountability, and Transparency*, 59–68. https://doi.org/10.1145/3287560.3287598

Sele, D., & Chugunova, M. (2024). Putting a Human in the Loop: Increasing Uptake, but Decreasing Accuracy of Automated Decision-Making. *PLOS ONE*, *19*(2), e0298037. https://doi.org/10.1371/journal.pone.0298037

Shahriar, S., Allana, S., Hazratifard, S. M., & Dara, R. (2023). A Survey of Privacy Risks and Mitigation Strategies in the Artificial Intelligence Life Cycle. *IEEE Access*, *11*, 61829–61854. IEEE Access. https://doi.org/10.1109/ACCESS.2023.3287195

Shanmugam, D., Diaz, F., Shabanian, S., Finck, M., & Biega, A. (2022). Learning to Limit Data Collection via Scaling Laws: A Computational Interpretation for the Legal Principle of Data Minimization. *Proceedings of the 2022 ACM Conference on Fairness,* *Accountability, and Transparency*, 839–849. https://doi.org/10.1145/3531146.3533148

Sharma, A., & Sharma, R. (2024). Comparative Analysis of Data Protection Laws and AI Privacy Risks in BRICS Nations a Comprehensive Examination. *Global Journal of Comparative Law*, *13*, 56–85. https://doi.org/10.1163/2211906X-13010003

Singh, A. K., & Mishra, N. K. (2018). Ethical Theory & Business a Study Based on Utilitarianism and Kantianism. *International Journal of Humanities and Social Development Research*, 107–124. https://doi.org/10.30546/2523-4331.2018.2.1.107

Smith, L. T., & Bochanski, S. J. (2022). Avoiding Unfair Bias in Insurance Applications of AI Models. *Society of Actuaries*. https://www.soa.org/resources/research-reports/2022/avoid-unfair-bias-ai/

Stahl, B. C., Antoniou, J., Ryan, M., Macnish, K., & Jiya, T. (2022). Organisational Responses to the Ethical Issues of Artificial Intelligence. *AI & Society*, *37*(1), 23–37. https://doi.org/10.1007/s00146-021-01148-6

Thierer, A. D. (2023). *Flexible, Pro-Innovation Governance Strategies for Artificial Intelligence* (SSRN Scholarly Paper 4423897). Social Science Research Network. https://papers.ssrn.com/abstract=4423897

Tracy, S. (2010). Qualitative Quality: Eight “Big-Tent” Criteria for Excellent Qualitative Research. *Qualitative Inquiry*, *16*, 837–851. https://doi.org/10.1177/1077800410383121

Van Bekkum, M., Zuiderveen Borgesius, F., & Heskes, T. (2024). *AI, Insurance, Discrimination and Unfair Differentiation. An Overview and Research Agenda* (SSRN Scholarly Paper 4918863). Social Science Research Network. https://doi.org/10.2139/ssrn.4918863

Van Den Boom, F. (2021). Regulating Telematics Insurance. In P. Marano & K. Noussia (Eds.), *Insurance Distribution Directive: A Legal Analysis* (pp. 293–325). Springer International Publishing. https://doi.org/10.1007/978-3-030-52738-9_12

Vazquez-Zapien, G. J., Mata-Miranda, M. M., Garibay-Gonzalez, F., & Sanchez-Brito, M. (2022). Artificial Intelligence Model Validation Before Its Application in Clinical Diagnosis Assistance. *World Journal of Gastroenterology*, *28*(5), 602–604. https://doi.org/10.3748/wjg.v28.i5.602

Wang, X., & Wu, Y. C. (2024). Balancing Innovation and Regulation in the Age of Generative Artificial Intelligence. *Journal of Information Policy*. https://doi.org/10.5325/jinfopoli.14.2024.0012

Zanke, P., & Sontakke, D. (2021). *Artificial Intelligence Applications in Predictive Underwriting for Commercial Lines Insurance*. https://thesciencebrigade.com/adlt/article/view/181

Zarifis, A., Holland, C. P., & Milne, A. (2019). Evaluating the Impact of AI on Insurance: The Four Emerging AI and Data-Driven Business Models. *Emerald Open Research*, *1*(1). https://doi.org/10.1108/EOR-01-2023-0001

Zoshak, J., & Dew, K. (2021). Beyond Kant and Bentham: How Ethical Theories Are Being Used in Artificial Moral Agents. *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, 1–15. https://doi.org/10.1145/3411764.3445102

APPENDIX A: PARTICIPANT INFORMATION SHEET

Dear Sir / Madam

My name is Mark Mkorongo I am a Master's student in Digital Business at the University of the Witwatersrand, Johannesburg. My supervisor is Prof. Patrick Ndayizigamiye. I am conducting a research study about the Mitigating Ethical Risks of Using Artificial Intelligence in the Insurance Industry.

I am inviting you to take part in an interview. If you decide to take part, your participation in this research study will last about 45 to 60 minutes.

With your permission, I would like to audio record the interview. This data will be stored in a password-protected device and only the researcher will have access to the data.

The interview will be confidential and anonymous. When I share the results of the research study, I will not include your name or anything else that could identify you. With your permission, other researchers may use the data collected from this research study, but your name and any personal information will not be used or passed on.

If you decide to take part in the research study, it should be because you want to volunteer. You do not have to take part. You can stop being in the study at any time. You do not have to answer any questions if you do not want to. You will not get any direct benefits if you choose to join the research study. You will not lose any services, benefits or rights you would normally have if you decide not to join. Taking part in the research study will not cost you anything. You will not be paid for being in this research study.

This research study will be written up as a research report. The report will be available on the university library website. If you would like to receive a summary of this report, I will be happy to send it to you.

If you have any questions during or afterwards about this research study, feel free to contact me or my supervisor on the details listed below. If you have any concerns or complaints about the ethical procedures of this research study, you are welcome to contact the University Human Research Ethics Committee (Non-Medical), telephone +27(0) 11 717 1408, email hrecnon-medical@wits.ac.za.

Yours sincerely,

Researcher:

Mark Mkorongo, 2702380@students.wits.ac.za, +27 (0) 78 052 7267

Supervisor:

Patrick Ndayizigamiye, pndayizigamiye@uj.ac.za

APPENDIX B: PARTICIPANT CONSENT FORM

Title of project: Mitigating Ethical Risks of Using Artificial Intelligence in the Insurance Industry

Name of researcher: Mark Mkorongo

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

I agree to the following:

(Please circle the relevant options below)

The research study was explained to me. I understand what this study is about.	YES	NO
I understand that I can volunteer to take part in the study	YES	NO
I agree that the interview may be audio recorded	YES	NO
I agree that direct quotations from my interview may be used by the researcher in their research report	YES	NO
I agree that my participation will remain anonymous (my name or other identifying data will not be used by the researcher in their research report)	YES	NO
I agree that other researchers may use the information I provide in my interview (depending on their own ethics clearance being obtained) but my name and any personal information will not be used or passed on	YES	NO

…………………………………… (signature)

…………………………………… (name of participant)

…………………………………… (date)

…………………………………… (signature)

…………………………………… (name of researcher/person seeking consent)

…………………………………… (date)

APPENDIX C: DATA COLLECTION INSTRUMENT

INTERVIEW GUIDE QUESTIONS

RESEARCH TITLE: Mitigating Ethical Risks of Using Artificial Intelligence in the Insurance Industry

Date	
Time	
Interviewee	
Interviewer	

Introduction

Thank you for taking the time to participate in this study. We are working on a project about the ethical implications of AI-driven decision-making within the insurance industry. To this end, we wish to collect data regarding experiences and knowledge related to the use of AI for decision-making in the insurance and its ethical implications.

Section 1: General Understanding of AI in the Insurance Industry

Q1.1 Can you please tell me about your role and experience in the insurance industry?

Notes:

Q1.2 Could you please describe your experience or knowledge related to the use of AI in decision-making processes within the insurance industry?

Notes:

Q1.3 What are some of the common applications of AI in your company/industry?

Notes:

Q1.4 In your opinion, how has AI changed the decision-making processes in the insurance industry?

Notes:

Section 2: Ethical Risks Arising from AI in Decision-Making

Q2.1 What ethical risks do you perceive with the use of AI in decision-making within the insurance industry?

Notes:

Q2.2 Can you provide any examples where these risks have manifested?

Notes:

Q2.3 How do these ethical risks affect insurance customers and policyholders?

Notes:

Q2.4 Are there specific groups of customers that are more vulnerable to these risks?

Notes:

Section 3: Addressing Ethical Risks

Q3.1 What strategies are currently employed by your company/industry to address the ethical risks associated with AI?

Notes:

Q3.2 How effective do you believe these strategies are?

Notes:

Q3.3 What are some of the challenges your company/industry faces in mitigating these ethical risks?

Notes:

Q3.4 How do you overcome these challenges?

Notes:

Section 4: Effectiveness of Current Strategies

Q4.1 How does your company/industry evaluate the effectiveness of the strategies used to address ethical risks?

Notes:

Q4.2 What metrics or benchmarks can be used in this evaluation?

Notes:

Q4.3 What improvements do you think are needed in the current strategies to better address the ethical implications of AI in decision-making?

Notes:

Section 5: Future Strategies and Recommendations

Q5.1 What strategies do you believe should be implemented to ensure AI technologies are used ethically in the insurance industry?

Notes:

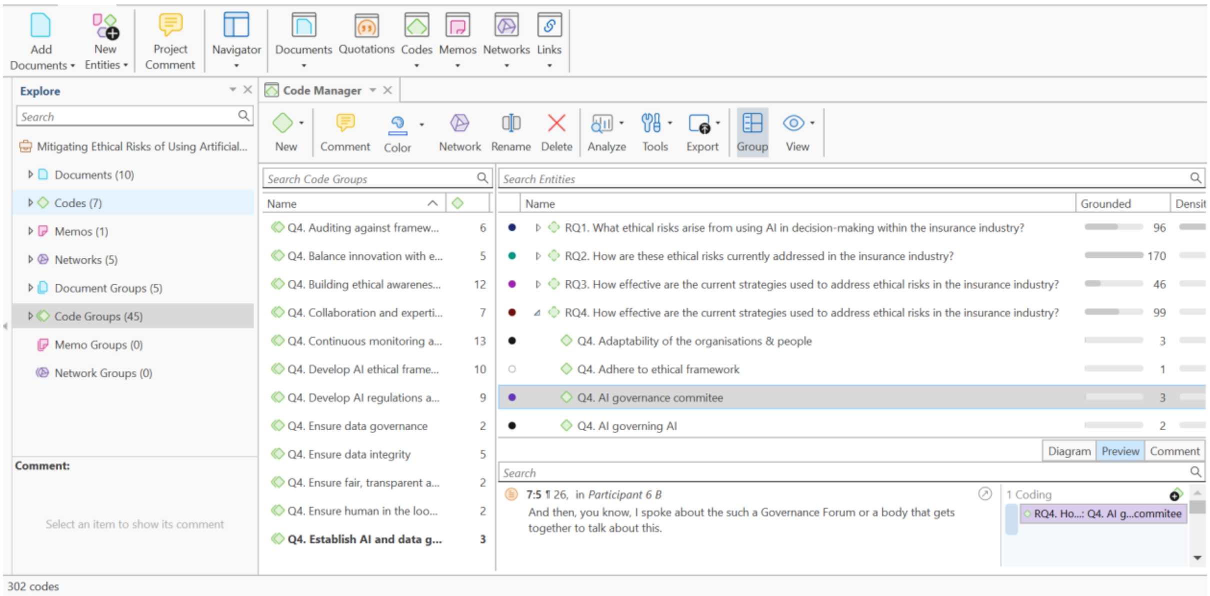
Q5.2 How can these strategies be effectively implemented and monitored?

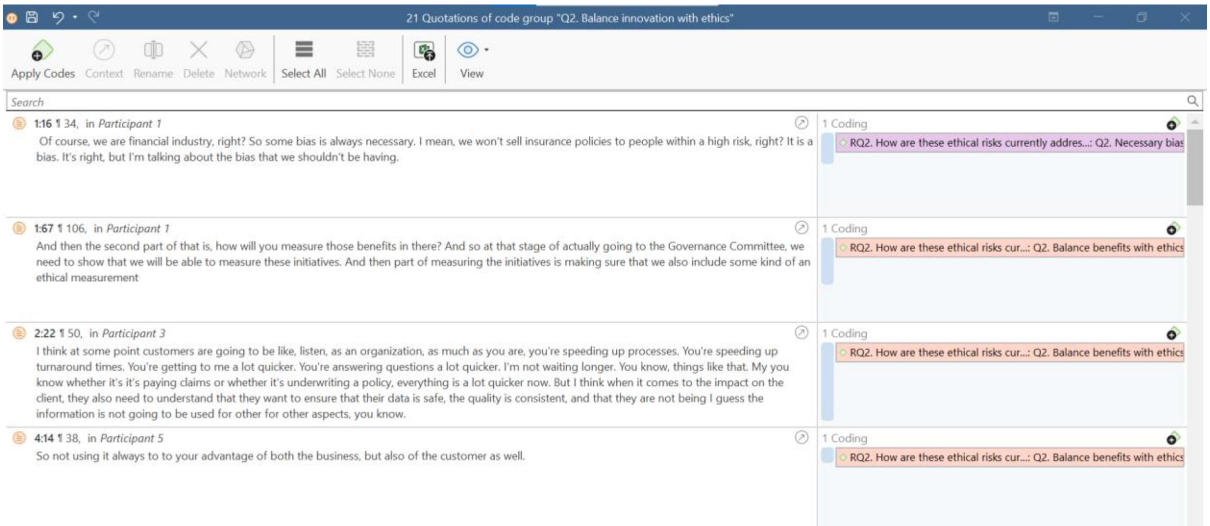
Notes:

Q5.3 Based on your experience, what recommendations would you provide to industry stakeholders to enhance the ethical use of AI in decision-making?

Notes:

APPENDIX D: QUALITATIVE CODING





APPENDIX E: ETHICAL CLEARANCE CERTIFICATE

Graduate School of Business Administration
University of the Witwatersrand, Johannesburg

Wits Business School

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

Ethics Clearance Certificate

Ethics protocol number:	WBS/DB2702380/510

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

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

Project title	Mitigating ethical risks of using artificial intelligence in the short-term insurance industry
Investigator / Researcher	Mr Mark Mkorongo
Nature of Project	MM (Digital Business)
Decision of the Committee	Approved, provided stakeholders and participants are guaranteed anonymity and confidentiality.
Issue Date of Certificate	02/09/2024
Expiry date	Date of submission of the project / research report
Chairperson	Dr Ayanda Magida ☎ +27 11 717 3953 ✉ ayanda.magida@wits.ac.za

Declaration by Researcher

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

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

	09/09/2024
Signature	Date:

APPENDIX F: PROOF OF EDITING LETTER

Athol Leach (Proofreading and Editing)

31 Park Rd
Fisherhaven
Hermanus 7200

Email: atholleach@gmail.com Cell: 0846667799

27 February 2025

To Whom It May Concern

This letter serves to confirm that I have edited the following research project report for the Degree of Master of Management in the field of Digital Business by **Mark Mkorongo** (2702380) titled:

"Mitigating Ethical Risks of Using Artificial Intelligence in the Insurance Industry"

The report was edited in terms of grammar, spelling, punctuation, and overall style. In doing so, use was made of MS Word's "Track changes" facility thus providing the student with the opportunity to reject or accept the changes made.

Please note that while I have checked the in-text references and those in the list of references for consistency of format (the latter as far as possible), I have not checked the veracity of the sources themselves.

The tracked document is on file.

Sincerely

Athol Leach
(MIS, Natal)