Managing AI Ethics Tensions: A Case Study at Multichoice

by

MERGANDRAN VELAYUDAN

A dissertation submitted in fulfilment of the requirements for the degree

M.A. (Philosophy)

in the Department of Philosophy at the

UNIVERSITY OF PRETORIA

FACULTY OF HUMANITIES

SUPERVISOR: PROF. E RUTTKAMP-BLOEM
CO-SUPERVISOR: DR H. ROBERTSON

MARCH 2023

© University of Pretoria

ACKNOWLEDGEMENT

I would like to acknowledge the many people who have played a role in assisting me throughout my studies.

- Professor Emma Ruttkamp-Bloem, my supervisor and guide through this process. Thank you for your unwavering support, your ideas and extensive knowledge in this field. This study reflects your commitment to building the field of AI ethics.
- Dr Helen Robertson, my second supervisor, thank you for the insightful comments, thought provoking suggestions and quick responses throughout this process. It's been a pleasure working with you.
- Mr Andries Masenge, Department of Statistics, University of Pretoria, thank you for inputs into the statistics presented.
- My remarkable wife, Dr Priya Rajaram, for being the amazing person that you are. You are my inspiration, my tower of strength, my north star and the source of so much that is good in my life. Thank you for building this life together with me, and for co-piloting our adventure with our incredible children, Heerania, Milind and Aarin.
- My mother, Rungamma Velayudan, and my late father, Arumogam Velayudan, for the sacrifices you've made over the years to get me this far. I am so grateful for your love and support.
- My family, from my siblings Vani Naicker, Anban Velayudan, Pridashan Velayudan and my late sister Nirmala Bissessur, to my brothers-in-law Rishan Bissessur and Mervyn Naicker and my sisters-in-law Anisha Rajkumar and Sharon Bissessur, and my nephews and nieces Akashni, Thamera, Riesh, Revina, Revaan and Satheesh, for being the stable platform from which everything else grows. I thank you for always being there for me.
- Finally, to my employer, Multichoice, and my work colleagues who contributed to this study and who supported me on this journey. Thank you for your willingness to be part of this process.

Sincerely,

Mergan Velayudan

© University of Pretoria

PLAGIARISM DECLARATION

DECLARATION

I, Mergandran Velayudan, declare that this dissertation is my own original work. Where secondary material is used, this has been carefully acknowledged and referenced in accordance with university requirements.

I understand what plagiarism is and am aware of university policy and implication in this regards.

Signature:

Mergandran Velayudan

Date: 21 March 2023

© University of Pretoria

ETHICS STATEMENT

I, Mergandran Velayudan, have obtained, for the research described in this work, the applicable research ethics approval.

Date of approval 04 August 2021
Research ethics number HUM005/0421

I, Mergandran Velayudan, declare that I have observed the ethical standards required in terms of the University of Pretoria's Code of Ethics for Researchers and Policy Guidelines for Responsible Research.

Signed

Mergandran Velayudan

Date: 21 March 2023

© University of Pretoria

ABSTRACT

Interest in Artificial Intelligence (AI) has never been greater – while we're at a likely tipping point in the adoption of AI into mainstream industry, we're still grappling with effective ways to manage the ethical concerns that AI surfaces on a regular basis. The purpose of this study is to shed some light into the types of AI applications being developed in the media industry in South Africa, investigate how AI ethics tensions surface and are managed when building these AI applications and provide recommendations for the management of AI ethics tensions in media organisations. Data was collected from respondents on the types of AI applications being developed, as well as the nature and characteristics of these projects including roles required to staff the projects, project duration, focus and business objectives, project outcomes, technologies used, and source of technologies used.

The study reviewed recent literature on AI ethics, and specifically research into the roles played by both individuals as well as the organisations they work for, in managing AI ethics considerations. Using the insights from the literature, as well as data collected during the study via a cross-sectional survey implementation, analysis was performed to determine associations between actions in the management of AI ethics tensions and the perceived outcomes and effectiveness. Several statistically significant associations, with both weak and medium effect size, were noted between the way AI ethics tensions were discovered and managed during projects, and the perceived outcomes and effectiveness of these actions. The associations noted potentially have implications for media organisations that are implementing AI solutions and are seeking to effectively manage AI ethics tensions.

Based on these analyses performed, recommendations are provided to inform the creation of effective frameworks to manage AI ethics tensions at media organisations developing AI solutions. Limitations of the study and further areas of research required are also discussed.

Key terms: artificial intelligence (AI), ethics, AI ethics, AI ethics tensions, cross-sectional survey, AI principles, AI ethics guidelines, AI ethics codes of conduct, AI solution development, ethical considerations

© University of Pretoria

TABLE OF CONTENTS

© University of Pretoria

© University of Pretoria

© University of Pretoria

© University of Pretoria

LIST OF TABLES

© University of Pretoria

© University of Pretoria

© University of Pretoria

LIST OF FIGURES

© University of Pretoria

LIST OF APPENDICES

© University of Pretoria

1. INTRODUCTION AND BACKGROUND

This study explores how artificial intelligence (AI) ethics considerations are incorporated into AI solution development at Multichoice, a media organisation in South Africa, with a specific view of exploring AI ethics tensions in recent AI solution development projects. The study will survey members of the AI community at Multichoice i.e., people who have contributed, either technically or non-technically, to recent AI solution development projects. The purpose of the study is to provide insights into the types of AI solutions being worked on in the media industry, including details such as the duration of these projects, how they are staffed, which technologies are used, the composition of project teams, the project objectives and focus as well as the outcomes of these projects. Furthermore, the prevalence of AI ethics tensions in such AI solution development projects will be explored, as well as individual and company roles in identifying and managing these AI ethics tensions. This study will only focus on recent AI solution development projects and is limited to respondents who currently work at the case organisation, Multichoice. Recommendations are made for the creation of a framework to effectively manage AI ethics tensions.

AI is a rapidly evolving field with the promise of remaking many aspects of our society, both for good as well as with potential negative consequences if not managed appropriately (Harari 2017). As a result of the widely acknowledged potential of AI solutions to introduce unintended consequences, a variety of organisations (public, private and NGOs) have taken the initiative to publish high level guidelines and principles which they believe will result in ethical implementations of AI that adhere to accepted societal norms. In a recent meta-analysis, Jobin, Ienca and Vayena (2019, p.389) found at least 84 unique sets of guidelines and principles and that number continues to increase on a monthly basis. While the number of published guidelines and principles continues to increase, these represent examples of soft law rather than actual legislation. Sossin and Smith (2003, p.869) note that soft law or “quasi-legislation” suffers from several limitations, including the lack of clarity about the intent and objectives of such guidelines as well as questionable ability to enforce such guidelines.

There is less effort being expended in understanding how to transform ideals into action compared with the effort being expended in defining guidelines and principles – the suspicion being that the publishing of the guidelines and principles are seen as the end goal by many organisations.

Mittelstadt (2019) notes that, aside from questions about the intent of AI ethics and guidelines and publications, their effectiveness should be brought into question. A principled approach is likely to be too far removed from day-to-day practice of AI development to be useful to practitioners and, with

© University of Pretoria

little to no legal framework to enforce the ethical guidelines and principles, it is unlikely that the situation will change soon if we continue with the current path of principles-focussed AI ethics.

Floridi (2019) notes that, given the amount of current interest in ethical considerations around digital technologies, it is the right time to start considering the "how" of digital ethics and not just the "what". He goes further and identifies five practical risk clusters that may arise when translating principles into practices, namely ethics shopping, ethics bluewashing, ethics lobbying, ethics dumping and ethics shirking. He suggests knowledge about these ethical risks reduces the probability of practitioners claiming ignorance and therefore limits the implementation risk to those based genuinely on "misunderstanding and misjudgements", not simple ignorance of digital ethical concerns.

To create a practical guide for machine learning engineers, designers and developers to help address the relatively large gap between "what" and "how", Morley et al (2019) creates an 'applied ethical AI typology' that focuses on the machine learning solution development process. The authors' framework maps tools available to the stages of the development process. While this is a promising start to the process of creating a comprehensive tool registry, the authors noted that there were already some limitations, including an unbalanced representation across AI ethics considerations (too much focus on explicability) and a general lack of usability (high skill level required to make use of the tools in many instances).

Although AI is accepted to be driving more and more innovations in the technology industry, there remains relatively few documented use cases which detail the specific nature of the application of AI within products – the concern being that many products claim to incorporate AI to leverage the hype in the market rather than to genuinely benefit users as a result of the incorporation of AI capabilities.

In the current study, AI solution development projects in a media organisation in South Africa will be explored. Details of the types of projects undertaken, which technologies are used, the staffing of these projects and the outcomes of the projects will be investigated. This study will contribute to the understanding of the types of AI solutions being developed today in a large multi-national organisation in the media industry.

In recent research, Whittlestone et al (2019a, p13) argue that exploring the tensions that arise when putting the AI ethics principles into practice could be more beneficial than simply focusing on AI ethics principles alone. This perspective takes into account the overlap between competing values that are described in high level principles or guidelines statements, as well as vague explanations of what these high-level values mean in practical applications. Accepting the notion that a situation presents as an ethical problem when there is no clear right and wrong answer also means that we need to accept

© University of Pretoria

that practitioners at various levels within AI solution development are faced with difficult trade-offs that need to be made – understanding some of these key tensions, or trade-offs, and how they are managed in organisations provides further insights into the current state of practice of AI ethics considerations in AI solution development.

Whittlestone et al (2019b, p199) identify four key tensions (quality of service versus autonomy; accuracy versus fairness; personalisation versus citizenship and automation versus dignity) that are likely to arise in AI applications but suggests that more work is required to determine exactly which tensions arise in each type of AI application being developed.

In the present study, practitioners forming part of the AI community at the case organisation, Multichoice, are surveyed to explore the extent to which the above four pairs of AI ethics tensions were present in recently complete AI solution development projects. In cases where respondents note the occurrence of AI ethics tensions, information is gathered on how these tensions were identified, how they were managed and what the outcome of managing these tensions was. This study contributes to a significant gap in the current literature, where no known studies exist that explore the prevalence, management and outcomes of managing AI ethics tensions in AI solution development projects.

© University of Pretoria

1.1 PROBLEM STATEMENT

Work is needed to determine the actual types of AI applications that are being deployed across various industries, whether AI ethics tensions are prevalent in AI solutions being developed now, and how AI ethics tensions are managed within organisations. In the current research study, some of these dimensions will be investigated in a case organisation, Multichoice. Multichoice represents a large corporate in the media industry in South Africa, with a significant portfolio of existing and completed AI solution development projects.

1.1.1 Research aim

The aim of this research is to identify and characterize the types of AI applications that are being developed within Multichoice as well as to identify and investigate the ethical tensions that arise while developing such AI applications. It is also intended to critically review and reflect on the individual versus organisational responsibility for implementing AI ethics principles within the context of AI solution development. Based on investigation of the situation in Multichoice, the research will aim to identify recommendations for the management of AI ethics tensions, thereby informing the creation of a framework to manage such AI ethics tensions.

1.1.2 Research objectives

1. To construct a survey which will identify:
    a. the types of AI applications that are being developed or deployed in a South African media organisation, including characteristics such as duration of projects, technologies used, staffing characteristics, project focus and project outcomes.
    b. Whether and which AI ethics tensions are present in AI solutions being developed at the case organisation.
    c. the nature and extent of support the organisation provides to practitioners to manage such AI ethics tensions, and the perceptions of practitioners on the nature and sufficiency of the support provided, as well as the impact of managing AI ethics tensions.
2. To critically review and reflect on the notion of individual versus organisational responsibility for implementing AI ethics principles.
3. To suggest recommendations for a framework to assist practitioners to manage these tensions.

© University of Pretoria

1.1.3 Research questions

1. What are the types of AI applications being worked on in a South African media company?
2. Which, if any, AI ethics tensions are present during the development of such AI applications?
3. How do individuals manage AI ethics tensions in current AI applications?
4. Can we propose recommendations for an effective framework to manage AI ethics tensions?

1.2 CHAPTER OUTLINES

The research report is presented according to the following chapter format.

Chapter one provides an orientation to the study and introduces the problem statement, research aims, research objectives and research questions. It outlines the contents of each of the chapters in the research report and describes key terminology and abbreviations used in the study.

Chapter two is the literature review. This chapter reviews the main body of literature related to the history and evolution of AI ethics related to the study, the management of AI ethics today and the focus on AI ethics tensions.

Chapter three discusses and presents the methods and procedures used in this study to answer the research questions. This includes determining the types of AI applications being developed today, the prevalence of AI ethics tensions in the development of AI applications and how these AI ethics tensions are being managed today.

This chapter includes a presentation of the research design selected, the research setting, discussions on reliability, validity, sampling and bias, participant selection criteria, recruitment of participants, ethical considerations, materials, data collection, data storage, data extraction and a discussion of the variables explored in this study.

© University of Pretoria

Chapter four presents the results of the study. The relevant results are presented, together with insights derived from the results. This chapter focuses on answering each of the research questions and achieving the main aims of the study.

Chapter five presents the findings from the study, both from the literature review as well as from the data analysis, as well as the conclusions, recommendations, limitations and areas of further study.

1.4 ABBREVIATIONS

Multichoice Group	MCG
Artificial Intelligence	AI
Machine Learning	ML
Research Question One	RQ1
Research Question Two	RQ2
Research Question Three	RQ3
Research Question Four	RQ4

1.5 SUMMARY

This chapter provided an introduction and background to the study. It served to introduce the rest of the research report. A list of terminology and a list of relevant abbreviations was provided. The following chapter will present literature relevant to this study and will position the current study relative to gaps in the existing literature.

© University of Pretoria

2. LITERATURE REVIEW

2.1 SCOPE OF THE CHAPTER

This chapter presents a review of the literature pertinent to this study, following the flow depicted in Figure 1.

Figure 1: Scope and flow of literature review

© University of Pretoria

The chapter begins by defining both artificial intelligence as well as the field of AI ethics. The subsequent sections discuss ethical considerations in new technologies as well as interest specifically in AI ethics. A discussion of codes of conduct, shortcomings of codes of conduct as well as the guidelines-based approach to AI ethics is included next. The impacts and limitations of codes of conducts is discussed before the focus shifts to the practical management of AI ethics. Some recent examples of practical guides in this area as well as the notion of AI ethics tensions are discussed next. The chapter concludes with some discussion on organizational versus individual responsibility for managing AI ethics considerations.

The current research study intends to assess a company's approach to the management of AI ethics tensions, and to propose recommendations that would inform the creation of a framework that can be used by practitioners to evaluate ethical considerations arising during the development of AI products - with a view towards mitigating potential ethical risks. The study will also undertake a critical reflection on the role of the individual versus organisational responsibility for implementing AI ethics principles in AI solutions.

2.2 DEFINITION AND HISTORY OF AI ETHICS

2.2.1 Definition of artificial intelligence

According to Müller's (2020) broadly encompassing definition, artificial intelligence is understood to refer to any computational system that shows intelligent behaviour. Interestingly, it suggests discarding the notion that we limit intelligence to human-like actions, which means that artificial intelligence, by this definition, includes both the performance of narrow-scope technical activities as well as wide-ranging "human" intelligence.

According to Bringsjord and Govindarajulu (2020), in their Stanford Encyclopedia of Philosophy Archive article, it is difficult and possibly even impossible, to attain a consensus view of a definition of artificial intelligence. The authors firstly suggest that the best we can do is to share proposed definitions of artificial intelligence and thereafter proceed to lean heavily on the definitions provided by Russell and Norvig in their widely popular text entitled "Artificial Intelligence: A Modern Approach" (Russell et al, 2010). According to Bringsjord and Govindarajulu (2020), AI may be defined in terms of goals falling along two dimensions. The first dimension is whether the goal is to match either human

© University of Pretoria

performance or an ideal rational performance while the second dimension relates to whether the system is built to reason or to act.

2.2.2 Definition of AI ethics

The ethics of AI is a field within applied ethics (Müller, 2020) that deals with the concerns that have become apparent as a result of the increasing adoption of artificial intelligence technologies in a wide range of applications, and where no clear answers exist to these concerns. The author suggests that we can divide these ethical issues, into considerations revolving around AI as objects (when we create AI systems for our use as tools) and subjects (when the AI itself can carry out actions independently as is the case of robots for example). The essence of this research study deals with the former consideration, where we are dealing with AI as an object created by and for humans to use.

2.2.3 Ethical considerations in new technologies

Although the explosion of interest in AI and consequently ethics of AI is relatively new, the challenges of understanding the ethical impact of new technologies is less so. Prior to the focus on AI ethics, the broader field of computer ethics enjoyed prominence as computers looked to change our ways of living. Moor (1985) suggests that problems in computer ethics arise precisely because of the newness of the applications of computer technology – as he puts it, novel applications of computer technologies create policy vacuums concerning how the technology should be used. And because computers possess “logical malleability”, the ability to be used for an almost limitless variety of applications, there will be an ongoing process of discovering new applications, which unearth new policy gaps that need to be addressed. The parallels between his concept of the “Computing Revolution” and the “AI Revolution” (Harari 2017), which promises to change the very fabric of society, are quite compelling.

Moor (1985) identifies the lack of transparency, or invisibility as he termed it, in computing technology as the source of ethical concerns. He further segments this invisibility into three elements namely invisible abuse (intentional unethical conduct), invisible programming values (bias arising from the specific perspective or value system of the programmer) and invisible complex calculations (computers performing calculations which are beyond human comprehension and therefore cannot

© University of Pretoria

be validated). Interestingly, all three of these invisibilities have already arisen in the field of artificial intelligence and are the subject of ongoing ethical concerns.

Johnson (2003) suggests that there are at least three ways to categorize ethical issues: according to the technology, according to the sector or according to ethical themes. In expanding on the approach of categorization using ethical themes, she identified some emerging issues which we are still grappling with today, namely ethics for computer professionals (we have multiple ethical guidelines but are they equivalent to law or merely ideals to strive for in a complex environment with competing priorities?), privacy (what data do we collect about people, how do we use it and how do we share it?), abuse (how do abuses in the cyber world compare to abuses in other contexts, and how should the punishments reflect the differences if any?) and internet issues (a catch all phrase referring to the question of whether the internet should be treated as distinctive from other contexts from an ethical perspective). As we shall see in this study, although much effort has been expended on research into emerging ethical issues, little in the way of concrete and unilateral answers have been arrived at.

2.2.4 Interest in AI ethics

Although there is a plethora of AI ethics guideline documents available presently, the explosion of interest in the study of AI ethics is relatively new and still topical. According to Wolf (2021), there has been a significant ramp-up in the interest in AI ethics since 2018 with the number of published studies only reaching significant volumes over the past few years since then. Table 1, reproduced from Wolf (2021) shows the count of Google Scholar citations with ("AI or "Artificial Intelligence") and ("ethics" or "ethical") in the title between 1985 and 2020:

Year	Count	Year	Count	Year	Count	Year	Count
1985	1	1994	0	2003	4	2012	7
1986	1	1995	0	2004	3	2013	5
1987	0	1996	0	2005	6	2014	12
1988	0	1997	0	2006	2	2015	10
1989	0	1998	3	2007	2	2016	21
1990	0	1999	1	2008	6	2017	45
1991	1	2000	4	2009	2	2018	128
1992	0	2001	0	2010	2	2019	334
1993	0	2002	1	2011	8	2020	342

Table 1:Google Scholar citations covering AI ethics keywords in the title between 1985 and 2020

© University of Pretoria

2.3 GUIDELINES AND PRINCIPLES TO MANAGE AI ETHICS

The past few years have seen a rapid growth in the number of guidelines and principles for AI ethics deployment. These have been published by organisations ranging from private organisations, NGOs, governments, research institutions as well as international bodies. In a recent meta-analysis, Jobin, Ienca and Vayena (2019, p.389) found at least 84 unique sets of guidelines and principles and noted that more than 80% of the documents were released after 2016. The authors found that, while the plethora of guideline documents indicate strong interest in AI ethics from both the private and public sectors, it is concerning that some regions (Africa among them) are not equal participants in the AI ethics debate.

The authors noted the difficulty in defining AI ethics uniformly by observing that no single principle emerged across all documents in the corpus researched. Five principles appeared in more than half of the sources consulted and, it is argued by the authors, these five principles represent a high-level consensus view of the AI ethics principles of most concern. The five principles are:

1. Transparency - refers to initiatives to improve explainability, interpretability or disclosure concerning AI initiatives.
2. Justice – is mentioned mostly in relation to fairness and the removal of bias or discrimination.
3. Non-maleficence – this refers to the avoidance of potential risks or harms e.g., hacking.
4. Responsibility – this refers to ensuring that AI products act with integrity and ensure that responsibility is clearly articulated (often interchanged with accountability).
5. Privacy – refers often to both privacy as a value as well as a right. In most contexts, this deals with data protection and data security.

It is further noted that, although there appears to be convergence towards these five principles, there is significant divergence in terms of relative importance, interpretation and links to actual implementation. There is a lack of clarity regarding how to resolve conflicts or tensions between ethical principles.

Fjeld et al (2020) studied 36 particularly visible and influential documents presenting AI principles. The authors found that, while these documents address a similar overall purpose, there is a significant variation in terms of intended audience, depth of content, scope and composition. Fjeld et al (2020) notes that this is based on, in part, the fact that these documents are authored by a wide variety of actors representing a diverse set of stakeholders and perspective. The authors note, however, that there appear to be eight themes that predominate across the documents, namely privacy,

© University of Pretoria

accountability, safety and security, transparency and explainability, fairness and non-discrimination, human control of technology, professional responsibility and promotion of human values. There is a further note in that study of the appearance that more recently released AI principles documents tend to incorporate all eight themes, pointing to an emergence of norms in the AI ethics field, at least among the included communities responsible for the creation of these documents. A caveat is that the documents covered by the study do not provide universal geographic coverage, with Africa being notable for its absence of any direct contribution to the documents reported on.

The current study will focus on the implementation phase of AI solution development and an element of the study will explore the extent to which the case organisation provides practical support to practitioners to navigate through competing ethical values as they arise during implementation.

Canca (2020) notes that there should be a distinction between core principles and instrumental principles. The author explains that core principles are derived from the widely used core principles of applied ethics, namely autonomy, beneficence and justice – these principles should be thought of as intrinsically valuable and are not derived from anything else. Instrumental principles should be considered to support the core principles and are only valuable to the extent that they help to uphold the core principles. The author contends that being able to identify core principles will make it easier to recognize whether an ethical challenge is being presented.

Canca (2020) further notes that when categorizing published AI principles into the core principles of autonomy, beneficence and justice, there is a relatively consistent picture of weighting being given to each of them across both regions as well as industries. The authors also state that AI principles are meant to provide a starting point for ethics review and will also help to ensure that specific ethical concerns are not overlooked – however, they are not meant to be a full decision-making system that removes the need for in-depth analysis of especially complex scenarios.

An interesting observation from Canca (2020) was the contention that there is still value in organisations setting up their own AI principles, despite the proliferation of existing guidelines and principles already in the public domain. The contention here is that organizations will be able to provide guidance on how to prioritise instrumental principles when core principles are in conflict.

© University of Pretoria

2.4 PROFESSIONAL CODES OF CONDUCT

Schwartz (2001) notes that various terminology such as codes of ethics, codes of conduct, mission statements and value statements have been used to refer to documents that provide guidelines for employee behaviour in the workplace. The author suggests that an accepted definition of a code of conduct is a "written, distinct, and formal document which consists of moral standards used to guide employee or corporate behaviour". Schwartz (2001) underlines the prevalence of such codes of conduct by citing that over 90% of large US corporations as well as 86% of large Canadian corporations had codes of conduct in place at the time of the study. Schwartz (2001) further states that codes of conduct serve the following purposes:

- Provide consistent standards for employees,
- Avoid legal consequences,
- Promote public image.

Schwartz (2001) found that employees did not comply with codes of conduct because of self-interest, dissatisfaction, environment, company priorities, company interests and ignorance while they complied with codes of conduct because of personal values, fear and loyalty to the company.

When looking into how codes influence behaviour, Schwartz (2001) introduced eight metaphors describing the ways in which codes of conduct influence behaviour:

- As a rulebook, to make explicit what constitutes acceptable behaviour.
- As a signpost, to give direction on who or what to consult.
- As a mirror, to confirm what is acceptable or not.
- As a magnifying glass, to indicate when caution may be required.
- As a shield, to resist unethical influences.
- As a smoke detector, to help warn others about unacceptable behaviour.
- As a fire alarm, to report issues.
- As a club, to enforce compliance is the face of potential sanctions.

© University of Pretoria

2.5 SHORTCOMINGS OF GUIDELINES AND CODES OF CONDUCT

In a comprehensive evaluation of the most prominent AI ethics guidelines published within the past five years, Hagendorff (2020) highlighted several significant shortcomings of the guidelines-based approach to AI ethics:

- Omissions: while the guidelines generally cover accountability, privacy and fairness, they mostly fail to cover issues of malevolent AI, political abuse of AI, reducing social cohesion, lack of diversity, social and ecological costs of AI among other issues.
- Ineffectiveness: according to empirical evidence, practitioners do not change their behaviour noticeably after reading AI ethics guidelines bringing into question the usefulness or impact of having guidelines at all
- Lack of enforcement mechanism: ethical guidelines and codes of practice mostly take the form of non-binding ideals that practitioners need to keep in mind rather than actual, enforceable regulations. It is up to individuals and organisations to balance speed against ethical considerations with the outcome that ethics is often an afterthought.
- Lack of clear accountability: with it generally being the case that no specific individual has final accountability for AI ethics decisions in AI product development, as well as a general lack of awareness of the bigger-picture consequences of technical decisions on society at large, AI practitioners do not appear to understand or take seriously the moral implications of work performed.
- Societal benefit is not the primary reason for implementing AI systems: unlike other fields such as medicine or education, the primary reason for implementing AI systems is economic benefit – therefore, in decision making contexts, any other considerations are easily cast aside in favour of economic benefit.
- Practitioners need to interpret exactly what is meant by each of the principles that should be adhered to, because of tenuous links between high level guidelines and actual practice.

In the same paper, the author suggests at least two remedies which will be explored further during this study:

- Introduce more technical explanations into the guidelines so that it becomes clearer what a specific principle or ideal means in terms of its impact on an AI implementation.
- Create a closer link between ethics and science research communities so that AI ethics can succeed in the “real world” of AI implementation.

© University of Pretoria

Mittelstadt (2019) noted that recent research suggested that AI ethics was converging on a set of principles that in some ways resembled the more established field of medical ethics. At first glance, this seems like a positive development since medical ethics are among the most well-known professional ethical principles. However, the author argues that principles may be insufficient to ensure AI ethics, citing the differences between the medical field and the AI field for this contention. These reasons are lack of professional norms (while medicine has a long history and a generally accepted view of what a practitioner should and shouldn't do, AI developers can come from a variety of educational and cultural background with limited common understanding of what best practices look like), lack of fiduciary responsibilities (medicine has a broad common aim of better patient health while the same commonality of purpose is absent from AI development), absence of ways to translate the principles into practice (as mentioned earlier in this section by Hagendorff (2020)) and a lack of legal accountability in the field of AI ethics, when compared with medical ethics (as a consequence of disparate self-regulation attempts from various independent sources in the field of AI ethics).

According to (Gogoll et al, 2021), one of the features of codes of conducts (CoCs) or guidelines is that they present values from a specific perspective – the authors identify the originator of the CoC, the addressed product and the target group of the CoC as being the three elements of perspectives that CoCs may pertain to and therefore reflect specific points of view or biases. In this vein, the authors' further point out that the existence of these perspectives also mean that analysing these CoCs would not be complete without also analysing the perspectives from which they are authored. Furthermore, the authors provide examples of how the same value could be interpreted differently and with differing emphasis, based on the perspective of the originator of the CoC.

Gogoll et al (2021) point out that the majority of CoCs agree on the core values yet differ on the details both in terms of definition of concepts as well as application specifics. The authors suggest that, because CoCs are based on values, and because these values are by their very nature high level and therefore underdetermined, a critical failing of CoCs is that they cannot be applied in practice. The authors further point out that CoC are usually overloaded with numerous different values, which often overlap, and practitioners are often able to "find" a convenient set of values within a specific CoC to fit their implementation into, irrespective of the nature of the practitioner's actions.

© University of Pretoria

Gogoll et al (2021) notes that the lack of detail in CoCs potentially leads to four different negative behaviours:

1. Cherry picking – in many instances, it is necessary to choose between competing values, with no clear guidelines on exactly what the boundary value is at which practitioners should choose one value over the other. The choice of one value over the other then becomes an arbitrary choice, left in the hands of the practitioner – but with enough justification to defend almost any position taken.
2. Risk of indifference – with broad sets of values and with the possibility that any action can be justified by even potentially contradictory values, there is a risk that practitioners may feel that they are allowed free rein to determine the outcome of any ethical choices based purely on their own intuition.
3. Ex-post orientation – due to CoCs not having the ability to be distilled into practical action that can be inserted into the development process, there is a risk that ethical deliberations only taken place once a project is completed and are concluded to fit the already-developed solution rather than influence the nature of the solution itself.
4. The desire for gut feelings – having a CoC may be as a result of a desire to reduce the ethical deliberation process into a heuristic, which is essentially a mental shortcut where we apply knowledge about similar situations from the past to deal with the current situation. The problem with applying heuristics in the moral domain is that previous situations may not fit exactly to the current situation and previously relied upon heuristics may therefore be ill-fitting to the current situation under analysis.

2.6 IMPACT OF CODES OF CONDUCT

While it is generally accepted that codes of conduct have become much more prevalent in corporate governance structures, the research community is less equivocal about the benefits of such codes of conduct. Schwartz summarized the results of 19 studies and reported that eight studies found that codes of conduct were effective, two studies found weak relationships between codes of conduct and behaviour and nine studies found that there was no significant relationship between codes of conduct and behaviour.

© University of Pretoria

The ACM's Code of Ethics is one of the more widely recognized code of ethics in existence for software development practitioners – in a widely cited study, McNamara et al (2018) undertook a behavioural study to measure whether software engineering students and professionals altered their decision making, based on an explicit request to consider the ACM Code of Ethics in making decisions. The study found that there was no evidence to support the contention that the Code of Ethics influences decision making in the software engineering context.

2.7 EVOLUTION TOWARDS MANAGING AI ETHICS IN PRACTICE

Floridi (2019) notes that it is the right time to focus on the "how" of digital ethics, now that there is a plethora of guidelines that cover the "what" of digital ethics and there is the seeming convergence towards common concepts if not a common vocabulary. He expands on the transition from high-level principles to detailed, practical guidance by introducing five specific clusters of digital ethics risks that need to be understood and mitigated. These are:

- Digital ethics shopping – since there are a multitude of AI ethics guidelines and mostly in the realm of soft law, this is a practice of retrofitting suitable principles ("mixing and matching" onto an existing process or behaviour to justify choices already made instead of making a genuine attempt to work towards a specific and public ethical standard
- Ethics bluewashing – implementing marketing tactics to give the impression that one is a lot more ethically responsible than is really the case, including setting up ethics boards and committees with no real authority, all while keeping behaviours unchanged.
- Ethics lobbying – publishing and pushing self-regulation guidelines in attempts to delay actual legislation that would create enforceable ethical principles and practices.
- Ethics dumping - exporting research activities from jurisdictions where there is relatively stronger legislation in place to jurisdictions where there is relatively lax legislation in place, with the intention of carrying out research that would be considered unacceptable in the original context.
- Ethics shirking – doing less ethical work if the perceived return from that work is lower in a given context.

© University of Pretoria

Floridi (2019) also notes that making practitioners aware of these ethical risks could mitigate the potential of such risks materializing – this is because, in his opinion, this removes the "ignorance defence" and leaves only the possibility of genuine misunderstanding and poor judgements as reasons for implementing ethically questionable solutions.

2.8 PRACTICAL STEPS TO MANAGING AI ETHICS

In a recent study, Morley et al (2019) set out to create a practical guide to assist machine learning developers to apply AI ethics considerations at every stage of the machine learning solution development process. Noting that there is still a significant gap between the "what" and the "how" of AI ethics, the authors constructed a framework that ties specific AI ethics tools to different stages of the machine learning solution development process. In doing so, the stated aim of the researchers was to create a reference work (not necessarily claiming completeness) to 1) encourage the move from principles to practices and 2) highlight the areas where more research is still required.

The authors noted that the initial results suggested:

- There is too much focus on explicability, at the expense of other AI ethics considerations. Furthermore, many of the tools available for explicability are available in the testing phase which suggests looking to explain as-built rather than a priori.
- There are few tools available to assess the impact of machine learning solutions on individuals and, fewer still are available to assess the impact on broader society. The link between a predicted outcome and a decision need not be automatic – it is still possible for people to translate the outcome of a predictive model into a decision and helping machine learning developers to build solutions that allow users more freedom to choose how predictions are used should enjoy increased focus.
- Although there is already a plethora of tools, there is also a general lack of usability associated with these tools. Whether because of poor documentation or a lack of real-world testing, the result is that these tools are used less than they should be.

Morley et al (2019) further suggests that a multi-disciplinary approach to AI ethics is likely to be beneficial on several fronts, including encouraging a more dynamic approach to ethics in line with machine learning system development processes as well as reducing the risks of designing in ethical

© University of Pretoria

principles but not actually implementing them in the final product. Among the research focus areas noted, the evaluation of tools currently available and how they actually perform in practice was highlighted as well as the creation of a knowledge base of how tools have been used at various stages of the machine learning development process and what impact they had on the application of AI ethics principles.

Mittelstadt (2019) further suggests that a top-down approach to AI Ethics (using principles and guidelines) should be complemented by case-studies of production AI systems, while also suggesting that the focus should shift from individual professional codes of AI ethics to organisational AI ethics where failures are attributed to institutions rather than individuals. Both observations related to AI ethics will be further explored in the current research study – RQ1 will provide insights into the types of AI applications being developed in a media company in South Africa while RQ3 will explore the extent to which organisational support is extended to individuals to manage AI ethics concerns.

2.8.1 Involvement of different levels of decisions makers

Gogoll et al (2021) highlight that there are multiple domains or levels of ethical decision making in the software development process (of which artificial intelligence solution development is a sub-set) and lists four levels of scope or domains of ethical decision making in an organisation:

- Politics – Is society interested in the solution and is it legal?
- Strategy – Is there a business case for the solution?
- Product design – What specific features should the solution provide?
- Development – how should the product features be implemented?

The authors point out that practitioners are at the end of a chain of decisions and, in that regard, are at the mercy of decisions made higher up in the value chain. The decisions whether a product is legal to build or not lies with the business decision makers and not with the system developers.

2.8.2 Involvement of other disciplines outside of computer science and engineering

In their opinion piece on the most important emerging AI ethics issues, Wolf et al (2021) note that the first issue of concern is which decision makers have a “seat at the table” when AI design decisions are being made. The authors point out that typical AI design decisions are made by people who have a background in computer science and engineering. Although useful for product development, Wolf et

© University of Pretoria

al (2021) suggest that these backgrounds do not adequately prepare people for the wide scope of decisions that need to be made to ensure that AI systems do not have wider negative societal impacts than designed for. As part of the research study, participants will provide input into both the composition of AI solution development teams as well as the levels at which decisions are made within AI solution development projects. Wolf et al (2021) suggest that, in addition to computer science and engineering professionals being involved in AI design decisions, there should be representation from other disciplines such as sociology, law and economics to provide valuable and different perspectives that may broaden ethical considerations.

While discussing the topics of fairness and equality in ML, Mittelstadt et al (2022) note that fairness in ML is a very active area of research. However, the authors also note that most of the effort spent on improving fairness of algorithms is being worked on in isolation, often without considering theories of equality and distributive justice that already exist in the fields of philosophy, politics and economics. The risk of this approach is that complex philosophical concepts are simplified into basic algorithmic implementation, which discard many of the nuances of the concepts that should be considered. In "sticking to what they know", computer scientists and engineers run the risk of limiting their considerations to matters within their control, such as performance and outcomes of algorithms, instead of looking at the broader context within which their solutions are meant to operate. The suggestion from Mittelstadt et al (2022) is that AI developers and deployers need to engage with "messy" details outside of their fields of expertise, so that better overall solutions are built in the quest for fair and equal AI implementations.

2.8.3 Introduction of ethical deliberation

In highlighting some of the challenges with and shortcomings of codes of conduct, Gogoll et al (2021) suggest that we should move away from using codes of conduct and should rather introduce the concept of ethical deliberation into the software development process. The authors raise the concern that ethical considerations should not be an afterthought in the development process, but that these considerations should be part and parcel of the development process. Their contention is that ethical deliberation should be practised by the team developing solutions and should not be practised by committees or councils that are not close to the actual development of the solution. Gogoll et all (2021) make the important point that software developers should work within the area of competence and should inform relevant decision makers of ethical issues that are outside of their scope. It is interesting to note that this will enable the "fire alarm" metaphor, which was noted by Schwatrz (2021) when discussing the potential benefits of codes of conduct.

© University of Pretoria

Gogoll et al (2021) note that introducing ethical deliberation into the software development process will encounter some challenges such as potential deceleration of time to market, questions regarding necessity and uncertainty over exactly where in the development process to implement such deliberations. On this positive side, the authors note that a product based on ethical deliberation may lead to better customer appreciation and loyalty and aligns well with the agile software development principle of reducing costs by preventing defects rather than fixing them.

2.8.4 Introduction of checklists to broaden ethical considerations

Canca (2020) made the point that complex ethical issues cannot be solved by merely relying on principles and guidelines. The author developed a tool called “The Box” that is interactive and navigates users through both a list of core principles as well as instrumental principles, to determine which of the principles are at play in each solution development scenario. Although the authors developed this checklist tool to help identify the relevant ethical concerns, their contention is that this is only a starting point and that complex scenarios require input from ethics experts that can draw on the relevant ethics theory.

2.8.5 Comprehensive organizational support

Winfield and Marina (2018) state that having professional standards and codes of conduct are an important starting point. However, they believe that this will only be effective if it is combined with a comprehensive ethical governance that includes strong institutional frameworks as well as principled leadership. The authors propose five starting points towards good ethical governance, namely:

1. Publishing an ethical code of conduct for the organisation – this ensures that individuals within the organisation are aligned on what is expected of them in the unique context of the specific organisation. The code of conduct should also include a mechanism to raise ethical concerns via a responsible manner.
2. Providing ethics and responsible innovation training for everyone in the organisation. Importantly, the authors make the point that ethics should not be added to existing processes, but should be integrated in a holistic manner into the ways of working. This is in line with the commentary of Gogoll et. al. (2021), covered in section 2.7.4, regarding the infusion of ethical deliberation into the software development process.

© University of Pretoria

3. Practising responsible innovation, including the participation of a wider range of stakeholders, the performance of ethical risk assessments, follow ups to action outcomes from ethical risk assessments and tools to undertake such assessments.
4. Practising transparency about ethical governance by not just claiming to be ethical but by showing exactly how the organisation is ethical in practice. Specific examples of this include publishing an ethical code, publicising membership of any ethics boards and sharing case studies of ethical risk assessments undertaken.
5. Valuing ethical governance rather than creating ethical governance frameworks as a smokescreen or to provide ethics-washing cover.

In this research study, RQ2 will explore the nature of the support provided to practitioners, at the case study organisation, to understand and to manage AI ethics considerations, the degree to which such considerations are managed at an individual level versus at an organisational level, as well as the perceived outcomes and effectiveness of managing such AI ethics concerns.

2.9 PROMOTING ETHICAL EXCELLENCE

In an essay on the importance of promoting ethical excellence in organisations, Cohen (2013) argues that it is much easier, though not necessarily more effective, to set a minimum ethical standard through the implementation of accountability systems than it is to foster a culture of responsible decision making. While accountability systems are important and serve the purpose of primarily detailing exactly how individuals met the basic compliance requirements for a specific situation, there are significant shortcomings with this approach. The author notes that setting up systems of accountability essentially set a baseline for the minimum acceptable compliance, without defining what excellence is represented by. This means that it is often possible to comply with the requirements of an accountability system while ignoring many pertinent items because they do not fall within the scope of what the accountability system requires from the individual concerned. Since these considerations are outside the scope of accountability systems, people are disincentivised from taking a broader look at any ethical considerations since it “doesn’t count for anything” in terms of how they are measured – in extreme cases, it may even lead to people ignoring known ethical concerns because they falls outside the scope of the accountability system.

© University of Pretoria

In a similar vein to the proposal to introduce ethical deliberation into the software development process, discussed in 2.7.3, Cohen (2013) suggests that people should be rather given responsibility for ensuring ethically responsible decisions are made, knowing that they have the support of their management instead of being given a formulaic methodology of addressing ethical concerns. In the current study, RQ3 will explore whether respondents resolve AI ethics concerns via pre-defined templates, or whether they exercised their own judgement to resolve such tensions. RQ3 will further explore whether perceived outcomes differed based on the approach followed.

2.10 THE ROLE OF WORKERS IN AI ETHICS

In their chapter on the role of AI workers in AI governance, Nedzhvetskaya and Tan (2022) propose a harm response model that seeks to highlight the dependencies and power dynamics in AI-related harm reporting. The model proposed is reproduced in Figure 2 below.

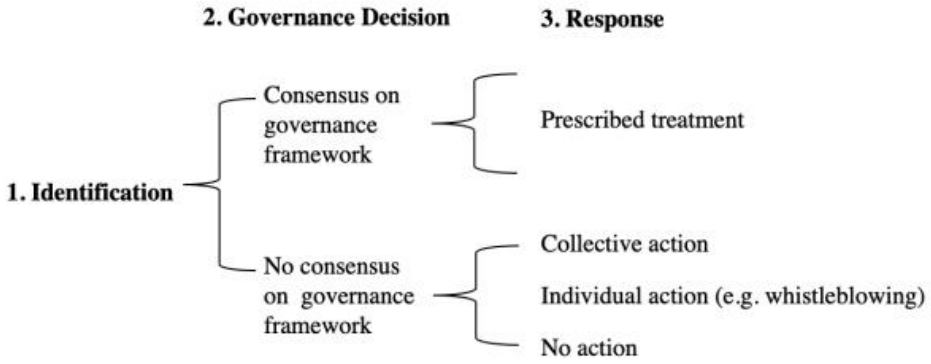

Figure 2: Nedzhevetskaya and Tan (2022) - Harm response model

According to their assessment, much of the research on AI ethics available today has focused on the identification of harms (Step 1 in their process model), through heuristics, toolboxes and checklists. However, additional work is required to assess the interdependencies between identification of harms, governance decisions and responses. As part of the research questions in this study, we will seek insights in the mechanics of all three stages of the harm response model in the case organisation. Questions in the research survey will cover the prevalence and identification of AI ethics tensions, the support provided by organisations to understand these tensions as well as to manage these tensions as well as the split between individual, team and external responsibilities in these processes.

© University of Pretoria

Furthermore, the survey will extend beyond the model above by exploring the perceived outcomes from implementing responses as well as the perceived effectiveness of responses.

2.11 THE PREVALENCE OF AI ETHICS TENSIONS

Whittlestone et al (2019a) suggest that it is encouraging that the research community seems to be converging on five key AI ethics principles because this allows for much broader and more complex ethical issues to be narrowed down into a few important themes. However, the authors note that principlism in applied ethics has some well-noted shortcomings:

- Principles are high level and general and lend themselves to a checklist but are not particularly suitable at a practical level in terms of guiding action.
- Principles are high level and can be broadly interpreted so they inevitably come into conflict with each other.
- As a result of their often-broad definitions, principles lend themselves to differing interpretations and are also subject to differing interpretations based on varying cultural norms and the specifics of individual scenarios.

The authors further suggest that the focus should rather shift from principles to exploring the tensions that arise when putting the principles into practice with a view to identifying, then exploring these tensions. Four reasons for focusing on tensions are provided:

1. Bridging the gap between high level principles and practice by highlighting how different values may come into conflict in practice and suggesting standards and regulations that help to apply principles in different scenarios.
2. Acknowledging that values may be interpreted differently by different groups and meanings of principles may differ among different groups.
3. Highlighting areas where there are gaps that need to be covered by researchers or policy makers.
4. Identifying ambiguities and knowledge gaps in our understanding of AI and its full effect on society so that we can refine research agendas with the hope of closing these gaps.

© University of Pretoria

Whittlestone et al (2019a, p199) elaborate on this by identifying four key tensions that are likely to arise in AI applications and which will be explored further in the current study:

Improving quality of services versus respecting privacy and autonomy

In AI/ ML applications, data is a required input, and the general rule is that the more data provided, the more accurate or useful the application will be. While these applications provide improvements in performance with additional data, this by necessity means that individual privacy is impacted by the increasing amount of data provided. Collecting more data also necessitates obtaining broader levels of consent which has negative implications for individual autonomy and privacy.

Enhancing accuracy of predictions versus ensuring equality and fairness

Algorithms can make better predictions and decisions with more data. However, having such improvements in accuracy available may result in unfair treatment of specific groups either because of a lack of data about groups or because of biases or skews in the process of collecting data about specific groups. For example, it is well documented that machine learning algorithms for facial detection perform better on light-skinned population groups due to skew in training data applied to most systems. Perkowitz (2021) provides a summary of some of the challenges inherent in facial recognition training data. Mittelstadt et al (2022) make the important point that equality and fairness need to be measured broadly, not only in terms of fairness to underrepresented groups but also fairness to overrepresented groups. The author explains that a quest to improve the lot of marginalized groups cannot be done at the expense of other groups – in other words, ensuring performance across the board is worse just to ensure equality of outcomes is a substandard outcome for AI solutions.

Increasing personalisation versus enhancing citizenship

One of the most often-touted benefits of machine learning systems is the ability to offer greater personalisation, whether this is for items to buy, television shows to watch or services to consume. Unquestionably, there is much consumer benefit to be harvested by the increase in personalisation. However, we run the risk of creating divisions in society if we increasingly segregate people into smaller-and-smaller groupings based on data gathered from them through their interactions with various systems and services.

Using more automation versus promoting dignity

In the corporate environment, automation of processes or tasks through the implementation of machine learning systems promises to improve efficiencies and reduce costs – which will have a

© University of Pretoria

positive impact on corporate profits. However, increasing automation may result in changes to the notion of creative tasks being the exclusive domain of humans e.g., creating visual art. Recent advances in generative AI solutions (e.g., Chat GPT) have already brought such questions to the fore of public debate (World Economic Forum, 2023). Increasing automation also runs the risk of accelerating the deskilling of people from jobs which are within the capabilities of AI systems.

RQ2 will assess the extent to which each of the above tensions are present in applications developed at a South African media company. RQ3 will focus on how these tensions are managed, both by individuals and the organisation, when they arise during such AI solution development initiatives.

Whittlestone et al (2019b, p13) also identify deeper understanding of current technological capabilities and real-world use cases as being important to both understanding the potential impact of AI as well to understanding the different perspectives of different groups regarding commonly used concepts and terminology. RQ1 will detail the types of AI applications being developed in a South African media company, with specific information on the roles involved in project delivery, the duration of projects, the types of technologies used, the team composition, project focus and project outcomes.

2.12 GAPS IN THE LITERATURE ADDRESSED BY THE CURRENT STUDY

AI ethics remains a field of research in its infancy at the point of authoring this report. While significant effort has gone into creating high level guidelines for codifying AI ethics principles, relatively little work has been produced on the practical implementation of such guidelines in organisations in the private sector.

Before we progress in the field of AI ethics, it is a pre-requisite that we have a better understanding of AI solution development itself, so that we know better the characteristics of what we are applying AI ethics principles and practices to in the first place. For competitive advantage and other reasons, relatively little information is disclosed publicly by large corporations about their AI solutions development efforts. Usually this is only done to support marketing of AI solutions, that are being sold as products. Very little is published by large corporations related to how they staff AI solution development projects, how long these solutions typically take to develop, what combinations of technologies are used, how corporates acquire the technologies and skills required to implement

© University of Pretoria

these solutions, what the primary focus of AI initiatives is and success and failure rates of such AI initiatives. With specific reference to geography, from an African and South African perspective, no empirical data has been published to date on AI solution development characteristics such as those mentioned above.

Thus, from an AI ethics perspective, very little empirical data is available on exactly how people manage AI ethics in practice in large corporations. We know about the AI ethics principles and guiding practical frameworks of the largest technology corporates (Pichai, 2018; Crampton, 2022). However, very little to nothing is shared by those same corporates about how they operationalize these frameworks on actual projects. As an example, the extent to which AI ethics tensions are present in AI solution development in industry has not been disclosed via any empirical studies at the time of writing in early 2023. Similarly, the manner of managing such tensions is not well documented and the split between organisational and individual responsibility for managing such ethics tensions has not been documented. Finally, the impact and outcomes of managing AI ethics tensions in the private sector has been very lightly covered by existing academic literature, if at all. The current study will aim to address some of these gaps in the literature to the degree possible with a case study of a large organisation in the media industry in South Africa. Empirical data has been collected on the topics of AI solution development as well as practical management of AI ethics tensions, which this research report presents.

© University of Pretoria

2.13 CONCLUSION

The literature included in this chapter presents a background to the study of AI ethics pertinent to the research topic. Foundational terminology is presented, after which the evolution of AI ethics is traced at a high level. The development of AI ethics guidelines is reflected upon, as well as the limitations of a guidelines-based approach to AI ethics.

The move towards managing AI ethics in practice is covered together with some practical interventions that have been proposed in the literature to manage AI ethics. The role of individual judgement versus compliance to ethics policies is discussed, as well as the role of workers in reporting AI harms. Finally, AI ethics tensions are explored with four tension pairs explained. Gaps in existing literature coverage are noted – the current study will aim to address some of these gaps in the published literature.

The following chapter will present the research methodology for this study, including the research aims and the research questions.

© University of Pretoria

3. RESEARCH METHODOLOGY

3.1 AIMS

The primary aim of this research study is to survey the prevalence and management of four pairs of AI ethics tensions in AI applications being developed in a South African media organisation. An additional aim of the study is to critically analyse, via a literature review, the splitting of responsibility for managing AI ethics tensions between individual practitioners and the organisations they are employed by and to use this analysis to inform the recommendations for the development of an AI ethics tensions management framework. In support of the primary aim, a further aim of this study is to identify the types of AI applications being developed in the case organisation as well as to identify pertinent characteristics of such AI applications to address existing gaps in the literature related to actual AI applications that are being developed now in industry.

3.2 RESEARCH DESIGN

The study will follow a cross-sectional survey design. Surveys will be self-administered via an online link that will be provided to participants. Cross-sectional surveys allow for the measurement of potentially large numbers of variables at a single point in time. Such surveys benefit from strong external validity due to the collection of data directly from the field but they have limited internal validity due to the point-in-time nature of this type of survey, resulting in an inability to characterize cause and effect relationships adequately (Bhattacherjee 2012, p.39). Bhattacherjee (2012) also points out that surveys suffer from several potential biases that have an impact on the validity of the research including non-response bias, sampling bias, social desirability bias, recall bias and common method bias (pp. 80-82).

© University of Pretoria

3.3 RESEARCH SETTING

Multichoice Group has been selected as the media organisation that will be the subject of this case study. Multichoice Group is considered a suitable research case for the following reasons:

- A large and active AI / machine learning community exists within the organisation, who may be able to participate in this study. This will provide a range of perspectives across the AI solution development lifecycle, from both technical as well as business-related perspectives.
- A diverse set of AI solutions are being developed or used within the organisation to address a multitude of business problems – this provides an opportunity to address some gaps in literature regarding actual AI use cases that are used by a South African media company.
- AI use cases are in various life cycle stages (in concept, in development and in production) which affords the opportunity to assess the perceived impact of ethical considerations at various points in the AI development process.

3.4 RELIABILITY

In social science research, reliability refers to the consistency of measurements (Drost, 2011) – when performed by different people, at different times and under different conditions. There should be an expectation that we get the same results if the measurement instruments have high reliability. Measurements can be subject to systemic errors or random errors – random errors will have the potential to impact reliability while systematic errors have the potential to impact validity.

According to Drost (2011), reliability is most estimated using measures of association via the correlation coefficient, otherwise referred to as a reliability coefficient. Reliability can be measured using a variety of options, including test-retest reliability, alternative forms, split-halves, inter-rater reliability and internal consistency (Drost, 2011). Each of these methods will be briefly described in the paragraphs below.

© University of Pretoria

3.4.1 Test-retest reliability

In test-retest reliability, the test is for stability across measurement sessions – i.e., temporal stability. The given instrument is administered to the same group of respondents twice, with the correlation between the attempts providing test-retest reliability (Drost, 2011). Test-retest as a technique suffers from limitations based on whether the interval between testing is too short (memory recall may affect the responses if remembered answers play a role in the second session of answers) or too long (the subject of the testing may be exposed to new information which may change their responses compared with the initial test session) (Drost, 2011).

3.4.2 Inter-rater reliability

Inter-rater reliability measures the reliability of each rater or the combined consistency of each rater's judgements (Drost, 2011). We are able to use the correlations between the ratings of individual judges to determine to reliability of each rater.

3.4.3 Split-half approach

In the split-half approach, two new tests and two new measures are created, each consisting of half of the original items, but both measuring the same behaviour (Drost, 2011). The advantages of the split-half method is that 1) it is not subject to memory effects, 2) it is usually cheaper to execute and 3) it does not require extended periods of time to collect test data (Drost 2011).

3.4.4 Internal consistency

Internal consistency refers to the degree of reliability of the constituents of the test instrument, it denotes how well a set of items measure a specific characteristic in a test (Drost, 2011). The measure of internal consistency is based on the average correlations between the single items in the test. Cronbach's alpha, or correlation alpha, is the most popular method of testing for internal consistency. Cronbach's alpha is useful for estimating reliability when the existence of a single factor has been determined (Drost, 2011).

© University of Pretoria

3.4.5 Factors affecting test reliability

Factors that affect the reliability of a test can broadly be divided into two categories, namely errors within a test and errors or variations between tests (Drost, 2011). The author further explains that errors within a test may be due to guessing, errors in sampling, administrative errors, missing out questions or failing to understand the actual question while errors between tests may be due to differences in content between two tests or changes to an individual's stance towards the test content between two iterations of a test.

3.4.6 Steps to improve reliability

Drost (2011) provides several suggestions to improve the reliability of a test. These suggestions are noted below, together with a summary of how these suggestions are incorporated in the current study to improve reliability of the test instrument.

Reliability improvement suggestion	**How this is implemented in the current study**
Write clearly	Tests are implemented using Microsoft Forms online to improve user experience
Include clear instructions	All sections have clear instructions, and some questions have hints
Provide training for respondents	Examples are provided to further illustrate concepts
Make tests longer	The survey consists of 53 questions

Table 2: Improving reliability of tests – suggested by Drost (2011)

Bhattacherjee (2012) also provides suggestions to improve the reliability of a test. These suggestions are noted below, together with a summary of how these suggestions are incorporated into the current study to improve reliability of the test instrument.

© University of Pretoria

Reliability improvement suggestion	How this is implemented in the current study
Replace data collection techniques that depend on researcher subjectivity	Questionnaires are implemented online, with data collected directly from participants
Ask only those questions that respondents may know the answers to	Sampling technique only selects people who have worked on AI projects previously at Multichoice
Avoid ambiguous items in your measure	Questions have been delineated as much as possible to answer only a single item each
Simplify the wording	Measure was tested on a pilot audience before deployment and suggestions incorporated

Table 3: Improving reliability in tests - suggested by Bhattacherjee (2012)

3.5 VALIDITY

While reliability measures the degree to which measurements are repeatable across test iterations, validity measures whether the test is meaningful i.e., whether the test is measuring what it sets out to measure (Drost, 2011). Each of the four types of validity are discussed briefly below.

Internal validity

Internal validity stems from the question of, given that there is a relationship between variables established by a study, is the relationship causal? Drost (2011) lists several threats to internal validity, including “history, maturation, testing, instrumentation, selection, mortality, diffusion of treatment and compensatory equalisation, rivalry and demoralisation”.

External validity

External validity refers to generalisability of relationships i.e., if a study finds a causal relationship between two constructs, how generalisable is this across different contexts, times and respondents (Dross, 2011). Generalisation to a specific target population is different to generalising across populations (Dross, 2011). If a specific target population has been mentioned in research objectives,

© University of Pretoria

external validity determines whether the research objectives have been met. External validity is also used to determine whether research findings can be generalised.

Construct validity

Construct validity refers to how well a research construct has been operationalised into a research instrument (Dross, 2011). The authors note that proving validity requires providing evidence across six types of validity, namely “face validity, content validity, concurrent and predictive validity, and convergent and discriminant validity”.

Statistical conclusion validity

This refers to whether a relationship exists between two variables (Drost, 2011). The author lists several threats to statistical conclusion validity, including “low statistical power, violation of assumptions, reliability of measures, reliability of treatment, random irrelevancies in the experimental setting, and random heterogeneity of respondents”.

3.6 SAMPLING

According to Bhattacherjee (2012), sampling refers to the statistical process of choosing a subset from a population of interest so that one can make observations as well as statistical inferences about the population. In Daniel (2012), sampling is defined as “the selection of a subset of a population for inclusion in a study”. Bhattacherjee (2012) notes that sampling is required to ensure feasibility and cost constraints are met in social science research – however, it is required that researchers ensure that the sample chosen is representative of the population of interest, to support generalization of findings. Invalid inferences may result if samples are incorrectly chosen or are biased.

© University of Pretoria

Population

Bhattacherjee (2012) defines a population as all units of analysis containing the characteristics that a researcher wishes to study or draw scientific inferences from. The author notes that a unit of analysis may be a “person, group, organization, country, object, or any other entity “.

Sampling design

Sampling design can be either single stage or multistage, also referred to as cluster sampling (Creswell, 2009). The author notes that single-stage sampling is appropriate when the researcher can sample the people directly (for example, if the researcher has access to the names of the people in the population) while cluster or multistage sampling is appropriate when groups need to be identified first, before names of people can be obtained from within those groups that will be sampled. In the present study, single-stage sampling was appropriate since names and contact details (email addresses) of individuals making up the population of interest were directly accessible to the researcher.

Sampling frame

The sampling frame refers to an accessible portion of the target population from which a sample can be drawn (Bhattacherjee, 2012). The author notes that this can be in the form of a list, with contact information of some sort. The purpose of a sampling frame is to denote a practical target sample for the study, when sampling the entire population is not realistic (Bhattacherjee, 2012). The author further notes that a sampling frame may not be representative of the population – in this case, one needs to be careful not to generalize findings from a sample to the entire population.

According to Fowler (2009), sampling frames have three characteristics that need to be understood and evaluated:

1. How comprehensive it is in covering the target population.

2. Whether it is possible to calculate the probability that a unit in the population will be selected.

3. The efficiency with which elements of the target population can be found within the frame.

The author notes that email addresses within a business context are a good example of a comprehensive source of population information, given that people employed by a business would almost universally have access to an email address provided by the employer.

© University of Pretoria

Sampling techniques

In the most general terms, sampling techniques either fall into the category of probability sampling or non-probability sampling. Bhattacherjee (2012) notes that probability sampling lends itself to generalization of results – however, non-probability sampling is justified in specific circumstances.

Probability sampling

In probability sampling, every unit in a population has a non-zero chance of being selected in the sample, and it is possible to determine this probability accurately (Bhattacherjee, 2012). The author notes that there are two attributes which are common to all types of probability sampling, namely 1) that every element in the population has a known, and non-zero, chance of being sampled and 2) there is random selection at some point in the sampling procedure. Bhattacherjee (2012) notes that types of probability sampling include simple random sampling (all subsets of a population have an equal probability of being selected), systematic sampling (elements are selected at regular intervals after the sampling frame has been ordered according to a selected criteria), stratified sampling (simple random sampling is performed after the sampling frame has been divided into non-overlapping subgroups), cluster sampling (dividing a population into more convenient clusters, select clusters randomly but sample all units in a cluster), matched-pairs sampling (dividing a population into two sub-groups based on a specific criterion then performing random sampling in one group with a suitable match from the second group) and multi-stage sampling (a combination of any of the above techniques in series in the same study, based on the requirements of the study).

Non-probability sampling

In non-probability sampling, some elements of a population have a zero chance of selection, or the probability of selection cannot be evaluated accurately (Bhattacherjee, 2012). The author lists convenience sampling, quota sampling, expert sampling and snowball sampling as examples of non-probability sampling techniques.

© University of Pretoria

3.7 PARTICIPANT SELECTION CRITERIA

The participant selection criteria are designed to ensure that 1) accessibility to participants for the purpose of this research project is reasonably practical within the time constraints of this research undertaking and, 2) that some of the bias mitigation strategies for managing survey research design are employed effectively.

With regards to accessibility of research participants, the criteria included are that each participant 1) must still be employed within Multichoice at the time of administering the survey and 2) must have access to a Multichoice domain email address via which all communications will be undertaken (so that they can be contacted initially as well as so that follow up communications can be easily implemented).

With regards to bias mitigation strategies, current employment at Multichoice may help with reducing non-response bias (through endorsement of the research from the Group CIO as well as the ease of implementing follow-up requests) while the requirement for participants to have been actively involved in at least one recently completed AI project within Multichoice may also help improve non-response bias through higher relevance of content (Bhattacherjee (2012, p. 81).

Direct involvement in an AI project, in the context of this research study, means that the participant must self-identify as having held one of the following roles during the AI project:

1. Data Scientist
2. Executive
3. Other Technical Professional
4. AI Engineer
5. Software Engineer
6. Business Stakeholder
7. Other Non-Technical Professional
8. Data Engineer

© University of Pretoria

To be included in the study, participants must also consent both to participate in the research study as well as to the sharing of their anonymised data for the research report as well as subsequent academic publications.

3.8 RECRUITMENT OF PARTICIPANTS

Participants were selected using expert sampling, a non-probabilistic sampling technique. According to Bhattacherjee (2012, p. 69), expert sampling is a technique that is applicable where respondents have a specific expertise on the phenomenon being studied. In this case, respondents were selected based on their prior exposure and participation in AI product development at Multichoice. Figure 3. shows the participant recruitment process followed for this research study.

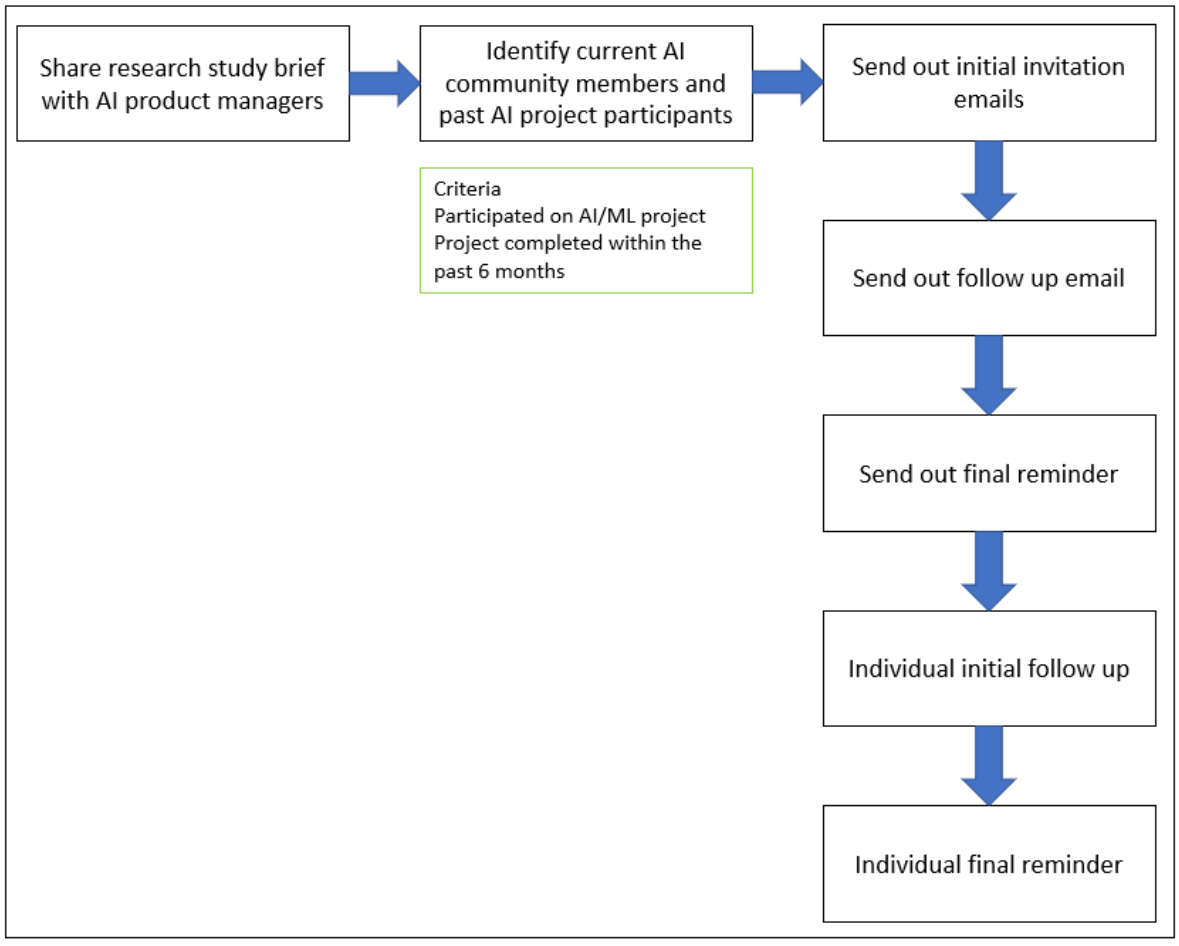

Figure 3: Participant selection and interaction flow

© University of Pretoria

AI product managers at the case organisation, Multichoice, are responsible for maintaining the roadmap of all AI initiatives undertaken across the organisations. They are therefore best positioned to provide insights into likely participants for this type of research study. After sharing the research study brief with the AI product managers, a list of employees who either form part of the AI community or who have recently participated in AI projects at Multichoice was identified.

Email invitations were sent out to this list of employees, requesting participation in the study. It was noted in the invitation that participation in the study was voluntary and that the results from the study would be shared with respondents at the conclusion of the study. Respondents were provided with a link at which the survey could be completed. Upon accessing the link, all respondents were required to confirm both consent to participate in the research study as well as consent to share data for the purposes of further analysis during the study, as well as for storage purposes according to the requirements of the University of Pretoria.

3.9 ETHICAL CONSIDERATIONS

An application was submitted online and was approved by the Research Ethics Committee of the Faculty of Humanities at the University of Pretoria (Reference number: HUM005/0421).

Permission to conduct the research study within Multichoice was provided by the Group CIO of Multichoice Group on the 24th of April 2021.

Each participant was invited to participate in the research study and provided with an explanation of the research context as well as the option to not participate in the research study. A letter of introduction and informed consent was supplied to each participant via email to their registered Multichoice domain email address. Furthermore, each online questionnaire contained a first section which requested confirmation of permission to participate in the study as well as confirmation that the participant's data could be used in an anonymised manner in subsequent academic publications.

To minimize risk of data leaks related to personal information and to comply with the POPI Act, no personally identifiable information, such as name, employee number or national identification details was collected from participants.

© University of Pretoria

3.10 MATERIALS

The purpose of a pilot phase in a research study is to ensure that the responses received are in line with what the researcher is expecting to achieve from the study i.e., ensure that questions perform as expected, but not pre-empt or predict the results of the study (Pickard, 2017). The survey was piloted with two participants, each participant occupying different roles within the organisation, and each having been involved in at least one previous AI product development project. It was confirmed that participants were able to understand the nature of AI ethics tension pairs from the examples provided at the beginning of each section related to the tension pairs.

The questionnaire consisted of seven sections.

Section One captured the informed consent confirmation from the participants as well as the permission to use anonymised participant data for future academic publications.

Section Two captured information about the participant's level of work experience, both generally as well as with AI solutions specifically, in addition to information about the participant's role in the identified AI product development project.

Section Three captured details about the nature of the AI product development project. This included information about the role fulfilled on the project, the project duration, duration of involvement, nature of AI technology used, nature of the business objectives of the project as well as the composition of the project team.

Sections Four to Seven captured information related to the participant's experience of each of the four identified AI ethics tension pairs during the AI product development project. A real-world example of each tension pair was provided at the beginning of each of the Sections Four to Seven to provide further context to the research participants.

3.11 DATA COLLECTION METHOD

The surveys were administered online using the Microsoft Office Forms online functionality, available at https://forms.office.com/ . Participants who indicated willingness to participate in the study were

© University of Pretoria

sent a link to the online survey. Individual and group reminders were sent to participants to help improve the response rate.

3.12 LIMITATIONS OF SURVEY DESIGN

Bhattacherjee (2012, p. 35) notes that the quality of a research design can be assessed in relation to four attributes:

1. Internal validity, which examines the nature of the cause/effect relationship between independent and dependent variables. It is noted that laboratory experiments may have strong internal validity due to the ability to manage temporal precedence effectively while field surveys have poor internal validity because of the inability to distinguish between cause and effect.

2. External validity, which refers to the ability to generalize observations from a sample to a larger population. Here, it is noted that survey research has strong external validity due to the data collection being based in the "real world" with a variety of participants.

3. Construct validity, which relates to how well the measurement scale is measuring the theoretical construct that is the subject of measurement.

4. Statistical conclusion validity, which examines the degree to which any conclusions derived using statistical processes is valid. This validity dimension is concerned with questions such as whether the correct statistical method was used as well as whether sample size requirements were met.

Bhattacherjee (2012, p.39) notes that field surveys have limited internal validity due to the point-in-time nature of this type of survey, resulting in an inability to characterize cause and effect relationships adequately. Bhattacherjee (2012, pp. 80-82) also points out that surveys suffer from several potential biases that may have an impact on the validity of research including non-response bias, sampling bias, social desirability bias, recall bias and common method bias. On the other hand, the author notes that field surveys have strong external validity, allow for the capture and control of many variables and allow for the study of a problem from multiple perspectives.

© University of Pretoria

In choosing a cross-sectional field survey as the research design for the current study, the researcher has noted the respective strengths and weaknesses of the survey research method while acknowledging the following actions which have been taken to minimize the impact of the drawbacks of this research method:

1. Non-response bias – refers to the challenge to validity of a research design if most respondents do not respond to the survey. Bhattacherjee (2012, p. 80) notes that this is a common concern in survey research, but also notes several strategies to mitigate this drawback. In the current study, the research workflow includes many of the listed strategies to improve the response rate, namely 1) advance notification; 2) relevance of content; respondent-friendly questionnaire; endorsement; follow-up requests; confidentiality and privacy.

2. Sampling bias – refers to the problem of selecting a sample that does not adequately represent the intended population. To mitigate this bias, the research workflow includes sampling all participants in qualifying AI product development projects, not just engineers or managers for example.

3. Social desirability bias – refers to the avoidance of controversial or negative responses in order to paint a better picture than the reality. While Bhattacherjee (2012, p.81) suggests that it is difficult to eliminate this bias, the anonymous nature of this survey provides as much opportunity as possible to minimize this effect.

4. Recall bias – refers to a subject's memory, motivation and ability (Bhattacherjee (2012, p.82). To mitigate this bias, the survey is designed to reference a specific project (participants are required to choose only one project to base their answers on). Participants are also required to capture the elapsed time between completing the project and answering the survey.

© University of Pretoria

3.13 DATA COLLECTION

Participant responses were collated on the Microsoft Office Forms online website. At the conclusion of the data collection period, the summary of the responses was downloaded from the website in Microsoft Excel format for further analysis.

The questionnaire consists of the following sections:

- Cover page, including description of the study, explanation of procedures, benefits, confidentiality, use of participant data, withdrawal without prejudice, further questions and follow-up as well as contact details of the author.
- Section One: Informed Consent – this is a mandatory section to complete. No further responses to the survey are recorded if the respondent does not agree to grant permission to participate in the survey as well as for responses to be used in the research report and subsequent publications.
- Section Two: Participant details – this section collects demographic data about the respondent.
- Section Three: Project details – this section collects information about the project reported on by the respondent.
- Sections Four-Seven: Captures information about the prevalence and management of the four tension pairs that are the subject of this research study.

A total of 53 questions are included, across all sections.

Mandatory response field to improve data reliability

The questionnaire has been configured such that all responses are mandatory – this means that all responses received have a complete questionnaire with no questions left blank or unanswered.

© University of Pretoria

3.14 DATA STORAGE AND SECURITY

During this research project, all data collected will be stored online in the personal Microsoft accounts of the principal researcher. Since the data will be collected using Microsoft Forms online functionality, it will be protected by the standard features of Microsoft Forms, which includes protection of data at rest as well as data in transit. Further details of the security used can be found at this link: https://support.microsoft.com/en-us/office/security-and-privacy-in-microsoft-forms-7e57f9ba-4aeb-4b1b-9e21-b75318532cd9

During submission of the final research report, all data collected during this research project will be shared with the University of Pretoria in a zipped and password-protected file. It is required that this data is stored for a period of 15 years.

3.15 DATA EXTRACTION AND ANALYSIS

The downloaded response file obtained from Microsoft Office Forms online website provides both a collation of all responses on a per-respondent basis as well as a collation of all responses on a question-by-question basis.

Descriptive statistics were used to analyse the data provided by participants in Section Two (Respondent Details) and Section Three (Project Details).

Inferential statistics was used to analyse the data provided by participants in Sections Four to Seven (AI Ethics Tensions Details).

3.16 VARIABLES: DESCRIPTIVE

Descriptive surveys serve the purpose of helping to characterise a situation and to look for patterns within the sample group (Pickard, 2017). According to the author, the data usually gathered is a

© University of Pretoria

combination of counts and measurements – we make use of descriptive statistics to analyse these types of surveys. Within the current research study, the descriptive element helps to answer the research question on the types of AI applications being developed today, as well as the prevalence of AI ethics tensions in AI application development.

3.16.1 Demographic information

To characterise the respondents to the survey i.e., people who are working or have worked on recent AI solution development, the following variables are noted. Descriptive statistics were performed on these variables, which will be discussed in Chapter Four. Each variable is listed below, together with a brief explanation of the variable.

Years of work experience

The overall number of years of work experience reported by the respondent.

Years of work experience at Multichoice Group

The total number of years of work experience specific to Multichoice Group.

Years of experience working with ML applications

The total number of years of work experience working with ML applications, across all organisations.

Primary role

The primary role held by the respondent.

Business unit reported to

The business unit that the respondent was part of at the time of completing the project that is being reported on.

© University of Pretoria

3.16.2 Project information

Answering RQ1 provides insights into the types of AI applications being worked on in a media company in South Africa. To fully characterise these AI applications, information has been gathered on the nature of the projects through which these AI applications were developed and delivered.

Project role

The role held by the respondent during the project being reported on. The skills gap in AI is well documented. According to Deloitte (2020), 68% of executives report a moderate-to-extreme skills gaps while a further 27% report an extreme skills gap. This often results in practitioners taking temporary roles on AI projects to fill in the gap in skill availability, only to return to their actual roles at the conclusion of the project. The purpose of this variable is to capture the role held on the project, which could be different from the respondent's permanent role.

Project duration

The total duration of the project.

Duration worked on project

The duration worked on the project by the respondent, noting that this could be shorter than the full project duration.

Business objective

The primary business goal to be achieved by the project.

Focus

The focus of the project – either externally (customer) facing or internally facing.

© University of Pretoria

Outcome

The outcome of the project – did the project meet expectations?

Technologies used

The technologies used on the project, as far as was visible to the respondent.

Technology acquisition method

The method used to acquire the technology on the project. The purpose of this variable is to provide insights into the proportions of AI solution development that consists of internally generated intellectual property versus purchased or licensed intellectual property.

Team composition

The composition of the project team, noting the split between contractors and permanent staff.

3.16.3 AI ethics tensions information

Prevalence

The occurrence of AI ethics tensions reported by respondents.

Method of discovery

The manner of discovery of the ethics tensions, reported by the respondent.

Understanding trade-offs

The extent of support provided by the organisation to the respondent, to understand the trade-offs represented by the AI ethics tension.

© University of Pretoria

Outcome

The perceived outcome of managing the AI ethics tension under consideration.

Level of management

The level at which the AI ethics tension was managed.

Appropriateness of level

The perception of the respondent on the appropriateness of the level at which the AI ethics tension was managed.

Organisational support

The organisational support provided to the respondent to manage the AI ethics tension under consideration in the response.

Effectiveness

The perceived effectiveness of managing the AI ethics tension under consideration.

3.17 VARIABLES: INFERENTIAL

3.17.1 Association between demographics and project characteristics, and tension prevalence

Inferential statistics were performed to determine whether demographics of the respondent or project characteristics impact on the prevalence of tensions reported. The following variables were tested for association / relationship:

- Work experience versus tension prevalence
- Project role versus tension prevalence

© University of Pretoria

- Total project duration versus tension prevalence

3.17.2 Association between tension management and perceived effectiveness of tension management

Inferential statistics were performed to determine whether there is an association between aspects of tension management (such as method of discovery, method of management, understanding trade-offs, perceived outcome, level of management, appropriateness of level of management and organisational support provided) and the perceived effectiveness in managing the tension.

3.17.3 Association between tension management and perceived outcome of tension management

Inferential statistics were performed to determine whether there is an association between aspects of tension management (such as method of discovery, method of management, understanding trade-offs, level of management, appropriateness of level of management and organisational support provided) and the outcome of managing the tension.

© University of Pretoria

4. DISCUSSION OF RESULTS

4.1 INTRODUCTION

The purpose of this chapter is to present the results of the survey conducted during this research study. The primary focus of this chapter is to answer the research questions posed in the previous chapter. For ease of reference, the research questions are listed below:

RQ1: What are the types of AI applications being worked on in a South African media company?

RQ2: Which, if any, AI ethics tensions are present during the development of such AI applications?

RQ3: How do individuals manage AI ethics tensions in current AI applications?

RQ4: Can we propose recommendations for an effective framework to manage AI ethics tensions?

While this chapter presents the results, an in-depth analysis of the results, as well as conclusions drawn, forms part of Chapter Five.

© University of Pretoria

4.2 SCOPE OF THIS CHAPTER

The figure below depicts the flow of the subsequent sections in this chapter.

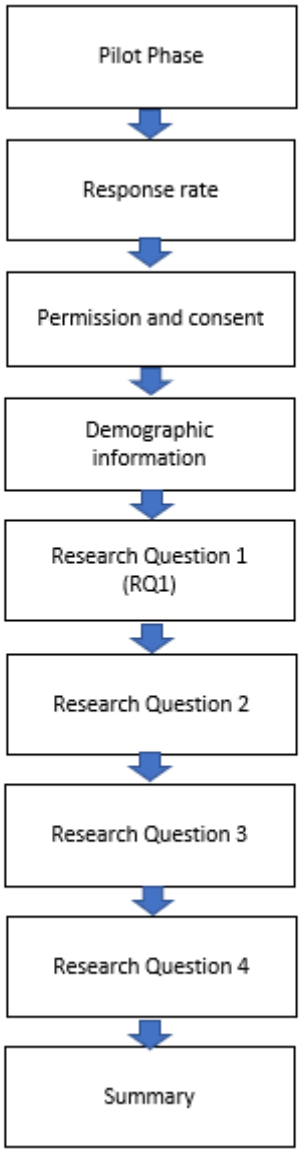

Figure 4: Flow of Chapter Four

The purpose and feedback from the pilot phase of the survey is discussed, followed by sections on response rate, and permission and consent. This is followed by a presentation of the results related to demographic information collected from the respondents. Note that the surveys were anonymous so

© University of Pretoria

no personally identifiable information was collected, neither was information collected about participant gender, race or sexual orientation.

Sections 4.7 – 4.9 seek to answer research questions one to three (RQ1, RQ2 and RQ3). Section 4.10 presents the findings relevant to answer research question 4, but the full implications for the creation of an AI ethics tension management framework are discussed in Chapter Five in the results section.

4.3 PILOT PHASE

The survey was piloted with two participants – the participants had differing levels of seniority within the organisation, and each had been involved in at least one previous AI product development project. The intentions of the pilot phase were:

- Determine whether the questions were easy to understand based on the prerequisites required for participation.
- Determine whether any answer options were missing for the multiple-choice questions.
- Determine whether the question hints were relevant and sufficient to gather the appropriate set of responses from participants.

Feedback received from the pilot phase participants was as follows:

1. Questions were clearly stated and easy to understand – explanations provided in the preamble of the questionnaire were sufficient.
2. Some business unit options as well as some role options were needed but not initially included – the final survey was updated to include these options.
3. Explanations of tension pairs needed more detail, which was added to the final questionnaire.

4.4 RESPONSE RATE

The survey was sent to 60 respondents – the request to complete the survey was sent via email. Multiple reminders to complete the survey were sent over a period of three months. Further individual

© University of Pretoria

follows ups were necessary to improve the response rate for completion of the survey. The final response rate was 73.33%, 44 completed surveys out of 60 people initially contacted.

4.5 PERMISSION AND CONSENT

In section one of the survey (questions one and two), participants were asked to agree to grant permission to participate in the survey and to confirm that they understood the purpose of the research. They were also asked to consent to the responses being used for the research and for subsequent academic publications. Since positive responses to both questions in section one was mandatory to proceed to the rest of the survey, all responses captured included positive responses to both questions. No respondents contacted the researcher to express any concerns either about participating in the survey or about any of the questions contained in the questionnaire.

4.6 DEMOGRAPHIC INFORMATION

The purpose of section two of the questionnaire was to collect demographic information about the participants in the research survey. Analysis of the demographic information provides some very important insights into the nature of participation in AI solution development projects, which could potentially have an impact on the efficacy of AI ethics interventions, and which will be discussed further in Chapter Five.

How many years of work experience do you have?

The years of experience reported by respondents ranged from 2.5 years to 32 years of experience with an average of 13.6 years' work experience and a median of 11.5 years' work experience.

For the purposes of further analysis in subsequent sections, the years of work experience was grouped into the following categories, in line with categorization used within the case organisation to reflect differing levels of seniority:

- Early career: 0-5 years' experience, 1 response.

© University of Pretoria

- Mid-career: 5-10 years' experience, 10 responses.
- Experienced: > 10 years' experience, 33 responses.

How many years of work experience at Multichoice do you have?

The years of work experience at Multichoice, for respondents to the survey, ranged from zero to fourteen years with an average tenure of 4.57 years and a median value of four years. This compares with an average tenure of Multichoice staff in general of seven years, a figure provided by a representative of the Multichoice HR department via email.

How many years of ML/AI experience do you have?

The reported number of years of ML/AI experience averaged to 3.39 years with a median of three years' experience in ML/AI.

Which role do you primarily identify with?

44 responses were received from people holding nine different roles, with the spread of responses shown in the table below.

Role	Number of Responses	Percentages
Executive	1	2%
Data Engineer	3	7%
Other Non-Technical Professional	3	7%
Software Engineer	3	7%
Business Stakeholder	4	9%
Product Manager	5	11%
AI Engineer	6	14%
Data Scientist	9	20%
Other Technical Professional	10	23%
Total	44	100%

Table 4:Roles held by survey respondents

© University of Pretoria

4.7 RESEARCH QUESTION ONE (RQ1)

RQ1: What are the types of AI applications being worked on in a South African media company?

In answering research question one, information about reporting lines as well as roles of practitioners are also included. These provide useful insights into how projects are run in the organisation and need to be considered when implementing any measures to manage AI ethics tensions.

Which business unit did you report to at the time of the project?

Survey responses were recorded from eight different business units, as shown in the table below.

Business Unit	Number of Responses	Percentages
Digital Enablement	1	2%
Finance	1	2%
Centre for Information and Insights	5	11%
Connected Video	6	14%
Customer Group	7	16%
Enterprise Business Systems	7	16%
Broadcast Technology	8	18%
AI Centre of Excellence	9	20%
Total	44	100%

Table 5: Business units originating AI projects

Although Multichoice operates an AI Centre of Excellence, which is responsible for building core AI expertise centrally, it is noteworthy that AI projects are originated across multiple divisions in the organisation. This has important implications for the implementation of any measures to manage AI ethics tensions. Both technical and non-technical business units need to be covered by any frameworks that manage AI ethics tensions. Furthermore, given the variety of roles involved in the development of AI solutions, as evidenced by the data in Table 1 above, it is equally clear that any interventions to manage AI ethics tensions should include traditional AI roles as well as other roles that are responsible for the delivery of projects in a typical organisation. Limiting any interventions to technical teams, or to business units that are traditionally associated with IT project delivery, runs the

© University of Pretoria

risk of missing out on many practitioners that are intricately involved in the development and delivery of AI solutions.

What role did you fulfil on this project?

According to their responses, people held eight different primary roles in the projects reported on.

Role	Number of Responses	Percentages
Data Engineer	2	5%
Other Non-Technical Professional	3	7%
Software Engineer	4	9%
Business Stakeholder	4	9%
AI Engineer	6	14%
Product Manager	7	16%
Data Scientist	9	20%
Other Technical Professional	9	20%
Total	44	100%

Table 6: Primary role performed on AI projects

While most respondents held the same role in the project as the primary role they identified with, a significant number of role changes were noted in the responses and are recorded in Table 4 below. In total, 11 out of 44 respondents (25% of respondents) noted that they changed primary roles for the duration of the project.

The frequency with which practitioners undertook role changes could be indicative of some realities facing organisations implementing AI solutions:

1. There is an often-reported shortage of skills in most markets across the globe in AI solution development. This results in people taking on roles temporarily within projects before returning to their “day jobs” at the conclusion of the project.

2. The average tenure of AI practitioners at Multichoice is significantly shorter than the average tenure of general staff at Multichoice, indicating that the likelihood of turnover during projects is higher for AI practitioners than for other practitioners – this again may lead to people taking on interim roles in projects to ensure completion of the work planned for a project.

© University of Pretoria

Original Identified Role	Project Role
Executive	Product Manager
Other Technical Professional	Data Scientist
Data Engineer	Software Engineer
Software Engineer	AI Engineer
AI Engineer	Software Engineer
Data Scientist	Product Manager
Other Non-Technical Professional	Business Stakeholder
Data Scientist	Other Technical Professional
Product Manager	Data Scientist
Business Stakeholder	Other Non-Technical Professional
Other Technical Professional	Product Manager

Table 7: Role changes within projects

What was the total duration of the project?

Responses for total project duration are shown in Table 5 below and was divided into four categories, ranging from less than three months to more than twelve months. While projects with less than three months duration were relatively less common, the rest of the categories were reasonably evenly distributed in the responses. In total, 65.9% of all AI solution development projects are reported to have duration of less than a year while only 34.1% of projects lasting from more than a year.

Total Project Duration	**Responses**	**Percentages**
Less than 3 months	5	11%
3 – 6 months	14	32%
6 – 12 months	10	23%
More than 12 months	15	34%
Total	44	100%

Table 8: Project duration for AI solution development projects

How long did you work on the project?

In Table 6 below, it is noted that thirteen of the forty-four responses (29.5%) differ between duration of the project and time worked on the project, suggesting that the respondents either joined the project after it started or didn't work on the project until it came to an end. Noting the shorter than average tenure of staff in AI roles versus those of general staff at Multichoice, it is important that this

© University of Pretoria

characteristic of staff working on AI solution development is considered when implementing any measures to manage AI ethics tensions. This consideration will be discussed further in Chapter Five.

Duration Worked on Project	Responses	Percentages
Less than 3 months	10	23%
3 – 6 months	14	32%
6 – 12 months	11	25%
More than 12 months	9	20%
Total	44	100%

Table 9: Duration worked on AI solution development projects

Which of the following responses best describe the primary business objective of this project?

Table 7 below shows that the most common business objectives of AI solution development projects are reported as customer retention, cost optimization and customer service. These responses account for 79.5% of all reported business objectives of AI solution development projects. While it may initially be surprising that revenue growth is not listed as one of the top three business objectives for AI projects, this is in line with several analyst reports which identify cost optimization and customer-related objectives as being more likely candidates for AI implementations.

Business Objective of the Project	Responses	Percentages
Process and governance	1	2%
Ethics of AI	1	2%
Fraud Identification	1	2%
Value added services	1	2%
Regulatory compliance	1	2%
Revenue growth	4	9%
Customer service improvement	9	20%
Cost optimization	10	23%
Customer retention	16	36%
Total	44	100%

Table 10: Business objective of AI solution development project

Which of the following responses best describe the focus of this project?

Table 8 below reports, somewhat surprisingly, that AI solution development projects are relatively evenly split between those that are internally facing and those that are customer facing. With AI being a new technology competence in most organisations, it is natural to assume that these projects would start off being mostly internally focussed to reduce risks of unintended impacts to customers, but respondents indicate that isn't the case.

© University of Pretoria

Project Focus	Responses	Percentages
Customer facing	19	43%
Internal	25	57%
Total	44	100%

Table 11: Focus of AI solution development projects

Which of the following responses best describe the outcome of this project?

In the table below, respondents report some relatively surprising findings. Firstly, less than seven percent of projects are reported as not having met objectives. This compares very favourably with generally reported statistics on the high failure rate attributed to traditional software development projects. For example, in the 2020 CHAOS report by the widely regarded Standish group, only 31% of modern software development projects are regarded as successful with 69% reported as either failed or significantly challenged.

Secondly, AI solution development projects are notorious for being easy to implement as POCs and pilots but rather more difficult to implement in production. However, respondents to the survey report that a full 61.36% of projects resulted in a full production deployment.

Project Outcome	Responses	Percentages
This project was limited to a pilot deployment to production and met this objective	2	5%
This project was limited to a pilot deployment to production but did not meet this objective	2	5%
This project was limited to a proof of concept or technology	12	27%
This project was meant to be implemented in production and met this objective	27	61%
This project was meant to be implemented in production but did not meet this objective	1	2%
Total	44	100%

Table 12: Outcomes of AI solution development projects

© University of Pretoria

Which of the following technologies and/or methods were used in this project?

For this question, respondents were provided with five default options ("computer vision", "natural language processing", "data science", "natural language understanding", unsure") as well as the opportunity to capture additional options not captured by the five default options. The responses to the five default options are captured in Table 13 below (note that a response in a particular category denotes that the technology was noted as being used, but not necessarily exclusively i.e., other technologies may also have been noted as being used on the same project). The finding from these responses is in line with what we expect to be reported – most AI solutions and tools being marketed today address data science and natural language problems. Most unstructured data available to corporates is in the form of text rather than images or videos, which may also explain why natural language use cases would be more common than computer vision use cases.

Technology Used	Responses	Percentages
Unsure	3	5%
Computer vision	5	9%
Natural language understanding	8	14%
Natural language processing	16	29%
Data science	24	43%
Total	56	100%

Table 13: Technologies used in AI solution developments

© University of Pretoria

In addition to the default choices, respondents were allowed to use a free text field to add other technologies that were used but did not fit into any of these categories. These technologies are listed in Table 14 below. It should be noted that many of the technologies listed below are not AI-specific technologies, reflecting the fact that some respondents appear to have taken a broader view of the question and reported back on both AI technologies as well as non-AI technologies used within projects that delivered AI solutions.

Other Technology Used (Free Text Field)
Cloud
Big data processing
Collaborative filtering
Mathematical modelling
User and asset clustering
Wagtail (Content Management System)
React
NodeJS
Postgres + MySQL
Feersum Engine
Recommendation(s)
Process and governance

Table 14: Other technologies used in AI solution developments

Of the 44 responses, three respondents were unsure of the technology used, 18 responses listed the use of a single technology, 11 responses listed the use of two technologies and five respondents listed the use of more than two technologies on the same project.

© University of Pretoria

Which response below would best characterize how the technology, used in the project, was developed or acquired?

In the table below, it is reported that the majority (63.64%) of AI solution development projects undertaken at Multichoice make use of either commercially available or open-source products or services. Given multiple reports which highlight the dominance of the USA and China in the AI landscape, it is unsurprising that there is a dependency on external products or services to implement AI solutions at Multichoice.

Technology Acquisition Characterisation	Responses	Percentage
Primarily made use of open-source products or services	5	11%
Primarily developed internally	16	36%
Primarily made use of commercial products or services	23	52%
Total	44	100%

Table 15: Technology acquisition characterisation for AI solution developments

Composition of the project team

The table below shows that more than 93% of all AI solution development projects are reported to have been implemented either with internal staff or with equal contribution from internal staff and consultants.

Project Team Composition	Responses	Percentage
Primarily consisted of consultants (external to the organisation)	3	7%
An even split between staff and consultants	12	27%
Primarily consisted of staff (internal to the organisation)	29	66%
Total	44	100%

Table 16: Project team composition for AI solution developments

© University of Pretoria

4.8 RESEARCH QUESTION TWO (RQ2)

RQ2: Which, if any, AI ethics tensions are present during the development of such AI applications?

The table below shows the aggregated frequencies of the prevalence of tensions reported by respondents to the research study.

			Tension type			
			Tension 1	**Tension 2**	**Tension 3**	**Tension 4**
Prevalence of the tension	Not encountered	Count	26	35	37	31
		Percentage	59.1%	79.5%	84.1%	70.5%
	Encountered once	Count	12	4	1	8
		Percentage	27.3%	9.1%	2.3%	18.2%
	Encountered more than once	Count	6	5	5	4
		Percentage	13.6%	11.4%	11.4%	9.1%
	This tension was only uncovered or acknowledged after the project was completed	Count	0	0	1	1
		Percentage	0	0.0%	2.3%	2.3%
	Total	Count	44	44	44	44
		Percentage	100.0%	100.0%	100.0%	100.0%

Table 17: Aggregated tension prevalence reported in the study

At a summary level, no tensions were encountered in 73.3% of the responses. At least one tension was encountered in 25.6% of responses and, in 1.1% of the responses, tensions were acknowledged after the completion of the project. It is important to note that tensions reported by respondents are based on the perceptions of the respondents and are not based on a ground truth that was established for each of the projects reported on by the respondents.

Prevalence of Individual tensions

The data shows that the tension between privacy and autonomy was encountered at least once by respondents across 40.9% of projects and was encountered more than once across 13.6% of all projects.

We also note that the tension between using algorithms to make decisions and ensuring fairness and equal treatment was encountered at least once across 20.5% of all projects and was encountered more than once across 11.4% of all projects.

The tension between increasing personalisation and enhancing solidarity and citizenship was encountered at least once in 15.9% of projects worked on by respondents, with this tension being encountered more than once across 11.4% of all projects worked on.

© University of Pretoria

Respondents encountered the tension between using automation to improve convenience and promoting dignity at least once across 29.5% of all projects, with the tension being encountered multiple times in a project 9.10% of the time.

4.9 RESEARCH QUESTION THREE (RQ3)

RQ3: How do individuals manage AI ethics tensions in current AI applications?

This section of the chapter will provide insights into how individuals manage AI ethics tensions in AI solutions being developed. Responses will cover various aspects of managing the tension ranging from how the tension was uncovered, to how it was managed and the respondents' perception of the effectiveness of managing the tension.

How were AI ethics tensions uncovered in projects?

Questions 19, 28, 37 and 46, which focused on how the four AI ethics tensions were uncovered, are discussed under this subsection. The table below shows an aggregated view of the responses across all four tensions.

			Tension type				
			Tension 1	Tension 2	Tension 3	Tension 4	Total
Uncovering tensions	The tension was not acknowledged during the project	Count	2	0	1	4	7
		Percentage	11.1%	0.0%	14.3%	30.8%	14.9%
	The tension was uncovered during a structured activity to identify ethical concerns, from within the project team	Count	9	3	3	7	22
		Percentage	50.0%	33.3%	42.9%	53.8%	46.8%
	The tension was uncovered during a structured activity to identify ethical considerations, from a party external to the project team	Count	1	1	0	0	2
		Percentage	5.6%	11.1%	0.0%	0.0%	4.3%
	The tension was uncovered serendipitously through a project team member	Count	6	4	3	2	15
		Percentage	33%	44.4%	42.9%	15.4%	31.9%
	The tension was uncovered serendipitously through an external party	Count	0	1	0	0	1
		Percentage	0%	11%	0%	0%	2.1%
	Total	Count	18	9	7	13	47
		Percentage	100.0%	88.9%	100.0%	100.0%	

Table 18: Discovery of tensions (aggregated view)

© University of Pretoria

It is encouraging to note that close to half of tensions were uncovered by project teams, engaged in a structured activity to identify ethical concerns. However, it is equally noteworthy that 34% of tensions were uncovered serendipitously – suggesting that much work is still to be done to embed ethical discovery processes into projects. A further 14.9% of tensions were not acknowledged at all during projects, highlighting the potential risk organisations face in ensuring that AI ethics considerations are adequately discovered and managed.

How were AI ethics tensions managed during a project?

Questions 20, 29, 38 and 47, which focus on how the four AI ethics tensions were managed, are discussed under this subsection. The table below shows an aggregated view of the responses across all four tensions.

			Tension type				
			Tension 1	Tension 2	Tension 3	Tension 4	Total
Managing tensions	I didn't encounter this tension	Count	0	0	1	0	1
		Percentage	0.0%	0.0%	12.5%	0.0%	2.1%
	I am not sure how the tension was managed	Count	3	1	1	4	9
		Percentage	16.7%	11.1%	12.5%	30.8%	18.8%
	This tension was managed by the team, with input from an external specialist team (such as legal, data management, data privacy, compliance, etc.)	Count	5	2	1	1	9
		Percentage	27.8%	22.2%	12.5%	7.7%	18.8%
	This tension was managed by the team, with no input from external specialists	Count	10	6	5	7	28
		Percentage	56%	66.7%	62.5%	53.8%	58.3%
	This tension was managed entirely by an external specialist team	Count	0	0	0	1	1
		Percentage	0%	0%	0%	8%	2.1%
	Total	Count	18	9	8	13	48
		Percentage	100.0%	100.0%	100.0%	100.0%	

Table 19: Process of managing tensions in projects (aggregated view)

In close to 60% of cases, AI ethics tensions were managed within project teams without explicit support from specialists outside of the development team while only 20.9% of AI ethics tensions were managed with assistance from specialist teams. While the literature suggests that teams would benefit from the inputs of AI ethicists, implementing this has potential cost implications for organisations. It is likely that these additional costs have not been factored into existing project budgets. This may be a practical obstacle to deeper involvement from AI ethics practitioners in managing AI ethics tensions in industry.

© University of Pretoria

How do you perceive the organisational support to understand the trade-off represented by the tension?

Questions 21, 30, 39 and 48, which focus on the level of organisational support provided to understand the trade-off represented by the tensions, are discussed under this subsection. The table below shows an aggregated view of the responses across all four tensions.

			Tension type				
			Tension 1	Tension 2	Tension 3	Tension 4	Total
Understanding the trade-offs	I didn't encounter this tension	Count	0	0	1	0	1
		Percentage	0.0%	0.0%	14.3%	0.0%	2.2%
	I am not sure how the tension was managed	Count	3	1	0	3	7
		Percentage	17.6%	11.1%	0.0%	23.1%	15.2%
	There was insufficient support to understand the trade-off represented by this tension	Count	2	1	1	2	6
		Percentage	11.8%	11.1%	14.3%	15.4%	13.0%
	There was sufficient support to understand the trade-off represented by this tension	Count	12	7	5	8	32
		Percentage	71%	77.8%	71.4%	61.5%	69.6%
	Total	Count	17	9	7	13	46
		Percentage	100.0%	100.0%	100.0%	100.0%	

Table 20: Understanding trade-offs (aggregated view)

It is encouraging that close to 70% of responses confirm that adequate support was provided to understand the trade-offs represented by the tensions reported.

© University of Pretoria

What was your perception of the outcome of managing the tension?

Questions 22, 31, 40 and 49, which focus on the outcome managing the tensions, are discussed under this subsection. The table below shows an aggregated view of the responses across all four tensions.

			Tension type				
			Tension 1	Tension 2	Tension 3	Tension 4	Total
Perceived outcome	I am not sure how the tension was managed	Count	2	1	0	1	4
		Percentage	11.1%	11.1%	0.0%	7.7%	8.5%
	The outcome was a good balance between benefiting the end user as well as the organisation	Count	4	1	5	7	17
		Percentage	22.2%	11.1%	71.4%	53.8%	36.2%
	The outcome was more beneficial to the end user than to the organisation	Count	7	3	2	2	14
		Percentage	38.9%	33.3%	28.6%	15.4%	29.8%
	The outcome was more beneficial to the organisation than to the end user	Count	5	4	0	3	12
		Percentage	28%	44.4%	0.0%	23.1%	25.5%
	Total	Count	18	9	7	13	47
		Percentage	100.0%	100.0%	100.0%	100.0%	

Table 21: Perceived outcome of managing the tension (aggregated view)

Responses are relatively evenly split between outcomes that favour the end user, outcomes that favour the organisation and outcomes that are well-balanced with both the end user and the organisation benefitting.

© University of Pretoria

What is your perception of the level at which the tension was managed?

Questions 23, 32, 41 and 50, which focus on the level at which the tensions were managed, are discussed under this subsection. The table below shows an aggregated view of the responses across all four tensions.

			Tension type				
			Tension 1	Tension 2	Tension 3	Tension 4	Total
Level of management of tension	I am not sure how the tension was managed / I didn't encounter this tension	Count	2	1	2	3	8
		Percentage	11.1%	11.1%	28.6%	23.1%	17.0%
	The tension was managed at an individual level	Count	5	3	1	2	11
		Percentage	27.8%	33.3%	14.3%	15.4%	23.4%
	This tension was managed at a team level	Count	0	5	4	4	13
		Percentage	0.0%	55.6%	57.1%	30.8%	27.7%
	This tension was managed at a management level	Count	11	0	0	4	15
		Percentage	61%	0.0%	0.0%	30.8%	31.9%
	Total	Count	18	9	7	13	47
		Percentage	100.0%	100.0%	100.0%	100.0%	

Table 22: Level at which tension was managed (aggregated view)

There is a relatively even split between tensions being managed at the individual, team and management level, with responses ranging between 23.4% and 31.9% for all three levels.

© University of Pretoria

What is your view of the appropriateness of the level at which the tension was managed?

Questions 24, 33, 42 and 51, which focus on the appropriateness of the level at which the tensions were managed, are discussed under this subsection. The table below shows an aggregated view of the responses across all four tensions.

			Tension type				
			Tension 1	Tension 2	Tension 3	Tension 4	Total
Appropriateness of level of management of tension	I am not sure how the tension was managed	Count	2	1	0	3	6
		Percentage	11.8%	11.1%	0.0%	23.1%	13.0%
	This tension should have been managed at a higher level in the hierarchy	Count	2	3	0	1	6
		Percentage	11.8%	33.3%	0.0%	7.7%	13.0%
	This tension was managed at an appropriate level	Count	13	5	7	7	32
		Percentage	76.5%	55.6%	100.0%	53.8%	69.6%
	This tension should have been managed at a lower level in the organisational hierarchy	Count	0	0	0	2	2
		Percentage	0%	0.0%	0.0%	15.4%	4.3%
	Total	Count	17	9	7	13	46
		Percentage	100.0%	100.0%	100.0%	100.0%	

Table 23: Appropriateness of the level at which tensions were managed (aggregated view)

Close to 70% of responses confirm that tensions were managed at an appropriate level. However, 13% of responses suggest that tensions were managed too low in the hierarchy and should have been managed at a higher level.

© University of Pretoria

What is your view of the organisational support to manage the tension?

Questions 25, 34, 43 and 52, which focus on the appropriateness of the level at which the tensions were managed, are discussed under this subsection. The table below shows an aggregated view of the responses across all four tensions.

			Tension type				
			Tension 1	Tension 2	Tension 3	Tension 4	Total
Organisational support to manage the tension	I am not sure how the tension was managed	Count	2	1	0	3	6
		Percentage	11.8%	11.1%	0.0%	20.0%	13.0%
	I had to use my own judgement to manage this tension	Count	2	3	0	1	6
		Percentage	11.8%	33.3%	0.0%	6.7%	13.0%
	Someone else on the project team had to use their own judgement to manage this specific tension	Count	13	5	7	7	32
		Percentage	76.5%	55.6%	100.0%	46.7%	69.6%
	We have a process in place to manage such tensions, but not specific supporting tools	Count	0	0	0	2	2
		Percentage	0%	0.0%	0.0%	13.3%	4.3%
	We have a process in place to manage such tensions, together with specific tools	Count	0	0	0	2	2
		Percentage	0%	0%	0%	13%	4.3%
	Total	Count	17	9	7	15	46
		Percentage	100.0%	100.0%	100.0%	100.0%	

Table 24: Organisational support to manage tensions (aggregated view)

In over 80% of cases, either the respondent or another team member had to use their judgement to manage the tension, without the aid of additional tools. Less than 10% of cases were resolved via a defined process to manage AI ethics considerations.

© University of Pretoria

What is your view of the effectiveness in managing the tension?

Questions 26, 35, 44 and 53, which focus on the effectiveness of managing the tensions, are discussed under this subsection. The table below shows an aggregated view of the responses across all four tensions.

			Tension type				
			Tension 1	Tension 2	Tension 3	Tension 4	Total
Effectiveness in managing tensions	I am not sure how the tension was managed / I did not encounter this tension	Count	1	2	2	2	7
		Percentage	5.6%	22.2%	28.6%	15.4%	14.9%
	Somewhat effective	Count	12	4	2	8	26
		Percentage	66.7%	44.4%	28.6%	61.5%	55.3%
	Very effective	Count	3	3	3	1	10
		Percentage	16.7%	33.3%	42.9%	7.7%	21.3%
	Not effective at all	Count	2	0	0	2	4
		Percentage	11%	0.0%	0.0%	15.4%	8.5%
	Total	Count	18	9	7	13	47
		Percentage	100.0%	100.0%	100.0%	100.0%	

Table 25: Effectiveness in managing tensions (aggregated view)

More than half of respondents report that processes to manage tensions were somewhat effective, while a further 21.3% report that processes to manage tensions were very effective. Less than 10% of respondents report that processes to manage tensions were ineffective.

© University of Pretoria

4.10 RESEARCH QUESTION FOUR (RQ4)

RQ4: Can we propose recommendations for an effective framework to manage AI ethics tensions?

While the first three research questions focused on providing details on the types of AI applications being developed at the case organisation, as well as the prevalence and management of AI ethics tension in the development of AI solutions, RQ4 is focused on providing recommendations for the creation of a framework to manage AI ethics tensions. In seeking to answer this research question, we will analyse the associations between data collected during this study. Specifically, the following associations will be explored:

- Association between demographics and tension prevalence

  Work experience versus tension prevalence

- Association between project characteristics and tension prevalence

  Project role versus tension prevalence

  Total project duration versus tension prevalence

- Association between tension management and perceived effectiveness of tension management

  Method of discovery, method of management, understanding trade-offs, perceived outcome, level of management, appropriateness of level of management and organisational support provided versus the perceived effectiveness in managing the tension.

- Association between tension management and perceived outcome of tension management

  Method of discovery, method of management, understanding trade-offs, level of management, appropriateness of level of management and organisational support provided versus the outcome of managing the tension.

© University of Pretoria

The associations are evaluated by performing a Chi-Square test. Although the Pearson Chi-Square test is the most widely used test for such associations, it is important to note that sample size limitations result in the Fisher-Freeman-Halton Exact test being more suited to this analysis. Specifically, we note that some cells have count below five and some cells also have a zero value. Combined with the overall sample size being 44 (below the recommended threshold for Pearson evaluation), the Fisher-Freeman-Halton exact test is reported on in the subsequent analyses.

4.10.1 Association between demographics and tension prevalence

We look at the association between years of work experience and tension prevalence to determine whether an individual's overall work experience plays a role in their experience of AI ethics tensions with AI solution development projects. The table below shows the results from the association testing – the p value obtained by the Fisher-Freeman-Halton Exact test is above 0.05 in all cases, denoting that there is no statistically significant association between years of work experience and tension prevalence, as reported by respondents, in AI solution development within the case organisation.

	Tension prevalence			
	Tension 1	Tension 2	Tension 3	Tension 4
Work experience	0.135	0.288	1.000	0.755

Table 26: Association between work experience and tension prevalence

4.10.2 Association between project characteristics and tension prevalence

We also looked at the association between project characteristics and tension prevalence. The two project characteristics explored were project role and project duration. Using the Fisher-Freeman-Halton Exact test, we tested the association between project characteristics and tension prevalence. As can be seen from the table below, there was no statistically significant association found between project characteristics and tension prevalence. It is important to note that tension prevalence is based on the reports from respondents but is not moderated by ground truth data for the projects being reported on.

	Tension prevalence

© University of Pretoria

	Tension 1	Tension 2	Tension 3	Tension 4
Project role	0.809	0.505	0.839	0.124
Project duration	0.236	1.000	1.000	0.875

Table 27: Association between project characteristics and tension prevalence

4.10.3 Association between tension management process and perceived effectiveness in managing the tension

In this section of the analysis, we look at the nature of the relationships between the ways in which tensions are managed (in terms of discovery, method of management, understanding trade-offs, organisational support, outcome, level of management and appropriateness of level of management) and the perceived effectiveness of the overall process of managing the tension.

Each of the cross tabulations is listed in Appendix E

Association between tension discovery and effectiveness of tension management

Appendix E1 details the crosstabulations for this evaluation. The Fisher-Freeman-Halton result shows a p-value of 0.105, which we interpret to mean that there is no statistically significant association between tension discovery method and the perceived effectiveness of the overall process to manage tensions. It is important to note that the discovery of tensions has not been moderated by ground truth data for projects that were reported on by respondents, and is based entirely on perceived prevalence of tensions.

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)	Exact Sig. (2-sided)
Pearson Chi-Square	18.053[a]	16	.321	.287
Likelihood Ratio	19.446	16	.246	.154
Fisher-Freeman-Halton Exact Test	22.200			.105
N of Valid Cases	47			

a. 23 cells (92.0%) have expected count less than 5. The minimum expected count is .02.

Table 28: Association between tension discovery and effectiveness of managing tensions

© University of Pretoria

Association between tension management procedure and effectiveness of tension management

Appendix E2 details the crosstabulations for this evaluation. As noted in the table below, the Fisher-Freeman-Halton result shows a p-value of 0.004, which we interpret to mean that there is a significant association between tension management procedure and the perceived effectiveness of the overall process to manage tensions.

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)	Exact Sig. (2-sided)
Pearson Chi-Square	28.197[a]	16	.030	.060
Likelihood Ratio	28.527	16	.027	.003
Fisher-Freeman-Halton Exact Test	30.391			.004
N of Valid Cases	47			

a. 23 cells (92.0%) have expected count less than 5. The minimum expected count is .02.

Table 29: Association between tension management procedure and effectiveness in managing tensions

We also consider whether the effect size is significant. Using Cramer's V, we obtain a value of 0.387 which is interpreted as a weak effect size.

Symmetric Measures

		Value	Approximate Significance	Exact Significance
Nominal by Nominal	Phi	.775	.030	.060
	Cramer's V	.387	.030	.060
N of Valid Cases		47		

Table 30: Effect size of association between understanding trade-offs and tension management effectiveness

© University of Pretoria

Association between level of support to understand trade-offs and effectiveness of tension management

Appendix E3 details the crosstabulations for this evaluation. As noted in the table below, the Fisher-Freeman-Halton result shows a p-value <0.001, which we interpret to mean that there is a significant association between level of support to understand trade-offs and the perceived effectiveness of the overall process to manage tensions.

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)	Exact Sig. (2-sided)
Pearson Chi-Square	35.320[a]	12	<.001	.006
Likelihood Ratio	35.672	12	<.001	<.001
Fisher-Freeman-Halton Exact Test	31.792			<.001
N of Valid Cases	47			

a. 18 cells (90.0%) have expected count less than 5. The minimum expected count is .04.

Table 31:Association between understanding trade-offs and effectiveness in managing tensions

We also consider whether the effect size is significant. Using Cramer's V, we obtain a value of 0.5 which is interpreted as a medium effect size.

Symmetric Measures

		Value	Approximate Significance	Exact Significance
Nominal by Nominal	Phi	.867	<.001	.006
	Cramer's V	.500	<.001	.006
N of Valid Cases		47		

Table 32: Effect size between understanding trade-offs and tension management effectiveness

© University of Pretoria

Association between outcome and effectiveness of tension management

Appendix E4 details the crosstabulations for this evaluation. As noted in the table below, the Fisher-Freeman-Halton result shows a p-value of 0.012, which we interpret to mean that there is a significant association between outcome of managing tensions and the perceived effectiveness of the overall process to manage tensions.

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)	Exact Sig. (2-sided)
Pearson Chi-Square	25.599[a]	12	.012	.011
Likelihood Ratio	24.640	12	.017	.014
Fisher-Freeman-Halton Exact Test	20.196			.012
N of Valid Cases	47			

a. 17 cells (85.0%) have expected count less than 5. The minimum expected count is .09.

Table 33: Association between outcome and effectiveness of tension management

We also consider whether the effect size is significant. Using Cramer's V, we obtain a value of 0.426 which is interpreted as a medium effect size.

Symmetric Measures

		Value	Approximate Significance	Exact Significance
Nominal by Nominal	Phi	.738	.012	.011
	Cramer's V	.426	.012	.011
N of Valid Cases		47		

Table 34: Effect size between outcome and tension management effectiveness

© University of Pretoria

Association between level of management and effectiveness of tension management

Appendix E5 details the crosstabulations for this evaluation. As noted in the table below, the Fisher-Freeman-Halton result shows a p-value of 0.04, which we interpret to mean that there is a significant association between level of management and the perceived effectiveness of the overall process to manage tensions.

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)	Exact Sig. (2-sided)
Pearson Chi-Square	28.793[a]	16	.025	.059
Likelihood Ratio	29.860	16	.019	.007
Fisher-Freeman-Halton Exact Test	27.539			.004
N of Valid Cases	47			

a. 23 cells (92.0%) have expected count less than 5. The minimum expected count is .04.

Table 35: Association between level of management and effectiveness in managing tensions

We also consider whether the effect size is significant. Using Cramer's V, we obtain a value of 0.391 which is interpreted as a weak effect size.

Symmetric Measures

		Value	Approximate Significance	Exact Significance
Nominal by Nominal	Phi	.783	.025	.059
	Cramer's V	.391	.025	.059
N of Valid Cases		47		

Table 36: Effect size between level of management and tension management effectiveness

© University of Pretoria

Association between appropriateness of level of management and effectiveness of tension management

Appendix E6 details the crosstabulations for this evaluation. As noted in the table below, the Fisher-Freeman-Halton result shows a p-value of <0.001, which we interpret to mean that there is a significant association between appropriateness of level of management and the perceived effectiveness of the overall process to manage tensions.

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)	Exact Sig. (2-sided)
Pearson Chi-Square	33.796[a]	12	<.001	.009
Likelihood Ratio	33.177	12	<.001	<.001
Fisher-Freeman-Halton Exact Test	28.476			<.001
N of Valid Cases	47			

a. 18 cells (90.0%) have expected count less than 5. The minimum expected count is .04.

Table 37: Association between appropriateness of level of management and effectiveness in managing tensions

We also consider whether the effect size is significant. Using Cramer's V, we obtain a value of 0.490 which is interpreted as a medium effect size.

Symmetric Measures

		Value	Approximate Significance	Exact Significance
Nominal by Nominal	Phi	.848	<.001	.009
	Cramer's V	.490	<.001	.009
N of Valid Cases		47		

Table 38: Effect size between appropriateness of level of management and tension management effectiveness

© University of Pretoria

4.10.4 Association between tension management and perceived outcome of tension management

Association between tension discovery and outcome of tension management

Appendix E7 details the crosstabulations for this evaluation. The Fisher-Freeman-Halton result below shows a p-value of 0.027, which we interpret to mean that there is a statistically significant association between tension discovery method and the perceived outcome of the tension management process.

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)	Exact Sig. (2-sided)
Pearson Chi-Square	20.164[a]	12	.064	.062
Likelihood Ratio	22.759	12	.030	.021
Fisher-Freeman-Halton Exact Test	18.738			.027
N of Valid Cases	47			

a. 16 cells (80.0%) have expected count less than 5. The minimum expected count is .09.

Table 39: Association between tension discovery and outcome of tension management

We also consider whether the effect size is significant. Using Cramer's V, we obtain a value of 0.378 which is interpreted as a weak effect size.

Symmetric Measures

		Value	Approximate Significance	Exact Significance
Nominal by Nominal	Phi	.655	.064	.062
	Cramer's V	.378	.064	.062
N of Valid Cases		47		

Table 40: Effect size between tension discovery and tension management outcome

© University of Pretoria

Association between tension management procedure and outcome of tension management

Appendix E8 details the crosstabulations for this evaluation. The Fisher-Freeman-Halton result below shows a p-value of 0.013, which we interpret to mean that there is a statistically significant association between tension management procedure and the perceived outcome of the tension management process.

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)	Exact Sig. (2-sided)
Pearson Chi-Square	26.644[a]	12	.009	.009
Likelihood Ratio	23.997	12	.020	.009
Fisher-Freeman-Halton Exact Test	20.331			.013
N of Valid Cases	47			

a. 17 cells (85.0%) have expected count less than 5. The minimum expected count is .09.

Table 41: Association between tension management procedure and tension management outcome

We also consider whether the effect size is significant. Using Cramer's V, we obtain a value of 0.435 which is interpreted as a medium effect size.

Symmetric Measures

		Value	Approximate Significance	Exact Significance
Nominal by Nominal	Phi	.753	.009	.009
	Cramer's V	.435	.009	.009
N of Valid Cases		47		

Table 42: Effect size between tension management procedure and tension management outcome

© University of Pretoria

Association between organisational support to understand trade-offs and outcome of tension management

Appendix E9 details the crosstabulations for this evaluation. The Fisher-Freeman-Halton result below shows a p-value of <0.001, which we interpret to mean that there is a statistically significant association between organisational support to understand trade-offs management procedure and the perceived outcome of the tension management process.

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)	Exact Sig. (2-sided)
Pearson Chi-Square	34.658[a]	9	<.001	<.001
Likelihood Ratio	30.244	9	<.001	<.001
Fisher-Freeman-Halton Exact Test	23.514			<.001
N of Valid Cases	47			

a. 13 cells (81.3%) have expected count less than 5. The minimum expected count is .17.

Table 43: Association between organisational support to understand trade-offs and tension management outcome

We also consider whether the effect size is significant. Using Cramer's V, we obtain a value of 0.496 which is interpreted as a medium effect size (table below).

Symmetric Measures

		Value	Approximate Significance	Exact Significance
Nominal by Nominal	Phi	.859	<.001	<.001
	Cramer's V	.496	<.001	<.001
N of Valid Cases		47		

Table 44: Effect size between organisational support to understand trade-offs and tension management outcome

© University of Pretoria

Association between level at which tension was managed and outcome of tension management

Appendix E10 details the crosstabulations for this evaluation. The Fisher-Freeman-Halton result below shows a p-value of 0.02, which we interpret to mean that there is a statistically significant association between level at which tension was managed and the perceived outcome of the tension management process.

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)	Exact Sig. (2-sided)
Pearson Chi-Square	33.305[a]	12	<.001	.002
Likelihood Ratio	26.111	12	.010	.013
Fisher-Freeman-Halton Exact Test	19.168			.020
N of Valid Cases	47			

a. 17 cells (85.0%) have expected count less than 5. The minimum expected count is .17.

Table 45: Association between level at which tension was managed and tension management outcome

We also consider whether the effect size is significant. Using Cramer's V, we obtain a value of 0.486 which is interpreted as a medium effect size (table below).

Symmetric Measures

		Value	Approximate Significance	Exact Significance
Nominal by Nominal	Phi	.842	<.001	.002
	Cramer's V	.486	<.001	.002
N of Valid Cases		47		

Table 46: Effect size between level at which tension was managed and tension management outcome

© University of Pretoria

Association between appropriateness of level at which tension was managed and outcome of tension management

Appendix E11 details the crosstabulations for this evaluation. The Fisher-Freeman-Halton result below shows a p-value of <0.01, which we interpret to mean that there is a statistically significant association between appropriateness of level at which tension was managed and the perceived outcome of the tension management process.

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)	Exact Sig. (2-sided)
Pearson Chi-Square	42.576[a]	9	<.001	<.001
Likelihood Ratio	32.475	9	<.001	<.001
Fisher-Freeman-Halton Exact Test	25.347			<.001
N of Valid Cases	47			

a. 13 cells (81.3%) have expected count less than 5. The minimum expected count is .17.

Table 47: Association between appropriateness of level at which tension was managed and tension management outcome

We also consider whether the effect size is significant. Using Cramer's V, we obtain a value of 0.550 which is interpreted as a medium effect size (table below).

Symmetric Measures

		Value	Approximate Significance	Exact Significance
Nominal by Nominal	Phi	.952	<.001	<.001
	Cramer's V	.550	<.001	<.001
N of Valid Cases		47		

Table 48: Effect size between appropriateness of level at which tension was managed and tension management outcome

© University of Pretoria

Association between organisational support to manage tension and outcome of tension management

Appendix E12 details the crosstabulations for this evaluation. The Fisher-Freeman-Halton result below shows a p-value of <0.01, which we interpret to mean that there is a statistically significant association between organisational support to manage tension and the perceived outcome of the tension management process.

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)	Exact Sig. (2-sided)
Pearson Chi-Square	32.683[a]	15	.005	.[b]
Likelihood Ratio	27.802	15	.023	.029
Fisher-Freeman-Halton Exact Test	21.756			.035
N of Valid Cases	47			

a. 23 cells (95.8%) have expected count less than 5. The minimum expected count is .09.

b. Cannot be computed because there is insufficient memory.

Table 49: Association between organisational support to manage tension and tension management outcome

We also consider whether the effect size is significant. Using Cramer's V, we obtain a value of 0.481 which is interpreted as a medium effect size (table below).

Symmetric Measures

		Value	Approximate Significance	Exact Significance
Nominal by Nominal	Phi	.834	.005	.[c]
	Cramer's V	.481	.005	.[c]
N of Valid Cases		47		

c. Cannot be computed because there is insufficient memory.

Table 50: Effect size between organisational support and tension management outcome

© University of Pretoria

4.11 SUMMARY

In this chapter, results of the research study are presented – the primary purpose of this chapter is to provide answers to the research questions posed earlier in the report. We provide answers to RQ1, RQ2 and RQ3 directly in this chapter. In addition to this, we provide results on the associations between various data collected in the study that will be used to summarize the recommendations for the creation of a framework to manage AI ethics tensions, in Chapter Five.

© University of Pretoria

5. CONCLUSIONS, EVALUATIONS AND RECOMMENDATIONS

5.1 INTRODUCTION

The aims of this research study were to identify and characterize the types of AI applications that are being developed within Multichoice as well as to identify and characterise the ethical tensions that arise while developing such AI applications. Furthermore, based on the data analysis, we intended to provide recommendations for the creation of an AI ethics tension management framework. The method to accomplish this was through a literature review, to understand both the current state of understand of AI ethics as well as to analyse the split of responsibility between individuals and organisations in the implementation of AI ethics practices, as well as a survey, to collect data on the current practices, experiences and perceptions of practitioners working on AI solution development.

This study addresses several important gaps in the existing literature. Firstly, there is a paucity of published information on the types of AI applications being developed in industry – this study aims to provide details on the types of AI applications being developed, the technologies used and the staffing methods for these AI solution developments. This is a valuable contribution to the existing body of knowledge regarding the types of AI applications that ethics considerations need to apply to. Secondly, there is very little to no empirical data on the types of AI ethics considerations practitioners are experiencing in AI solution development as well as how they are being equipped to address these considerations by the organisations they work for. This study aims to provide empirical data from a South African media organisation on how often these AI ethics considerations arise, how they are dealt with, and the perceived outcomes of the processes used to manage these AI ethics considerations. This study provides valuable first contributions in these areas in some cases, especially with regards to the experience and management of AI ethics tensions. Thirdly, this study aims to provide recommendations for the management of AI ethics tensions in practical applications – this will be a notable empirical contribution to the literature on AI ethics.

© University of Pretoria

5.2 FINDINGS FROM THIS STUDY

5.2.1 Findings from the literature review

Interest in AI ethics has never been greater

It was noted in the literature review that there has been a significant increase in the quantity of published literature in the area of AI ethics over the past four years. The recent, and highly publicised, release of first Chat GPT by Open AI and then Bard by Google, together with detailed write-ups on their shortcomings, has placed immense mainstream media attention on the role of AI ethics in the rapid advancement of capability in the AI landscape. While this provides a wonderful opportunity for AI researchers to have a voice in the development of AI technologies, it also means that the commercial stakes are much higher now and these place added burden on practitioners within AI solution development – see, as an example of the additional commercial stakes, the $100b stock market loss suffered by Google and attributed to failures by its Bard chatbot (Wittenstein, 2023) shortly after the launch of the product.

Codes of conduct may not serve the purpose intended

While empirical research has provided mixed results on the value of codes of conduct in influencing behaviour of practitioners, some researchers still recommend that tailored guidelines within organisational contexts may at least serve to provide alignment between stakeholders on how to manage AI ethics considerations or, at the very least, ensure that there is awareness of the types of AI ethics considerations that may be at play within AI solution development.

We need to distinguish between minimum compliance and ethical excellence

The argument against codes of conduct centres on the notion that setting up a code of conduct in other industries and professions (such as accounting) has resulted in a minimum compliance mentality rather than an ethical excellence mentality. The suggestion is that people should rather be asked to take open-ended accountability for ethics rather than be given a set of items to check off on a compliance list.

Practical steps can be taken to improve ethical standards in AI development

© University of Pretoria

Researchers have provided several recommendations concerning practical steps to take to improve addressing of AI ethics considerations. These include the involvement of a broad range of stakeholders, the including of different levels of management, the introduction of ethical deliberation, the introduction of checklists and the provision of comprehensive organisational support to assist practitioners to effectively address AI ethics considerations.

Managing ethical considerations is everyone's job

While AI/ML engineers are closest to the AI solution development process, the range of decisions required to conceptualize, build and deploy an AI solution necessitates the involvement of a wide range of stakeholders – decisions taken at an organisational, team and individual level all can impact on the ethical position of an AI implementation. Managing AI ethics considerations therefore cannot be left to an ethics officer or any other individual in the process – this needs to be a shared and ongoing responsibility.

Workers in AI ethics need assistance to effectively navigate AI ethics considerations in projects

It has been widely reported that AI ethics training is generally not included in the curriculum of engineering professionals – yet they are often at the frontlines of managing AI ethics considerations in projects. Comprehensive organisations support is required to assist AI workers to be more effective in managing AI ethics considerations – this extends from awareness, to training (among other topics, in ethical reasoning and analysis), to well-defined ethics processes and finally to expert assistance from ethicists who can bring different perspectives to ethically difficult scenarios.

5.2.2 Findings from the data analysis

Extensive analysis of data was performed on the responses received via the questionnaire, a summary of which will be noted in the subsequent paragraphs.

Descriptive statistics on the types of AI applications being worked on in Multichoice

In depth information was provided on the types of AI applications that are being developed at a South African media organisation, Multichoice. The types of AI solution development projects being worked

© University of Pretoria

on, the types of roles that were assigned to these projects, the duration of projects as well as the tenure of people involved in these projects was noted. Furthermore, information was uncovered on the types of technologies used in AI solutions, as well as the type of typical staffing modes used in these projects. All the aforementioned information provides valuable insights into the projects that drive the development of real-world applications of AI. This is a unique contribution to the literature from both a South African as well as pan-African perspective.

Association between demographics and tension prevalence

The association between work experience and reporting of AI ethics tensions was tested using the Fisher-Freeman-Halton Exact test. The resulting p-value was interpreted to mean that there was no significant association between work experience and reporting of AI ethics tensions on AI solution development projects. It is important to note that this finding is based on reports from respondents, without the benefit of an independent review of each project to validate the accuracy of responses. This finding emphasises the importance of AI ethics considerations being everyone's responsibility in AI solution development projects on an ongoing basis. Leaving ethics considerations to senior team members, team managers or executives, and typically, until the end of AI processes and adoption, may result in ethics considerations being missed in the AI solution development process.

Association between project characteristics and tension prevalence

The association between project characteristics and tension prevalence was also tested using the Fisher-Freeman-Halton Exact test. The resulting p-value was interpreted to mean that there was no significant association between project characteristics and reporting of AI ethics tensions on AI solution development projects. The project characteristics tested consisted of the duration of the project as well as the project role held by the respondent.

One take-away from this finding suggests that technical and non-technical project stakeholders are equally likely to report AI ethics tensions, lending further credence to the suggestions from researchers that AI ethics should be managed by a wide variety of stakeholders across disciplines and not just from a computer science or engineering background. Another take-away from this finding suggests that even short-duration AI solution development projects may pose AI ethics risks and organisations should be vigilant in ensuring that they adequately identify and address AI ethics risks no matter how small or large their AI project portfolio is. It is important to note that this finding is based on reports from respondents, without the benefit of an independent review of each project to validate the accuracy of responses.

© University of Pretoria

Association between tension management approach and tension management effectiveness

The association between tension management approach and tension management effectiveness was tested using the Fisher-Freeman-Halton Exact test. As mentioned earlier in Chapter Four, this test was used due to limitations of the Pearson Chi-Square test when dealing with small samples as well as when dealing with counts less than five in the sample data.

The resulting p-values and effect size calculated showed that there were several significant associative relationships within the study data. The table below provides a summary of the associations noted.

		Tension management effectiveness
Tension management process	Tension discovery	Not significant
	Tension management procedure	Highly significant, weak effect size
	Level of support to understand trade-offs	Highly significant, medium effect size
	Outcome of tension management process	Highly significant, medium effect size
	Level of management	Highly significant, weak effect size
	Appropriateness of level of management	Highly significant, medium effect size

Table 51: Significance of associations between tension management process and tension management effectiveness

The data indicates that several factors in the tension management process are associated with perceived effectiveness of the tension management process. While the way tensions are discovered does not associate with tension management effectiveness, tension management procedure, support to understand trade-offs, outcomes of the process either favouring the end user, the company or both, the level of management of the tension and the perceived appropriateness of the level at which the tension is managed, all associate with perceived tension management effectiveness.

Association between tension management approach and tension management outcomes

The association between tension management approach and perceived tension management outcome was tested using the Fisher-Freeman-Halton Exact test. As mentioned earlier in Chapter Four, this test was used due to limitations of the Pearson Chi-Square test when dealing with small samples as well as when dealing with counts less than five in the sample data.

The resulting p-values and effect size calculated showed that there were several significant associative relationships within the study data. The table below provides a summary of the associations noted.

© University of Pretoria

		Tension management outcome
Tension management process	Tension discovery	Highly significant, weak effect size
	Tension management procedure	Highly significant, medium effect size
	Level of support to understand trade-offs	Highly significant, medium effect size
	Level of management	Highly significant, medium effect size
	Appropriateness of level of management	Highly significant, medium effect size

Table 52:Associations between tension management process and tension management outcome

The data indicates that several factors in the tension management process are associated with the perceived outcome of managing the tension. This includes how a tension was discovered, tension management procedure, support to understand trade-offs, the level of management of the tension and the perceived appropriateness of the level at which the tension is managed.

This suggests that multiple touch points within the process used to detect and subsequently manage AI ethics tensions may play a role in the perceived outcome of managing those tensions i.e., does the outcome represent the best interests of the organisation, the end user of the application, or is it equally favourable to both the organisation as well as the end user.

5.3 CONCLUSIONS FROM THE FINDINGS

Findings from the literature review have informed parts of the questionnaire that was implemented in this study. It was noted that AI ethics tensions provide a unique challenge to AI workers, who may not necessarily have the skills required to understand and effectively manage these ethical considerations. It was suggested in the literature that AI ethics considerations may be best managed through a multi-intervention approach, considering the varied roles participating in AI projects, the different levels of management at which decisions are made in organisations, widely varying project roles, different project objectives, varied process maturity levels within the organisation, and different project delivery methodologies within projects. These factors, and their association with perceived

© University of Pretoria

effectiveness in managing AI ethics tensions, were explored in the current study via an in-depth data analysis.

Findings from the data analysis performed during this study suggest that demographic factors may play a limited role in the identification of AI ethics tensions on AI solution development projects. This supports the existing literature, where researchers have suggested that effective management of AI ethics considerations needs the contribution from practitioners from across disciplines and levels of hierarchy within organisations developing AI solutions. It is important to note that this finding is based on reports from respondents, without the benefit of an independent review of each project to validate the accuracy of responses.

Further findings indicate the project roles also play a limited role in indicating ability to detect AI ethics tensions, adding to the body of knowledge proposing that both technical and non-technical role players should be involved in managing AI ethics considerations. It was also found that project duration does not associate strongly with the AI ethics tension prevalence, lending support to the notion that projects of all sizes and durations are at risk of delivering solutions containing undiagnosed AI ethics tensions – it is not sufficient for organisations to only assess their largest, longest or most important projects for AI ethics tensions, this has to be done (on an ongoing basis through all stages of the system lifecycle) across all projects to ensure that the risk of unmanaged ethics considerations is minimized. It is important to note that, because responses were not validated against a ground truth of known project characteristics, there may be other factors which contribute to the lack of reporting of prevalence of AI ethics tensions on projects e.g., lack of ability of respondents to discern AI ethics tensions.

Associations have been noted between the tension management process and the perceived effectiveness of managing AI ethics tensions, as perceived by practitioners that have been involved in AI solution development projects. This suggests that organisations would be well served in defining clearly what their approach is to managing AI ethics considerations in AI solution development projects to ensure that they are effectively identifying and managing AI ethics tensions as well as other AI ethics considerations.

Associations have also been noted between the tension management process and the perceived outcomes of managing AI ethics tensions – keeping in mind that the outcome of the process is either detrimental or favourable to the end user. Organisations face a real risk of user backlash if they fail to identify and manage the AI ethics considerations appropriately and develop solutions which are later found to have disadvantaged their end users knowingly.

© University of Pretoria

5.4 RECOMMENDATIONS

From the findings of the literature review, as well as the subsequent findings from the data analysis based on the questionnaire responses, several recommendations can be made that should be considered when creating a framework to management AI ethics considerations such as AI ethics tensions in AI solution development projects – these are listed in the sub-sections below. It is important to note that these findings are based on reports from respondents, without the benefit of an independent review of each project to validate the accuracy of responses against an established ground truth.

5.4.1 All AI workers need to be included in the scope of the AI ethics framework developed

No significant association was found between years of work experience and experience of AI ethics tensions in AI solution development. This may suggest that even inexperienced workers are able to understand and detect AI ethics tensions in the projects they work on. However, it is important to note the limitation of this project that ground truth for tension prevalence on projects was not established – therefore, we need to consider that accuracy of responses on tension prevalence may also play a role in this association. Furthermore, no significant association was found between project role and AI ethics tensions which may suggest that both technical as well as non-technical roles are equally important in managing AI ethics considerations in organisations. Again, it is important to note that the absence of a ground truth on tension prevalence should temper the conclusion we draw from this finding. Respondent data, taken at face value, points to frameworks developed within an organisation to manage AI ethics tensions needing to encompass all AI workers and not just a subset such as engineers or computer scientists or just technical business units within the organisation. However, to ensure validity of the findings, further studies should extend the current study methodology to include ground truth data on the prevalence of AI ethics tensions.

5.4.2 All projects need to be covered within the scope of AI ethics frameworks

No significant association was found between project duration and the prevalence of AI ethics tensions. This suggests that both short and long duration projects are equally likely to present situations where ethical dilemmas arise and need to be dealt with – therefore, any framework that

© University of Pretoria

manages AI ethics should ensure that all projects undertaken within an organisation are included in the scope of an AI ethics framework. Assuming that only key projects, for example, should be subject to full ethical review, may lead to AI ethics tensions going unchecked and unmanaged across several other projects. It is important to note that this finding is based on reports from respondents, without the benefit of ground truth data for each project.

5.4.3 Ensure that the procedure to manage tensions is clearly defined

From the data analysis, we note that there is a significant association between the procedure to manage tensions and the perceived effectiveness in managing tensions. Organisations will be well served in ensuring that their internal procedures to identify and manage AI ethics tensions are well documented and that all staff are aware and trained on the procedures that apply to AI solution development initiatives.

5.4.4 Ensure that tensions are managed at the correct level

We also note from the data analysis that there exists a significant association between level at which tensions are managed and both the perceived effectiveness as well as the outcomes of managing these tensions. Considering this, organisations should ensure that not only are there clear guidelines on how tensions should be managed, but there should also be clear delineations between the various levels within the organisation and how each level is expected to manage AI ethics tensions. As reflected upon in the literature review, a wide scope of decisions is taken while bringing AI solutions to market and these decisions need to be taken at the appropriate level to ensure that 1) accountability is managed appropriately within the organisation and 2) people making decisions have the full decision-making context relevant to their level in the organisation before they provide inputs.

5.4.5 Ensure that teams are trained on how to assess the trade-offs between competing ethical values

Comprehensive organisational support has been noted, in the literature review, as a key success factor in managing AI ethics considerations. In the data analysis, organisational support to understand trade-offs is strongly associated with both perceived tension management outcomes as well as perceived tension management effectiveness. This forms one of the pillars of comprehensive organisational support to manage AI ethics considerations. It is therefore important that adequate support is

© University of Pretoria

provided to all staff members so that they can identify and effectively manage ethical considerations that may arise within the AI solution development projects they are working on.

5.5 CONCLUSION, LIMITATIONS AND AREAS FOR FURTHER RESEARCH

The moment for AI ethics in academic research is clearly now – and with the recent publicity surge around generative AI, the clamour to understand this nascent area of AI will only intensify. While much foundational work has been done to set the scene for various approaches to managing AI ethics, it remains clear that feedback from industry on how AI product development is managed ethically has been patently missing.

This study has provided a contribution to understanding how AI solutions are being developed in a media company in South Africa. Further studies are needed, not just in South Africa and across Africa, but world-wide and across industry boundaries to present more insights into how large enterprises are managing AI ethics considerations.

Through both the literature analysis, as well as the subsequent data analysis of the responses received, insights have been provided on practitioner views on how AI ethics tensions are identified and managed today, as well as the outcomes and effectiveness of these management actions. Furthermore, recommendations have been provided, based on the results of the analysis, that should inform the creation of AI ethics tension management frameworks in organisations.

While the study has provided new insights into the prevalence and management of AI ethics tensions, limitations and areas for further research should be noted. The current study used a cross-sectional survey design with purposive sampling – it would help answer generalizability if future studies could be extended across organisations and industries to determine whether the same conclusions are arrived at. To enhance the reliability of the study, it would also be a useful extension to the current study to survey multiple practitioners working on the same AI solution development project to determine whether the concepts of AI ethics tensions are equally well understood across practitioners both in the same organisation as well as across organisations and industries. This would require some element of anonymity being given up by both respondents as well as organisations participating in the study.

© University of Pretoria

It would also be useful to perform longitudinal studies to evaluate the impact of implementing the recommendations listed in the previous section, to better characterise and understand the management of AI ethics tensions in organisations.

AI ethics is an area of growing concern as well as attention – while there are many opportunities for organisations to remake themselves and their industries through the implementation of AI solutions, this needs to be done within a framework of comprehensive understanding and support for the ethical management of AI.

© University of Pretoria

REFERENCES

1. Bhattacherjee, A. (2012). *Social Science Research: Principles, Methods, and Practices*. University of South Florida, Digital Commons, Tampa Library. https://digitalcommons.usf.edu/oa_textbooks/3

2. Bringsjord, S and Govindarajulu, N.S. (2022). Artificial Intelligence*. The Stanford Encyclopedia of Philosophy (Summer 2020 Edition)*, Zalta, E.N. (ed.), https://plato.stanford.edu/archives/sum2020/entries/artificial-intelligence/

3. Canca, C. (2020) "Operationalizing AI Ethics Principles," Communications of the ACM, 63(12), pp. 18–21. https://doi.org/10.1145/3430368

4. Cohen, S. (2013) "Promoting Ethical Judgment in an Organisational Context," Journal of Business Ethics, 117(3), pp. 513–523.

5. Crampton, N., 2022, "Microsoft's framework for building AI systems responsibly", Microsoft. Date accessed:25 February 2023. <https://blogs.microsoft.com/on-the-issues/2022/06/21/microsofts-framework-for-building-ai-systems-responsibly/>

6. Creswell, John W. 2009. *Research Design Qualitative, Quantitative, and Mixed Methods Approaches*. Los Angeles: Sage Publications Inc.

7. Daniel, Johnnie. 2012. *Sampling Essentials Practical Guidelines for Making Sampling Choices*. Los Angeles: Sage Publications Inc.

8. Deloitte, 2020, "Talent and workforce effects in the age of AI", Deloitte Centre for Technology, Media and Telecommunications. Date accessed: 02 March 2023. < https://www2.deloitte.com/content/dam/insights/us/articles/6546_talent-and-workforce-effects-in-the-age-of-ai/DI_Talent-and-workforce-effects-in-the-age-of-AI.pdf>

9. Drost, E.A. 2011, "Validity and Reliability in Social Science Research", Education Research and Perspectives, vol. 38, no. 1, pp. 105-124.

10. European Commission, Ethics guidelines for trustworthy AI | Shaping Europe's digital future [WWW Document], n.d. URL https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai (accessed 6.2.22).

© University of Pretoria

11. Fjeld, J., Achten, N., Hilligoss, H., Nagy, A. and Srikumar, M., Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-Based Approaches to Principles for AI (January 15, 2020). Berkman Klein Center Research Publication No. 2020-1, Available at SSRN: https://ssrn.com/abstract=3518482 or http://dx.doi.org/10.2139/ssrn.3518482

12. Floridi, L. (2019). 'Translating Principles into Practices of Digital Ethics: Five Risks of Being Unethical', Philosophy & Technology, volume 32, p. 185-193;

13. Fowler, Floyd J. (2009). Survey Research Methods (4th ed.). Thousand Oaks, CA: SAGE Publications, Inc. Available at < https://doi.org/10.4135/9781452230184.> [Accessed 25 March 2023].

14. Greene, J. C., Caracelli, V. J., & Graham, W. F. (1989). Toward a Conceptual Framework for Mixed-Method Evaluation Designs. Educational Evaluation and Policy Analysis, 11(3), 255–274. https://doi.org/10.2307/1163620

15. Gogoll, J., Zuber, N., Kacianka, S., Greger, T., Pretschner, A., Nida-Rümelin, J., 2021. Ethics in the Software Development Process: from Codes of Conduct to Ethical Deliberation. Philos. Technol. 34, 1085–1108. https://doi.org/10.1007/s13347-021-00451-w

16. Harari, Y.N.(2017) 'Reboot for the AI Revolution', Nature, Vol 550, p. 324-327;

17. Hagendorff, T. (2020). 'The Ethics of AI Ethics: An Evaluation of Guidelines', Minds and Machines, volume 30, p. 99-120;

18. Jobin, A. Ienca, M. Vayena, E. (2019) 'The global landscape of AI ethics guidelines', Nature Machine Intelligence, volume 1, p. 389 – 399

19. Johnson, D.G. (2003). 'Computer Ethics' Floridi, L. (ed.)in *The Blackwell Guide to the Philosophy of Computing and information.* Chichester: John Wiley & Sons, Incorporated, pp. 65-75. Available from ProQuest Ebook Central (accessed 18 May 2020)

20. McNamara, A., Smith, J., & Murphy-Hill, E. (2018). Does ACM's code of ethics change ethical decision making in software development? In Proceedings of the 2018 26th ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering (pp. 729–733).

21. Mittelstadt, B.(2019) 'Principles alone cannot guarantee ethical AI', Nature Machine Intelligence, volume 1, p.501-507

© University of Pretoria

22. Mittelstadt, Brent and Wachter, Sandra and Russell, Chris, The Unfairness of Fair Machine Learning: Levelling down and strict egalitarianism by default (January 20, 2023). Available at SSRN: https://ssrn.com/abstract=4331652

23. Moor, J.H. (1985) 'What is Computer Ethics', Metaphilosophy, Vol 15, No.4, p.266-274;

24. Morley, J. Floridi, L. Kinsey, L. Elhalal, A.(2019) 'From What to How: An Initial Review of Publicly Available AI Ethics Tools, Methods and Research to Translate Principles into Practices', Science and Engineering Ethics, vol. 26 p. 2141-2168

25. Müller, Vincent C., "Ethics of Artificial Intelligence and Robotics", The Stanford Encyclopedia of Philosophy (Winter 2020 Edition), Edward N. Zalta (ed.), URL = https://plato.stanford.edu/archives/win2020/entries/ethics-ai/

26. Nedzhvetskaya, Nataliya, and J. S. Tan, "The Role of Workers in AI Ethics and Governance", in Justin B. Bullock and others (eds), The Oxford Handbook of AI Governance (online edition, Oxford Academic, 14 Feb. 2022), https://doi.org/10.1093/oxfordhb/9780197579329.013.68 , accessed 24 Feb. 2023.

27. OECD, State of Implementation of the OECD AI Principles: Insights from National AI Policies: Accessed from https://www.oecd-ilibrary.org/docserver/1cd40c44-en.pdf?expires=1627133509&id=id&accname=guest&checksum=366CD233838F111941BB91253E5AC2C8 on 24 July 2021

28. Perkowitz, S., 2021. The bias in the machine: Facial recognition technology and racial disparities. MIT Case Studies in Social and Ethical Responsibilities of Computing https://doi. org/10.21428/2c646de5, 62272586(5), p.15.

29. Pichai, S., 2018, "AI at Google: our principles", Google. Date accessed:25 February 2023. <https://blog.google/technology/ai/ai-principles/>

30. Pickard, AJ 2017, Research Methods in Information, Facet Publishing, London. Available from: ProQuest Ebook Central. [2 March 2023].

31. Russell, Stuart J., et al. 2010, Artificial Intelligence: A Modern Approach. 3rd ed. Upper Saddle River, NJ, Prentice Hall.

32. Schwartz, M. The Nature of the Relationship between Corporate Codes of Ethics and Behaviour. Journal of Business Ethics 32, 247–262 (2001). https://doi-org.uplib.idm.oclc.org/10.1023/A:1010787607771

© University of Pretoria

33. Sossin, L & Smith, CW 2003, 'Hard Choices and Soft Law: Ethical Codes, Policy Guidelines and the Role of the Courts in Regulating Government', Atlanta Law Review, vol. 40, no. 1, pp. 867-894

34. Taylor, Steve, et al., 2018, "Responsible AI – Key Themes, Concerns & Recommendations for European Research and Innovation: Summary of Consultation with Multidisciplinary Experts", June. [Taylor, et al. 2018 available online]

35. Whittlestone, J. Nyrup, R. Alexandrova, A. Cave, S. (2019) 'The role and limits of principles in AI ethics: towards a focus on tensions', Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society, p195-200

36. Whittlestone, J. Nyrup, R. Alexandrova, A. Dihal, K. Cave, S. (2019) *Ethical and societal implications of algorithms, data, and artificial intelligence: a roadmap for research.* London: Nuffield Foundation.

37. Winfield Alan F. T. and Jirotka Marina. (2018). Ethical governance is essential to building trust in robotics and artificial intelligence systemsPhil. Trans. R. Soc. A.376: 20180085, https://doi.org/10.1098/rsta.2018.0085.

38. Winfield, Alan, Katina Michael, Jeremy Pitt, and Vanessa Evers (eds.), 2019, *Machine Ethics: The Design and Governance of Ethical AI and Autonomous Systems*, special issue of *Proceedings of the IEEE*, 107(3): 501–632.

39. Wittenstein, J., 2023," A Factual Error by Bard AI Chatbot Just Cost Google $100 Billion", Time. Date accessed 10 March 2023. < https://time.com/6254226/alphabet-google-bard-100-billion-ai-error/>

40. Wolf, M. J. et al. (2021) "Ai Ethics: A Long History and a Recent Burst of Attention," Computer, 54(1), pp. 96–102. doi: 10.1109/MC.2020.3034950.

41. World Economic Forum, 2023, "Generative AI: a game-changer that society and industry need to be ready for", World Economic Forum. Date accessed: 24 February 2023. <https://www.weforum.org/agenda/2023/01/davos23-generative-ai-a-game-changer-industries-and-society-code-developers/>

42. Paul Chamness Iida. *(2020). The Concise APA Handbook: APA 7th Edition: Vol. APA 7th edition*. Information Age Publishing.

© University of Pretoria

APPENDICES

APPENDIX A – RESEARCH ETHICS COMMITTEE APPROVAL LETTER

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Faculty of Humanities
Fakulteit Geesteswetenskappe
Lefapha la Bomotho

Humanities 100. Since 1919

4 August 2021

Dear Mr M Velayudan

Project Title:	Managing AI Ethics Tensions: A Case Study at Multichoice
Researcher:	Mr M Velayudan
Supervisor(s):	Prof EB Ruttkamp-Bloem
Department:	Philosophy
Reference number:	26473586 (HUM005/0421)
Degree:	Masters

I have pleasure in informing you that the above application was **approved** by the Research Ethics Committee on 29 July 2021. Data collection may therefore commence.

Please note that this approval is based on the assumption that the research will be carried out along the lines laid out in the proposal. Should the actual research depart significantly from the proposed research, it will be necessary to apply for a new research approval and ethical clearance.

1. The Letters of introduction and informed consent must be presented on a department of Philosophy letterhead, that also indicates the department's contact details.
2. The letters must state that the data will be stored in a password-protected format at the Department of Philosophy for a minimum of 15 years.
3. The Committee notes the permission letter from Multichoice. The letter however does not explicitly state that the name of the organisation can be used in the research. The organisation name is already reflected in the research title. It is advised that a revised permission letter be obtained to this effect.

We wish you success with the project.

Sincerely,

Prof Innocent Pikirayi
Deputy Dean: Postgraduate Studies and Research Ethics
Faculty of Humanities
UNIVERSITY OF PRETORIA
e-mail: PGHumanities@up.ac.za

© University of Pretoria

APPENDIX B – LETTER OF PERMISSION FROM MULTICHOICE

MultiChoice
ENRICHING LIVES

From:
Multichoice Support Services (Pty) Ltd
P.O. Box 1502
Randburg, 2125
South Africa

22 April 2021

To:
Research Ethics Committee
University of Pretoria

Re: Permission to conduct research at Multichoice

This letter serves as confirmation that Mergandran Velayudan has requested permission, and been granted such permission, to conduct research at Multichoice Support Services (Pty) Ltd, related to the research study entitled "A case study of AI Ethics at Multichoice", as a post graduate student at the University of Pretoria.

This permission includes the necessary access to:

1. Identify qualifying projects that meet the research criteria
2. Identify and contact staff members who meet the research criteria
3. Send out questionnaires to staff members who meet the research criteria
4. Perform interviews of staff members who meet the research criteria
5. Make relevant observations related to the research study

It is further noted that all responses provided by staff members will be captured anonymously and that no personally identifiable information will be captured in either research reports or subsequent academic submissions.

Brad Eliot
Group CIO
Multichoice Group

DStv GOtv showmax DStv Media Sales irdeto

© University of Pretoria

APPENDIX C – LETTER OF INTRODUCTION AND INFORMED CONSENT

(Double-click to open full pdf document)

professionals to be more effective in the management of ethical AI tensions.

Use of participant data

Participant data will only be used in aggregate form to analyze and report on findings. No personally-identifiable information will be captured or processed - there will be complete anonymity for all responses provided. This will apply to the initial research report as well as to any subsequent academic publications of the research findings.

Withdrawal without Prejudice

Participation in this study is voluntary; refusal to participate will involve no penalty. You are free to withdraw consent and discontinue participation in this project at any time without prejudice or penalty. You are also free to refuse to answer any question we might ask you.

Further Questions and Follow-Up

You are welcome to ask the researchers any questions that occur to you during the survey or interview. If you have further questions once the interview is completed, you are encouraged to contact the researchers using the contact information given below.

For any questions regarding the University of Pretoria Ethics Clearance Process for post-graduate research, please contact the Faculty of Humanities at the University of Pretoria.

Mergan Velayudan
mergan.velayudan@gmail.com
+27 76 423 9507

* Required

Section 1 - Informed Consent

1. By selecting the "Agree" option, I grant my permission for participation in this survey. I hereby confirm that I have understood the purpose of this research. *

- ◯ Agree
- ◯ Disagree

2. By selecting the "Agree" option, I give consent to my responses being used anonymously in this research report and in any subsequent academic publications that stem from this research report and findings *

- ◯ Agree
- ◯ Disagree

© University of Pretoria

APPENDIX D – RESEARCH QUESTIONNAIRE

(Double-click to open full pdf document)

Managing AI Ethics Tensions: A Case Study at Multichoice

Study Title: Managing AI Ethics Tensions: A Case Study at Multichoice

Researchers:
M Velayudan
Prof E. Ruttkamp-Bloem
Dr. H. Robertson

Before agreeing to participate in this research, we strongly encourage you to read the following explanation of this study. This statement describes the purpose and procedures of the study. Also described is your right to withdraw from the study at any time. This study has been approved by the Research Ethics Committee of the Faculty of Humanities at the University of Pretoria.

Explanation of Procedures

This study is designed to examine the ways in which ethical AI tensions are managed in typical AI developments in a media organisation. We are conducting this study to learn more about this question since it has not been studied much in the past. Participation in the study involves completion of a survey that asks you basic questions about yourself as well as the relevant work you have undertaken, and may include a face-to-face interview, which will last for approximately thirty minutes to one hour. The interviews will be conducted by one of the researchers, recorded and later transcribed for the purpose of data analysis.

Benefits

The anticipated benefit of participation is the opportunity to discuss your opinions, concerns and actions related to the management of ethical AI tensions, and to contribute to understanding of decision-making and actions during the process of developing AI solutions, as it specifically relates to the management of ethical AI tensions. There is no financial or other remunerative benefit to participating in this survey and neither is there any obligation to participate in this research study. Should you wish participate in this study, you are also free to withdraw participation at any point subsequently.

Confidentiality

The information gathered during this study will remain confidential at all times, both during this project and after conclusion of this project. Only the researchers will have access to the study data and information. There will not be any identifying names on the surveys or interview transcripts; they will be coded and the key to the code will be kept securely by the researcher. Your names and any other identifying details will never be revealed in any publication of the results of this study. The recordings will be destroyed at the completion of the study. The results of the research will be published in the form of a research paper and may be published in a professional journal or presented at professional meetings. The knowledge obtained from this study will be of great value in guiding

© University of Pretoria

APPENDIX E - CROSSTABULATIONS

Appendix E1: Discovery of tensions x effectiveness in managing tensions

Which response below best describes how this tension was uncovered? * Which response below best describes your opinion of the effectiveness of the approach to manage this tension? Crosstabulation

			Which response below best describes your opinion of the effectiveness of the approach to manage this tension?					
			I am not sure how this tension was managed	I didn't encounter this tension	Not effective at all	Somewhat effective	Very effective	Total
Which response below best describes how this tension was uncovered?	The tension was not encountered or acknowledged in this project	Count	2	0	2	3	0	7
		% of Total	4.3%	0.0%	4.3%	6.4%	0.0%	14.9%
	The tension was uncovered during a structured activity to identify ethical concerns, from within the project team	Count	2	0	0	16	4	22
		% of Total	4.3%	0.0%	0.0%	34.0%	8.5%	46.8%
	The tension was uncovered during a structured activity to identify ethical considerations, from a party external to the project team	Count	0	0	0	1	1	2
		% of Total	0.0%	0.0%	0.0%	2.1%	2.1%	4.3%
	The tension was uncovered serendipitously through a project team member	Count	2	1	2	6	4	15
		% of Total	4.3%	2.1%	4.3%	12.8%	8.5%	31.9%
	The tension was uncovered serendipitously through an external party	Count	0	0	0	0	1	1
		% of Total	0.0%	0.0%	0.0%	0.0%	2.1%	2.1%
Total		Count	6	1	4	26	10	47
		% of Total	12.8%	2.1%	8.5%	55.3%	21.3%	100.0%

Appendix E2: Tension management procedure x effectiveness in managing tensions

Which response below best describes how this tension was managed when it was encountered? * Which response below best describes your opinion of the effectiveness of the approach to manage this tension? Crosstabulation

			Which response below best describes your opinion of the effectiveness of the approach to manage this tension?					
			I am not sure how this tension was managed	I didn't encounter this tension	Not effective at all	Somewhat effective	Very effective	Total
Which response below best describes how this tension was managed when it was encountered?	I am not sure how this tension was managed	Count	4	0	3	0	1	8
		% of Total	8.5%	0.0%	6.4%	0.0%	2.1%	17.0%
	I didn't encounter this tension	Count	0	0	0	1	0	1
		% of Total	0.0%	0.0%	0.0%	2.1%	0.0%	2.1%
	This tension was managed by the team, with input from an external specialist team (such as legal, data management, data privacy, compliance, etc.)	Count	0	0	0	6	3	9
		% of Total	0.0%	0.0%	0.0%	12.8%	6.4%	19.1%
	This tension was managed by the team, with no input from external specialists	Count	2	1	1	18	6	28
		% of Total	4.3%	2.1%	2.1%	38.3%	12.8%	59.6%
	This tension was managed entirely by an external specialist team	Count	0	0	0	1	0	1
		% of Total	0.0%	0.0%	0.0%	2.1%	0.0%	2.1%
Total		Count	6	1	4	26	10	47
		% of Total	12.8%	2.1%	8.5%	55.3%	21.3%	100.0%

© University of Pretoria

Appendix E3: Organisational support to understand trade-offs x effectiveness in managing tensions

Which response best describes your opinion of the level of support provided by your organisation to understand the trade-off represented by the tension? * Which response below best describes your opinion of the effectiveness of the approach to manage this tension? Crosstabulation

			Which response below best describes your opinion of the effectiveness of the approach to manage this tension?					
			I am not sure how this tension was managed	I didn't encounter this tension	Not effective at all	Somewhat effective	Very effective	Total
Which response best describes your opinion of the level of support provided by your organisation to understand the trade-off represented by the tension?	I am not sure how this tension was managed	Count	4	0	2	1	0	7
		% of Total	8.5%	0.0%	4.3%	2.1%	0.0%	14.9%
	I didn't encounter this tension	Count	1	0	0	0	1	2
		% of Total	2.1%	0.0%	0.0%	0.0%	2.1%	4.3%
	There was insufficient support to understand the trade-off represented by this tension	Count	0	0	2	4	0	6
		% of Total	0.0%	0.0%	4.3%	8.5%	0.0%	12.8%
	There was sufficient support to understand the trade-off represented by this tension	Count	1	1	0	21	9	32
		% of Total	2.1%	2.1%	0.0%	44.7%	19.1%	68.1%
Total		Count	6	1	4	26	10	47
		% of Total	12.8%	2.1%	8.5%	55.3%	21.3%	100.0%

Appendix E4: Outcome x effectiveness in managing tensions

Which response best describes your opinion of the outcome of the process of managing this tension? * Which response below best describes your opinion of the effectiveness of the approach to manage this tension? Crosstabulation

			Which response below best describes your opinion of the effectiveness of the approach to manage this tension?					
			I am not sure how this tension was managed	I didn't encounter this tension	Not effective at all	Somewhat effective	Very effective	Total
Which response best describes your opinion of the outcome of the process of managing this tension?	I am not sure how this tension was managed	Count	2	0	2	0	0	4
		% of Total	4.3%	0.0%	4.3%	0.0%	0.0%	8.5%
	The outcome was a good balance between benefiting the end user as well as the organisation	Count	1	0	0	10	6	17
		% of Total	2.1%	0.0%	0.0%	21.3%	12.8%	36.2%
	The outcome was more beneficial to the end user than to the organisation	Count	2	0	0	10	2	14
		% of Total	4.3%	0.0%	0.0%	21.3%	4.3%	29.8%
	The outcome was more beneficial to the organisation than to the end user	Count	1	1	2	6	2	12
		% of Total	2.1%	2.1%	4.3%	12.8%	4.3%	25.5%
Total		Count	6	1	4	26	10	47
		% of Total	12.8%	2.1%	8.5%	55.3%	21.3%	100.0%

© University of Pretoria

Appendix E5: Level of support and effectiveness in managing tensions

Which response best describes the level at which the tension was managed from an organisational hierarchy perspective? * Which response below best describes your opinion of the effectiveness of the approach to manage this tension? Crosstabulation

			Which response below best describes your opinion of the effectiveness of the approach to manage this tension?					
			I am not sure how this tension was managed	I didn't encounter this tension	Not effective at all	Somewhat effective	Very effective	Total
Which response best describes the level at which the tension was managed from an organisational hierarchy perspective?	I am not sure how this tension was managed	Count	3	0	2	1	0	6
		% of Total	6.4%	0.0%	4.3%	2.1%	0.0%	12.8%
	I didn't encounter this tension	Count	1	0	0	1	0	2
		% of Total	2.1%	0.0%	0.0%	2.1%	0.0%	4.3%
	This tension was managed at a management level, outside the scope of the team	Count	1	0	1	3	0	5
		% of Total	2.1%	0.0%	2.1%	6.4%	0.0%	10.6%
	This tension was managed at a team level	Count	0	0	1	16	6	23
		% of Total	0.0%	0.0%	2.1%	34.0%	12.8%	48.9%
	This tension was managed at an individual practitioner level	Count	1	1	0	5	4	11
		% of Total	2.1%	2.1%	0.0%	10.6%	8.5%	23.4%
Total		Count	6	1	4	26	10	47
		% of Total	12.8%	2.1%	8.5%	55.3%	21.3%	100.0%

Appendix E6: Appropriateness of level of management x effectiveness in managing tensions

Which response best describes your opinion about the appropriateness of the level at which this tension was managed? * Which response below best describes your opinion of the effectiveness of the approach to manage this tension? Crosstabulation

			Which response below best describes your opinion of the effectiveness of the approach to manage this tension?					
			I am not sure how this tension was managed	I didn't encounter this tension	Not effective at all	Somewhat effective	Very effective	Total
Which response best describes your opinion about the appropriateness of the level at which this tension was managed?	I am not sure how this tension was managed	Count	4	0	2	0	0	6
		% of Total	8.5%	0.0%	4.3%	0.0%	0.0%	12.8%
	This tension should have been managed at a higher level in the hierarchy	Count	0	0	2	4	1	7
		% of Total	0.0%	0.0%	4.3%	8.5%	2.1%	14.9%
	This tension should have been managed at a lower level in the organisational hierarchy	Count	0	0	0	2	0	2
		% of Total	0.0%	0.0%	0.0%	4.3%	0.0%	4.3%
	This tension was managed at an appropriate level	Count	2	1	0	20	9	32
		% of Total	4.3%	2.1%	0.0%	42.6%	19.1%	68.1%
Total		Count	6	1	4	26	10	47
		% of Total	12.8%	2.1%	8.5%	55.3%	21.3%	100.0%

© University of Pretoria

Appendix E7: Tension discovery and outcome of managing tensions

Crosstab

			Which response best describes your opinion of the outcome of the process of managing this tension?				
			I am not sure how this tension was managed	The outcome was a good balance between benefiting the end user as well as the organisation	The outcome was more beneficial to the end user than to the organisation	The outcome was more beneficial to the organisation than to the end user	Total
Which response below best describes how this tension was uncovered?	The tension was not encountered or acknowledged in this project	Count	2	3	0	2	7
		% of Total	4.3%	6.4%	0.0%	4.3%	14.9%
	The tension was uncovered during a structured activity to identify ethical concerns, from within the project team	Count	0	10	9	3	22
		% of Total	0.0%	21.3%	19.1%	6.4%	46.8%
	The tension was uncovered during a structured activity to identify ethical considerations, from a party external to the project team	Count	0	0	2	0	2
		% of Total	0.0%	0.0%	4.3%	0.0%	4.3%
	The tension was uncovered serendipitously through a project team member	Count	2	4	3	6	15
		% of Total	4.3%	8.5%	6.4%	12.8%	31.9%
	The tension was uncovered serendipitously through an external party	Count	0	0	0	1	1
		% of Total	0.0%	0.0%	0.0%	2.1%	2.1%
Total		Count	4	17	14	12	47
		% of Total	8.5%	36.2%	29.8%	25.5%	100.0%

© University of Pretoria

Appendix E8: Tension management procedure x outcome of managing tensions

Crosstab

			Which response best describes your opinion of the outcome of the process of managing this tension?				
			I am not sure how this tension was managed	The outcome was a good balance between benefiting the end user as well as the organisation	The outcome was more beneficial to the end user than to the organisation	The outcome was more beneficial to the organisation than to the end user	Total
Which response below best describes how this tension was managed when it was encountered?	I am not sure how this tension was managed	Count	4	2	0	2	8
		% of Total	8.5%	4.3%	0.0%	4.3%	17.0%
	I didn't encounter this tension	Count	0	1	0	0	1
		% of Total	0.0%	2.1%	0.0%	0.0%	2.1%
	This tension was managed by the team, with input from an external specialist team (such as legal, data management, data privacy, compliance, etc.)	Count	0	3	3	3	9
		% of Total	0.0%	6.4%	6.4%	6.4%	19.1%
	This tension was managed by the team, with no input from external specialists	Count	0	10	11	7	28
		% of Total	0.0%	21.3%	23.4%	14.9%	59.6%
	This tension was managed entirely by an external specialist team	Count	0	1	0	0	1
		% of Total	0.0%	2.1%	0.0%	0.0%	2.1%
Total		Count	4	17	14	12	47
		% of Total	8.5%	36.2%	29.8%	25.5%	100.0%

Appendix E9: Level of support to understand trade-offs x outcome of managing tensions

Crosstab

			Which response best describes your opinion of the outcome of the process of managing this tension?				
			I am not sure how this tension was managed	The outcome was a good balance between benefiting the end user as well as the organisation	The outcome was more beneficial to the end user than to the organisation	The outcome was more beneficial to the organisation than to the end user	Total
Which response best describes your opinion of the level of support provided by your organisation to understand the trade-off represented by the tension?	I am not sure how this tension was managed	Count	4	2	0	1	7
		% of Total	8.5%	4.3%	0.0%	2.1%	14.9%
	I didn't encounter this tension	Count	0	0	1	1	2
		% of Total	0.0%	0.0%	2.1%	2.1%	4.3%
	There was insufficient support to understand the trade-off represented by this tension	Count	0	2	0	4	6
		% of Total	0.0%	4.3%	0.0%	8.5%	12.8%
	There was sufficient support to understand the trade-off represented by this tension	Count	0	13	13	6	32
		% of Total	0.0%	27.7%	27.7%	12.8%	68.1%
Total		Count	4	17	14	12	47
		% of Total	8.5%	36.2%	29.8%	25.5%	100.0%

© University of Pretoria

Appendix E10: Level of management x outcome of managing tensions

Crosstab

			Which response best describes your opinion of the outcome of the process of managing this tension?				
			I am not sure how this tension was managed	The outcome was a good balance between benefiting the end user as well as the organisation	The outcome was more beneficial to the end user than to the organisation	The outcome was more beneficial to the organisation than to the end user	Total
Which response best describes the level at which the tension was managed from an organisational hierarchy perspective?	I am not sure how this tension was managed	Count	4	2	0	0	6
		% of Total	8.5%	4.3%	0.0%	0.0%	12.8%
	I didn't encounter this tension	Count	0	1	1	0	2
		% of Total	0.0%	2.1%	2.1%	0.0%	4.3%
	This tension was managed at a management level, outside the scope of the team	Count	0	2	1	2	5
		% of Total	0.0%	4.3%	2.1%	4.3%	10.6%
	This tension was managed at a team level	Count	0	9	8	6	23
		% of Total	0.0%	19.1%	17.0%	12.8%	48.9%
	This tension was managed at an individual practitioner level	Count	0	3	4	4	11
		% of Total	0.0%	6.4%	8.5%	8.5%	23.4%
Total		Count	4	17	14	12	47
		% of Total	8.5%	36.2%	29.8%	25.5%	100.0%

Appendix E11: Appropriateness of level of management x outcome of managing tensions

Crosstab

			Which response best describes your opinion of the outcome of the process of managing this tension?				
			I am not sure how this tension was managed	The outcome was a good balance between benefiting the end user as well as the organisation	The outcome was more beneficial to the end user than to the organisation	The outcome was more beneficial to the organisation than to the end user	Total
Which response best describes your opinion about the appropriateness of the level at which this tension was managed?	I am not sure how this tension was managed	Count	4	1	0	1	6
		% of Total	8.5%	2.1%	0.0%	2.1%	12.8%
	This tension should have been managed at a higher level in the hierarchy	Count	0	1	1	5	7
		% of Total	0.0%	2.1%	2.1%	10.6%	14.9%
	This tension should have been managed at a lower level in the organisational hierarchy	Count	0	2	0	0	2
		% of Total	0.0%	4.3%	0.0%	0.0%	4.3%
	This tension was managed at an appropriate level	Count	0	13	13	6	32
		% of Total	0.0%	27.7%	27.7%	12.8%	68.1%
Total		Count	4	17	14	12	47
		% of Total	8.5%	36.2%	29.8%	25.5%	100.0%

© University of Pretoria

Appendix E12: Organisational support to manage the tension x outcome of managing tensions

Crosstab

			Which response best describes your opinion of the outcome of the process of managing this tension?				
			I am not sure how this tension was managed	The outcome was a good balance between benefiting the end user as well as the organisation	The outcome was more beneficial to the end user than to the organisation	The outcome was more beneficial to the organisation than to the end user	Total
Which response best describes the organisation's support provided to manage this tension?	I am not sure how this tension was managed	Count	4	2	1	1	8
		% of Total	8.5%	4.3%	2.1%	2.1%	17.0%
	I didn't encounter this tension	Count	0	0	0	1	1
		% of Total	0.0%	0.0%	0.0%	2.1%	2.1%
	I had to use my own judgement to manage this tension	Count	0	6	4	6	16
		% of Total	0.0%	12.8%	8.5%	12.8%	34.0%
	Someone else on the project team had to use their own judgement to manage this specific tension	Count	0	3	1	3	7
		% of Total	0.0%	6.4%	2.1%	6.4%	14.9%
	We have a process in place to manage such tensions, but not specific supporting tools	Count	0	5	5	1	11
		% of Total	0.0%	10.6%	10.6%	2.1%	23.4%
	We have a process in place to manage such tensions, together with specific tools	Count	0	1	3	0	4
		% of Total	0.0%	2.1%	6.4%	0.0%	8.5%
Total		Count	4	17	14	12	47
		% of Total	8.5%	36.2%	29.8%	25.5%	100.0%

© University of Pretoria