Gordon Institute of Business Science
University of Pretoria

Trust in artificial intelligence for its adoption and use in organisational decision-making

23022907

A research project submitted to the Gordon Institute of Business Science, University of Pretoria, in partial fulfilment of the requirements for the degree of Master of Philosophy (Corporate Strategy).

25th November 2024

Abstract

Considering the recent developments and mainstream attention on Artificial Intelligence, organisations are facing increased pressure to realise the potential benefits which this new generation of tools and techniques seek to unlock. However, to responsibly leverage and benefit from the advantages which AI promises to offer, those responsible for decision-making in organisations need to be willing to trust the technology. For these reasons, this qualitative research study focused on trust in artificial intelligence for its adoption and use in organisational decision-making, and the key related concepts of explainable AI (xAI) and transparency.

The theoretical relevance of this research was to develop insights into, and new understanding of how trust in AI is formed for decision-making in organisations, as well as to reveal new insights and understanding of the relationship between the key concepts of xAI and transparency which lead to trust in AI.

The study followed a qualitative, exploratory design with a phenomenological approach, to explore the lived experiences of the research phenomena from the perspective of individuals responsible for organisational decision-making. A total of 19 semi-structured interviews were conducted, with participants who were exposed to or had experience of Artificial Intelligence and its impact on their organisations. The participants were drawn from a setting of worldwide organisations, across 16 diverse sectors, from healthcare and financial services to defence and aviation. Rich insights and understanding of the main theoretical concepts and research phenomena were revealed through a systematic, thematic analysis.

Several similarities were identified between the findings of the study and the literature, adding to the theoretical body of knowledge, while eight nuances of difference provided potential refinements, and three distinct differences highlighted potential extensions. Lastly, a conceptual framework was developed and refined through each stage of the research, culminating in a view of the potential research contributions in relation to extant theory.

The research outcomes extended the theoretical understanding of trust formation in AI, for its adoption and use in organisational decision-making environments, while leading to actionable insights for organizations aiming to build trust in AI technologies.

Keywords:

Artificial Intelligence, Trust in AI, explainable AI; xAI; Transparency; Adoption and Use

Declaration

I declare that this research project is my own work. It is submitted in partial fulfilment of the requirements for the degree of Master of Philosophy in Corporate Strategy at the Gordon Institute of Business Science, University of Pretoria. It has not been submitted before for any degree or examination in any other University. I further declare that I have obtained the necessary authorisation and consent to carry out this research.

23022907 23rd November 2024

List of Figures

List of Tables

Table of Contents

Chapter 1: Introduction to the Research Problem

This qualitative research study focused on trust in artificial intelligence (AI) for its adoption and use for decision-making in organisations. The business and theoretical relevance, as well as the problem, are well summarised by the following quotation from literature:

"*Given that AI offers tremendous potential for industry, organizations are keen to know how to reap the benefits of AI systems. This can only be achieved if users are willing to trust an artificial agent*" (Sullivan et al., 2022, p. 542) .

Through a systematic review of extant literature from top academic journals, key concepts were identified and explored as relevant to trust engenderment, namely explainable AI (xAI) and Transparency. Through investigation of the scholarship discussions surrounding these concepts, areas for further research were identified as having potential to provide additional insights and understanding as to how this trust in AI is formed.

1.1 Background: Business Relevance

The World Economic Forum (2024) explained that for organisations to unlock AI's potential, trust in AI is a key requirement, and that we are currently in an era of foundation-building to ensure AI's responsible adoption and use. Gartner (2024) found that a lack of trust in AI is a factor presenting a "significant obstacle to AI success and value realisation" (p. 1), while the OECD cited trust in AI as one of the conditions which may promote favourable use of AI, but also noted that "AI adoption is still limited and uneven across firms and sectors" (OECD, 2024, p. 5). As part of the OECD's recommended *values-based principles*, subsequently endorsed by the European Union, trustworthy AI is outlined as a foundational principle informing recommendations for policymakers (OECD, 2024, p. 59) in the "development and adoption of AI" (OECD, 2024, p. 5).

In a global study report (17 countries and 17,193 respondents) published by KPMG, in collaboration with the university of Queensland (Gillespie et al., 2023), several interesting insights are presented. Firstly, 61% of people surveyed were unwilling to trust AI and the study found a high positive correlation between trust in AI and its acceptance (p. 13). Secondly, the OECD "*trustworthy AI management and governance principles"* were endorsed by 96-99% of respondents, and of these eight (8) principles, transparency and explainability of AI were one grouping (Gillespie et al., 2023, p. 40). This speaks to the relevance of the concepts focused on in the proposed research.

Related to the above discussion of trust in AI, several additional points supporting business relevance of explainable AI (xAI) for AI adoption in decision-making are worth mentioning. Firstly, in a recent cross-industry survey report of 1684 participants, McKinsey &

Company (2023, p. 6) show that *explainability* is in the top-5 generative AI-related risks, with *inaccuracy* being number 1. McKinsey & Company (2023) also found that “more than two thirds of respondents expect their organisations to increase their AI investment over the next three years” (p. 19). In addition, OECD (2024) states that “AI solutions may not be explainable, impacting accountable evidence-based decision-making” (p. 20), and goes further to explain that the lack of understanding of “black-box” AI algorithms can “grow to a lack of trust” (p. 20). Lastly, to demonstrate economic importance, GlobalData (2023, p. 3) forecasts growth for the global AI market of 295.3% from 2022, to USD323.3Billion by 2027. The above evidence shows relevance to the key theoretical concepts of this research, and an urgency for further insights and understanding at a practical level, while demonstrating a significant global market for AI which organisations may seek to benefit from.

1.2 The Research Problem: Theoretical Relevance

The primary focus of this exploratory research centred around the problem that without trust, AI would encounter resistance to its adoption in organisations, which is a key interest to both “practitioners and academics”, (Vanneste & Puranam, 2024, p. 13). Additionally, Glikson & Woolley (2020) highlighted that there remained a lack of understanding surrounding transparency as a contributor to emotional trust in AI, and there was a paucity of studies on trust in AI and its adoption and use at an organisational level identified by both Wong et al. (2024) and Enholm et al. (2021). The scholarship also called for an understanding of how these trust issues could be overcome to facilitate AI’s use in decision-making (Shrestha et al., 2019), and it remained uncertain for organisations how they would be able to realise the cross-industry contribution which AI stands to make (Sullivan et al., 2022). The above points created an urgency for further exploration, understanding and insights at a theoretical level.

Related to the above problem surrounding trust in AI were the concepts of *xAI* and *transparency*, as they were widely regarded across the scholarship as being foundational to trust engenderment, as evidenced by three systematic literature reviews (Haque et al., 2023; Laato et al., 2022; Glikson & Woolley, 2020). It is still worth noting, however, that the concept of xAI was still new in academia and had only begun garnering mainstream attention in scholarship since 2018 (Arrieta et al., 2019) and had not yet been fully explored.

1.2.1 Trust in AI:

The literature review of Trust in AI revealed an important contextual progression from interpersonal trust to trust in technology, and ultimately to trust in AI, establishing foundational links between trust, organizational behaviour, and decision-making (McAllister, 1995; Hoff & Bashir, 2015; Wang et al., 2016; Glikson & Woolley, 2020). Building on this foundation, trust in AI was recognized as important for its adoption and use in organizations. Scholars similarly

identified a lack of trust as a significant barrier to its adoption for AI-based decision-making (Enholm et al., 2021; Wang & Ding, 2024; Sabharwal et al., 2024). However, in addition to this recognition, an area remained underexplored that while cognitive trust in AI had been studied extensively, the influence of transparency on emotional trust had been less examined (Glikson & Woolley, 2020; Wang et al., 2016).

Insights from the literature also revealed that further research at the organizational level was needed to clarify how trust influences AI adoption and use across varied contexts (Wong et al., 2024; Enholm et al., 2021; Chen et al., 2021). Here, explainable AI (xAI) emerged as a potential enabler, suggesting the importance of xAI's role in bridging trust gaps and fostering broader AI adoption within organizations (Wang et al., 2016; Weber et al., 2023), speaking to its importance as a central concept in this research. Finally, a focus on data-driven decision-making underscored the relevance of xAI, transparency, and trust as being central to effective organizational decision-making processes (Sabharwal et al., 2024; Enholm et al., 2021). Several areas for further research were identified and are highlighted in the Chapter 2 section conclusion, from which aspects of the research questions were derived.

1.2.2 Explainable AI (xAI):

The literature highlighted the "black-box" nature of AI as a significant barrier to AI adoption, due to its inscrutability (Berente et al., 2021) and lack of transparency (Rai, 2020), which limits its use in decision-making contexts. This common concern and need to understand the processing steps of the "black-box" becomes particularly important in regulated fields where sector-specific standards require explainability (Weber et al., 2023). To overcome these concerns, Rai (2020) suggested that xAI could improve user engagement, while Berente et al. (2021) emphasized explainability as important for engendering trust. A need for standardised metrics and guidelines for evaluating effectiveness of xAI methods also emerged, with a call for further research by Silva et al. (2023) and Haque et al. (2023).

Given this broad application potential, a cross-disciplinary call for extending xAI research into varied application domains (Rai, 2020; Sabharwal et al., 2024) further supported the suggestion that xAI holds promise for fostering trust and enhancing decision-making in diverse sectors. Furthermore, an understanding of the factors which contribute to trust in AI are still needed to build a foundation for additional practical and academic development of xAI (Pumplun et al., 2023; Balasubramanian et al., 2022; Sullivan et al., 2022; Abedin, 2022; Lukyanenko et al., 2022). Finally, researchers extended the exploration of xAI to include social and managerial dimensions of trust, with Berente et al. (2021) suggesting that studies should examine not only cognitive aspects but also how social trust factors impact organizational decision-making. This view was similar to the cognitive and emotional trust dimensions previously discussed by Glikson and Woolley (2020).

1.2.3 Transparency:

The scholarship underscored transparency as a cornerstone of AI trust research, establishing a foundational role for transparency in fostering trust in AI (Laato et al., 2022; Haque et al., 2023; Glikson & Woolley, 2020). Laato et al. (2022) identified that a lack of transparency negatively impacts "trustworthiness," emphasizing transparency's role in fostering reliability and user confidence. Sabharwal et al. (2024) further highlighted that "transparency is likely to be the foundation of trust in AI" (p. 6) adding that human values should be central to the human connection to AI technologies. In broader terms, transparency was viewed as essential not only for trust in AI, but also as a requirement for AI adoption (Choung et al., 2022).

Regarding transparency for AI decision-making, a lack of transparency and understanding were noted by Lindebaum et al. (2020) as a limitation of AI which could become a liability for decision-making (p. 256). With the introduction of data protection regulations, Kim et al. (2023) recommended further research to promote higher "standards of transparency and accountability" to foster judicious decision-making (p. 1306). Additionally, Vanneste & Puranam (2024) identified transparency as an intervention to clarify the reasoning and understanding underpinning an algorithm's decisions.

Transparency's role in building both cognitive and emotional trust was another important theme identified by Glikson & Woolley (2020). The scholars identified transparency as a contributor to cognitive trust while noting that its influence on emotional trust remained unexplored. Building on the previous point, Glikson & Woolley (2020) and Sullivan et al. (2022) emphasized a need for further research into transparency as an emotional dimension of trust in AI. They suggested that understanding how transparency affects emotional trust in AI could provide deeper insights into the dynamics of trust in AI and also in addressing the "uncanny feeling".

1.2.4 xAI and Transparency:

The relationship between xAI and transparency was interpreted in varying ways, with some scholars viewing them as interchangeable concepts while others proposed a causal link between them. Rai (2020) and Gregor (2024) similarly discussed the concepts of transparency and explainability (in relation to xAI) interchangeably and did not draw significant distinction from one to the other. On the other hand, several scholars viewed transparency as an outcome or an implication of explanation (Kim et al., 2023; Haque et al., 2023; Laato, 2022), while Sabharwal et al. (2024) suggested that explainability of AI is caused by transparency. Additionally, Glikson & Woolley, (2020) provided another perspective, that explainability is an "aspect" of transparency. Importantly from the review of xAI and transparency, Kim et al.

(2024) identified a need for interdisciplinary theoretical development in xAI outcomes, including within “organizational science” (p. 1306), to deepen understanding across different disciplines.

These perspectives revealed some ambiguity in the understanding of both concepts, highlighting a need for further research to clarify whether transparency leads to explainability or vice versa. Despite these variations, both xAI and transparency were portrayed as sharing the overarching goal of transforming AI from “black box” to “glass box” systems, fostering trust and confidence by making AI more interpretable and reliable (Rai, 2020).

1.2.5 Summary:

Through the sections and sub-sections of the Chapter 2 literature review, a conceptual framework of the scholarship landscape surrounding trust in AI, xAI, transparency, AI decision-making and the adoption and use of AI in organisations was unveiled. What became apparent through the review of the extant literature was that each of these concepts was interconnected and could not be viewed in isolation from one another, or exclusively.

This was demonstrated by the three SLRs (Glikson & Woolley, 2020; Laato et al., 2022; Haque et al., 2023) in combination with the additional literature presented. The conceptual framework which was developed is provided in Figure 7 for clarity, and to draw together the individual concepts from the preceding sections to highlight the identified relationships and key insights between them.

1.3 The Research Question(s)

The research questions were formulated through an overarching understanding of the needs for additional research observed across the reviewed literature, found in the respective section conclusions, 2.2.4, 2.3.4 and 2.4.4. The questions were formulated to generate additional insights and understanding in the areas of overlap between trust in AI, xAI and transparency, to broaden the foundational knowledge of their interaction (Pumplun et al., 2023; Balasubramanian et al., 2022; Sullivan et al., 2022; Abedin, 2022; Lukyanenko et al., 2022). These research needs were highlighted through an analysis and interpretation of the literature related to each presented concept of Trust in AI, xAI and Transparency, and the connections between them.

Research Question 1: How does trust in AI lead to its adoption and use for organisational decision-making? (Vanneste & Puranam, 2024; Sullivan et al., 2022; Wong et al., 2024; Enholm et al., 2021)

Research Question 2: How does xAI engender trust in AI for decision-making in organisations? (Rai, 2020; Sabharwal et al., 2024; Lukyanenko et al., 2022) And, by

Extension: ***What are the factors that contribute to trust in AI?*** (Sullivan et al., 2022; Balasubramanian et al., 2022; Abedin, 2022; Glikson & Woolley, 2020; Lukyanenko et al., 2022; Pumplun et al., 2023)

Research Question 3: How does transparency influence emotional trust in AI for organisational decision-making? (Glikson & Woolley, 2020; Wang & Ding, 2024; Wang et al., 2016; Berente et al., 2021)

1.4 Aims of the Research

The overarching aims of this research are to develop insights and new understanding of trust in AI for decision-making in organisations, as well as to reveal new insights and understanding of the relationship between the key concepts of xAI and transparency which lead to trust in AI. Therefore, the aims as they relate to the research questions are:

The aims of RQ1 are to develop new insights and understanding into the mechanisms of trust formation in AI in organisations, and to therefore develop new insights as to how trust in AI leads to adoption and beneficial use of AI for organisational decision-making.

The aims of RQ2 are firstly to explore and understand how xAI contributes to the formation of trust in AI for organisational decision-making. Secondly, to explore additional insights and understanding of the factors which the individual responsible for organisational decision-making finds relevant to engendering their trust in AI for decision-making.

The aim of RQ3 was to gain additional insights and understanding of transparency as an emotional dimension of trust in AI, in the context of the study around organisational decision-making.

An additional aim of the research was to develop a conceptual framework as an outcome of the overall study.

1.5 Potential Contributions of the Research

This research study aimed to extend the current understanding and to provide additional insights into trust in artificial intelligence for its adoption and use in organisational decision-making. The apparent research contributions identified are presented as potential refinements to the extant literature (Crane et al., 2016). These potential contributions are summarised below in three categories: potential additions to the body of knowledge, potential refinements to the body of knowledge, and potential extensions to the body of knowledge.

1.5.1 Potential Additions to Theory

On the basis of the research conducted in this study, and as discussed in Section 7.2, and summarised in Section 7.3.1, several apparent areas of similarity to the literature were revealed, which are proposed as potential **additions** to the body of knowledge. Namely:

- Trust takes time to develop, which also speaks to the importance of change management for trust in AI.
- AI should inform and enhance human decision-making, not replace it. Therefore, AI is most effective and trusted as a supportive tool to enhance decision-making, which must also be overseen by humans.
- There was an awareness and shared concern of the “Black-Box” nature of AI. Transparency, understanding and explainability were all recognised to build trust and to overcome this “Black-Box” concern.
- Explainability should address “How” AI arrives at outcomes, which creates an understanding of “How” AI works and how it processes data. This was viewed as important for building trust in AI.
- Fear is a barrier to adoption, which is an emotional dimension of trust in AI and is primarily related to job security.
- AI has limited ability to identify human nuance, which inhibits trust.
- Transparency and Understanding are closely related and are directly linked to trust in AI for decision-making. xAI and Transparency both drive understanding, which promotes trust in AI decision-making.

1.5.2 Potential Refinements of Theory

On the basis of the research conducted in this study, and as discussed in Section 7.2, and summarised in Section 7.3.2, eight apparent nuances of difference to the literature were revealed, which are proposed as potential **refinements** to the body of knowledge. Namely:

- Trust exists as a spectrum or on a scale, and is not a simple uni-variate concept, but potentially consists of many independent variables, which contribute to overall trust development.
- Trust potentially develops at different rates in different industries, which emphasises the potential need for tailored explanations and transparency requirements, specific to the particular use-case.
- Gradual AI adoption and experimentation is potentially important for creating trust in AI.
- Human verification of AI outputs is needed, which potentially demonstrates reliability, and aids in developing trust.

- Explanation must be consistent throughout the data lifecycle, to demonstrate integrity of inputs, processing techniques within the model, and also the resulting outputs. Explanations and xAI design should potentially focus on the full AI system, not just inputs or outputs.
- AI Self-Criticism is potentially necessary for trust and confidence in AI systems for decision-making. If an AI can point to areas where it may be wrong or inaccurate, this would potentially engender improved trust from the user.
- Transparency is required to potentially alleviate fear and anxiety (strong human emotions) and to foster trust.
- There are potentially different priorities for transparency, depending on the role of the individual responsible for decision-making within the organization.

1.5.3 Potential Extensions to Theory

On the basis of the research conducted in this study, and as discussed in Section 7.2, and summarised in Section 7.3.2, three apparent distinct differences to the literature were revealed, which are proposed as potential **extensions** to the body of knowledge. Namely:

- There was a tension between the decision-making groups regarding “blindly-trusting” AI. Some supported it while others rejected it. This tension reveals that while some groups are inclined to trust AI without reservation, others strongly reject such an approach, highlighting a potential divide in comfort levels and perceived risks associated with AI trust for adoption and use.
- It is potentially important to contain the possible negative outcomes of AI use for decision-making, to engender trust. If it can be demonstrated that consequences can be easily remedied or contained within a “blast-radius”, it would potentially engender greater trust in AI and facilitate adoption and use for decision-making.
- Provenance and Data Integrity are potentially both antecedents to trust in AI and its subsequent adoption. Demonstrating prior successes and suitability of AI for an application or purpose may potentially engender greater trust. Likewise, demonstrating the integrity of the data being used may also engender greater trust in AI.

1.5.4 Potential New Insights and Understanding for the Research Questions

1.5.4.1 Research Question 1: How does trust in AI lead to its adoption and use for organisational decision-making?

The potential new insights and understanding which were revealed from the potential additions, refinements and extensions to theory, in answering RQ1 are summarised below:

- When trust is developed gradually, is included in a change management process, and acts as a support tool for individuals responsible for decision-making, it potentially leads to its adoption and use for organisational decision-making.
- Understanding that trust potentially exists on a spectrum and develops at different rates in different industries provides potentially new insights into how trust in AI leads to adoption and use for organisational decision-making.
- Developing trust gradually and using humans to verify its outputs also potentially creates trust in AI, which can lead to adoption and use for organisational decision-making.
- Trusting that AI can contain its possible negative impacts potentially engenders trust for its adoption and use for organisational decision-making, while blindly trusting AI can lead to adoption, but may also cause unintended consequences.

1.5.4.2 Research Question 2: How does xAI engender trust in AI for decision-making in organisations? And: What are the factors that contribute to trust in AI?

The potential new insights and understanding which were revealed from the potential additions, refinements and extensions to theory, in answering RQ2 are summarised below:

- Not just xAI, but also transparency and understanding, all collectively address the “black-box” and explain how AI arrives at outcomes, thereby engendering trust in AI for decision-making in organisations.
- There is a seemingly circular/cross-supportive relationship between xAI, transparency and understanding, which potentially creates a clearer picture of the “interplay” / circularity / cross-support between the three.
- Consistent explanation of data and its integrity throughout the AI lifecycle, demonstration of provenance of AI and creation of self-critical AI explanations, are all factors which potentially engender trust for decision-making in organisations.

1.5.4.3 Research Question 3: How does transparency influence emotional trust in AI for organisational decision-making?

The potential new insights and understanding which were revealed in answering RQ3 are summarised below:

- Transparency and understanding are closely related and potentially alleviate the human emotions of fear and anxiety, which are primarily related to job-security, thus influencing emotional trust in AI for organisational decision-making.
- AI has a limited ability to identify human nuance, which might potentially be alleviated by transparency, allowing the connection of humans to how AI operates.

1.5.5 Conceptual Framework Development

As one of the aims of this study, a conceptual framework was also developed, edited and refined as the research progressed. The final version of the conceptual framework is presented in section 7.2.4. as an outcome of the research. A larger A4 version is included in **Appendix J**.

1.6 Scope of the Research

1.6.1 Theoretical Scope

The theoretical scope of the study was limited to the literature on trust in AI, xAI and transparency within an organisational context, and was focused on decision-making and adoption. However, due to the exploratory nature of qualitative studies, additional insights in related areas were also revealed, for example, conjoined agency and change management. Although machine learning, big data and data-driven decision-making are all acknowledged as important aspects of AI, the research questions and aims were centred on trust in AI, transparency and explainability (xAI) more broadly.

1.6.2 Physical Scope

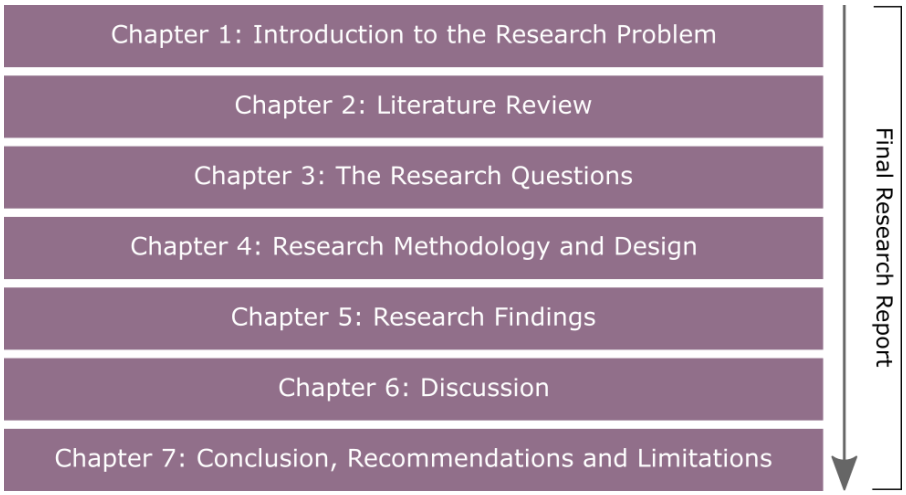
The physical scope of the research was worldwide organisations, across diverse sectors. The final sectors covered by the research were: Business consulting, Chemical industry, telecommunications, pharmaceutical manufacturing, transport & logistics, engineering consulting, legal advisory, retail and consumer, defence, aviation, transport & rail infrastructure, human resource management, software development, healthcare, retail banking and investment banking. The scope was not geographically limited due to the global accessibility of AI and was also defined by the literature which calls for exploration in new contexts (Abedin, 2022) and domains (Rai, 2020; Haque et al., 2023; Glikson & Woolley, 2020; Rabiee et al., 2024). Organisations based in South Africa, Australia, Norway, the Netherlands, the United Kingdom and the United States were included in the research study.

1.7 Research Report Overview

This research report follows the GIBS “Purple Pages” recommendations for the structure and layout and is therefore presented over seven (7) chapters, as per Figure 1.

Figure 1:

Research Report Overview



Note: Author’s own.

Chapter 1 covers a summary introduction of the conducted research, including its business and theoretical relevance, which is followed by a detailed literature review in Chapter 2, of the key concepts relevant to the research. From this review of the literature, the research questions and aims were developed and are articulated in Chapter 3. An overview of the proposed methodological approach and design, which was employed to answer the research questions and to address their aims, is then given in Chapter 4. Next, the findings of the research are presented in Chapter 5, followed by a discussion of them relative to the extant literature in Chapter 6. In conclusion, Chapter 7 presents the potential theoretical contributions of the research and a related final conceptual framework (Figure 30). This is followed by management and stakeholder recommendations, the limitations of the study and suggestions for future research.

Chapter 2: Literature Review

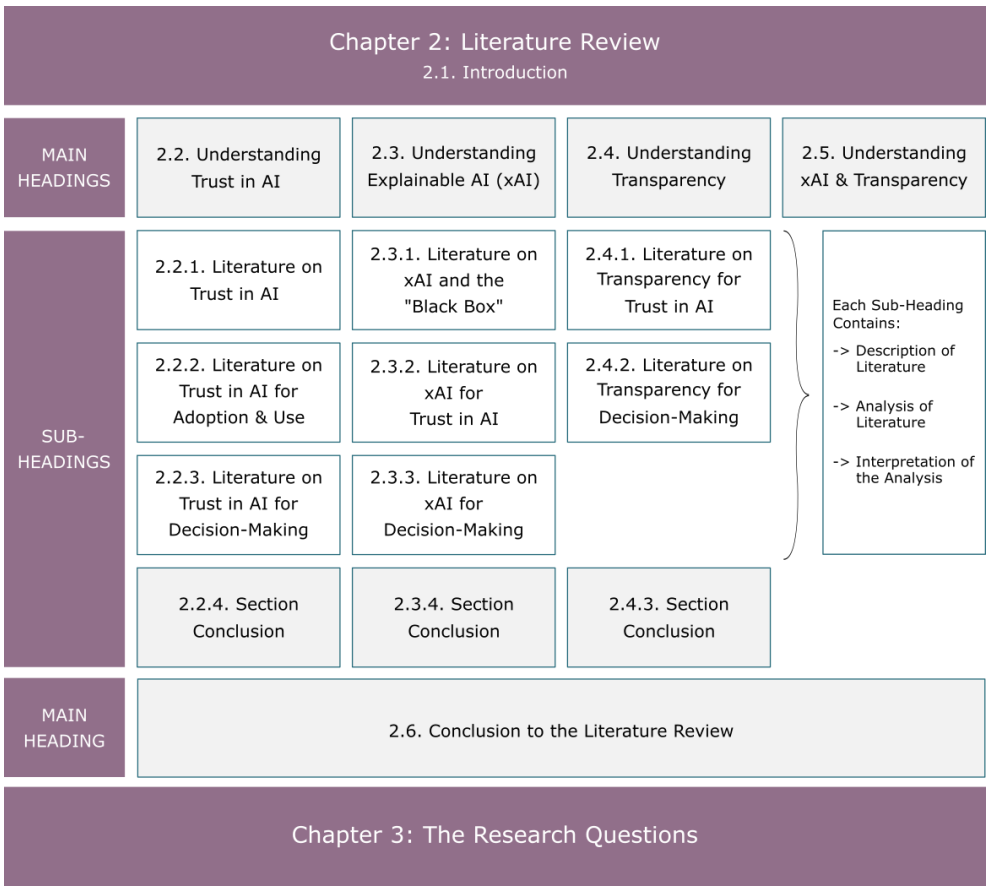
In the following section, a review of extant literature on artificial intelligence and it's use in organisational decision-making is provided, and evidence of several key concepts from the related scholarship are presented and discussed. Each sub-section discussion is followed by an analysis and interpretation, leading to the individual concept conclusion, which draws attention to calls for further research and areas for exploration and further insights. These section conclusions then combine to form the overarching conclusion to the literature review, which leads to the proposed research questions.

The review of the literature involved several steps. First, a Google Scholar and database (ProQuest & EBSCOhost) search of top management and technology journals was performed for only the past five (5) years (2019-2023/4). This recent and narrow date range ensured the current relevance of the literature and developed scholarship. Primary keywords used were "artificial intelligence", "AI in organisations", "AI in management", "AI decision-making", and "AI adoption". This revealed additional forks in the recent literature to "trust in AI", "explainable AI" and "xAI" which provided further concepts for investigation. These additional concepts were then also explored using the aforementioned method. Secondly, the gathered articles were filtered by 4* journal rating using the Chartered ABS AJG guide, which was then followed by a further search of "citing articles" in Google Scholar for the same keyword set, and again a ranking check in AJG. The result was a collection of 31 articles of 4* and 4-star rating, and 28 articles of 3-star rating. Some links and articulations of the key concepts were found in 2 Star journals, but those were cross-checked in Scopus for percentile and citation scores, of which only 99th Percentile articles and above, relevant to technology and management, were used. Majority 4* and 4 AJG journals have been included, with additional 3 AJG journals. All other journals have been used sparingly and only as appropriate to strengthen a particular analysis. This has ensured a high *quality* of review and relevance to the research questions.

A matrix of the literature review, denoting a roadmap of the following section is provided in Figure 2, overleaf, for an overview and ease of navigation. The identified concepts have been arranged from left to right and top to bottom, in such order as to create a logical flow through the literary landscape.

Figure 2:

Literature Review Matrix



Note: Author's own.

2.1 Introduction

As a result of the literature review process identified above, three recent systematic literature reviews (SLRs) were identified which provided an overarching view of the scholarship landscape. The first, by Laato et al. (2022) identified five goals of xAI, of which transparency and trustworthiness are two, in addition to understandability, fairness and controllability. Haque et al. (2023) posit "five effects of xAI systems" (p. 2), which again include transparency, trust, understandability, and fairness, but provide usability as opposed to controllability. Lastly, Glikson & Woolley (2020) explored extant literature on trust in AI and identified "tangibility, transparency, reliability, task characteristics and immediacy behaviours" (p. 627) as the five main dimensions of cognitive trust, with tangibility, anthropomorphism and immediacy behaviours being components of emotional trust. Glikson & Woolley (2020) noted that transparency, despite being a dimension of cognitive trust required further exploration in relation to emotional trust. This shall become relevant during the remainder of this chapter.

From the above SLRs, the common scholarly discussions are around the concepts of transparency, trust in AI (Laato et al., 2022; Haque et al., 2023; Glikson & Woolley, 2020) and

xAI (Laato et al., 2022; Haque et al., 2023). These three concepts formed the key areas of focus for this proposed research, due to the commonality of scholarly discussion and numerous calls for further exploration.

2.2 Understanding Trust in AI

2.2.1 Literature on Trust in AI

2.2.1.1. Discussion of Literature on Trust in AI: The foundation for understanding trust in AI (as opposed to just trust or trust in technology) for this proposed research is primarily built on the SLR by Glikson and Woolley (2020) and their two main dimensions of cognitive trust in AI and emotional trust in AI. This view frames the primary trust dimensions relevant for AI-specific applications in organizations, and their work draws from several key pieces of seminal literature. McAllister (1995), studied "affect and cognition-based trust as foundations for interpersonal cooperation in organisations" (p. 24). McAllister's work grounds the exploration of trust in foundational interpersonal dimensions, applicable to organizational AI contexts. Dunn et al. (2012) later consider affective and cognitive trust in the context of inter-personal trust and social comparisons within organisations, which helped to extend the trust framework from interpersonal to organizational levels, a relevant backdrop for AI use.

This work by Dunn et al. (2012) was referenced by Wang et al. (2016) who begin using the terminology of affective/emotional and intellectual/cognitive interchangeably, which was then further referenced and built upon by Glikson & Woolley (2020). Wang et al. (2016) therefore introduced the terminology that bridges interpersonal trust research with AI-specific contexts. Hoff & Bashir (2015) identified the importance of trust for not only inter-personal engagement, but also for "interaction with technology", (p. 3). Their findings usefully linked emotional trust with technology interactions, which may be of potential importance in this study to understanding user engagement with AI. They bring in a discussion of the relationship of trust to automated decision-making, as well as highlighting the importance of emotion-driven trust for technology adoption. These observations are relevant to gaining insights and understanding as to the importance of emotional trust for AI acceptance and are directly relevant to AI deployment in organizations.

Wang et al. (2016) extended these concepts to "recommendation agents" (RA's) or "software-based decision aids" (p. 1), thereby applying the trust dimensions to automated tools, making the leap to AI-reliant decision aids. As far back as 2016, Wang et al. (2016) were highlighting the lack of consideration of the differences between the emotional and cognitive aspects of trust with regard to these RA's, which is recently echoed in calls for further research by Glikson & Woolley (2020), now in the context of transparency having an effect on emotional trust in AI. Wang et al. (2016) also called for further research around affect-based (emotional-

based) trust in recommendation agents, which Glikson & Woolley (2020) are still calling for regarding the relationship between transparency and emotional trust in AI.

When compared and viewed together, this scholarship identified the importance of understanding how transparency impacts emotional trust in AI adoption, which was therefore explored further in this study. Another piece of important evidence which Wang et al. (2016) highlighted was an influence of *explanations* on cognitive and emotional trust, which speaks to the construct of xAI, yet to be outlined in a sub-section of this chapter. This evidence supported the potential role of explainable AI (xAI) in fostering both emotional and cognitive trust, which may be of relevance to the broader context of this study, once explored further.

2.2.1.2. Analysis of Literature on Trust in AI: There was a progression which can be seen in the presentation of the literature above. This is further outlined below before providing the meaning of the analysis in the closing paragraph.

There is a well-established scholarly use of the cognitive and emotion-based trust components, which has been built over a period spanning 30 years, in both an organisational, social, and technological context. McAllister's (1995) seminal work established dual dimensions of trust within inter-personal organizational contexts, laying a foundation upon which subsequent studies could apply and adapt these constructs to technology. Hoff & Bashir (2015) extended this foundation to the human-technology interface, affirming the applicability of these dimensions in the context of automated systems while emphasizing the significance of emotion-driven trust in technology adoption. Their work broadened the theoretical landscape, marking a shift from interpersonal trust frameworks to trust frameworks relevant to human-machine interactions, while Wang et al. (2016) extended the body of knowledge by applying the trust dimensions to automated tools, which allowed for the later leap to AI-reliant decision aids. Glikson & Woolley (2020) further advanced this line of inquiry, positioning emotional and cognitive trust as distinct yet interrelated components critical to AI engagement, also aligning these dimensions with transparency.

The comparison between McAllister (1995), Hoff & Bashir (2015), Wang et al. (2016) and Glikson & Woolley (2020) therefore revealed a progression of these trust components from an inter-human to a human-technology and later, a human-AI perspective. This progression from inter-human trust to human-technology trust and human-AI trust reflected an ongoing scholarly evolution towards understanding trust's role in AI. We also saw the formation of early links between trust, decision-making, organisations, technology, and explanation, which feed into the current discourse on Trust in AI. However, while cognitive trust had been explored in depth, emotional trust remained less examined within AI contexts,

despite Wang et al. (2016) and Glikson & Woolley (2020) highlighting it as important for AI acceptance.

2.2.1.3. Interpretation of Literature on Trust in AI: For the purpose of this research, it was important to have demonstrated the development of the concept of inter-personal trust, to a concept of trust in technology and, ultimately, a concept of trust in AI. Exploring the above has also revealed important foundational links between trust, organisational behaviour, and decision-making, which become relevant in the subsequent sub-sections of this chapter. A clear progression in scholarly focus was observed, showing an evolution in focus from cognitive trust to an under-explored domain of emotional trust in AI.

A potential area for further research was also revealed surrounding transparency and emotion-based trust in AI. This gap in emotional trust research, specifically in relation to transparency and explainability, underscored a research need, which this study aimed to address by investigating how transparency influences emotional trust in AI for organizational decision-making.

2.2.2 Literature on Trust in AI for Adoption and Use

2.2.2.1. Discussion of Literature on Trust in AI for Adoption and Use:

In their exploration of agency perception and its effect on trust in AI, Vanneste & Puranam (2024) suggested that adoption and use of an AI system may be impacted by a lack of trust. This highlighted the potential impact of trust in AI on adoption and emphasized trust as a determining factor in whether organizations will integrate AI into their workflows. This sentiment was echoed by Sullivan et al. (2022) who established that gaining benefit from AI can only be achieved through trust in AI. They therefore reinforced trust as essential for AI's practical value in organizations, supporting the need for trust-focused studies.

Additionally, Sullivan et al. (2022) called for further research into the different dimensions of trust and their measurement. Lastly, another important argument which the authors made regarding harms and injustice was that "cognitive appraisals" directly cause "emotional appraisals" (Sullivan et al., 2022, p. 529). This suggested that emotional responses may be influenced by cognitive judgments and similar to Glikson & Woolley, (2020), identified a research need around the emotional dimension of trust in AI. In a discussion on employee-AI trust, Enholm et al. (2021) suggested a direct link between a lack of trust in AI and its adoption and use in organisations, calling it an inhibitor. This insight positioned a lack of trust as a potential barrier to AI adoption, which the research aimed to investigate in an organisational setting.

Leading into the discussion of xAI in the following sections, Wang & Ding (2024) concluded that adoption of AI required improved levels of trust in the explanation of its function.

Importantly to this research, this underscored the influence of xAI in trust engenderment, speaking to the relevance of xAI's inclusion as a construct for the study. Further supporting this point, Weber et al. (2023) also alluded to xAI possibly increasing trust and thereby increasing adoption rates of AI systems. In related work, a paucity of literature and studies on trust in AI was identified by Wong et al. (2024) across the levels of the organisation, business unit and individual user. A further "lack of understanding of how AI is adopted and used in organisations" and trust in AI possibly inhibiting AI use was highlighted by Enholm et al. (2021, p. 1709). These literature gaps aligned with the study's aim to address trust in AI across organizational levels and to further understand its influence on adoption.

Lastly, Glikson & Woolley (2020) identified the importance of trust in AI for its use in organisations in the future and suggested that this trust would be foundational to AI's organisational role. Glikson & Woolley (2020) went further to call for future studies to "address the emotional and cognitive aspects of trust in AI together", and to explore the relationship between them (p. 649). Both observations positioned trust as a fundamental requirement for AI's role in organisations, while highlighting the need to explore the emotional aspects of trust in AI.

2.2.2.2. Analysis of Literature on Trust in AI for Adoption and Use: In analysing the above evidence, comparison between Sullivan et al. (2022) and Glikson & Woolley (2020) revealed a similarity in their reference to the cognitive and emotional aspects of trust. Interestingly, a key difference was that Sullivan et al. (2022) inferred a causal link between the cognitive and emotional dimensions, whereas Glikson and Woolley (2020) did not make this connection explicitly. Glikson & Woolley (2020) did, however, call for further research in this regard. The identified difference suggested a gap in understanding of the relationship between cognitive and emotional trust, which indicated that further research could help clarify how these dimensions interact to foster trust in AI.

There was also similarity across several scholars who identified a causal link between trust in AI and its adoption and use. (Vanneste & Puranam, 2024; Sullivan et al., 2022; Enholm, et al., 2021; Wang & Ding, 2024; Glikson & Woolley, 2020). Their observations that trust is a necessary precondition for AI adoption and use was strengthened by their shared view, which supported the focus of this study.

Comparison of Wang & Ding (2024) and Weber et al. (2023) alluded to a link between xAI, trust in AI and the subsequent adoption and use of AI technologies in organisations, which is further explored in subsequent sections of this chapter. This similarity supported the importance of xAI in fostering trust, suggesting that xAI could act as a bridge to user trust, thereby promoting AI adoption. Wong et al. (2024) and Enholm et al. (2021) similarly identified

a lack of knowledge surrounding trust in AI and adoption and use at an organisational level, calling for further research to explore the relationship further. Their shared identification of this gap indicated a limited understanding of trust in AI at an organisational level and also how trust might impact AI adoption.

2.2.2.3. Interpretation of Literature on Trust in AI for Adoption and Use:

What was revealed in the analysis of the above literature was that trust in AI is foundational to its adoption and use, however there was still an identified gap in the literature regarding the interaction between the cognitive and emotional dimensions of trust in AI. The analysis also indicated a possible causal link between cognitive and emotional trust, which may be important for understanding how to engender trust in AI. Furthermore, there was another identified area for future research at the organisational level, regarding studies on Trust in AI and its subsequent adoption and use. The role of xAI as a potential enabler of trust also emerged as an insight, suggesting it may bridge trust gaps and encourage broader adoption within organizations.

2.2.3 Literature on Trust in AI for Decision-Making

2.2.3.1. Discussion of Literature on Trust in AI for Decision-Making:

In a literature review of AI and business value, Enholm et al. (2021) identified that at an organisational level, employees may face a future requirement to base their decisions on AI systems, and that this requires them to form a level of trust in AI. This established a direct link between trust and the practical use of AI in decision-making. They then went further in highlighting the difficulty of "building trust between humans and machines" (p. 1718), which emphasised the complexity of AI adoption into decision-making processes. Enholm et al. (2021) also warned that AI is data-driven and that it could therefore be compromised by biases if the input data is unbalanced or discriminatory, which is relevant to the study as a dimension which might affect trust in AI's use for decision-making.

The uptake of AI for decision support was highlighted by Chen et al. (2021), specifically during the Covid-19 pandemic, as a key area of organisational readiness. However, they also emphasised trust in AI as a fundamental pre-requisite to use of decision support systems which are built on AI. The implication here was that uptake of AI is needed for organisational readiness, but that it cannot be achieved without trust in AI. The authors then called for further research on human-centred AI frameworks across different industries (Chen et al., 2021), which in part speaks to the chosen setting of this research. Additionally, Chen et al. (2021) focused on "human-in-the-loop" type systems (p. 1), while Enholm et al. (2021) emphasized the focus of AI as being an "assistive role" which supports, rather than replaces humans (p. 1720).

In an article exploring xAI and its impact on “human-AI trust and decision performance”, Wang & Ding (2024, p. 1) identified a lack of trust in AI as a key barrier to decision-making. Although xAI is not a direct focus of this section’s discussion, it does identify a relationship between xAI, trust in AI and decision-making. Furthermore, Sabharwal et al. (2024) asserted that AI decision support would not be possible without managerial adoption, for use in their decision-making. Another point which the authors raised was that explanation of AI would boost transparency to “alleviate trust difficulties in data-driven decision-making” (p. 1), which was important to the research questions and related to the study’s broader focus on trust in AI.

Lastly, Shrestha et al. (2019) concluded their article on AI and organisational decision-making with a call for further research to discover how “trust concerns” (p. 78) can be overcome in order to leverage the rising importance of AI-based decision-making. This final observation underscored the importance of addressing trust issues for AI’s future in decision-making and highlighted a gap which this research aims to explore.

2.2.3.2. Analysis of Literature on Trust in AI for Decision-Making:

Comparison of the above literature revealed similarity in an apparent link between trust in AI and organisational decision-making, with trust in AI appearing as a pre-requisite to AI-based decision-making (Enholm et al., 2021; Chen et al., 2021; Shrestha et al., 2019). This similarity across sources highlighted that trust is an important component for enabling practical use of AI in organizational decision-making. Furthermore, it reinforced the idea that trust must be established before AI is integrated into decision-making processes within organisations.

Both Enholm et al. (2021) and Sabharwal et al. (2024) emphasized the data-driven nature of AI and its importance in decision-making; with Sabharwal et al. (2024) inferring that the key concepts of this research (trust in AI, explainability (xAI) and transparency) may assist with trust engenderment in data-driven decision-making. While data is certainly at the core of AI’s functionality, it was one part of the broader focus of this research, with the research questions being centred on trust in AI, transparency and explainability (xAI) more broadly.

Trust in AI also similarly appeared as a barrier and a difficulty to the adoption and use of AI for decision-making, across several articles. (Enholm et al., 2021; Wang & Ding, 2024; Sabharwal et al., 2024). The recurrence of this theme suggested that trust-related challenges are acknowledged as an obstacle to AI adoption in decision-making contexts. This underlined a need to generate new insights and understanding of how to mitigate these trust barriers. Although there was similarity in the literature that trust in AI is required for decision-making, further insights are still required. This was supported by the call for further research into “trust concerns” by Shrestha et al. (2019, p. 78) and by Chen et al. (2021) on human-centred AI frameworks across different industries.

2.2.3.3. Interpretation of Literature on Trust in AI for Decision-Making:

The literature established trust in AI as foundational for its use in organisational decision-making, with scholars highlighting that trust must be built for AI to become a reliable part of organizational decision-making processes. The observations on data-driven decision-making added an additional perspective to the focus of the research questions, which are centered on trust in AI, transparency and xAI more broadly. A lack of trust in AI was revealed as a potential barrier to the adoption and use of AI for organisational decision-making, with scholars calling for further research in this regard. The above interpretations of the literature are directly reflected in the aims of this study and its research questions.

2.2.4 Section Conclusion

This section explored the concept of trust in relation to AI, highlighting its importance to the adoption and use of AI in organisations for decision-making. Presentation of various sources of evidence from literature, their analysis and subsequent interpretation, identified several areas for further research, which are presented below:

a.) The emotional aspect of trust and its relationship to trust in RA's had not been answered since identification of the gap by Wang et al. (2016) or the subsequent call by Glikson & Woolley (2020) to explore the effect of transparency on emotional trust in AI.

b.) There was still an identified gap in the literature regarding the interaction between the cognitive and emotional dimensions of trust in AI (Glikson & Woolley, 2020).

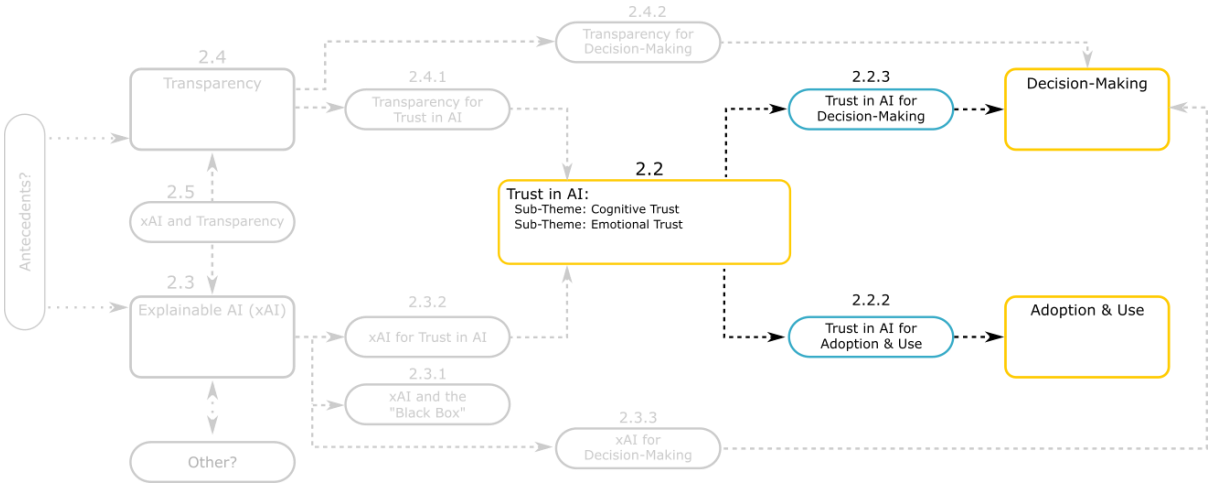
c.) There was a need for studies on trust in AI and its subsequent adoption and use across organisational, business unit and individual levels across different sectors (Wong et al., 2024; Enholm et al., 2021; Chen et al., 2021).

d.) Lastly, there was a need to discover how trust concerns can be overcome to leverage the rising importance of AI-based decision-making (Shrestha et al., 2019).

Therefore, conclusion could be drawn that Trust in AI is important both to the adoption and use of AI in organisations, but also for its use in organisations for AI-based decision-making. There were still areas for development of additional insights, both theoretically and in practice. The presented scholarship also alluded to a relationship between xAI and trust, as well as a relationship between xAI and decision-making, which are both discussed further in the following section. Figure 3 shows the conceptual relationships developed thus-far through the discussion, analysis, interpretation, and conclusion of section 2.2.

Figure 3:

Conceptual Framework Development Contribution of Section 2.2



Note: Author's Own

2.3 Understanding Explainable AI (xAI)

2.3.1 Literature on xAI and the "Black Box"

2.3.1.1. Discussion of Literature on xAI and the "Black Box": An important observation from the related literature was that one of the key drawbacks of AI is its inscrutability (Berente et al., 2021) and the "black-box" nature of AI models and tools (Rai, 2020). This inherent lack of transparency identified by the authors was considered a primary limitation of AI systems, which supported the further exploration of xAI as a solution. In the context of marketing research, Rai (2020) identified the importance of explanation in "black-box" AI models. The author suggested that xAI holds the potential to affect "whether or not the user takes action based on the prediction" (p. 139) and provides "rationale for the decision-making process" (p. 138). This positioned xAI as a tool for enhancing user engagement and decision confidence, both relevant to AI adoption.

In an SLR of xAI in the financial sector, Weber et al. (2023) identified a lack of explainability as a key reason for the lack of uptake of AI systems. They intimated that sectors such as finance, which have strict regulatory requirements, need to be able to understand the processing steps within the "black box". Their observations revealed a potential regulatory necessity for explainability, which speaks to its importance in creating additional transparency of the "black-box".

In the field of healthcare predictive analytics (HPA), Kim et al. (2023) stated that a lack of transparency was attributable to the "black box structure" (p. 1) of AI models and suggested a method for explaining "black-box" predictions. Wang & Ding (2024) further suggested that explanation of the AI algorithm improves transparency, which can aid in addressing the "black box" issue by improving trust. These two observations potentially linked transparency and

trust, framing xAI as a tool to foster both. Ågerfalk (2020) also explained that for AI systems to be accountable, xAI could "un-black box alogorithmic decision-making" (p. 6), which emphasised a potential connection between xAI and improved accountability.

Additional scholars (Choung et al., 2022; Vanneste & Puranam, 2024; Grover et al., 2022) all discussed the concepts of explainability, transparency and the "black box" as a primary barrier to AI uptake for decision-making. Their similarity of views reinforced xAI's role as an important tool for improving AI's usability in decision contexts. Rai (2020) also suggested xAI as a tool for "black box" to "glass box" (Rai, 2020, p. 138); a metaphor which emphasised xAI's role in creating transparency in AI systems, which may be of importance to trust and adoption.

Lastly, Lukyanenko et al. (2022) suggested the importance of understanding what goes on within AI systems and highlighted both xAI and transparency as fertile ground for future AI research, while stressing the need to establish a further foundational understanding of all three concepts. This last observation, also supported by the literature discussed above, identified an ongoing research gap, positioning this study within a broader discussion and research agenda on xAI and transparency.

2.3.1.2. Analysis of Literature on xAI and the "Black Box": From a comparison of the above literature, it was clear that a similar view existed that the "black box" is a fundamental barrier to the uptake of AI (Weber et al., 2023; Ågerfalk, 2020; Berente et al., 2021; Rai, 2020). The similarity highlighted a shared concern about the lack of transparency of AI systems and their internal workings, which limits AI's adoption across different fields and industries. In particular, Rai (2020), Weber et al. (2023), and Kim et al. (2023) emphasized this issue in marketing, finance, and healthcare contexts, respectively, indicating sector-specific concerns where transparency is essential for regulatory compliance and user trust. This potentially showed varied demands of xAI across contexts.

Also observed was the similarity across not only the SLR of Weber et al. (2023) but also all of the above-cited scholars, that xAI and explainability are a key element in addressing the "black-box" problem. Their shared views reinforced xAI's role in making AI systems more transparent, which was identified as being important for user trust and acceptance. Most interesting was the similarity of the apparent centrality of the "black box" problem to the concepts/constructs of transparency, xAI, trust in AI and decision-making discussed throughout this chapter, as well the required foundational understanding of what goes on within it (Lukyanenko et al., 2022; Berente et al., 2021; Rai, 2020).

2.3.1.3. Interpretation of Literature on xAI and the "Black Box": We thus identified the "black box" as a barrier to AI adoption and use for decision-making, providing relevance

to the previously-presented section evidence in this chapter. We also concluded that being able to explain what is going on within a “black-box” AI system is key to giving individuals responsible for decision-making the required visibility and transparency of the inner-workings of that system.

From a perspective of personal accountability for individuals responsible for decision-making, explainability may have connection to the emotional aspects of trust in AI, but this was not certain and required further insights and enquiry. Additionally, the cross-sectoral demand for explainability, particularly in regulated fields, reinforced the importance of xAI in aligning AI tools with industry-specific transparency standards. These observations demonstrated the relevance of the xAI construct to the proposed research.

2.3.2 Literature on xAI for Trust in AI

2.3.2.1. Discussion of Literature on xAI for Trust in AI: Explainability of AI was a key dimension highlighted by Berente et al. (2021) to address its inscrutability and thus engender trust, positioning it as a solution for overcoming AI’s opacity. In a wide-reaching quantitative study by Silva et al. (2023), a strong positive correlation between explainability of xAI systems and human trust was revealed. However, despite the correlation, the authors noted that progress in the xAI field was “hindered” and called for standardised measurement to “evaluate explainability methods” (p. 2). A gap was highlighted by this, as reliable measures of explainability are important for evaluating trust outcomes from xAI.

In another study, Wang & Ding (2024) explored whether xAI can improve trust in AI but only found partial support for xAI improving trust in AI. Their work therefore suggested that future research should further consider initially-estimated trust and the “role of AI in the dynamic trust calibrating process” (p. 14). Rai (2020) identified that xAI provided two levels of trust, namely trust in the AI model and trust in the AI prediction. Despite the research being focused in the marketing sector, Rai (2020) called for further research on xAI in these two areas of trust formation but in “different application domains” (p. 140). The setting of this research was also therefore positioned to potentially reveal domain insights into xAI and trust in AI across varied contexts.

Sabharwal et al. (2024) identified xAI as key to trust engenderment and managerial adoption of AI and called for further research on the topic, while (Pumplun et al., 2023) called for a further understanding of how to design xAI for end-users. Both therefore demonstrated a current and relevant need for further research around xAI. This was also true of Balasubramanian et al. (2022) who identified the need to identify the factors which drive human trust in machine learning (ML), a sub-set of AI, and Abedin (2022) similarly called for future study on explainability factors (p. 447). Of interest from these authors was a seemingly

practical focus on understanding the “criteria” which actually need to go into creating xAI tools which could engender trust in AI.

Additionally, the SLR work by Haque et al. (2023) identified trust in AI as one of the five effects of xAI, showing commonality among the scholarship as to the relationship between the two. Haque et al. (2023) also called for identification of xAI development guidelines in their proposed future research agenda (p. 10). Their SLR positioned trust in AI as a consistent outcome of xAI, suggesting that standardized development practices could enhance trust outcomes. In the second SLR by Laato et al. (2022), trustworthiness was identified as a goal of xAI and the importance of identification and use of xAI features and guidelines by xAI developers was emphasised. In financial services, Esmaeilzadeh & Vaezi (2022) pointed to AI needing to explain itself to satisfy regulatory requirements and to “prove lack of bias” (p. 560), a sentiment which was echoed by Weber et al. (2023). This highlighted xAI’s potential to address compliance needs, particularly in fields where transparency and trust are regulatory concerns.

Lastly, Pumplun et al. (2023) presented useful research on the use of xAI in clinical decision support systems (CDSSs) which emphasised the importance firstly of trust in the technology (clearly required when using it as a basis for clinical decisions), but also of the importance of "explainability" in the engenderment of trust in AI systems. This underscored xAI’s relevance in critical contexts, where trust in AI impacts decision outcomes.

2.3.2.2. Analysis of Literature on xAI for Trust in AI: Comparing the variety of cross-industry examples in the evidence, all calling for explainability in AI, or a deeper understanding of xAI, an urgency emerged for further foundational research of the construct. Such widespread calls across sectors highlighted the need to deepen the understanding of xAI’s role in fostering trust in AI, indicating its broad relevance. Firstly, across two SLRs (Haque et al., 2023; Laato et al., 2022), there was a demonstrated similarity in the scholarship that xAI leads to trust in AI (Haque et al., 2023) and also that trust is a goal of xAI (Laato et al., 2022). However, there was a notable exception of Wang & Ding (2024) who found only partial support of the relationship, and subsequently called for further research in this regard. The difference suggested that xAI’s trust-building potential might depend on context and pointed to a need for studies on specific factors influencing the relationship.

There were also similar calls for further research across the scholarship (Silva et al., 2023; Wang & Ding, 2024; Rai, 2020; Sabharwal et al., 2024) around the various aspects of the relationship between xAI and Trust in AI. There was also a call for further research by Rai (2020) in different application domains, and by Sabharwal et al. (2024) around the engenderment of trust in AI in managers through xAI. This spoke to Pumplun et al. (2023)

identifying the need for design guidelines for xAI, as well as Abedin (2022) who called for future research on factors of “explainability practices” (p. 447). These similar and recurring themes underscored a shared view on the importance of xAI for trust engenderment, which aligns closely with the study’s focus.

2.3.2.3. Interpretation of Literature on xAI for Trust in AI: What became apparent from the evidence and analysis was that the use of xAI is an opportunity to engender trust in AI across various domains. There are also calls from scholarship to explore the relationship further and to develop further insights around the relationship between xAI and trust in AI in different settings. The analysis also indicated that xAI's trust-building potential may vary by context, pointing to a need for sector-specific guidelines and evaluation methods.

It was clear that there was excitement around this burgeoning concept and that additional work was required to develop further insights for its foundational development and deployment. Further refinement of xAI design and measurement practices emerged as critical for achieving its full trust-enhancing potential. This leaves areas for development both in theory and practice, with a clear call to understand the foundational factors of xAI.

2.3.3 Literature on xAI for Decision-Making

2.3.3.1. Discussion of Literature on xAI for Decision-Making: In addition to the observations that xAI contributes to trust in AI, Rai (2020) also drew attention to the use of xAI in illuminating the rationale within the AI system that allows its use in decision-making. Their work was introduced by highlighting a paucity of understanding of the decisions made by AI systems (Rai, 2020). Both of these insights potentially identified xAI as a mechanism for clarifying AI logic, suggesting that explainability is important for trust in AI decision-making. This further suggested a significant gap in understanding AI outputs, underscoring the need for xAI to address inherent opacity in AI decision-making.

The importance of xAI providing insights into the reasons for AI-recommended decisions was reinforced by Rabiee et al. (2024), who also stressed the importance of supervised feature selection, for such models to enhance explainability. Vanneste & Puranam (2024) concurred with the momentum developing behind the importance of explanation for AI decision-making, while xAI was again highlighted by Ågerfalk (2020) as important to the same outcome. These three sources similarly affirmed xAI’s role in fostering understandable AI decisions, reinforcing its relevance in AI decision-making contexts.

Berente et al. (2021) drew attention to the importance of managers in decision-making, supported by AI, and suggested that explanation “of AI decisions is not only cognitive but also social, which has been overlooked by literature” (p. 1444). Their perspective broadened the view of xAI, emphasizing the social context of understanding AI-driven decisions, and

highlighted a possibly overlooked dimension. In a related discourse on AI & Data Network Effects, Gregory et al. (2021) repeatedly emphasised the importance of explainability as it relates to AI-predictions and subsequent decision-making, suggesting that it “must be considered strategic” (p. 24) due to its relationship to “perceived user value” (p. 24). The interesting insight from their work was their view that xAI was important enough to be considered strategic in AI-supported decisions.

A further study on xAI, trust and decision-performance by Wang & Ding (2024) presented a finding that, in a sales setting, “xAI can improve human decision accuracy” (p. 1), and called for research around xAI and “trust in AI for different decision-makers” (p. 14). Lastly, in the fields of management and finance, Sabharwal et al. (2024) noted that xAI explanations for AI decision-support systems were “not well established in the academic literature” (p. 4). Both research articles highlighted a gap in the literature on xAI’s role in structured decision-support, which underlined the need for further research in this area, supporting the focus areas of this study.

2.3.3.2. Analysis of Literature on xAI for Decision-Making: Rai (2020), Rabiee et al. (2024), Vanneste & Puranam (2024) and Ågerfalk (2020) all similarly emphasized the important role of explainability in AI. They agreed that understanding the reasoning behind AI decisions increases trust and can improve decision-making performance. Such alignment suggested that xAI was widely viewed as important to AI-supported decision-making, reinforcing its importance in organizational contexts.

Wang & Ding (2024) and Sabharwal et al. (2024) pointed out that despite the recognized importance of xAI, there was a need for more research, especially concerning how xAI affects trust among different individuals responsible for decision-making. This identified gap indicated that xAI’s trust engenderment might vary depending on the role of the individual responsible for decision-making and reinforced the need for further study.

Lastly, Berente et al. (2021) focused on the role of managers in AI decision-making processes, arguing that the social context of explanations was often overlooked in favour of cognitive ones, a similar theme to the dimension of emotion presented by Glikson & Woolley (2020) in the preceding section on trust in AI. The potential relevance of social factors in xAI’s role within decision-making was underscored by this similarity.

2.3.3.3. Interpretation of Literature on xAI for Decision-Making: Overall, while there was a strong consensus on the importance of xAI for enhancing trust and improving decision-making, it was apparent that the scholarly understanding was still developing and left room for exploratory work to garner new insights. The analysis also suggested that xAI’s

impact on decision-making may vary depending on the role of the individual responsible for decision-making, highlighting a potential need for context-specific application of xAI.

Of particular interest was the recurring theme of cognitive versus social trust in AI which may be related to the affective and emotional factors of trust in AI previously discussed. However, this required further exploration and insight development. Additionally, the potential strategic role of xAI in decision-making underscored its importance beyond technical support, positioning it as an essential component for aligning AI outputs with organizational goals.

2.3.4 Section Conclusion

The review of literature in this section underscored the pivotal role of xAI in addressing the challenges posed by AI's "black box" nature. Scholars across various domains, ranging from marketing and finance to healthcare, consistently highlighted the inscrutability of AI systems as a significant barrier to their adoption and trustworthiness. The centrality of xAI to enhancing transparency, fostering trust, and improving decision-making efficacy was widely acknowledged. As per the preceding section, several areas for further research were identified, and are presented below as an extension from section 2.2:

e.) Both Silva et al. (2023) and Haque et al. (2023) emphasized the need for standardised metrics and guidelines to evaluate the effectiveness of xAI methods.

f.) Rai (2020) and Sabharwal et al. (2024) called for extension of xAI research to different application domains. This would lead to a better understanding of how xAI can foster trust and improve decision-making across different fields.

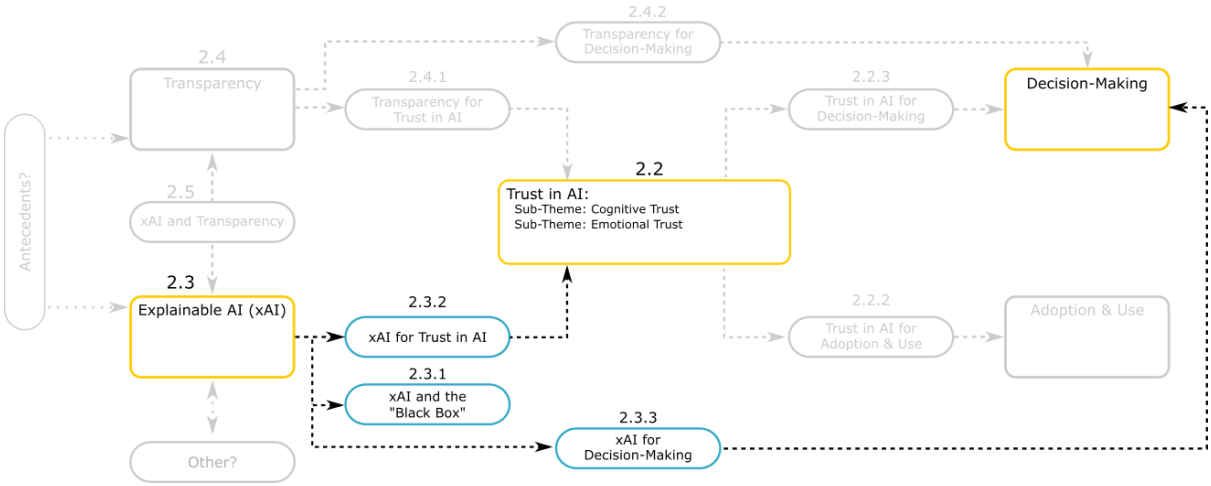
g.) Pumplun et al. (2023), Blasubramanian et al. (2022), Sullivan et al. (2022), Abedin (2022) and Lukyanenko et al. (2022) called for further understanding of the factors contributing to trust in AI and those which required consideration for design of xAI systems.

h.) Berente et al. (2021) emphasized exploring not just cognitive trust, but also the social aspect and managerial impacts of xAI, a theme similar to that of cognitive and emotional trust in AI presented by Glikson & Woolley (2020). Particularly, how these factors influence trust and decision-making in organisational settings.

Figure 4 shows the conceptual relationships developed thus-far through the discussion, analysis, interpretation, and conclusion of section 2.3.

Figure 4:

Conceptual Framework Development Contribution of Section 2.3



Note: Author's Own

2.4 Understanding Transparency

2.4.1 Literature on Transparency for Trust in AI

2.4.1.1. Discussion of Literature on Transparency for Trust in AI: Beginning from the three SLRs of Laato, et al. (2022), Haque et al. (2023) and Glikson & Woolley (2020) it was possible to develop a comprehensive overview of the scholarly landscape and the relationship between transparency and trust in AI. Together, these SLRs established transparency as a cornerstone in AI trust research, forming a foundation for further exploring its role in fostering trust in AI.

Laato et al. (2022) identified that a lack of transparency has a negative impact on "trustworthiness", with both being two of the five goals of xAI. The role of transparency here emphasized its function in fostering reliability and user confidence, while underscoring its priority within xAI goals. Haque et al. (2023), on the other hand, considered both trust in AI and transparency as effects of xAI systems, while Glikson & Woolley (2022) presented transparency as a dimension which builds cognitive trust in AI. Importantly, Glikson & Woolley, (2020) highlighted that transparency was yet to be explored as an emotional dimension of trust in AI, opening avenues for further investigation. The relationship between xAI and transparency is further elaborated on in the following Section, 2.5.

Sabharwal et al. (2024) provided a statement on the relationship that "transparency is likely to be the foundation of trust in AI" and emphasised the importance of "human values" needing to be central to the human connection to AI technologies (p. 6). By tying transparency to human values, their work expanded transparency's relevance, positioning it as essential not only for trust in AI, but also for user-aligned AI adoption. Transparency leading to trust in AI was supported by Wang & Ding (2024), who also called for further research into the "impact

of transparency on human trust in AI" (p. 14), in different scenarios, and with different "types of decision-makers" compared to those in their study (p. 14). Their emphasis on varied contexts and types of individuals responsible for decision-making pointed to the nuanced effects which transparency may have across different settings.

The work by Sullivan et al. (2022) suggested that transparency may address an emotional or moral "uncanny feeling" associated with trust in AI and called for further research in this regard. This insight introduced a unique angle, proposing that transparency could counteract discomfort and moral unease, which speaks to human emotion and could be important for user acceptance. Lastly, in an investigation of ethical principles and trust in AI, Choung et al. (2022) called for future exploration of transparency, among other factors, as a critical requirement for adoption of AI technologies. This closing observation positioned transparency not only as a facilitator of trust in AI, but also as a fundamental ethical requirement for responsible AI adoption and use.

2.4.1.2. Analysis of Literature on Transparency for Trust in AI: The similarity between the three SLRs and additional articles highlighted a widespread acknowledgment among scholars that transparency is not merely a feature but a critical necessity for building trust in AI. (Laato et al., 2022; Haque et al., 2023; Glikson & Woolley, 2020; Sabharwal et al., 2024; Wang & Ding, 2024; Sullivan et al., 2022; Choung et al., 2022). Such broad agreement underscored transparency's foundational role, reinforcing its importance to the development of trust in AI. Furthermore, Sabharwal et al. (2024) and Wang & Ding (2024) similarly suggested that without transparency, it would be challenging to build a reliable foundation for trust in AI technologies. Their shared view further reinforced that transparency is central to trust in AI.

Laato et al. (2022), Haque et al. (2023), and Glikson & Woolley (2020) all explored how transparency impacts trust, but with differing emphasis. Laato et al. (2022) and Haque et al. (2023) discussed it as a direct effect of xAI, while Glikson & Woolley (2020) considered it within the framework of cognitive trust. They also suggested that work was needed to understand the emotional aspect of trust in AI and how transparency affects it. These differences in emphasis pointed to transparency's multi-dimensional role in trust formation, hinting at both cognitive and emotional pathways.

Glikson & Woolley (2020) and Sullivan et al. (2022) both similarly discussed the emotional aspects of transparency and trust in AI, which were less emphasized in other studies. Their similar focus on emotional aspects added depth to the trust discussion, suggesting that transparency could mitigate emotional concerns linked to AI use.

2.4.1.3. Interpretation of Literature on Transparency for Trust in AI: There was a well-established link between transparency and trust in AI in the scholarship, with transparency increasing trust. However, transparency as an emotional dimension of trust in AI was not adequately explored by the scholarship and was a gap in the extant literature which may be a potential area for future research which could reveal valuable insights. The analysis further suggested that transparency's impact may vary across cognitive and emotional trust dimensions, indicating that different approaches to transparency may be required to address both aspects effectively. Additionally, there was an emerging view that transparency may mitigate emotional discomfort or moral unease, potentially making it a valuable tool for addressing broader ethical and emotional concerns in AI adoption.

2.4.2 Literature on Transparency for AI Decision-Making

2.4.2.1. Discussion of Literature on Transparency for AI Decision-Making: Lindebaum et al. (2020) had highlighted a lack of transparency and understanding as a significant limitation of AI which could become a liability in decision-making (p. 256). They also further identified data quality and self-reinforced negative feedback as two key vulnerabilities of AI algorithms. These insights pointed to transparency as not only a feature of AI systems, but a safeguard, essential for mitigating inherent risks in AI-driven decisions. With the introduction of new data protection regulations, Kim et al. (2023) recommended that researchers should rise to the challenge of “raising standards of transparency and accountability” (p. 1306), which might engender judicious decision-making. This emphasis on regulatory compliance for accountability underscored transparency as both a legal and ethical standard, integral to responsible AI decision-making.

Vanneste & Puranam (2024) identified transparency as an intervention to clarify the reasoning and understanding underpinning an algorithm’s decisions. They also considered the agency aspects related to trust and adoption of AI for decision-making. Their work highlighted transparency as a means to increase interpretability and trustworthiness, particularly as AI’s role in decision-making expands. Additionally, Sabharwal et al. (2024) similarly agreed that in order to use AI’s advantages for decision-making, transparency was a necessary requirement for responsible adoption, reinforcing its importance.

However, Vanneste & Puranam (2024) also raised an important issue that modern AI was becoming more “agentic” at the same time that it was increasing in complexity, exacerbating the difficulty with transparency in decision-making. This observation introduced a challenge as AI evolves, suggesting that transparency will be more difficult to achieve as systems grow more autonomous and intricate. Glikson & Woolley (2020) also drew attention to the same dynamic between complexity and transparency and suggested that, especially in the context of complicated AI systems, this would lead to an increase in the necessity for

transparency. Their perspective added weight to Vanneste & Puranam, (2024)'s argument that as AI complexity rises, so does the imperative for clear explanations in decision processes.

2.4.2.2. Analysis of Literature on Transparency for AI Decision-Making: From the above evidence, the similarities in scholarship were readily visible and demonstrated common agreement in the requirement of transparency in AI to strengthen confidence in AI decision-making (Lindebaum et al., 2020; Kim et al., 2023; Sabharwal et al., 2024). This alignment reinforced transparency's role as a foundational element in enhancing trust for AI decision-making. Both Vanneste & Puranam (2024) and Glikson & Woolley (2020) made similar observations regarding the inherently increasing complexity trajectory of AI. Their shared concern underscored the challenge that increasing complexity of AI may pose to transparency, suggesting that trust-building through AI explanations could become more difficult as systems evolve.

A difference in focus also emerged, as Vanneste & Puranam (2024) discussed transparency primarily in terms of interpretability and trustworthiness of AI, while Kim et al. (2023) emphasized regulatory compliance of AI as a driver and Lindebaum et al. (2020) discussed its use to mitigate risk. This variation indicated that transparency's application to engender trust in AI decision-making might shift depending on context; which could be to enhance usability, address ethical concerns, satisfy legal standards, or to address risk, for example.

2.4.2.3. Interpretation of Literature on Transparency for AI Decision-Making: The current trend of AI complexity is ever-increasing as new computational power becomes available to system architects. It was an important observation and consideration therefore, that the emphasis on transparency for decision making requires urgent attention. A lack of transparency was also identified as a liability to AI decision-making by Lindebaum et al. (2020), which would infer a future scenario of increased liability in AI decision-making unless transparency gets sufficient focus. The analysis further suggested that as AI systems become more agentic and complex, achieving transparency might require more sophisticated approaches to meet regulatory, ethical, and trust-related needs. Lastly, there was agreement that transparency assists in AI decision-making, which also reinforced transparency as not only a trust enabler but also as a necessary safeguard to maintain accountability in high-stakes decision environments.

2.4.3 Section Conclusion

This section explored the role of transparency in fostering trust in AI and facilitating AI-based decision-making. It also underscored the importance of transparency for the

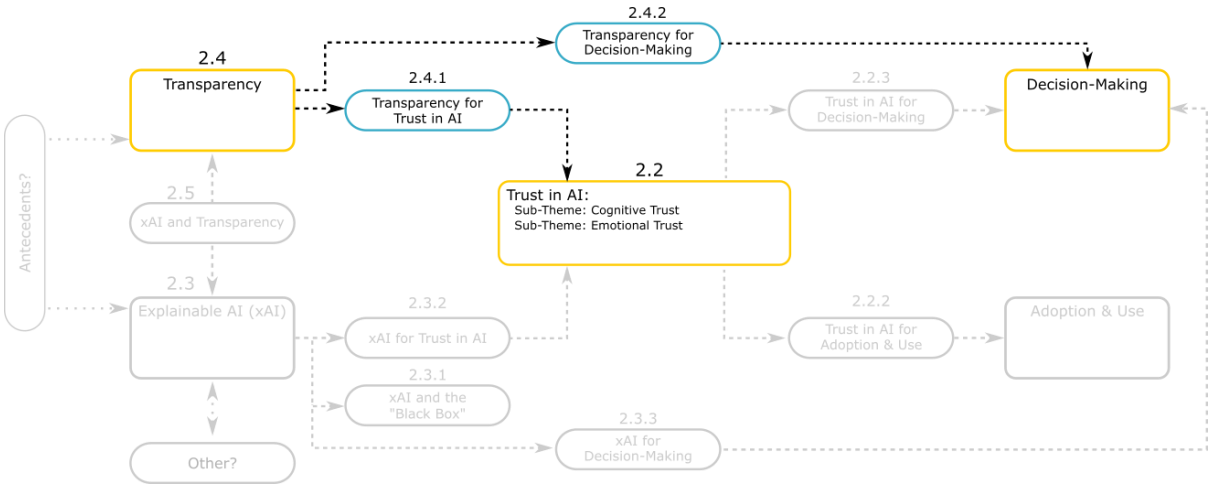
responsible adoption of high-complexity AI agents or systems in decision-processes. Another area for further research was identified and given below, extending from section 2.3:

i.) Glikson & Woolley (2020) and Sullivan et al. (2022) emphasized a need for further research into transparency as an emotional dimension of trust in AI, suggesting that understanding how transparency affects emotional trust in AI could provide deeper insights into the dynamics of trust in AI and also in addressing the “uncanny feeling”.

Figure 5 shows the conceptual relationships developed thus-far through the discussion, analysis, interpretation, and conclusion of section 2.4.

Figure 5:

Conceptual Framework Development Contribution of Section 2.4



Note: Author’s Own

2.5 Understanding the Relationship between xAI and Transparency

2.5.1 Literature on the Relationship between xAI and Transparency

2.5.1.1. Discussion of Literature on the Relationship Between xAI and Transparency: Thus far, the concepts of xAI and transparency were presented mostly in isolation in previous sections of this chapter, with brief insight that they were in some way connected. This section attempts to position the two in relation to each other.

In their article, Sabharwal et al. (2024) discussed transparency and xAI mostly interchangeably, but suggested that explainability of xAI was caused by transparency. Their interpretation implied a causal relationship, where transparency serves as the foundation for explainability within xAI systems. Rai (2020) discussed xAI and transparency interchangeably, in the context of the model itself, but also in the context of the resulting prediction (p. 140). The author further proposed that the overarching goal of xAI is to transition “black box” AI systems to “glass box” AI systems (p. 139).

Haque et al. (2024)'s SLR presented transparency as an outcome or an effect of an xAI system, while Laato et al. (2022)'s SLR represented transparency as a goal of xAI. These differing views underscored a conceptual ambiguity, positioning transparency either as an end result or as a guiding principle of xAI. However, Kim et al. (2023) positioned xAI as a precursor to transparency in "black box" systems, and also called for urgent inter-disciplinary theoretical development in the outcomes areas of xAI (p. 1306). This highlighted xAI's role in enhancing transparency, especially in complex systems, and suggested a need for broader research.

Furthermore, Glikson & Woolley (2020), within their preamble definition of transparency in their SLR, positioned explainability of AI as an "important aspect of transparency" (p. 631). The definition which they gave framed explainability as a subset or part of transparency, suggesting that clear explanations are integral to achieving transparency in AI. Lastly, Gregor (2024) discussed both concepts interchangeably, albeit as separate "principles" (p. 52), which presented transparency and xAI as complementary but not entirely overlapping constructs.

2.5.1.2. Analysis of Literature on the Relationship Between xAI and Transparency: Comparing the manner in which the above literature presented and discussed the two concepts of xAI and Transparency relative to each other, revealed some interesting insights into the relationship.

Sabharwal, et al. (2024), Rai (2020) and Gregor (2024) similarly discussed the concepts of transparency and explainability (in relation to xAI) interchangeably and did not draw significant distinction from one to the other. This similarity suggested some degree of conceptual overlap, implying that transparency and explainability might often be seen as either complementary concepts, or actually the same concept.

On the other hand, several scholars viewed transparency as an outcome or an implication of explanation (Glikson & Woolley, 2020; Kim et al., 2023; Haque et al., 2023; Laato, 2022), while Sabharwal et al. (2024) suggested the opposite, that explainability of AI is caused by transparency. These differences revealed some ambiguity in the understanding of xAI and transparency. The literature did not appear to agree whether the concepts were similar, different or the same, or which one might lead to or cause the other.

2.5.1.3. Interpretation of Literature on the Relationship Between xAI and Transparency: The analysis revealed a conceptual ambiguity in how xAI and transparency are defined and applied across contexts, suggesting that further research could clarify whether one is a subset of the other or if they operate as distinct yet complementary elements. The overall observation was that the two concepts were closely related. Although there may be slight differences in the positioning of the concepts relative to each other, what was clear was

that they have the same overarching goal of transitioning “black box” AI systems to “glass box” AI systems (Rai, 2020, p. 139). Their shared objective of fostering clarity in AI systems highlighted a unified purpose, positioning both as essential for increasing user trust and confidence in AI.

2.5.2 Section Conclusion

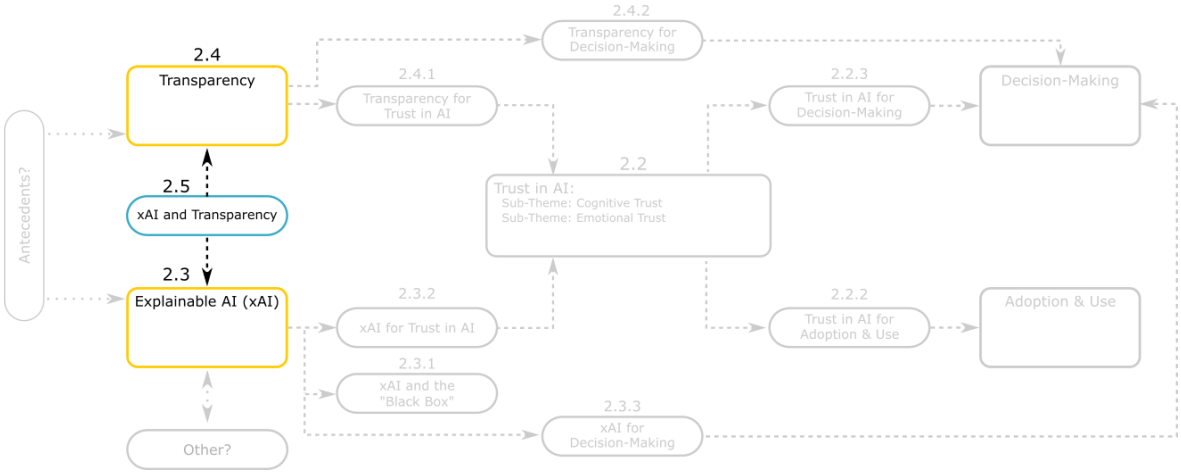
When viewed in the context of the preceding sections, we found that both xAI and Transparency are closely related concepts. Both shared the goal of addressing opacity of AI systems by enhancing trust in AI of those individuals responsible for decision-making. Additionally, an extra area for further research was identified, adding to section 2.4:

j.) Kim et al. (2024) called for additional theoretical development in the outcome areas of xAI, across different disciplines, including “organisational science” (p. 1306).

Figure 6 shows the conceptual relationships developed thus-far through the discussion, analysis, interpretation, and conclusion of section 2.5.

Figure 6:

Conceptual Framework Development Contribution of Section 2.5



Note: Author’s Own

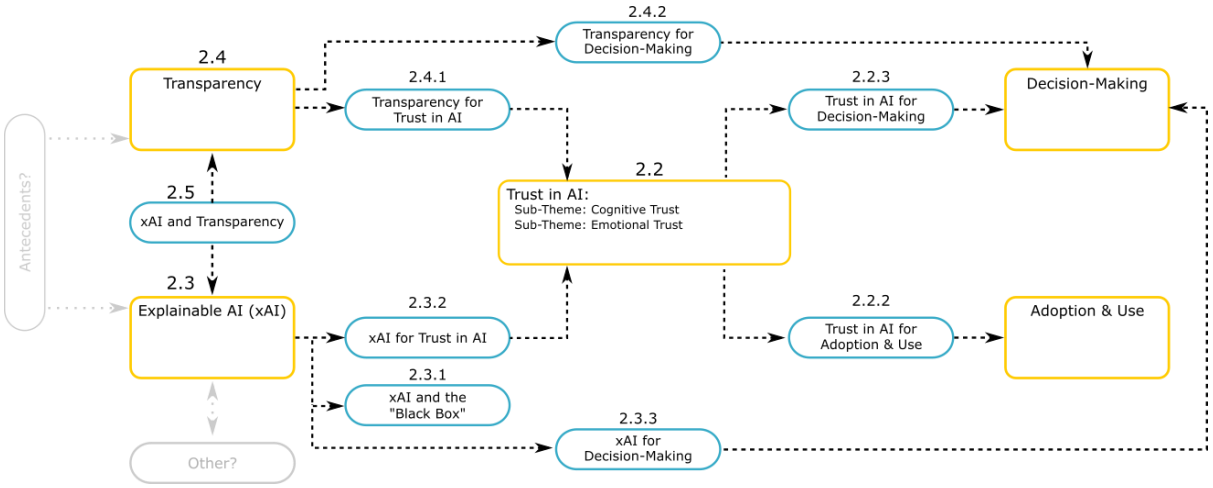
2.6 Conclusion to the Literature Review

Through the sections and sub-sections of this literature review a conceptual framework of the scholarship landscape surrounding trust in AI, xAI, transparency, AI decision-making and the adoption and use of AI in organisations was unveiled. What became apparent through the review of the extant literature was that each of these concepts was interconnected and could not be viewed in isolation from one another, or exclusively. This was demonstrated by the three SLRs (Glikson & Woolley, 2020; Laato et al., 2022; Haque et al., 2023) in combination with the additional literature presented. The conceptual framework which was developed through this review is provided in Figure 7 for clarity, and to draw together the

individual concepts from the preceding sections to highlight the identified relationships and key insights between them.

Figure 7:

Conceptual Framework of Extant Literature.



Note: Author's own

A common theme throughout the analysis and section conclusions was the important role which transparency plays in building cognitive trust in AI systems, while the impact of transparency on emotional trust in AI remains unexplored for the most part (Figure 7, 2.4.1 and 2.2). xAI was also identified as a key component of engendering trust in AI (Figure 7, 2.3 and 2.3.2) and is related with the concept of transparency (Figure 7, 2.5). These two elements were shown to have a relationship both in creating trust in AI, as well as aiding in its adoption and use in organisations, and for decision-making (Figure 7, 2.2.3; 2.2.2; 2.3.3; and 2.4.2).

Transparency and xAI are not only foundational for trust in AI, but they bridge the gap between the complexity of AI algorithms and "black box" AI systems and enable stakeholders to develop trust in AI outputs, decisions, and recommendations. However, the scholarship had identified and called for further understanding of the factors driving trust in AI agents which may lead to additional insights beyond these main concepts, either as antecedents to their development (Figure 7), or as other observations or findings in addition to explainability and transparency. Due to the exploratory nature of the research, it was important to note that the research questions might also have revealed additional insights beyond the immediate dimensions of the conceptual framework.

Having prepared the conceptual framework in Figure 7 as an overview of the reviewed literature, the relationships between the key concepts (trust in AI, xAI and transparency) became clearer. Combining the conceptual framework with the identified areas for further research (designated "**a**" through "**j**") from each of the preceding sections, allowed focused and relevant development of the research questions, ensuring alignment with the literature.

This approach ensured *quality*, *rigor* and *relevance* to the theoretical literature and the research problem. As a result of this approach, three main research questions were developed to address areas of paucity in each of the main concepts, while remaining open-ended so as to gain further insights and understanding into the relationship between each. The research questions are presented in Chapter 3 which follows.

Chapter 3: The Research Questions

The primary research problem, as outlined in Section 1, informed the literature review and the three identified concepts which were subsequently investigated (As presented in chapters 2.2, 2.3, 2.4 and 2,5). Through extensive research of these concepts, several areas in the scholarship were identified where knowledge was either not yet well established or where scholars were calling for additional contribution. These areas for further research were then summarised in the conclusions of the preceding sections and used to develop a set of proposed research questions. The research questions were developed to potentially reveal additional insights which may possibly contribute to the theoretical literature, and aid in addressing the research problem.

3.1 Research Question 1: How does trust in AI lead to its adoption and use for organisational decision-making? (Vanneste & Puranam, 2024; Sullivan et al., 2022; Wong et al., 2024; Enholm et al., 2021)

The aims of RQ1 are to develop new insights and understanding into the mechanisms of trust formation in AI in organisations, and to therefore develop new insights as to how trust in AI leads to adoption and beneficial use of AI for organisational decision-making.

3.2 Research Question 2: How does xAI engender trust in AI for decision-making in organisations? (Rai, 2020; Sabharwal et al., 2024; Lukyanenko et al., 2022) And, by Extension: ***What are the factors that contribute to trust in AI?*** (Sullivan et al., 2022; Balasubramanian et al., 2022; Abedin, 2022; Glikson & Woolley, 2020; Lukyanenko et al., 2022; Pumplun et al., 2023)

The aims of RQ2 are firstly to explore and understand how xAI contributes to the formation of trust in AI for organisational decision-making. Secondly, to explore additional insights and understanding of the factors which the individual responsible for organisational decision-making finds relevant to engendering their trust in AI for decision-making.

3.3 Research Question 3: How does transparency influence emotional trust in AI for organisational decision-making? (Glikson & Woolley, 2020; Wang & Ding, 2024; Wang et al., 2016; Berente et al., 2021)

The aim of RQ3 is to gain additional insights and understanding of transparency as an emotional dimension of trust in AI, in the context of the study around organisational decision-making.

An additional aim of the research was to develop a conceptual framework as an outcome of the overall study.

Chapter 4: Research Methodology

A matrix of the methodology section, denoting section headings and sub-headings is provided in Figure 8, for an overview and ease of navigation.

Figure 8:

Research Methodology Matrix

Note: Author's Own

4.1 Choice of Methodology

4.1.1 Ontology

Trust in AI is formed from the beliefs and experiences of individuals (Vanneste & Puranam, 2024; Lee & See, 2004, p. 53), and the proposed research questions all seek to develop additional insights and understanding of how this trust in AI is experienced and *constructed* by individuals responsible for organisational decision-making, through their own "meaning-making" and subsequent "actions" (Bell et al., 2019, p. 12). The research questions are therefore underpinned by an ontological stance of *social constructionism.* Bell et al. (2019, p. 13) explain that "constructionism suggests that the categories people use to understand the natural and social world are in fact social products", or that individuals create their reality. Therefore, our understanding of what the individuals responsible for organisational decision-making revealed as their experiences and/or factors contributing to their trust in AI was influenced by the social constructs surrounding them and their individual interpretations and experiences which lead to them acting, reinforcing the ontological choice of *social constructionism*.

4.1.2 Epistemology

Saunders et al. (2009, p. 121) suggested that "interpretivism is the way we as humans attempt to make sense of the world around us", which is aligned with a phenomenological approach that explores actions from the perspective of the actor (Bell et al., 2019, p. 16) and that "refrains from judgements of reality until they are demonstrated" (Creswell, 2007, p. 58).

It is important to note that this study uses a *phenomenological approach*, which is not the same as *phenomenology*. Creswell, (2007) describes a *phenomenological approach* as being suited to research which seeks to "understand several individuals' common or shared experiences of a phenomenon" (p. 60). Whereas Saunders et al. (2009) explain *phenomenology* as "the way in which we as humans make sense of the world around us" (p. 116). A phenomenological approach therefore borrows principles from phenomenology (Creswell, 2007) to explore lived experiences, while phenomenology itself is a deeper philosophical way to understand these experiences.

This phenomenological approach was appropriate to understanding how individuals responsible for organisational decision-making interact with and experience AI in relation to the world around them. To address the research questions, the proposed methodology therefore followed an *interpretivist* epistemology which was aligned with understanding and exploration, as opposed to a positivist approach which would have endeavoured to explain (Bell et al., 2019). Lastly, the interpretivist view was applicable to this research as it asked "how" and "what" type questions, which were aligned with the explorative approach of the research design (Bell et al., 2019; Braun & Clarke, 2006).

4.1.3 Research Approach

Given the proposed research's foundation in the interpretivist paradigm and having a phenomenological approach, it employed a qualitative, cross-sectional design, comprising primary data (semi-structured interviews). As such, the design allowed for triangulation of data and insights between multiple interviews, which ensured validity (Schwandt et al., 2007), *credibility*, *dependability*, and *rigour*. An inductive and deductive reasoning approach was followed to make the "conceptual leap" between the collected data and extant theory (Klag & Langley, 2013, p. 149). This allowed development of understanding as per the research aims, while it further allowed for comparison with, and potential refinement or extension of, extant literature (Crane et al., 2016, p. 4). The field of AI in management and strategy scholarship is still in its infancy, and as such, theory is still developing, and studies were few (Wong et al., 2024; Enholm et al., 2021; Glikson & Woolley, 2020). These choices therefore allowed rich analysis of factors contributing to AI trust to be explored, potentially contributing to the existing theoretical knowledge base (Crane et al., 2016).

4.2. Research Setting

The setting is the context in which the study is to be conducted, as associated with the research question (Bell et al., 2019). The research setting chosen was therefore one of worldwide organisations. The setting was not geographically limited due to the global accessibility of AI, which is not country-specific. The research questions focused on an

organisational level of exploration, as the study focused on an organisational level of decision-making.

Further, the setting included diverse sectors such as (but not limited to) financial services, healthcare, manufacturing, and telecommunications. The diversity of sectors allowed for unique insights through triangulation of the data (Bell et al., 2019) across the different domains. Bell et al. (2019) indicate that it is useful to triangulate data across different domains, therefore the choice of conducting the study across different sectors. This is also called for in the reviewed literature (Rai, 2020; Sabharwal et al., 2024; Chen et al., 2021).

4.3 Level and Unit of Analysis

The level of analysis was described by Siebert et al. (2004) as the "specific focal unit under consideration" (p. 334). As the research questions focused on organisational decision-making (the focal unit), the **level of analysis** chosen was the organisational level. In addition, the overarching research problem identified by Sullivan et al. (2022) related to organisational adoption of AI and therefore further supports the organisational level of decision-making as the focus of enquiry.

Bogie, (2024) defined the unit of analysis as, "The unit of analysis is who or what you use to gather data…" (p. 12). This meant that the **unit of analysis** was the individual, because interviews were conducted with individuals who had knowledge and experience of the research phenomenon, as outlined in the research design which assumed a phenomenological approach.

4.4 Sampling Method, Size and Criteria

4.4.1 Sampling Method

For this qualitative study, ***purposive sampling*** ensured appropriate focus to address the research question(s), as it is ***non-probabilistic*** and ensured relevance of the selected sample to the designed research Bell et al. (2019). This also ensured that the selected sample group had the required experience to provide insights relevant to the research. Additionally, Saunders et al. (2009) identified ***snowball sampling*** as a method to extend the research group through referral by the original invitees. Although not a preferred strategy to generate the required sample size, due to the potential introduction of *bias*, this technique was identified to be subsequently employed if initial response rates were problematic (Saunders et al., 2009).

No difficulties were experienced in reaching the target sample size (outlined below), however, a few participants suggested including others from their professional networks with related experience of the research phenomena. These opportunities for additional insights were not overlooked, and therefore an "indirect" snowball sampling occurred which added an additional four participants to the dataset. Diversity criteria did not need to be applied to

mitigate the identified risk of *bias* (Braun & Clarke, 2021) as the resulting sample set was sufficiently diverse in and of itself.

4.4.2 Sample Size

For qualitative research using purposive sampling, there was no fixed rule for sample size, however, Bell et al. (2019) suggested a size guideline range of between eight (8) to 15 interviews as being appropriate for student research (p. 119). A “matrix approach” (Bogie, 2024) was therefore used, with four (4) organisations of five (5) participants each initially being targeted, giving an indicative sample size of 20 participants, in line with the recommendations of Bogie, (2024).

While the planned sampling was expected to yield between 16 to 20 participants from four organisations and four sectors, the final sample yielded 19 participants from a more diverse range of 18 organisations and 16 sectors. The associated interviews of the participants therefore allowed for exploration of the experiences and understanding of trust in AI, of individuals responsible for organisational decision-making, across diverse sectors and organisations. This broader sample remained well suited to exploratory research which values a diversity of perspectives (Braun & Clarke, 2021). The list of participants and their organisational sectors is provided in Table 1 below.

Table 1:

Overview of Participant Decision-Making Levels and Sectors.

	Participant	Sector
1	Participant 1	Business Consulting
2	Participant 2	Chemical Industry
3	Participant 3	Telecommunications
4	Participant 4	Pharmaceutical Manufacturing
5	Participant 5	Transport & Logistics
6	Participant 6	Engineering Consulting
7	Participant 7	Legal Advisory
8	Participant 8	Retail and Consumer
9	Participant 9	Defence
10	Participant 10	Aviation
11	Participant 11	Transport & Rail Infrastructure
12	Participant 12	Human Resource Management
13	Participant 13	Software Development
14	Participant 14	Healthcare
15	Participant 15	Retail Banking
16	Participant 16	Telecommunications
17	Participant 17	Investment Banking
18	Participant 18	Investment Banking
19	Participant 19	Retail Banking

Note: Author’s Own.

4.4.3 Sample Criteria

A ***heterogeneous*** sample of individuals responsible for organisational decision-making, from various sectors and with diverse experiences, was viewed as appropriate to the interpretive paradigm, because of the type of the research questions, and the focus on “key themes” (Saunders et al., 2009, p. 234). This allowed for rich cross-sectoral comparison of findings and identification of similarities and differences, which led to additional insights and understanding. The purposive sampling criteria which were proposed for identification of participants with experience relative to the research questions (Bell et al., 2019) were:

- Must be from the research setting of worldwide organisations.
- Must be from a range of sectors such as (but not limited to) financial services, healthcare, manufacturing, and telecommunications.
- Must be an individual responsible for organisational decision-making within the above organisations.
- Must be exposed to and/or have experience of artificial intelligence and its influence on their organisation.

By focusing on individuals responsible for organisational decision-making who were most-likely to influence trust in AI, the research design ensured *relevance* of the qualitative data gathered. The sampling approach also aimed to cover a broad spectrum of industry contexts, while controlling for unwanted influences specific to each industry. For example, finance may have had a higher adoption rate of AI than manufacturing, but possibly with the same factors influencing levels of trust. Therefore, the selected population enhanced the ability to identify both unique and common factors influencing trust in AI, across different organisational environments. The deliberate decision to study firms from different industries spoke to the need for *external validity* and *transferability* of the findings, which ensured that the results of this study and the factors of trust which were identified could be *generalised* to other senior decision-making processes in different contexts (Bell et al., 2019, p. 40). By selecting firms across industries, common trust factors emerged which were independent of the sectoral setting.

4.5 Research Instrument

For the identified unit of analysis, the focus of all the research questions was on the personal experiences of trust in AI of the individual responsible for organisational decision-making. Therefore, the research instrument employed was a ***semi-structured interview*** (protocol in **Appendix A**) of the participants (for a primary data source) who were selected through the purposive sampling criteria. This was appropriate to qualitative research and an interpretivist approach (Bell et al., 2019).

The research protocol was developed to avoid the possibility of the researcher asking the questions differently or inconsistently between the participants, thus adding to the reliability of data (Bell et al., 2019). However, the choice of semi-structured interviews did allow some flexibility in choice and ordering of the question set depending on the conversation flow (Saunders et al., 2009). In line with Josselson (2013, p. 35) and "know[ing] the whole person in relation to some question," each interview allowed for exploration of the participant's, experiences, and understanding of trust in AI as it pertained to each research question, and what they saw as being of importance and relevance to the question (Bell et al., 2019). Broad and general questions were preferred "so that the participants can construct the meaning of a situation" (Creswell, 2007, p. 21) and particular attention was given to avoiding the "*Hawthorne Effect*". This is explained in Bell et al. (2019, p. 47) as the effect of the interviewer or topic on the responses of the interviewee.

With the Interview protocol being designed in line with the recommendations of Josselson (2013), questions were chosen to be open-ended to "invite exploration" (p. 51). An opening and closing *little q* question was developed to firstly set the context and to begin the interview and lastly, to reveal any closing insights which the *big Q* questions did not manage to extract (Josselson, 2013). A total of nine (9) big-Q questions were used, which spanned the three (3) research questions. Each interview question was carefully considered, to allow a degree of overlap in the participant responses with respect to each research question, and to reveal insights and understanding of the relationships *between* the research concepts of trust in AI, xAI and transparency. Additionally, probing questions and prompts were used to "follow up" particular answers and to "explore further what has been said" (Bell et al., 2019, p. 153). A "pilot interview" was conducted to get a feeling for the flow of the questions and the transcription process, which identified any potential errors in question logic ahead of the main interviews (Josselson, 2013).

4.6 Data Gathering Process – (Semi-Structured Interviews)

The research instrument described in Section 4.5 was used to gather data from all the participants, allowing triangulation during the analysis and findings, which supported quality and credibility of the research (Bell et al., 2019). The data gathering process only began following confirmation of ethical clearance having been granted by the GIBS Research Ethics Committee. Ethical clearance approval was granted, with no required amendments, and was received via email notification on the 16th of July 2024. A copy of the ethical clearance approval is included for reference in **Appendix B**. Following receipt of the ethical clearance approval, interviews with participants, selected using the sample criteria in section 4.4.3, were then arranged.

Starting on 17th July, potential participants were emailed to invite them to participate in the research (Sample email in **Appendix C**). The email included a list of options for potential interview dates and times, so that the participant could select a convenient slot on the first attempt, thus avoiding the need for lengthy coordination and back-and-forth, and respecting their time. Most participants agreed on the first attempt, and a minimum of one week was allowed before sending a follow-up request to those who hadn't yet replied. If no response was received to the follow-up email, the participant was not contacted again, and their privacy was respected.

On receipt of confirmation of a suitable date and time, a Microsoft Teams invite was sent via Microsoft Outlook, which included confirmations of the confidentiality and anonymity of the research, and the informed consent letter (see **Appendix D**) as an attachment. Conducting the interviews online over Microsoft Teams meant that the participants did not need to install any software (Saunders et al., 2009). A request to return the informed consent letter ahead of the interview was also included in the invite. An example of the calendar invite is included in **Appendix E**. If the signed informed consent letter had not been received by the day before the scheduled interview, a polite reminder was sent, with a last reminder two hours before the interview started. No participants failed to provide the required informed consent, and none re-scheduled for any reason.

All interviews were scheduled between 26th July and the 23rd of August, with one final interview being added later on the 9th September. The interview process was limited to a single interview per participant, which was scheduled not to exceed 90minutes in length. The researcher successfully kept to this single interview to avoid breaking the flow of the conversations, thereby to avoiding the need to revisit the prior discussion as this would have been (inadvertently) leading, impacting *quality* and *rigour*. In total, 19 interviews were conducted, with an average length of 45minutes. The shortest interview was only 26minutes in length but still yielded valuable insights and new codes, while the longest interview was 83minutes long.

Data was recorded on two separate audio recording devices and was transcribed and redacted within an as short as practicable time window after conclusion of the interview (Bell et al., 2019). Each interview was video recorded (with the permission of the participant) using the built-in functionality of Microsoft Teams, with a duplicate audio recording being made using a separate mobile phone voice recording app.

The built-in voice to text transcription of Microsoft Teams was also used to generate an MS word file of the interview data and an independent transcriber was not used. Therefore, the non-disclosure agreement which was included in the Ethical clearance application was not

required. The researcher opted instead to review each of the transcripts themself, editing any garbled or inaccurate wording by reference to the original video recording. During this review process, the researcher redacted and anonymised names of participants, the participant's colleagues, organisational names, internal organisational tool names (like apps and software) and any other potential identifiers, in keeping with the requirements of the informed consent and ethical clearance. This immersion in the data strengthened the researcher's familiarity and understanding further (Braun & Clarke, 2006, pp. 87-93).

Audio and video files from the interviews were deleted following completion of the Chapter 5 data analysis. Only the transcribed, anonymised and redacted primary data sets were submitted to GIBS. Data shall be stored for a minimum period of 10 years on a physical USB flash drive with fingerprint (biometric) access, to negate the risk of forgetting any passwords and to ensure security of data access. Cloud storage was and shall not be used, to avoid any potential risk of data breach. No local copies of data have been retained on any work or personal computer drives or desktops, and all emails and invites to the participants have been deleted to avoid any risk of harm related to any identifiers included in them (e.g. email addresses, participant names and organisation names.)

4.7 Data Analysis Approach

4.7.1 Thematic Analysis

Thematic analysis (Braun & Clarke, 2021) was considered suitable for analysing the data from this qualitative study. Its analytical approach is a common method for analysing interview data, and it is well aligned with the requirements of exploratory research. Thematic analysis is described as a reflexive method, employing flexibility of interpretation which "seeks to develop 'patterns' (themes, categories) across cases" in qualitative research (Braun & Clarke, 2021, p. 37). This makes it suitable for exploratory studies where understanding recurring themes is important. In addition to its flexibility, another advantage of thematic analysis is its ability to "highlight similarities and differences across the data set" (Braun & Clarke, 2006, p. 97). While Bell et al. (2019) raised a concern about the lack of clarity regarding the procedures of thematic analysis, they did support that its flexibility would be an advantage to this type of qualitative interpretivist research. By designing a systematic and structured approach to the data analysis, this interpretive flexibility allowed the researcher to identify patterns of meaning and themes without "rigid rules" (Braun & Clarke, 2006, p. 82), in relation to the research questions, using their own interpretation and understanding of the data.

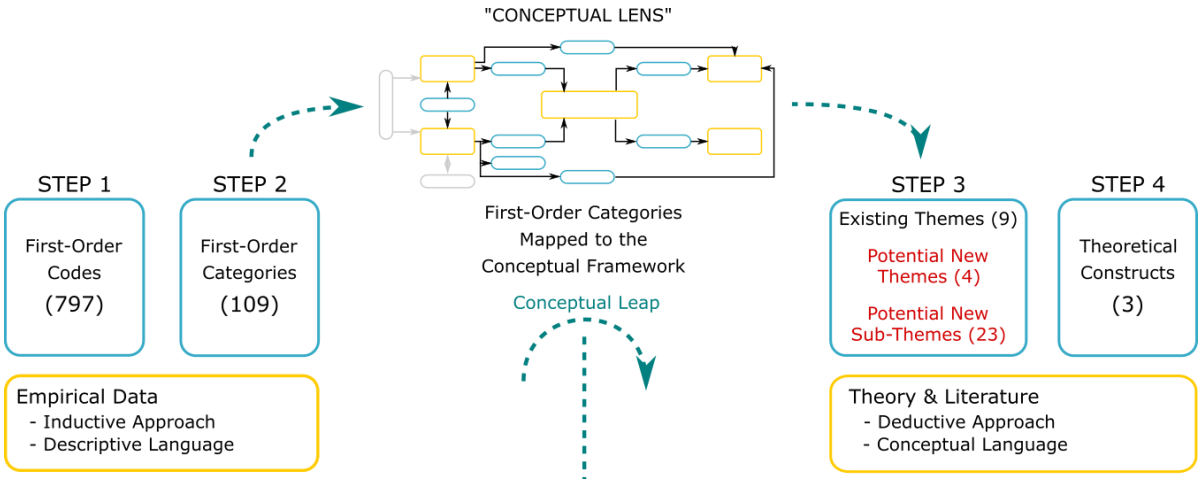
To begin the analysis, the interview transcripts were imported into a new project file in ATLAS.ti and re-named by participant number. The researcher spent time reading through each transcript, while watching the related video recordings back as a process of

familiarisation, before the subsequent analysis. During this process, the built-in editing function of Atlas.ti was used to further redact any previously-missed possible identifiers, and to clean up any further inarticulate areas of the data (with reference to the original recordings). Braun & Clarke (2006, pp. 87-93) outlined the process of "familiarisation, initial coding, theme search, review and naming and subsequent report production", followed by interpretive analysis, which creates structure to the analysis, ensuring *rigour* while creating *transferability*. While reading, the researcher also actively checked for any additional redactions/anonymisations and ensured that the transcripts accurately reflected the participant's experiences and contributions. This familiarisation and review ensured *quality* while addressing any gaps and required corrections between the videos and transcripts (Braun & Clarke, 2019).

After becoming familiar with and organising the data, the data analysis process began. This data analysis process followed four structured steps, which combined both inductive and deductive approaches (Braun & Clarke, 2021; Klag & Langley, 2013). The Conceptual Framework, Figure 7 from the Chapter 2 literature review, was developed through the researcher's understanding of the extant literature on Trust in AI, xAI and Transparency. This Conceptual Framework was therefore used as a specific "conceptual lens", through which to organise and interpret the findings in relation to the extant literature, and served as the "underlying structure, the scaffolding or frame" of the study (Merriam & Tisdell, 2016, p. 85). The Conceptual Framework incorporated the key concepts of Trust in AI, xAI and Transparency and was directly related to the research interest. This bridged the gap between the empirical data and theory, thereby facilitating the "conceptual leap" (Klag & Langley, 2013) from the descriptive language of the codes and categories to the conceptual language of the literature. An overview of the four-step process is provided in Figure 9, and discussed further as follows.

Figure 9:

Overview of the Four-Step Process Used for the Thematic Analysis



Note: Author's Own, prepared with consideration of Klag & Langley, (2013) and Merriam & Tisdell, (2016).

This approach and combination of steps resulted in a robust structure for both the presentation of the Chapter 5 findings and the subsequent Chapter 6 discussion of the findings in relation to the literature. A further, more detailed example of the step 1 to step 4 flow is provided in **Appendix G,** showing the actual themes, sub-themes and constructs, as later discussed in Chapters 5 and 6.

The first and second steps of this four-step process were primarily performed using ATLAS.ti. These two steps were descriptive, using business language, and inductively drew from the data. The first step was to systematically work through and inductively analyse the semi-structured interview transcripts to identify codes, patterns, and themes (Braun & Clarke, 2006) related to the literature and research questions. This analysis step was appropriate to an exploratory, qualitative study (Bell et al., 2019; Braun & Clarke, 2021; Klag & Langley, 2013) and was therefore relevant to the proposed research. Researcher bias was acknowledged, and emphasis was placed on letting the data speak, as opposed to fitting it to preconceived notions. Each code captured a "unit of meaning" from the participants' experiences and understanding of the research phenomena. Across the 19 participants, this process captured a total of 829 quotations to which 797 first-order codes were assigned to capture the underlying units of meaning.

In the second step, the first-order codes were reviewed alongside their associated quotations. Codes that highlighted similar findings were then grouped into first-order categories using the ATLAS.ti code group manager. From this second step a total of 109 first-order categories were created from the 797 first-order codes. The book of "Codes and Categories" is included for reference in **Appendix F**.

At this same stage, the transcript documents were grouped into their associated participant groups, as identified in Table 1. This grouping allowed ease of cross-comparison of the codes and categories between participant groups using the ATLAS.ti “Code-Document Analysis” tool. This was valuable in creating deeper insights and understanding, by structuring the participant data for both in-case and cross-case comparison and presentation of the findings in Chapter 5. A summary of the participant groups is provided in Table 2 below. Due to diversity of sectors and for ease of analysis, the participants were grouped by decision-making level within the organisation, i.e. Executive, Senior and Operational.

Table 2:

Participant Groupings Assigned for the Data Analysis and Presentation of Findings

	Participant	Sector	Participant Group
1	Participant 1	Business Consulting	Executive
2	Participant 10	Aviation	Executive
3	Participant 11	Transport & Rail Infrastructure	Executive
4	Participant 13	Software Development	Executive
5	Participant 14	Healthcare	Executive
6	Participant 17	Investment Banking	Executive
7	Participant 18	Investment Banking	Executive
8	Participant 2	Chemical Industry	Senior
9	Participant 3	Telecommunications	Senior
10	Participant 4	Pharmaceutical Manufacturing	Senior
11	Participant 9	Defence	Senior
12	Participant 12	Human Resource Management	Senior
13	Participant 16	Telecommunications	Senior
14	Participant 5	Transport & Logistics	Operational
15	Participant 6	Engineering Consulting	Operational
16	Participant 7	Legal Advisory	Operational
17	Participant 8	Retail and Consumer	Operational
18	Participant 15	Retail Banking	Operational
19	Participant 19	Retail Banking	Operational

Note: Author’s Own.

After identifying the initial codes and categories, the analysis moves from the descriptive language of business and the interviewees to the abstract academic language, theoretical concepts and themes found in the extant literature. This is described by Klag & Langley, (2013), as the "conceptual leap"; the transformative process of moving from empirical observations to abstract theoretical insights, bridging the "world of the field" with "the world of ideas" (p. 150). This approach was considered suitable for an exploratory study, as it allowed for exploration of participants’ experiences while relating the findings to established literature or theory (Bell et al., 2019, p. 55).

The third step involved applying the conceptual lens by mapping the identified first-order categories to the Conceptual Framework. Using this approach, the 109 first-order categories were mapped to nine (9) existing themes from the Conceptual Framework, with four (4) potential new themes and 23 potential new sub-themes emerging from the process. The four potential new themes represented possible distinct differences (extensions) to the literature, and the 23 potential new sub-themes indicated possible nuances of difference (refinements) to the literature (Crane et al., 2016), as discussed in Chapters 6 and 7 to follow.

The fourth and final step mapped the identified existing themes, as well as the potentially new themes and sub-themes to the theoretical constructs and their related research questions. This ensured relevance of the identified themes and topics to the research phenomena being explored and built a robust framework for discussion between the findings and the literature in Chapter 6. The approach was useful in identifying/providing the key themes related to the research questions, across a broad view of the extant literature and allowed further development of the conceptual framework (Saunders et al., 2009, p. 61) from Chapter 2.

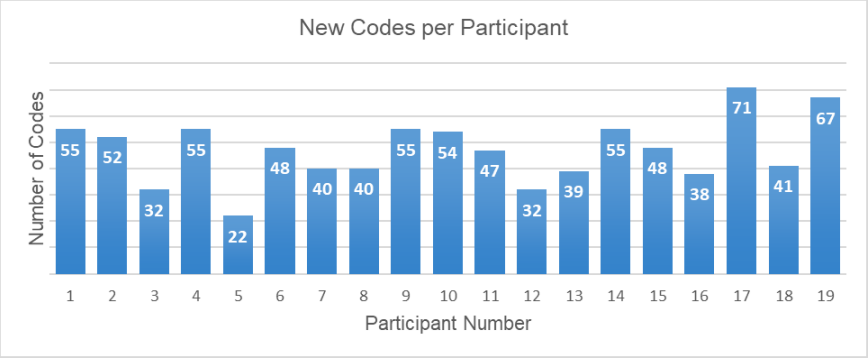
4.8 Data Saturation

Data saturation is described by Bell et al. (2019) as having been achieved when no new codes are being generated from the data. This view is similar to Saunders et al. (2009) who describe data saturation as “the stage when any additional data collected provides few, if any, new insights” (p. 590). Although 19 interviews were conducted, new codes continued to emerge. A possible reason may be due to the richness of insights, and the diversity in participant perspectives in the data.

Data saturation was therefore not achieved, and within the time constraints of this masters-level research, it was not possible to conduct more than 19 interviews. However, a level of detail was achieved which allowed the analysis to proceed with a meaningful depth of codes and a rich set of code groups. Figure 10 shows the total number of codes identified per participant interview.

Figure 10:

Number of Codes Revealed per Participant Interviewed



Note: Author's Own. Generated in Microsoft Excel, using output from ATLAS.ti.

4.9 Quality and Rigour

The research methodology and design approach was chosen to ensure *reliability*, *replicability*, and *validity*, the three main "*quality* criteria in business and management research" (Bell et al., 2019, p. 39). By using purposive sampling and the identified selection criteria, the research ensured a high-*quality* data set which was *reliable*, *replicable*, *valid,* and directly relevant to the research questions (Braun & Clarke, 2006). The systematic and defined steps in the proposed data collection and analysis process added *rigour* to the study (Magnani & Gioia, 2023). Additionally, the use of multiple sources of primary data in the thematic analysis allowed for triangulation, thereby strengthening the credibility of the research (Bell et al., 2019). However, Bell et al. (2019) also highlighted differences between qualitative and quantitative requirements and proposed parallels for qualitative evaluation (pp. 89-91), as drawn from Schwandt et al. (2007, p. 12). These are summarised below (quantitative dimensions in brackets) with the features of this proposed research which seek to address them:

Credibility (Internal Validity): Triangulation between semi-structured interview data. A systematic research design was employed, based on literature from high quality academic journals, ranked 4*, 4 and 3 on AJG.

Transferability (External Validity): Purposive sampling was used to ensure a good fit of participants to the research questions and the aims of the proposed research. "Thick description" was used to provide context for the "degree of fit" of the data (Schwandt et al., 2007, p. 19), and a well-constructed *Semi-Structured Interview Protocol* (Josselson, 2013), was used and is provided in **Appendix A**. Lastly, a *Consistency Matrix* was developed and is provided in **Appendix F**.

Dependability (Reliability): All primary data and interview transcripts were retained, as well as submitted to GIBS for scrutiny (Bell et al., 2019, p. 91).

Confirmability (Objectivity): The researcher ensured that they remained neutral and objective throughout the interview process and did not allow “personal values or theoretical inclinations” (Bell et al., 2019, p. 92) to influence the research. Attention was paid to a well-written final report, which was coherently and logically laid out using established academic structure, to demonstrate *rigour* and *quality*.

These dimensions ensured that the proposed research was methodologically sound and was capable of producing valid and reliable findings which may potentially contribute meaningfully to the understanding of the research questions.

4.10 Ethical Considerations

Prior to engaging in outward-facing research, GIBS ethical clearance was attained. In addition, all interviewees and their organisations were afforded anonymity, informed consent, and privacy (Bell et al., 2019) to mitigate any possibility of harm. No known or potential conflicts of interest were identified, and therefore none were disclosed as highlighted in the proposal to ensure transparency.

4.10.1 Additional considerations for ethical clearance:

Question 13 - An independent transcriber was not used, and no transcriber was therefore required to sign the non-disclosure ahead of them receiving any audio recordings. However, the required non-disclosure agreement was still included with the Ethical clearance application at the time of submission, as the need for an external transcriber had not yet been decided.

Question 14 – To ensure confidentiality and/or anonymity no names of individuals or organisations were reported, and this was achieved through redaction of the primary data transcripts. The researcher was also aware of the potential for certain phrases and/or comments (such as a purpose statement) to become an identifier. Therefore, during data preparation and in reading transcripts with this awareness, attention was paid to removing/redacting these potential identifiers at the same time as anonymisation and preparation of the collected and transcribed data. Audio and video files from the interviews were deleted following completion of the Chapter 5 data analysis. All e-mails and digital interaction records with interviewees were deleted on completion of the interviews and data preparation.

Question 15 – No organisational consent was required because any organisation that met the criteria set out in paragraph 4.4.3, under purposive sampling was appropriate for the study. Therefore, no particular organisations needed to be targeted, and the information and data returned was not required to be organisation-specific.

Question 25 – Data shall be stored for a minimum period of 10 years on a physical USB flash drive with fingerprint (biometric) access, to negate the risk of forgetting any passwords and to ensure security of data access. Cloud storage shall not be used to avoid any potential risk of data breach. No local copies of data have been retained on any work or personal computer drives or desktops.

4.11 Limitations of the Research Design

The research had some inherent limitations due to time constraints, geographic setting, and methodological choices, and simply by virtue of being a qualitative study, which is often viewed as being "too subjective" due to the researcher and their "unsystematic views" (Bell et al., 2019, p. 96). However, some of these concerns were mitigated by the research design being structured and systematic (Magnani & Gioia, 2023), and by the application of the *quality* and *rigour* criteria previously discussed. Notwithstanding, the following limitations of the research design remained:

- The research was performed at a master's level and was limited to a time window of five (5) months, from ethical clearance to final submission.
- The researcher was a first-time master's student and despite a previous honours qualification had not yet practiced the subtleties of qualitative interviewing and data collection. In mitigation, an interview protocol was developed using Josselson (2013) as a guide, and best-practice principles were identified and adhered to.
- Due to the use of purposive sampling in the proposed methodology, and the level of analysis and sectoral settings chosen, the findings of the research may not be generalizable across other settings (Bell et al., 2019), inherent to an inductive approach (Saunders et al., 2009, p. 127). However, this was in part mitigated by the use of an inductive-deductive approach (Klag & Langley, 2013) to allow generalisation to theory through deduction, also by application of a "conceptual lens" (Merriam & Tisdell, 2016).
- A sample size was selected which should ensure adequate data saturation, but a wider, more comprehensive study may yet reveal further insights.

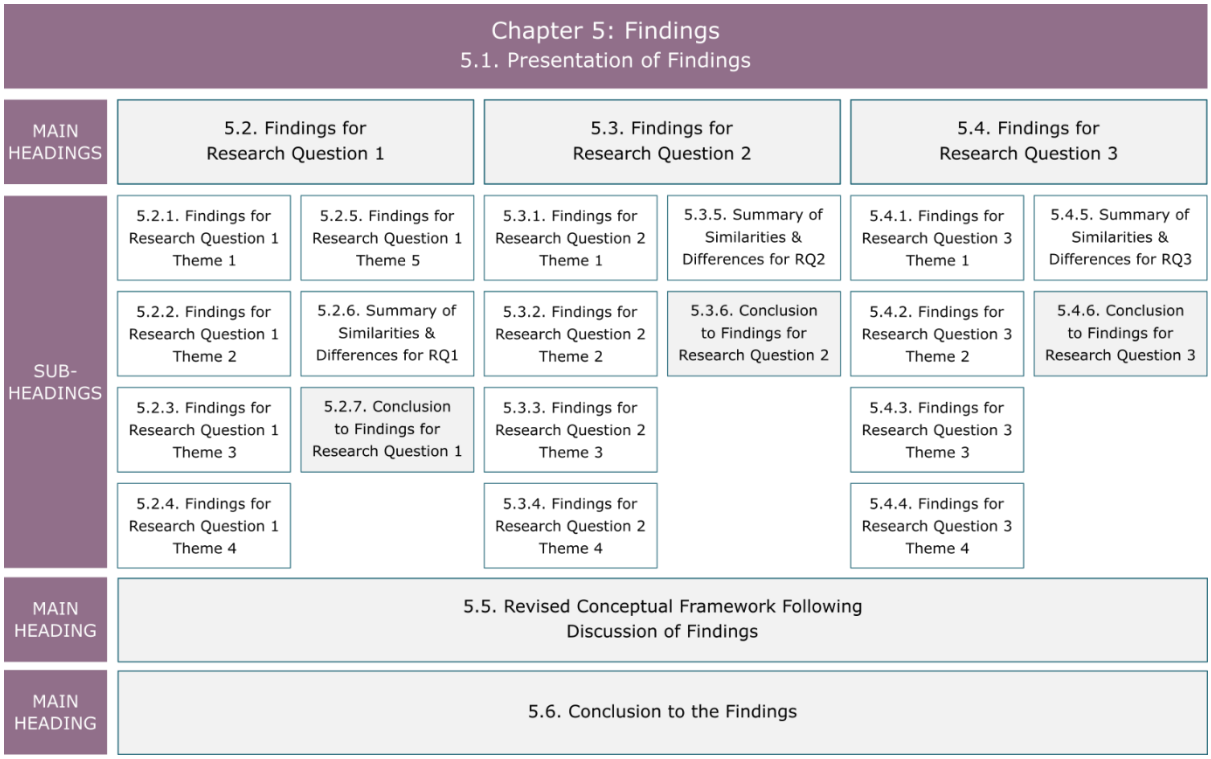
Chapter 5: Findings

5.1 Presentation of Findings

This section presents the findings from the analysis of the participant interviews. The findings which follow are presented in relation to, and in order of the Research Questions discussed in Chapter 3. An outline of the sections and subsections, of the Chapter 5 findings which follow, is provided in Figure 11 below.

Figure 11:

Matrix of Sections and Sub-sections for Chapter 5



Note: Author's Own

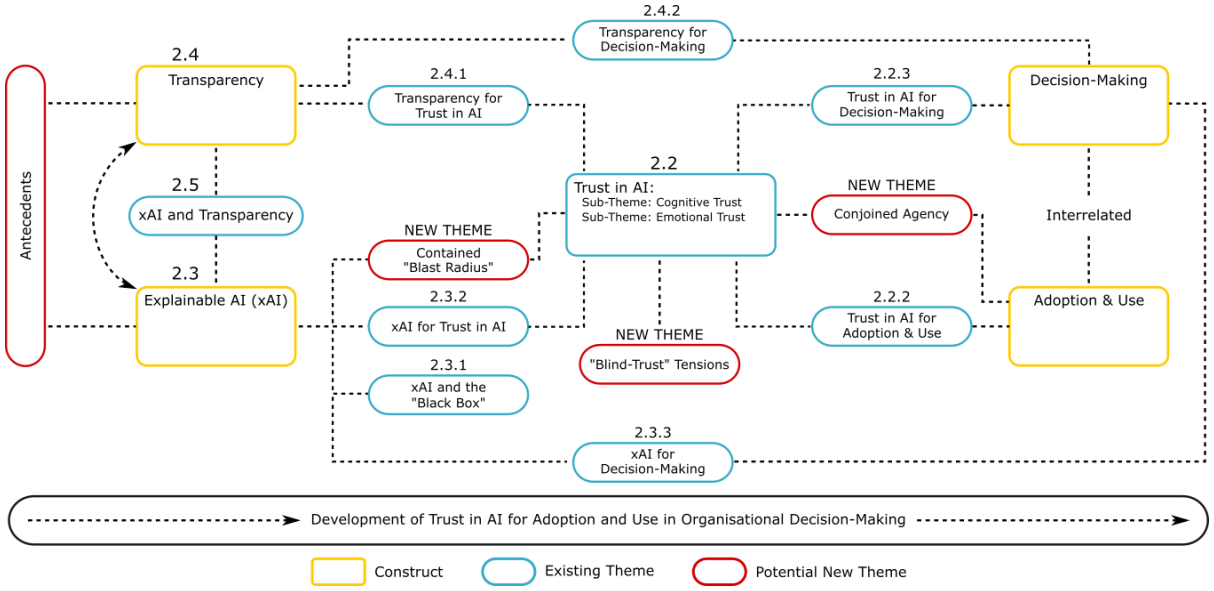
Following the coding and analysis of the data, as described in Chapter 4, Section 4.2.6, a total of 13 themes were mapped to the Conceptual Framework, previously envisaged with nine themes in conclusion to Chapter 2.

The selected themes, presented under each Research Question, were identified as having the potential to provide new insights and understanding of the chosen research topic. Step 3 and step 4 of the mapping performed during the "conceptual leap" (Described in Chapter 4) resulted in revisions to the Chapter 2 Conceptual Framework and an updated representation is provided below (Figure 12), including four (4) potentially new main themes, to guide discussion of the findings. At the end of Chapter 5, a further update to the conceptual framework is provided, introducing the potential new sub-themes discussed in this chapter. Lastly, another update to this same Conceptual Framework is presented at the end of Chapter

6, following comparison of the findings with the extant literature, to confirm which of the potential new themes and sub-themes remain as possible contributions to the body of research.

Figure 12:

Revised Conceptual Framework Showing Potential New Themes



Note: Author's Own

In line with Chapter 4, Table 1, participant groups are marked with their own unique colours as identifiers for ease of reference and navigation, namely: Executive Decision-Making (Blue Group), Senior Decision-Making (Yellow Group) and Organisational Decision-Making (Green Group). Throughout this chapter, the evidence tables are colour-coded accordingly.

At the beginning of each research question section, a table is provided to summarise the themes and sub-themes identified as related to the particular research question. The formatting of these summary tables is consistent with the Conceptual Framework, with potential new themes shown in red text.

In addition to the above, each Research Question section begins with a summary table showing the "frequency of occurrence" of each theme and related topic, by participant group. The researcher notes that depiction of these "frequencies of occurrence" do not reflect any significance but are only included to highlight the most prevalent topics of discussion from the findings, in relation to the research questions. These topics were examined to attain further insights and understanding into the phenomenon explored through their associated Research Question.

5.2 Findings for Research Question 1: *How does trust in AI lead to its adoption and use for organisational decision-making?*

The themes related to the research question will be discussed in this section. As discussed in Chapter 4, differences that arose from the analysis identified some distinct differences and some nuances of difference. The former are shown, in Table 3, as potential new themes and the latter as potential new sub-themes. For RQ1, five (5) main themes emerged from the data analysis, and two (2) of these are potentially new. A further eight (8) potential new sub-themes also emerged.

Table 3:

Themes Emerging from RQ1

RQ1 Theme	Similarities	Distinct Differences	Nuance of Difference	New Topics / Sub-Themes
	Existing Theme	New Theme	Sub-Theme	
Trust in AI	X		XX	Yes (2 New Sub-Themes)
Trust in AI for Adoption & Use	X		XX	Yes (2 New Sub-Themes)
Trust in AI for Decision-Making	X		XX	Yes (2 New Sub-Themes)
Blind Trust Tensions		X	X	Yes (1 New Main Theme and 1 associated sub-theme)
Blast Radius		X	X	Yes (1 New Main Theme and 1 associated sub-theme)

Note: Author's Own

In addition to the above, Table 4 below provides a summary of the frequency of occurrence of each of the five themes, while highlighting the topics most prevalent in the data from the different participant groups. The frequencies of mention are presented not in numbers, but as "Many", "Some", or "Few", as these descriptors are more suitable for qualitative analysis.

Table 4:

Frequency and Main Topics Related to RQ1

RQ1			
SELECTED THEMES	**Executive**	**Senior**	**Operational**
Trust in AI	Many	Some	Few
Main Topics	*Trust Develops Over Time; Trust as a Spectrum*	*Trust Develops Over Time; Trust as a Spectrum*	*Trust Develops Over Time; Trust as a Spectrum*
Trust in AI for Adoption and Use	Many	Many	Many
Main Topics	*Early Adoption & Experimentation; Change Management Engenders Trust*	*Early Adoption & Experimentation; Change Management Engenders Trust*	*Early Adoption & Experimentation; Change Management Engenders Trust*
Trust in AI for Decision-Making	Many	Many	Many
Main Topics	*Only to Inform Decisions; Output Reliability & Trust*	*Only to Inform Decisions; Output Reliability and Trust*	*Only to Inform Decisions; Output Reliability and Trust*
"Blind-Trust" Tensions	Few	Some	None
Main Topics	*Blind-Trust "Yes"; Blind-Trust "No"*	*Blind-Trust "Yes"; Blind-Trust "No"*	*None*
"Blast Radius"	Many	Some	Many
Main Topic	*"Blast Radius"*	*"Blast Radius"*	*"Blast Radius"*

Note: Author's own.

5.2.1 RQ1: Theme 1 – Trust in Artificial Intelligence

This theme is a main theme from the Chapter 2 literature review and is reflected in Figure 12 at the start of Chapter 5. The evidence gathered from the participants provided a rich set of insights. The use of colours connects the evidence to the respective participant groups, while the evidence itself was selected for discussion of the main topics shown in table 4. The two main topics discussed are "*Trust Develops Over Time*" and "*Trust as a Spectrum*".

5.2.1.1 Evidence of Trust in Artificial Intelligence.

Table 5:

Evidence of Trust in Artificial Intelligence

RQ1: Theme 1 – Trust in Artificial Intelligence
With a focus on the topics "*Trust Develops over Time*" and "*Trust as a Spectrum*"
Participant 1: *"Okay, so I think that you get two ends of the scale of trust. You get the misplaced trust, where people just blind the trust without knowledge, and then you get the other end of the scale where, or in the middle, you get organizations rightfully don't deploy because they don't trust necessarily the results."*
Participant 10: *"So there's degrees of trust, and there's degrees of comfort."*
Participant 11: *"I would trust it, but as long as the answer didn't get beyond what I'll call a complex. Horizon, you know, if the answers beyond the complexity horizon that I can't understand, then I start to become nervous that I'm making decisions that I don't know why, and that's, that's a concern."*
Participant 10: *"I think it's going to take time still, some time, to be completely trustworthy."*
Participant 11: *"Because we trust what we become familiar with, but to become familiar with it, we need to spend some time with it, get to know it."*
Participant 3: *"Although I describe a spectrum, I think we are very much at the bottom end, there is very little [trust] right, even for when we do use it, it's their caveats around."*
Participant 4: *"Trust is a function of credibility, reliability, intimacy, divided by self-orientation. What is it? Credibility, reliability, you know, it's, you know, there's, there's that there's a trust equation."*
Participant 9: *"I think it's going to be a slow, slow burn definitely."*
Participant 15: *"At this point, when you say trust, your trust can be defined in a couple of different ways."*
Participant 15: *"but leaving the decision, or completely basing your decision on AI, I think It's going to take a long time, in my opinion."*
Participant 19: *"So I think it's going to take, it's going to take a while to get there. I think maybe, depending, again, on the industry, I think in certain industries, they probably can get there much quicker."*

Note: Author's Own.

5.2.1.2 In-Case Analysis of the Evidence.

The Executive Decision-Making group of participants exhibited a similar pattern of thought that Trust in AI is not a single-value concept, but that it has an element of magnitude or breadth to it. Although all participants in this group articulated the concept in different words.

Participant 1 explained this as “trust being on a scale with two ends to it”. Using a different description, but with similarity of meaning, Participant 10 expressed different “degrees” of trust, while Participant 11 viewed trust as being appropriate only within a “complexity horizon”. Participant 10 provided further insight on an additional topic, that they saw Trust in AI requiring time to develop. Participant 11 provided similar insight that they understood trust as requiring familiarity, and that this takes time to achieve.

The Senior Decision-Making group also exhibited some similarity of insight, in which Participant 3 expressed trust as being a “spectrum” which we haven’t yet progressed from the bottom end of. Participant 4, on the other hand, explained trust as a function… mentioning a “trust equation”, with trust as the dependent variable and the independent variables of credibility, reliability, intimacy and self-orientation all contributing to a multi-variate trust construct. Lastly, Participant 9 again similarly referred to trust as requiring slow development, over time.

The experiences of the Operational Decision-Making group showed trust as being defined in different ways by Participant 15, but they did not elaborate further. The same participant also offered their views on the time dimension of trust, stating that to trust AI with a decision is still going to take a long time. This view of trust taking time to develop was supported by Participant 19, but with the difference that they saw certain industries being able to trust AI with decisions sooner than others.

5.2.1.3 Cross-Case Analysis of the Evidence.

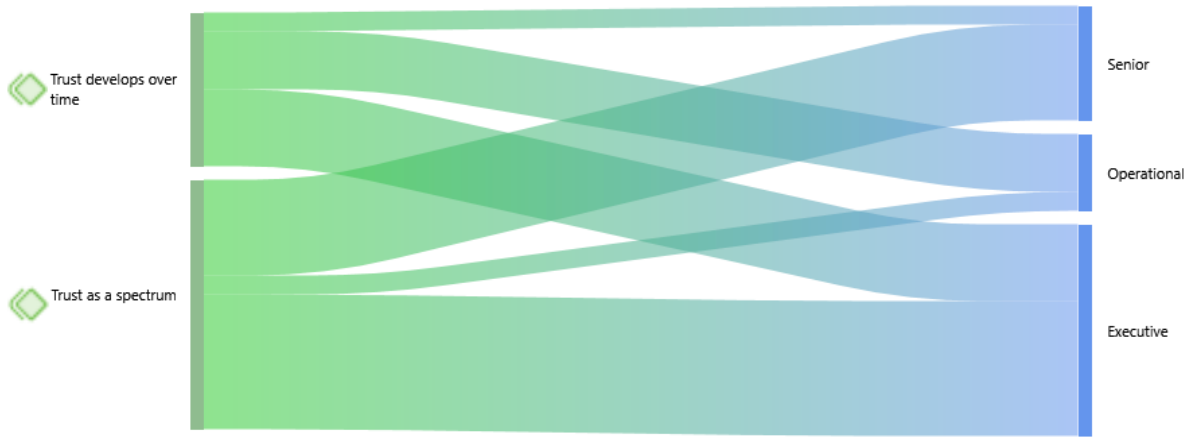
The Executive Decision-Making group’s experience, highlighted by Participant 1 with their view that trust is a scale with two ends to it, is similar to the Senior Decision-Making group who viewed trust as a “spectrum”, as illustrated by Participant 3. Although the Operational Decision-Making group described the concept slightly differently, they also identified trust as not being a single-value concept. The Executive Decision-Making group expressed trust as requiring time to develop through familiarity, while the Senior Decision-Making group, although not mentioning the need for familiarity, similarly said that trust requires time to develop. The Operational Decision-Making group however, through Participant 19, provided a unique insight that trust may possibly develop at different rates in different industries.

5.2.1.4 Distribution of Sub-Themes by Participant Group.

The “Sankey” diagram below provides a visual representation of the distribution and flow of the sub-themes across different participant groups. The diagram is prepared from the total Atlas.ti code set related to the two sub-themes in and provides a visual overview of the flow of discussion and density of quotations from the participant data.

Figure 13:

Sub-Theme Distribution: "Trust as a Spectrum" and "Trust Development Over Time"



Note: Author's own, generated from primary data in Atlas.ti software.

Trust develops over time: This theme has connections across all three groups, with the most significant flow towards the Executive Decision-Making group, followed by the Operational Decision-Making group, and the least to the Senior Decision-Making group. This potentially indicates that the perception of how trust evolves over time was most recognized among the Executive Decision-Making group and was not as prevalent with the Senior or Operational Decision-Making groups.

Trust as a spectrum: The Operational Decision-Making group showed the least engagement with this theme, indicating that while they may recognize the concept, it was not necessarily as central to their discussions as the idea of trust developing over time. Both the Executive and Senior Decision-Making groups demonstrated an awareness of the sub-theme, showing that both groups discussed or acknowledged Trust in AI as having various levels or forms. The Executive Decision-Making group discussed the sub-theme more than the Senior Decision-Making group.

5.2.1.5 Conclusion on Trust in Artificial Intelligence.

Across the participant groups, there was a similarity in the view that trust in AI is not a single-value concept but rather exists on a spectrum or scale. All the participant groups recognised this concept of a multi-faceted nature of trust. While they used slightly different terminology, the shared recognition of trust being more than just binary was clear. Additionally, all groups similarly echoed the idea that trust in AI develops over time, which underscored a consistent belief that trust is something that requires gradual building, rather than an being granted immediately.

One difference which emerged between the groups was how they described the nature of trust, with a variation in terms such as trust as a "spectrum", trust on a two-sided scale, or

trust as an equation. However, a second, more distinct difference emerged that trust in AI might develop at different rates across different industries. This insight into differing industry timelines is a unique observation, signalling a potential industry-dependent relationship between time and trust formation.

Key Concepts Identified: Trust takes time to develop. Trust exists as a spectrum or on a scale. Trust develops at different rates in different industries.

5.2.2 RQ1: Theme 2 – Trust in Artificial Intelligence for Adoption and Use

This theme is a main theme from the Chapter 2 literature review and is reflected in Figure 12 at the start of Chapter 5. The evidence gathered from the participants provides a rich set of insights. The use of colours connects the evidence to the respective participant groups, while the evidence itself is selected for discussion of the main topics shown in table 4. The two main topics discussed are "*Early Adoption and Experimentation*" and "*Change Management Engenders Trust*".

5.2.2.1 Evidence of Trust in Artificial Intelligence for Adoption and Use.

Table 6:

Evidence of Trust in Artificial Intelligence for Adoption and Use

RQ1: Theme 2 – Trust in Artificial Intelligence for Adoption and Use
With a focus on the topics *"Early Adoption and Experimentation"* and *"Change Management Engenders Trust".*
Participant 11: *"So I think what would probably want to do is a managed approach, where we start to use it and build trust in AI, and I think if we built the trust, I think it would, it could be useful."*
Participant 17: *"But if you walk the guys through it, if you get them involved, and maybe start them with, you know, broadly, what is the thinking around this thing, what it can't do, what it can't do, and where we think is what's possible."*
Participant 17: *"Therefore, should I work with it or should I undermine it if we don't get that change management journey."*
Participant 1: *"So I think you'll find that the early adopting organizations will put executives in charge who know the technology intimately, and who can kind of be the bulls#$tometer at a very, very high level in the organization."*
Participant 10: *"It was actually not something that many people knew about, but we were looking at some AI and that we, I think we had chat GPT two was out, if I'm not mistaken, and I remember pottering around a bit on it, feeding it some data sets and seeing what spewed out. I think we'd have to start in small, small control studies and really just test the viability of it."*
Participant 14: *"Yes, definitely. We are currently, you know, there's quite a lot of pilots used by white pilot where we are actually looking at going into more sort of complex decision making, you know, scenarios."*
Participant 14: *"You know, you have the early adopters, then you've got the laggards, and I think, trust will, you know, will grow."*
Participant 17: *"We've got quite a number of irons in the fire. Some of those pilots in the fire are in sort of, let's call it closed pilots. So we are test driving a number of initiatives."*

Participant 3: *"I was thinking before the call, has it got something to do with change management as well, this whole adoption of it? Because where are you in that cycle of change? I'm just thinking pragmatically, if it makes sense to use it, I'm happy to use it. But I'm not just going to throw my eggs into one basket and say, oh, look, something that would normally take six months, this program did in two days."*
Participant 12: *"You would need to find the right timing for that. I think it would be... So say, for example, you've been trying for a year to get people to use chatGPT. You don't throw it in right at the beginning, because, like, hey, this thing just popped up two weeks ago. And why am I the last one to be using it? You know what I mean?"*
Participant 2: *"So, you know, for the instance, the other day I tried to, I was putting together a brand, like a sales strategy for our emerging markets regions, and to draw, let's say, inspiration for some of the points to make sure I've covered everything, you know, I used chat GPT to basically, you know, draw me out a sales strategy for Zambia and Botswana, etc. It gave me extremely useful guiding points to be able to elaborate on and to think about."*
Participant 9: *"I think it would open the gate to allow us to do some, some sandpit style decision making. So if that trust was there to start with, that would open the door."*
Participant 9: *"So the early adopters, I think, would have a much lesser impact on the decision making than the later adopters, where that trust is built over time, that we do have some consistency and some evidence-based data of good decision making and appropriate and correct outcomes."*
Participant 9: *"You're going to have those sort of Trailblazer organizations that that jump straight in and they might have success in some spaces, but I think there will be those that that are in spaces that are not suited very well, right?*
Participant 19: *"So change management, I think it's going to be critical to help everyone along on this journey to really understand what AI is and how it can actually assist us from a business standpoint."*
Participant 19: *"And maybe these two things would have interlinked in terms of what I was saying from change management perspective, getting people buy into the journeys, etc, and then getting them to be able to trust the model outcomes."*
Participant 15: *"So we started sort of experimenting, should I say, a little bit with AI. Maybe when covid hit, and, and we've, we've done quite a bit now, quite a bit more now, where we just used a lot of AI to make things processes more efficient in terms of decision making."*
Participant 19: *"I mean, we've, I would say, fairly newly adopted it. I would say, still in the exploratory phase, we have been able to deliver a couple of use cases that currently is running in the environment."*
Participant 19: *"But I think we really are at the point where many organizations are now starting to explore AI."*

Note: Author's Own.

5.2.2.2 In-Case Analysis of the Evidence.

Within the Executive Decision-Making group, Participant 1 suggested that early adopters place executives who know the technology intimately, in charge, to critically evaluate AI. Participant 10 recounted small-scale experiments with AI tools like ChatGPT, emphasizing the importance of small control studies, while Participant 11 highlighted a managed, step-by-step approach to building trust in AI. Participant 14 mentioned ongoing pilot projects aimed at integrating AI into complex decision-making scenarios, noting the progression differences of trust between early adopters and "laggards". Lastly, Participant 17 emphasized involving individuals in a change management process to build trust and understanding of AI's

capabilities and limitations. This group showed a similar approach to gradual AI adoption, with a focus on building trust through familiarity and structured experimentation to minimise risk.

In the Senior Decision-Making group, Participant 2 described using AI to support the creation of sales strategies, highlighting the role of AI as a tool for inspiration and efficiency, while Participant 3 questioned the role of change management in the AI adoption process. Participant 3 also reflected on how to align AI adoption with their current change cycle. Participant 9 emphasized "sandpit-style" decision-making and noted that early adopters would likely see less immediate impact compared to later adopters who benefit from established trust and evidence-based outcomes. Similarly, Participant 12 also discussed the timing of AI introduction, suggesting that AI should be implemented strategically once the workforce is ready. The senior decision-making group focused on gradual adoption, in line with broader strategic goals and ensuring that the timing of its introduction aligns with the organisational change cycle.

In the Operational Decision-Making group, Participant 15 described experimenting with AI technologies since the onset of COVID-19, focusing on enhancing process efficiency and decision-making. Participant 19 similarly echoed the importance of change management, stating that it is critical for ensuring understanding and buy-in during AI implementation. They also noted that their organization was in an exploratory phase, with multiple use cases currently being tested. This observation also indicated a gradual, measured adoption. The overall emphasis within this group was on integrating AI gradually, ensuring the operational workforce is prepared and engaged in the transition.

5.2.2.3 Cross-Case Analysis of the Evidence.

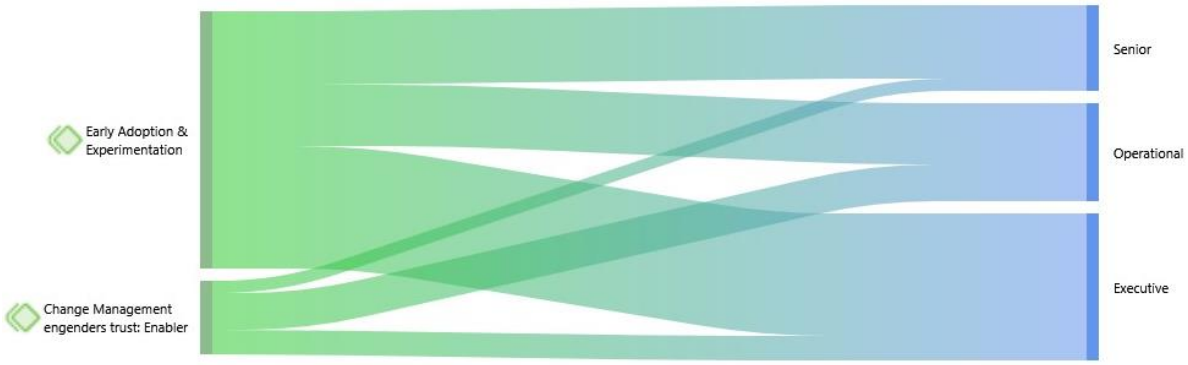
Comparing the groups revealed similarities and differences in their approach to AI adoption and change management. The Executive Decision-Making group emphasized structured and small-scale experimentation to build familiarity and trust gradually, while the Operational Decision-Making group focused on experimentation and the importance of change management to ensure buy-in and understanding during implementation. The Senior Decision-Making group similarly acknowledged the need for change management but was more focused on aligning AI implementation with the change cycle to maximize acceptance. All groups recognized the importance of a gradual, change-managed approach to AI adoption.

5.2.2.4 Distribution of Sub-Themes by Participant Group.

The “Sankey” diagram below provides a visual representation of the distribution and flow of the sub-themes across different participant groups. The diagram is prepared from the total Atlas.ti code set related to the two sub-themes and provides a visual overview of the flow of discussion and density of quotations from the participant data.

Figure 14:

Sub-Theme Distribution: "Early Adoption & Experimentation" and "Change Management".



Note: Author's own, generated from primary data in Atlas.ti software.

Early Adoption & Experimentation: This theme is most prominently associated with the Executive Decision-Making group, as indicated by the widest flow towards them. It suggests that the Executive Decision-Making group was engaged in discussions around early adoption and experimentation. The flow towards the Senior Decision-Making group is slightly narrower, indicating that while this group also discussed early adoption and experimentation, it was not as central as for the Executive Decision-Making group. The smallest flow is directed towards the Operational Decision-Making group, implying that operational management did not focus as much on this aspect compared to the other groups.

Change Management Engenders Trust: This theme shows a relatively balanced flow between the Operational and Executive Decision-Making groups, indicating that both levels engaged in discussions around how change management serves as an enabler of Trust in AI. The flow from these groups suggests that this theme is recognized and valued in their roles. In contrast, the Senior Decision-Making group had the smallest flow, potentially indicating that while they acknowledged that change management is a trust enabler, it was less central in their interview responses compared to the other two groups.

5.2.2.5 Conclusion on Trust in Artificial Intelligence for Adoption and Use.

All groups similarly recognized the importance of gradual AI adoption, experimentation, and change management. There was emphasis on leading structured and controlled pilot projects through experimentation to build trust, while there was additional focus on ensuring workforce readiness and buy-in through change management. Both would gradually increase familiarity and trust in AI systems, allowing organizations to evaluate and integrate AI into their operations in small steps.

AI adoption through change management was also seen to play an important role in the adoption and use of AI across different organizational levels. It was consistently emphasized as important for building trust, while aligning with the current change cycle and

securing buy-in during the implementation of AI technologies. A difference across the groups lay in their focus, ranging from oversight and risk management to timing and strategic integration, and finally to practical implementation and efficiency gains. In summary, the findings underlined that AI adoption requires a balanced combination of gradual adoption, structured experimentation, and change management.

Key Concepts Identified: The importance of gradual AI adoption, experimentation and change management for trust in AI.

5.2.3 RQ1: Theme 3 – Trust in Artificial Intelligence for Decision-Making

This theme is a main theme from the Chapter 2 literature review and is reflected in Figure 12 at the start of Chapter 5. The evidence gathered from the participants provides a rich set of insights. The use of colours connects the evidence to the respective participant groups, while the evidence itself is selected for discussion of the main topics shown in table 4. The two main topics discussed are "*Output Reliability and Trust*" and "*Only to Inform Decisions*".

5.2.3.1 Evidence of Trust in Artificial Intelligence for Decision-Making.

Table 7:

Evidence of Trust in Artificial Intelligence for Decision-Making

RQ1: Theme 3 – Trust in Artificial Intelligence for Decision-Making
With a focus on the topics *"Output Reliability and Trust"* and *"Only to Inform Decisions".*
Participant 11: *"But I don't think I would have a problem using AI to make more informed decisions."*
Participant 13: *"I would use it to help inform, okay, so, like, maybe a data analytics tool or a way to, you know, the way, say, like, a second set of eyes, or, you Know, bounce ideas, leverage the processing capability of something that has an artificial intelligence to, like, inform scenarios and things like that."*
Participant 14: *"So that's why I'm saying it's in terms of, you know, those kinds of functions, and then where we use it, in terms of decision-making algorithms is sort of more in the clinical space. So if we are looking at, you know, treatment plans that have been submitted to us for pre authorization. So, you know, we use AI to basically, instead of humans, you know, to basically look at the treatment plan based on our, you know, guidelines and protocols, funding guidelines, and, you know, it can make a funding decision."*
Participant 11: *"I'd always want the backstop of a wizened practitioner to look at the outcomes and double check whether they're reasonable."*
Participant 13: *"To encourage more trust in it, well, there would need to be a reliability. You'd want to see, like a consistency of that, an accuracy."*
Participant 14: *"So basically, it's basically to trust whether you know the outputs [are] accurate and reliable, you know, obviously you want, I mean, there must be consistency in the decision making."*
Participant 3: *"I just see it as just one tool in an arsenal of products you would use to inform a decision. I don't see it as being superior to any of those on the toolbox. I just see it as another one."*

Participant 4: *"It can sort inform me day to day, if I if I can look at an app and say, Hey, this is where your team engagement is today. And you know, this person's a bit low. I mean, I think, I think we're going to get to a model which is not far from that."*
Participant 12: *"No, yes to inform, but not to make decisions. Okay, so interesting. So to use, so using it to gather information, not gather but, but to maybe, yeah, actually gather information and data, which can help to inform decisions, but not make decisions."*
Participant 3: *"I've found it useful at a high level, but you still have to go through and check. It's not foolproof by any stretch of the imagination..."*
Participant 4: *"I kind of look at it and I see that that's probably as good as what we could have done. And as a professional, I can judge the effectiveness of that of course, you know?"*
Participant 16: *"You very quickly get to a stage, [...] where you have to trust the model, you know, you can do some fundamental checks. And, at some stage, I almost felt like I'm rechecking, like people that just adopt Excel the first time, we'd recalculate everything manually again just to check if Excel is right, you know?"*
Participant 6: *"But ultimately, all the decision making and all the quality control around decisions is made by people, supported by data that's been screened and also verified [by] people that understand the tools or the models that have been built."*
Participant 7: *"So I think our organization will use it to inform decisions, maybe data driven decisions, but certainly not to make them."*
Participant 15: *"Yeah, to inform your decisions, not to make your decisions. And maybe it's because I am old fashioned. Maybe it's because of the generation that you come from as well, where we did things, you know, you sat and you really filtered through information, and you went through presentation fatigue, and everybody presented, from finance to HR to risk and compliance, everything that can go wrong with this decision before you made the decision right."*
Participant 6: *"It has to be, you've narrowed it down, what does it mean? And then have a human check it."*
Participant 7: *"So for us, it, I think trust would be reliability of information, so not missing the things that we're looking for and not giving wrong sort of advice."*
Participant 19: *"So I think from an end user perspective, it would be, you know, having the ability to really trust and rely on the outcomes that AI is giving you."*

Note: Author's Own.

5.2.3.2 In-Case Analysis of the Evidence.

Within the Executive Decision-Making group, Participant 11 expressed openness to using AI for informed decision-making but emphasized the need for human oversight to verify outcomes. Participant 13 viewed AI as a supplemental tool, comparing it to a "second set of eyes" to leverage data and inform scenarios. Similarly, Participant 14 described the use of AI in clinical and funding decision contexts, where AI outputs are trusted when used against established guidelines and protocols. The executives in this group had a similar approach of using AI as a supportive tool, requiring human verification to build trust.

Within the Senior Decision-Making group, Participant 3 saw AI as just one tool among many, emphasizing that while it is useful, it is not foolproof and requires human checking. Participant 4 acknowledged the potential of AI to inform daily activities, such as team engagement, but stressed the need for professional judgment to validate AI's effectiveness. Participant 12 expressed a clear preference for AI to gather information and inform decisions

but not to make them autonomously. Participant 16 drew a parallel with initial Microsoft Excel use, illustrating the hesitance in trusting AI completely without double-checking outputs manually. The Senior Decision-Making group similarly focused on using AI as an informative tool while maintaining human oversight for validation.

In the Operational Decision-Making group, Participant 6 highlighted that AI provides data support, but human checking is essential for final quality control and decision validation. Participant 7 noted that while their organization might use AI to guide decisions, humans still make the final call. Similarly, Participant 15 emphasized that final decision-making should remain a human task. In this group, the operational focus remained on ensuring the reliability and accuracy of AI outputs, supported by human expertise.

5.2.3.3 Cross-Case Analysis of the Evidence.

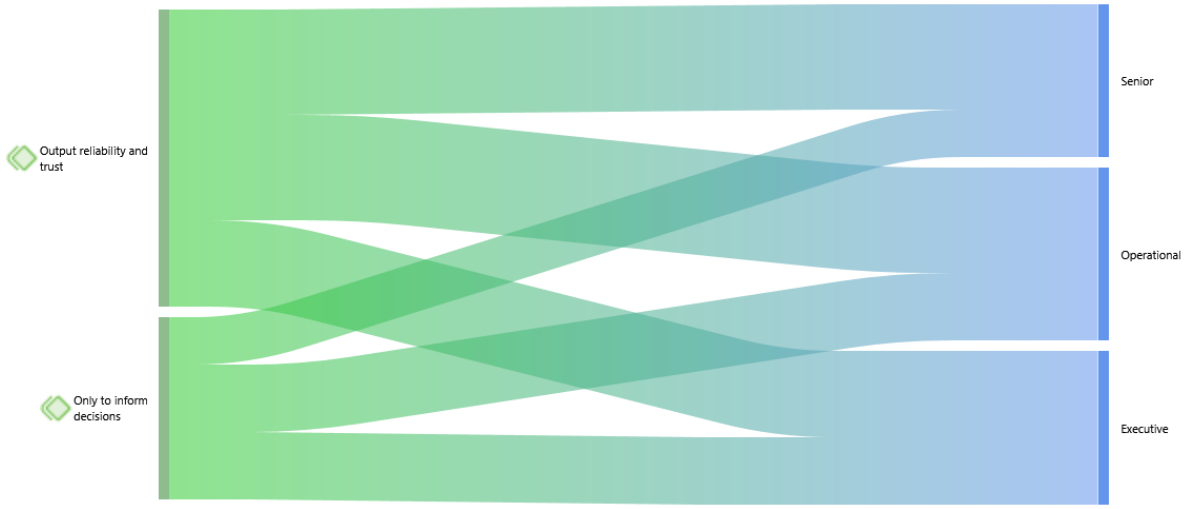
Comparing the insights across groups revealed similar themes in the use of AI primarily to inform and supplement decision-making, rather than to replace human judgment. Both Executive and Senior Decision-Making groups emphasized the importance of human verification or checking, to ensure trust in AI outputs. Executives focused on using AI as a supplemental tool that enhances existing decision-making processes, while the Senior Decision-Making group highlighted the importance of integrating professional judgement to assess AI's effectiveness. The Operational Decision-Making group similarly underscored the importance of human intervention for quality control and decision validation, stressing the need for reliable information from AI systems. All groups recognized the role of AI as a valuable but supplemental tool, with trust built through human oversight.

5.2.3.4 Distribution of Sub-Themes by Participant Group.

The "Sankey" diagram below provides a visual representation of the distribution and flow of the sub-themes across different participant groups. The diagram is prepared from the total Atlas.ti code set related to the two sub-themes and provides a visual overview of the flow of discussion and density of quotations from the participant data.

Figure 15:

Sub-Theme Distribution: "Only to Inform Decisions" and "Output Reliability and Trust"



Note: Author's own, generated from primary data in Atlas.ti software.

Output Reliability and Trust: This theme featured relatively evenly between all groups, indicating that the selected decision-making groups emphasized the concept of output reliability, and its association with trust, evenly regardless of their seniority. There was slightly less focus in the Executive Decision-Making group, which might possibly reflect less focus at this management level for actual verification of outputs, a task usually performed by the management tiers below.

Only to Inform Decisions: This theme was most mentioned in the Executive and Operational Decision-Making groups, and they discussed the concept of using AI outputs solely to inform decision-making evenly between their groups. The Senior Decision-Making group had a more moderate focus on the sub-theme during the interviews, indicating that this concept was less central in discussions with them.

5.2.3.5 Conclusion on Trust in Artificial Intelligence for Decision-Making.

All groups similarly viewed AI as a supportive tool to inform and enhance decision-making, rather than a replacement for human judgment. They all emphasized the necessity for human verification, recognizing that trust in AI is contingent on its ability to provide consistent, accurate, and reliable outputs. There was a collective understanding that AI's role is valuable in contributing to the decision-making process, but it must be sense-checked by human oversight to ensure decisions align with organizational standards and expectations.

While the groups shared this overall perspective, there was a difference in how they framed the relationship between AI and human oversight. Some focused on the importance of verifying AI outputs against established protocols. Others placed more emphasis on balancing AI with other tools, using professional judgment to assess its utility in broader decision-making

contexts. Additionally, there was a focus on ensuring the day-to-day accuracy and quality of AI outputs, with human intervention playing a key role in final control. These nuanced differences reflect varying areas of focus, from strategic oversight to balancing AI with existing tools, and ensuring accuracy.

Key Concepts Identified: AI should inform and enhance human decision-making, not replace it. Human verification of AI outputs is needed.

5.2.4 RQ1: Theme 4 – "Blind-Trust" Tensions

This theme is a potential new main theme and a distinct difference (on first analysis) from the Chapter 2 literature review and is reflected in Figure 12 at the start of Chapter 5. The evidence gathered from the participants provides a rich set of insights. The use of colours connects the evidence to the respective participant groups, while the evidence itself is selected for discussion of the main topics shown in table 4. The main topics discussed are *"Blind-Trust "Yes""* and *"Blind-Trust "No""*.

5.2.4.1 Evidence of "Blind-Trust" Tensions.

Table 8:

Evidence of "Blind-Trust" Tensions

RQ1: Theme 4 – "Blind-Trust" Tensions
With a focus on the topics *"Blind-Trust "Yes""* and *"Blind-Trust "No"".*
Participant 1: *"Fire and forget and put information in, and you get out processed information which probably doesn't, doesn't engender a lot of trust."*
Participant 1: *"I think with AI, the people who kind of go in blindly and think that it's a panacea are going to end up getting burnt. And it happens. Like absolutely, you know, it happens. It's happened in our organization, where people have had, they've generated client communications using open AI and or [Internal AI Tool X], and they haven't read and checked them, you know?"*
Participant 10: *"Because they just kept following the formula without considering the external sources. Now I remember going deep into it, and that's the danger with blindly following something. Would we get to a point where we blindly trust it so much that we blindly follow it and it could be used against us?"*
Participant 1: *"I think that if you don't know how it works, and if you all you see is a black box, you are not well served by blindly trusting it."*
Participant 17: *"You've got a population with structurally low literacy levels and even lower sort of deep technical understanding of technology. We've got a high technology adoption rate, but many people use it blindly."*
Participant 4: *"Um, that's an interesting one, because probably, I trust it too much. I mean, I don't know, like, I'm not, I'm not asking for a lot of stuff ahead of making a decision."*
Participant 4: *"Okay, I like that. Maybe I'll use that or not so, so I'm probably not asking a lot of questions, as much as, as much as maybe we should."*

Participant 16: *"I mean, most people probably don't understand exactly the mathematical formulas and how, you know? I don't know some Einstein is way beyond me. I sometimes I've some idea what he says, but even my lecturers at varsity said, sometimes I understand, and sometimes they don't, you know, but you somehow trust the work that was done, and the work that was done is validated by experts in the field and so forth."*
Participant 16: *"So let's go there for a second. If there's no way to validate AI, and we just have to continue with this kind of trust, just trust that it works, you know, I don't think AI is going to be adopted in large scale."*
Participant 2: *"If there was a trust, let's say a blind trust, like we trust Excel to calculate our sheets for us correctly. Or we trust an email that's being sent, that it would be sent. Or we trust but you know, even like Power BI reports to pull data, you know, like, effectively, I guess that's what trust in AI would be, you know, like it, you wouldn't question it. You would use it as a tool [...], you wouldn't have to go back and think over or recheck or sense check."*
Participant 4: *"I think we take a face validity, and then if it looks reasonable, it's okay. Well, I don't need to be questioning it, probably, if it didn't make anything, you know?"*
Participant 9: *So the more you aim to control, the more trust you can put into it, because you have that control. The less control and boundary you apply, the longer it'll take to build the trust."*
Participant 12: *"I don't need to know how it works. I trust [it], because, humans have created so many, aircrafts, everything, you know? I trust that how this tool was created, the most brilliant minds have come together to create these things. I'm more interested in, you know, what can I get out of it? How is it going to impact me? How's it going to improve my life? How's it going to improve my day to day?"*
Participant 12: *And so far, I haven't lost it. Hasn't given me a reason not to trust it. But it's because I've also had to learn how to use it."*
Participant 16: *"AI and Generative AI, there is a necessity to trust the model that is built by the machine learning most blindly, and I think it's an inherent requirement."*
Participant 16: *"AI requires you to believe, you know, machine learning, anything like recognizing patterns beyond the human conception level."*

Note: Author's Own.

5.2.4.2 In-Case Analysis of the Evidence.

In the Executive Decision-Making group, Participant 1 highlighted the risk associated with blindly trusting AI. They referenced instances within their organization where such blind trust led to client communications not being verified, resulting in negative consequences. Participant 10 emphasized the danger of following AI outputs without critical evaluation, stressing the potential for AI being used against the organisation if it is blindly trusted. Participant 17 further underscored the risk of blind adoption in populations with limited technical literacy, which highlighted that while technology adoption rates are high, the understanding of how it works is often low. This group similarly advocated for a critical approach to AI, emphasizing the importance of understanding how AI works and generates outputs, to avoid the pitfalls of blind trust.

The Senior Decision-Making group, however, presented a mixed perspective. Participant 4 admitted to sometimes trusting AI outputs too much, acknowledging the need for

more questioning before relying on AI-driven decisions. Participant 16 discussed the difficulty of understanding the mathematical formulae underlying AI. They also illustrated that while blind trust might be necessary at times, there will be resistance to adoption if there is no means for validation. Participant 2 expressed a pragmatic view, equating Trust in AI with trust in other familiar tools (e.g., Excel or Power-BI), suggesting a degree of trust based on the assumption that experts have validated these systems. However, a different perspective was given by Participant 12, who highlighted the need for understanding AI's functionality and its impact on day-to-day activities. The Senior Decision-Making group thus revealed a tension between caution and convenience, with some acknowledging the inherent risks of blind-trust and others leaning towards accepting AI based on precedent and familiarity.

Notably, there was **no evidence** from the Operational Decision-Making group regarding their stance on blind-trust in AI. This absence was significant as it indicated that, at the operational level, the discussions or concerns surrounding blind-trust in AI may not be as prevalent or prioritized as at the Executive and Senior Decision-Making levels, where strategic decision-making and risk management are more central.

5.2.4.3 Cross-Case Analysis of the Evidence.

Across the groups, there was a clear difference in their view towards blind-trust in AI. The Executive Decision-Making group were unanimous in their rejection of blind-trust, emphasizing the importance of understanding AI mechanisms, verifying outputs, and avoiding uncritical reliance on AI technologies. Their collective stance was significant, suggesting that at the highest decision-making level, there was a strategic emphasis on maintaining control and ensuring oversight when adopting AI technologies. In contrast, the Senior Decision-Making group presented a more varied perspective. While some senior members acknowledged the risks of blind-trust, others leaned towards a more pragmatic approach. They expressed a degree of trust in AI, based on its perceived reliability and the expertise behind its development.

This difference reflected a tension between scepticism and convenience at the Senior Decision-Making level, which potentially reveals that there may be a need to balance trust in AI with the need for efficiency and practicality. The absence of evidence from the Operational Decision-Making group highlighted a gap in the discourse at this level, which could indicate that discussions around the risks of blind-trust in AI are not as prioritized in their roles. This lack of focus may be due to the nature of their work, which might be centered more on implementation rather than strategic oversight and risk evaluation.

5.2.4.4 Distribution of Sub-Themes by Participant Group.

The “Sankey” diagram below provides a visual representation of the distribution and flow of the sub-themes across different participant groups. The diagram is prepared from the total Atlas.ti code set related to the two sub-themes and provides a visual overview of the flow of discussion and density of quotations from the participant data.

Figure 16:

Sub-Theme Distribution: “Blind-Trust “Yes”” and “Blind-Trust “No””

Note: Author’s own, generated from primary data in Atlas.ti software.

Blind-Trust “Yes”: The discussion with the Senior Decision-Making participant group showed more inclination towards blind-trust than the Executive Decision-Making group. However, an almost-equal amount of the interview discourse suggested that the Senior Decision-Making group would also not favour “blind-trust”. This created both an in-case and cross-case tension between these management levels.

Blind Trust “No”: The Executive Decision-Making group unanimously rejected “blind-trust”, indicating that the discussion with them in the interviews showed no support for it. This might potentially imply that the Executive Decision-Making group preferred a more cautious or evidence-based approach, actively avoiding situations where trust is given without question.

Note: There was a notable absence of mention for either support or rejection of “blind-trust” in the Operational Decision-Making group interviews, with the broader topic of “blind-trust” not being discussed.

5.2.4.5 Conclusion on “Blind-Trust” Tensions.

There was some similarity in evidence between the participant groups, with some being opposed to “blind-trust”, emphasizing the need for verified and controlled use of AI technologies. Others showed more risk tolerance, expressing caution, and some supported the convenience of “blind-trust” based on perceived reliability and expert verification. The

similarity was therefore a shared recognition of the risks associated with “blind-trust” in AI, with participants acknowledging the potential pitfalls of reliance on AI technologies.

In terms of differences, there was a clear tension regarding "blindly-trusting" AI. Some participants responsible for decision-making rejected blind trust altogether, reflecting a strategic, risk-averse stance, while others demonstrated more varied perspectives, with some showing a higher tolerance for trusting AI without full verification. Interestingly, the Operational Decision-Making group did not raise the topic of “blind-trust”, highlighting that there appears to be a potential gap in awareness or prioritization of this issue at an operational level.

Key Concepts Identified: There was a tension between the decision-making groups regarding “blindly-trusting” AI. Some supported it, while others rejected it.

5.2.5 RQ1: Theme 5 – “Blast-Radius”

This theme is a potential new main theme and a distinct difference (on first analysis) from the Chapter 2 literature review and is reflected in Figure 12 at the start of Chapter 5. The evidence gathered from the participants provides a rich set of insights. The use of colours connects the evidence to the respective participant groups, while the evidence itself is selected for discussion of the main topics shown in table 4. The main topic discussed is “*Blast-Radius*”.

5.2.5.1 Evidence of “Blast-Radius”.

Table 9:

*Evidence of “*B*last-*R*adius”*

RQ1: Theme 5 – “Blast-Radius”
Participant 1: *"You get organizations rightfully don't deploy because they don't trust necessarily the results, because they don't know how it's derived, and they don't know how to set guardrails on the technology."*
Participant 14: *"Again, we've kind of set boundaries, yes, it's stuff that would really go through without...[pause] You know, it's very easy to make it to say, yes, pay, no, don't pay. But you know, the moment you have more complex decision making that says yes, pay, but only if you know, then it starts to get a little bit more complex, you know. So we've started off with a very simple, you know, literally, a machine that can just run with a few rules here and there."*
Participant 14: *"We can, you know, kind of give me explanations as to the pros and cons and why this route... you know, that the final recommendation is really in my best interest, before I can make that decision."*
Participant 17: *"And we are slowly converting them into leverage, your sort of buzzwords, general purpose, sort of AI's that we've got available because we're shifting the architecture to start leveraging common frameworks, common, sort of guard-rails, common sorts of styles of engagement."*
Participant 17: *"But again, a phased roll out to manage the blast radius. So start internally curated audience."*

Participant 17: *"I'm still leaning towards this thing must be constrained to, like, effectively, an employee number. It must have limited sort of mandates. Uh, yes, it could be a super employee type mandate, but it must be limited. And I think it's almost just to curb the runaway effect."*
Participant 17: *"So we must come up with ways also to, like, just also create safety nets in the hopefully miniscule likelihood events, yeah, that these things are compromised, or there's some form of runaway effect you want to I keep on using this term managed blast radius, so it keeps that thing accountable."*
Participant 17: *"So that helped us in that sort of adoption, in terms of implementation, that trust we always had starting out with use cases where we can manage the blast radius, minimize the negative impact."*
Participant 2: *"What steps could it take to rectify itself, or, you know, rectify the strategy or the move?"*
Participant 2: *"If I see that it's not working, how quickly could I undo what it's done? Or, you know, like, once I've implemented it? Can I now circumvent it and carry on without it there?*
Participant 2: *"I would also then understand where I could haul in safety mechanisms to ensure that whatever it does beneath that, there's some sort of, what's the word backstop, should something go wrong."*
Participant 3: *"So when they make that decision, where in this model are they playing, because that would impact your risk tolerance, like where you are on your scale of risk, if you did some kind of a two by two, yeah, you want to plot it. It's like any strategy outline you would have to look at the different scenarios that you've looked at. You wouldn't put all your eggs on the one basket. You'd have to trade off risks and benefits."*
Participant 9: *"However, if I was to sort of surmise what that would look like I mentioned earlier, we would certainly sandpit that. Which, by nature, puts a boundary around the impact that that could potentially have."*
Participant 6: *"So a lot of the big organizations are actually starting to bring in an AI code of conduct. Right? Those code and conduct terms are around what AI could be used for. It's around the limitations of AI. It's around the limitations of what data is shared with others who controls the tools."*
Participant 6: *"I think there's going to be a lot of mistrust, a lot of mistrust, and I think what they will do is trust in AI is going to be incrementally increased in our day-to-day personal lives."*
Participant 15: *"So you would probably test it in a safe environment where there isn't going to be too much impact, or you would test its decision-making ability without allowing you to make the decision and having a human filter to before the decision is processed to say yay or nay."*

Note: Author's Own.

5.2.5.2 In-Case Analysis of the Evidence.

Within the Executive Decision-Making group, there was a strong emphasis on controlling the potential negative impact of AI by establishing guardrails and phased rollouts. Participant 1 highlighted the need for setting boundaries or guardrails on AI to mitigate risks, explaining why some organizations are hesitant to deploy it without such safeguards. Participant 14 discussed how they simplified initial AI deployments to minimize complexity, to ensure control over decision outcomes. Participant 17 explicitly used the term "blast radius" to describe their phased rollout strategy, indicating an intentional effort to limit AI's impact to an internally controlled environment before broader implementation. The same participant also emphasized the importance of creating safety nets and setting constraints on AI capabilities,

such as limiting its decision-making authority or mandate. The Executive Decision-Making group demonstrated a similar approach to risk mitigation, focusing on managing potential fallout by setting boundaries and creating controlled environments.

The Senior Decision-Making group also emphasized damage control and containment strategies. Participant 2 discussed the importance of implementing safety mechanisms and ensuring the ability to quickly reverse or adjust AI's actions if things go wrong. They also stressed the need for "backstops" to mitigate potential fallout. Participant 3 talked about assessing risks using scenario planning to weigh trade-offs, suggesting a strategic approach to managing risk exposure. Participant 9 mentioned using "sandpits" or controlled environments as a way to contain the impact of AI testing, similar to the phased approach described by the Executive Decision-Making group. The Senior Decision-Making group displayed a focus on creating environments and processes that minimize AI's negative impact, ensuring that any fallout remains contained.

The Operational Decision-Making group also acknowledged the importance of managing AI's impact but focused more on practical limitations and incremental trust-building. Participant 6 highlighted how organizations were implementing codes of conduct to define AI's boundaries, thus ensuring that usage was controlled and monitored. They also emphasized that trust in AI would build slowly as people grow more comfortable with AI in their day-to-day lives. Participant 15 reinforced the idea of testing AI in safe, low-impact settings, using human oversight to validate AI's outputs, before allowing it to make autonomous decisions. This group showed a similar understanding of containing AI's impact by establishing limitations, rules, and safe testing environments.

5.2.5.3 Cross-Case Analysis of the Evidence.

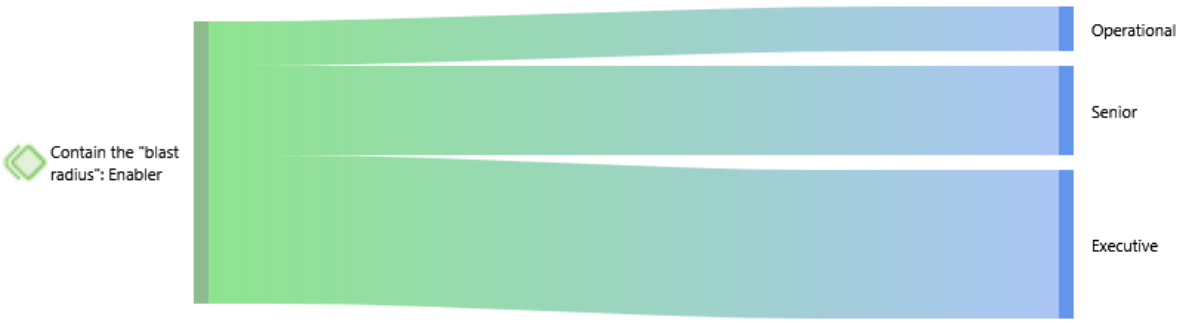
Across the groups, there was a similar emphasis on minimizing AI's potential negative impact, but the approach and focus varied slightly between them. The Executive Decision-Making group prioritized structured containment strategies such as phased rollouts, setting boundaries, and building safety nets to manage the "blast-radius" explicitly. Their focus was on deploying AI in ways that ensure any fallout remained manageable and predictable. The Senior Decision-Making group also aligned with the executives, using methods such as safety mechanisms and "sandpit" testing environments to manage and contain AI's impact. They emphasized risk management by suggesting scenario planning to understand potential outcomes and trade-offs. The Operational Decision-Making group shared a similar focus on containment, but emphasized establishing practical limitations, such as codes of conduct and low-impact testing in controlled environments.

5.2.5.4 Distribution of the Main Theme by Participant Group.

The “Sankey” diagram below provides a visual representation of the distribution and flow of the sub-themes across different participant groups. The diagram is prepared from the total Atlas.ti code set related to the two sub-themes and provides a visual overview of the flow of discussion and density of quotations from the participant data.

Figure 17:

Sub-Theme Distribution of the “Blast-Radius” Theme.



Note: Author’s own, generated from primary data in Atlas.ti software.

“Blast Radius”: All of the participant groups discussed the containment of risk or negative effects of using AI for decision-making. However, of interest from the above diagram is that the interview data collected was more prevalent in the Executive Decision-Making group, followed by the Senior Decision-Making group and lastly the Operational Decision-Making group. This potentially demonstrates a heightened awareness of the need for risk-aversion at the more senior levels in the organisation.

5.2.5.5 Conclusion on “Blast-Radius”.

Across all participant groups, there was the similarity of a consistent emphasis on managing the potential negative impact, or “blast-radius,” of AI in decision-making. Each group recognized the importance of containment strategies to control the risks associated with AI use. All similarly advocated for methods to ensure that AI’s impact remains predictable and manageable, whether through phased rollouts, safety mechanisms, or controlled environments. Their shared concern highlighted the similarity of focus on mitigating risks across the organization, demonstrating a collective view of ensuring that AI adoption does not lead to unintended consequences.

There was a difference in how containment strategies were approached. Some prioritized a structured, proactive approach to risk control, using phased rollouts and safety nets to manage potential fallout. Others emphasized scenario planning and trade-offs, adapting to management of different risk levels. Another approach focused on practical, incremental trust-building through gradual testing and codes of conduct, reflecting a step-by-

step emphasis on trust and practical limitations. These differences highlighted varying priorities in how containment of AI's "Blast-Radius" was managed.

Key Concepts Identified: The importance of containing the potential negative outcomes of AI use in decision-making, to engender trust.

5.2.6 Summary of Similarities and Differences in the Findings for RQ1

Table 10 below is provided as a summary and convenient reference for the findings related to the sub-themes discussed under the main themes of RQ1. The table was prepared from the in-case and cross-case analyses and conclusions for each of the five themes presented in Section 5.2.

Table 10:
Summary of Similarities and Differences in the Findings, Across Themes, for RQ1.

RQ1: *How does trust in AI lead to its adoption and use for organisational decision-making?*		
	Similarities	**Differences**
Theme 1: Trust in Artificial Intelligence	• Trust takes time to develop. • Trust exists as a spectrum or on a scale.	• Trust develops at different rates in different industries • The “spectrum” of trust was described in varying ways between participant groups.
Theme 2: Trust in Artificial Intelligence for Adoption and Use	• Importance of gradual AI adoption, experimentation and change management.	• Focus varied across groups, from oversight and risk management to timing and strategic integration, to practical implementation and efficiency gains.
Theme 3: Trust in Artificial Intelligence for Decision-Making	• AI should inform and enhance human decision-making, not replace it. • Human verification of AI outputs is needed.	• Difference in framing the relationship between AI and human oversight. • Varying areas of focus, including strategic oversight, balancing AI with other tools, and day-to-day accuracy and quality of output.
Theme 4: "Blind-Trust" Tensions	• Shared recognition of the risks associated with “blind-trust” of AI.	• There is a tension between the Decision-Making groups regarding “blindly-trusting” AI. Some support it while others reject it. • Operational Decision-Making group did not raise the topic of “blind-trust”, highlighting a potential gap in awareness.
Theme 5: "Blast-Radius"	• Containing the potential negative outcomes of AI use for decision-making.	• Varying approaches were used in describing how containment of AI's "Blast-Radius" was managed. • Phased roll-outs with safety nets vs. adaptation to risk through scenario planning, vs. practical limitations and controlled testing environments.

Note: Author’s Own.

5.2.7 Conclusion to the Findings for RQ1: Summary of Key Concepts

In conclusion to the findings for RQ1, the similarities and differences from each of the discussed themes and their conclusions were reviewed, to derive the “Key Concepts” which emerged from the theme, topic and sub-theme findings. These key concepts were derived from both the similarities and differences which emerged and are presented in Table 11.

Table 11:

Summary of Key Concepts from the Findings, by Theme for RQ1.

RQ1: *How does trust in AI lead to its adoption and use for organisational decision-making?*		
	Topics / Sub-Themes	**Key Concepts Identified**
Theme 1: Trust in Artificial Intelligence	• *Trust Develops over Time* • *Trust as a Spectrum*	• Trust takes time to develop. • Trust exists as a spectrum or on a scale. • Trust develops at different rates in different industries.
Theme 2: Trust in Artificial Intelligence for Adoption and Use	• *Early Adoption & Experimentation* • *Change Management Engenders Trust*	• Importance of gradual AI adoption and experimentation • Importance of change management for trust in AI.
Theme 3: Trust in Artificial Intelligence for Decision-Making	• *Only to Inform Decisions* • *Output Reliability & Trust*	• AI should inform and enhance human decision-making, not replace it. • Human verification of AI outputs is needed.
Theme 4: "Blind-Trust" Tensions	• *"Blind-Trust" Tensions*	• There is a tension between the Decision-Making groups regarding "blindly-trusting" AI. Some support it while others reject it.
Theme 5: "Blast-Radius"	• *"Blast-Radius"*	• Importance of containing the potential negative outcomes of AI use for decision-making, to engender trust.

Note: Author's own.

5.3 Findings for Research Question 2: *How does xAI engender trust in AI for decision-making in organisations? And: What are the factors that contribute to trust in AI?*

The themes related to the research question will be discussed in this section. As discussed in Chapter 4, differences that arose from the analysis identified some distinct differences and some nuances of difference. The former are shown, in Table 12, as potential new themes and the latter as potential new sub-themes. For RQ2, four (4) main themes emerged from the data analysis, and one (1) of these is potentially new. A further eight (8) potential new sub-themes also emerged, plus additional insights for the main theme "*xAI for Trust in AI*".

Table 12:

Themes Emerging from RQ2

` RQ2 Theme	**Similarities**	**Distinct Differences**	**Nuance of Difference**	**New Topics / Sub-Themes**
	Existing Theme	**New Theme**	**Sub-Theme**	
xAI and the "Black-Box"	**X**		XXX	Yes (2 New Sub-Themes)
xAI for Trust in AI	**X**		XX	Yes (1 new Sub-Theme and Additional Insights)
xAI for Decision-Making	**X**		XX	Yes (1 New Sub-Theme)
Antecedents		**X**	XX	Yes (New Main Theme, 2 New Sub-Themes)

Note: Author's Own.

In addition to the above, Table 13 below provides a summary of the frequency of occurrence of each of the five themes, while highlighting the topics most prevalent in the data from the different participant groups. The frequencies of mention are presented not in numbers, but as "Many", "Some", or "Few", as these descriptors are more suitable for qualitative analysis.

Table 13:

Frequency and Main Topics Related to RQ2

RQ2			
SELECTED THEMES	**Executive**	**Senior**	**Operational**
xAI and the "Black-Box"	Many	Few	Few
Main Topics	*Understanding and The "Black-Box"; The "Black-Box"; Transparency and the "Black-Box"*	*Understanding and The "Black-Box"; The "Black-Box"; Transparency and the "Black-Box"*	*Understanding and The "Black-Box"; The "Black-Box"; Transparency and the "Black-Box"*
xAI for Trust in AI	Many	Some	Many
Main Topics	*Explanation of Data; Explanation of "How?"*	*Explanation of Data; Explanation of "How?"*	*Explanation of Data; Explanation of "How?"*
xAI for Decision-Making	Few	Some	Few
Main Topic	*Self-Criticism of AI*	*Self-Criticism of AI*	*Self-Criticism of AI*
Antecedants	Many	Many	Many
Main Topics	*Provenance for Trust; Importance of Data*	*Provenance for Trust; Importance of Data*	*Provenance for Trust; Importance of Data*

Note: Author's own.

5.3.1 RQ2: Theme 1 – xAI and the Black Box

This theme is a main theme from the Chapter 2 literature review and is reflected in Figure 12 at the start of Chapter 5. The evidence gathered from the participants provided a rich set of insights. The use of colours connects the evidence to the respective participant groups, while the evidence itself is selected for discussion of the main topics shown in table 13. The three main topics discussed are "*Understanding and the "Black-Box"*", "*The "Black-Box"*" and "*Transparency and the "Black-Box*".

5.3.1.1 Evidence of *xAI and the "Black-Box"*.

Table 14:

Evidence of xAI and the "Black-Box"

RQ2: Theme 1 – xAI and the Black Box
With a focus on the topics "*Understanding and the "Black-Box"*", "*The "Black-Box"*" and "*Transparency and the "Black-Box"*.
Participant 1: *"I think that if you don't know how it works, and if you all you see is a black-box, you are not well served by blindly trusting it."*
Participant 1: *"And once again, you get to the whole black box question and obviously, the more proprietary things are, the more challenging it is to, you know, they don't want to give away their secret sauce."*

Participant 17: *"So I suppose the first thing we must not end up in a situation where the de facto language around the topic is a black-box. You know, as soon as you go down that path of it's a black box. It's already telling you something. Yeah, I don't know how it does it. So that's the first telltale sign."*
Participant 17: *"So I think if you've got those types of elements in there, the language never goes to "black-box" because of that transparency, explainability, etcetera."*
Participant 18: *"I suspect most our systems are black boxes, I assume, because it's probably an IP element around it. So the guys don't want to disclose what it's, what's in it, but I think some element of transparency probably would, would give you that level of level of comfort, external endorsement, you know, kind of the things that I've mentioned before."*
Participant 1: *"And you generally find the mis-trust is either born through a total misunderstanding of the technology or seeing it as a black-box."*
Participant 14: *"It must be transparent. You need to understand what went in there, in the processing that led to the outcome. If that's not transparent, it's like a "black-box". You know how we often talk about the "black-box"."*
Participant 3: *"As soon as there's clouds of shades of grey around it or questions of the result, you won't get that mass market adoption, people will retreat back to their little insular world."*
Participant 4: *"I think that's probably the key one around the quality and integrity of the data as we think about whatever, what is the, what is the, the source of the of the data that's being used in some of these LLMs and others that is sometimes a black box."*
Participant 4: *"I mean, so I think it… no one might pretend to understand it. I get the outputs of it, but it is a black box a lot of the time."*
Participant 9: *"So we've talked a lot, we haven't done it yet, about using AI to help us unpack what a suite of options would be, as opposed to us trying to build into more traditional models, like come up with... what does the in-between look like, and I personally would love to see It become a trusted augmenter, rather than a this sort of dark magic that sits in the background that no one understands."*
Participant 16: *"I can say so I firstly, I had to understand the method, you know, the neural network method and so forth, the how it works. Now, luckily that that was my background. So, you know, I, I trusted that quite quickly."*
Participant 6: *"It's a bit like a "black-box", right? If you put s#$t in, no matter how many times you know, what are we going to get? We're going to get, excuse the language, we're going to get s#$t out, right?"*
Participant 7: *"I think specific, especially for someone like me who's got no knowledge of it, and it kind of feels like you know the Wizard of Oz, the man behind the curtains, where you don't really know what you're dealing with."*
Participant 15: *"So if you've got full transparency, and you trust the output AI is giving you, then, yeah, however, decision-making is based on more… there's context, and there's experience, and there's the grey areas I was talking about."*

Note: Author's Own.

5.3.1.2 In-Case Analysis of the Evidence.

Within the Executive Decision-Making group, there was a shared concern about the "black-box" nature of AI and the need for transparency to build trust. Participant 1 expressed that a lack of understanding, or the perception of AI as a "black-box leads to mis-trust". They emphasized that Intellectual Property or proprietary know-how often limits transparency, making it challenging to fully trust AI technology. Similarly, Participant 18 acknowledged that many systems remain "black-boxes" due to Intellectual Property but suggested that some

degree of transparency could increase “comfort”. Participant 14 stressed the importance of understanding the inputs and processing within AI systems, likening the lack of transparency to the “black-box” issue, while Participant 17 discussed how transparency and explainability avoids the “black-box” language altogether.

The Senior Decision-Making group discussed the “black-box” issue from a more strategic perspective. Participant 3 noted that the lack of clarity or "shades of grey" around AI leads to resistance in widespread adoption, alluding to explanation being needed to gain user acceptance. Participant 4 referred to the data sources for AI models as being part of the “black-box”. Participant 9 expressed a desire for AI to be an understandable and trusted tool rather than remaining a “sort of dark magic”, which similarly highlights the importance of transparency of AI in developing trust. Lastly, Participant 16 emphasized the need for understanding the underlying methods of AI, noting that familiarity with technical aspects helps to build trust quickly. The Senior Decision-Making group showed a similar preference for explanation and clarity and emphasized the importance of understanding AI.

In the Operational Decision-Making group, participants also expressed concerns about the “black-box” nature of AI. Participant 6 described AI as a “black-box”, highlighting the concept of "garbage in, garbage out" and stressing that if the input data quality is poor, the outputs will be equally unreliable. Participant 7 likened the experience of working with AI to the "Wizard of Oz," where the “man behind the curtains” leads to discomfort. Participant 15 similarly agreed that transparency helps to build trust, but that decision-making still requires additional context beyond AI outputs.

5.3.1.3 Cross-Case Analysis of the Evidence.

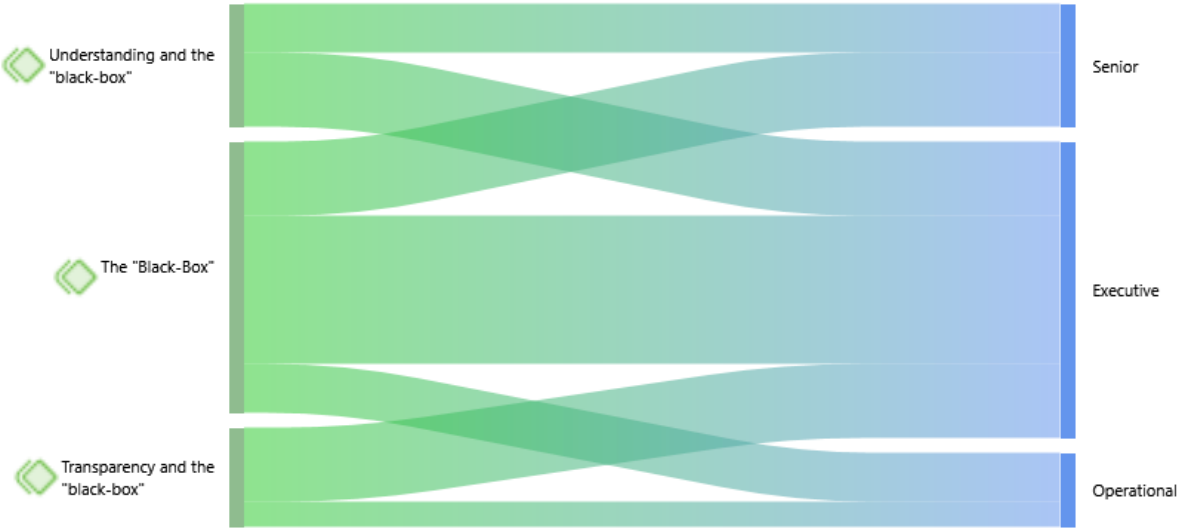
The comparison across groups highlighted a shared awareness of the “black-box” nature of AI and the approach to address it varied across groups. The Executive Decision-Making group mentioned all of explainability, transparency and understanding to address the “black-box” and emphasised explainability and understanding as a means to build trust. The Senior Decision-Making group mentioned the need for explanation and understanding, particularly of AI’s data sources and the methods the AI used, to build confidence and acceptance and to address the “black-box”. However, they did not specifically mention transparency to address the ‘black-box”. The Operational Decision-Making group similarly emphasized the importance of transparency, expressing discomfort when AI systems are opaque and also stressed that decision-making is based on context beyond just the outputs of AI. However, they did not mention either understanding nor explainability to address the “black-box”.

5.3.1.4 Distribution of the Sub-Themes by Participant Group.

The “Sankey” diagram below provides a visual representation of the distribution and flow of the sub-themes across different participant groups. The diagram is prepared from the total Atlas.ti code set related to the two sub-themes and provides a visual overview of the flow of discussion and density of quotations from the participant data.

Figure 18:

Sub-Theme Distribution: “Understanding & The “Black-Box””, “The “Black-Box””, “Transparency & The “Black-Box””



Note: Author’s own, generated from primary data in Atlas.ti software.

Understanding & the “Black-Box”: Of the participant groups, the Executive and Senior Decision-Making groups provided new insights closely linking understanding to the “black-box” and discussed this in their responses.

The “Black-Box”: Although the “black-box” was not specifically mentioned in the interview protocol, all of the participant groups picked up on the term or the concept and discussed it in their responses. This demonstrated an awareness of the phenomena of the “black-box”, with the Executive and Senior Decision-Making groups referring to explainability in relation to it. The Executive Decision-Making group was the only group who referred to all of explainability, transparency and understanding with respect to addressing the “black-box” nature of AI to create trust.

Transparency & the “Black-Box”: Both the Executive and Operational Decision-Making groups referred to and discussed transparency in relation to the “black-box”.

5.3.1.5 Conclusion on *xAI and the "Black-Box"*.

Across all participant groups, there was a similarity of concern regarding the "black-box" nature of AI, and the need for transparency was mutually recognized as important for building trust. All decision-making groups similarly acknowledged that the opacity of AI systems can lead to discomfort and mis-trust. While there were common concerns about the "black-box," all groups emphasized the importance of addressing this issue through various approaches. They specifically highlighted the role of not just explainability (xAI), but also transparency and understanding in ensuring trust in AI and its subsequent adoption.

As alluded to, there were differences in how each group approached the "black-box" problem. Some emphasized a broader set of tools, citing transparency, explainability, and understanding as being important to building trust and addressing the "black-box" issue. Others focused primarily on explanation and understanding, particularly of data sources and AI methods, though they did not emphasize transparency as much. The remainder focused mainly on transparency to resolve their discomfort but did not specifically highlight the need for explainability or understanding. These differences indicated that while all groups recognized the importance of tackling the "black-box" issue, some took a more comprehensive approach, using all of explainability, transparency and understanding to address it.

Key Concepts Identified: Awareness and shared concern of the "Black-Box" nature of AI. Transparency, understanding and explainability were all recognised to build trust and overcome the "Black-Box".

5.3.2 RQ2: Theme 2 – xAI for Trust in AI

This theme is a main theme from the Chapter 2 literature review and is reflected in Figure 12 at the start of Chapter 5. The evidence gathered from the participants provides a rich set of insights. The use of colours connects the evidence to the respective participant groups, while the evidence itself is selected for discussion of the main topics shown in table 13. The two main topics discussed are "*Explanation of Data*", and "*Explanation of "How?"*".

5.3.2.1 Evidence of *xAI for Trust in AI.*

Table 15:

Evidence of xAI for Trust in AI

RQ2: Theme 2 – xAI for Trust in AI
With a focus on the topics "*Explanation of Data*", and "*Explanation of "How?"*".
Participant 1: *"I think it would be, well, if there's one thing it would know be understanding how the model was trained."*

Participant 10: *"So from that perspective, I think you would need to have, you need to have a good explanation of the underlying assumptions of how it works things out, and what it calculates and what its subsets are, and what algorithms it does to decide what is reality and maybe it needs to explain which reality it's speaking from and which reality was the most prevalent."*
Participant 13: *"So I think it's, I think one, there's who, who kind of delivers the service. That's one, how are they using, say, your data or interaction? These are common problems we've got."*
Participant 17: *"So when this thing lands, if you can demonstrate that it is attached and correlated with those sort of things in in terms of how it works, it helps a great deal."*
Participant 10: *"What explanations? We would have to know what data set it was taking the information from."*
Participant 14: *"First, I need information, and I need to know that that is reliable information. I need to know that that information is relevant to my circumstances. I need that kind of thing, and that's why I'm saying I'd need that before decision-making."*
Participant 18: *"Because I'm obviously, I guess, kind of more mathematical, you know, statistical, I'd want some level of confidence in the quality of the data and the credibility of the underlying data."*
Participant 2: *"And would it be able to assure me that that decision? How could it assure me that the decision is actually the best decision for my business [...]?"*
Participant 3: *"And how has the AI informed each of those scenarios? You'd want to lay that out and explain it, […]."*
Participant 12: *"Um, I think, actually, the other thing is, is around how to use it. So what helped me get over that was I was invited to some training sessions by our marketing folks, and they actually showed, showed us how to use chat, GPT, and how to write prompts."*
Participant 2: *"So ultimately, if I've asked it to make a decision, and it gives me a decision, I'd want to know from it, which data sets did it pull From? What data did it did it use?"*
Participant 9: *"So in order to apply that level of trust, the absolute clarity of what that information or where that information goes and how it's handled through full life cycle needs to be absolute. So there is no open avenues."*
Participant 12: *"Yeah, well, then we need to say that. […] I've looked at your question and I've collected information from multiple sources, or something like that, where it's not just a single view."*
Participant 6: *"Showing it's methodology, showing how it's derived, the actual bit of data, derived its response, showing it. So I think showing that, not just providing it, right."*
Participant 8: *"So in implementation or use of AI, it'll have to explain to me how it's going to make the action, the process, whatever I'm doing, more efficient. How is it going to make whatever I'm doing efficient and efficient in in two ways, if it could save money, or can it save time, you know?"*
Participant 19: *"Yeah, so obviously, how it arrived at a certain decision."*
Participant 19: *"So, very clearly being able to show me how we had arrived at those outcomes, what data points it had taken into account, which data points potentially had higher weightings over others, [...] I think that would definitely help to increase the trust in the model."*
Participant 6*:"Yeah, so first, first thing would be visibility, or. Of the data visibility of the logic, right? So where, where is it getting the data from."*
Participant 15: *"What explanations I need to understand... [...] I need to understand what data it's used and what the source of the data was, and the integrity of the data, to provide the decision it made?"*

Note: Author's Own.

5.3.2.2 In-Case Analysis of the Evidence.

Within the Executive Decision-Making group, there was a similar emphasis on understanding both the methodology, or the "How?" behind AI systems and the data used to

generate outputs. Participant 1 highlighted the importance of understanding how the model was trained, while Participant 17 noted that demonstrating how AI outcomes correlate with business-relevant factors enhances trust. Participant 10 described the need for explanations about underlying assumptions and data sets, and also indicated that a clear understanding of "how it works things out" was important for their acceptance. Additionally, Participant 13 described a need to understand how data is used in the interaction with AI. Lastly, Participant 18 stressed the importance of explaining the credibility and quality of the data to give confidence. The Executives similarly showed that both the explanation of “how” and the reliability of data sources are central to building trust in AI systems.

In the Senior Decision-Making group, the focus was similarly on understanding both the AI’s methodology (again, the “How?”) and its data sources to ensure confidence in its decisions. Participant 2 stressed the need for assurance of how decisions are made, specifically wanting clarity on which data sets were used. Participant 9 emphasized the necessity for absolute clarity throughout the data life cycle to ensure trust, while Participant 3 suggested that explanations on how AI informs different scenarios are important. Participant 12 pointed out that training sessions that explain how to use AI effectively can help build understanding and acceptance. Participant 12 also described a need of the AI to explain its data sources. This group illustrated that explanation of the data sources and the methodology (the “How?”) behind AI’s decision-making process is important for adoption and trust in decision-making.

In the Operational Decision-Making group, there was a focus on practical and visible explanations of “how” AI functions as well as the transparency of its data sources. Participant 6 highlighted the importance of showing the methodology and how data and logic are used to generate AI outputs. Participants 8 and 19 respectively underscored the need for AI to explain how its decisions can improve efficiency and then how outcomes were arrived at, using what data. Participant 19 also mentioned that providing explanation of clear data points and weightings, would enhance trust. Similarly, Participant 15 reinforced the importance of understanding the integrity and source of the data used by AI to make a decision. This group emphasized that practical, detailed explanations and data transparency are important to trust and effective implementation of AI tools.

5.3.2.3 Cross-Case Analysis of the Evidence.

Across the groups, there was a similar emphasis on the need for transparency in both the AI methodology (the "How?") and the data used to build trust in AI systems. The Executive Decision-Making group prioritized both the methodology and data credibility. They similarly stressed that a clear explanation of “how” AI operates, and the reliability of data sources is central for trusting AI decisions. The Senior Decision-Making group similarly emphasized the

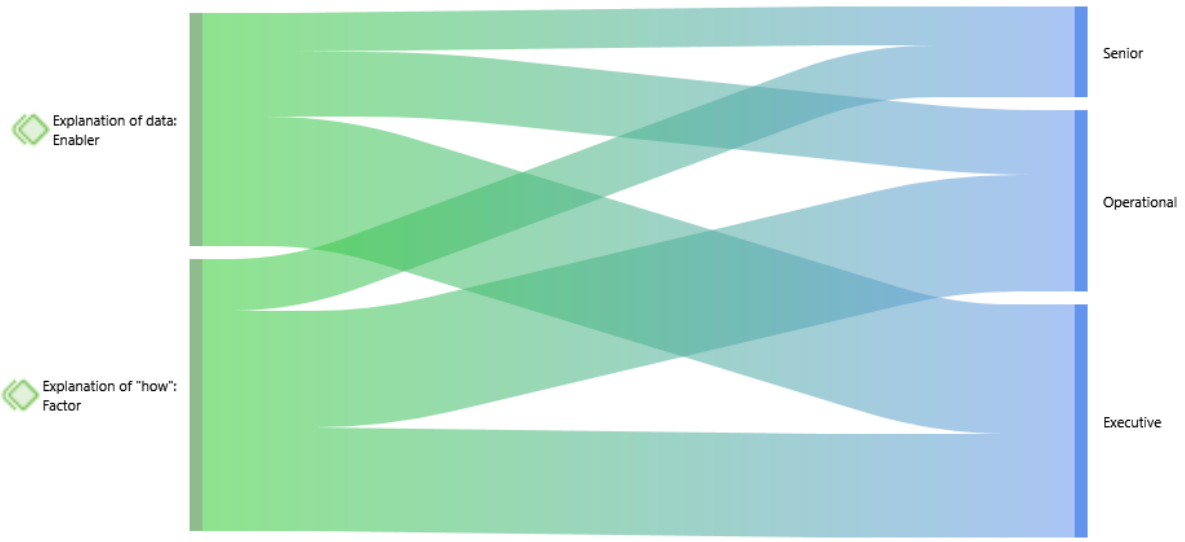
importance of explaining both the AI's methodology and its data sources and focused on these elements as essential for gaining trust and acceptance of AI. They stressed the need for clarity throughout the AI data life-cycle and also the need for explanations to demonstrate that AI's scenarios and decisions are trustworthy for effective decision-making. The Operational Decision-Making group concentrated on the practical aspects of transparency, also requiring understandable explanations of AI processes and data. They also focused on explanation of the efficiency gains from AI and validating the integrity of the data used as central to trusting AI decisions.

5.3.2.4 Distribution of the Sub-Themes by Participant Group.

The "Sankey" diagram below provides a visual representation of the distribution and flow of the sub-themes across different participant groups. The diagram is prepared from the total Atlas.ti code set related to the two sub-themes and provides a visual overview of the flow of discussion and density of quotations from the participant data.

Figure 19:

Sub-Theme Distribution: "Explanation of Data" and "Explanation of "How?""



Note: Author's own, generated from primary data in Atlas.ti software.

Explanation of Data: All of the participant groups discussed explanation of the data used by AI, with the most discussion coming from the Executive Decision-Making group. Although it was also discussed by the Senior and Operational Decision-Making groups, it was to a lesser extent.

Explanation of "How?": All of the participant groups similarly discussed a need for explanation of how AI achieves an output or decision, with almost equal emphasis between the Executive and Operational Decision-Making groups. There was less discussion or mention of the topic with the Senior Decision-Making group.

5.3.2.5 Conclusion on *xAI for Trust in AI.*

All participant groups similarly emphasized the importance of explaining both “how” AI systems operate and also the credibility of the data used, highlighting the importance of explanation and transparency of AI’s methodology and data sources to build trust. This shared recognition across the groups reinforced that understanding AI's processes and the data it relies on are important to fostering trust and facilitating AI adoption.

There was a difference in focus regarding what aspects of AI explanation were prioritized. Some groups emphasized both explanation of the AI methodology (the “How”) and explanation of data credibility as important to trusting AI decisions. Others concentrated on the need for clarity throughout the entire AI decision-making process, ensuring that all stages are transparent to support effective decision-making. Meanwhile, another group focused on practical aspects, such as how AI improves efficiencies and ensures data integrity. These variations reflected differing priorities in how transparency and explainability (xAI) are viewed as essential for building trust in AI systems.

A further distinct difference was from the Senior Decision-Making group who emphasised the importance of clarity (through explanation) throughout the data lifecycle.

Key Concepts Identified: Explainability should address “How” AI arrives at outcomes. Explanation must be consistent throughout the data lifecycle. Explanation requirements were varied across decision-making levels.

5.3.3 RQ2: Theme 3 – xAI for Decision-Making.

This theme is a main theme from the Chapter 2 literature review and is reflected in Figure 12 at the start of Chapter 5. The evidence gathered from the participants provides a rich set of insights. The use of colours connects the evidence to the respective participant groups, while the evidence itself is selected for discussion of the main topics shown in table 13. The main topic discussed is “*Self-Criticism of AI*”.

5.3.3.1 Evidence of *xAI for Decision-Making*.

Table 16:

Evidence of xAI for Decision-Making

RQ2: Theme 3 – xAI for Decision-Making
With a focus on the topic “Self-Criticism of AI”
Participant 10: *"It would almost question itself, right? Question itself, yeah, yeah. So almost do, like a "think-again" analysis."*

Participant 10: *"In a strange way, [it would be] comforting, if maybe in a very basic level, there were forums where AI then shared where it had made mistakes, yeah, or it was collated where it had made mistakes, and you look at the frequency of those errors and how it is self-corrected, potentially [...] I think that level of transparency or maybe it's not transparency, and maybe it's a collection of data sets of the sh#$%y outcomes of AI and where it's where it lied, yeah, and how it lied and why it lied, and then we understand the logic around it, yeah."*
Participant 11: *"And then the second factor is, how do I understand residual errors through machine learning training? So what reports could it give me to understand it's, you know, any self-diagnostics that it can do, or any self-verification that it could do."*
Participant 2: *"How could it assure me that the decision is actually the best decision for my business, and that it is not getting into like matrix territory, but it's not trying to sabotage me for some other entity or party's benefit, you know [...] so I'd like to understand what it's considered."*
Participant 3: *"Um, I work in a governance area, and I'm very conscious of audit trails, [...] If they're wanting to use AI, I think they would need to refer back to models and structures and how they've used the AI in the decision-making."*
Participant 2: *"Can it explain to what certainty or extent it could be wrong? Like, could it inform me of, like, its own limitations in that decision? Now I'd like to understand if it knows, maybe it can. Maybe can also explain to me its limitations or what it feels it's not good at. At the outset, what shouldn't I press it too much on?"*
Participant 2: *"Is it, yeah, again, is it aware of its own limitations."*
Participant 12: *"I think, because it's not, maybe it's not always going to be the perfect outcome, and that should also be communicated, you know, where it's worked, where it's not worked. [...] Failure is part of a [human] learning process. And I kind of maybe would like to see, you know, AI in that same way too. "*
Participant 6: *"If AI could self-interrogate, if you could actually go and say, Look, I've looked into what you've put up for me here, you've given me a particular set of data, if I could go in and say to whatever the tool was, why in this particular logic on this section, have you decided to choose this? Explain..."*

Note: Author's Own.

5.3.3.2 In-Case Analysis of the Evidence.

Within the Executive Decision-Making group, there was a focus on AI's ability to self-critique as a means to build trust and confidence in its decision-making. Participant 10 discussed the idea of AI performing a "think-again" analysis, suggesting that it would be reassuring if AI could openly report its errors and self-correct based on past mistakes. Participant 10 also emphasized the value of understanding when AI goes wrong and the frequency of its errors as a means of building understanding of AI's operation. Participant 11 highlighted the importance of AI offering self-diagnostics or self-verification reports, expressing a need for AI to demonstrate awareness of residual errors and to self-assess its performance.

In the Senior Decision-Making group, there was a similar emphasis on AI being able to acknowledge and communicate its limitations. Participant 2 wanted assurance that AI could recognize and explain its own weaknesses and potential errors in decision-making, stressing the importance of AI's awareness of its own limitations for building trust. Participant 3, with a governance perspective, indicated that having audit trails and being able to trace AI's models

and structures could serve as a form of self-assessment for accountability. Participant 12 viewed AI's ability to communicate both its successes and failures as part of a transparent learning process, likening it to human learning and emphasizing that such transparency would enhance their confidence in using AI.

In the Operational Decision-Making group, the emphasis was on actionable self-criticism. Participant 6 suggested the need for AI tools to be able to explain and justify specific decisions, demonstrating a capability to interrogate and validate its own logic. This participant stressed that AI should be able to provide explanations when questioned about its decisions, indicating a link with explainability as part of its self-assessment processes. These perspectives highlighted that AI's self-criticism must be explained for practical use.

5.3.3.3 Cross-Case Analysis of the Evidence.

All groups similarly emphasized the importance of AI self-criticism. The Executive Decision-Making group prioritized AI's ability to explore and communicate errors as a way to build confidence and credibility, framing self-criticism as being important for understanding. The Senior Decision-Making group emphasized AI's recognition and communication of its own limitations, linking this capability to trust and governance. They highlighted that AI almost needs to be self-aware and accountable for its decision-making to gain user acceptance. The Operational Decision-Making group focused on practical self-assessment and explanation, needing AI to justify its decisions in way that enhances its use in operational contexts.

5.3.3.4 Distribution of the Sub-Themes by Participant Group.

The "Sankey" diagram below provides a visual representation of the distribution and flow of the sub-themes across different participant groups. The diagram is prepared from the total Atlas.ti code set related to the two sub-themes and provides a visual overview of the flow of discussion and density of quotations from the participant data.

Figure 20:

Sub-Theme Distribution: "Self-Criticism of AI"

Note: Author's own, generated from primary data in Atlas.ti software.

Self-Criticism of AI: The diagram above shows that all of the participant groups mentioned some form of self-criticism, assessment or verification by AI to help engender trust for decision-making. The most prominent mention of this need for self-criticism was in the Senior Decision-Making group, followed by the Executives and then the Operational Decision-Making group.

5.3.3.5 Conclusion on *xAI for Decision-Making.*

All participant groups similarly showed that AI's ability for self-criticism is important to building trust and confidence in its decision-making. Some emphasized the need for AI to explore and communicate its errors as a way to build confidence, credibility, and understanding. Others similarly focused on AI's capacity to recognize and communicate its limitations, linking this self-awareness to trust and governance, and stressing the importance of accountability for AI's decision-making processes. The remainder prioritized practical self-assessment, requiring AI to justify its decisions to enhance operational use.

While all groups aligned on the importance of AI self-criticism, there was a difference in focus. Some highlighted self-criticism as important to fostering understanding, while others stressed a governance perspective and ability to communicate successes and failures, with a focus on transparency and accountability. The remainder concentrated on practical validation, requiring AI's self-criticism to be actionable and relevant to their operational context. These variations in perspective suggested that for xAI to be truly effective across an organization, it must integrate mechanisms for self-assessment, error reporting, and transparent communication of limitations.

Key Concepts Identified: AI self-criticism is necessary for trust and confidence in AI systems for decision-making.

5.3.4 RQ2: Theme 4 – Antecedents.

This theme is a potential new main theme and a distinct difference (on first analysis) from the Chapter 2 literature review and is reflected in Figure 12 at the start of Chapter 5. The evidence gathered from the participants provides a rich set of insights. The use of colours connects the evidence to the respective participant groups, while the evidence itself is selected for discussion of the main topics shown in table 13. The two main topics discussed are "*Provenance for Trust*", and "*Importance of Data*".

5.3.4.1 Evidence of *Antecedents*.

Table 17:

Evidence of Antecedents

RQ2: Theme 4 – Antecedents
With a focus on the topics "*Provenance for Trust*", and "*Importance of Data*".
Participant 10: *"Again, for me, I'm hitting on data sets, because that's, in the end, what it needs to take its information from."*
Participant 14: *"Well, I mean, I think it would be number one… can you trust the data, right? Is it, you know, rigorously tested?"*
Participant 18: *"I think [my organisation] would [use AI for decision-making], assuming we manage the quality of the data, I think it would."*
Participant 10: *"So that's also good, yeah, interesting. It would have to be a proven track-record."*
Participant 18: *"So, for example, once I asked that chat GPT thing, you know, who is [Participant] at [Company X], for example, okay, right? And it came back with, a whole lot of [#$%#], to be honest. Like that made me then have a fundamental concern about everything else that I'd ever asked it, because it came up with like, stuff that I knew was entirely incorrect."*
Participant 4: *"Trust is linked to the quality and integrity of the data, and then, like some sense of [...] are there inherent biases in it, and where does the data come from? That's determined some of the insights around how we make decisions..."*
Participant 4: *"It's not the AI necessarily, I think it's the quality and it's the kind of data we have in there isn't yet right for us to be, growing [adoption]. So I think that would [matter], because there is a high level of trust in some of the other areas."*
Participant 9: *"So trust in the in the data would go a long way to enable ongoing trust in in models now your model outcomes would be I think it would be time based like it would just like any, any adoption of new technology, right?"*
Participant 2: *"But, like, that's, that's one external factor that, I guess, you know, it'll be in a, oh, in a global conflict, where they are coming from? Is it coming from, you know, like the communist regime, or, ultimately, dictatorship, or someone, some country that's stealing data. So it's one aspect of it, I guess."*
Participant 4: *"That's right, yeah, that's provenance. Yeah, they have some credibility of something that they've built. They've got some, something in the bank that you can then... [...] There's a trust that comes with it, and then can be expanded to other areas of AI."*
Participant 9: "*Trust is built over time that that we do have some consistency and some evidence-based data of good decision making and appropriate and correct outcomes."*
Participant 5: *"If the trust was in place, and I know that this data is solid. I've got controls in place to monitor and manage the fact that this data reconciles. And, you know, knowing all of that, I can then use it with trust."*
Participant 15: *"So, that level of trust is there because of the data integrity behind it."*
Participant 19: *"Like I said earlier, seeing source references, being able to glance over them, validate them, and be sure that come from reliable sources that would be sufficient for me, you know, for that particular use case."*
Participant 5: *"So I mean, in terms of the trust, you know, we're old-fashioned in terms of trust. It has to be something that is well entrenched in the market."*
Participant 5: *"It's not, we're not going to be the first-time adopters on any package of product. It has to be something that has gained its traction, it's got stellar reviews. And then will we be the adopters to go, okay, it works, and now we can be open to start trusting it, but we won't be the first ones to implement and get burned and develop trust in that way."*

Participant 8: *"I'd need to then test, to do extensive testing to see if the results are kind of within our expected parameters."*
Participant 19: *"So, very clearly being able to show me how we derived at those outcomes, what data points it had taken into account, which data points potentially had higher weightings over others, yeah, all of the Yeah, it goes back to just seeing the data points that it utilized in deriving that decision. I think that would definitely help to in to increase the trust in the model."*

Note: Author's Own.

5.3.4.2 In-Case Analysis of the Evidence.

In the Executive Decision-Making group, the emphasis was on the provenance of the AI tool and integrity of data as important antecedents to building trust in AI. Participant 10 highlighted the need for reliable data sets and a proven track record for AI to gain their confidence and trust. Participant 14 questioned whether the data used by AI is rigorously tested, indicating that the quality of inputs possibly influences their level of trust. Participant 18 mentioned that trust in AI would depend on managing data quality effectively, and also noted that when an AI provided incorrect information, it damaged their overall trust in the system. This group showed that for Executives, trust in AI possibly depends on the quality of source data and the proven reliability, and historical use of the AI tool.

The Senior Decision-Making group similarly emphasized the importance of data integrity and provenance of the AI tool for trust. Participant 4 linked trust directly to the quality and biases in the data, suggesting that the origins and consistency of the data shape decision-making processes. Participant 9 indicated that trust builds over time, through consistent good decision-making and correct outcomes. Participant 2 added that data is particularly important when external influences (e.g., political regimes) might impact the integrity of information. This group underscored that understanding the source and consistency of data is important for fostering long-term trust in AI's outputs.

In the Operational Decision-Making group, there was a similar focus on practical, tested AI and good data integrity for building trust. Participant 5 emphasized that for them to trust AI, it must be backed by well-established market credibility and solid, verified data. Participant 8 highlighted the need for extensive testing to validate AI results, reinforcing the idea that trust depends on credible performance and alignment with expected outcomes. Participant 19 noted that having access to source references and being able to validate or verify the origin of data directly impacts their level of trust. This group similarly emphasized that a proven track record of the AI and data integrity are important for operational adoption and trust in AI.

5.3.4.3 Cross-Case Analysis of the Evidence.

Across all groups, the focus on AI provenance and data integrity as important factors for trust was similar. The Executive Decision-Making group emphasized the need for good

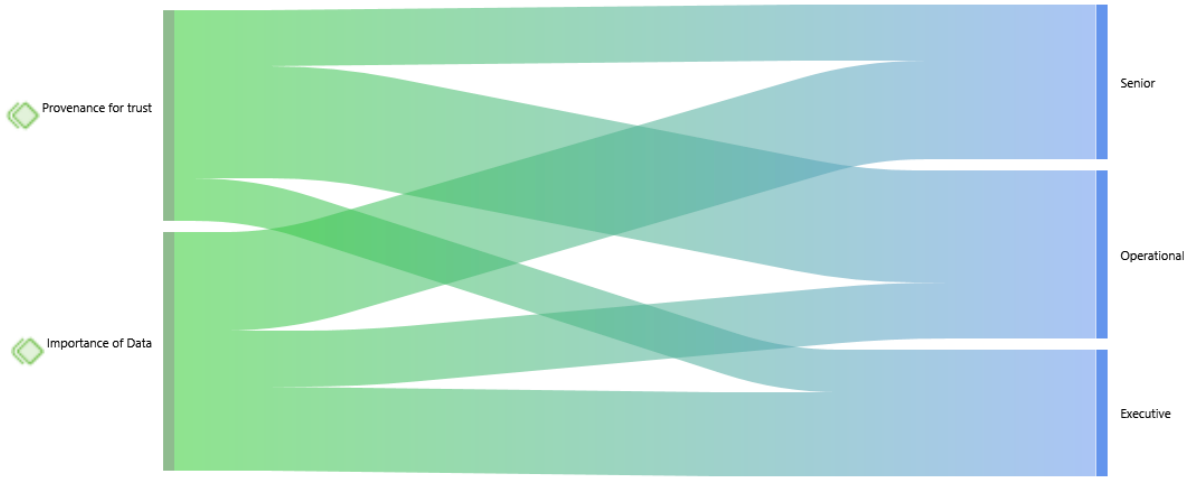
quality, rigorously tested data and a proven track record for the AI tool itself. They therefore associated trust with both the historical reliability of the technology and the quality of its data sources. The Senior Decision-Making group also prioritized data integrity, linking it directly to the tool's ability to deliver consistent, accurate outcomes over time. They stressed that understanding the source and consistency of the data was important, especially when external factors could influence data quality. Similarly, the Operational Decision-Making group focused on practical validation, market credibility, and the need for verifiable data. They highlighted that testing and proven reliability are important for them to adopt and trust AI tools in their daily operations.

5.3.4.4 Distribution of the Main Theme by Participant Group.

The "Sankey" diagram below provides a visual representation of the distribution and flow of the sub-themes across different participant groups. The diagram is prepared from the total Atlas.ti code set related to the two sub-themes and provides a visual overview of the flow of discussion and density of quotations from the participant data.

Figure 21:

Sub-Theme Distribution: "Provenance for Trust" and "Importance of Data"



Note: Author's own, generated from primary data in Atlas.ti software.

Provenance for Trust: The majority of discussion for this theme was found in the Operational Decision-Making group's interviews, followed by Senior Decision-Making group and then the Executives. The sub-theme was notably discussed with all three decision-making groups, indicating their awareness of the topic.

Importance of Data: This sub-theme was mostly discussed with the Executive and Senior Decision-Making groups, with an almost equal prevalence between them. Again, it was interesting to note that the sub-theme was discussed across all three decision-making groups, again showing broad awareness.

5.3.4.5 Conclusion on *Antecedents*.

The significance of both provenance and data integrity was similarly emphasized across the participant groups as foundational for building trust in AI. The participant groups similarly recognized the need for validated and credible data, as well as a proven track record, to trust AI systems. This shared focus highlighted the importance of ensuring reliable data inputs and a history of successful AI deployment. This alluded to trust in AI having the antecedents of both quality of data and the tool's demonstrated performance over time.

All groups shared this perspective, but there was a difference in their focus. Some emphasized the need for rigorously tested data and a proven historical reliability of the AI tool itself, linking trust to both data quality and the technology's track record. Others similarly prioritized data integrity but stressed the importance of understanding the data's origin, especially when external factors could influence its credibility. The remainder focused more on practical validation, requiring verifiable data and extensive testing to ensure reliable performance in their operations. These variations suggested that while provenance and data integrity are similarly important, the emphasis shifted slightly between strategic oversight and operational use.

Key Concepts Identified: Provenance is an antecedent to adoption. Data integrity is an antecedent to adoption.

5.3.5 Summary of Similarities and Differences in the Findings for RQ2

Table 18 below is provided as a summary and convenient reference for the findings related to the sub-themes discussed under the main themes of RQ2. The table is prepared from the in-case and cross-case analyses and conclusions for each of the four themes presented in Section 5.3.

Table 18:

Summary of Similarities and Differences in the Findings, Across Themes, for RQ2.

RQ2: *How does xAI engender trust in AI for decision-making in organisations? And: What are the factors that contribute to trust in AI?*		
	Similarities	**Differences**
Theme 1: xAI and the "Black-Box"	• Awareness and shared concern of the "Black -Box" nature of AI. • Transparency recognised across all groups to build trust and overcome the "Black-Box".	• Some mentioned explainability, transparency and understandability to overcome the "Black-Box", while others only mentioned Explanation and Understanding. The remainder only mentioned transparency to overcome the "Black-Box".

Theme 2: xAI for Trust in AI	• Shared view that explainability should address "How" AI arrives at outcomes. • Shared view that explainability of data sources and credibility is required. • Explanation must be consistent throughout the data lifecycle.	• The Senior Decision-Making group was the only group to stress the need for clarity throughout the data lifecycle. • Focus varied between the participant groups. Some wanted explanation of methodology and data credibility, others wanted explanation of the decision-making process, and the remainder focused on practical aspects of explanation.
Theme 3: xAI for Decision-Making	• AI Self-Criticism is necessary for trust and confidence in AI systems for decision-making.	• AI must self-criticise for confidence and understanding. • AI self-criticism must communicate limitations. • AI self-criticism must validate AI decisions.
Theme 4: Antecedents	• Provenance of prior or successful use is an antecedent to adoption. • Data integrity is an antecedent to adoption.	• While provenance and data integrity are universally important, the emphasis shifts between historical reliability, external factors influencing data quality and practical focus requiring extensive testing.

Note: Author's Own.

5.3.6 Conclusion to the Findings for RQ2: Summary of Key Concepts

In conclusion to the findings for RQ2, the similarities and differences from each of the discussed themes and their conclusions were reviewed, to derive the "Key Concepts" which emerged from the theme, topic and sub-theme findings. These key concepts are derived from both the similarities and differences which emerged and are presented in Table 19 below.

Table 19:

Summary of Key Concepts from the Findings, by Theme for RQ2.

RQ2: *How does xAI engender trust in AI for decision-making in organisations? And: What are the factors that contribute to trust in AI?*		
	Topics / Sub-Themes	**Key Concepts Identified**
Theme 1: xAI and the "Black-Box"	• *The "Black-Box"* • *Understanding and the "Black-Box"* • *Transparency and the "Black-Box"*	• Awareness and shared concern of the "Black -Box" nature of AI. • Transparency, understanding and explainability were all recognised to build trust and overcome the "Black-Box".
Theme 2: xAI for Trust in AI	• *Explanation of Data* • *Explanation of "How?"*	• Explainability should address "How" AI arrives at outcomes. • Explanation must be consistent throughout the data lifecycle.
Theme 3: xAI for Decision-Making	• *Self-Criticism of AI*	• AI Self-Criticism is necessary for trust and confidence in AI systems for decision-making.
Theme 4: Antecedents	• *Provenance for Trust* • *Importance of Data*	• Provenance is an antecedent to adoption. • Data integrity is an antecedent to adoption.

Note: Author's Own.

5.4 Findings for Research Question 3: *How does transparency influence emotional trust in AI for organisational decision-making?*

The themes related to the research question will be discussed in this section. As discussed in Chapter 4, differences that arose from the analysis identified some distinct differences and some nuances of difference. The former are shown, in Table 20, as potential new themes and the latter as potential new sub-themes. For RQ3, four (4) main themes

emerged from the data analysis, and one (1) of these is potentially new. A further seven (7) potential new sub-themes also emerged, as well as additional insights for the main theme "*Transparency for Decision-Making*".

Table 20:

Themes Emerging from RQ3

RQ3 Theme	Similarities	Distinct Differences	Nuance of Difference	New Topics / Sub-Themes
	Existing Theme	New Theme	Sub-Theme	
Transparency for Trust in AI	X		XX	Yes (2 New Sub-Theme)
Transparency for Decision-Making	X		XX	Yes (1 new Sub-Theme and Additional Insights)
xAI and Transparency	X		XX	Yes (2 New Sub-Theme)
Conjoined Agency		X	XX	Yes (New Main Theme, 2 New Sub-Themes)

Note: Author's Own.

In addition to the above, Table 21 below provides a summary of the frequency of occurrence of each of the five themes, while highlighting the topics most prevalent in the data from the different participant groups. The frequencies of mention are presented not in numbers, but as "Many", "Some", or "Few", as these descriptors are more suitable for qualitative analysis.

Table 21:

Frequency and Main Topics Related to RQ3

RQ3			
SELECTED THEMES	**Executive**	**Senior**	**Operational**
Transparency for Trust in AI	Many	Many	Many
Main Topics	*Fear as a Barrier; AI Identification of Human Nuance*	*Fear as a Barrier; AI Identification of Human Nuance*	*Fear as a Barrier; AI Identification of Human Nuance*
Transparency for Decision-Making	Few	Some	Few
Main Topics	*Transparency and Understanding*	*Transparency and Understanding*	*Transparency and Understanding*
xAI and Transparency	Many	Some	Many
Main Topics	*Understanding of "How?"; Understanding and Trust*	*Understanding of "How?"; Understanding and Trust*	*Understanding of "How?"; Understanding and Trust*
Conjoined Agency	Many	Some	Many
Main Topics	*Human-AI Conjoined Agency; AI Assistance of Humans*	*Human-AI Conjoined Agency; AI Assistance of Humans*	*Human-AI Conjoined Agency; AI Assistance of Humans*

Note: Author's own.

5.4.1 RQ3: Theme 1 – Transparency for Trust in AI

This theme is a main theme from the Chapter 2 literature review and is reflected in Figure 12 at the start of Chapter 5. The evidence gathered from the participants provided a rich set of insights. The use of colours connects the evidence to the respective participant groups, while the evidence itself is selected for discussion of the main topics shown in table

21. The two main topics discussed are “*Fear as a Barrier*”, and “*AI Identification of Human Nuance*”.

5.4.1.1 Evidence of *Transparency for Trust in AI*.

Table 22:

Evidence of Transparency for Trust in AI

RQ3: Theme 1 – Transparency for Trust in AI
With a focus on the topics “*Fear as a Barrier*”, and “*AI Identification of Human Nuance*”.
Participant 13: *"As we currently see with humans, not all are equally intelligent, but the toughest thing in the industry is distilling requirements and architecting things, those kind of I don't see AI being anywhere near it yet on that, and if it were, it might be a little bit copy and paste or lacking the nuance in the situation at this point."*
Participant 17: *"I suppose, the nature of the language, the language model or the personality model, it must feel authentic."*
Participant 17: *"But if you are an AI representing a brand that needs to engage with people, in good times, in bad times, and moments of vulnerability, then you almost want an AI that strikes the right balance and tone and temperament so that when it engages with you, the responses are fit for purpose for all those scenarios. So that’s one thing, the tonality around it."*
Participant 1: *"I think there's a lot of fear. If you're talking about the general public, the first one, of course, is security of employment and tenure. There's anxiety around potential job losses."*
Participant 11: *"Yeah, if we know what it is, and then we know either there's no need to be afraid of it, or there is need to be afraid of it in certain contexts, but that learning needs to kind of be there."*
Participant 13: *"I think it's because of the industries that I'm working within. Some are creative, [...] so in that sense, there's a bit of fear that AI could be replacing what people do."*
Participant 17: *"I think what I have seen in most cases, it's a vulnerability topic. Those vulnerabilities play out a lot. One, they don't truly understand what AI can do in general, but more specifically, and this is where the fear comes in. They do because they don't know that [AI]."*
Participant 2: *"So, like, has it got human interests? Like, has it really thought of the human element of it? I would like to know, you know that I would like to know that it's considering that, you know, if I'm going to really feel that it is trustworthy."*
Participant 2: *"The one, the one part that, that I think it's going to battle with is, like, the ultimately, that human element of just the importance of relationships and business."*
Participant 3: *"You need to have safe adoption. Um, and if I think back to the launch of an iPhone, for example, yeah, what caused it to be used by the masses? It was dead simple, it was intuitive. It felt very low risk. It helped you. It enhanced your life." [without the perception of low risk] You'd go back to people being more insular and more fearful of using it.*
Participant 16: *"So the next thing was that the actual data sets did make intuitive sense, you know. So I could believe it came up with the 18 variables it came up with. […] Well, okay, the location at night is also the location where people most likely live. And that made sense, you know."*
Participant 2: *"What actually could happen if we brought this in, you know? So I think based on that alone, I think because no one, people are scared, I guess, in any organization to make a mistake."*
Participant 12: *"I still link it back to fear you I think once, once people have moved beyond that state, they would be able to trust the tool and it will help them to, you know, make informed decisions."*

Participant 16: *"It's perhaps the inherent fear of lack of trust, you know, and so forth, that you have to understand the method that it was derived, rather than the actual output of it, you know."*
Participant 6: *"But to take over that human level of understanding and our ability to look at the detail of every single thing, a question that I can't see it at the moment, be powerful enough to do that just yet, right?"*
Participant 7: *"Again, I think the problem, and probably the reason why it's not used so much, is because, like we said, of the nuance and that personality that is missing, that is needed in our industry, and there's no insight that it can do that."*
Participant 15: *"And there's some nuances, I think, especially in our business and maybe other businesses where everything is fully automated. There might be a higher level of trust, but where there are legacy systems and manual processes involved and you supplementing it with AI, you can't, I think, wholly trust AI to make your decisions on what your AI says."*
Participant 5: *"I mean, if you ask me, where do I see AI going? I honestly don't think it's going to be like when we introduced spreadsheets. A lot of people said, Oh, you're going to take away jobs. But it didn't take away jobs. It enhanced what people were doing in their job function."*
Participant 6: *"But then it's also around protecting how staff and employees and others working within jobs see the AI not as a threat, but as a tool."*
Participant 8: *"So I think it's key in keeping fears at bay. There's always going to be a little bit of a hesitation, especially at the beginning. So you need it needs to be transparent."*
Participant 19: *"I think there's still quite a huge fear of AI […] fear that AI is going to take over our jobs. […] So you do have this resistance, you know, from an adoption point of view, and I've heard it even, you know, at the boardroom table. I've heard that roles might become redundant."*
Participant 19: *"I mean, there's still a lot of fear around just sharing your data, because you don't know how your data is going to be used, how it's going to be protected. People are worried about sharing it, because it can obviously get out there."*

Note: Author's Own.

5.4.1.2 In-Case Analysis of the Evidence.

Within the Executive Decision-Making group, there was a notable concern regarding fear as a barrier to AI adoption. Participant 1 pointed to job security and employment tenure as significant sources of anxiety. Participant 13 expressed fear specifically in creative industries where AI could potentially replace human roles. Participant 17 linked fear to vulnerabilities, particularly when people do not fully understand AI's capabilities. This group therefore showed that fear, centered on job security, acts as an emotional barrier to trust and adoption. Regarding AI's ability to identify human nuances, Participant 17 emphasized the importance of AI accurately reflecting tone and temperament in AI interactions, suggesting that AI must adapt its responses to fit the scenario. Participant 13 noted that AI still struggles with nuances, indicating a gap between current AI capabilities and human experience and industry understanding.

The Senior Decision-Making group expressed similar concerns about fear as a barrier. Participant 3 linked fear to risk perception, using the example of the iPhone's simple, intuitive design that reduced perceived risk and facilitated mass adoption. Participant 12 related fear directly to the lack of trust in AI, suggesting that overcoming fear is important for trust and

informed decision-making. In terms of AI's understanding of human nuance, Participant 2 expressed scepticism about AI's ability to grasp human interests and relationship elements in business, stating that without these, trust would be challenging. Participant 16 similarly added that for AI outputs to be trusted, they must make intuitive sense.

In the Operational Decision-Making group, fear was also evident, particularly concerning job loss and data privacy. Participant 19 noted the fear of job redundancy and reluctance to share data due to fears over its protection. Participant 8 stressed that transparency is key to managing these fears, particularly during initial adoption phases. The group also similarly highlighted the current limitations of AI in capturing human nuances. Participant 6 noted that AI does not yet have the ability to handle detailed human understanding effectively, while Participant 7 mentioned that AI's inability to replicate human nuance and personality remains a barrier in their industry. Lastly, Participant 15 added that AI struggles to navigate legacy systems and manual processes, which further impacts trust.

5.4.1.3 Cross-Case Analysis of the Evidence.

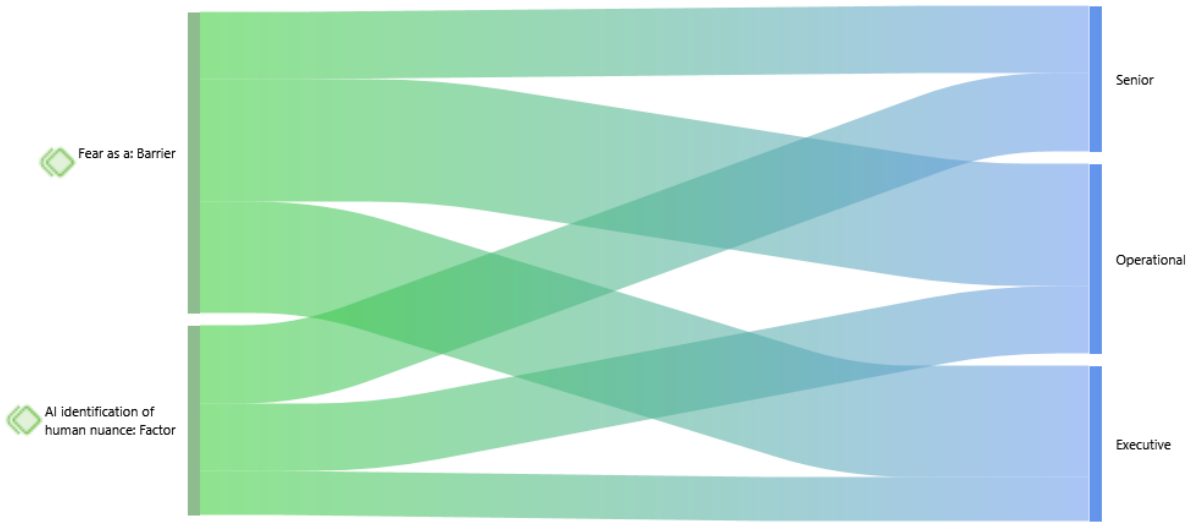
Across all groups, "fear as a barrier" was a consistent theme, with job security, data privacy, and AI's inability to handle elements of human nuance being central concerns. The Executive Decision-Making and Operational Decision-Making groups highlighted job security as a main anxiety, with executives focusing on industry-specific fears (e.g., in creative sectors) and operational staff emphasizing concerns around job redundancy and data protection. The Senior Decision-Making group similarly highlighted fear but expanded their focus to include risk perception and lack of trust. Regarding "AI's identification of human nuance", all groups similarly acknowledged its current limitations but with difference in focus. The Executive and Senior Decision-Making groups emphasized AI's need to identify tone, temperament, and relational elements respectively, to build trust. The Operational Decision-Making group stressed AI's inability to match human nuance and personality as an important barrier to trust.

5.4.1.4 Distribution of the Main Theme by Participant Group.

The "Sankey" diagram below provides a visual representation of the distribution and flow of the sub-themes across different participant groups. The diagram is prepared from the total Atlas.ti code set related to the two sub-themes and provides a visual overview of the flow of discussion and density of quotations from the participant data.

Figure 22:

Sub-Theme Distribution: "Fear as a Barrier" and "AI Identification of Human Nuance"



Note: Author's own, generated from primary data in Atlas.ti software.

***Fear as a Barrier*:** The sub-theme was notably discussed with all three decision-making groups indicating their awareness of the topic. There was an almost even split in focus between the Operational and Executive Decision-Making groups, with the Senior Decision-Making group discussing the Sub-them only slightly less.

***AI Identification of Human Nuance*:** This sub-theme was also discussed across all three participant groups, mostly at the Senior and Operational Decision-Making levels, with a similar prevalence between them. Again, it was interesting to note that the sub-theme was discussed across all three decision-making groups, again showing broad awareness of its importance.

5.4.1.5 Conclusion on *Transparency for Trust in AI*.

"Fear as a Barrier" and "AI Identification of Human Nuance" are important sub-themes that influence transparency and trust in AI across participant groups. Fear was similarly highlighted as a barrier to adoption across all participant groups. This was particularly regarding job security, data privacy, and AI's ability to manage human relationships. All groups also similarly recognized that transparency and clear communication are essential to reducing these fears and fostering trust in AI. Lastly, there was a similar view that AI has limited ability to identify human nuance and that this is an inhibitor to trust. The similar narratives suggested that being transparent about AI's capabilities and limitations may reduce fear and is important to alleviating anxieties. In turn, this could pave the way for greater acceptance and trust of AI.

Differences emerged in how each group framed these sub-themes. Most participants emphasized job security, with some focusing on industry-specific fears, such as those in

creative fields, and others extended the discussion to relationship-based fears and risk perceptions. There was also a focus on AI's inability to match human insight and personality. Each group addressed the sub-themes through improved AI design and enhanced transparency. Therefore, more intuitive interaction could increase trust and acceptance of AI across organizational levels.

Key Concepts Identified: Fear is a barrier to adoption and is primarily related to job security. Transparency is required to alleviate fear and anxiety and foster trust. AI has limited ability to identify human nuance, which inhibits trust.

5.4.2 RQ3: Theme 2 – Transparency for Decision-Making

This theme is a main theme from the Chapter 2 literature review and is reflected in Figure 12 at the start of Chapter 5. The evidence gathered from the participants provides a rich set of insights. The use of colours connects the evidence to the respective participant groups, while the evidence itself is selected for discussion of the main topics shown in table 21. The main topic discussed is "*Transparency and Understanding*".

5.4.2.1 Evidence of *Transparency for Decision-Making*.

Table 23:

Evidence of Transparency for Decision-Making

RQ3: Theme 2 – Transparency for Decision-Making
With a focus on the topic "*Transparency and Understanding*"
Participant 10: *"[When asked about transparency for trust - Q8] For me, it's understanding, again, the source of the information, and understanding how it manipulates the data, how it runs through it, what the algorithm is, and we wouldn't necessarily understand it to its infinite detail, but getting a broad-based understanding of what logic is used there."*
Participant 11: *"So if there was transparency about what it is, and what you were getting, and I understood the steps along the way […] I think it can improve it [trust], because it [AI] becomes a relatable thing."*
Participant 14: *"It must be transparent. You need to understand what went in there, in the processing that led to the outcome. If that's not transparent, it's like a black-box. You know how we often talk about the black-box?"*
Participant 2: *"So, the transparency gives the leadership, [...] or the organization, the transparency would give the leadership a baseline, in a way, to test, like to gage... It's to gage what's behind it, who's behind it, [...] it comes down to sense checking, because you're trying to understand what's behind it, [...] It's the intention behind the software."*
Participant 2: *"I would understand what its limitations are, because it'd be transparent as to what it can do, what it can't do, so I would know how far I could stretch it and still expect good results."*
Participant 4: *"So as part of transparency, I think it's important to understand some of the basics and to have some understanding of like, what these models are, how they're built, and what they mean and how they work."*
Participant 9: *"Yeah, I mean this, this does go and overlap a little bit with my previous answer, in that the transparency would go a long way to facilitate understanding. Understanding allows and facilitates trust."*

Participant 9: *"So allowing transparency in an AI would allow me to understand what its data sources are, how it interprets that data source, whether that interpretation is the same or aligned with what my objectives and my understanding are, how it intends to manipulate that data set and how it arrives at outcome. So for me, that is transparency."*
Participant 9: *"It's that understanding and the explanation which I don't think the people who are developing those broadly used models are willing to divulge at this point because it, it's their competitive advantage."*
Participant 16: *"I don't think it's just transparency for humans to understand. At some stage you're going to get to some state, you know, "so what?" you have transparency but no one understands anyway."*
Participant 7: *"And I think just understanding that in more detail, I guess I can liken it to a human relationship with a friend or a boyfriend or a husband or whatever. And how you build trust there. It's transparency, and it's mutual understanding."*
Participant 15: *"So I need to understand what's being fed into the model, [...] So the transparency in terms of how it's been built and what information is going in there, and how the information is being used and manipulated by the AI as well, to make those decisions is very important in trusting the AI, otherwise there's going to be no trust."*
Participant 15: *"The transparency is critical. Understanding what's going in is critical. I can trust this information, but I must test the information, yes."*

Note: Author's Own.

5.4.2.2 In-Case Analysis of the Evidence.

In the Executive Decision-Making group, there was a strong emphasis on transparency as a foundation for understanding AI and, consequently, building trust. Participant 10 expressed the need to understand the source of information and the underlying logic used by AI. They indicated that this would develop a broad understanding of how the system operates. Participant 11 highlighted that transparency in the AI process could make it more relatable and improve trust. Participant 14 similarly emphasized the necessity of understanding the inputs and processes behind AI outcomes. Interestingly, they highlighted again as per a previously-presented sub-theme that a lack of transparency leads to a "black-box" that limits trust. This group showed that for executives, transparency is important in gaining an understanding of AI, which is directly linked to fostering trust in its decision-making.

The Senior Decision-Making group also stressed the relationship between transparency and understanding. Participant 2 noted that transparency provides a baseline for testing and sense-checking AI's outputs and to offer clarity on the tool's limitations and scope of capability. Participant 4 mentioned that understanding the basics of how AI models are built is important, while Participant 9 reiterated that transparency facilitates understanding, which, in turn, fosters trust. However, they also expressed concern that AI developers might be unwilling to divulge key details, as this transparency could impact their competitive advantage. Participant 16 added an important view, suggesting that transparency alone was insufficient if the users cannot comprehend the information provided. This group similarly highlighted that transparency must lead to meaningful understanding for it to be effective in building trust.

In the Operational Decision-Making group, transparency and understanding were also similarly highlighted as important for trust. Participant 15 emphasized that a practical understanding of what information is fed into the model and how it is manipulated by AI is important to trust. They stressed that without transparency and a clear understanding, trust would not be possible. Participant 7 compared transparency in AI to building trust in human relationships, where mutual understanding is an important component. This group similarly identified transparency and clear understanding as being important to trust.

5.4.2.3 Cross-Case Analysis of the Evidence.

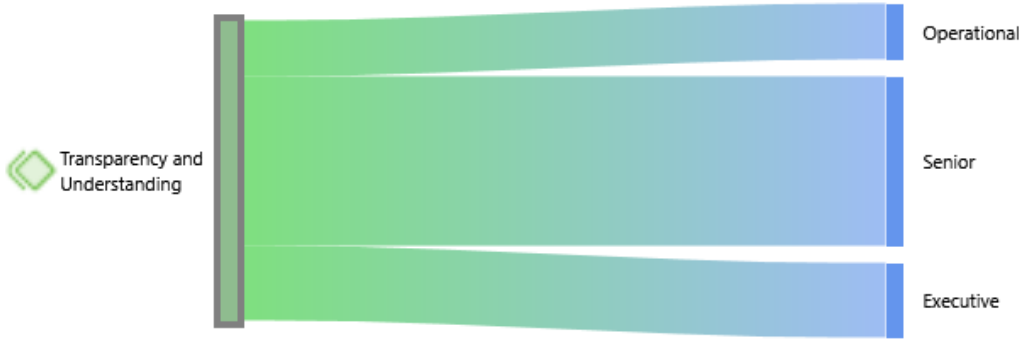
The Executive Decision-Making group saw transparency as being important for understanding AI functionality, which directly influences trust. Similarly, the Senior Decision-Making group viewed transparency as a mechanism for clarity but stressed that it must lead to a meaningful understanding for it to be effective. The Operational Decision-Making group emphasized transparency and clear practical understanding being important to trust, noting that without transparency and clear practical understanding, trust cannot be established at an operational level. This relationship between transparency and understanding suggested that for AI to be trusted across all levels of an organization, transparency must not only be available but also tailored to each group's needs. This would ensure that it results in a comprehensive and meaningful understanding of AI's capabilities and limitations.

5.4.2.4 Distribution of the Main Theme by Participant Group.

The "Sankey" diagram below provides a visual representation of the distribution and flow of the sub-themes across different participant groups. The diagram is prepared from the total Atlas.ti code set related to the two sub-themes and provides a visual overview of the flow of discussion and density of quotations from the participant data.

Figure 23:

Sub-Theme Distribution: "Transparency and Understanding"



Note: Author's own, generated from primary data in Atlas.ti software.

Transparency and Understanding: The majority of discussion for this theme was found in the Senior Decision-Making group's interviews, followed by the Executive and then

Operational Decision-Making groups. The sub-theme was notably discussed with all three decision-making groups, indicating their awareness of the topic.

5.4.2.5 Conclusion on *Transparency for Decision-Making*.

Across all participant groups, there was a shared emphasis on the importance of transparency in fostering trust in AI for decision-making. All groups agreed that transparency is important for understanding how AI functions, which should allow for better risk assessment and informed decision-making. The consistent message was that without transparency, trust cannot be established, as users need to understand how AI processes data and produces outcomes to feel confident in relying on its decisions.

There was some difference in how each group viewed transparency. Some saw transparency as essential to understanding AI's functionality, directly linking it to trust in decision-making. Others highlighted the need for transparency to provide clarity, but they also stressed that transparency must lead to meaningful understanding for it to be truly effective. The remainder focused more on practical understanding, emphasizing that trust at the operational level depends on clear and easily understood explanations of AI processes. These differences suggested that while all groups valued transparency, they prioritized it in slightly different ways depending on their role within the organization.

Key Concepts Identified: Transparency and Understanding are closely related and directly linked to trust in AI for decision-making. Different priorities for transparency, depending on the role of the individual responsible for decision-making within the organization.

5.4.3 RQ3: Theme 3 – xAI and Transparency

This theme is a main theme from the Chapter 2 literature review and is reflected in Figure 12 at the start of Chapter 5. The evidence gathered from the participants provides a rich set of insights. The use of colours connects the evidence to the respective participant groups, while the evidence itself is selected for discussion of the main topics shown in table 21. The two main topics discussed are "*Understanding of "How?"*", and "*Understanding and Trust*".

Note: It should be noted that for this theme, answers were grouped per RQ3 which related to transparency (in keeping with the structure of the report), and the discussion focused on the topics around understanding, to reveal new insights related to transparency. However, some of the evidence was collected from answers to the questions related to xAI and what explanations were required. In their responses, the participants referred to understanding and transparency as part of those explanations. Hence, in part, the section conclusion refers back to explanations and xAI, as relates to the theme.

5.4.3.1 Evidence of *xAI and Transparency*.

Table 24:

Evidence of xAI and Transparency

RQ3: Theme 3 – xAI and Transparency
With a focus on the topics "*Understanding of "How?"*" and "*Understanding and Trust*".
Participant 1: *"Organizations rightfully don't deploy because they don't trust necessarily the results, because they don't know how it's derived."*
Participant 1: *"So you want to understand how those models are actually generating their information, right? […] Know what question you want to ask and have some understanding of how the platform's actually doing it."*
Participant 14: *"And for us it really is, is exactly about that if you don't know what informs what comes out of it, you just don't know, like, how are you going to explain it? How are you going to have the trust that it's reliable?"*
Participant 1: *"And then you generally, what we find is that you've got the those who don't know enough about it to trust it, and they're making a good decision, because you shouldn't do anything that you don't fully understand."*
Participant 1: *"[For Trust] Everyone is basically required to understand the at least the basics, the intermediate of how AI works, and what the limitations are, the ethical challenges, and I think that the executives, the executive layer in our business, [should be] all exquisitely well informed on it, okay?"*
Participant 10: *"I have no human I have no history, I have no shared trust. I have no shared work pattern or ethic or understanding of how it works. Maybe it's an understanding thing."*
Participant 10: *"[For Trust] For me, it's understanding again, the source of the information, and understanding how it manipulates the data, how it runs through it, what the algorithm is, […] getting a broad-based understanding of what logic is used there."*
Participant 14: *"If I'm using AI, then it better be, you know, relevant to me and applicable to me, you know, yeah, yeah. So I think that's part of it, of needing to understand that, and that builds trust."*
Participant 3: *"My personal view is, no way would I trust it okay, because what's informing that decision? Have you given it all the right parameters and the right information?"*
Participant 16: *"The actual understanding of the how it actually goes and takes all this data and puts it into a model, right? That was quite important. Otherwise, you wouldn't have trusted it."*
Participant 9: *"With understanding comes trust. So how the model goes about interrogating data, framing it and manipulating it, and then giving us some output, I think that understanding of how that works would be helpful in gaining trust and the ability to adopt it."*
Participant 7: *"I think we need to see more of it and understand it. Like, for me, there's a lot of question marks behind the tech behind this, and there's no knowledge of how it works."*
Participant 15: *"So I need to understand what's being fed into the model, […] So the transparency in terms of how it's been built and what information is going in there, and how the information is being used and manipulated by the AI as well, to make those decisions is very important in trusting the AI, otherwise there's going to be no trust."*
Participant 19: *"But when we started giving them the outcomes that the model has given and say, Okay, these are the factors that the model are considering and saying, […] Then, you know, internally they started having those aha moments as well. But like, actually it makes sense, because now it's starting to make sense."*

Participant 6: *"Why is it giving something different? And you need to understand why. And then if you looked in and understood, well, okay, right? It's actually in the last few months, it's learned that it can get faster, it doesn't need as much data, or it's got better quality data, and then you could go, okay, so it's logical that the version two is offering us the right solution, and version one was a less accurate solution."*
Participant 7: *"Developers and stuff sit with our business and understand what our requirements and nuances and issues are, yeah and sort of answer for those things that would make me feel a lot more trust."*
Participant 8: *"But I think the guys that understand the technology would trust it. The guys that don't yet, they wouldn't trust results."*
Participant 15: *"Um, as people maybe understand more how AI learns and how it how it makes its decisions and based on what criteria it evaluates and makes the decisions […], the trust will increase, yes."*

Note: Author's Own.

5.4.3.2 In-Case Analysis of the Evidence.

In the Executive Decision-Making group there was a similar emphasis on understanding how AI operates as important for building trust. Firstly, Participant 1 expressed that organizations hesitate to deploy AI because they lack explanation and understanding of how AI models generate results. Secondly, they highlighted that trust is tied to knowledge of the underlying mechanism, or the AI's logic. Thirdly, they emphasized that a basic or intermediate understanding of AI's operations is necessary for executives to build trust. Therefore, explanation of both would be useful. Participant 10 reinforced this view, pointing out the need to comprehend the source of AI information, the algorithm, and the logic used. Similarly, Participant 14 noted that without understanding how AI produces its outcomes, it becomes difficult to explain it and to trust its reliability. This group similarly showed that explanations to provide understanding of AI's internal processes is important for engendering trust, and this transparency provides the necessary context for informed decision-making.

In the Senior Decision-Making group, participants similarly highlighted the relationship between understanding AI's workings and trust. Participant 3 showed scepticism, questioning the trustworthiness of AI decisions if the parameters and information feeding the model are unclear. Participant 9 elaborated that understanding "how" AI models manipulate, and frame data is important to gaining trust and improving adoption. Participant 16 similarly emphasized that understanding how AI processes data is "quite important" to establishing trust. This group similarly highlighted that for trust to develop, explanation and transparency about "how" AI functions must be provided, would allow them to verify and comprehend its outputs.

The Operational Decision-Making group stressed the importance of practical understanding and transparency for trust. Participant 15 highlighted that understanding the data fed into AI models, and "how" these inputs are used, is important for building confidence in the AI's outputs. Participant 7 mentioned that having developers understand business requirements and nuances, and to the have the developers explain how AI processes work to

them would enhance trust. Participant 6 pointed out that observing AI's learning and performance improvements over time helps in understanding and trusting the system. Lastly, both Participants 8 and 19 similarly added that users who are familiar with and understand AI's technology are more likely to trust it. This group suggests the need for explanations that are understandable, practical and aligned with business requirements to develop trust.

5.4.3.3 Cross-Case Analysis of the Evidence.

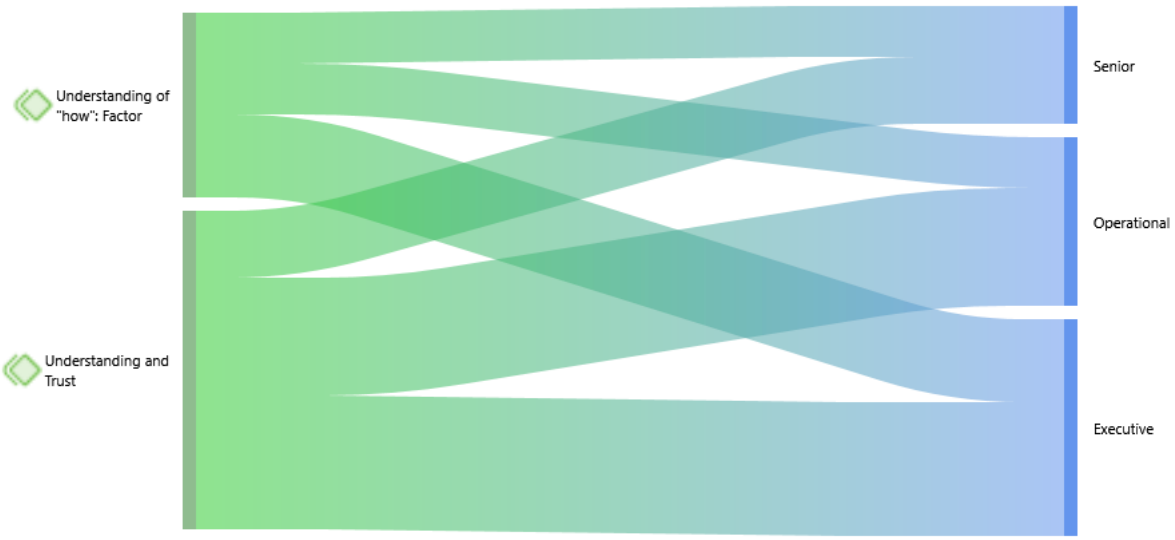
All groups shared a similar view that explanations to understand "how" AI works is important for building trust, but they focused on slightly different dimensions. The Executive Decision-Making group prioritized explanation to understand the technical aspects, such as algorithms and logic, seeing transparency as important for explaining decision-making. The Senior Decision-Making group similarly emphasized the need for transparency but focused on explanation to understand how AI manipulates and frames data. The Operational Decision-Making group highlighted the importance of practical understanding and transparency that aligns with business requirements and emphasised continuous engagement over time to maintain trust. While all groups connected explanation to understand "how" AI works to create trust, they varied slightly in their focus.

5.4.3.4 Distribution of the Main Theme by Participant Group.

The "Sankey" diagram below provides a visual representation of the distribution and flow of the sub-themes across different participant groups. The diagram is prepared from the total Atlas.ti code set related to the two sub-themes and provides a visual overview of the flow of discussion and density of quotations from the participant data.

Figure 24:

Sub-Theme Distribution: "Understanding of "How?"" and "Understanding and Trust"



Note: Author's own, generated from primary data in Atlas.ti software.

Understanding of "How?": The majority of discussion for this theme was found in the Executive Decision-Making group's interviews, followed by the Senior and Operational Decision-Making groups in a lower, but equal proportion. The sub-theme was notably discussed with all three decision-making groups, indicating their awareness of the topic, but was most prevalent in the Executive group.

Understanding and Trust: This sub-theme was mostly discussed with the Executive and Operational Decision-Making groups, with an almost equal prevalence between them. Again, it was interesting to note that the sub-theme was discussed across all three decision-making groups, again showing broad awareness. However, there was less prevalence in the Senior Decision-Making group.

5.4.3.5 Conclusion on *xAI and Transparency*.

There was a similarity of recognition across all participant groups that transparency and explanation to understand "how" AI works are important for building trust. All groups similarly agreed that transparency and understanding play an important role in fostering confidence in AI's decision-making. Furthermore, the groups showed similarity that explanation to understand how AI processes data, its algorithms, and the logic behind its outputs are important factors for trust. It would appear that these factors then enable informed decision-making and promote trust across organizational levels. The shared emphasis on transparency and understanding highlighted the important connection between explanation to understand AI's internal workings and developing trust in its use.

There were slight differences in focus among the groups. Some prioritized technical clarity, focusing on understanding the algorithms and logic behind AI, to explain AI decision-making. Others emphasized the manipulation and framing of data, with transparency seen as necessary for assessment of AI's processes and outputs. The remainder focused on practical understanding that aligned with business requirements and highlighted the need for continuous observation over time. These variations suggest that for xAI to be effective, it must provide clear and transparent explanations tailored to the needs and priorities of each group to create understanding.

Key Concepts Identified: xAI and Transparency drive understanding, which promotes trust in AI decision-making. Understanding "How" AI works and processes data is important for building trust.

5.4.4 RQ3: Theme 4 – Conjoined Agency

This theme is a potential new main theme and is a distinct difference (on first analysis) from the Chapter 2 literature review and is reflected in Figure 12 at the start of Chapter 5. The evidence gathered from the participants provides a rich set of insights. The use of colours

connects the evidence to the respective participant groups, while the evidence itself is selected for discussion of the main topics shown in table 21. The two main topics discussed are "*Human-AI Conjoined Agency*", and "*AI Assistance of Humans*".

5.4.4.1 Evidence of *Conjoined Agency*.

Table 25:

Evidence of Conjoined Agency

RQ3: Theme 4 – Conjoined Agency
With a focus on the topics "*Human-AI Conjoined Agency*", and "*AI Assistance of Humans*"
Participant 1: *"And it was great, to sort of actually say to [AI], put together a very short summary, a 10 bullet point numbered summary, and then the interesting points... Please give me more information on bullet four, etc, and this helped me to drive my decisions around how I had the conversation with the client in their workshop."*
Participant 13: *"So I'd say, Yeah, AI should be leveraged as a tool and not a replacement for, you know, human smarts. I think for me, I trust it to a point."*
Participant 14: *"And there it is, really to assist, you know, the human talent you know, to it's used mainly, I guess, for efficiency, and not only efficiency, you know, also in terms of sort of the decision-making algorithms."*
Participant 17: *"I'm still leaning towards this thing must be constrained to, like, effectively, an employee number. It must have limited sort of mandates. Uh, yes, it could be a super employee type mandates, but it must be limited. And I think it's almost just to curb the runaway effect."*
Participant 2: *"It would be, you know, a companion where you would say, give me the top three highlights of this year's results, and it suggests to me four different levers that we can pull as a company to improve or get, you know, better results out of it, and whatever it tells you would almost be, in a way, taken as that's what we need to do."*
Participant 3: *"So it might be to get part way to solve a problem, you directing it, maybe on data gathering or trying to streamline the analysis, but you still going to have to have a person to fact check it, it would be doing the grunt work."*
Participant 4: *"And I think that it's kind of co-pilot for what it means to be a manager. [...] So I'm using AI more, some from a personal level. I use it more for, report-writing, emails, synthesizing information, and I've got quite a high level of trust for that."*
Participant 12: *"I have just used "Co-Pilot" [Google Product] a couple of times myself, but really it's tools that are then not to replace people, but to help and elevate and remove redundancies."*
Participant 12: *"Still requires human intervention, and it requires human skill to sift out what is true and what is not true. Because, chatGPT, for example requires human skill to understand the prompts and to get quality data."*
Participant 5: *"I honestly think I don't think it's going to be like when we introduced spreadsheets. A lot of people said, Oh, you're going to take away jobs. But it didn't take away jobs. It enhanced what people were doing in their job function."*
Participant 6: *"I think it's narrow stuff that could support activities that enable a more effective, efficient use of resources. So I think they're open to stuff along the lines of, can it take over someone being a scribe in a meeting? Can it analyse huge amounts of data?"*
Participant 7: *"So when they're making a sale, we've got AI tools like [Tool X], which sort of go through the documents. You train it to look for certain clauses or certain consents or certain provisions that would be triggered on a sale. So it speeds up our process, indeed."*
Participant 15: *"There might be a higher level of trust, but where there are legacy systems and manual processes involved and you're supplementing it with AI, you can't, I think, wholly trust AI to make your decisions with your AI says."*

Participant 19: *"So how do we then actually use AI to rather augment and not replace certain tasks and functions within the business? I think that would also help to start building trust in AI."*
Participant 19: *"Um so one of our first use cases was a model that we used to actually assist us in fraud detection."*

Note: Author's Own.

5.4.4.2 In-Case Analysis of the Evidence.

In the Executive Decision-Making group, there was a focus on leveraging AI as a tool to assist, rather than to replace, human capabilities. Participant 1 described using AI to create summaries and to gather information to support decision-making. This demonstrated how AI can enhance human agency. Participant 13 expressed trust in AI to a certain extent, as long as it remains a supportive tool rather than a replacement. Participant 14 emphasized AI's role in improving efficiency and assisting with decision-making algorithms, while Participant 17 noted that AI should be constrained within limited mandates, functioning like a "super employee" with a defined role. This group similarly highlighted that AI, when used as an assistant, enhances human capabilities and supports decision-making, which in turn helps to build trust.

In the Senior Decision-Making group, the emphasis was on the concept of AI as a "co-pilot" or companion in the decision-making process. Participant 2 described AI as a tool that provides actionable insights, which humans can use to guide business strategies. Participant 3 pointed out that while AI can gather and streamline data, human oversight is necessary for fact-checking AI's outputs. Participant 4 referred to AI as a "co-pilot", noting that it can assist in tasks like report-writing and information synthesis, and that they have high trust in the AI's ability for that. Participant 12 echoed this by highlighting AI's role in removing redundancies while still requiring human intervention for effective use. This group showed that AI's effectiveness and trustworthiness increase when it operates alongside humans, augmenting their decision-making abilities rather than replacing them.

In the Operational Decision-Making group, the focus was on using AI to enhance efficiency and streamline processes. Participant 5 likened AI's impact to that of spreadsheets, which enhanced human tasks rather than replacing them. Participant 6 discussed AI's potential in supporting narrow tasks, like data analysis or taking meeting notes, to optimize resources. Participant 7 mentioned AI tools that automate document review, speeding up business processes. Participant 19 also reinforced that AI should be used to augment rather than replace functions, citing an example of AI assisting in fraud detection. This group emphasized that AI is trusted more when it is applied as a supportive tool that integrates into existing human workflows.

5.4.4.3 Cross-Case Analysis of the Evidence.

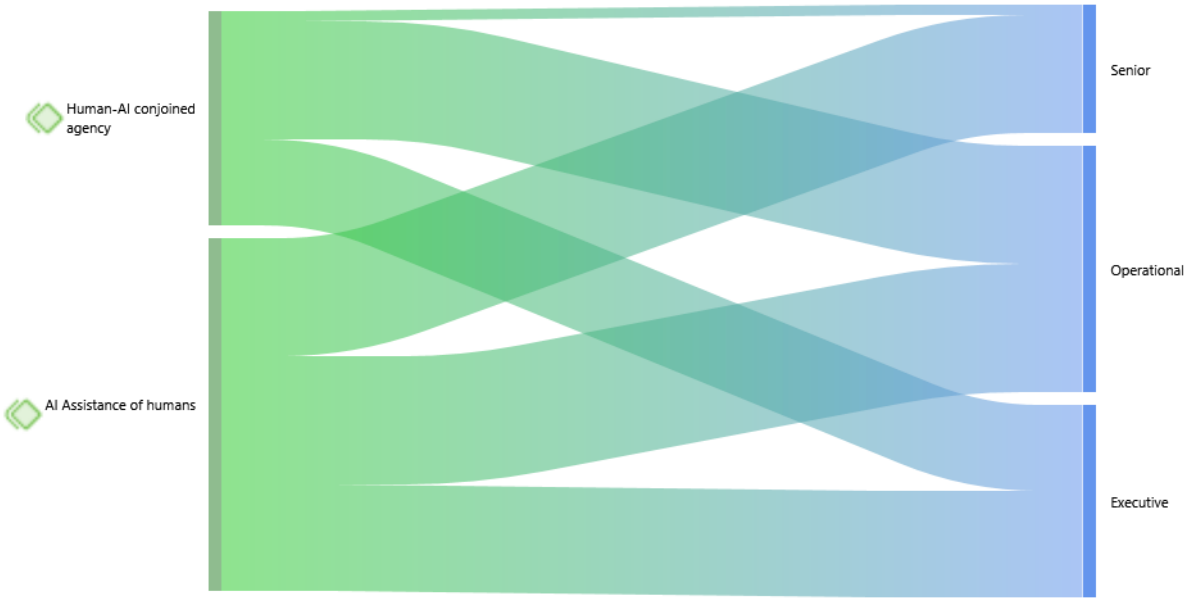
All groups similarly highlighted the importance of AI as a supportive tool that enhances human capabilities, but their emphasis on its application and integration differed slightly. The Executive Decision-Making group focused on using AI to enhance efficiency and aid decision-making processes, while maintaining human control over its functions. The Senior Decision-Making group emphasized the concept of AI as a "co-pilot" that works in tandem with humans. This provides guidance and support but requires human intervention for oversight. The Operational Decision-Making group focused on AI's role in optimizing processes and performing supportive tasks. They stressed that trust is built when AI augments rather than replaces human roles. Across all groups, there was a similar view that AI could assist humans to improve decision-making and efficiency, and that trust is built through collaboration and conjoined agency.

5.4.4.4 Distribution of the Main Theme by Participant Group.

The "Sankey" diagram below provides a visual representation of the distribution and flow of the sub-themes across different participant groups. The diagram is prepared from the total Atlas.ti code set related to the two sub-themes and provides a visual overview of the flow of discussion and density of quotations from the participant data.

Figure 25:

Sub-Theme Distribution: "Human-AI Conjoined Agency" and "AI Assistance of Humans"



Note: Author's own, generated from primary data in Atlas.ti software.

Human-AI Conjoined Agency: The majority of discussion for this theme was found in the Operational Decision-Making group's interviews, followed by the Executive Decision-Making group. There was little awareness or emphasis noted in the participant interviews from

the Senior Decision-Making group. The sub-theme was, however, still discussed with all three decision-making groups.

AI Assistance of Humans: This sub-theme was discussed evenly across all participant groups, showing a broad awareness and also highlighting their exposure to some form of AI assistance in their roles.

5.4.4.5 Conclusion on *Conjoined Agency*.

"Human-AI Conjoined Agency" and "AI Assistance of Humans" are closely related themes that emphasize trust in AI through a collaborative, supportive role. All participant groups shared similar views that AI is most effective and trustworthy when it enhances human decision-making, while remaining under human oversight and control. This shared agency reinforced their similar view that AI should serve as a tool to augment human capabilities, not to replace them, which fosters trust and confidence in AI's use. Across all groups, the collaborative approach of AI functioning as an assistant, rather than a replacement, emerged as central to building trust in AI-driven decision-making.

Again, there was a slight difference in how each group perceived this collaboration with AI. Some focused on AI's role in improving efficiency and supporting decision-making processes. However, they also advocated maintaining strict human control. Others highlighted the idea of AI as a "co-pilot," working in tandem with humans, where AI assists but requires human oversight for accuracy and accountability. The remainder emphasized AI's practical application in streamlining tasks and optimizing workflows, stressing that trust is built when AI integrates into existing human processes without threatening job roles. These differences highlighted the importance of conjoined agency in reinforcing AI-human, collaborative partnership across organizational levels.

Key Concepts Identified: AI is most effective and trusted as a supportive tool to enhance decision-making, which must be overseen by humans.

5.4.5 Summary of Similarities and Differences in the Findings for RQ3

Table 26 below is provided as a summary and convenient reference for the findings related to the sub-themes discussed under the main themes of RQ3. The table is prepared from the in-case and cross-case analyses and conclusions for each of the four themes presented in Section 5.4.

Table 26:

Summary of Similarities and Differences in the Findings, Across Themes, for RQ3.

RQ3: *How does transparency influence emotional trust in AI for organisational decision-making?*		
	Similarities	**Differences**
Theme 1: Transparency for Trust in AI	• Shared view that fear is a barrier to adoption. • Fear is related to job security. • Transparency is required to alleviate anxiety and foster trust. • AI has limited ability to identify human nuance, which inhibits trust.	• Difference in focus between groups, with some focusing on industry-specific fears. Others highlighted relationship and risk-based fears. Others identified AI's inability to match human insight and personality is a barrier.
Theme 2: Transparency for Decision-Making	• Transparency and Understanding are closely related, directly linked to trust in AI for decision-making. • Without transparency, trust cannot be established.	• While all groups value transparency, each group prioritizes it in slightly different ways depending on their role within the organization. • Transparency to understand AI functionality, leading to trust. • Transparency for clarity and risk assessment. Must lead to actionable insights. • Transparency comprises practical explanations of AI processes.
Theme 3: xAI and Transparency	• xAI and Transparency drive understanding, which promotes trust in AI decision-making. • Understanding "How" AI works is important for building trust. • Understanding how data is processed is important for trust.	• Differences in focus indicate that for xAI to be effective, it must provide clear and transparent explanations tailored to the needs and priorities of each group to create understanding.
Theme 4: Conjoined Agency	• AI most effective and trusted as a supportive tool to enhance decision-making. • AI should support humans but not replace them. • Decision-making must be overseen by humans.	• Difference in focus between groups, with some focusing on enhanced efficiency and decision-making support. Others highlighted AI as a "co-pilot, working in tandem with humans. The remainder focused on practical tasks like workflow optimisation.

Note: Author's Own.

5.4.6 Conclusion to the Findings for RQ3: Summary of Key Concepts

In conclusion to the findings for RQ3, the similarities and differences from each of the discussed themes and their conclusions were reviewed, to derive the "Key Concepts" which emerged from the theme, topic and sub-theme findings. These key concepts are derived from both the similarities and differences which emerged and are presented in Table 27.

Table 27:

Summary of Key Concepts from the Findings, by Theme for RQ3.

RQ3: *How does transparency influence emotional trust in AI for organisational decision-making?*		
	Topics / Sub-Themes	**Key Concepts Identified**
Theme 1: Transparency for Trust in AI	• *Fear as a Barrier* • *AI Identification of Human Nuance*	• Fear is a barrier to adoption and is primarily related to job security. • Transparency is required to alleviate fear and anxiety and foster trust. • AI has limited ability to identify human nuance, which inhibits trust.
Theme 2: Transparency for Decision-Making	• *Transparency and Understanding*	• Transparency and Understanding are closely related, directly linked to trust in AI for decision-making. • Different priorities for transparency, depending on the role of the individual responsible for decision-making within the organization.
Theme 3: xAI and Transparency	• *Understanding of "How?"* • *Understanding and Trust*	• xAI and Transparency drive understanding, which promotes trust in AI decision-making. • Understanding "How" AI works and processes data is important for building trust.
Theme 4: Conjoined Agency	• *Human-AI Conjoined Agency* • *AI Assistance of Humans*	• AI is most effective and trusted as a supportive tool to enhance decision-making, which must be overseen by humans.

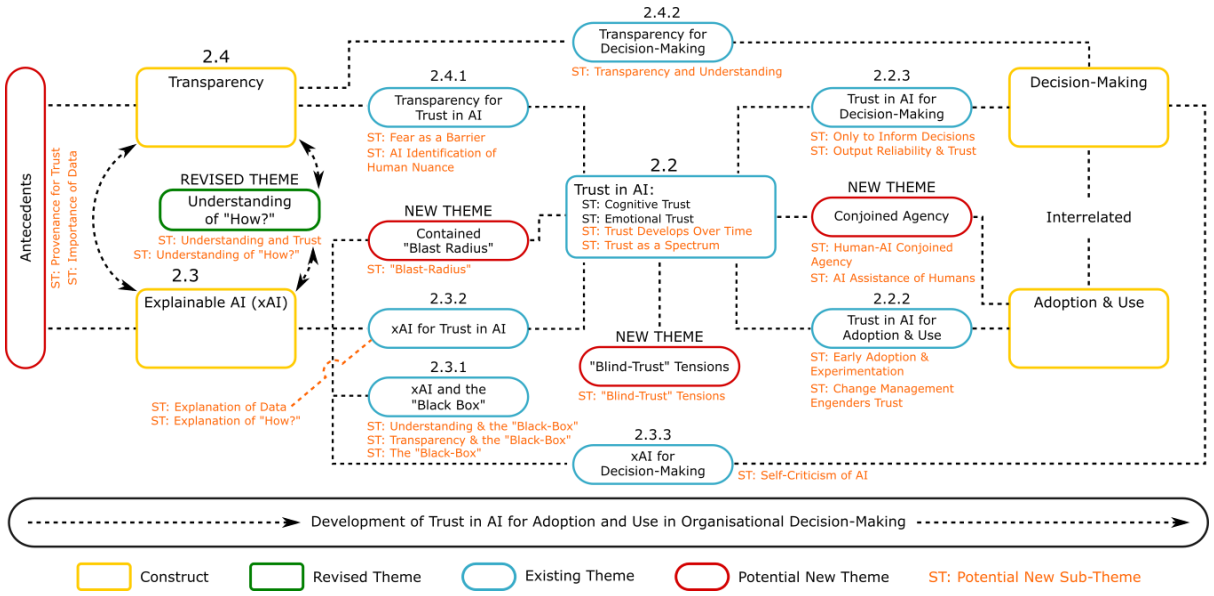
Note: Author's Own.

5.5 Revised Conceptual Framework Following Discussion of Findings

Having discussed and presented the findings through Chapter 5, another revision to the Conceptual Framework is provided in Figure 26 below. Also now included in this framework are the 23 potential new sub-themes which were discussed throughout the chapter for each Research Question. This revised Conceptual Framework will serve as a reference for Chapter 6, where each potential new theme and sub-theme is analysed against the extant literature to determine potential contribution. A point to note in the diagram is the revision of one of the original themes “2.5 xAI and Transparency” to “Understanding of “How?””. This is depicted in green and will be analysed in Chapter 6.

Figure 26

Revised Conceptual Framework Showing Potential New Themes and Sub-Themes.



Note: Author’s Own

5.6 Conclusion to the Findings: Summary of Key Concepts

In conclusion to Chapter 5, the similarities and differences from each of the discussed themes and their conclusions were reviewed, to derive the "Key Concepts" which emerged from the theme findings. These key concepts are derived from both the similarities and differences which emerged, and are presented in Table 28 below, for comparison to literature in the Chapter 6 discussion.

Table 28:

Summary of Key Concepts from the Findings, by Theme.

RQ1: *How does trust in AI lead to its adoption and use for organisational decision-making?*		
	Topics / Sub-Themes	**Key Concepts Identified**
Theme 1: Trust in Artificial Intelligence	• *Trust Develops over Time* • *Trust as a Spectrum*	• Trust takes time to develop. • Trust exists as a spectrum or on a scale. • Trust develops at different rates in different industries.
Theme 2: Trust in Artificial Intelligence for Adoption and Use	• *Early Adoption & Experimentation* • *Change Management Engenders Trust*	• Importance of gradual AI adoption and experimentation • Importance of change management for trust in AI.
Theme 3: Trust in Artificial Intelligence for Decision-Making	• *Only to Inform Decisions* • *Output Reliability & Trust*	• AI should inform and enhance human decision-making, not replace it. • Human verification of AI outputs is needed.
Theme 4: "Blind-Trust" Tensions	• *"Blind-Trust" Tensions*	• There is a tension between the Decision-Making groups regarding "blindly-trusting" AI. Some support it while others reject it.
Theme 5: "Blast-Radius"	• *"Blast-Radius"*	• Importance of containing the potential negative outcomes of AI use for decision-making, to engender trust.
RQ2: *How does xAI engender trust in AI for decision-making in organisations? And: What are the factors that contribute to trust in AI?*		
	Topics / Sub-Themes	**Key Concepts Identified**
Theme 1: xAI and the "Black-Box"	• *The "Black-Box"* • *Understanding and the "Black-Box"* • *Transparency and the "Black-Box"*	• Awareness and shared concern of the "Black -Box" nature of AI. • Transparency, understanding and explainability were all recognised to build trust and overcome the "Black-Box".
Theme 2: xAI for Trust in AI	• *Explanation of Data* • *Explanation of "How?"*	• Explainability should address "How" AI arrives at outcomes. • Explanation must be consistent throughout the data lifecycle.
Theme 3: xAI for Decision-Making	• *Self-Criticism of AI*	• AI Self-Criticism is necessary for trust and confidence in AI systems for decision-making.
Theme 4: Antecedents	• *Provenance for Trust* • *Importance of Data*	• Provenance is an antecedent to adoption. • Data integrity is an antecedent to adoption.
RQ3: *How does transparency influence emotional trust in AI for organisational decision-making?*		
	Topics / Sub-Themes	**Key Concepts Identified**
Theme 1: Transparency for Trust in AI	• *Fear as a Barrier* • *AI Identification of Human Nuance*	• Fear is a barrier to adoption and is primarily related to job security. • Transparency is required to alleviate fear and anxiety and foster trust. • AI has limited ability to identify human nuance, which inhibits trust.

Theme 2: Transparency for Decision-Making	• *Transparency and Understanding*	• Transparency and Understanding are closely related, directly linked to trust in AI for decision-making. • Different priorities for transparency, depending on the role of the individual responsible for decision-making within the organization.
Theme 3: xAI and Transparency	• *Understanding of "How?"* • *Understanding and Trust*	• xAI and Transparency drive understanding, which promotes trust in AI decision-making. • Understanding "How" AI works and processes data is important for building trust.
Theme 4: Conjoined Agency	• *Human-AI Conjoined Agency* • *AI Assistance of Humans*	• AI is most effective and trusted as a supportive tool to enhance decision-making, which must be overseen by humans.

Note: Author's own.

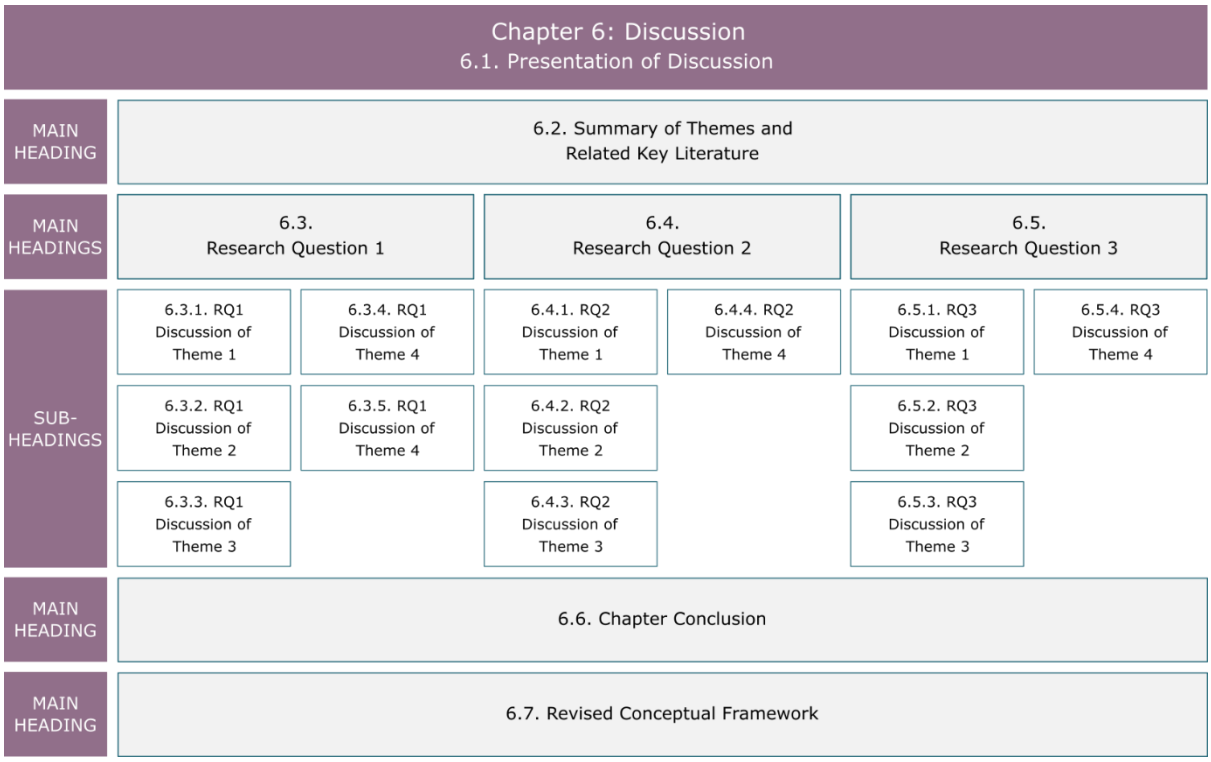
Chapter 6: Discussion

6.1 Presentation of Discussion

This section presents the discussion of the research findings from Chapter 5 and compares them to the Literature from Chapter 2. As per the structure of Chapter 5, the discussion is presented by Research Question, with a focus on the themes and sub-themes presented therein. An outline of the sections and sub-sections of the Chapter 6 discussion which follows, is provided in Figure 27 below.

Figure 27:

Matrix of Sections and Sub-Sections for Chapter 6



Note: Author's Own.

This chapter contributes an additional level of analysis, to determine whether the findings are present in existing literature and/or whether they reveal any new insights and understanding. Each of the 13 themes is discussed and compared to literature using a structured and systematic approach. The comparison first considers the similarities between the findings and the literature. Where similarities are found, these are noted in the relevant conclusion, and the theme / sub-theme is retained as supporting extant literature, and as an expected outcome. Where differences are found, three main steps are followed to systematically draw a conclusion as to any potential contribution. The three steps taken are explained as follows:

6.1.1 Description of the 3-Step Process to Identify Potential Contributions

Step 1: A targeted keyword search is performed, for each of the key articles (and their scholars) in Chapter 2, which were cited for the related Research Question. The keywords searched are related to the key concepts in the theme or sub-theme identified as a potential difference. Where no matches are found, the next step is applied.

Step 2: A broadening of the above targeted keyword search to three pieces of additional literature presented in Chapter 2 is performed. Similarly, the keywords searched are related to the key concepts in the theme or sub-theme which was identified as a potential difference. In the absence of any matches being found, the next step is applied.

Step 3: In this step, a Google Scholar search is performed for any new articles by up to three top scholars from the related research question over the past two years. If any recent publications related to the topic are identified, a keyword search of that article is performed, and the result is analysed for any similarity or difference to the findings.

Only after having completed the above three steps, with no matches identified, is the theme / sub-theme considered to indicate a potential difference and therefore a potential contribution. This process is completed for each theme and its sub-themes. This systematic approach, although not fully exhaustive, was chosen to provide a consistent review and comparison of the findings to the literature. This chapter once again concludes with a further revision to the previous Conceptual Framework given in Figure 26, at the end of Chapter 5. The newly revised conceptual framework identifies which of the potential new themes and sub-themes from Chapter 5 survived the checks against extant literature and which remain as potential contributions to be presented in the Chapter 7 conclusion.

6.2 Summary of Themes and Related Key Literature

Table 29 provides an overview of the key focus areas for this chapter. This is designed to assist the reader in following the logic for the comparison Steps 1-3 as highlighted in Section 6.1.1 above. The table presents the key scholars cited for each research question, to be used in Step 1 as well as additional scholars per theme, for the potential Step-2 search. In addition, the table also presents the primary top scholars applicable to the research question for ease of reference, should the search of new articles by them be required as outlined for Step 3.

Table 29:

Overview of Research Questions, Themes and Related Key Literature for Steps 1-3

RQ1: *How does trust in AI lead to its adoption and use for organisational decision-making?*		
Scholars Cited in RQ1 - For Step 1 (where required)		
(Vanneste & Puranam, 2024); (Sullivan et al., 2022); (Wong et al., 2024); (Enholm et al., 2021)		
Themes from RQ1	**Related Key Literature for Step 2 (where required)**	**Top Scholars for Step 3 (where required)**
Theme 1: Trust in Artificial Intelligence	• (Glikson and Woolley, 2020); (Chen et al., 2021); (Wang & Ding, 2024); (Weber et al., 2023);	• Phanish Puranam; Ella Glikson; Anita Woolley
Theme 2: Trust in Artificial Intelligence for Adoption and Use	• (Wang & Ding, 2024); (Weber et al., 2023); (Glikson and Woolley, 2020)	• Phanish Puranam; Ella Glikson; Anita Woolley
Theme 3: Trust in Artificial Intelligence for Decision-Making	• (Chen et al., 2021); (Wang & Ding, 2024); (Sabharwal et al., 2024); (Shrestha et al., 2021)	• Phanish Puranam; Ella Glikson; Anita Woolley
Theme 4: "Blind-Trust" Tensions	• To be identified through Steps 1-3	• Phanish Puranam; Ella Glikson; Anita Woolley
Theme 5: "Blast-Radius"	• To be identified through Steps 1-3	• Phanish Puranam; Ella Glikson; Anita Woolley
RQ2: *How does xAI engender trust in AI for decision-making in organisations? And: What are the factors that contribute to trust in AI?*		
Scholars Cited in RQ2 - For Step 1 (where required)		
(Rai, 2020); (Sabharwal et al., 2024); (Lukyanenko et al., 2022)		
Themes from RQ2	**Related Key Literature for Step 2 (where required)**	**Top Scholars for Step 3 (where required)**
Theme 1: xAI and the “Black-Box”	• (Berente et al., 2021); (Weber et al., 2023); (Kim et al., 2023); (Choung et al., 2022)	• Arun Rai; Nicholas Berente; Roman Lukyanenko
Theme 2: xAI for Trust in AI	• (Berente et al., 2021); (Silva et al., 2023); (Wang & Ding, 2024); (Abedin, 2022)	• Arun Rai; Nicholas Berente; Roman Lukyanenko
Theme 3: xAI for Decision-Making	• (Rabiee et al., 2024); (Vanneste & Puranam, 2024); (Wang & Ding, 2024)	• Arun Rai; Nicholas Berente; Roman Lukyanenko
Theme 4: Antecedents	• To be identified through Steps 1-3	• Arun Rai; Nicholas Berente; Roman Lukyanenko
RQ3: *How does transparency influence emotional trust in AI for organisational decision-making?*		
Scholars Cited in RQ3 - For Step 1 (where required)		
(Glikson & Woolley, 2020); (Wang & Ding, 2024); (Wang et al., 2016); (Berente et al., 2021)		
Themes from RQ3	**Related Key Literature for Step 2 (where required)**	**Top Scholars for Step 3 (where required)**
Theme 1: Transparency for Trust in AI	• (Laato et al., 2022); (Haque et al., 2023); (Choung et al., 2022)	• Anita Woolley; Samuli Laato; Nicholas Berente
Theme 2: Transparency for Decision-Making	• (Kim et al., 2023); (Lindebaum et al., 2020); (Vanneste & Puranam, 2024)	• Anita Woolley; Samuli Laato; Nicholas Berente
Theme 3: xAI and Transparency	• (Rai, 2020); (Laato et al., 2022); (Haque et al., 2023)	• Anita Woolley; Samuli Laato; Nicholas Berente
Theme 4: Conjoined Agency	• To be identified through Steps 1-3	• Anita Woolley; Samuli Laato; Nicholas Berente

Note: Author’s Own.

6.3 Research Question 1: *How does trust in AI lead to its adoption and use for organisational decision-making?*

The aims of RQ1 were to develop new insights and understanding into the mechanisms of trust formation in AI in organisations, and to therefore develop new insights as to how trust in AI leads to adoption and beneficial use of AI for organisational decision-making. There are three themes which were mapped directly to RQ1, with two potential new themes, all of which are discussed in this section. Table 30 below provides a summary of the themes, topics / sub-themes and key concepts from the Chapter 5 findings which will be discussed under RQ1.

Table 30:

Themes Emerging from RQ1 and their Key Concepts Identified in the Findings.

RQ1: *How does trust in AI lead to its adoption and use for organisational decision-making?*		
Themes from RQ1	**Topics / Sub-Themes**	**Key Concepts Identified**
Theme 1: Trust in Artificial Intelligence	• *Trust Develops over Time* • *Trust as a Spectrum*	• Trust takes time to develop. • Trust exists as a spectrum or on a scale. • Trust develops at different rates in different industries.
Theme 2: Trust in Artificial Intelligence for Adoption and Use	• *Early Adoption & Experimentation* • *Change Management Engenders Trust*	• Importance of gradual AI adoption and experimentation • Importance of change management for trust in AI.
Theme 3: Trust in Artificial Intelligence for Decision-Making	• *Only to Inform Decisions* • *Output Reliability & Trust*	• AI should inform and enhance human decision-making, not replace it. • Human verification of AI outputs is needed.
Theme 4: "Blind-Trust" Tensions	• *"Blind-Trust" Tensions*	• There is a tension between the Decision-Making groups regarding "blindly-trusting" AI. Some support it while others reject it.
Theme 5: "Blast-Radius"	• *"Blast-Radius"*	• Importance of containing the potential negative outcomes of AI use for decision-making, to engender trust.

Note: Author's Own.

6.3.1 RQ1: Discussion of Theme 1 – Trust in Artificial Intelligence

This theme is a main theme from the Chapter 2 literature review and is reflected in Figure 12 at the start of Chapter 5. The two main topics discussed in the findings were "*Trust Develops Over Time*" and "*Trust as a Spectrum*". The key concepts which emerged from these main topics are compared to the literature in the sections which follow, and conclusions are made as to their potential contribution.

6.3.1.1 Recap of the findings on Trust in Artificial Intelligence.

The findings on "Trust in Artificial Intelligence" across participant groups highlighted several key concepts. First, there was a shared understanding that trust in AI is not a single-value concept but exists on a spectrum or scale. All the participant groups recognised this concept of a multi-faceted nature of trust. While they used slightly different terminology, the shared recognition of trust being more than just binary was clear.. Another key insight was

that trust develops over time, with the participants consistently emphasizing that trust in AI is a gradual process that evolves, rather than being something granted immediately. A third finding was from a difference in views between groups, that trust in AI might develop at different rates across different industries. This insight into differing industry timelines was a unique observation, signalling a potential industry-dependent relationship between time and trust formation.

6.3.1.2 Recap of the literature on Trust in Artificial Intelligence.

The literature on trust in AI showed a well-established scholarly foundation in both cognitive and emotion-based components of trust, which has developed over the last 30 years. Comparison of McAllister (1995), Hoff & Bashir (2015), and Glikson & Woolley (2020), revealed a progression from inter-human trust to trust in technology and, more recently, trust in AI. This evolution also reflected early connections between trust, decision-making, and explanation, which contributed to the contemporary discourse on trust in AI. The systematic literature review by Glikson & Woolley, (2020) explored the dynamic nature of trust development over time, for different AI representations, namely: robotic AI, embedded AI, and virtual AI. However, the emotional aspect of trust and its relationship to trust in AI had not been answered since identification of the gap by Wang et al. (2016) or the subsequent echo by Glikson & Woolley (2020). Exploring the above revealed important foundational links between trust, organisational behaviour, and decision-making, which also become relevant in the subsequent research questions and their associated themes.

6.3.1.3 Discussion of the findings and literature.

Both the findings and the literature similarly identified trust development over time as an important concept. Glikson & Woolley, (2020) described trust as either increasing or decreasing over time, based on the type of AI concerned (Robotic, Virtual or Embedded). While the participants were commenting more in the domain of virtual and embedded AI they did not offer insight as to trust either increasing or decreasing, but simply supporting literature's insight that it develops over time. Comparing the finding that *trust in AI exists on a spectrum or scale* had some similarity to literature's description of the cognitive and emotional aspects of trust (Glikson & Woolley, 2020), however this was different to the key concept outlined in the findings, which indicated a multi-variate trust, beyond just cognitive and emotional. The finding was therefore retained as a difference and will be analysed further using the 3-step process outlined in Section 6.1.1. Lastly, the concept of trust developing at different rates in different industries was not found in Chapter 2 literature and was also retained for further analysis.

6.3.1.4 Further discussion of the differences between findings and literature.

STEP 1: The articles which were identified for use in Step 1 were those cited in the related research question, as highlighted in Table 24: Vanneste & Puranam, (2024), Sullivan et al. (2022), Wong et al. (2024) and Enholm et al. (2021). The keywords used across both remaining key concepts were “trust spectrum”, “trust scale”, “AI trust rates”, “AI trust in industries” and “industry trust”. The word searches on the four selected articles did not yield any matches, therefore Step 2 was performed.

STEP 2: Considering trust’s existence on a spectrum or scale, Step 2 was followed by extending the word search from Step 1 to three pieces of additional literature presented in the discussions from Chapter 2. Using the keywords “trust spectrum” and “trust scale”, literature by Chen et al. (2021), Weber et al. (2023) and Wang & Ding, (2024) was searched for any evidence of similarity, with none revealed. This finding was therefore retained for further analysis in step 3. Following the same process, the concept of trust developing at different rates in different industries was further explored (Using the keywords “AI trust rates”, “AI trust in industries” and “industry trust”), with no similarities revealed in either Chen et al. (2021), Weber et al. (2023) or Wang & Ding, (2024). This finding was therefore also retained for further analysis in step 3.

STEP 3: For the final step, three top scholars from the related research question and theme were selected, and a Google Scholar search was performed for the keywords above, on their most recent publications from the past two years. The authors selected were Phanish Puranam; Ella Glikson and Anita Woolley. For all the three authors, no newer publications in top journals were identified as related to the key concepts, with the literature already included in Chapter 2 being the most recent.

6.3.1.5 Interpretation and conclusion on Trust in Artificial Intelligence.

For the first key concept identified in the findings, that *trust takes time to develop*, similarity was found in the extant literature, thus supporting existing discussions as an expected outcome. Having followed the three subsequent steps to analyse the key concept, *trust exists as a spectrum or on a scale*, no apparent similarity was found in literature. This was therefore seen as a nuance of difference to the main theme (Trust in Artificial Intelligence) and was retained as a potential new sub-theme, “***Trust as a Spectrum***”. Lastly, (also having followed the 3 Step process) no evidence was found in literature of the key concept, *trust develops at different rates in different industries*, and this was therefore also considered a nuance of difference and will be retained as a potential new sub-theme “***Trust Rates in Industries***”. Both the potential new sub-themes identified will be included in the revised conceptual framework at the end of Chapter 6.

Table 31:

Summary of Themes, Sub-Themes and Key Concepts for RQ1, Theme 1

Themes from RQ1	Topics / Sub-Themes	Key Concepts Identified
Theme 1: Trust in Artificial Intelligence	• *Trust Develops over Time* • *Trust as a Spectrum* • *Trust Rates in Industries*	• Trust takes time to develop. • Trust exists as a spectrum or on a scale. • Trust develops at different rates in different industries.

Note: Author's Own.

The above table provides a visual summary for ease of reference for which themes, sub-themes and/or key concepts are retained as potential insights or potential contributions (green), and which were found to support the existing scholarship (black).

6.3.2 RQ1: Discussion of Theme 2 – Trust in AI for Adoption and Use

This theme is also a main theme from the Chapter 2 literature review and is reflected in Figure 12 at the start of Chapter 5. The two main topics discussed in the findings were "*Early Adoption & Experimentation*" and "*Change Management Engenders Trust*". The key concepts which emerged from these main topics are compared to the literature in the sections which follow, and conclusions are made as to their potential contribution.

6.3.2.1 Recap of the findings on Trust in AI for Adoption and Use.

The findings on "Trust in Artificial Intelligence for Adoption and Use" highlighted that participants across all groups recognized the importance of a *gradual and strategic approach to AI adoption*. The participants also identified trust in AI as being fundamental for AI adoption. Across the cases, there was an emphasis on building trust through structured and controlled experimentation, often in small-scale pilot projects. The approach aimed to increase familiarity and trust in AI systems, allowing organizations to evaluate and integrate AI into their operations in small steps.

The findings also highlighted that *change management* plays an important role in the adoption and use of AI across different organizational levels. It was consistently emphasized as important for building trust, while aligning with the current change cycle and securing buy-in during the implementation of AI technologies. A difference across the groups lay in their focus, ranging from oversight and risk management to timing and strategic integration, and finally to practical implementation and efficiency gains. In summary, the findings underlined that AI adoption requires a balanced combination of gradual adoption, structured experimentation, and change management.

6.3.2.2 Recap of the literature on Trust in AI for Adoption and Use.

The literature on trust in AI for adoption and use emphasized that trust is important for integrating AI systems in organizations. Vanneste & Puranam (2024) and Enholm et al. (2021) identified a lack of trust as being a barrier to adoption, while Sullivan et al. (2022) highlighted

the importance of understanding and measuring both cognitive and emotional dimensions, suggesting a causal link between them. Wang & Ding (2024) and Weber et al. (2023) proposed that explainable AI (xAI) can enhance trust and increase adoption rates. However, Wong et al. (2024) and Enholm et al. (2021) pointed to a limited understanding of trust in AI across organizational levels. Lastly, Glikson & Woolley (2020) underscored trust as being foundational for AI's use in organisations and advocated for research addressing the cognitive and emotional aspects together. Despite these insights, gaps remained in exploring the relationship between these dimensions and their impact on AI adoption.

6.3.2.3 Discussion of the findings and literature.

Both the literature and the findings were similar in that they underscored the importance of trust and its role in AI adoption within organisations. However, there was no evidence in the Chapter 2 literature review on "Trust in artificial intelligence for adoption and use" which pointed towards the two key concepts identified in the findings. Therefore, both findings were retained as potential differences and the 3-Step process, described in Section 6.1.1 above, was followed to analyse both further.

6.3.2.4 Further discussion of the differences between findings and literature.

STEP 1: The articles which were identified for use in Step 1 were those cited in the related research question, as highlighted in Table 24: Vanneste & Puranam, (2024), Sullivan et al. (2022), Wong et al. (2024) and Enholm et al. (2021). The keywords used across both remaining key concepts were "change management", "early adoption" and "experimentation". Although Enholm et al. (2021) was included in the literature review, change management was not a key area of interest within the theme at the time, and was not previously captured. However, upon further review as part of this step, although "change management" was not specifically mentioned, there was a strong theme of change in business process, organisational structure, and employee-AI Trust already identified by the authors (Enholm et al., 2021, p. 1718). The key concept of "change management engenders trust" was therefore similar to and supported the existing body of knowledge. The concept of "early adoption and experimentation", however was not identified and was analysed further in Step 2.

STEP 2: Considering "early adoption and experimentation", Step 2 was followed by extending the word search from Step 1 to three pieces of additional literature presented in the discussions from Chapter 2. Using the keywords "early adoption" and "experimentation", literature by Wang & Ding, (2024), Weber et al. (2023) and Glikson and Woolley, (2020) was searched for any evidence of similarity, with none revealed. This finding was therefore retained for further analysis in step 3.

STEP 3: For the final step, three top scholars from the related research question and theme were selected, and a Google Scholar search was performed for the keywords above, on their most recent publications from the past two years. The authors selected were Phanish Puranam, Ella Glikson and Anita Woolley. For all the three authors, no newer publications in top journals were identified as being related to the key concept, with the literature already included in Chapter 2 being the most recent.

6.3.2.5 Interpretation and conclusion on Trust in AI for adoption and use.

For the first key concept identified in the findings, that *change management engenders trust*, similarity was found in the extant literature, and it was retained as an expected outcome, supporting the existing discourse. Having followed the three subsequent steps to analyse the key concept, *early adoption and experimentation*, no apparent similarity was found in literature. This was therefore seen as a nuance of difference to the main theme (Trust in Artificial Intelligence for Adoption and Use) and was retained as a potential new sub-theme, "***Early Adoption and Experimentation***". The potential new sub-theme identified will be included in the revised conceptual framework at the end of Chapter 6.

Table 32:

Summary of Themes, Sub-Themes and Key Concepts for RQ1, Theme 2

Themes from RQ1	Topics / Sub-Themes	Key Concepts Identified
Theme 2: Trust in Artificial Intelligence for Adoption and Use	• *Early Adoption & Experimentation* • *Change Management Engenders Trust*	• Importance of gradual AI adoption and experimentation • Importance of change management for trust in AI.

Note: Author's Own.

The above table provides a visual summary for ease of reference for which themes, sub-themes and/or key concepts are retained as potential insights or potential contributions (green), and which were found to support the existing scholarship (black).

6.3.3 RQ1: Discussion of Theme 3 – Trust in AI for Decision-Making

This theme is another main theme from the Chapter 2 literature review and is reflected in Figure 12 at the start of Chapter 5. The two main topics discussed in the findings were "*Only to Inform Decisions*" and "*Output Reliability and Trust*". The key concepts which emerged from these main topics are compared to the literature in the sections which follow, and conclusions are made as to their potential contribution.

6.3.3.1 Recap of the findings on Trust in AI for Decision-Making.

The findings on "Trust in Artificial Intelligence for Decision-Making" across participant groups highlighted two key concepts. First, there was a shared understanding that AI should inform and enhance human decision-making, but not replace it. Across all groups, AI was

consistently viewed as a supportive tool to inform and enhance decision-making, rather than a replacement for human judgment. Secondly, there was an emphasis on the necessity for human verification and that trust in AI depends on consistent, accurate and reliable outputs. AI's role was seen more as contributing to decision-making but that humans were still needed to ensure alignment with organisational requirements. To summarise, AI was recognized as a valuable contributor to decision-making, but human oversight was stressed as being essential to align AI-driven decisions with organizational standards and expectations. This would, in turn, build trust in AI outputs through consistency, reliability, and verification.

6.3.3.2 Recap of the literature on Trust in AI for Decision-Making.

The literature on trust in AI for decision-making highlighted that trust is important for integrating AI-based decision-making within organizations. Enholm et al. (2021) identified that as employees may increasingly be required to rely on AI for decisions, building trust between humans and AI remains a challenge. Similarly, Chen et al. (2021) emphasized the rising adoption of AI for decision support, particularly during the Covid-19 pandemic, and stressed that trust is a fundamental prerequisite for using these AI-based decision support systems. Additionally, Chen et al. (2021) focused on "human-in-the-loop" type systems (p. 1), and Enholm et al. (2021) emphasized the focus of AI as being an "assistive role" which supports, rather than replaces humans (p. 1720). They also called for further research into human-centered AI frameworks tailored to different industries. Both Enholm et al. (2021) and Sabharwal et al. (2024) emphasized the data-driven nature of AI and its importance in decision-making; with Sabharwal et al. (2024) inferring that the key concepts of this research (trust in AI, explainability (xAI) and transparency) may assist with trust engenderment in data-driven decision-making. Wang & Ding (2024) identified a lack of trust as a barrier to decision-making with AI, and Sabharwal et al. (2024) suggested that AI decision support for use in decision-making would not be possible without managerial adoption. Additionally, Shrestha et al. (2019) underscored the importance of addressing "trust concerns" to fully capitalize on the potential of AI's future in decision-making. Comparing the literature revealed a shared understanding that trust in AI is a prerequisite for its adoption in decision-making, yet a lack of trust was also seen as a challenge to be overcome.

6.3.3.3 Discussion of the findings and literature.

Both the findings and the literature showed similarities on the role of AI as a supportive tool that should enhance human decision-making, rather than replacing it. The findings indicated that AI was consistently viewed across all participant groups as a supplemental tool that informs and supports human judgment in decision-making. Similarly, the literature by Enholm et al. (2021) and Chen et al. (2021) emphasized the role of AI in decision support.

Chen et al. (2021) focused on “human-in-the-loop” type systems (p. 1), and Enholm et al. (2021) emphasized the focus of AI as being an “assistive role” which supports, rather than replaces humans (p. 1720). This finding was therefore viewed as supporting the body of extant literature. Regarding human verification of AI outputs, the findings highlighted the potential to engender trust in AI for decision-making support, by ensuring human verification and checking of AI outputs. The literature of Shrestha et al. (2019) and Chen et al. (2021) called for research to discover how “trust concerns” can be overcome. The findings were viewed as potentially providing insights to this call from literature and this finding was therefore retained as a potential new insight to be checked for using the 3-Step process.

6.3.3.4 Further discussion of the differences between findings and literature.

STEP 1: The articles which were identified for use in Step 1 were those cited in the related research question, as highlighted in Table 24: Vanneste & Puranam, (2024), Sullivan et al. (2022), Wong et al. (2024) and Enholm et al. (2021). The keywords used in the search of the literature related to the remaining key concept were “human verification”, “verification”, “checking”, “output reliability” and “AI outputs”. The word searches on the four selected articles did not yield any matches, therefore Step 2 was performed.

STEP 2: Considering human verification of AI outputs, Step 2 was followed by extending the word search from Step 1 to four pieces of additional literature presented in the discussions from Chapter 2. Using the keywords “human verification”, “verification”, “checking”, “output reliability” and “AI outputs”, literature by Chen et al. (2021), Sabharwal et al. (2024), Shrestha et al. (2021) and Glikson & Woolley, (2020) was searched for any evidence of similarity, with none revealed. Although Glikson & Woolley, (2020) did discuss AI reliability, they did not discuss it in relation to human verification of AI outputs as engendering trust in AI for decision-making. This finding was therefore considered a nuance of difference to extant literature and is retained for further analysis in Step 3.

STEP 3: For the final step, three top scholars from the related research question and theme were selected, and a Google Scholar search was performed for the keywords above, on their most recent publications from the past two years. The authors selected were Phanish Puranam, Ella Glikson and Anita Woolley. For all of the three authors, no newer publications in top journals were identified, with the literature already included in Chapter 2 once again being the most recent.

6.3.3.5 Interpretation and conclusion on Trust in AI for Decision-Making.

For the first key concept identified in the findings, that *AI should only inform and enhance human decision-making,* similarity was found in the extant literature, thus further

supporting existing discussions. Having followed the three subsequent steps to analyse the key concept, *human verification of AI outputs*, no apparent similarity was found in literature. This was therefore seen as a nuance of difference to the main theme (Trust in Artificial Intelligence for Decision-Making) and was retained as a potential new sub-theme, "***Output Reliability and Trust***". As before, the potential new sub-theme identified will be included in the revised conceptual framework at the end of Chapter 6.

Note: The nuance of difference of *human verification of AI outputs*, was noted as a potential direct insight into the (non-RQ) call for research by Shrestha et al. (2019) and Chen et al. (2021) to discover how "trust concerns" can be overcome. This sub-theme "***Output Reliability and Trust***" may therefore provide potential additional insight to a call for further research which was not initially scoped as one of the Research Questions for this dissertation.

Table 33:

Summary of Themes, Sub-Themes and Key Concepts for RQ1, Theme 3

Themes from RQ1	Topics / Sub-Themes	Key Concepts Identified
Theme 3: Trust in Artificial Intelligence for Decision-Making	• *Only to Inform Decisions* • *Output Reliability & Trust*	• AI should inform and enhance human decision-making, not replace it. • Human verification of AI outputs is needed.

Note: Author's Own.

The above table provides a visual summary for ease of reference for which themes, sub-themes and/or key concepts are retained as potential insights or potential contributions (green), and which were found to support the existing scholarship (black).

6.3.4 RQ1: Discussion of Theme 4 – "Blind-Trust" Tensions

This theme is a potentially new main theme which was not identified in the Chapter 2 literature review. It is reflected as a potential new theme in Figure 12 at the start of Chapter 5. The main topic discussed in the findings was "*Blind-Trust Tensions*". The key concept which emerged from the main topic is compared to the literature in the sections which follow, and conclusions are made as to its potential contribution.

6.3.4.1 Recap of the findings on "Blind-Trust" Tensions.

A new theme of "blind-trust" tensions emerged during the analysis of the findings in Chapter 5, which revealed an interesting insight across different decision-making levels. The underlying meaning of the theme was captured in the identified key concept that there is a tension between decision-making groups regarding "blindly-trusting" AI. Some supported it while others rejected it. For clarity, "blind-trust" in this context was identified by the participants as the action of simply accepting and trusting AI at face-value, with no checks or questions as to its ability, reliability or consistency. The findings revealed that there was a similar recognition

of the risks associated with “blind-trust” in AI, with participants acknowledging the potential pitfalls of reliance on AI technologies.

However, in terms of differences, there was a clear tension regarding "blindly-trusting" AI. Some participants responsible for decision-making rejected blind trust altogether, reflecting a strategic, risk-averse stance, while others demonstrated more varied perspectives, with some showing a higher tolerance for trusting AI without full verification. Interestingly, the Operational Decision-Making group did not raise the topic of “blind-trust”, highlighting that there appears to be a potential gap in awareness or prioritization of this issue at an operational level. This variation in perspectives pointed to a broader tension (across all decision-making levels) within organizations between maintaining control of implementation and/or adoption risks and seeking practical trust in AI for decision-making.

6.3.4.2 Recap of the literature on “Blind-Trust” Tensions.

There was no evidence presented in the Chapter 2 literature review on *“Blind-Trust” Tensions*. Therefore, the finding was retained as a potential difference and the 3-Step process, described in Section 6.1.1 above, was followed to analyse it further.

6.3.4.3 Further discussion of the differences between findings and literature.

STEP 1: The articles which were identified for use in Step 1 were those cited in the related research question, as highlighted in Table 24: Vanneste & Puranam, (2024), Sullivan et al. (2022), Wong et al. (2024) and Enholm et al. (2021). The keywords used to search for evidence of the key concept were “blind-trust”, “blind”, “unconditional trust”, “questioning” and “trust tension”. The word searches on the four selected articles did not yield any matches, therefore Step 2 was performed.

STEP 2: Considering the absence of *“blind-trust” tensions* from the literature review, a more exhaustive Step 2 was followed by extending the word search from Step 1 to **not three, but six** pieces of additional literature, presented in the broader discussions from Chapter 2. Using the same keywords identified for Step 1 above, literature by Chen et al. (2021), Weber et al. (2023) and Wang & Ding, (2024), Glikson & Woolley, (2020); Shrestha et al. (2021) and Berente et al. (2021) was searched for any evidence of similarity, with none revealed. This finding was therefore retained for further analysis in step 3.

STEP 3: For the final step, three top scholars from the related research question and theme were selected, and a Google Scholar search was performed for the keywords above, on their most recent publications from the past two years. The authors selected were Phanish Puranam, Ella Glikson and Anita Woolley. For all the three authors, no newer publications in

top journals were identified as related to the key concept, with the literature already included in Chapter 2 being the most recent.

6.3.4.4 Interpretation and conclusion on "Blind-Trust" Tensions.

For the key concept identified, that *there is a tension between decision-making groups regarding "blindly-trusting" AI*, no apparent similarity was found in the extant literature, thus revealing a potential new insight related to trust development in AI for its adoption and use in decision-making. This was therefore seen as a distinct difference to the extant literature and was retained as a potential new theme, "***Blind-Trust Tensions***". This potential new theme will be included in the revised conceptual framework at the end of Chapter 6.

Table 34:

Summary of Themes, Sub-Themes and Key Concepts for RQ1, Theme 4

Themes from RQ1	Topics / Sub-Themes	Key Concepts Identified
Theme 4: "Blind-Trust" Tensions	• *"Blind-Trust" Tensions*	• There is a tension between the Decision-Making groups regarding "blindly-trusting" AI. Some support it while others reject it.

Note: Author's Own.

The above table provides a visual summary for ease of reference for which themes, sub-themes and/or key concepts are retained as potential insights or potential contributions (green), and which were found to support the existing scholarship (black).

6.3.5 RQ1: Discussion of Theme 5 – "Blast-Radius".

This theme is a potentially new main theme which was not identified in the Chapter 2 literature review. It is reflected as a potential new theme in Figure 12 at the start of Chapter 5. The main topic discussed in the findings was "*Blast-Radius*". The key concept which emerged from the main topic was compared to the literature in the sections which follow, and conclusions were made as to its potential contribution.

6.3.5.1 Recap of the findings on "Blast-Radius".

A new theme of "*blast-radius*" emerged during the analysis of the findings in Chapter 5, which revealed another interesting insight across different decision-making levels. The underlying meaning of the theme is captured in the key concept, which highlighted the importance of containing the potential negative outcomes of AI use in decision-making, to engender trust. Each group recognized the importance of containment strategies to control the risks associated with AI use. All similarly advocated for methods to ensure that AI's impact remains predictable and manageable, whether through phased rollouts, safety mechanisms, or controlled environments. Their shared concern highlighted the similarity of focus on

mitigating risks across the organization, demonstrating a collective view of ensuring that AI adoption does not lead to unintended consequences.

However, the groups suggestions varied in their approach to containment. Some prioritized a structured, proactive approach to risk control, using phased rollouts and safety nets to manage potential fallout. Others emphasized scenario planning and trade-offs, adapting to management of different risk levels. Another approach focused on practical, incremental trust-building through gradual testing and codes of conduct, reflecting a step-by-step emphasis on trust and practical limitations. These differences highlighted varying priorities in how containment of AI's "Blast-Radius" was managed. This demonstrated a diverse but consistent focus on mitigating risks associated with AI, ensuring that any negative impacts are contained while fostering trust and confidence in AI's use in decision-making.

6.3.5.2 Recap of the literature on "Blast-Radius".

There was no evidence presented in the Chapter 2 literature review on *"blast-radius"*. Therefore, the finding was retained as a potential difference and the 3-Step process, described in Section 6.1.1 above, was followed to analyse it further.

6.3.5.3 Further discussion of the differences between findings and literature.

STEP 1: The articles which were identified for use in Step 1 were those cited in the related research question, as highlighted in Table 24: Vanneste & Puranam, (2024), Sullivan et al. (2022), Wong et al. (2024) and Enholm et al. (2021). The keywords used to search for evidence of the key concept were "blast-radius", "containment", "risk mitigation", "boundary" and "negative outcomes". Sullivan et al. (2022) did not discuss any of the keywords in the context of the key concept itself, but highlighted a need for further research to "identify and mitigate potential risks associated with artificial agents", in relation to their identified concept of "uncanny feeling" and trust in AI (p. 542). The key concept of setting a "blast-radius" to contain potential AI risk may therefore possibly highlight an insight to the request for further research by Sullivan et al. (2022). However, the word searches on the four selected articles did not yield any matches directly related to the key concept, or evidence of the containment of potential negative AI outcomes, therefore Step 2 was still performed.

STEP 2: Considering the absence of *"blast radius"* from the literature review, a more exhaustive Step 2 was followed by extending the word search from Step 1 to **not three, but six** pieces of additional literature, presented in the broader discussions from Chapter 2. Using the same keywords identified for Step 1 above, literature by Chen et al. (2021), Weber et al. (2023) and Wang & Ding, (2024), Glikson & Woolley, (2020); Shrestha et al. (2021) and Berente et al. (2021) was searched for any evidence of similarity, with none revealed. This finding was therefore retained for further analysis in step 3.

STEP 3: For the final step, three top scholars from the related research question and theme were selected, and a Google Scholar search was performed for the keywords above, on their most recent publications from the past two years. The authors selected were Phanish Puranam, Ella Glikson and Anita Woolley. For all the three authors, no newer publications in top journals were identified as related to the key concept, with the literature already included in Chapter 2 being the most recent.

6.3.5.4 Interpretation and conclusion on "Blast-Radius".

For the key concept identified, *the importance of containing the potential negative outcomes of AI use in decision-making, to engender trust*, no apparent similarity was found in the extant literature, thus revealing a potential new insight related to trust development in AI for its adoption and use in decision-making. This insight was therefore seen as a distinct difference to the extant literature and was retained as a potential new theme, "*Blast-Radius*". The potential new theme will be included in the revised conceptual framework at the end of Chapter 6.

Note: Step 1 of the 3-Step process revealed that the key concept of *containing the potential negative outcomes of AI use in decision-making, to engender trust,* may potentially address an aspect of the call for further research by Sullivan et al. (2022) to "identify and mitigate potential risks associated with artificial agents" (p. 542). This was therefore also highlighted as a potential new insight for their request. This theme "***Blast-Radius***" may therefore provide potential additional insight to a call for further research which was not initially scoped as one of the Research Questions for this dissertation.

Table 35:

Summary of Themes, Sub-Themes and Key Concepts for RQ1, Theme 5

Themes from RQ1	Topics / Sub-Themes	Key Concepts Identified
Theme 5: "Blast-Radius"	• *"Blast-Radius"*	• Importance of containing the potential negative outcomes of AI use for decision-making, to engender trust.

Note: Author's Own.

The above table provides a visual summary for ease of reference for which themes, sub-themes and/or key concepts are retained as potential insights or potential contributions (green), and which were found to support the existing scholarship (black).

6.4 Research Question 2: *How does xAI engender trust in AI for decision-making in organisations?* And: *What are the factors that contribute to trust in AI?*

The aims of RQ2 were to firstly explore and understand how xAI contributes to the formation of trust in AI for organisational decision-making. Secondly, to explore additional insights and understanding of the factors which the organisational end-user finds relevant to engendering their trust in AI for decision-making. There are three themes which were mapped directly to RQ2, with one potential new theme, all of which are discussed in this section. Table 34 below provides a summary of the themes, topics / sub-themes and key concepts from the Chapter 5 findings which will be discussed under RQ2.

Table 36:

Themes Emerging from RQ2, and their Key Concepts Identified in the Findings.

RQ2: *How does xAI engender trust in AI for decision-making in organisations? And: What are the factors that contribute to trust in AI?*		
Themes from RQ2	**Topics / Sub-Themes**	**Key Concepts Identified**
Theme 1: xAI and the "Black-Box"	• *The "Black-Box"* • *Understanding and the "Black-Box"* • *Transparency and the "Black-Box"*	• Awareness and shared concern of the "Black -Box" nature of AI. • Transparency, understanding and explainability were all recognised to build trust and overcome the "Black-Box".
Theme 2: xAI for Trust in AI	• *Explanation of Data* • *Explanation of "How?"*	• Explainability should address "How" AI arrives at outcomes. • Explanation must be consistent throughout the data lifecycle.
Theme 3: xAI for Decision-Making	• *Self-Criticism of AI*	• AI Self-Criticism is necessary for trust and confidence in AI systems for decision-making.
Theme 4: Antecedents	• *Provenance for Trust* • *Importance of Data*	• Provenance is an antecedent to adoption. • Data integrity is an antecedent to adoption.

Note: Author's Own.

6.4.1 RQ2: Discussion of Theme 1 – xAI and the "Black-Box".

This theme is a main theme from the Chapter 2 literature review and is reflected in Figure 12 at the start of Chapter 5. The three main topics discussed in the findings were "*The "Black-Box"*", "Understanding and the "Black-Box"" and "Transparency and the "Black-Box"". The key concepts which emerged from these main topics are compared to the literature in the sections which follow, and conclusions are made as to their potential contribution.

6.4.1.1 Recap of the findings on xAI and the "Black-Box".

Two key concepts were identified in the findings, namely: an *awareness and shared concern of the "Black-Box" nature of AI* and secondly, the insight that *transparency, understanding and explainability were all recognised to build trust and overcome the "Black-Box"*. The concept of the "black-box" in AI emerged as a shared concern across all decision-making groups, highlighting the challenges posed by the opacity of AI systems. Across the groups, there was recognition that a lack of clarity can lead to discomfort and mistrust, underscoring the need to address that lack of clarity to build confidence in AI. Across all

groups, transparency, explainability (xAI), and understanding were identified as key elements to overcoming the “black-box” issue and ensuring trust in AI, although the emphasis on each varied.

For some groups, tackling the “black-box” problem involved a more comprehensive approach. Some emphasized a broader set of tools, citing transparency, explainability, and understanding as being important to building trust and addressing the "black-box" issue. Others focused primarily on explanation and understanding, particularly of data sources and AI methods, though they did not emphasize transparency as much. The remainder focused mainly on transparency to resolve their discomfort but did not specifically highlight the need for explainability or understanding. Despite the differences in emphasis, there was a shared recognition of the need to “demystify” AI systems. All groups pointed to either or all of transparency, explainability, and understanding as important to mitigating the “black-box” nature of AI and fostering trust in its outputs. This consistent acknowledgment underscored the importance of clear and interpretable AI systems to encourage broader adoption and trust within organizations.

6.4.1.2 Recap of the literature on xAI and the “Black-Box”.

The literature on xAI and the “black box” identified the inscrutability of AI systems as a barrier to adoption and trust. Rai (2020) and Berente et al. (2021) emphasized that the “black-box” nature of AI models hinders decision-making. Explainable AI (xAI) was presented as a potential solution to this challenge, providing clarity and rationale for AI decisions. Rai (2020) further suggested that xAI can shift AI from a “black box” to a more transparent “glass box,” influencing users’ willingness to act on AI predictions (p. 139), while Weber et al. (2023) also highlighted explainability as essential in regulated sectors like finance, where understanding AI processing steps is important. Similarly, Kim et al. (2023) focused on the lack of transparency in healthcare predictive analytics, linking it to the “black-box” structure of AI models. Wang & Ding (2024) further suggested that explainability (xAI) can improve transparency and accountability, thus addressing the “black box” problem for decision-making. Lukyanenko et al. (2022) emphasized the need for a foundational understanding of xAI, transparency, and the mechanisms within AI systems. Collectively, the literature viewed the “black box” as a fundamental barrier to AI uptake, with xAI and transparency being important elements in building trust, understanding, and accountability in AI-driven decision-making.

6.4.1.3 Discussion of the findings and literature.

Both the findings and the literature similarly recognized the “black-box” nature of AI as a barrier to adoption and trust. In the findings, decision-making groups acknowledged the challenges posed by the opacity of AI systems and the discomfort it creates. Similarly, the

literature pointed out that the inscrutability of AI models hinders decision-making and erodes trust (Berente et al., 2021; Rai, 2020; Lukyanenko et al., 2023). A further similarity was that both the findings and the literature emphasized that transparency, explainability (xAI), and understanding are key elements in overcoming the "black-box" problem. The findings indicated that these factors were also seen as important for building trust and fostering acceptance of AI across decision-making groups. The literature also discussed xAI and transparency and understanding as important to improving understanding and accountability (Weber, 2023; Lukyanenko et al., 2022). Lastly, both the findings and the literature underscored the role of explainability (xAI) in fostering trust and ensuring that AI systems are clear and interpretable. The findings described a consistent recognition of xAI as a tool to build confidence in AI. Similarly, the literature emphasized xAI's potential to shift AI from a "black-box" to a more transparent and understandable "glass-box" (Rai, 2020; Ågerfalk, 2020).

6.4.1.4 Interpretation and conclusion on xAI and the "Black-Box".

For the first key concept identified in the findings, *an awareness and shared understanding of the "black-box and AI*, similarity was found between the findings and the extant literature, thus supporting existing discussions in the scholarship. This was therefore seen as a similarity to the extant literature and was not retained as a potential new sub-theme. For the second key concept identified in the findings, *that transparency, understanding and explainability were all recognised to build trust and overcome the "black-box"*, similarity was also found between the findings and the extant literature, thus supporting existing discussions in the scholarship. This was therefore also seen as a similarity to the extant literature and was not retained as a potential new sub-theme. However, in relation to the main theme, *xAI and the "Black-Box"*, there was a further insight from the findings that it's not just xAI that can address the "Black-Box", but also Transparency and Understanding. There was therefore a nuance of difference which did not necessarily warrant a new sub-theme (due to the overall similarity) but was of importance to understanding the seemingly circular/cross-supportive relationship between xAI, transparency and understanding. In summary, the literature identifies each topic / sub-theme individually, but a picture of the "interplay"/circularity/cross-support between them becomes somewhat clearer.

Table 37:

Summary of Themes, Sub-Themes and Key Concepts for RQ2, Theme 1

Themes from RQ2	Topics / Sub-Themes	Key Concepts
Theme 1: xAI and the "Black-Box"	• *The "Black-Box"* • *Understanding and the "Black-Box"* • *Transparency and the "Black-Box"*	• Awareness and shared concern of the "Black -Box" nature of AI. • Transparency, understanding and explainability were all recognised to build trust and overcome the "Black-Box".

Note: Author's Own.

Table 37 provides a visual summary for ease of reference for which themes, sub-themes and/or key concepts are retained as potential insights or potential contributions (green), and which were found to support the existing scholarship (black).

6.4.2 RQ2: Discussion of Theme 2 – xAI for Trust in AI

This theme is a main theme from the Chapter 2 literature review and is reflected in Figure 12 at the start of Chapter 5. The two main topics discussed in the findings were "*Explanation of Data*" and "*Explanation of "How?"*". The key concepts which emerged from these main topics are compared to the literature in the sections which follow, and conclusions are made as to their potential contribution.

6.4.2.1 Recap of the findings on xAI for Trust in AI.

Two key concepts were identified in the findings, namely: that *Explainability should address "How" AI arrives at outcomes* and secondly, the insight that *Explanation must be consistent throughout the data lifecycle*. Across all groups, there was a shared emphasis on explaining both "how" AI systems operate and the credibility of the data they use. This collective focus highlighted the importance of explanation and transparency of AI's methodology and data sources to build trust. This shared recognition across the groups reinforced that understanding AI's processes and the data it relies on are important to fostering trust and facilitating AI adoption.

There was a difference in focus regarding what aspects of AI explanation were prioritized. Some groups emphasized both explanation of the AI methodology (the "How") and explanation of data credibility as important to trusting AI decisions. Others concentrated on the need for clarity throughout the entire AI decision-making process, ensuring that all stages are transparent to support effective decision-making. Another perspective focused more on practical aspects, such as how AI enhances efficiency and maintains data integrity.

A further distinct difference was from the Senior Decision-Making group who emphasised the importance of clarity (through explanation) throughout the data lifecycle. Overall, the key concepts identified included the importance of addressing the "How?" behind AI outcomes, ensuring consistent explanations of data credibility throughout the data lifecycle, and recognizing that explanation requirements might vary across different decision-making levels.

6.4.2.2 Recap of the literature on xAI for Trust in AI.

The analysis of literature on xAI and trust in AI highlighted a growing consensus across multiple studies that explainability is key to fostering trust in AI systems. Haque et al. (2023) and Laato et al. (2022) demonstrated that xAI not only leads to trust but also positions trust as a primary goal of explainable AI. However, Wang & Ding (2024) presented a different view,

showing only partial support for the connection between xAI and trust, which underscored the need for further research. Scholars such as Silva et al. (2023), Rai (2020), and Sabharwal et al. (2024) called for more exploration of xAI's role in trust development across different application domains and managerial contexts. Esmaeilzadeh & Vaezi (2022) pointed to AI needing to explain itself to satisfy regulatory requirements and to "prove lack of bias" (p. 560), a sentiment which was echoed by Weber et al. (2023). There were also appeals to better understand the design and implementation of xAI (Pumplun et al., 2023; Abedin, 2022) and to establish design guidelines for xAI for its use by end-users. The literature therefore suggested that xAI presents an opportunity to engender trust in AI across industries and domains, but also indicated that more foundational research is needed. The scholars advocated for further investigation to clarify the relationship between xAI and trust.

6.4.2.3 Discussion of the findings and literature.

Although there was a high-level similarity between the findings and literature that xAI or explainability is central to enhancing or engendering trust in AI (Haque et al., 2023; Laato et al., 2022), the two main topics of "*Explanation of Data*" and "*Explanation of "How?"* were not specifically mentioned or alluded to in the literature review. The two main topics may therefore be relevant to the calls by Pumplun et al. (2023) and Abedin, (2022) to better understand the requirements for design and implementation of xAI. Considering the above, the two topics and the highlighted key concepts were retained for further analysis using the 3-Step process.

6.4.2.4 Further discussion of the differences between findings and literature.

STEP 1: The articles which were identified for use in Step 1 were those cited in the related research question (RQ2), as highlighted in Table 24: (Rai, 2020); (Sabharwal et al., 2024); (Lukyanenko et al., 2022). The keywords used across both remaining key concepts were "data explanation", "how", "explanation of how", "data lifecycle", "data sources" and "AI methodology". Although there were numerous matches to the "data" keyword in all three articles, and some mention of data sources, none of them mentioned or alluded to "Explanation of data", let alone of the concept being of any importance to xAI. Therefore, this finding was retained as a difference for further analysis in Step 2. Regarding the second topic "Explanation of "How?"", A definition was found in Rai, (2020) which stated, "*Explainable AI (XAI) is the class of systems that provide visibility into **how** an AI system makes decisions and predictions and executes its actions*" (p. 137). The second concept "Explanation of "How?"" therefore supported extant literature and was not analysed further in subsequent steps.

STEP 2: Considering "*Explanation of Data*", Step 2 was followed by extending the word search from Step 1 to three pieces of additional literature presented in the related discussions

from Chapter 2. Using the keywords “data explanation” and “data sources” and “data lifecycle”, literature by Berente et al. (2021), Wang & Ding (2024) and Abedin, (2022) was searched for any evidence of similarity. Again, none of the included authors identified “data explanation”, “data sources” or “data lifecycle” as important to xAI. The “*Explanation of Data*” finding was therefore retained for further analysis in step 3.

STEP 3: For the final step, three top scholars from the related research question and theme were selected, and a Google Scholar search was performed for the keywords above, on their most recent publications from the past two years. The authors selected were Arun Rai, Nicholas Berente and Roman Lukyanenko. For all the three authors, no newer publications in top journals were identified as related to the key concepts, with the literature already included in Chapter 2 being the most recent.

6.4.2.5 Interpretation and conclusion on xAI for Trust in AI.

For the first key concept identified in the findings, that *explainability should address “How” AI arrives at outcomes*, similarity was found in the extant literature, thus supporting existing discussions. Having followed the three subsequent steps to analyse the second key concept, *explanation must be consistent throughout the data lifecycle*, no apparent similarity was found in literature. This was therefore seen as a nuance of difference to the extant literature and was retained as a potential new sub-theme, “***Explanation of Data***”. The potential new sub-theme identified will be included in the revised conceptual framework at the end of Chapter 6.

Note: The discussion of the findings and literature revealed that the sub-theme may therefore be relevant to the calls by Pumplun et al. (2023) and Abedin, (2022) to better understand the requirements for design and implementation of xAI. The sub-theme “***Explanation of Data***” may therefore provide potential additional insight to their call for further research which was not initially scoped as one of the Research Questions for this dissertation.

Table 38:

Summary of Themes, Sub-Themes and Key Concepts for RQ2, Theme 2

Themes from RQ2	Topics / Sub-Themes	Key Concepts
Theme 2: xAI for Trust in AI	• *Explanation of Data* • *Explanation of “How?”*	• Explanation must be consistent throughout the data lifecycle. • Explainability should address “How” AI arrives at outcomes.

Note: Author’s Own.

The above table provides a visual summary for ease of reference for which themes, sub-themes and/or key concepts are retained as potential insights or potential contributions (green), and which were found to support the existing scholarship (black).

6.4.3 RQ2: Discussion of Theme 3 – xAI for Decision-Making

This theme is another main theme from the Chapter 2 literature review and is reflected in Figure 12 at the start of Chapter 5. The main topic discussed in the findings was "*Self-Criticism of AI*". The key concept which emerged from the main topic is compared to the literature in the sections which follow, and conclusions are made as to its potential contribution.

6.4.3.1 Recap of the findings on xAI for Decision-Making.

The key concept identified, which emerged from the findings was, that *AI Self-Criticism is necessary for trust and confidence in AI systems for decision-making*. Participants emphasized that AI must explore and openly communicate its errors, enhancing confidence, credibility, and understanding of its processes. Self-awareness and self-assessment were viewed as important to demonstrating reliability and transparency. This ability of AI to communicate or explain its self-criticism to the user might enhance understanding of its processes and operation, which was highlighted as being important for building trust and confidence in its decision-making.

Furthermore, there was a similar focus between groups on AI's capacity to recognize and communicate its limitations, linking this self-awareness to governance, accountability, and gaining user acceptance. Additionally, participants stressed that AI's self-criticism needed to be actionable, offering clear justifications for its decisions to improve use in operations. Overall, for explainable AI (xAI) to be effective, it must incorporate mechanisms for ongoing self-assessment, error reporting, and transparent communication of limitations to ensure trust, transparency, and accountability in decision-making processes.

6.4.3.2 Recap of the literature on xAI for Decision-Making.

The literature on xAI for decision-making underscored its role in enhancing trust and providing clarity in AI-supported decision-making. Rai (2020) highlighted that xAI can illuminate the rationale within AI systems, addressing the limited understanding of AI decisions. Rabiee et al. (2024) reinforced the importance of xAI by emphasizing supervised feature selection to enhance explainability. Similarly, Vanneste & Puranam (2024) and Ågerfalk (2020) emphasized the increasing significance of xAI in decision-making, while Berente et al. (2021) argued that AI explanations should not only address cognitive aspects but also include often overlooked social contexts (p. 1444). Gregory et al. (2021, p. 24) considered explainability as strategic due to its impact on perceived user value. Wang & Ding (2024) found that xAI can "improve decision accuracy" in sales settings (p. 1), while Sabharwal et al. (2024) noted that xAI explanations for AI decision-support systems remain underdeveloped in management and finance literature. Despite the consensus on xAI's importance, scholars such as Wang & Ding (2024) and Sabharwal et al. (2024) called for more

research to understand xAI's influence on structured decision support (in the context of individuals responsible for decision-making). The literature also suggested a need for exploring cognitive versus social trust in AI, which may be linked to emotional factors of trust, but this required further investigation (Berente et al., 2021; Glikson & Woolley, 2020).

6.4.3.3 Discussion of the findings and literature.

There was no apparent similarity observed between the findings and the literature presented in the Chapter 2 literature, regarding *"AI Self-Criticism"*. Therefore, the finding was retained as a potential nuance of difference and the 3-Step process, described in Section 6.1.1 above, was followed to analyse it further.

6.4.3.4 Further discussion of the differences between findings and literature.

STEP 1: The articles which were identified for use in Step 1 were those cited in the related research question, as highlighted in Table 24: (Rai, 2020); (Sabharwal et al., 2024); (Lukyanenko et al., 2022). The keywords used in the search of the literature related to the key concept were "self-criticism", "criticism", "self-verification" and "limitations". The word searches on the three selected articles yielded some matches in Lukyanenko et al. (2022) to discussion of AI limitations, but nothing related to the AI itself explaining or communicating its limitations to the user. The rest of the keyword searches yielded no results, therefore Step 2 was performed.

STEP 2: Considering AI self-criticism, Step 2 was followed by extending the word search from Step 1 to four pieces of additional literature presented in the discussions from Chapter 2. Using the keywords "self-criticism", "criticism", "self-verification" and "limitations", literature by Rabiee et al. (2024); Vanneste & Puranam (2024); Wang & Ding (2024) was searched for any evidence of similarity, with none revealed. The finding was therefore retained for further analysis in Step 3.

STEP 3: For the final step, three top scholars from the related research question and theme were selected, and a Google Scholar search was performed for the keywords above, on their most recent publications from the past two years. The authors selected were Arun Rai, Nicholas Berente and Roman Lukyanenko. For all of the three authors, no newer publications in top journals were identified, with the literature already included in Chapter 2 once again being the most recent.

6.4.3.5 Interpretation and conclusion on xAI for Decision-Making.

For the key concept identified in the findings, that *AI Self-Criticism is necessary for trust and confidence in AI systems for decision-making,* no apparent similarity was found in the extant literature after having followed the 3-step process above. This was therefore seen

as a nuance of difference to the main theme (xAI for Decision-Making) and was retained as a potential new sub-theme, "***AI Self-Criticism***". The potential new sub-theme identified will be included in the revised conceptual framework at the end of Chapter 6.

Table 39:

Summary of Themes, Sub-Themes and Key Concepts for RQ2, Theme 3

Themes from RQ2	Topics / Sub-Themes	Key Concept
Theme 3: xAI for Decision-Making	• *Self-Criticism of AI*	• AI Self-Criticism is necessary for trust and confidence in AI systems for decision-making.

Note: Author's Own.

The above table provides a visual summary for ease of reference for which themes, sub-themes and/or key concepts are retained as potential insights or potential contributions (green), and which were found to support the existing scholarship (black).

6.4.4 RQ2: Discussion of Theme 4 – Antecedents.

This theme is a potentially new main theme which was not identified in the Chapter 2 literature review. It is reflected as a potential new theme in Figure 12 at the start of Chapter 5. The main topics discussed in the findings were "*Provenance for Trust*" and "*Importance of Data*". The key concepts which emerged from the main topics are compared to the literature in the sections which follow, and conclusions are made as to their potential contribution.

6.4.4.1 Recap of the findings on Antecedents.

The key concepts identified, which emerged from the findings were that *provenance of the AI is an antecedent to adoption* and also that *data integrity is an antecedent to adoption*. The significance of AI tool provenance and data integrity emerged as important factors in building trust in AI systems. Across the participants, there was consistent emphasis on the need for validated and credible data, as well as a proven track record, to trust AI systems. This shared focus highlighted the importance of ensuring reliable data inputs and a history of successful AI deployment to engendering trust. This alluded to trust in AI having the antecedents of both quality of data and the tool's demonstrated performance over time.

All groups shared this perspective, but there was a difference in their focus. Some emphasized the need for rigorously tested data and a proven historical reliability of the AI tool itself, linking trust to both data quality and the technology's track record. Others similarly prioritized data integrity but stressed the importance of understanding the data's origin, especially when external factors could influence its credibility. The remainder focused more on practical validation, requiring verifiable data and extensive testing to ensure reliable performance in their operations. These variations suggested that while provenance and data

integrity are similarly important, the emphasis shifted slightly between strategic oversight and operational use.

6.4.4.2 Recap of the literature on Antecedents.

There was no evidence presented in the Chapter 2 literature review on either "*Provenance for Trust*" or "*Importance of Data*" as either antecedents to AI trust, or as factors requiring explanation to the user specifically. Therefore, the findings were retained as potential differences and the 3-Step process, described in Section 6.1.1 above, was followed to analyse them both further.

6.4.4.3 Further discussion of the differences between findings and literature.

STEP 1: The articles which were identified for use in Step 1 were those cited in the related research question, as highlighted in Table 24: (Rai, 2020); (Sabharwal et al., 2024); (Lukyanenko et al., 2022). The keywords used in the search of the literature related to *Provenance for Trust* were "provenance", "prior success", "explanation of success", "antecedent" and "historic use". The word searches on the three selected articles did not yield any matches directly related to the key concept, or evidence of provenance of AI being an antecedent to trust formation, therefore the concept was retained as a potential difference and Step 2 was still performed. The keywords used in the search of the literature related to the *Importance of Data* were "data integrity", "data confidence", "data credibility", "antecedent" and "data verification". Although there were many matches related to data, the word searches on the three selected articles did not yield any matches directly related to the importance of data as an antecedent to trust, therefore Step 2 was still performed.

STEP 2: Considering the absence of both topics from the literature review, a more exhaustive Step 2 was followed by extending the word search from Step 1 to **not three, but six** pieces of additional literature, presented in the broader discussions from Chapter 2. Using the same keywords identified for Step 1 above, literature by Rabiee et al. (2024), Vanneste & Puranam (2024), Wang & Ding (2024), Berente et al. (2021), Weber et al. (2023), Kim et al. (2023) and Choung et al. (2022) was searched for any evidence of similarity. Vanneste & Puranam, (2024) did reveal matches to the keyword "integrity", but these were related to moral integrity of AI, not data integrity, and furthermore, not as an antecedent to trust. Both findings were therefore retained for further analysis in step 3.

STEP 3: For the final step, three top scholars from the related research question and theme were selected, and a Google Scholar search was performed for the keywords above, on their most recent publications from the past two years. The authors selected were Arun Rai, Nicholas Berente and Roman Lukyanenko. For all the three authors, no newer

publications in top journals were identified as related to the key concept, with the literature already included in Chapter 2 being the most recent.

6.4.4.4 Interpretation and conclusion on Antecedents.

For the two key concepts identified, *that provenance of the AI is an antecedent to adoption and also that data integrity is an antecedent to adoption*, no apparent similarity was found in the extant literature. This revealed potential new insights related to trust development in AI for its adoption and use in decision-making. These insights were therefore seen as nuanced differences to the extant literature and were therefore retained as potential new sub-themes under the potential new theme, "*Antecedents*". The potential new theme and the potential new sub-themes will all be included in the revised conceptual framework at the end of Chapter 6.

Table 40:

Summary of Themes, Sub-Themes and Key Concepts for RQ2, Theme 4

Themes from RQ2	Topics / Sub-Themes	Key Concepts
Theme 4: Antecedents	• *Provenance for Trust* • *Importance of Data*	• Provenance is an antecedent to adoption. • Data integrity is an antecedent to adoption.

Note: Author's Own.

The above table provides a visual summary for ease of reference for which themes, sub-themes and/or key concepts are retained as potential insights or potential contributions (green), and which were found to support the existing scholarship (black).

6.5 Research Question 3: *How does transparency influence emotional trust in AI for organisational decision-making?*

The aim of RQ3 is to gain additional insights and understanding of transparency as an emotional dimension of trust in AI, in the context of the study around organisational decision-making. There are three themes which were mapped directly to RQ3, with one potential new theme, all of which are discussed in this section. Table 41 below provides a summary of the themes, topics / sub-themes and key concepts from the Chapter 5 findings which will be discussed under RQ3.

Table 41:

Themes Emerging from RQ3 and their Key Concepts Identified in the Findings.

RQ3: *How does transparency influence emotional trust in AI for organisational decision-making?*		
Themes from RQ3	**Topics / Sub-Themes**	**Key Concepts Identified**
Theme 1: Transparency for Trust in AI	• *Fear as a Barrier* • *AI Identification of Human Nuance*	• Fear is a barrier to adoption and is primarily related to job security. • Transparency is required to alleviate fear and anxiety and foster trust. • AI has limited ability to identify human nuance, which inhibits trust.
Theme 2: Transparency for Decision-Making	• *Transparency and Understanding*	• Transparency and Understanding are closely related, directly linked to trust in AI for decision-making. • Different priorities for transparency, depending on role of the individual responsible for decision-making within the organization.
Theme 3: xAI and Transparency	• *Understanding of "How?"* • *Understanding and Trust*	• xAI and Transparency drive understanding, which promotes trust in AI decision-making. • Understanding "How" AI works and processes data is important for building trust.
Theme 4: Conjoined Agency	• *Human-AI Conjoined Agency* • *AI Assistance of Humans*	• AI is most effective and trusted as a supportive tool to enhance decision-making, which must be overseen by humans.

Note: Author's Own.

6.5.1 RQ3: Discussion of Theme 1 – Transparency for Trust in AI.

This theme is a main theme from the Chapter 2 literature review and is reflected in Figure 12 at the start of Chapter 5. The two main topics discussed in the findings were "*Fear as a Barrier*" and "*AI Identification of Human Nuance*". The key concepts which emerged from these main topics are compared to the literature in the sections which follow, and conclusions are made as to their potential contribution.

6.5.1.1 Recap of the findings on Transparency for Trust in AI.

Two key concepts related to the topic "Fear as a Barrier" emerged from the findings, namely: *Fear is a barrier to adoption and is primarily related to job security* and, *transparency is required to alleviate fear and anxiety and foster trust*. An additional key concept related to the topic "AI Identification of Human Nuance" emerged, that *AI has limited ability to identify human nuance, which inhibits trust*. Across the participant groups, fear was consistently

recognized as a barrier to AI adoption. This was particularly regarding job security, data privacy, and AI's ability to manage human relationships. All groups also similarly recognized that transparency and clear communication are essential to reducing these fears and fostering trust in AI. The similar narratives suggested that being transparent about AI's capabilities and limitations may reduce fear and is important to alleviating anxieties. In turn, this could pave the way for greater acceptance and trust of AI.

There was also a similar view that AI has limited ability to identify human nuance and that this is an inhibitor to trust. The findings suggested that addressing fears through better AI communication, transparency, and more relatable AI systems could build trust and pave the way for broader AI adoption across organizational contexts.

6.5.1.2 Recap of the literature on Transparency for Trust in AI.

The literature on transparency for trust in AI highlighted its important role in fostering trust in AI systems. Laato et al. (2022), Haque et al. (2023), and Glikson & Woolley (2020) provided an overview of how transparency and trust are interrelated, identifying transparency as a key goal of xAI. Laato et al. (2022) found that a lack of transparency negatively affects AI's trustworthiness, while Haque et al. (2023) considered transparency and trust as outcomes of xAI. Glikson & Woolley (2020) framed transparency as important for building cognitive trust but noted that its emotional dimensions were underexplored. Sabharwal et al. (2024) emphasized that transparency serves as the foundation for trust in AI and argued for the centrality of human values in AI (p. 6). Additionally, Wang & Ding (2024) called for further research into how transparency influences human trust in AI across different decision-making contexts, while Sullivan et al. (2022) similarly suggested that transparency may address an emotional discomfort or "uncanny feeling" in trusting AI, calling for further research. Lastly, Choung et al. (2022) underscored transparency as an essential ethical requirement for AI adoption. Across the literature, scholars agreed on the necessity of transparency to build trust in AI, but also identified gaps, particularly in exploring its emotional dimensions.

6.5.1.3 Discussion of the findings and literature.

In the findings, decision-making groups acknowledged AI's inability to identify Human nuance, and the need for AI to understand and manage human relationships. This was similar to the literature by Sabharwal et al. (2024) who argued for centrality of human values in AI. The topic of "*AI Identification of Human Nuance"* therefore supports Sabharwal et al. (2024) who stated, with respect to trust in AI, that "if humans cannot connect to how machines operate, trust falters" (p. 6). By extension, the lack of ability of AI to identify human nuance might potentially be alleviated by transparency, by allowing the connection of humans to how AI operates. There was also further similarity between the findings and Laato et al. (2022) and

Haque et al. (2023) that transparency is important for enhancing the trustworthiness of AI. What was identified by the findings, but not specifically mentioned in the literature review in Chapter 2 was the concept that *fear, particularly of job security, is a barrier to AI adoption*. A further nuance of difference and therefore a potential insight between the findings and Chapter 2 literature review was that *transparency is required to alleviate fear and anxiety and foster trust.* Both findings were therefore retained as potential nuances of difference to the literature and the 3-Step process, described in Section 6.1.1 above, was followed to analyse them both further.

6.5.1.4 Further discussion of the differences between findings and literature.

STEP 1: The articles which were identified for use in Step 1 were those cited in the related research question, as highlighted in Table 24: (Glikson & Woolley, 2020); (Wang & Ding, 2024) and (Berente et al., 2021). The keywords used in the search of the literature, as related to the key concepts and the topic "*Fear as a Barrier*" were "fear of AI", "job security", "fear barrier", "anxiety" and "transparency to alleviate". The word searches on the three selected articles revealed some similarity to Glikson & Woolley, (2020) who identified fear as a negative emotion related to emotional trust in robotic AI. They also called for further research into understanding how transparency influences emotional trust in AI for organisational decision-making (the focus of RQ3). However, they did not discuss the link revealed in the findings that transparency may be required to alleviate anxiety and foster trust in AI, and this was retained as a potential insight for further exploration in Step 2. An additional similarity was identified in Berente et al. (2021) that AI and "any new frontier of automation has brought forth fears of deskilling and labour substitution" (p. 1444). Therefore, the concept of "*fear of job security*" was similar to and supported extant literature and was not discussed further in subsequent steps.

STEP 2: Considering the concept of "*transparency is required to alleviate anxiety and foster trust"* being potentially different to extant literature in relation to the topic of "*Fear as a Barrier*", Step 2 was followed by extending the word search from Step 1 to four pieces of additional literature presented in the discussions from Chapter 2. Using the same keywords "fear of AI", "job security", "fear barrier", "anxiety" and "transparency to alleviate", literature by Laato et al. (2022), Haque et al. (2023) and Choung et al. (2022) was searched for any evidence of similarity, with none revealed. The finding was therefore retained for further analysis in Step 3.

STEP 3: For the final step, three top scholars from the related research question and theme were selected, and a Google Scholar search was performed for the keywords above, on their most recent publications from the past two years. The authors selected were Anita

Woolley, Samuli Laato and Nicholas Berente. For all of the three authors, no newer publications in top journals were identified, with the literature already included in Chapter 2 once again being the most recent.

6.5.1.5 Interpretation and conclusion on Transparency for Trust in AI.

For the first key concept identified in the findings, *Fear is a barrier to adoption and is primarily related to job security*, similarity was found between the findings and the extant literature, thus confirming existing discussions in the scholarship. This was therefore seen as a similarity to the extant literature and was not retained as a potential new sub-theme.

For the second key concept identified in the findings, that *transparency is required to alleviate fear and anxiety and foster trust*, no apparent similarity was found between the findings and the extant literature. Furthermore, this finding might have the potential to provide additional insights to the call for further research by Glikson & Woolley, (2020) considered in RQ3. Glikson & Woolley, (2020) specifically identify fear as an important human emotion. The observation that transparency might alleviate fear and anxiety therefore provides direct insight as to how transparency influences emotional trust in AI for organisational decision-making (RQ3). This finding was therefore retained as a potential new sub-theme and will be added to the conceptual framework at the end of Chapter 6. The initial name given to the sub-theme, "*Fear as a Barrier*" was also adjusted to "***Transparency to Mitigate Fear and Foster in AI***", to better reflect the identified concept/finding.

For the third and final key concept identified in the findings, *AI has limited ability to identify human nuance, which inhibits trust,* similarity was found in the extant literature and supports Sabharwal et al. (2024). The related topic was not carried forward as a potential new sub-theme. It must be noted, however, that the lack of ability of AI to identify human nuance might potentially be alleviated by transparency, by allowing the connection of humans to how AI operates.

Table 42:

Summary of Themes, Sub-Themes and Key Concepts for RQ3, Theme 1

Themes from RQ3	Topics / Sub-Themes	Key Concepts
Theme 1: Transparency for Trust in AI	• *Fear as a Barrier* • *Transparency to Mitigate Fear of AI* • *AI Identification of Human Nuance*	• Fear is a barrier to adoption and is primarily related to job security. • Transparency is required to alleviate fear and anxiety and foster trust. • AI has limited ability to identify human nuance, which inhibits trust.

Note: Author's Own.

The above table provides a visual summary for ease of reference for which themes, sub-themes and/or key concepts are retained as potential insights or potential contributions (green), and which were found to support the existing scholarship (black).

6.5.2 RQ3: Discussion of Theme 2 – Transparency for Decision-Making.

This theme is a main theme from the Chapter 2 literature review and is reflected in Figure 12 at the start of Chapter 5. The main topic discussed in the findings was "*Transparency and Understanding*". The key concept which emerged from this main topic is compared to the literature in the sections which follow, and conclusions are made as to its potential contribution.

6.5.2.1 Recap of the findings on Transparency for Decision-Making.

There were two key concepts related to the topic "*Transparency and Understanding*" which emerged from the findings, namely: that *Transparency and Understanding are closely related and are directly linked to trust in AI for decision-making.* Secondly, that there are *different priorities for transparency, depending on the role of the individual responsible for decision-making within the organization*. All participant groups emphasized the importance of transparency in fostering trust in AI for decision-making. All groups agreed that transparency is important for understanding how AI functions, which should allow for better risk assessment and informed decision-making. The consistent message was that without transparency, trust cannot be established, as users need to understand how AI processes data and produces outcomes to feel confident in relying on its decisions.

There was some difference in how each group viewed transparency. Some saw transparency as essential to understanding AI's functionality, directly linking it to trust in decision-making. Others highlighted the need for transparency to provide clarity, but they also stressed that transparency must lead to meaningful understanding for it to be truly effective. Participants responsible for operational decision-making focused more on practical understanding, emphasizing that trust at the operational level depends on clear and easily understood explanations of AI processes. These differences suggested that while all groups valued transparency, they prioritized it in slightly different ways depending on their role within the organization.

6.5.2.2 Recap of the literature on Transparency for Decision-Making.

The literature on transparency for AI decision-making underscored the need for transparency, to enhance confidence and trust in AI's decision-making processes. Lindebaum et al. (2020) highlighted lack of transparency and understanding as a significant limitation and potential liability in AI decision-making (p. 256), emphasizing data quality and self-reinforced negative feedback as key vulnerabilities of AI algorithms. Kim et al. (2023) recommended raising standards of transparency and accountability, especially with new data protection regulations, while Vanneste & Puranam (2024) argued that transparency clarifies the reasoning and understanding behind an AI algorithm's decisions. Sabharwal et al. (2024) were of a similar view that transparency is essential for responsible AI adoption. However, Vanneste & Puranam (2024) and Glikson & Woolley (2020) also pointed out that the increasing

complexity of AI systems complicates transparency, making it even more important. There was broad similarity in the literature that transparency strengthens trust and is necessary for effective AI decision-making.

6.5.2.3 Discussion of the findings and literature.

The findings emphasized that transparency allows users to comprehend AI processes, which facilitates trust and informed decision-making across all levels. This aligned with the literature, where Lindebaum et al. (2020) and Vanneste & Puranam (2024) similarly identified transparency and understanding as important for building trust, particularly by clarifying AI's reasoning and improving confidence in its decision-making. The concept that *Transparency and Understanding are closely related and are directly linked to trust in AI for decision-making* was therefore similar to and supported extant literature and was not carried forward to the 3-Step process. However, the findings introduced a potential new insight regarding *different priorities for transparency, depending on the role of the individual responsible for decision-making within the organization*. This key concept was not explicitly mentioned or revealed in the Chapter 2 literature. The literature emphasized more general themes, such as data quality (Lindebaum et al., 2020) and the need for transparency in complex AI systems (Glikson & Woolley, 2020). There was therefore a difference to extant literature in that it did not mention how transparency requirements may differ across organizational roles. This finding was therefore retained as potential nuance of difference to the literature and the 3-Step process, described in Section 6.1.1 above, was followed to analyse it further.

6.5.2.4 Further discussion of the differences between findings and literature.

STEP 1: The articles which were identified for use in Step 1 were those cited in the related research question, as highlighted in Table 24: (Glikson & Woolley, 2020); (Wang & Ding, 2024) and (Berente et al., 2021). The keywords used in the search of the literature, as related to the key concept and the topic "*Transparency and Understanding*" were "transparency priority", "transparency focus", "transparency levels", "transparency requirements" and "transparency differences", as well as a much broader keyword, "role". The latter keyword was included to search for any mention of transparency in relation to the word "role". The word searches on the three selected articles revealed no matches, therefore Step 2 was performed.

STEP 2: Considering the concept of "*different priorities for transparency, depending on the role of the individual responsible for decision-making within the organization"* being potentially different to extant literature in relation to the topic of "*Transparency and Understanding*", Step 2 was followed by extending the word search from Step 1 to four pieces of additional literature presented in the discussions from Chapter 2. Using the same keywords

"transparency priority", "transparency focus", "transparency levels", "transparency requirements", "transparency differences" and "role", literature by Kim et al. (2023); Lindebaum et al. (2020); Vanneste & Puranam (2024) was searched for any evidence of similarity, with none revealed. The finding was therefore retained for further analysis in Step 3.

STEP 3: For the final step, three top scholars from the related research question and theme were selected, and a Google Scholar search was performed for the keywords above, on their most recent publications from the past two years. The authors selected were Anita Woolley, Samuli Laato and Nicholas Berente. For all of the three authors, no newer publications in top journals were identified, with the literature already included in Chapter 2 being the most recent.

6.5.2.5 Interpretation and conclusion on Transparency for Decision-Making.

For the first key concept identified in the findings, *Transparency and Understanding are closely related and are directly linked to trust in AI for decision-making*, similarity was found between the findings and the extant literature, thus supporting existing discussions in the scholarship. This was therefore seen as a similarity to the extant literature and was not retained as a potential new sub-theme. For the second key concept identified in the findings, that there are *different priorities for transparency, depending on the role of the individual responsible for decision-making within the organization*, no apparent similarity was found between the findings and the extant literature. The initial name given to the sub-theme, "*Transparency and Understanding*" was also adjusted to "***Role-Dependent Transparency***", to better reflect the identified concept/finding. This was also therefore retained as a potential new sub-theme and will be added to the conceptual framework at the end of Chapter 6.

Table 43:

Summary of Themes, Sub-Themes and Key Concepts for RQ3, Theme 2

Themes from RQ3	Topics / Sub-Themes	Key Concepts
Theme 2: Transparency for Decision-Making	• *Transparency and Understanding* • *Role-Dependent Transparency*	• Transparency and Understanding are closely related, directly linked to trust in AI for decision-making. • Different priorities for transparency, depending on the role of the individual responsible for decision-making within the organization.

Note: Author's Own.

The above table provides a visual summary for ease of reference for which themes, sub-themes and/or key concepts are retained as potential insights or potential contributions (green), and which were found to support the existing scholarship (black).

6.5.3 RQ3: Discussion of Theme 3 – xAI and Transparency.

This theme is a main theme from the Chapter 2 literature review and is reflected in Figure 12 at the start of Chapter 5. The two main topics discussed in the findings were "*Understanding of "How?"*" and "*Understanding and Trust*". The key concepts which emerged from these main topics are compared to the literature in the sections which follow, and conclusions are made as to their potential contribution.

6.5.3.1 Recap of the findings on xAI and Transparency.

A key concept related to the topics, which emerged from the findings, was that *xAI and transparency drive understanding, which promotes trust in AI decision-making.* An additional key concept from the findings which relates to the topics, was that *Understanding "How" AI works and processes data is important for building trust.* There was a similarity of recognition across all participant groups that transparency and explanation to understand "how" AI works are important for building trust. Furthermore, the groups showed similarity that explanation to understand how AI processes data, its algorithms, and the logic behind its outputs are important factors for trust. The shared emphasis on transparency and understanding highlighted the important connection between explanation to understand AI's internal workings and developing trust in its use.

There were slight differences in focus among the groups. Some prioritized technical clarity, focusing on understanding the algorithms and logic behind AI, to explain AI decision-making. Others emphasized the manipulation and framing of data, with transparency seen as necessary for assessment of AI's processes and outputs. The remainder focused on practical understanding that aligned with business requirements and highlighted the need for continuous observation over time. These variations suggest that for xAI to be effective, it must provide clear and transparent explanations tailored to the needs and priorities of each group to create understanding.

6.5.3.2 Recap of the literature on xAI and Transparency.

The literature on the relationship between xAI and transparency highlighted that while the two concepts are closely related, scholars positioned them in slightly different ways. Sabharwal et al. (2024), Rai (2020) and Gregor (2024) discussed them interchangeably, whereas Glikson & Woolley, (2020), Haque et al. (2024) and Laato et al. (2022) presented transparency as an outcome or implication of explainability. Kim et al. (2023) also called for further interdisciplinary research to understand the outcome areas of xAI (p. 1306). Glikson & Woolley (2020) identified explainability as an "important aspect of transparency" (p. 631), while Gregor (2024) treated them as separate but related "principles" (p. 52). Despite these varying perspectives, there was similarity in the scholarship that both concepts shared the same

overarching objective: transitioning opaque "black box" AI systems into "glass box" systems (Rai, 2020, p. 139). This shared goal emphasized the important role of xAI and transparency in enhancing trust and accountability in AI decision-making.

6.5.3.3 Discussion of the findings and literature.

There were similarities between the findings and the literature on xAI and transparency, in recognizing that both concepts play an important role in trust in AI decision-making. For the concept that *xAI and transparency drive understanding, which promotes trust in AI decision-making,* both the findings and literature emphasized that explainability (xAI) and transparency work together (Sabharwal et al., 2024; Rai, 2020; Gregor, 2024), in building trust. However, a nuance of difference emerged between the findings and literature that explainability (xAI) and transparency drive understanding, which then engenders trust. For the second key concept that *understanding "How" AI works and processes data is important for building trust*, the literature review in Chapter 2 did not identify any similarity as discussed. Therefore, both findings were retained as potential differences to the literature and the 3-Step process, described in Section 6.1.1 above, was followed to analyse them further.

6.5.3.4 Further discussion of the differences between findings and literature.

STEP 1: The articles which were identified for use in Step 1 were those cited in the related research question, as highlighted in Table 24: (Glikson & Woolley, 2020), (Wang & Ding, 2024) and (Berente et al., 2021). The keywords used in the search of the literature, as related to the key concepts, were "understanding", "understanding how" and "understanding and trust". The word searches on the three selected articles revealed some similarity to Glikson & Woolley, (2020, p. 641) who discuss the importance of "understanding how", in relation to explanations, transparency and trust building. There was therefore similarity to and support of extant literature and the topic "Understanding of "How?" was not explored further. However, Glikson & Woolley, (2020) do not identify the concept from the findings that *xAI and transparency drive understanding, which then leads to trust*. No additional matches were found in the other articles either, therefore this finding was retained for further analysis in Step 2.

STEP 2: Considering the concept that "*xAI and transparency drive understanding, which promotes trust in AI decision-making"* being potentially different to extant literature in relation to the topic of "*Understanding and Trust*", Step 2 was followed by extending the word search from Step 1 to three pieces of additional literature presented in the discussions from Chapter 2. Using the same keywords "understanding", "understanding how" and "understanding and trust", literature by Rai (2020), Laato et al. (2022) and Haque et al. (2023) was searched for any evidence of similarity. A match was found in Laato et al. (2022), who discussed transparency of an AI system enabling users to "better understand it" (p. 10), and

they also identified that xAI should enable users to "understand" the systems which they are using.

Note: This interestingly also supported the findings from RQ2, Theme 1 regarding the "interplay"/circularity/cross-support of the three concepts xAI, transparency and understanding.

6.5.3.5 Interpretation and conclusion on xAI and Transparency.

For the first key concept identified in the findings, *xAI and transparency drive understanding, which promotes trust in AI decision-making*, similarity was found between the findings and the extant literature, thus supporting existing discussions in the scholarship. Although the exact mechanism of xAI and transparency driving understanding (then leading to trust) was not identified specifically, there was similarity that transparency leads to understanding or explanation (xAI) leads to understanding. This was therefore seen as a similarity to the extant literature and was not retained as a potential new sub-theme.

However, for the second key concept identified in the findings, that *Understanding "How" AI works and processes data is important for building trust*, similarity was found between the concept literature, therefore supporting discussion in the scholarship. Neither of the topics "*Understanding of "How?"*" and "*Understanding and Trust*" were therefore included in the Conceptual Framework as potential contributions, but the "interplay"/circularity/cross-support of the three concepts xAI, transparency and understanding was noted as potentially insightful.

Table 44:

Summary of Themes, Sub-Themes and Key Concepts for RQ3, Theme 3

Themes from RQ2	Topics / Sub-Themes	Key Concepts
Theme 3: xAI and Transparency	• *Understanding of "How?"* • *Understanding and Trust*	• xAI and Transparency drive understanding, which promotes trust in AI decision-making. • Understanding "How" AI works and processes data is important for building trust.

Note: Author's Own

The above table provides a visual summary for ease of reference for which themes, sub-themes and/or key concepts are retained as potential insights or potential contributions (green), and which were found to support the existing scholarship (black).

6.5.4 RQ3: Discussion of Theme 4 – Conjoined Agency.

This theme is a potentially new main theme which was not identified in the Chapter 2 literature review. It is reflected as a potential new theme in Figure 12 at the start of Chapter 5. The main topics discussed in the findings were "*Human-AI Conjoined Agency*" and "*AI Assistance of Humans*". The key concept which emerged from the main topics is compared to

the literature in the sections which follow, and conclusions are made as to its potential contribution.

6.5.4.1 Recap of the findings on Conjoined Agency.

The key concept identified, which emerged from the findings was that *AI is most effective and trusted as a supportive tool to enhance decision-making, which must be overseen by humans*. These themes are closely related and emphasize trust in AI through a collaborative, supportive role. All participant groups shared similar views that AI is most effective and trustworthy when it enhances human decision-making, while remaining under human oversight and control. This shared agency reinforced their similar view that AI should serve as a tool to augment human capabilities, not to replace them, which fosters trust and confidence in AI's use. Across all groups, the collaborative approach of AI functioning as an assistant, rather than a replacement, emerged as central to building trust in AI-driven decision-making.

Some difference emerged in how this collaboration was perceived. Some participants focused on AI's role in improving efficiency and supporting decision-making processes. However, they also advocated maintaining strict human control. Others highlighted the idea of AI as a "co-pilot," working in tandem with humans, where AI assists but requires human oversight for accuracy and accountability. The remainder emphasized AI's practical application in streamlining tasks and optimizing workflows, stressing that trust is built when AI integrates into existing human processes without threatening job roles. These perspectives collectively highlighted the importance of conjoined agency in reinforcing AI-human, collaborative partnership across organizational levels.

6.5.4.2 Recap of the literature on Conjoined Agency.

There was no evidence presented in the Chapter 2 literature review on either “*Human-AI Conjoined Agency*” or “*AI Assistance of Humans*”. Therefore, the findings were retained as potential differences and the 3-Step process, described in Section 6.1.1 above, was followed to analyse them both further.

6.5.4.3 Further discussion of the differences between findings and literature.

STEP 1: The articles which were identified for use in Step 1 were those cited in the related research question, as highlighted in Table 24: (Glikson & Woolley, 2020); (Wang & Ding, 2024) and (Berente et al., 2021). The keywords used in the search of the literature, related to *Conjoined Agency* were “conjoined”, “assistance”, “assistant”, “shared”, “agency”, “augment” and “support”. Matches were found in Berente et al. (2021) against the keyword “conjoined”, referring to “conjoined agency between AI and humans” as a contemporary frontier in facets of AI autonomy (p. 1443). The reference list in Berente et al. (2021) also

revealed an article in the 4* *Academy of Management Review* journal by Murray, Rhymer & Sirmon (2021) titled “Humans and Technology: Forms of Conjoined Agency in Organizations”. Further searches of the authors in the aforementioned journal revealed an active discourse in extant literature, related to both *Human-AI Conjoined Agency* and *AI Assistance of Humans*. Therefore, the additional analysis steps were not followed, as the findings supported an existing scholarly discussion.

6.5.4.4 Interpretation and conclusion on Conjoined Agency.

For the key concept identified, that *AI is most effective and trusted as a supportive tool to enhance decision-making, which must be overseen by humans*, similarity was found in the extant literature to the topic of conjoined agency. The insight was therefore seen as similar to, and supporting the extant literature, and neither the theme nor its topics/sub-themes will be presented as potentially new in the revised conceptual framework at the end of Chapter 6.

Table 45:

Summary of Themes, Sub-Themes and Key Concepts for RQ3, Theme 4

Themes from RQ2	Topics / Sub-Themes	Key Concept
Theme 4: Conjoined Agency	• *Human-AI Conjoined Agency* • *AI Assistance of Humans*	• AI is most effective and trusted as a supportive tool to enhance decision-making, which must be overseen by humans.

Note: Author’s Own.

The above table provides a visual summary for ease of reference for which themes, sub-themes and/or key concepts are retained as potential insights or potential contributions (green), and which were found to support the existing scholarship (black).

6.6 Chapter Conclusion

This chapter used a process of systematic comparison between the Chapter 5 findings and the literature, using the 3-Steps presented. Of the initial 13 themes and 23 topics / sub-themes identified in Chapter 5, three (3) themes were retained as potentially contributing new insights, as well as 12 of the topics / sub-themes. The remaining themes and sub-themes were confirmed as part of the extant literature, therefore supporting and adding to the body of knowledge. Table 46, below, provides an overview summary of the outcomes of the Chapter 6 discussion.

Table 46:

Summary of the Outcomes of Comparison Between the Findings and Literature.

RQ1: *How does trust in AI lead to its adoption and use for organisational decision-making?*		
Themes from RQ1	**Topics / Sub-Themes**	**Key Concepts Identified**

Theme 1: Trust in Artificial Intelligence	• *Trust Develops over Time* • *Trust as a Spectrum* • *Trust Rates in Industries*	• Trust takes time to develop. • Trust exists as a spectrum or on a scale. • Trust develops at different rates in different industries.
Theme 2: Trust in Artificial Intelligence for Adoption and Use	• *Early Adoption & Experimentation* • *Change Management Engenders Trust*	• Importance of gradual AI adoption and experimentation • Importance of change management for trust in AI.
Theme 3: Trust in Artificial Intelligence for Decision-Making	• *Only to Inform Decisions* • *Output Reliability & Trust*	• AI should inform and enhance human decision-making, not replace it. • Human verification of AI outputs is needed.
Theme 4: "Blind-Trust" Tensions	• *"Blind-Trust" Tensions*	• There is a tension between the Decision-Making groups regarding "blindly-trusting" AI. Some support it while others reject it.
Theme 5: "Blast-Radius"	• *"Blast-Radius"*	• Importance of containing the potential negative outcomes of AI use for decision-making, to engender trust.

RQ2: *How does xAI engender trust in AI for decision-making in organisations? And: What are the factors that contribute to trust in AI?*		
Themes from RQ2	**Topics / Sub-Themes**	**Key Concepts Identified**
Theme 1: xAI and the "Black-Box"	• *The "Black-Box"* • *Understanding and the "Black-Box"* • *Transparency and the "Black-Box"*	• Awareness and shared concern of the "Black -Box" nature of AI. • Transparency, understanding and explainability were all recognised to build trust and overcome the "Black-Box".
Theme 2: xAI for Trust in AI	• *Explanation of Data* • *Explanation of "How?"*	• Explanation must be consistent throughout the data lifecycle. • Explainability should address "How" AI arrives at outcomes.
Theme 3: xAI for Decision-Making	• *Self-Criticism of AI*	• AI Self-Criticism is necessary for trust and confidence in AI systems for decision-making.
Theme 4: Antecedents	• *Provenance for Trust* • *Importance of Data*	• Provenance is an antecedent to adoption. • Data integrity is an antecedent to adoption.

RQ3: *How does transparency influence emotional trust in AI for organisational decision-making?*		
Themes from RQ3	**Topics / Sub-Themes**	**Key Concepts Identified**
Theme 1: Transparency for Trust in AI	• *Fear as a Barrier* • *Transparency to Mitigate Fear of AI* • *AI Identification of Human Nuance*	• Fear is a barrier to adoption and is primarily related to job security. • Transparency is required to alleviate fear and anxiety and foster trust. • AI has limited ability to identify human nuance, which inhibits trust.
Theme 2: Transparency for Decision-Making	• *Transparency and Understanding* • *Role-Dependent Transparency*	• Transparency and Understanding are closely related, directly linked to trust in AI for decision-making. • Different priorities for transparency, depending on the role of the individual responsible for decision-making within the organization.
Theme 3: xAI and Transparency	• *Understanding of "How?"* • *Understanding and Trust*	• xAI and Transparency drive understanding, which promotes trust in AI decision-making. • Understanding "How" AI works and processes data is important for building trust.
Theme 4: Conjoined Agency	• *Human-AI Conjoined Agency* • *AI Assistance of Humans*	• AI is most effective and trusted as a supportive tool to enhance decision-making, which must be overseen by humans.

Note: Author's Own.

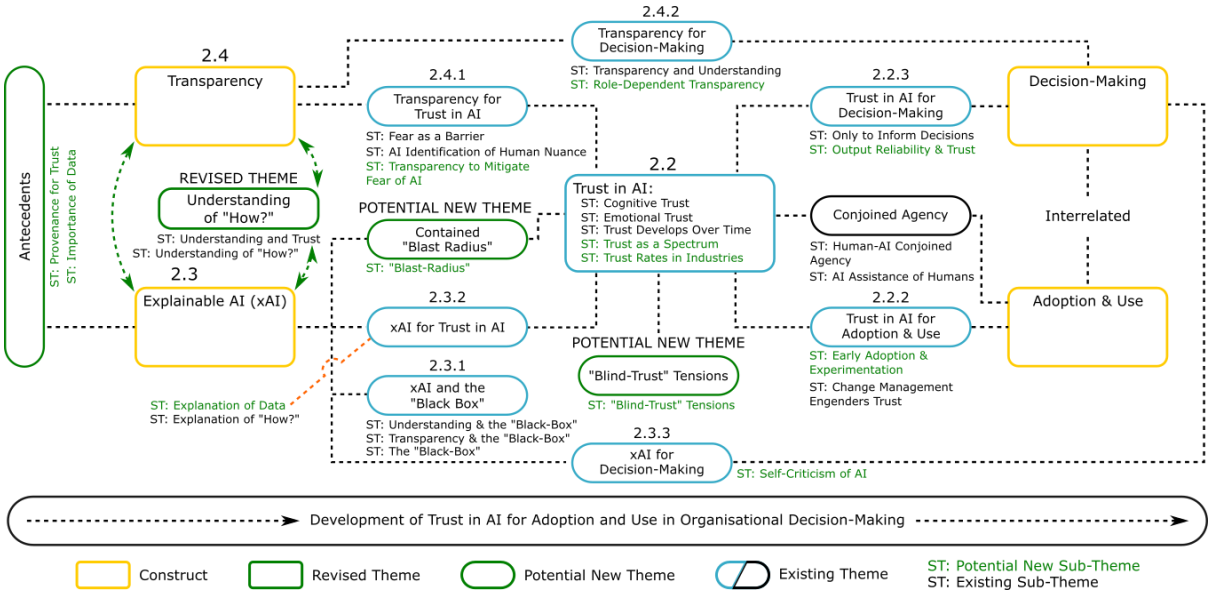
The above table provides a visual summary (for ease of reference) of which themes, sub-themes and/or key concepts are retained as potential insights or potential contributions (green), and which were found to support the existing scholarship (black).

6.7 Revised Conceptual Framework Following Discussion with Literature

The themes and sub-themes identified in Table 46 above are shown as a visual summary in the revised Conceptual Framework provided below as Figure 28. This final conceptual framework will be simplified in Chapter 7 as a more presentable and user-centric practical overview framework. The revision now depicts three (3) potentially new themes, one (1) revised theme, and 12 potentially new sub-themes in (green), with the rest having been identified as part of the extant literature and the body of knowledge (shown in Black) through the 3-step process which was followed.

Figure 28:

Revised Conceptual Framework, Following the Chapter 6 Discussion



Note: Author's Own

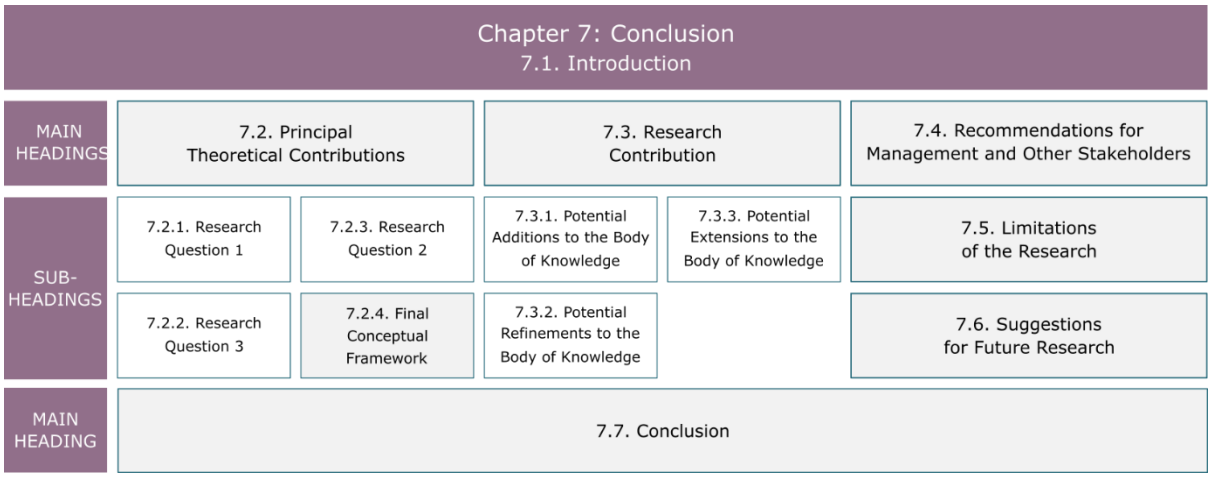
Chapter 7: Conclusion

7.1 Introduction

As with the preceding chapters, this chapter is organized by research question for ease of navigation. The theoretical conclusions for each of the research questions are discussed in terms of their similarities, nuances of difference and distinct differences, to highlight the potential extensions or refinements extant literature which were revealed in Chapter 6. Recommendations for management are then provided, followed by the limitations of this study and subsequent suggestions for further research. An outline of the sections and sub-sections, of the Chapter 7 conclusion which follows, is provided in Figure 29 below.

Figure 29:

Matrix of Sections and Sub-Sections for Chapter 7



Note: Author's Own.

The purpose of this chapter is to present a consolidated view of the research outcomes which were revealed through the comparative analysis of the findings and literature in Chapter 6. The overall purpose of the research was to gain additional insights and understanding of trust in artificial intelligence for its adoption and use in organisational decision-making. The setting of the research was worldwide organisations across diverse sectors, including but not limited to, financial services, healthcare, manufacturing and telecommunications. Three main participant groups were included in the research, namely, Executive, Senior and Operational individuals responsible for decision-making. All of the participants selected for the semi-structured interviews had exposure to (or experience of) artificial intelligence and its influence on their organisation.

Following presentation of the research outcomes, the final conceptual framework, which was developed as a build throughout the preceding chapters, (building up to Figure 28 in the Chapter 6 conclusion) is edited and included as a visual landscape of the research outcomes in Figure 30, in Section 7.2.4.

7.2 Principal Theoretical Conclusions

7.2.1 Research Question 1: How does trust in AI lead to its adoption and use for organisational decision-making?

The aims of RQ1 were to develop new insights and understanding into the mechanisms of trust formation in AI in organisations, and to therefore develop new insights as to how trust in AI leads to adoption and beneficial use of AI for organisational decision-making. The research outcomes for RQ1 were structured into five main themes. Three of the themes were derived from the Chapter 2 literature review and were mapped directly to RQ1, while the remaining two themes were identified as being potentially new. All five of the identified themes are discussed in this section, and the similarities, nuances of difference and distinct differences to literature are presented collectively, as related to Research Question 1.

7.2.1.1 Similarities to Literature for Research Question 1

Concerning the theme, “trust in artificial intelligence”, the research outcomes were consistent with the literature regarding *trust taking time to develop*. Glikson & Woolley, (2020) described trust as either increasing or decreasing over time, based on the type of AI concerned (Robotic, Virtual or Embedded). Both the findings and the literature similarly identified trust development over time as an important concept, which is therefore retained as an expected outcome, supporting the existing discourse.

Regarding the theme, “trust in artificial intelligence for adoption and use”, the research outcomes were consistent with the literature that *change management is important to trust in AI*. Enholm et al. (2021) who identified a theme of organisational structure and business process change in relation to employee-AI trust. The findings and the literature similarly identified the importance of change management to trust engenderment, which is therefore retained as an expected outcome, supporting the existing discourse.

With respect to the theme, “trust in artificial intelligence for decision-making”, the research outcomes were consistent with the literature that *AI should only inform and enhance human decision-making.* Enholm et al. (2021) and Chen et al. (2021) both emphasized the role of AI in decision-support, with Enholm et al. (2021) focusing on “human-in-the-loop” type systems, and Chen et al. (2021) emphasizing the focus of AI being an “assistive role” which supports, rather than replacing humans (p. 1720). The findings and the literature similarly identified that AI should only inform and enhance human decision-making, and this is therefore retained as an expected outcome, supporting the existing discourse.

7.2.1.2 Nuances of Difference to Literature for Research Question 1

In relation to the theme, “trust in artificial intelligence”, (Glikson and Woolley, 2020; Chen et al., 2021; Wang & Ding, 2024; Weber et al., 2023), two nuances of difference were identified. Firstly, the finding that trust in AI exists on a spectrum or scale showed some similarity to Glikson & Woolley, (2020)’s description of the cognitive and emotional aspects of trust, but described trust differently as a much broader, multi-variate concept, not just contained to emotional and cognitive dimensions. The second nuance identified was that trust develops at different rates in different industries, and this was also not found in Chapter 2 literature. Both were therefore identified as areas of potential refinement to the literature and are captured as the potential new sub-themes “Trust as a Spectrum” and “Trust Rates in Industries”.

With respect to the theme, “trust in artificial intelligence for adoption and use” (Wang & Ding, 2024; Weber et al., 2023; Glikson and Woolley, 2020), the importance of gradual AI adoption and experimentation was revealed as an apparent nuance of difference to the literature which was reviewed. This outcome is therefore captured as a potential new sub-theme “Early Adoption and Experimentation”.

An additional nuance of difference with regard to the theme, “trust in artificial intelligence for decision-making” (Chen et al., 2021; Wang & Ding, 2024; Sabharwal et al., 2024; Shrestha et al., 2021), was identified. The need for human verification of AI outputs to engender trust did not appear in the literature which was reviewed and is therefore retained as the potential new sub-theme “Output Reliability and Trust”.

In addition, another insight related to human verification of AI outputs was that, through human verification, trust concerns can be overcome. This is highlighted because Shrestha et al. (2019) and Chen et al. (2021) recognised that overcoming trust concerns was not covered in extant literature. Human verification of AI outputs may therefore provide a potential additional insight and new understanding in this regard.

7.2.1.3 Distinct Differences to Literature for Research Question 1

For the potentially new theme, “blind-trust tensions”, a key concept was identified, that there is a tension between decision-making groups regarding “blindly-trusting” AI. No apparent similarity was found in the extant literature, thus revealing a potential new insight related to trust development in AI for its adoption and use in decision-making. This was therefore seen as a distinct difference to the extant literature and is retained as a potential new theme, “Blind-Trust Tensions”.

For the potentially new theme, “blast-radius”, a key concept was identified, regarding the importance of containing the potential negative outcomes of AI use in decision-making, to

engender trust. No apparent similarity was found in the extant literature, thus revealing an additional potential new insight related to trust development in AI for its adoption and use in decision-making. This insight was therefore seen as a distinct difference to the extant literature and is retained as a potential new theme, “Blast-Radius”.

This potential new theme, “Blast-Radius”, highlighted a key concept in the need to contain the potential negative outcomes of AI use in decision-making, to engender trust. This is highlighted because Sullivan et al. (2022) recognised the need for further research to “identify and mitigate potential risks associated with artificial agents” (p. 542). The concept of containment of the “blast-radius” related to AI adoption and use may therefore also provide additional insights or understanding in relation to this extant literature.

7.2.1.4 Conclusions for Research Question 1

In conclusion to Research Question 1, the study identified several similarities to existing literature, including the gradual development of trust in AI, the importance of change management for fostering trust, and the role of AI as a support tool for human decision-making rather than a replacement. These similarities are offered as potential additions to the body of knowledge.

Nuances of difference were also revealed and highlighted further potential new insights, such as trust in AI existing on a spectrum, varying trust development rates across industries, the importance of gradual AI adoption, and the necessity for human verification of AI outputs to reinforce trust. The nuances of difference are offered as areas of potential refinement to the extant literature.

Distinct differences from the literature included insights into "blind-trust tensions," which highlighted differences in trust approaches across decision-making groups, and also the “blast-radius” concept, which emphasised the importance of containing AI’s potential negative impacts to build trust. These distinct differences are offered as potential extensions to the body of knowledge.

The similarities are retained as additions to the body of knowledge in the final conceptual framework (Figure 30), the nuances of difference are included as potential new sub-themes, and the distinct differences are included as potential new themes.

The research outcomes addressed the aims of RQ1 by developing new insights and understanding into the mechanisms of trust formation in AI in organisations, and also by developing new insights and understanding as to how trust in AI leads to its adoption and use for organisational decision-making. The potentially new insights and understanding which were revealed in answering RQ1 are summarised below:

Research Question 1: How does trust in AI lead to its adoption and use for organisational decision-making?

- When trust is developed gradually, is included in a change management process, and acts as a support tool for individuals responsible for decision-making, it potentially leads to its adoption and use for organisational decision-making.
- Understanding that trust potentially exists on a spectrum and develops at different rates in different industries provides potentially new insights into how trust in AI leads to adoption and use for organisational decision-making.
- Developing trust gradually and using humans to verify its outputs also potentially creates trust in AI, which can lead to adoption and use for organisational decision-making.
- Trusting that AI can contain its possible negative impacts potentially engenders trust for its adoption and use for organisational decision-making, while blindly trusting AI can lead to adoption, but may also cause unintended consequences.

7.2.2 Research Question 2: How does xAI engender trust in AI for decision-making in organisations? And: *What are the factors that contribute to trust in AI?*

The aims of RQ2 were to firstly explore and understand how xAI contributes to the formation of trust in AI for organisational decision-making. Secondly, to explore additional insights and understanding of the factors which the organisational end-user finds relevant to engendering their trust in AI for decision-making. The research outcomes for RQ2 were structured into of four main themes. Three of the themes were derived from the Chapter 2 literature review and were mapped directly to RQ2, while the remaining theme was identified as being potentially new. All four of the identified themes are discussed in this section, and the similarities, nuances of difference and distinct differences to literature are presented collectively, as related to Research Question 2.

7.2.2.1 Similarities with Literature for Research Question 2

Regarding the theme “xAI and the “black-box””, the research outcomes were consistent with the literature regarding an awareness and shared understanding of *the “black-box”* nature of AI as discussed by Rai, (2020), Berente et al. (2021), Kim et al. (2023), Lukyanenko et al. (2023) and others. Both the findings and the literature similarly identified the “black-box” as an important concept in explainability of AI, and it is therefore retained as an expected outcome, supporting the existing discourse.

Also, regarding the theme “xAI and the “black-box””, the research outcomes were consistent with the literature that transparency, explainability (xAI), and understanding are key elements in overcoming the “black-box” problem (Rai, 2020; Ågerfalk, 2020; Kim et al., 2023;

Wang & Ding, 2024; Choung et al., 2022). The literature also discussed xAI and transparency and understanding as essential to improving clarity and accountability (Weber, 2023; Lukyanenko et al., 2022). Both the findings and the literature similarly identified that transparency, understanding and explainability are all recognised to build trust and overcome the “black-box”. This is therefore retained as an expected outcome, further supporting the existing discourse.

In relation to the theme “xAI for trust in AI”, the findings were consistent with the literature that “Explanation of “How?”” is an important concept. A definition from Rai, (2020) states that “*Explainable AI (XAI) is the class of systems that provide visibility into **how** an AI system makes decisions and predictions and executes its actions*” (p. 137). Both the findings and the literature similarly identified that explainability should address “How” AI arrives at outcomes, and the finding is therefore retained as an expected outcome, supporting the existing discourse.

Considering the similarities identified above, the study revealed areas of similarity with the extant literature of a shared understanding of the “black-box” nature of AI and that explainability (through xAI) should address “how” AI arrives at outcomes. These two outcomes are presented as potential additions to the body of knowledge.

7.2.2.2 Nuances of Difference to Literature for Research Question 2

Considering the main theme, “xAI and the “Black-Box”*”* (Berente et al., 2021; Weber et al., 2023; Kim et al., 2023; Choung et al., 2022), there was a further insight that it’s not just xAI that can address the “Black-Box”, but also Transparency and Understanding. This does not necessarily warrant a new sub-theme (due to the overall similarity) but is of importance to understanding the seemingly circular/cross-supportive relationship between xAI, transparency and understanding. In summary, while the literature identifies each topic individually, this potentially new insight creates a clearer picture of the “interplay”/circularity/cross-support between the three.

Having explored the theme "xAI for Trust in AI” (Berente et al., 2021; Silva et al., 2023; Wang & Ding, 2024; Abedin, 2022), a potentially new insight emerged from the analysis of the key concept that data explanation must be consistent throughout the data lifecycle. No apparent similarity was identified in the existing literature, making this outcome a potential nuance of difference which is retained as a potential new sub-theme, titled “Explanation of Data.” In addition, the discussion of the findings and literature also revealed that the sub-theme may be relevant for providing further insight and understanding of the requirements for design and implementation of xAI, as called for by Pumplun et al. (2023) and Abedin, (2022).

Regarding the theme "xAI for Decision-Making" (Rabiee et al., 2024; Vanneste & Puranam, 2024; Wang & Ding, 2024), a potentially new insight emerged from the findings that AI Self-Criticism is necessary for trust and confidence in AI systems for decision-making. No apparent similarity was identified in the existing literature, making this outcome a nuance of difference to the main theme, which was retained as a potential new sub-theme, titled "AI Self-Criticism."

Under the potentially new main theme **"Antecedents"**, the findings highlighted two key concepts: that provenance of the AI and data integrity are crucial antecedents to adoption. As no apparent similarity to these two concepts was identified in the existing literature, they reveal potential new understanding of trust development in AI for its adoption in decision-making applications. Consequently, these nuanced differences are retained as potential new sub-themes, "Provenance for Trust" and "Importance of Data".

7.2.2.3 Distinct Differences to Literature for Research Question 2

Further to the two potentially new sub-themes "Provenance for Trust" and "Importance of Data", a distinct difference was identified that "antecedents" were not specifically discussed in the related literature. The potentially new main theme "Antecedents" was therefore also considered to be an apparent new insight under which to place the two sub-themes and was added to the conceptual framework.

7.2.2.4 Conclusions for Research Question 2

In conclusion to Research Question 2, the study identified several similarities to existing literature, including: an awareness and shared concern of the "black-box" nature of AI; the recognition of xAI, transparency and understanding all being important to overcoming the "black-box"; and the importance of explanation of "how" AI arrives at outcomes.

Nuances of difference were also revealed and highlighted further potential new insights, such as: the need for consistent explanation of data throughout the data lifecycle; the necessity of AI self-criticism for trust and confidence in AI systems for decision-making; the importance of demonstrated provenance of AI for trust and adoption of AI; and the importance of data and data integrity for trust and adoption of AI.

A distinct difference from the literature included insights into "antecedents" to trust engenderment for the adoption and use of AI in decision-making, including the factors "provenance for trust" and "importance of data" which appeared to lead to trust engenderment in AI for its adoption and use in organisational decision-making.

The above similarities are retained as additions to the body of knowledge in the final conceptual framework (Figure 30), the nuances of difference are included as potential new sub-themes, and the distinct differences are included as potential new themes.

The research outcomes addressed the aims of RQ2 by developing new insights and understanding into how xAI engenders trust in AI for decision-making in organisations, as well as by contributing new insights as to certain factors which may contribute to that trust in AI. The potentially new insights and understanding which were revealed in answering RQ2 are summarised below:

Research Question 2: How does xAI engender trust in AI for decision-making in organisations? And: What are the factors that contribute to trust in AI?

- Not just xAI, but also transparency and understanding, all collectively address the "black-box" and explain how AI arrives at outcomes, thereby engendering trust in AI for decision-making in organisations.
- There is a seemingly circular/cross-supportive relationship between xAI, transparency and understanding, which potentially creates a clearer picture of the "interplay" / circularity / cross-support between the three.
- Consistent explanation of data and its integrity throughout the AI lifecycle, demonstration of provenance of AI and creation of self-critical AI explanations, are all factors which potentially engender trust for decision-making in organisations.

7.2.3 Research Question 3: How does transparency influence emotional trust in AI for organisational decision-making?

The aim of RQ3 is to gain additional insights and understanding of transparency as an emotional dimension of trust in AI, in the context of the study around organisational decision-making. The research outcomes for RQ3 were structured into of four main themes. Three of the themes were derived from the Chapter 2 literature review and were mapped directly to RQ2, while the remaining theme was identified as being potentially new. All four of the identified themes are discussed in this section, and the similarities, nuances of difference and distinct differences to literature are presented collectively, as related to Research Question 3.

7.2.3.1 Similarities with Literature for Research Question 3

Regarding the theme "transparency for trust in AI", the research outcome that *fear is a barrier* to adoption which is primarily related to job security was consistent with literature by Berente et al. (2021) that AI and "any new frontier of automation has brought forth fears of deskilling and labour substitution" (p. 1444). The findings and the literature similarly identified fear as a barrier to adoption of AI as an important concept, and this is therefore retained as an expected outcome, supporting the existing discourse.

Also, regarding "transparency for trust in AI", the research outcome on the importance of *AI identification of human nuance* was consistent with Sabharwal et al. (2024) who argued

for centrality of human values in AI, stating with respect to trust in AI that “if humans cannot connect to how machines operate, trust falters” (p. 6). The lack of ability of AI to identify human nuance might potentially be alleviated by transparency, by allowing the connection of humans to how AI operates. This outcome was also therefore retained as an expected outcome, supporting the existing discourse.

In relation to the theme “transparency for decision-making”, the research outcome that *transparency and understanding are closely linked and are directly related to trust in AI for decision-making* was consistent with literature by Lindebaum et al. (2020) and Vanneste & Puranam (2024) who identified transparency and understanding as essential for building trust, particularly by clarifying AI’s reasoning and improving confidence in its decision-making. The research outcome and the literature both identified the close link between transparency, understanding and trust in AI for decision-making and the outcome was therefore retained as expected, supporting existing discourse.

Considering the theme “xAI and transparency”, the research outcome that *xAI and transparency drive understanding, which promotes trust in AI decision-making* was consistent with literature by Laato et al. (2022). They discussed transparency of an AI system enabling users to “better understand it” (p. 10), while also identifying that xAI should enable users to “understand” the systems which they are using which would then lead to trust. This interestingly also supports the insight from RQ2, Theme 1 regarding the “interplay”/circularity/cross-support of the three concepts xAI, transparency and understanding. The outcome was also therefore retained as expected, supporting existing discourse, while highlighting an additional potential new insight.

Also considering “xAI and transparency”, the research outcomes were consistent with the literature that *understanding “How” AI works and processes data is important for building trust.* Glikson & Woolley, (2020, p. 641) similarly discussed the importance of “understanding how”, in relation to explanations, transparency and trust building, and the outcome was therefore retained as an expected outcome, supporting the existing discourse.

In relation to the theme of “conjoined agency”, the research outcome that *AI is most effective and trusted as a supportive tool to enhance decision-making, which must be overseen by humans,* was consistent with Berente et al. (2021) who referred to “conjoined agency between AI and humans” as a contemporary frontier in facets of AI autonomy (p. 1443). Their reference list also revealed an article in the 4* *Academy of Management Review* journal by Murray, Rhymer & Sirmon (2021) titled “Humans and Technology: Forms of Conjoined Agency in Organizations”. Therefore, the outcome was retained as expected, supporting existing discourse.

Several similarities to literature were identified in the study. Fear as a barrier to AI adoption, the importance of AI identification of human nuance, and the close link between transparency, understanding and trust in AI for decision-making were the first set of expected outcomes. These were complimented by xAI and transparency driving trust in AI for decision-making, the importance of understanding “how” AI works in building trust and that AI should be a supportive tool, overseen by humans. All six outcomes are thus presented as potential additions to the body of knowledge.

7.2.3.2 Nuances of Difference to Literature for Research Question 3

With respect to the main theme “Transparency for Trust in AI” (Laato et al., 2022; Haque et al., 2023; Choung et al., 2022), a potentially new insight emerged from the findings, identifying that transparency is essential to alleviate fear and anxiety (both human emotions) and thereby foster trust. As no apparent similarity was found between this finding and the extant literature, this outcome was retained as a potential new sub-theme. The initial sub-theme title, “Fear as a Barrier,” has been refined to “Transparency to Mitigate Fear and Foster Trust in AI” to more accurately capture the identified concept.

Furthermore, this finding is highlighted as having the potential to provide additional insights as to how transparency influences emotional trust in AI for organisational decision-making (Glikson & Woolley, 2020), as considered in Research Question 3.

Under the main theme “Transparency for Decision-Making” (Kim et al., 2023; Lindebaum et al., 2020; Vanneste & Puranam, 2024), the second key concept identified, that *transparency priorities vary according to the role of the individual responsible for decision-making within the organization*, no apparent similarity with the extant literature was shown. This nuanced difference provided insight that transparency needs may be role-dependent. Consequently, the initial sub-theme, “Transparency and Understanding,” was also refined to “Role-Dependent Transparency” to more accurately reflect this finding.

7.2.3.3 Distinct Differences to Literature for Research Question 3

No distinct differences were identified between the findings and the literature for Research Question 3.

7.2.3.4 Conclusions for Research Question 3

In conclusion to Research Question 3, the study identified several similarities to existing literature, including: Fear being a barrier to AI adoption which is primarily related to job-security; AI having limited ability to identify human nuance, which inhibits trust; the insight that transparency and understanding are closely related and directly linked to trust in AI for decision-making; xAI and transparency drive understanding, which promotes trust in AI

decision-making; and that understanding “how” AI works and processes data is important for building trust.

Nuances of difference were also revealed and highlighted further potential new insights, that: Transparency is required to alleviate fear and anxiety (both human emotions) and to foster trust; and that there may potentially be different priorities for transparency, depending on the role of the individual responsible for decision-making within the organisation.

The above similarities are retained as additions to the body of knowledge in the final conceptual framework (Figure 30), and the nuances of difference are included as potential new sub-themes.

The research outcomes addressed the aims of RQ3 by developing new insights and understanding into how transparency influences emotional trust in AI for organisational decision-making. The potentially new insights and understanding which were revealed in answering RQ3 are summarised below:

Research Question 3: How does transparency influence emotional trust in AI for organisational decision-making?

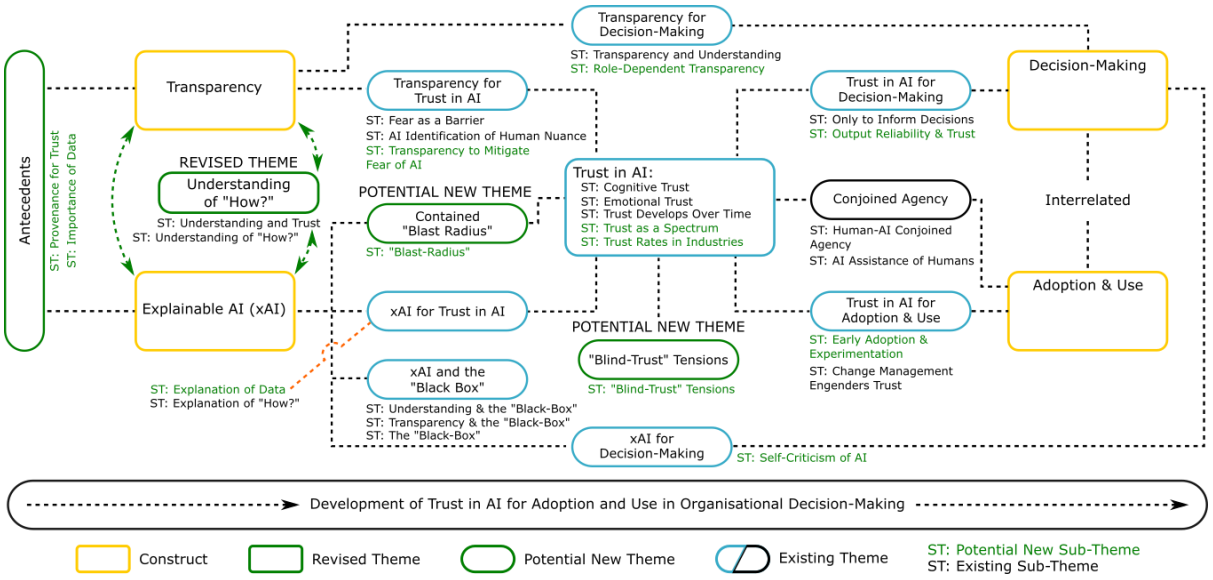
- Transparency and understanding are closely related and potentially alleviate the human emotions of fear and anxiety, which are primarily related to job-security, thus influencing emotional trust in AI for organisational decision-making.
- AI has a limited ability to identify human nuance, which might potentially be alleviated by transparency, allowing the connection of humans to how AI operates.

7.2.4 Final Conceptual Framework

The final conceptual framework presented in Figure 30 below is a re-work of the version presented at the end of Chapter 6 (Figure 28), with some minor adjustments made to provide better legibility and visibility for Section 7.3 to follow. An A4 full-page version is included in **Appendix J**.

Figure 30:

Final Conceptual Framework



Note: Author's Own

7.3 Research Contribution

This research study aimed to extend the current understanding and to provide additional insights into trust in artificial intelligence for its adoption and use in organisational decision-making. The apparent research contributions identified are presented as potential refinements to the literature (Crane et al., 2016) in this section, in three separate sub-sections and categories, namely: potential additions to the body of theory derived from similarities to literature, potential refinements to the theory derived from nuances of difference to literature, and potential extensions to the theory derived from distinct differences to literature.

7.3.1 Potential Additions to the Body of Knowledge

On the basis of the research conducted in this study, and as discussed in Section 7.2, several apparent areas of similarity to the literature were revealed. These are presented in Table 47 as potential additions to the body of knowledge, supporting the discussions in extant scholarship.

Table 47:

Summary Table of Similarities / Potential Additions to the Body of Knowledge

Existing Themes	Similarities / Potential Additions (Existing Sub-Themes)	Key Concepts
Trust in Artificial Intelligence	• *Trust Develops over Time*	• Trust takes time to develop
Trust in Artificial Intelligence for Adoption and Use	• *Change Management Engenders Trust*	• Importance of change management for trust in AI
Trust in Artificial Intelligence for Decision-Making	• *Only to Inform Decisions*	• AI should inform and enhance human decision-making, not replace it.
xAI and the "Black-Box"	• *The "Black-Box"* • *Understanding and the "Black-Box"* • *Transparency and the "Black-Box"*	• Awareness and shared concern of the "Black -Box" nature of AI. • Transparency, understanding and explainability were all recognised to build trust and overcome the "Black-Box".
xAI for Trust in AI	• *Explanation of "How?"*	• Explainability should address "How" AI arrives at outcomes.
Transparency for Trust in AI	• *Fear as a Barrier* • *AI Identification of Human Nuance*	• Fear is a barrier to adoption and is primarily related to job security. • AI has limited ability to identify human nuance, which inhibits trust.
Transparency for Decision-Making	• *Transparency and Understanding*	• Transparency and Understanding are closely related, directly linked to trust in AI for decision-making.
xAI and Transparency	• *Understanding of "How?"* • *Understanding and Trust*	• xAI and Transparency drive understanding, which promotes trust in AI decision-making. • Understanding "How" AI works and processes data is important for building trust.
Conjoined Agency	• *Human-AI Conjoined Agency* • *AI Assistance of Humans*	• AI is most effective and trusted as a supportive tool to enhance decision-making, which must be overseen by humans.

Note: Author's Own.

The above table provides a visual summary of the similarities to literature, which are retained as existing sub-themes and their related themes, as potential additions to the body of knowledge. The existing sub-themes and their related themes are included in the final conceptual framework in Figure 30. The key concepts which emerged from this study as potential **additions** are as follows:

- Trust takes time to develop, which also speaks to the importance of change management for trust in AI.
- AI should inform and enhance human decision-making, not replace it. Therefore, AI is most effective and trusted as a supportive tool to enhance decision-making, which must also be overseen by humans.
- There was an awareness and shared concern of the "Black-Box" nature of AI. Transparency, understanding and explainability were all recognised to build trust and to overcome this "Black-Box" concern.

- Explainability should address “How” AI arrives at outcomes, which creates an understanding of “How” AI works and processes data. This was viewed as important for building trust in AI.
- Fear is a barrier to adoption, which is an emotional dimension of trust in AI and is primarily related to job security.
- AI has limited ability to identify human nuance, which inhibits trust.
- Transparency and Understanding are closely related and are directly linked to trust in AI for decision-making. xAI and Transparency both drive understanding, which promotes trust in AI decision-making.

7.3.2 Potential Refinements to the Body of Knowledge

On the basis of the research conducted in this study, as discussed in Section 7.2, several nuances of difference to the literature were revealed. These are presented in Table 48 below as potential refinements to the body of knowledge, providing additional insights and understanding.

Table 48:

Summary of Existing Themes and Potential New Sub-Themes Offered as Potential Refinements to the Body of Knowledge.

Existing Themes	Nuances of Difference (Potential New Sub-Themes)	Key Concepts
Trust in Artificial Intelligence	• *Trust as a Spectrum* • *Trust Rates in Industries*	• Trust exists as a spectrum or on a scale. • Trust develops at different rates in different industries.
Trust in Artificial Intelligence for Adoption and Use	• *Early Adoption & Experimentation*	• Importance of gradual AI adoption and experimentation
Trust in Artificial Intelligence for Decision-Making	• *Output Reliability & Trust*	• Human verification of AI outputs is needed.
xAI for Trust in AI	• *Explanation of Data*	• Explanation must be consistent throughout the data lifecycle.
xAI for Decision-Making	• *Self-Criticism of AI*	• AI Self-Criticism is necessary for trust and confidence in AI systems for decision-making.
Transparency for Trust in AI	• *Transparency to Mitigate Fear of AI*	• Transparency is required to alleviate fear and anxiety and foster trust.
Transparency for Decision-Making	• *Role-Dependent Transparency*	• Different priorities for transparency, depending on the role of the individual responsible for decision-making within the organization.

Note: Author’s Own.

The above table provides a visual summary of the nuances of difference to literature, offered as potential new sub-themes, and potential refinements to the body of knowledge. The potential new sub-themes (green) and their related existing themes (black) are included in the final conceptual framework in Figure 30. The key concepts which emerged from this study as potential **refinements** are as follows:

- Trust exists as a spectrum or on a scale, and is not a simple uni-variate concept, but potentially consists of many independent variables, which contribute to overall trust development.
- Trust potentially develops at different rates in different industries, which emphasises the need for tailored explanations and transparency requirements specific to the particular use-case.
- Gradual AI adoption and experimentation is potentially important for creating trust in AI.
- Human verification of AI outputs is needed, which potentially demonstrates reliability and aids in developing trust.
- Explanation must be consistent throughout the data lifecycle, to demonstrate integrity of inputs, processing techniques within the model, and also the resulting outputs. Explanations and xAI design should potentially focus on the full AI system, not just inputs or outputs.
- AI Self-Criticism is potentially necessary for trust and confidence in AI systems for decision-making. If an AI can point to areas where it may be wrong or inaccurate, this would potentially engender improved trust from the user.
- Transparency is required to potentially alleviate fear and anxiety (strong human emotions) and to foster trust.
- There are potentially different priorities for transparency, depending on the role of the individual responsible for decision-making within the organization.

7.3.3 Potential Extensions to the Body of Knowledge

On the basis of the research conducted in this study, as discussed in Section 7.2, three distinct differences to the literature were revealed. These are presented in Table 49 below as potential **extensions** to the body of knowledge, providing additional insights and understanding.

Table 49:

Summary of Potential New Themes, Potential New Sub-Themes and Key Concepts Offered as Potential Extensions to the Body of Knowledge.

Potential New Themes	Potential New Sub-Themes	Key Concepts
"Blind-Trust" Tensions	• *"Blind-Trust" Tensions*	• There is a tension between the decision-making groups regarding "blindly-trusting" AI. Some support it while others reject it.
"Blast-Radius"	• *"Blast-Radius"*	• Importance of containing the potential negative outcomes of AI use for decision-making, to engender trust.
Antecedents	• *Provenance for Trust* • *Importance of Data*	• Provenance is an antecedent to adoption. • Data integrity is an antecedent to adoption.

Note: Author's Own.

The above table provides a visual summary of the apparent distinct differences to literature, offered as potential new themes, and potential extensions to the body of knowledge. The potential new themes (green) are included in the final conceptual framework in Figure 30. The key concepts which emerged from this study as potential extensions are as follows:

- There was a tension between the decision-making groups regarding “blindly-trusting” AI. Some supported it while others rejected it. This tension reveals that while some groups are inclined to trust AI without reservation, others strongly reject such an approach, highlighting a potential divide in comfort levels and perceived risks associated with trust in AI for its adoption and use.
- It is potentially important to contain the potential negative outcomes of AI use for decision-making, to engender trust. If it can be demonstrated that consequences can be easily remedied or contained within a “blast-radius”, it would potentially engender greater trust in AI and facilitate adoption and use for decision-making.
- Provenance and Data integrity are potentially both antecedents to trust in AI and subsequent adoption. Demonstrating prior successes and suitability of AI for an application or purpose may potentially engender greater trust. Likewise, demonstrating the integrity of the data being used may also engender greater trust in AI.

7.4 Recommendations for Management and Other Stakeholders

The recommendations for management and other stakeholders are drawn from the key concepts, theoretical conclusions and potential research contributions identified in this research study, as highlighted in Chapter 7. The insights and understanding which emerged from the outcomes of this research serve to inform the recommendations, which may potentially be of use to a variety of stakeholders involved in the adoption, implementation and use of AI in organisations, namely (but not limited to):

- Individuals at all levels in business, responsible for organisational decision-making.
- IT and data management teams.
- Human resource managers, trainers and educators.
- Risk and compliance officers.
- AI software developers and vendors.
- Industry regulators and policymakers.
- Consultants, external advisors and service providers.
- Regulatory or industry bodies.
- Government departments or agencies responsible for related legislation and oversight.

The primary recommendations are as follows:

- *Implementation of Gradual AI Adoption Strategies:* Encouraging gradual, phased AI adoption may allow for incremental trust development. This aligns with the importance of change management for trust in AI and mitigates concerns about AI replacing human roles. Also, this allows for the finding that trust takes time to develop, and at potentially different rates in different industries.
- *Investment in Explainable AI (xAI) and Transparent Systems:* Prioritising AI solutions that offer transparency of data processes, algorithms, and decision logic may help users understand AI's functionality. This thereby builds trust in AI and supports more informed decision-making. This recommendation is further supported by the insights that explanation, transparency and understanding are all important to overcoming the "black-box" nature of AI.
- *Development of Role-Specific Transparency Protocols:* Tailoring transparency requirements, to meet the needs of different individuals responsible for decision-making within the organization and providing role-appropriate explanations of AI processes may foster trust across operational and strategic levels. Consideration should also be given to the actual aspects of an AI tool which would lead the end-user to trust it, adopt it and use it.
- *Encouragement of "Human-in-the-Loop" Decision Models:* Positioning AI as an assistive tool rather than a replacement for human decision-making, while maintaining human oversight in AI-driven processes, may reinforce trust in AI and also accountability in AI decisions. This recommendation may create the right balance between leveraging the advantages of AI speed and processing ability, without "blindly-trusting" AI for critical organisational decisions.
- *Facilitation of Regular AI Training and Awareness Programs:* Offering ongoing education around AI's capabilities and limitations may enable employees at all levels to better understand AI's processes, data usage, and the need for human verification. This might in turn support trust in AI for subsequent adoption.
- *Implementation of AI Self-Criticism Features:* Ensuring that AI systems are designed with self-assessment and error-reporting mechanisms may enhance confidence in their reliability and perceived "honesty" and would allow users to critically assess AI outputs before basing any decisions on them. This might also alleviate certain aspects of AI's inability to identify human nuance, thereby engendering additional trust.

- *Addressing Concerns of Job Security (related to AI adoption):* Proactively communicating AI's intended role in the organization and emphasizing how it supports, rather than replaces human tasks, may alleviate fears and anxieties related to job security, thereby enhancing trust.
- *Highlighting Prior Success, Provenance and Data Integrity Standards:* Explaining prior successes, the provenance of AI tools and the standards employed to ensure data integrity, may all serve as trust-building measures to engender AI adoption and use. All of these observations speak to risk mitigation as well as informed risk tolerance for the individual responsible for organisational decision-making.
- *Mitigating AI's "Blast-Radius" of Potential Negative Outcomes:* Developing risk mitigation strategies for any adverse outcomes of AI adoption, may engender improved trust in AI technologies. By proactively addressing possible issues ahead of deployment, and by setting appropriate "safety nets" and safeguards, the organisation may be more willing to readily adopt AI solutions and tools for decision-making.
- *Monitoring and Adapting to Industry-Specific Trust Dynamics:* Recognising that trust in AI may develop at different rates across different industries and sectors might help to tailor solutions which are more readily-trusted. Continued monitoring, assessment and adaptation of AI tools to industry-specific requirements might also ensure a more sustainable approach to adoption and use.

7.5 Limitations of the Research

The limitations in this section are noted as different to the limitations of the research design and methods discussed in Chapter 4, section 4.11. Discussed below are six identified limitations of the overall research study.

- This study focused on Trust in AI, xAI and Transparency, but not on other areas of literature such as machine learning, data analytics, big data and data-driven decision-making.
- This study recognised the cognitive and emotional elements of Trust in AI, but it did not address the psychological factors.
- This study covered only certain sectors and not others, such as Education, Cybersecurity, Agriculture, Energy and Utilities, Construction, Mining, Engineering, Media and Entertainment, etc.
- This study was limited to a total of 19 participants from the executive, senior and operational decision-making levels, but did not explore the experiences of middle-management and ordinary employees as related to the research phenomena.

- This study identified three potential new themes, as distinct differences to extant literature, but did not explore them in any further detail.
- Similarly, this study identified eight potential new sub-themes, and these were also not explored in further detail.

7.6 Suggestions for Future Research

Based on the conclusions and limitations of the research, additional areas for potential future exploration are outlined below:

- The theoretical scope of this research focused on the literature related to Trust in AI, xAI and Transparency within an organisational context, and was further focused on AI adoption and use for decision-making. Future research may wish to extend this scope to additional areas of literature. Possibly literature on change-management, conjoined agency, big data and data-driven decision-making.
- The physical scope of the research explored 17 worldwide organisations across 16 different sectors, with a total sample size of 19 participants. Future studies may consider either expansion of the sample to more participants across additional sectors, or expansion of the sample within focused sectors. Additionally, future studies might explore the experiences of middle-management and ordinary employees related to the research phenomena.
- The research identified three potentially new themes, derived from the differences between the findings and the extant literature respectively, namely:
    - “Blind-Trust”,
    - “Blast-Radius”, and
    - “Antecedents”.

However, the study did not cover any of these potentially new themes in detail as it was not an aim of the research to explore them further. Therefore, it would be useful for future research to explore them in more detail to gain additional insights and understanding into their potential contribution to a further extension of the literature.

- Similarly, the research identified eight potentially new sub-themes, derived from the nuances of difference between the findings and the literature, namely:
    - “Trust as a Spectrum”,
    - “Trust Rates in Industries”,
    - “Early Adoption & Experimentation”,
    - “Output Reliability & Trust”,

- “Explanation of Data”,
- “Self-Criticism of AI”,
- “Transparency to Mitigate Fear of AI”, and
- “Role-Dependent Transparency”.

The study also did not cover any of these potentially new sub-themes in detail as it was not an aim of the research to explore them further. Therefore, it would similarly be useful for future research to explore them in more detail to gain additional insights and understanding into their potential contribution to a further refinement of the literature.

8.0 References

Abedin, B. (2022). Managing the tension between opposing effects of explainability of artificial intelligence: A contingency theory perspective. *Internet Research*, 32(2), 425-453. https://doi.org/10.1108/INTR-05-2020-0300

Ågerfalk, P. J. (2020). Artificial intelligence as digital agency. *European Journal of Information Systems*, 29(1), 1-8. https://doi.org/10.1080/0960085X.2020.1721947

Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., Garcia, S., Gil-Lopez, S., Molina, D., Benjamins, R., Chatila, R., & Herrera, F. (2020). Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion*, 58, 82-115. https://doi-org.uplib.idm.oclc.org/10.1016/j.inffus.2019.12.012

Balasubramanian, N., Ye, Y., & Xu, M. (2022). Substituting human decision-making with machine learning: Implications for organizational learning. *Academy of Management Review*, 47(3), 448-465. https://doi-org.uplib.idm.oclc.org/10.5465/amr.2019.0470

Bell, E., Bryman, A. & Harley, B. (2019). *Business research methods*. (2nd int. ed.). Oxford University Press.

Berente, N., Gu, B., Recker, J., & Santhanam, R. (2021). Managing artificial intelligence. *MIS Quarterly*, 45(3). https://doi-org.uplib.idm.oclc.org/10.25300/MISQ/2021/16274

Bogie, J. (2024). *Day 3: MPhil corporate strategy: Research methodology 2024* [Powerpoint slides]. Aspire. http://gibs.blackboard.com

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research In Psychology*, *3*(2), 77-101. https://doi.org/10.1191/1478088706qp063oa

Braun, V., & Clarke, V. (2021). Can I use TA? Should I use TA? Should I not use TA? Comparing reflexive thematic analysis and other pattern-based qualitative analytic approaches. *Counselling & Psychotherapy Research*, 21(1), 37–47. https://doi-org.uplib.idm.oclc.org/10.1002/capr.12360

Chen, J., Lim, C. P., Tan, K. H., Govindan, K., & Kumar, A. (2021). Artificial intelligence-based human-centric decision support framework: an application to predictive

maintenance in asset management under pandemic environments. *Annals of Operations Research*, 1-24. https://doi.org/10.1007/s10479-021-04373-w

Choung, H., David, P., & Ross, A. (2023). Trust and ethics in AI. *AI & Society*, 38(2), 733-745. https://doi.org/10.1007/s00146-022-01473-4

Crane, A., Henriques, I., Husted, B. & Matten, D. (2016). Publishing country studies in business & society: Or, do we care about CSR in Mongolia? *Business & Society*, 55(1), 3-10. https://doi.org/10.1177/0007650315619507

Creswell, J. W. (2007). *Qualitative inquiry and research design: Choosing among five approaches*. (2nd ed.). Sage Publications.

Dunn, J., Ruedy, N. E., & Schweitzer, M. E. (2012). It hurts both ways: How social comparisons harm affective and cognitive trust. *Organizational Behavior and Human Decision Processes*, *117*(1), 2-14. http://dx.doi.org/10.1016/j.obhdp.2011.08.001

Enholm, I. M., Papagiannidis, E., Mikalef, P., & Krogstie, J. (2022). Artificial intelligence and business value: A literature review. *Information Systems Frontiers*, *24*(5), 1709-1734. https://doi.org/10.1007/s10796-021-10186-w

Esmaeilzadeh, H., & Vaezi, R. (2022). Conscious empathic AI in service. *Journal of Service Research*, 25(4), 549-564. https://doi-org.uplib.idm.oclc.org/10.1177/10946705221103531

Gartner. (2024). *The pillars of a successful artificial intelligence strategy*. In Gartner (No. G00805859). Retrieved May 18, 2024, from https://www.gartner.com/document/5373763?ref=solrResearch&refval=412348043&

Gillespie, N., Lockey, S., Curtis, C., Pool, J., & Akbari, A. (2023). *Trust in Artificial Intelligence: A Global Study.* The University of Queensland and KPMG Australia. https://doi.org/10.14264/00d3c94

Glikson, E., & Woolley, A. W. (2020). Human trust in artificial intelligence: Review of empirical research. *Academy of Management Annals*, 14(2), 627-660. https://doi.org/10.5465/annals.2018.0057

GlobalData. (2023). *GlobalData industry profile: Global artificial intelligence* (NAICS:

541715). Retrieved February 6, 2024, from https://explorer-globaldata-com.uplib.idm.oclc.org/Analysis/details/global-artificial-intelligence

Gregor, S. (2024). Responsible artificial intelligence and journal publishing. *Journal of the Association for Information Systems*, *25*(1), 48-60. https://doi-org.uplib.idm.oclc.org/10.17705/1jais.00863

Gregory, R. W., Henfridsson, O., Kaganer, E., & Kyriakou, H. (2021). The role of artificial intelligence and data network effects for creating user value. *Academy of Management Review*, *46*(3), 534-551. https://doi-org.uplib.idm.oclc.org/10.5465/amr.2019.0178

Grover, P., Kar, A. K., & Dwivedi, Y. K. (2022). Understanding artificial intelligence adoption in operations management: Insights from the review of academic literature and social media discussions. *Annals of Operations Research*, 308(1), 177-213. https://doi.org/10.1007/s10479-020-03683-9

Haque, A. B., Islam, A. N., & Mikalef, P. (2023). Explainable artificial intelligence (XAI) from a user perspective: A synthesis of prior literature and problematizing avenues for future research. *Technological Forecasting and Social Change*, *186*, 122120. https://doi.org/10.1016/j.techfore.2022.122120

Hoff, K. A., & Bashir, M. (2015). Trust in automation: Integrating empirical evidence on factors that influence trust. *Human Factors*, *57*(3), 407-434. https://doi-org.uplib.idm.oclc.org/10.1177/0018720814547570

Josselson, R. (2013). *Interviewing for qualitative inquiry: A relational approach*. Guilford Press.

Kim, B. (Raymond), Srinivasan, K., Kong, S. H., Kim, J. H., Shin, C. S., & Ram, S. (2023). Rolex: A novel method for interpretable machine learning using robust local explanations. *MIS Quarterly*, 47(3), 1303–1332. https://doi-org.uplib.idm.oclc.org/10.25300/MISQ/2022/17141

Klag, M., & Langley, A. (2013). Approaching the conceptual leap in qualitative research. *International Journal of Management Reviews*, 15(1), 149–166.

https://doi.org/10.1111/j.1468-2370.2012.00349.

Laato, S., Tiainen, M., Najmul Islam, A. K. M., & Mäntymäki, M. (2022). How to explain AI systems to end users: a systematic literature review and research agenda. *Internet Research*, *32*(7), 1-31. https://doi.org/10.1108/INTR-08-2021-0600

Lindebaum, D., Vesa, M., & den Hond, F. (2020). Insights From "The Machine Stops" to better understand rational assumptions in algorithmic decision making and its implications for organizations. *Academy of Management Review*, 45(1), 247–263. https://doi-org.uplib.idm.oclc.org/10.5465/amr.2018.0181

Lukyanenko, R., Maass, W., & Storey, V. C. (2022). Trust in artificial intelligence: From a foundational trust framework to emerging research opportunities. *Electronic Markets*, 32(4), 1993–2020. https://doi-org.uplib.idm.oclc.org/10.1007/s12525-022-00605-4

Magnani, G., & Gioia, D. (2023). Using the Gioia methodology in international business and entrepreneurship research. *International Business Review*, 32(2), 102097. https://doi.org/10.1016/j.ibusrev.2022.102097

McAllister, D. J. (1995). Affect-and cognition-based trust as foundations for interpersonal cooperation in organizations. *Academy of Management Journal*, *38*(1), 24-59. https://doi-org.uplib.idm.oclc.org/10.2307/256727

McKinsey & Company. (2023). *The state of AI in 2023: Generative AI's breakout year*. Retrieved February 6, 2024, from https://www.mckinsey.com/~/media/mckinsey/business%20functions/quantumblack/our%20insights/the%20state%20of%20ai%20in%202023%20generative%20ais%20breakout%20year/the-state-of-ai-in-2023-generative-ais-breakout-year-v3.pdf?shouldIndex=false

Merriam, S., & Tisdell, E. (2016). *Qualitative Research. A Guide to Design and Implementation.* Fourth edition. San Francisco, CA: Jossey-Bass

OECD. (2024). *Collective action for responsible AI in health: OECD artificial intelligence papers.* OECD Publishing. Retrieved February 6, 2024, from https://www.oecd-ilibrary.org/deliver/f2050177-en.pdf?itemId=%2Fcontent%2Fpaper%2Ff2050177-

en&mimeType=pdf

Pumplun, L., Peters, F., Gawlitza, J. F., & Buxmann, P. (2023). Bringing machine learning systems into clinical practice: A design science approach to explainable machine learning-based clinical decision support systems. *Journal of the Association for Information Systems*, 24(4), 953–979. https://doi-org.uplib.idm.oclc.org/10.17705/1jais.00820

Rabiee, M., Mirhashemi, M., Pangburn, M. S., Piri, S., & Delen, D. (2024). Towards explainable artificial intelligence through expert-augmented supervised feature selection. *Decision Support Systems*, 181, N.PAG. https://doi-org.uplib.idm.oclc.org/10.1016/j.dss.2024.114214

Rai, A. (2020). Explainable AI: From black box to glass box. *Journal of the Academy of Marketing Science*, 48(1), 137–141. https://doi-org.uplib.idm.oclc.org/10.1007/s11747-019-00710-5

Sabharwal, R., Miah, S. J., Wamba, S. F., & Cook, P. (2024). Extending application of explainable artificial intelligence for managers in financial organizations. *Annals of Operations Research*, 1-31. https://doi.org/10.1007/s10479-024-05825-9

Saunders, M., Lewis, P., & Thornhill, A. (2009). *Research methods for business students*. (5th ed.). Pearson Education.

Schwandt, T. A., Lincoln, Y. S., & Guba, E. G. (2007). Judging interpretations: But is it rigorous? Trustworthiness and authenticity in naturalistic evaluation. *New Directions for Evaluation*, *2007*(114), 11-25. https://doi-org.uplib.idm.oclc.org/10.1002/ev.223

Shrestha, Y. R., Ben-Menahem, S. M., & Von Krogh, G. (2019). Organizational decision-making structures in the age of artificial intelligence. *California Management Review*, *61*(4), 66-83. https://doi.org/10.1177/0008125619862257

Seibert, S. E., Silver, S. R., & Randolph, W. A. (2004). Taking empowerment to the next level: A multiple-level model of empowerment, performance, and satisfaction. *Academy of management Journal*, *47*(3), 332-349. https://doi-org.uplib.idm.oclc.org/10.5465/20159585

Silva, A., Schrum, M., Hedlund-Botti, E., Gopalan, N., & Gombolay, M. (2023). Explainable artificial intelligence: Evaluating the objective and subjective impacts of xAI on human-agent interaction. *International Journal of Human–Computer Interaction*, *39*(7), 1390-1404. https://doi-org.uplib.idm.oclc.org/10.1080/10447318.2022.2101698

Sullivan, Y., de Bourmont, M., & Dunaway, M. (2022). Appraisals of harms and injustice trigger an eerie feeling that decreases trust in artificial intelligence systems. *Annals of Operations Research*, *308*, 525-548. https://doi.org/10.1007/s10479-020-03702-9

Vanneste, B. S., & Puranam, P. (2024). Artificial intelligence, trust, and perceptions of agency. *Academy of Management Review*, (ja), amr-2022. https://doi.org/10.5465/amr.2022.0041

Wang, P., & Ding, H. (2024). The rationality of explanation or human capacity? Understanding the impact of explainable artificial intelligence on human-AI trust and decision performance. *Information Processing & Management*, *61*(4), 103732. https://doi.org/10.1016/j.ipm.2024.103732

Wang, W., Qiu, L., Kim, D., & Benbasat, I. (2016). Effects of rational and social appeals of online recommendation agents on cognition- and affect-based trust. *Decision Support Systems*, 86, 48–60. https://doi-org.uplib.idm.oclc.org/10.1016/j.dss.2016.03.007

Weber, P., Carl, K. V., & Hinz, O. (2023). Applications of explainable artificial intelligence in finance—a systematic review of finance, information systems, and computer science literature. *Management Review Quarterly*, 1-41. https://doi.org/10.1007/s11301-023-00320-0

Wong, L. W., Tan, G. W. H., Ooi, K. B., & Dwivedi, Y. (2024). The role of institutional and self in the formation of trust in artificial intelligence technologies. *Internet Research*, *34*(2), 343-370. https://doi.org/10.1108/INTR-07-2021-0446

World Economic Forum. (2024, January 2). *Trust in AI: Why the right foundations will determine its future*. Retrieved May 10, 2024, from https://www.weforum.org/agenda/2024/01/davos24-trust-ai-right-foundations-

determine-its-future/

9.0 Appendices

9.1 Appendix A – Semi-Structured Interview Protocol

Table 50:

Semi-Structured Interview Protocol

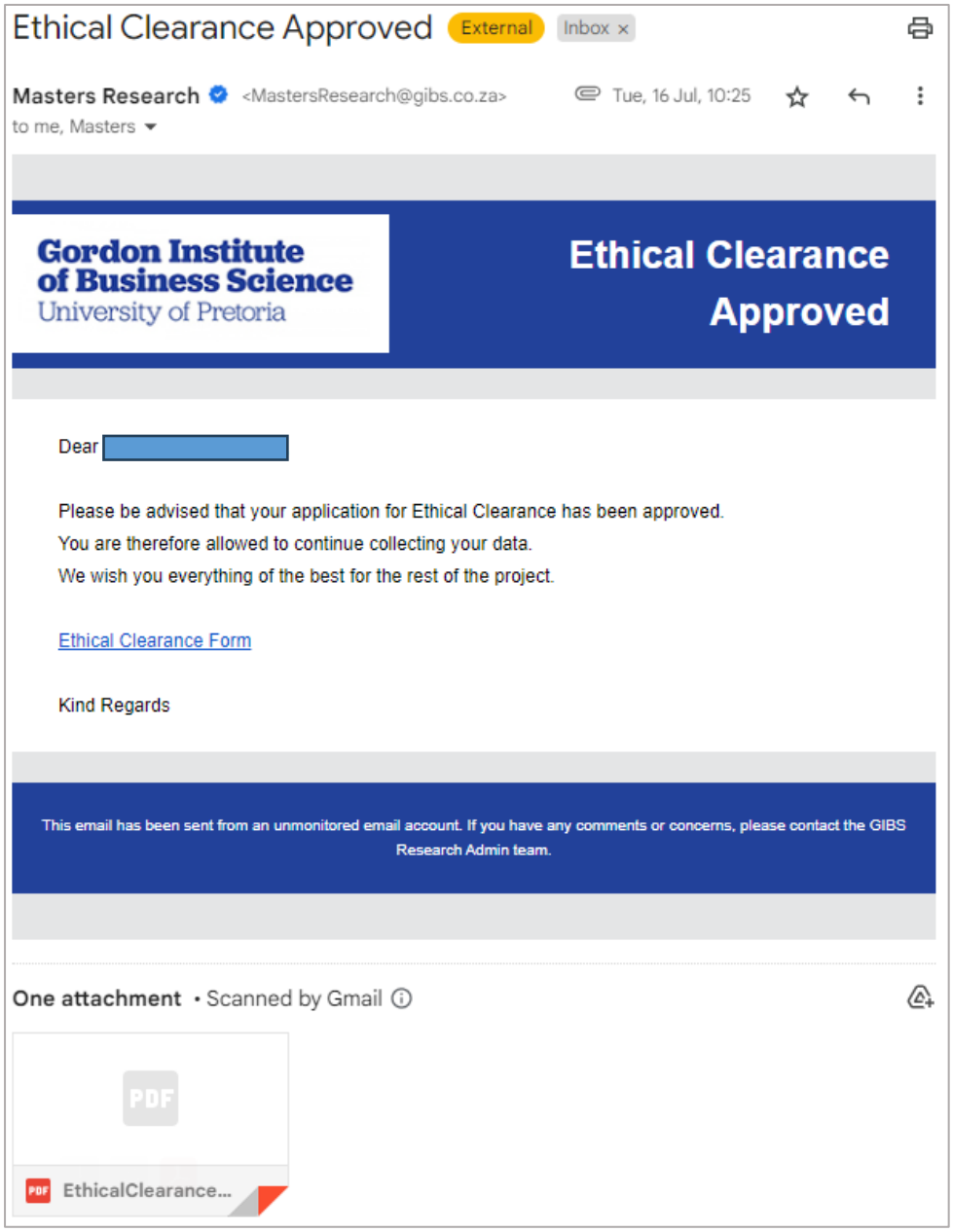
Research Questions	Interview Questions
Opening Question: Familiarisation and orientation of the interviewee.	**Little q1:** Please could you tell me about your experience and understanding of AI in organisations and your work environment?
RQ 1: How Does trust in AI lead to its adoption and use for organisational decision-making?	**Question 1:** Please would you tell me what the concept of Trust in AI means to your organisation?
	Question 2: In your experience, would your organisation use AI to make or inform decisions?
	Question 3: How would trust in AI influence the adoption and use of AI in your organisation for making or informing decisions?
RQ 2: How does xAI engender trust in AI for decision-making in organisations?	**Question 4:** In your experience, what explanations are needed to make AI more trustworthy?
And (Extension to RQ2): What are the factors that contribute to trust in AI?	**Question 5:** What are the factors external to the organisation which contribute to trust in AI?
	Question 6: In your experience, what explanations do you need from the AI system ahead of making any decisions?
	Question 7: What specific factors of an AI system would make you feel it is trustworthy?
RQ 3: How does transparency influence emotional trust in AI among organisational decision-makers?	**Question 8:** In your experience and understanding, how would transparency of an AI system influence trust in AI?
	Question 9: How does transparency influence trust in AI for organisational decision-making?
Closing Question: To understand any other insights which the interviewee finds relevant to the broad discussion.	**Little q2:** To close off our interview, I wonder if you could tell me how you see this developing going forward?
Additional questions: To either calrify or illicit further explanation and detail.	*Probing q1:* Please could you please elaborate further? *Probing q2:* Could you give me an example of what you mean? *Probing q3:* Where or when did this happen, could you explain the situation? *Probing q4:* Please would you give me an example to illustrate that? *Probing q5:* And could you tell me what your experience of that was?

Note: Author's own.

9.2 Appendix B – Ethical Clearance Approval

Figure 31:

Copy of Ethical Clearance Approval

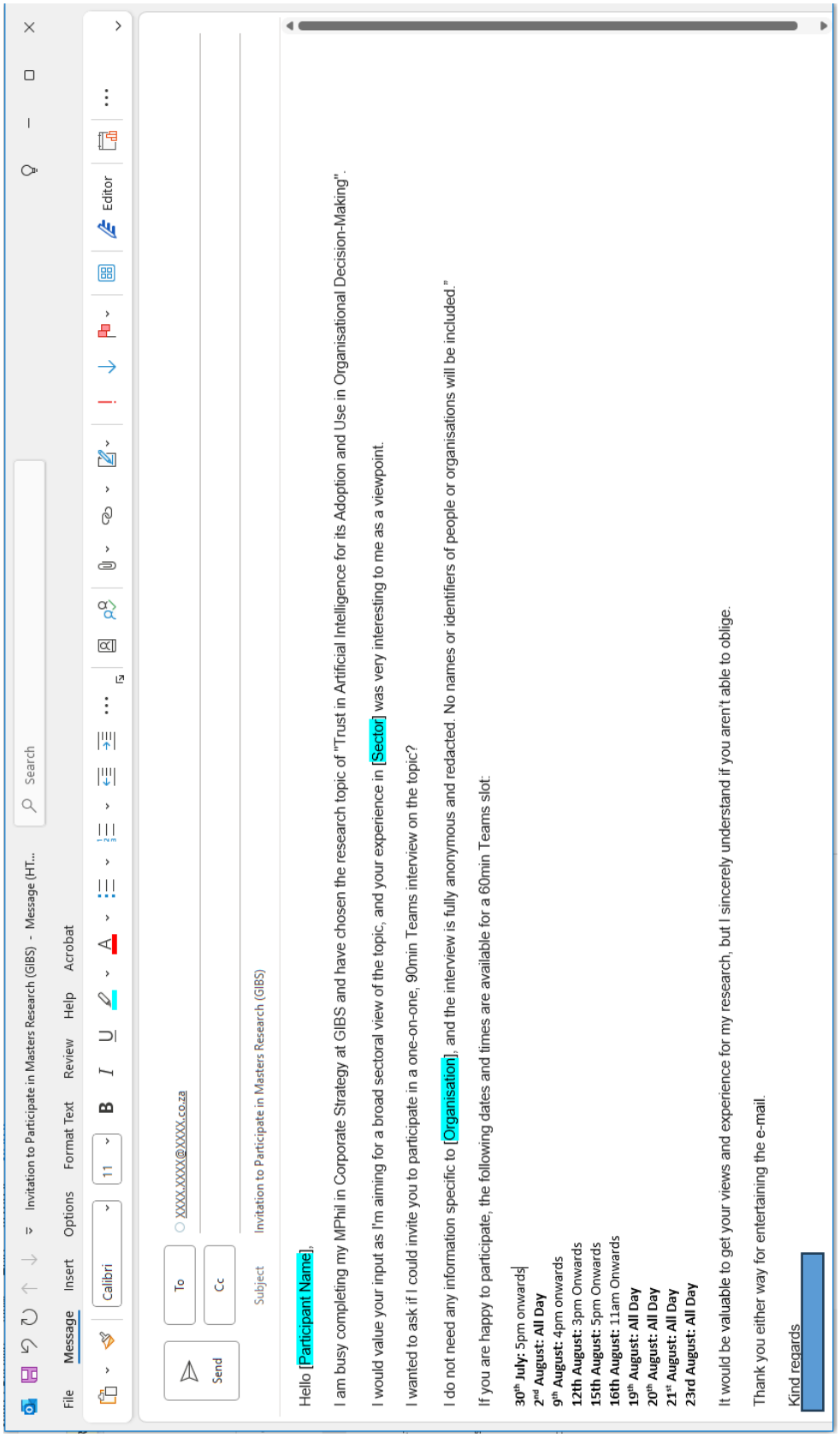


Note: Author's Own.

9.3 Appendix C – Sample e-Mail Invitation

Figure 32:

Sample e-Mail Invitation



Note: Author's Own.

9.4 Appendix D – Sample Informed Consent Letter

Figure 33:

Sample Informed Consent Letter

INFORMED CONSENT FOR INTERVIEWS

To whom it may concern,

I am conducting research on ***Trust in Artificial Intelligence for its Adoption and Use in Organisational Decision-Making***. Our interview is expected to last ***60minutes to 90minutes***, and will help us understand *how Trust in AI leads to its adoption and use for organisational decision-making*. Your participation is voluntary and you can withdraw at any time without penalty. By signing this letter, you are indicating that you have given permission for:

- The interview to be recorded;
- The recording to be transcribed by a third-party transcriber, who will be subject to a standard non-disclosure agreement;
- Verbatim quotations from the interview may be used in the report, provided they are not identified with your name or that of your organisation;
- The data to be used as part of a report that will be publicly available once the examination process has been completed; and
- All data to be reported and stored without identifiers.

If you have any concerns, please contact my supervisor or me. Our details are provided below.

Researcher	**Supervisor**
XXXX	XXXX
xxxxxxxx@xxxx.co.za	xxxxxx@xxxx.co.za
+xx xx xxx xxxx	

Signature of participant:_____ Date:_____

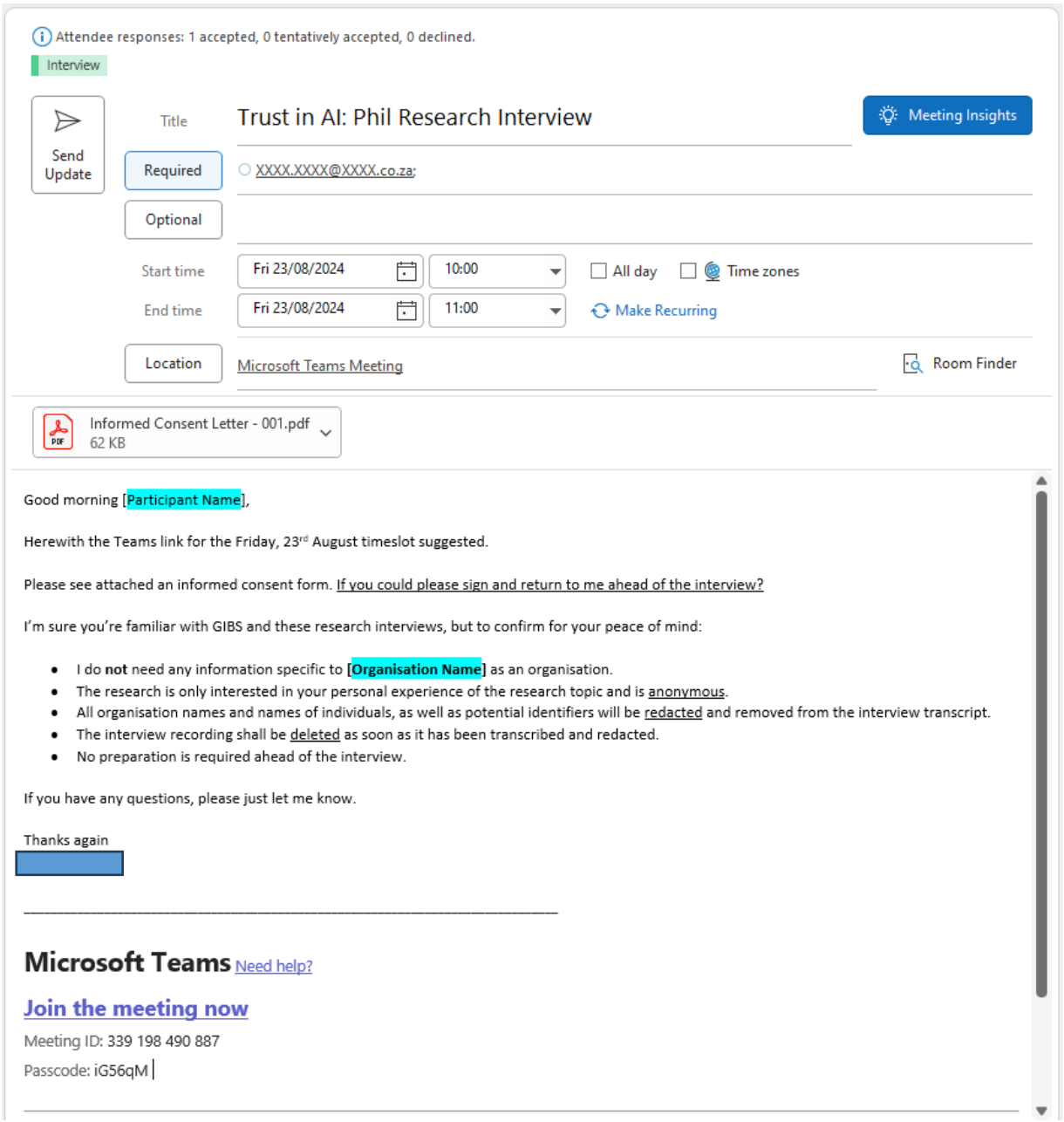
Signature of researcher: [redacted] Date: 23/07/2024

Note: Author's Own.

9.5 Appendix E – Sample Calendar Invite

Figure 34:

Sample Calendar Invite



Note: Author's Own.

9.6 Appendix F – Code Book

Table 51:

Code Book Export from ATLAS.ti

First-Order Codes	First-Order Categories: Code Groups
○ 100% Absolute trust would be irresponsible	Trust in AI General Trust as a spectrum
○ 100% Trust will never be achieved	Trust in AI General Trust as a spectrum
○ 3rd party verification of correctness is a factor	Verification: Factor Verification of outputs: Factor
○ Ability of AI to predict changes is a factor	Change Management engenders trust: Enabler
○ Ability of AI to understand human interests engenders trust	AI identification of human nuance: Factor
○ Ability to accurately interpret data engenders trust	AI data interpretation ability: Factor
○ Ability to calibrate engenders trust	Ability to calibrate & interrogate AI: Factor
○ Ability to interrogate and adjust is a factor	Verification: Factor Ability to calibrate & interrogate AI: Factor
○ Ability to interrogate logic creates understanding	Ability to calibrate & interrogate AI: Factor
○ Ability to interrogate the model is a factor	Verification: Factor Ability to calibrate & interrogate AI: Factor
○ Ability to learn is a factor	Continued adoption and trust: Enabler
○ Ability to tailor the AI to user preference	Personalisation to user
○ Accountability for AI use a factor in trust for adoption	Accountability and responsibility: Barrier
○ Accuracy, bias and ethics are part of the playing field	Bias and ethical concerns: Factor Ethics
○ Adjustability and visibility are factors	Ability to calibrate & interrogate AI: Factor
○ Adopting too fast will cause revolt / rejection	Early Adoption & Experimentation
○ Adoption and use for speed and efficiency	AI for speed and efficiency
○ Adoption for analysis, but not strategic control	Use in Strategy
○ Adoption not for adoption's sake	Adoption requires intent
○ AI ability to self-back-test is a factor	Self-criticism of AI
○ AI ability to understand the user would build trust	AI Understanding of Humans: Factor
○ AI adoption for speed to market	AI for speed and efficiency
○ AI adoption requires intent	Adoption requires intent
○ AI adoption to improve legacy systems	Influence of legacy systems
○ AI as a companion	AI Assistance of humans
○ AI as an assistance tool for productivity	AI Assistance of humans
○ AI as an assistance tool is easier to trust	AI Assistance of humans
○ AI can enhance decision-making turnaround speed not actual decision	AI for speed and efficiency
○ AI data interpretation for better decision-making	AI data interpretation ability: Factor
○ AI decision for human approval	Human-AI conjoined agency
○ AI decision-making risk of absolving responsibility	Accountability and responsibility: Barrier
○ AI decisions only within the understanding and complexity horizon of the user	Trust as a spectrum
○ AI decisions should mimic/match humans in explanation	Explanation of "how": Factor Explanation for Decision-Making
○ AI developer understanding of organisational needs is a factor	Trust in AI developers: Factor
○ AI enables faster decision-making	AI for speed and efficiency
○ AI for decision-making support but not complete autonomy	Only to inform decisions
○ AI immaturity to detect nuance like humans do	AI identification of human nuance: Factor
○ AI Infancy and Trust	AI infancy and Immaturity: Barrier
○ AI just a tool in a toolbox	AI Assistance of humans
○ AI may struggle with human relationships	AI identification of human nuance: Factor
○ AI more to inform than to make decisions	Only to inform decisions
○ AI must also understand the user	AI Understanding of Humans: Factor
○ AI must provide evidence for decision-making	Human Oversight of Decisions: Factor

First-Order Codes	First-Order Categories: Code Groups
○ AI mutual understanding to build trust	AI Understanding of Humans: Factor
○ AI needs discretion and understanding for situational awareness	AI Understanding of Humans: Factor
○ AI not subject to emotional influence	Emotion & Empathy: Factor
○ AI overlay on trusted legacy systems for decision-making	Influence of legacy systems
○ AI personalisation to user will increase trust	Personalisation to user
○ AI removes human subjectivity in decision-making	Human decision weakness
○ AI requires blind belief because of inherent complexity	"Blind-trust Tension" Blind Trust "Yes"
○ AI risk appetite would improve adoption for informing decisions	Risk and Adoption
○ AI roadshows for industry visibility and awareness?	Explanation of existing use: Enabler
○ AI self-awareness would impact trust either positively of negatively	Self-criticism of AI
○ AI self-interrogation and transparency engenders trust	Self-criticism of AI
○ AI self-learning creates challenges to explanation	Explanation challenges: Barriers
○ AI should inform, not make decisions	Only to inform decisions
○ AI should understand financial consequences	AI Understanding of business needs: Factor
○ AI standards wouldn't help to earn trust	Standards and Trust
○ AI to inform decisions or create efficiencies only	Only to inform decisions
○ AI to inform decisions, but verified by humans	Verification of outputs: Factor Only to inform decisions
○ AI tone and temperament for trust	AI identification of human nuance: Factor
○ AI Trust serves to democratise service offerings	Service Provider trust and stability: Factor
○ AI trust versus human trust is important	Human decision weakness
○ AI trusted for insights gathering	AI Assistance of humans
○ AI trusted for pre-cursory decisions	Only to inform decisions
○ AI understanding also important	AI Understanding of Humans: Factor
○ AI understanding of particular business needs	AI Understanding of business needs: Factor
○ AI understanding of user will increase trust	Understanding and Trust
○ AI understanding the user as a factor	AI Understanding of Humans: Factor
○ AI use of initial trust determines ongoing trust	Continued adoption and trust: Enabler
○ AI used to inform humans	Only to inform decisions
○ AI-Assisted decision-making	AI Assistance of humans
○ AI's ability to reveal trends and hidden patterns from data	AI data interpretation ability: Factor
○ Algorithm explanation as a factor	Explanation of "how": Factor
○ Alluding to black-box	The "Black-Box"
○ Application of transparency to organisational decision-making	Organisational transparency
○ Are output decisions biased by the inputs of the user	Bias and ethical concerns: Factor
○ Ask the right questions for the right output	Junk-in Junk-out
○ Assistance as an early use case	Early Adoption & Experimentation AI Assistance of humans
○ Assistance to mitigate fear of job losses	Fear as a: Barrier AI Assistance of humans
○ Assistance with Information interrogation	AI Assistance of humans
○ At some point, blind trust is required	"Blind-trust Tension" Blind Trust "Yes"
○ Augmentation of AI and human	Human-AI conjoined agency
○ Augmented use to enable trust	AI Assistance of humans
○ Authenticity as factor for trust	
○ Avoid black-box perception to engender trust	Trust and public sentiment / perception The "Black-Box"
○ Baby-steps for early adoption	Early Adoption & Experimentation
○ Bad human communication is a factor	Human decision weakness
○ Barrier could be inability to develop benchmark examples	Benchmarking for trust: Enabler
○ Benchmarking and scrutiny to form trust	Benchmarking for trust: Enabler
○ Benchmarking and testing overcomes hype	Testing and Trust: Factor Benchmarking for trust: Enabler

First-Order Codes	First-Order Categories: Code Groups
○ Better explanation of process increases trust	Process visibility
○ Better outputs from better understanding	Understanding of the AI tool's method: Factor
○ Black box leads to blind trust	"Blind-trust Tension" The "Black-Box" Blind Trust "No"
○ Black box, junk in junk out	Junk-in Junk-out The "Black-Box"
○ Blind trust an inherent requirement of AI	"Blind-trust Tension" Blind Trust "Yes"
○ Blind trust is an issue	"Blind-trust Tension" Blind Trust "No"
○ Blind trust is not sustainable at scale	"Blind-trust Tension" Blind Trust "No"
○ Blind trust will get you burnt	"Blind-trust Tension" Blind Trust "No"
○ Breadth of data an important factor	Transparency of data: Factor
○ Breadth of response as a factor to engender trust	
○ Broad adoption leads to trust	Trust and adoption
○ Broader use leads to adoption	Broader use and adoption (Network Effect?): Factor
○ Build confidence and trust over time	Trust develops over time
○ C-Suite fear from lack of understanding inhibits adoption	Fear as a: Barrier
○ Can it explain self-certainty or self-limitation?	Self-criticism of AI
○ Can't have blind trust without understanding. Problematic.	"Blind-trust Tension" Understanding and Trust Blind Trust "No"
○ Cant delegate accountability to AI is a barrier to adoption	Accountability and responsibility: Barrier
○ Cant have trust without knowing what informed the decision.	Understanding of the AI tool's method: Factor Understanding of "how": Factor
○ Cant yet identify human nuance	AI identification of human nuance: Factor
○ Cant yet replace human judgement	Human experience for trust
○ Case studies required for trust	Case studies for Trust: Enabler
○ Case study an important factor	Case studies for Trust: Enabler Broader use and adoption (Network Effect?): Factor
○ Catch 22 between understanding and regulation. Chicken and egg.	Regulation and adoption
○ CEO set ambition for grand adoption	Adoption requires intent
○ Certainty of output correctness needs to be demonstrated	Quality and Correctness of Outputs: Factor
○ Certification of credibility gives comfort for decision-making	Certification requirements
○ Challenge in measuring trust	Measurement of trust
○ Challenges of adoption in low literacy contexts	
○ Change management as enabler of understanding	Change Management engenders trust: Enabler
○ Change management enables buy-in, enables trust	Change Management engenders trust: Enabler Stakeholder buy-in
○ Change management for internal trust	Change Management engenders trust: Enabler
○ Change Management for Trust	Change Management engenders trust: Enabler
○ Clean data as a factor for trust	Data management: Factor and Enabler
○ Clean data requires vetting, causes diminishing returns	Data management: Factor and Enabler
○ Clean, internal data easier to trust	Internal data trust
○ Clear communication about AI potential	Change Management engenders trust: Enabler
○ Clear data with specific focus creates uses	Importance of Data
○ Closed data sets as a factor	Data management: Factor and Enabler
○ Clouds of grey prevent mass market adoption	The "Black-Box"
○ Code of conduct for boundaries and blast radius	Contain the "blast radius": Enabler
○ Comfort as a proxy for trust	
○ Comfort that expert is checking	Output reliability and trust Checking of outputs: Factor
○ Comparison of human and AI outputs for trust	Output reliability and trust Checking of outputs: Factor
○ Comparison of outcomes for adoption and implementation confidence/trust	Verification of outputs: Factor
○ Competitor use engenders trust	Broader use and adoption (Network Effect?): Factor

First-Order Codes	First-Order Categories: Code Groups
○ Competitor use of AI will drive adoption	Broader use and adoption (Network Effect?): Factor
○ Complexity forces trust engenderment	"Blind-trust Tension" Blind Trust "No"
○ Confidence in outcomes engenders trust which leads to adoption	Output reliability and trust Outcome confidence: Factor
○ Confidence in output leads to use in decision making	Quality and Correctness of Outputs: Factor
○ Confidentiality of data is external factor in trust	Data privacy and security for trust: Factor and Enabler
○ Configurability to enhance trust	Ability to calibrate & interrogate AI: Factor
○ Conjoined agency for human discretion	Human-AI conjoined agency
○ Conjoined agency reduces risk of poor decision-making	Human-AI conjoined agency
○ Conjoined decision-making	Human-AI conjoined agency
○ Conjoined human-AI agency	Human-AI conjoined agency
○ Conjoined use to drive decisions	Human-AI conjoined agency
○ Consequence makes trust a long-term earned outcome	Output reliability and trust Outcome confidence: Factor
○ Contained testing of decision-making ability to engender trust	Contained testing for trust
○ Contained testing to minimise blast radius	Contain the "blast radius": Enabler Contained testing for trust
○ Contained testing to minimise impact engenders trust	Contained testing for trust
○ Contained use to minimise impact	Contain the "blast radius": Enabler
○ Contextual awareness for trust	AI identification of human nuance: Factor
○ Continued adoption will overcome initial mis-trust for decision-making	Continued adoption and trust: Enabler
○ Continued evolution builds trust	Continued adoption and trust: Enabler
○ Continued use builds trust	Continued adoption and trust: Enabler
○ Continuous monitoring and testing for currency builds trust	Continuous testing required
○ Continuous testing for decision validity	Continuous testing required
○ Continuous testing through AI life is a factor	Continuous testing required
○ Contractual obligation of information security	Data privacy and security for trust: Factor and Enabler
○ Control of data sources	Importance of Data Importance of data source
○ Corporate governance inhibits willingness to trust	Governance and trust
○ Correct outputs engender trust	Quality and Correctness of Outputs: Factor
○ Credibility and trust in service providers engenders trust	Service Provider trust and stability: Factor
○ Credibility built by testing	Testing and Trust: Factor
○ Credibility of accuracy and reliability impacts trust	Output reliability and trust
○ Credible test cases important to engendering trust	Case studies for Trust: Enabler
○ Cultural influence and bias is evidence of importance of inputs	Bias and ethical concerns: Factor
○ Currency and evidence of updated-ness engenders trust	Continuous testing required
○ Currency of transparency and explanation	Explainability and Transparency
○ Current senior decision-making made on poor data	Human decision weakness
○ Current senior decision-making worse than AI	Human decision weakness
○ Cyber security as an external factor	Importance of privacy and security: Factor
○ Data correctness for trust	Data and input reliability: Factor and Enabler
○ Data integrity and provenance important for trust	Provenance for trust
○ Data integrity engenders trust	Importance of Data
○ Data integrity for trust	Importance of Data
○ Data manipulation erodes use for decision-making	Data management: Factor and Enabler
○ Data privacy and security for trust	Importance of privacy and security: Factor Data privacy and security for trust: Factor and Enabler
○ Data privacy for trust	Data privacy and security for trust: Factor and Enabler
○ Data privacy, governance and control	Importance of privacy and security: Factor
○ Data provenance as a trust factor	Provenance for trust
○ Data quality and integrity for trust	Importance of Data
○ Data quality key to use in decision making	Importance of Data
○ Data relevance for trust	Data management: Factor and Enabler

First-Order Codes	First-Order Categories: Code Groups
○ Data reliability and accuracy	Data and input reliability: Factor and Enabler
○ Data security an important factor	Data privacy and security for trust: Factor and Enabler
○ Data security and privacy as external trust factors	Importance of privacy and security: Factor Data privacy and security for trust: Factor and Enabler
○ Data security and safeguards as a factor	Safety and trust
○ Data security for trust	Data privacy and security for trust: Factor and Enabler
○ Data sources as a factor	Importance of Data Importance of data source
○ Data trust as a factor engenders AI trust	Importance of Data
○ Data verification as external factor	Verification: FactorVerification of data: Factor
○ Data verification important to trust	Verification of data: Factor
○ Decision-making requires full visibility and trust of information handling	Understanding of the AI tool's method: Factor
○ Decisions made at "a very human level"	Human Oversight of Decisions: Factor
○ Defined boundaries and firewalls for information	Importance of privacy and security: Factor
○ Degrees of trust	Trust as a spectrum
○ Delayed deliveries inhibit trust	Service Provider trust and stability: Factor
○ Dependency and cost tie-in risk an adoption barrier	Risk and Adoption
○ Developer understanding of nuance to build trust	Lack of AI industry nuance: Barrier Understanding and Trust
○ Developing trust in external partner frameworks	Trust in AI developers: Factor
○ Different levels of transparency	Transparency levels
○ Different types of trust	Trust as a spectrum
○ Difficult for AI to have nuanced context of specific fields	Lack of AI industry nuance: Barrier
○ Difficult to achieve organisational transparency to adopt AI. Barrier	Organisational transparency
○ Do the outputs add value	Value and trust Quality and Correctness of Outputs: Factor
○ Does it do what it says it does? Engenders trust	Validation of outputs: Factor
○ Does the AI have vested interests would impact trust	AI identification of human nuance: Factor
○ Don't let it be a black box	In vivo codes The "Black-Box"
○ Due diligence before decision-making	Human Oversight of Decisions: Factor
○ Early adoption, trailblazers, success cases; a catalyst for trust and adoption	Early Adoption & Experimentation
○ Early adoption	Early Adoption & Experimentation
○ Early adoption exploration	Early Adoption & Experimentation
○ Early adoption leads to trust	Early Adoption & Experimentation
○ Early adoption success	Early Adoption & Experimentation
○ Early adoption trust for decision-making weaker than later adopters	Early Adoption & Experimentation
○ Early adoption will jump-start trust and further use	Early Adoption & Experimentation
○ Early experimentation reveals shortcomings	Early Adoption & Experimentation
○ Early exploration for adoption readiness	Early Adoption & Experimentation
○ Early internal exploration	Early Adoption & Experimentation
○ Early sandpit creates boundaries for blast radius	Contain the "blast radius": Enabler
○ Early testing and trials to identify use cases	Contained testing for trust
○ Ease of undoing an action is a factor	Contain the "blast radius": Enabler
○ Easier to trust internal AI than external AI	Internal use easier
○ Emotion and human element in decision-making	Emotion & Empathy: Factor
○ Empathy and dignity for trust	Emotion & Empathy: Factor
○ Empathy as a factor for trust	Emotion & Empathy: Factor
○ Emphasis on outputs and results for trust	Quality and Correctness of Outputs: Factor
○ Employ internal people who understand, to create use	Need internal experts
○ End-user trust in AI outcomes	Output reliability and trust Outcome confidence: Factor
○ Endorsement of AI credibility, provides confidence, key to adoption	External endorsement: Enabler
○ Environment too complex for full transparency	Transparency levels

First-Order Codes	First-Order Categories: Code Groups
○ Equal access and competition as a factor for trust and adoption	Competitive advantage drives "black-box": Barrier
○ Erosion of trust jeopardises business models	
○ Examples of adoption success engender trust	Success and trust: Factor Adoption and success
○ Executive layer must be well-informed	Informed Leadership
○ Existing trust in data interrogation and interpretation	AI data interpretation ability: Factor
○ Existing use in industry engenders trust	Explanation of existing use: Enabler
○ Expert fall-back to check validity of outcomes	Output reliability and trust Outcome confidence: Factor
○ Expert validation engenders trust	Validation of outputs: Factor
○ Explain data verification process	Explanation of data: Enabler
○ Explain how AI achieves the objective	Explanation of "how": Factor
○ Explain source of data	Explanation of data: Enabler
○ Explain trends in training data	Explanation of data: Enabler
○ Explainability at developer level, not user	Explanation challenges: Barriers
○ Explainability vs Transparency	Explainability and Transparency
○ Explainability not different to transparency	Explainability and Transparency
○ Explanation by case study or benchmark examples	Case studies for Trust: Enabler Explanation of existing use: Enabler Benchmarking for trust: Enabler
○ Explanation no longer needed after trust achieved by proven success	Success and trust: Factor
○ Explanation of "how"	Explanation of "how": Factor
○ Explanation of "how" scenarios were developed	Explanation of "how": Factor
○ Explanation of "how" the AI got the output	Explanation of "how": Factor
○ Explanation of bias and logic	Bias and ethical concerns: Factor
○ Explanation of compliance rigor	Explanation of Provenance: Enabler
○ Explanation of confidence level & credibility of underlying data	Explanation of data: Enabler
○ Explanation of credibility and existing use engenders trust	Explanation of existing use: Enabler
○ Explanation of data anomalies would provide comfort	Explanation of data: Enabler
○ Explanation of data confidence level for decision-making	Explanation of data: Enabler Explanation for Decision-Making
○ Explanation of data provenance	Explanation of Provenance: Enabler
○ Explanation of data source credibility	Explanation of data: Enabler
○ Explanation of data sources	Explanation of data: Enabler
○ Explanation of depth of substantiating data for decision-making	Explanation of data: Enabler Explanation for Decision-Making
○ Explanation of expected output accuracy level based on input data	Explanation of data: Enabler
○ Explanation of grounding and principles	Explanation of "how": Factor
○ Explanation of how AI arrives at a decision	Explanation of "how": Factor Explanation for Decision-Making
○ Explanation of how AI filtered fact from nonsense	Explanation of "how": Factor
○ Explanation of how and what	Explanation of "how": Factor
○ Explanation of how it can help	Explanation of "how": Factor
○ Explanation of how the AI will help	Explanation of "how": Factor
○ Explanation of how the decision was made	Explanation of "how": Factor Explanation for Decision-Making
○ Explanation of how to use the AI tool engenders trust	Explanation of "how": Factor
○ Explanation of input, output and a verification examples	Verification of outputs: Factor
○ Explanation of inputs, model, outputs creates understanding	Understanding of the AI tool's method: Factor
○ Explanation of methodology and process used for an output	Explanation of "how": Factor Process visibility
○ Explanation of model confidence level for decision-making	Explanation of Provenance: Enabler Explanation for Decision-Making
○ Explanation of potential downsides	Explanation of risk
○ Explanation of potential mitigations of bad decision outcome	Contain the "blast radius": Enabler Explanation for Decision-Making
○ Explanation of previous results	Explanation of Provenance: Enabler

First-Order Codes	First-Order Categories: Code Groups
○ Explanation of pros and cons before making a decision	Contain the "blast radius": Enabler Explanation for Decision-Making
○ Explanation of provenance in use case application	Case studies for Trust: Enabler Explanation of Provenance: Enabler
○ Explanation of references and source data	Explanation of data: Enabler
○ Explanation of scale of risk	Explanation of risk Contain the "blast radius": Enabler
○ Explanation of source credibility	Explanation of Provenance: Enabler
○ Explanation of source verification	Explanation of data: Enabler
○ Explanation of sources	Explanation of Provenance: Enabler
○ Explanation of the AI's evolution over time	Explanation challenges: Barriers
○ Explanation of the depth of source data	Explanation of data: Enabler
○ Explanation of traceability	Explanation of Provenance: Enabler
○ Explanation of underlying assumptions	Explanation of "how": Factor
○ Explanation of what data was used	Explanation of data: Enabler
○ Explanation of what was considered for the decision	Explanation of "how": Factor Explanation for Decision-Making
○ Explanation of who developed the AI	Trust in AI developers: Factor
○ Explanation that the problem is understood	AI Understanding of business needs: Factor
○ Explanation through user-relevant/understood example	Explanation of existing use: Enabler
○ Explanation to understand "how" the model was trained	Explanation of "how": Factor
○ Explanation to understand biases	Bias and ethical concerns: Factor
○ Explanation to understand source of data influencing decision-making	Explanation of data: Enabler Explanation for Decision-Making
○ Explanations of provenance ahead of decisions	Explanation of Provenance: Enabler Explanation for Decision-Making
○ Explanations should build understanding	Explainability and Transparency
○ Exploratory use	Early Adoption & Experimentation
○ Extent of reliability engenders trust	Output reliability and trust
○ External adoption engenders trust	Broader use and adoption (Network Effect?): Factor
○ External endorsement builds AI credibility and trust	External endorsement: Enabler
○ External endorsement for adoption in decision-making	External endorsement: Enabler
○ External factors contribute more to mis-trust than trust	Trust and public sentiment / perception
○ External influence a barrier to trust	Trust and public sentiment / perception
○ External verification and review as a factor	Verification: Factor Verification of outputs: Factor
○ Face validity easier to trust	"Blind-trust Tension" Blind Trust "Yes"
○ Factor is fit our residual error of output	Output reliability and trust
○ Fake news is an external factor	Trust and public sentiment / perception
○ False external narratives impact trust in AI	Trust and public sentiment / perception
○ Familiarity builds trust but requires time	Trust develops over time
○ Far away from AI being a "source of truth"	Data and input reliability: Factor and Enabler
○ Fear and trust are related	Fear as a: Barrier
○ Fear is a barrier to trust and adoption for decision-making	Fear as a: Barrier
○ Fear of AI misuse can reduce trust	Fear as a: Barrier
○ Fear of data sharing for trust	Fear as a: Barrier
○ Fear of making a mistake as a barrier to adoption	Fear as a: Barrier
○ Feedback and iteration accelerate trust and adoption	Feedback and trust: Enabler
○ Filtering inconsistencies from external information adds trust	Consistency and Trust
○ Fire and forget doesn't create trust	"Blind-trust Tension" Blind Trust "No"
○ First use accuracy engenders trust	Provenance for trust
○ First-time adoption needs provenance	Provenance for trust
○ Five key risk factors to trust in AI	Risk and AI Trust
○ Fully automated systems easier to trust	AI Assistance of humans

First-Order Codes	First-Order Categories: Code Groups
○ General public fear and anxiety	Fear as a: Barrier
○ Generational differences in understanding and therefore adoption	Generational trust differences
○ Governance and regulation as external trust factors	Regulation and adoption
○ Governance as a factor	Governance and trust
○ Governance frameworks to improve trust	Governance and trust
○ Governance vs transparency importance	Governance and trust
○ Greater adoption engenders trust	Trust and adoption
○ Greater adoption through trust from proven ability	Broader use and adoption (Network Effect?): Factor
○ Greater trust leads to AI decision-making	Adoption requires intent
○ Having understanding, easier to trust	Understanding and Trust
○ Hesitance to give control to user a barrier	
○ High adoption with low understanding risks	Risk and Adoption
○ High level of trust for human support tasks	AI Assistance of humans
○ Higher success rate would lead to adoption	Adoption and success
○ Higher trust in internal data	Internal data trust
○ How data is managed is a factor	Data management: Factor and Enabler
○ How do you trust the explanation?	Explanation challenges: Barriers
○ How it works for other companies is a factor	Broader use and adoption (Network Effect?): Factor
○ How to mitigate bias to get a balanced view	Bias and ethical concerns: Factor
○ Human ability to control AI still needed for trust	Human experience for trust
○ Human Assistance engenders trust	AI Assistance of humans
○ Human assistance for speed	AI Assistance of humans
○ Human element critical in bigger decisions (risk)	Human Oversight of Decisions: Factor
○ Human endorsement engenders trust	Verification of outputs: Factor
○ Human experience also needed for trust	Human experience for trust
○ Human experience is a barrier to AI adoption for decision-making	Human experience for trust
○ Human experience, understanding and empathy is essential for trust	AI Understanding of Humans: Factor
○ Human intelligence augmentation	Human-AI conjoined agency
○ Human oversight to build confidence	Human Oversight of Decisions: Factor
○ Human-AI conjoined agency	Human-AI conjoined agency
○ Humans being informed to remain relevant	Informed Leadership
○ Humans not more accurate than AI	Human decision weakness
○ Humans required for the decision	Human Oversight of Decisions: Factor
○ Humans scared to execute, even at 100% trust	Verification of outputs: Factor
○ Humans still have to take responsibility for AI	Accountability and responsibility: Barrier Human Oversight of Decisions: Factor
○ Humans will always require some small degree of human approval	Verification of outputs: Factor
○ If success was verifiable, adoption for redundancies of people	Adoption and success
○ Immature AI understanding of risk for decision-making	AI Understanding of business needs: Factor
○ Importance of AI ability to understand risk appetite	AI Understanding of business needs: Factor
○ Importance of change management	Change Management engenders trust: Enabler
○ Importance of consequence of lack of trust	Contained testing for trust
○ Importance of different levels of transparency	Transparency levels
○ Importance of feedback loop with users	Feedback and trust: Enabler
○ Importance of governance, bias and ethics	Bias and ethical concerns: Factor Ethics
○ Importance of initial success in building trust	Success and trust: Factor
○ Importance of piloting before adoption	Testing and adoption: Factor
○ Importance of prompts for junk in junk out	Junk-in Junk-out
○ Importance of safeguards to remove early adoption trust concerns	Safety and trust
○ Importance of structured vs un-structured data in AI trust	Data management: Factor and Enabler
○ Importance of the dataset as information source	Importance of Data Importance of data source

First-Order Codes	First-Order Categories: Code Groups
○ Importance of transparency to engender trust	Transparency and trust
○ Importance of understanding AI methods	Understanding of the AI tool's method: Factor
○ Importance of understanding model derivation, not just outputs	Understanding of the AI tool's method: Factor
○ Important to keep humans involved for risk appetite	Human-AI conjoined agency
○ Improved efficiency and happiness post-adoption	AI for speed and efficiency
○ Inability to perform degrades trust	Success and trust: Factor
○ Increase in use	Broader use and adoption (Network Effect?): Factor
○ Increase in use at organisational level	Data and input reliability: Factor and Enabler
○ Increasing importance of trust with increased adoption	Trust and adoption Continued adoption and trust: Enabler
○ Incremental increase in AI trust exposure will mitigate mistrust	Contain the "blast radius": Enabler
○ Independent certification of credibility creates transparency	Certification requirements
○ Industry tailoring as a factor	Lack of AI industry nuance: Barrier
○ Information Vetting	Verification of data: Factor
○ Input parameter and information correctness	Input correctness and security: Factor
○ Interaction readiness for early adoption and user feedback	Early Adoption & Experimentation Feedback and trust: Enabler
○ Internal & External Change management	Change Management engenders trust: Enabler
○ Internal adoption easier to trust	Internal use easier
○ Internal AI models easier to adopt	Internal use easier
○ Internal and external trust self-reinforcing	Need internal experts
○ Internal audience for early adoption	Early Adoption & Experimentation
○ Internal data easier to adopt and test	Internal data trust
○ Internal development of use cases	Internal use easier
○ Internal governance of acceptable use case in organisation	Governance and trust
○ Internal models mitigate external factors of trust	Internal use easier
○ Internal training to fast-track adoption	Need internal experts
○ Internal trust easier than external trust	Internal use easier
○ Intimacy is a factor in trust	Emotion & Empathy: Factor
○ Intuitive sense of outputs engenders trust	AI identification of human nuance: Factor
○ Intuitive sense-making for trust	AI identification of human nuance: Factor
○ Issues in the adoption phase reduce trust	Trust and adoption
○ Its more about confidence in inputs giving you confidence in outputs than explanation	Input correctness and security: Factor
○ Junk in, junk out	Junk-in Junk-out
○ Knowledge of AI tool determines strategic influence	Use in Strategy
○ Lack of confidence prohibiting adoption	Adoption requires intent
○ Lack of explanation a barrier to use	Explanation challenges: Barriers
○ Lack of explanation diminishes trust	Explanation challenges: Barriers
○ Lack of nuance of personality is a barrier to use	AI identification of human nuance: Factor
○ Lack of regulation and ethics causes mis-trust	Regulation and adoption
○ Lack of trust due to fear and uncertainty	Fear as a: Barrier
○ Lack of trust due to fear of job losses	Fear as a: Barrier
○ Lack of understanding input to output conversion	Understanding of the AI tool's method: Factor
○ Lack of understanding needs to see results to trust	Success and trust: Factor
○ Lack of validation prevents adoption at scale	Validation of outputs: Factor
○ Language & personality as factors for trust	AI identification of human nuance: Factor
○ Late adoption benefits from trust over time	Early Adoption & Experimentation
○ Leadership values a barrier that AI can't replace	AI identification of human nuance: Factor
○ Legacy data can be incomplete and inaccurate	Influence of legacy systems Data management: Factor and Enabler
○ Legacy data can skew AI outcomes and erode trust	Influence of legacy systems Data management: Factor and Enabler
○ Legacy Machine Learning decision-making	Influence of legacy systems

First-Order Codes	First-Order Categories: Code Groups
○ Legacy systems inhibit AI adoption	Influence of legacy systems
○ Legacy systems more nuanced, less trust	AI identification of human nuance: Factor Influence of legacy systems
○ Legacy trust is a barrier to AI entry	Influence of legacy systems
○ Level of transparency just for internal or external audit	Transparency levels
○ Leveraging common frameworks and guardrails	Contain the "blast radius": Enabler
○ Leveraging legacy decision-making models for trust	Influence of legacy systems
○ Limited adoption for decision-making	Only to inform decisions
○ Link to dynamic capabilities, vs regulation & control	Regulation and adoption
○ Literally, the more trust, the more use	Use and trust
○ Logical and credible are factors of trust	Provenance for trust
○ Long term service stability a factor	Service Provider trust and stability: Factor
○ Low understanding and blind use of technology	"Blind-trust Tension" Blind Trust "No"
○ Lower trust in external AI service providers	Service Provider trust and stability: Factor
○ Managed approach to build trust	Change Management engenders trust: Enabler
○ Market credibility as a factor	External endorsement: Enabler
○ Measurement of trust by measurement of success rate	Measurement of trust Success and trust: Factor
○ Media negativity drives workplace resistance	Trust and public sentiment / perception
○ Minimise the blast radius engenders trust	Contain the "blast radius": Enabler
○ Mis-placed trust leads to failure	Success and trust: Factor
○ Mis-trust from mis-understanding and seeing AI as a black box	Understanding and the "black-box"
○ Model trust vs output trust	
○ Moral provenance an external factor	Provenance for trust
○ More transparency, more trustworthy, more trust	Transparency and trust
○ Must have fail-safe to create trust	Safety and trust
○ Must have knowledge of what the model is doing	Understanding and the "black-box"
○ Must know how to set guard rails	Only to inform decisions
○ Must understand boundaries, but boundaries limit effectiveness too	Contain the "blast radius": Enabler
○ Natural barrier to using externally-developed models	Trust in AI developers: Factor
○ Natural hesitance to adopt	Trust as a spectrum
○ Need continuity of understanding to prevent trust regression	Contained testing for trust
○ Need for external AI certification & authentication	Certification requirements
○ Need for internal AI experts	Need internal experts
○ Need humans in specialised vs commoditised markets	Lack of AI industry nuance: Barrier
○ Need relevant, reliable information before decision-making. Must be explained	Explanation of data: Enabler
○ Need to see more and understand "how"	Understanding of "how": Factor
○ Need to sense-check AI decisions	Output reliability and trust Checking of outputs: Factor
○ Need to understand how decision was reached	Only to inform decisions
○ Need understanding to build trust	Understanding and Trust
○ Need universal standard to check model validity	Standards and Trust
○ Network effect impacts trust and adoption	Broader use and adoption (Network Effect?): Factor
○ No controls and guards on externally developed models a barrier	Trust in AI developers: Factor
○ No deployment from not knowing "how"	Understanding of "how": Factor
○ Non-linear trust development	Trust as a spectrum
○ Nuance of human experience important	AI identification of human nuance: Factor Human experience for trust
○ Ongoing comparison of outcomes	Output reliability and trust Outcome confidence: Factor
○ Open collaboration and governance for trust	Governance and trust
○ Open source vs proprietary an important factor	Importance of Data Importance of data source
○ Open-source to mitigate bias engenders trust	Bias and ethical concerns: Factor

First-Order Codes	First-Order Categories: Code Groups
○ Organisational business models predicated on trust	
○ Organisational entry of external AI is a factor	Importance of privacy and security: Factor
○ Organisational exploration of AI	Early Adoption & Experimentation
○ Organisational transparency of AI use	Organisational transparency
○ Organisational transparency vs AI transparency	Organisational transparency
○ Outcome validation to build trust	Validation of outputs: Factor
○ Outcomes are more important than understanding	Output reliability and trust Outcome confidence: Factor
○ Outcomes still require human oversight	Output reliability and trust Outcome confidence: Factor
○ Output consistency enables trust	Consistency and Trust
○ Output reliability, consistency and accuracy engenders trust	Output reliability and trust Consistency and Trust
○ Output trust delegated to AI managers	Output reliability and trust Checking of outputs: Factor
○ Output verification for trust	Verification of outputs: Factor
○ Outputs are questioned due to black box	The "Black-Box"
○ Outside the complexity horizon of user, AI decisions too risky	Risk and AI Trust
○ Overcome fear and resistance by making AI exciting and appealing	Fear as a: Barrier
○ Package the test-case success to convince stakeholders	Case studies for Trust: Enabler Success and trust: Factor
○ Parallel to initial e-Commerce mistrust	Stakeholder buy-in
○ Part of company strategy	Use in Strategy
○ Passive AI as a shortcut to adoption	Continued adoption and trust: Enabler
○ Paucity of industry baseline or benchmarking standards	Benchmarking for trust: Enabler Standards and Trust
○ Paucity of industry knowledge and nuance prevents adoption	Lack of AI industry nuance: Barrier
○ People buy-in to enable trust	Stakeholder buy-in
○ Perceived complexity a possible barrier	Trust and public sentiment / perception
○ Perceived lag in AI adoption	Trust and public sentiment / perception
○ Perception of AI's purpose affects trust	Trust and public sentiment / perception
○ Political bias of the model is a factor	Bias and ethical concerns: Factor
○ Polluted data pollutes decisions	Importance of Data
○ Pop culture and Sci-Fi a barrier to adoption	Trust and public sentiment / perception
○ Pre-use user validation to build trust	Validation of outputs: Factor
○ Prevailing sentiment can drive trust	Trust and public sentiment / perception
○ Privacy and security for trust and adoption	Importance of privacy and security: Factor
○ Productivity and trust after the "hype cycle"	
○ Proof of concept and test-case as first-adoption barrier	Testing and adoption: Factor
○ Proof of outcome over perfect accuracy and confidence level	Output reliability and trust Outcome confidence: Factor
○ Proof of prior success and use	Success and trust: Factor
○ Proprietary and competitive advantage engenders black-box	Competitive advantage drives "black-box": Barrier The "Black-Box"
○ Proven success removes need for transparency	Success and trust: Factor
○ Provenance and track record for trust	Provenance for trust
○ Provenance and user feedback important to trust	Feedback and trust: Enabler Provenance for trust
○ Provenance of consistency, correct outcomes engenders trust	Consistency and Trust Provenance for trust
○ Providing certification requires transparency	Certification requirements
○ Provision of additional insights as a factor	Only to inform decisions
○ Public opinion and fear as external factors	Fear as a: Barrier
○ Public perception an external factor	Trust and public sentiment / perception
○ Public understanding an external factor	Trust and public sentiment / perception
○ Quality of data engenders trust	Importance of Data
○ Quality of output engenders trust	Quality and Correctness of Outputs: Factor

First-Order Codes	First-Order Categories: Code Groups
○ Quality of the data, not the AI	Importance of Data
○ Quality, integrity and source of data as a black box	Importance of Data The "Black-Box"
○ Readily adopted	Testing and adoption: Factor
○ Ready adoption for insights for decisions	Only to inform decisions
○ Regulation & Governance for adoption in decision-making	Regulation and adoption
○ Regulation as a challenge to adoption	Regulation and adoption
○ Regulatory compliance	Regulation and adoption
○ Regulatory requirements inhibit AI-only decision-making	Regulation and adoption
○ Reliability an important factor	Output reliability and trust
○ Reliability engenders trust	Output reliability and trust
○ Reliability engenders trust for decision-making	Output reliability and trust
○ Reliability of input data for decisions	Data and input reliability: Factor and Enabler
○ Reliability, consistency and accuracy reinforce validity	Output reliability and trust Consistency and Trust
○ Removal of biases will engender trust	Bias and ethical concerns: Factor
○ Removal of human emotion in decision making	Emotion & Empathy: Factor
○ Removal of human from decision-making	Human decision weakness
○ Removing empathy and emotion	Emotion & Empathy: Factor
○ Reputation element is a barrier	Stakeholder buy-in
○ Requirement for vetting of AI solutions	Human experience for trust
○ Resistance to adoption due to fear of job losses	Fear as a: Barrier
○ Response time as a factor to engender trust	AI for speed and efficiency
○ Retention of legacy trust into AI trust	Influence of legacy systems
○ Rigorous testing as a factor engenders trust	Testing and Trust: Factor
○ Rigorous testing important to trust	Testing and Trust: Factor
○ Risk aversion inhibits trust	Risk and AI Trust
○ Risk management of deployed solutions	Risk and Adoption
○ Risk of AI service provider dependence diminishes trust	Risk and AI Trust
○ Risk of complete mistrust to absolute trust	Risk and AI Trust
○ Risk of data manipulation can influence outputs, inhibit trust	Risk and AI Trust
○ Risk of longevity of service provider	Service Provider trust and stability: Factor
○ Risk of reliance is an external factor which diminishes trust	Risk and AI Trust
○ Risk of strategic accountability between CEO and CTO	Accountability and responsibility: Barrier Use in Strategy
○ Risk or fear of AI-dependence is a barrier to adoption	Fear as a: Barrier
○ Risk aversion is a barrier to adoption	Risk and Adoption
○ Risk that insufficient training degrades outputs and diminishes trust	Risk and AI Trust
○ Risk to independent thought diminishes trust	Risk and AI Trust
○ Role of AI in data aggregation & analysis	AI data interpretation ability: Factor
○ Safety nets to create accountability and trust	Accountability and responsibility: Barrier Safety and trust
○ Same information to humans or AI can lead to different outputs. Understanding the steps understands the output	Understanding of the AI tool's method: Factor
○ Scale of use impacts trust	Use and trust
○ Scepticism and checking of outputs	Output reliability and trust Checking of outputs: Factor
○ Sectoral endorsement for adoption	External endorsement: Enabler
○ Security and integrity of input data important to decision-making trust	Input correctness and security: Factor
○ Security important to trust	Data privacy and security for trust: Factor and Enabler
○ Self orientation	Self-criticism of AI In vivo codes
○ Self-explanation of limitations	Self-criticism of AI
○ Self-questioning is an important factor	Self-criticism of AI
○ Senior leaders need to see inside the black box	Informed Leadership The "Black-Box"

First-Order Codes	First-Order Categories: Code Groups
○ Senior supervision of early adoption	Early Adoption & Experimentation
○ Sense-checking of outputs	Output reliability and trust Checking of outputs: Factor
○ Shared trust between organisation and customer	Trust as a spectrum
○ Shouldn't trust what you don't understand	Understanding and Trust
○ Silly AI mistakes erode trust	Success and trust: Factor
○ Simple rating showing current validity of outputs	Quality and Correctness of Outputs: Factor
○ Simplicity and intuitive to use are factors	AI identification of human nuance: Factor
○ Situational awareness of AI	AI Understanding of business needs: Factor
○ Size of business decision impacts level of interrogation	Ability to calibrate & interrogate AI: Factor
○ So you've got misplaced trust and you've got informed trust	In vivo codes
○ Some people just don't trust	Trust and adoption
○ Sometimes humans can't explain	Explanation challenges: Barriers
○ Source is difficult to understand	Importance of Data Importance of data source
○ Speed of AI to mitigate failed use cases	AI for speed and efficiency
○ Speed to answers	AI for speed and efficiency
○ Stakeholder buy-in engenders trust	Stakeholder buy-in
○ Stakeholder trust leads to adoption, needs test-cases	Testing and adoption: Factor Stakeholder buy-in
○ Standardisation and evaluation criteria for AI trust	Standards and Trust
○ Still too early to adopt, despite benefits	Early Adoption & Experimentation
○ Sub-factors of trust	Trust as a spectrum
○ Success prevents trust degradation	Success and trust: Factor
○ Successes build trust. Failures cause regression of trust	Success and trust: Factor
○ Superficial outputs diminish trust potential	Quality and Correctness of Outputs: Factor
○ Sustained value supports continued adoption	Value and trust Continued adoption and trust: Enabler
○ Technology acceptance model?	Stakeholder buy-in
○ technology readiness levels	In vivo codes
○ temperature	Ability to calibrate & interrogate AI: Factor
○ Testimonials engender trust	Explanation of Provenance: Enabler
○ Testing as a factor for trust	Testing and Trust: Factor
○ The more use, the greater the trust	Use and trust
○ There's a trust equation	Trust as a spectrum
○ Those with AI knowledge require transparency, those without require explanations	Explainability and Transparency
○ To deploy a solution, must know what its doing	Understanding of the AI tool's method: Factor
○ To inform but not to make decisions	Only to inform decisions
○ Too early in AI development for clear policy	Regulation and adoption
○ Too much trust requirement can impede effectiveness and utility	"Blind-trust Tension" Blind Trust "Yes"
○ Tool for efficiency	AI for speed and efficiency
○ Training limitations cause need for human sense-checking	Verification of outputs: Factor
○ Transparency absolutely necessary, particularly at the start	Transparency and trust
○ Transparency aligns human and AI objectives through understanding	Transparency and Understanding
○ Transparency allows introduction of fall-backs	Contain the "blast radius": Enabler
○ Transparency allows output comparison between multiple models	Output reliability and trust Checking of outputs: Factor
○ Transparency allows sense-check of adoption	Transparency for adoption
○ Transparency allows understanding of many factors	Transparency and Understanding
○ Transparency allows understanding of risks	Transparency and Understanding
○ Transparency and clear explainability important to adoption	Transparency for adoption
○ Transparency and explainability mitigate the black-box	Transparency and the "black-box"
○ Transparency and mutual understanding build human trust	Transparency and Understanding

First-Order Codes	First-Order Categories: Code Groups
○ Transparency and understanding facilitate each other	Transparency and Understanding
○ Transparency answers "how", explanation gives detail	Explainability and Transparency
○ Transparency as a factor for trust	Transparency and trust
○ Transparency as a factor!?	Transparency and trust
○ Transparency can rationalise fear of use	Fear as a: Barrier
○ Transparency can't "perfectly" inform trust	Transparency and trust
○ Transparency creates comfort for decision-makers	Transparency and decision-making
○ Transparency creates relatability for users	Transparency for adoption Transparency and Understanding
○ Transparency creates understanding	Transparency and Understanding
○ Transparency creates understanding and adoption	Transparency for adoption
○ Transparency critical for understanding in decision-making	Transparency and decision-making
○ Transparency critical to explain what informs outputs	Transparency of outputs: Factor
○ Transparency critical to trust in AI	Transparency and trust
○ Transparency demonstrates intent of use	Transparency of use: Factor
○ Transparency explains how	Explanation of "how": Factor Explainability and Transparency
○ Transparency facilitates understanding, understanding facilitates trust	Transparency and Understanding
○ Transparency for AI certification bodies	Certification requirements
○ Transparency for understanding is critical	Transparency and Understanding
○ Transparency for user understanding	Transparency and Understanding
○ Transparency from service providers for trust	Service Provider trust and stability: Factor
○ Transparency gives confidence for first adoption	Transparency for adoption
○ Transparency gives understanding of limitations	Transparency and Understanding
○ Transparency gives understanding of logic	Transparency and Understanding
○ Transparency gives understanding, mitigates the black box.	Understanding and the "black-box" Transparency and the "black-box" Transparency and Understanding
○ Transparency important factor to what's behind the decision-making	Transparency and decision-making
○ Transparency important for accountability	Accountability and responsibility: Barrier
○ Transparency important for technical implementation	Transparency for adoption
○ Transparency is an enabler of trust	Transparency and trust
○ Transparency is nothing without understanding!	Transparency and Understanding
○ Transparency key for trust	Transparency and trust
○ Transparency necessary for trust	Transparency and trust
○ Transparency of AI function	Transparency and trust
○ Transparency of AI mistakes and errors	Self-criticism of AI
○ Transparency of AI outputs allows others to understand decisions	Transparency and decision-making
○ Transparency of Back-stops	Contain the "blast radius": Enabler
○ Transparency of evidence and data for trust and decision-making	Transparency of data: Factor Transparency and decision-making
○ Transparency of fallibility engenders trust	Self-criticism of AI
○ Transparency of how the model works to build trust	Transparency and trust
○ Transparency of how the output was achieved	Transparency of outputs: Factor
○ Transparency of industry nuance handling important for trust	Lack of AI industry nuance: Barrier
○ Transparency of Organisational use	Organisational transparency
○ Transparency of organisational use important for trust	Organisational transparency
○ Transparency of process important to trust	Process visibility
○ Transparency of removal of bias engenders trust	Bias and ethical concerns: Factor
○ Transparency of success factors will help adoption	Transparency for adoption
○ Transparency of track-record impacts trust	Success and trust: Factor
○ Transparency of underlying data	Transparency of data: Factor
○ Transparency of use cases engenders trust	Transparency of use: Factor
○ Transparency of who has access engenders trust	Transparency of use: Factor

First-Order Codes	First-Order Categories: Code Groups
○ Transparency provides comfort for use in decision-making	Transparency and decision-making
○ Transparency reduces risk of bias in decision-making	Bias and ethical concerns: Factor
○ Transparency removes grey areas in decision-making	Transparency and the "black-box"
○ Transparency should involve relatable explanations	Explainability and Transparency
○ Transparency to address black-box trust issues	Transparency and the "black-box"
○ Transparency to keep fears at bay	Fear as a: Barrier
○ Transparency to overcome early adoption hesitation	Transparency for adoption
○ Transparency would massively increase trust	Transparency and trust
○ Treat AI as an employee with limitation of mandate	AI Assistance of humans Contain the "blast radius": Enabler
○ Trust and adoption through gradual exposure	Trust and adoption
○ Trust and stakeholder buy-in engender adoption	Trust and adoption Stakeholder buy-in
○ Trust as a driver for adoption and use	Trust and adoption
○ Trust as a scale	Trust as a spectrum
○ Trust as the main driver of adoption	Trust and adoption
○ Trust build through proof and verification	Verification of outputs: Factor
○ Trust built through open and transparent processes	Process visibility
○ Trust comes from provenance and credibility	Provenance for trust
○ Trust defined as not needing to change outputs	Quality and Correctness of Outputs: Factor
○ Trust depends on internal vs external development	Trust in AI developers: Factor
○ Trust depends on use case	Use and trust
○ Trust development is a slow process	Trust develops over time
○ Trust develops over time, with verification	Verification of outputs: Factor
○ Trust engendered by data compliance monitoring	Data management: Factor and Enabler
○ Trust for use in decisions linked to disclosure of inherent bias and data quality / integrity	Importance of Data Bias and ethical concerns: Factor
○ Trust from adding value	Value and trust
○ Trust from expert sense-checking output	Output reliability and trust Checking of outputs: Factor Validation of outputs: Factor
○ Trust from humility	AI identification of human nuance: Factor
○ Trust from self-awareness of limitations	Self-criticism of AI
○ Trust from testing inputs	Testing and Trust: Factor
○ Trust grows with transparency and data integrity	Transparency of data: Factor
○ Trust in accuracy and usefulness	Value and trust
○ Trust in advice	Only to inform decisions
○ Trust in AI needs to be earned	Trust in AI General Trust develops over time
○ Trust in AI outcomes	Trust in AI General
○ Trust in data leads to trust in models	Importance of Data
○ Trust in data security	Data privacy and security for trust: Factor and Enabler
○ Trust in input data engenders AI trust	Input correctness and security: Factor
○ Trust in input information security	Input correctness and security: Factor
○ Trust in outcomes	Output reliability and trust Outcome confidence: Factor
○ Trust in outcomes vs trust in service providers & practitioners	Service Provider trust and stability: Factor
○ Trust in output accuracy and reliability	Output reliability and trust
○ Trust in outputs is different to trust in input information	Input correctness and security: Factor
○ Trust in process integrity	Process visibility
○ Trust in process privacy	Importance of privacy and security: Factor
○ Trust is a first step to repeated use	Use and trust
○ Trust is a first step, followed by value/benefit	Value and trust
○ trust is a function of credibility, reliability, intimacy	Output reliability and trust In vivo codes Trust as a spectrum

First-Order Codes	First-Order Categories: Code Groups
○ Trust is a spectrum based on use case	Trust as a spectrum
○ Trust is about overcoming fear	Fear as a: Barrier
○ Trust is lost, not earned	
○ Trust is not in the tool but in the people providing it	Service Provider trust and stability: Factor
○ Trust is proven through testing	Testing and Trust: Factor
○ Trust key to adoption in organisations	Trust and adoption
○ Trust linked to understanding	Understanding and Trust
○ Trust needs market provenance	Provenance for trust
○ Trust opens door for early experimentation	Early Adoption & Experimentation
○ Trust requires clarity of information	Explanation of data: Enabler
○ Trust takes time	Trust develops over time
○ Trust to overcome decision-making adoption barrier	Trust and adoption
○ Trust underpinned by ethical considerations	Ethics
○ Trust will grow with expertise and experience	Trust in AI General Trust develops over time Human experience for trust
○ Trust would be blind-trust, no need to check	"Blind-trust Tension" Blind Trust "Yes"
○ Truth from lies is a factor	Trust and public sentiment / perception
○ Two streams of explanation, technical vs non-technical	Explanation challenges: Barriers
○ Uncertainty in trusting existing AI	Trust in AI General
○ Uncertainty of credibility due to AI infancy	AI infancy and Immaturity: Barrier
○ Understand how information is generated	Understanding of "how": Factor
○ Understand inputs ahead of providing outputs	Input correctness and security: Factor
○ Understanding "how" the AI uses data is fundamental to trust	Understanding of "how": Factor
○ Understanding & transparency used as competitive advantage is a barrier	Competitive advantage drives "black-box": Barrier Transparency and Understanding
○ Understanding AI model evolution builds trust in outputs	Understanding and Trust
○ Understanding and dark magic (black box)	Understanding and the "black-box"
○ Understanding as an imposed requirement of the organisation	Understanding and Trust
○ Understanding boundaries as a factor for trust	Understanding and Trust
○ Understanding can't be a closed system	Understanding of the AI tool's method: Factor
○ Understanding engenders comfort and trust	Understanding and Trust
○ Understanding engenders trust	Understanding and Trust
○ Understanding engenders trust, engenders adoption	Understanding and Trust
○ Understanding is part of transparency	Transparency and Understanding
○ Understanding leads to trust	Understanding and Trust
○ Understanding mitigates vulnerability and fear	Fear as a: Barrier
○ Understanding of "how"	Understanding of "how": Factor
○ Understanding of "how" for decision-making	Understanding of "how": Factor
○ Understanding of AI benefit engenders trust	Understanding and Trust
○ Understanding of AI decision-making criteria will engender trust	Understanding and Trust
○ Understanding of bias	Bias and ethical concerns: Factor
○ Understanding of consequence from a human's perspective	AI Understanding of Humans: Factor
○ Understanding of data engenders trust	Understanding and Trust
○ Understanding of limitations	Understanding of the AI tool's method: Factor
○ Understanding of relevance and applicability of AI builds trust	Understanding and Trust
○ Understanding of system rules and variables is a factor	Understanding of the AI tool's method: Factor
○ Understanding the tool enables visibility	Understanding of the AI tool's method: Factor
○ Understanding what's "in" the AI (black box) quickly leads to trust	Understanding and the "black-box"
○ Understanding, consistency, repeatability and accountability engender trust	Accountability and responsibility: Barrier Consistency and Trust
○ Unsure that explanation is what is needed	Explanation challenges: Barriers
○ Up-time risk is an external factor	Risk and AI Trust

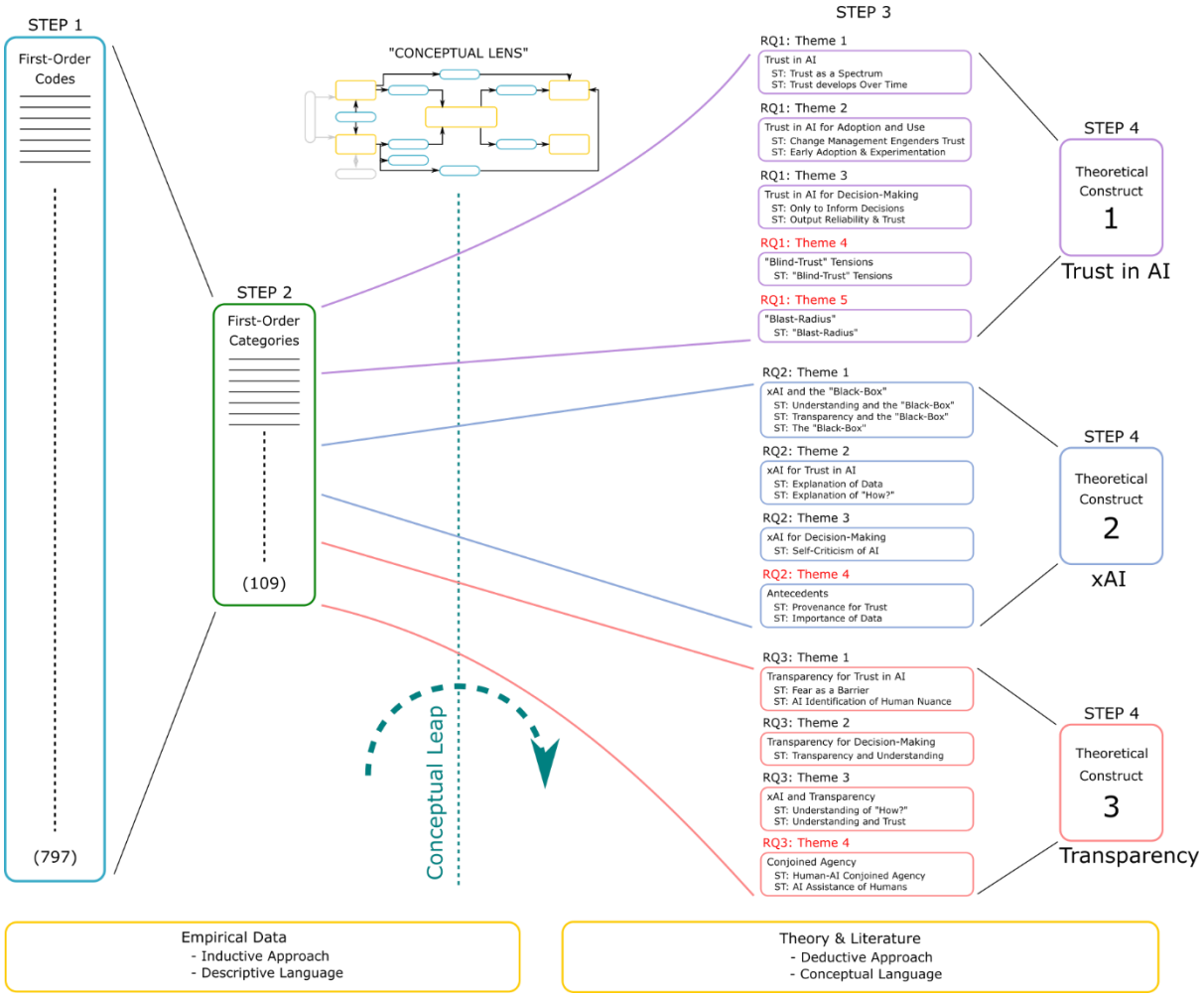
First-Order Codes	First-Order Categories: Code Groups
○ Use case delay from trust paucity	Use and trust
○ Use for efficiency	AI for speed and efficiency
○ Use for internal data mining and presentation	Internal data trust
○ Use for operational, but not yet strategic decisions	Use in Strategy
○ Use in back-end decision-making easier than human-facing	AI Assistance of humans
○ Use in source-comparison for decision-making	
○ Use of copyrighted information in model training a hinderance to trust	Data privacy and security for trust: Factor and Enabler
○ Use of internally-customised, external models	Internal use easier
○ Use to inform decisions	Only to inform decisions
○ User curiosity gives blind trust	"Blind-trust Tension" Blind Trust "Yes"
○ User intuition in explanation	AI identification of human nuance: Factor
○ User validation against known correct answers may build strong trust	Validation of outputs: Factor
○ User value engenders trust	Value and trust
○ Verification by those with understanding	Verification of outputs: Factor
○ Verification of success leads to adoption	Adoption and success
○ Vetting of inputs and outputs	Input correctness and security: Factor
○ Visibility of model logic for trust	Understanding of the AI tool's method: Factor
○ Visual explanation of "how" is a factor	Explanation of "how": Factor
○ What external data sets are a factor	Data and input reliability: Factor and Enabler
○ What method or algorithm was used is a factor	Understanding of the AI tool's method: Factor
○ What's informing the decision?	Understanding of "how": Factor
○ Who controls the safety is a factor	Safety and trust
○ Who created the AI engenders trust	Trust in AI developers: Factor
○ Who developed the model as an external factor	Trust in AI developers: Factor
○ Who or what did the programming is a factor	Service Provider trust and stability: Factor
○ Who retains data is a factor	Transparency of data: Factor
○ Will take a long time for full AI decision-making autonomy	Trust develops over time
○ Will use AI to inform, but not make decisions	Only to inform decisions
○ Without understanding "how" you exit quickly	Understanding and Trust
○ Wo is and how are they using AI needs explanation	Explanation of "how": Factor
○ Worked examples by user group to build trust	Use and trust
○ Wouldn't trust AI to apply judgement	Human experience for trust
○ You need the human for the decision-making	Human Oversight of Decisions: Factor
○ Younger generation more likely to trust AI	Generational trust differences

Note: Author's Own.

9.7 Appendix G – Example of Step 1 to 4 Data Analysis

Figure 35:

Example of Step 1 to 4 Data Analysis Process



Note: Author's Own.

9.8 Appendix H – Consistency Matrix

Table 52:

Consistency Matrix

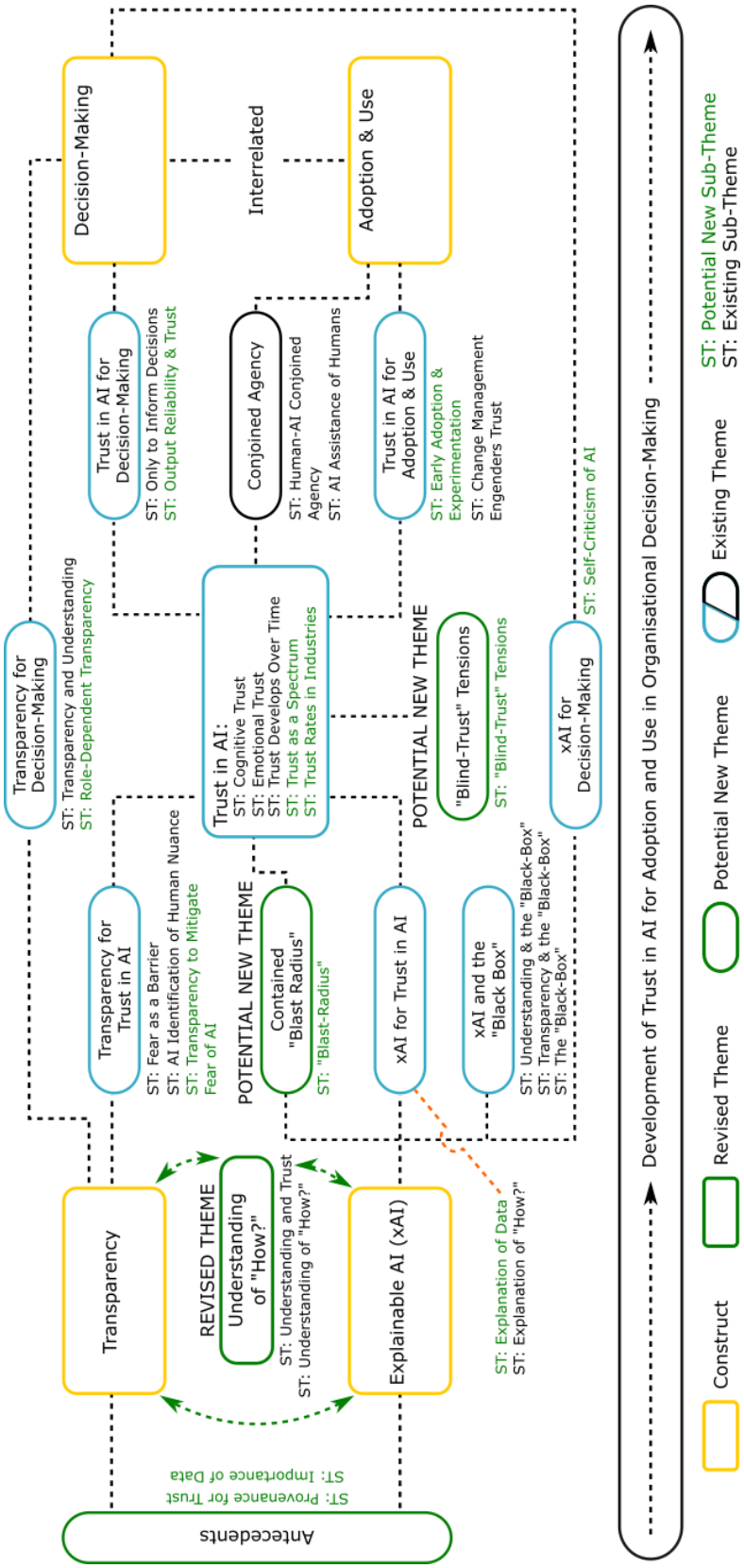
Research Questions	Literature Review	Data Collection Tool	Data Analysis
RQ 1: How Does trust in AI lead to its adoption and use for organisational decision-making?	Vanneste & Puranam, (2024); Sullivan et al., (2022); Wong et al., (2024); Enholm et al., (2021)	Semi-Structured Interview, Questions 1, 2 and 3	Transcription, thematic analysis for new understanding (Bell et al., 2019)
RQ 2: How does xAI engender trust in AI for decision-making in organisations?	Rai et al., (2020); Sabharwal et al., (2024); Lukanyenko et al., (2022)	Semi-Structured Interview, Question 4	Transcription, thematic analysis for new understanding (Bell et al., 2019)
And (Extension to RQ2): What are the factors that contribute to trust in AI?	Sullivan et al., (2022); Balasubramanian et al., (2022); Abedin, (2022); Glikson & Woolley, (2020); Lukanyenko et al., (2022); Pumplun et al., (2023)	Semi-Structured Interview, Questions 5, 6 and 7	Transcription, thematic analysis for new understanding (Bell et al., 2019)
RQ 3: How does transparency influence emotional trust in AI among organisational decision-makers?	Glikson & Woolley, (2020); Wang & Ding, (2024); Wang et al., (2016); Berente et al., (2021)	Semi-Structured Interview, Questions 8 and 9	Transcription, thematic analysis for new understanding (Bell et al., 2019)

Note: Author's Own

9.9 Appendix J – Final Conceptual Framework

Figure 36:

Final Conceptual Framework, A4 Scale



Note: Author's Own.