



SCHOOL OF
ECONOMICS AND
MANAGEMENT

The Defensive Equilibrium

*Balancing between Algorithmic Efficiency and
Human-Centric Sustainability*

by

Tuan Ngo & Emilia Wamsler

June 2026

Master's Programme in International Strategic Management

Supervised by Magnus Johansson

Abstract

The integration of Artificial Intelligence (AI) into corporate sustainability promises to alleviate the cognitive boundaries of the Strategic Decision-Making Process (SDMP) by synthesising massive, complex datasets. However, the inherent friction between the reductionist, instrumental logic of algorithmic models and the pluralistic, value-laden logic of sustainability creates profound operational and ethical paradoxes. Employing a qualitative, multi-case study of the Swedish manufacturing sector and utilising the Gioia methodology, this thesis investigates how corporate decision-makers navigate the tension between algorithmic efficiency and human control.

The findings reveal a stark literature-practice gap: rather than experiencing seamless "agentic delegation," decision-makers face severe socio-technical friction driven by fragmented sustainability data architectures. Consequently, managers experience extreme epistemic uncertainty and actively engage in "strategic non-use." They assign AI to administrative synthesis while retaining strict, defensive human authorisation over high-stakes sustainability choices. This study conceptualises this phenomenon as the Defensive Equilibrium, a path-dependent state where locally rational risk-aversion, combined with domain-neutral IT governance, collectively traps the firm in a state of systemic algorithmic under-use.

To dismantle this equilibrium, the thesis proposes a Human-Centric AI Governance Approach that maps domain-specific ethical guardrails and continuous "algorithmic interrogation" protocols directly onto the phase-gates of the SDMP. Ultimately, this research argues that the strategic value of AI in sustainability is entirely contingent upon resolving foundational data constraints and formally institutionalising human moral accountability.

Keywords: *Artificial Intelligence, Strategic Decision-Making, Corporate Sustainability, Algorithmic Governance, Algorithm Aversion, Bounded Rationality, Socio-Technical Systems.*

Acknowledgement

This thesis marks the culmination of an intensive, challenging, and profoundly rewarding academic journey at the Lund University School of Economics and Management (LUSEM). Conducting this research required navigating complex intersections of technology, strategy, and human behaviour, and it would not have been possible without the support of many individuals at LUSEM.

We would also like to extend our sincere appreciation to our supervisor, Magnus Johansson, for his invaluable guidance, critical feedback, and continuous encouragement. His insights pushed us to elevate our theoretical approach and refine our academic arguments.

A profound thank you goes to the sustainability managers, data experts, and operational users across the Swedish manufacturing sector who graciously participated in our study. While you must remain anonymous, your willingness to speak candidly about the operational realities and friction of algorithmic implementation provided the essential foundation for our research.

On a personal note, we wish to express our deepest love and gratitude to our friends and families in Sweden and Vietnam. Our journey has been made infinitely easier by your patience and your ability to keep us grounded during the most challenging phases of this research. Distance has never diminished your support, and we share this milestone wholeheartedly with all of you.

Table of Content:

1.0 Introduction.....	7
1.1 Background and Practical Motivation.....	7
1.2 Problematisation: The AI Paradox.....	7
1.3 Research Question.....	9
1.4 Purpose and Contributions.....	10
1.5 Delimitations.....	11
1.6 Thesis Outline.....	11
2.0 Literature Review.....	13
2.1 Strategic Decision-Making:.....	13
2.1.1 The Structure and Phases of the SDMP.....	13
2.1.2 Rationality and Power Dynamics in SDM.....	13
2.1.3 Heuristics, Mental Shortcuts, and Cognitive Overload.....	15
2.2 Decision-Making within Sustainability.....	16
2.3 Artificial Intelligence in Decision-Making.....	17
2.3.1 Definition and Types.....	17
2.3.2 Mitigating Behavioural Traps of Decision-Making.....	17
2.3.3 New Traps, Algorithmic Biases, and Negative Sides of AI.....	18
2.3.4 Managerial Attitude: Usage & Risk.....	19
2.4 Organisational AI Governance.....	20
2.4.1 The Need for Governance.....	20
2.4.2 Ensuring Accountability.....	20
2.4.3 The Scope of Governance.....	21
2.4.4 Implementation Mechanisms and Policies.....	21
2.4.5 Temporal and Maturity Dynamics.....	22
2.5 The Effect of AI in Sustainability Decision-Making.....	23
2.5.1 AI Heuristics linked to Sustainability.....	23
2.5.2 Sustainability Heuristics in SDM.....	23
2.6 Interpretation of Literature.....	24
3.0 Methodology.....	26
3.1 Research design.....	26
3.2 Data Collection.....	27
3.2.1 Interview Protocol.....	27
3.2.3 Data Recording and Transcription.....	29
3.3 Sampling and Participant Selection.....	29
Table 1: Table of Respondents.....	31
3.3.1 Exceptions to the Interview Protocol.....	32

3.4 Data Analysis.....	34
Table 2: Gioia Method Chart.....	36
3.5 Validity and Reliability.....	37
3.6 Ethical Considerations.....	37
3.7 Limitations.....	38
4.0 Empirical Data.....	40
4.1 1st Order Concept Findings.....	40
4.1.1 Heavy need human oversight & control.....	40
4.1.2 Sustainability data quality as a bottleneck.....	41
4.1.3 AI as efficiency tool not strategic sustainability driver.....	42
4.1.4 Governance focused AI implementation.....	43
4.1.5 Focus on safety, compliance & regulation within sustainability.....	44
4.1.6 Mistrust in AI outputs.....	45
4.1.7 Learning curve hindering usage.....	46
4.1.8 Sustainability nuance & complexity.....	46
4.1.9 Internal sustainability AI systems for control.....	47
4.2 Summary of 1st Order Concepts.....	49
Table 3: 1st Order Concepts Respondent Data.....	49
5.0 Analysis and Discussion.....	50
5.1 Interpreting 2nd Order Themes into Aggregate Dimensions.....	50
5.1.1 Defensive Personal Attitudes Towards AI.....	51
5.1.2 Socio-Technical Friction in Sustainability Strategy.....	52
5.1.3 Organisational Risk-Averse Governance Orientation.....	53
5.2 Analysis of Aggregate Dimensions.....	55
5.3. Cross-Case patterns and Internal Contradictions.....	59
5.4 Revisiting the Paradox: the Literature-Practice Gap.....	60
5.4.1 The Transfer of Cognitive Overload and Defensive Heuristics.....	60
5.4.2 Agentic AI is Rare in Sustainability Practice.....	60
5.4.3 Algorithmic Appreciation and Aversion versus Non-Use.....	61
5.4.4 Neutrality of AI Governance and the Need for Multidimensional Governance.....	61
5.4.5 Data Governance should be Considered Together with AI Governance in Sustainability.....	62
5.5 Re-Interpreting the SDMP: AI-Sustainability Governance and Human Agency.....	62
Figure 1: Human-Centric AI Governance for Sustainability Decision-Making.....	64
5.5.1 Explanation of Figure 1.....	64
5.5.2 Conditions and Limitations.....	65
6.0 Conclusion.....	68
6.1 Practical Implications.....	69

6.2 Limitations and Future Research.....	70
7.0 References.....	72
8.0 Declaration of AI Use.....	81
9.0 Appendices.....	82

List of Figures

Figure 1: Human-Centric AI Governance for Sustainability Decision-Making	64
--	----

List of Tables

Table 1: Table of Respondents	31
Table 2: Gioia Method Chart	36
Table 3: 1st Order Concepts Respondent Data	49

List of Appendices

9.1 Interview Guide: Sustainability Manager.....	82
9.2 Interview Guide: Data Expert.....	84
9.3 Interview Guide: Operational User.....	87
9.4 1st Order Concepts.....	90

List of Abbreviations

Abbreviations	Reference
AI	Artificial Intelligence
BI&A	Business Intelligence and Analytics
CDO	Chief Digital Officer
CIT	Critical Incident Technique
CSO	Chief Sustainability Officer
CSRD	Corporate Sustainability Reporting Directive
ESG	Environmental, Social, Governance

EU	European Union
FMCG	Fast-Moving Consumer Goods
GDPR	General Data Protection Regulation
IP	Intellectual Property
IS	Information Systems
IT	Information Technology
LLM	Large Language Model
SDM	Strategic Decision-Making
SDMP	Strategic Decision-Making Process
SME	Small-Medium Enterprise
STS	Socio-Technical Systems

1.0 Introduction

1.1 Background and Practical Motivation

Corporate discussions surrounding sustainability have evolved significantly in recent years. Organisations are progressively moving away from only maintaining manual, retrospective compliance reporting, opting instead to adopt data-driven strategies designed to identify long-term competitive levers (Collins et al., 2024). This profound transition aligns with the emergence of “Industry 5.0”, a paradigm that requires the merging of human-centric sustainability objectives with advanced technological systems to reset economic and social balances (Wang et al., 2025). Consequently, corporate sustainability is no longer perceived merely as an exercise in environmental, social, and governance (ESG) compliance or a superficial public relations effort. Rather, it is increasingly viewed as a core strategic imperative essential for firm resilience, survival, and sustained competitive advantage (George et al., 2021; Nidumolu et al., 2009). Organisations that effectively implement sustainable practices are better positioned to tackle grand challenges, such as climate change and resource scarcity, while simultaneously uncovering new opportunities for strategic innovation.

However, the pursuit of corporate sustainability introduces immense strategic complexity, placing a severe cognitive burden on managers. Decision-makers face fundamental paradoxes, in which they must simultaneously balance traditional economic performance driven by instrumental logic and short-term profit maximisation, with pressing, long-term environmental and social imperatives dictated by integrative sustainability logics (Ferraro et al., 2015; Hahn et al., 2014). Synthesising the massive, heterogeneous datasets required to evaluate these interconnected sustainability dimensions frequently leads to information overload. This can severely constrain the efficacy of the Strategic Decision-Making Process (SDMP) (Csaszar et al., 2024).

1.2 Problematisation: The AI Paradox

Despite its promise to alleviate cognitive burdens and enhance operational efficiency, the integration of Artificial Intelligence (AI) into strategic decision-making introduces a profound paradox. While AI attempts to solve the problem of information overload by processing vast

datasets at near-instantaneous speeds, it concurrently introduces critical "black box" risks (Berente et al., 2021). These risks prominently include algorithmic bias, poor data quality, and a severe lack of decision opacity, which can inadvertently lead to discriminatory, unethical, or fundamentally flawed strategic outcomes (Berente et al., 2021; Zhao & Gómez Fariñas, 2022). Modern AI tools, particularly those based on complex deep learning and machine learning algorithms, often generate results without providing any underlying reasoning that is readily understandable to human users (Berente et al., 2021; Lebovitz et al., 2022). Because AI models learn from historical data, they are inherently past-oriented and can reproduce or even amplify the existing prejudices and inequalities embedded within that data (Martin, 2019; Zhao & Gómez Fariñas, 2022).

This opacity poses a severe threat to sustainable decision-making. When managers cannot explain how an algorithm reached its conclusion, they face intense epistemic uncertainty (Lebovitz et al., 2022). This uncertainty frequently forces professionals into one of two detrimental behavioural extremes. On the one hand, managers may exhibit "algorithm aversion," where they consciously or unconsciously reject the superior computational insights of the machine in favour of their own flawed, intuitive heuristics (Burton et al., 2020). On the other hand, managers may display "automation bias" or unengaged augmentation, blindly trusting the AI without critical oversight (Lebovitz et al., 2022). In this latter scenario, professionals abdicate their moral and strategic responsibilities, allowing the algorithm to make decisions without human interrogation (Martin, 2019).

Furthermore, relying heavily on algorithms risks marginalising the nuanced human empathy, contextual understanding, and qualitative judgment essential for socio-ethical considerations (Kellogg et al., 2020; Raisch & Krakowski, 2021). Human decision-makers possess the unique capacity for substantive, value-rational reflection, which is necessary to balance the pluralistic goals of sustainability, whereas AI systems operate on an intensely instrumental logic designed to optimise predefined utility functions (Gao & Bansal, 2013). When agentic Information Systems (IS) artefacts, which can independently develop protocols and select actions, are granted expansive decision-making latitude, the centre of agency shifts away from humans (Baird & Maruping, 2021). This shift risks creating a technocratic regime of control where algorithms

bypass the ethical heuristics and interpersonal negotiations that workers typically use to make complex moral decisions (Kellogg et al., 2020).

Finally, the use of AI itself carries a severe ecological cost that managers often overlook. Due to a "heuristics gap" in AI ethics, stakeholders frequently treat AI as a weightless, virtual entity, entirely ignoring the massive energy consumption and physical infrastructure required to train complex "Red AI" models (Tamburrini, 2025). Consequently, unregulated or poorly governed AI systems represent a socio-ethical and environmental blind spot, threatening to undermine the very sustainable development goals they are deployed to achieve (Raisch & Krakowski, 2021).

1.3 Research Question

Synthesising these streams reveals a distinct empirical and methodological gap in the literature. While the macro-level implications of AI on firm performance, sustainability reporting, and the theoretical underpinnings of corporate sustainability are increasingly well-documented (Mustafa et al., 2025), the micro-foundational perspective remains largely unexplored. The microfoundations movement in strategy emphasises the need to avoid macro-to-macro causal claims, advocating instead for an understanding of how individual-level actions, interactions, and cognitive mechanisms mediate organisational outcomes (Felin et al., 2015).

Specifically, there is a profound lack of research regarding how individual decision-makers behave when faced with the paradoxical questions of AI use while simultaneously navigating the complex, competing demands of sustainable strategic decisions. Moreover, as technologies advance to become agentic participants in organisational routines, very little is understood about how this shifting agency affects the execution of these routines. However, this is contingent on the firm desire to implement agentic AI. Unpacking this black box requires observing how managers deal with conjoined agency and negotiate the boundaries between algorithmic recommendations and human ethical judgment (Murray et al., 2021).

To address this gap, this thesis poses the following research question: *How do decision-makers navigate the tension between algorithmic efficiency and human control within sustainability?*

1.4 Purpose and Contributions

The purpose of this study is to investigate how AI influences the decision-making for sustainability, exploring whether AI acts merely as an isolated functional tool or as an agentic participant in organisational decision-making (Berente et al., 2021). It examines how the integration of self-learning algorithms alters firm governance, particularly concerning the ethical implications, security, and accountability of delegated decisions (Martin, 2019; Powell et al., 2011). By unpacking the human-machine dynamic, the ultimate goal of this research is to propose a human-centric governance approach that clarifies responsibility and reliance boundaries, ensuring that AI enhances, rather than compromises, a firm's sustainable competitive advantage.

This study makes several key contributions to the literature. Theoretically, it contributes to the field of behavioural strategy by grounding strategic management in realistic assumptions about human cognition and human-algorithm interaction (Powell et al., 2011). It moves beyond the dichotomous debate of automation versus augmentation by illustrating how managers practically enact "AI interrogation practices" to reconcile divergent knowledge claims and mitigate algorithmic opacity. Furthermore, the study adds nuances to IS delegation theory by conceptualising the human-agentic IS artifact dyad within the context of sustainability.

Practically, this research offers critical insights for executives and policymakers. It demonstrates that the responsibility for the ethical implications of algorithms cannot be fully outsourced to the machine; developers, managers, and top management teams remain permanently accountable for the design and deployment of these systems (Schwaeke et al., 2025). By providing a structured approach for managing the inscrutability and generative autonomy of AI, this study equips organisational leaders with the digital readiness and governance strategies required to implement AI responsibly. It highlights the absolute necessity of hybrid metrics and active stakeholder management to integrate AI successfully into daily operations without succumbing to algorithmic bias (Mustafa et al., 2025). Ultimately, it guides firms in achieving a synergistic "Human+" collaboration, where AI empowers human experts to formulate highly effective, ethically sound sustainability strategies.

1.5 Delimitations

To maintain a focused and rigorous scope, this study establishes several clear delimitations regarding what it will not address. First, the research conducts no technical evaluation of AI systems. The study explicitly excludes any examination of AI coding, software engineering, computational mathematics, or technical architecture. It adopts a sociotechnical perspective, focusing entirely on how knowledge workers interact with and utilise algorithms in practice, rather than how those algorithms are technically constructed or mathematically optimised (Anthony, 2021; Leonardi, 2012; Pachidi et al., 2021). Second, the study does not engage in an end-to-end mapping of the entire organisational strategic process. The focus remains squarely on the discrete process of decision-making and human-machine interaction when making sustainability decisions, abstracting away from broader change management, public relations rollout, or supply chain execution sequences. Finally, there is no measurement of post-execution impact. The research does not attempt to quantitatively measure the long-term financial, environmental, or operational results of the decisions after they are implemented. Evaluating the long-term changes in carbon footprint or ESG ratings falls outside the scope of this investigation. Instead, the analysis is strictly confined to the cognitive, behavioural, and structural dynamics of the decision-making process itself at the managerial level. This study does not isolate or compare regulatory and stakeholder pressure differentials across manufacturing sub-sectors. While these external pressures are acknowledged as part of the socio-technical context, the analysis focuses on the cross-cutting cognitive and governance dynamics at the firm level rather than sector-specific compliance regimes.

1.6 Thesis Outline

To systematically navigate the complexities of algorithmic integration within sustainable SDM, this thesis is structured into six sequential chapters.

Chapter 1: Introduction has established the empirical and theoretical background, problematised the automation-augmentation paradox within corporate sustainability, and defined the research question guiding this study.

Chapter 2: Literature Review constructs the theoretical foundation. It unpacks Mintzberg et al.'s (1976) SDMP and examines the unique cognitive burdens of corporate sustainability. Crucially, it synthesises the "clash of logics" between the instrumental nature of artificial intelligence and the integrative, pluralistic demands of sustainable strategy.

Chapter 3: Methodology details the empirical architecture of the study. It outlines the qualitative, multi-case study design deployed across the Swedish manufacturing sector. Furthermore, it explains the application of the Gioia methodology, visualising how 1st-order informant narratives were systematically elevated into 2nd-order themes and overarching aggregate dimensions.

Chapter 4: The Empirical Findings section presents the results of the qualitative analysis, reporting the interviewees' accounts as first-order concepts grouped under second-order themes. This structure demonstrates how the raw data collection directly resulted in the discovered concepts.

Chapter 5: Discussion elevates the empirical findings into grounded theory. It conceptualises the central phenomenon of the Defensive Equilibrium, a state of systemic algorithmic under-use. The chapter culminates by answering the research question through the introduction of a novel, Phase-Gated Human-Centric AI Governance Approach, designed specifically to resolve the observed socio-technical friction across the SDMP.

Chapter 6: Conclusion synthesises the core arguments of the thesis. It clearly delineates the theoretical contributions to behavioral strategy and information systems literature, and identifies critical areas for future research, including the overarching ecological paradox of deploying energy-intensive AI for sustainability objectives.

2.0 Literature Review

2.1 Strategic Decision-Making:

Strategic Decision-Making (SDM) is defined by choices that are infrequent, non-routine, and complex, critically impacting an organisation's long-term health and survival (Shepherd et. al., 2024; Elbanna, 2006). These decisions are characterised by novelty, open-endedness, and a lack of a predetermined set of ordered responses, leading scholars to categorise the SDMP as "unstructured" (Mintzberg et. al., 1976; Elbanna et. al., 2020). Organisations typically begin this process with only a vague understanding of the problem and an equally vague concept of the eventual solution, necessitating a recursive and discontinuous approach to reaching a final commitment to action (Mintzberg et. al., 1976).

2.1.1 The Structure and Phases of the SDMP

The core structure of the SDMP is generally understood through three central phases: identification, development, and selection (Mintzberg et. al., 1976; Papadakis et. al., 1998). Within these phases, Mintzberg et al. (1976) identify seven central routines supported by communication, decision control, and political activity. The identification phase involves decision recognition and diagnosis, where stimuli are cumulated until a threshold is reached and cause-effect relationships are determined. The development phase, which typically consumes the greatest amount of organisational resources, revolves around search routines to find ready-made solutions and design routines to develop custom solutions. Finally, the selection phase involves screening a large number of alternatives, evaluation-choice routines to investigate feasibility, and an authorisation routine where the decision is ratified. Crucially, research indicates that these phases do not occur in a simple sequential order; rather, the process is dynamic, involving constant branching, cycling, and recycling as decision-makers move through various stages to comprehend complex issues (Eisenhardt & Zbaracki, 1992; Mintzberg et. al., 1976).

2.1.2 Rationality and Power Dynamics in SDM

A central theme in SDM literature is the tension between different forms of rationality. The rational model of choice assumes that actors have known objectives and gather complete information to select an optimal alternative (Eisenhardt & Zbaracki, 1992; Shepherd et. al., 2024). Procedural rationality, a key construct in this domain, refers to the extent to which the

decision process involves the systematic collection and analysis of information relevant to the decision (Dean & Sharfman, 1996; Papadakis et. al., 1998). However, empirical research consistently highlights cognitive limits to this ideal, leading to the paradigm of bounded rationality (Eisenhardt & Zbaracki, 1992). Under bounded rationality, decision-makers lack the capacity to process all possible data and instead "satisfice," seeking solutions that are "good enough" rather than optimal (Dean & Sharfman, 1996; Eisenhardt & Zbaracki, 1992). The relationship between rationality and effectiveness remains a subject of debate. Some research suggests that comprehensive processes correlate positively with performance in stable environments but negatively in unstable ones (Fredrickson & Mitchell, 1984; Eisenhardt & Zbaracki, 1992). Conversely, other studies find that procedural rationality is positively related to decision effectiveness regardless of the environment (Dean & Sharfman, 1996). However, rationality has been increasingly framed as multidimensional, suggesting that strategic decision-makers are rational in some ways but not others, adjusting their levels of analysis based on the specific characteristics of the decision and the complexity of the environment (Eisenhardt & Zbaracki, 1992; Papadakis et. al., 1998; Elbanna et al., 2020).

Parallel to rational analysis is the influence of power dynamics and political behaviour (Eisenhardt & Zbaracki, 1992). Organisations are accurately portrayed as political systems composed of coalitions of individuals with partially conflicting objectives (Shepherd et. al., 2024; Mintzberg et. al., 1976; Elbanna, 2006). Politics in the SDMP refers to those observable, often covert actions by which people enhance their power to influence a choice, such as coalition formation, lobbying, and the manipulation of critical information channels (Dean & Sharfman, 1996). Mintzberg et al. (1976) identified the "political supporting routine," noting that these activities serve to clarify power relationships or build consensus. The impact of politics on the SDMP is largely viewed as detrimental to effectiveness (Dean & Sharfman, 1996). Power aspects can slow the process, cause major delays, or distort information flows as managers prioritise parochial interests over organisational goals (Eisenhardt, 1989; Mintzberg et. al., 1976). Eisenhardt (1989) notes that centralised power in SDM is not necessarily beneficial, as it can limit the information flows required for quick methodical decision-making. Furthermore, contemporary research indicates that the source of political behaviour is often found in executive

coalitions and their need for control, with power struggles primarily catalysed by decision uncertainty and environmental hostility (Elbanna et. al., 2020; Shepherd et. al., 2024).

2.1.3 Heuristics, Mental Shortcuts, and Cognitive Overload

The transition from rational analysis to the use of heuristics, or mental shortcuts, is driven by the cognitive limitations of decision-makers and the information-processing demands of the environment (Atanasiu et. al., 2025; Phillips-Wren & Adya, 2020). Heuristics are defined as cognitive shortcuts that serve as potential aids to decision-making by focusing attention on particular aspects of information and allowing for quicker choices (Hodgkinson et al., 2023; Atanasiu et. al., 2025). They are applicable to strategic decisions because such choices are inherently judgmental, non-routine, and complex (Shepherd et. al., 2024; Maitland & Sammartino, 2015). The literature presents two divergent perspectives on heuristics. The "heuristics-and-biases" approach focuses on innate heuristics used reflexively, which can lead to systematic deviations from rational judgment (biases) and suboptimal outcomes (Hodgkinson et al., 2023; Atanasiu et. al., 2025). Conversely, the "fast-and-frugal" and "simple-rules" perspectives frame heuristics as "rational" adaptive tools (Bingham & Eisenhardt, 2011; Atanasiu et. al., 2025). These learned heuristics are distilled from experience and stored in portfolios that help executives allocate scarce resources to advantageous opportunities (Bingham & Eisenhardt, 2011; Atanasiu, 2021). The reliance on heuristics is often a direct response to cognitive overload. Information overload occurs when information demands exceed an individual's finite cognitive capacity, which can distort evaluation routines and lead to "paralysis by analysis" (Phillips-Wren & Adya, 2020). This overload acts as a primary decision stressor, negatively impacting decision quality and increasing the likelihood of cognitive errors (Phillips-Wren & Adya, 2020; Shepherd et. al., 2024). In these high-pressure situations, decision-makers are more likely to fall back on heuristics to reduce mental effort and reach closure on major decisions (Atanasiu et. al, 2025). Understanding how these mental shortcuts interact with rational steps and political dynamics is essential for a realistic view of how strategic decisions actually occur.

2.2 Decision-Making within Sustainability

Sustainable decisions require these diverse spatial and temporal considerations, the process for sustainability fundamentally differs from traditional decision-making. The process must become highly integrative, paradoxical, and synthesised, enabling managers to accommodate contradictory forces rather than eliminating them (Hahn et al., 2014; Slawinski & Bansal, 2015).

Considering the sustainable decisions, they have begun to transcend mere ESG compliance filters used for corporate reporting and public relations. Instead, they represent profound strategic shifts aimed at transforming a firm's sustainable value creation, such as implementing circular economy practices, waste minimisation, and the systemic reduction of carbon footprints (Guzmán et al., 2025; Teh & Corbitt, 2015). Because of this multidimensionality, sustainable decisions differ fundamentally from traditional strategic choices. Traditional strategy relies on an "instrumental logic," which operates on the assumption that the singular purpose of a firm is to maximise economic value (Gao & Bansal, 2013). This singular focus allows for a linear, reductionist, and highly analytical decision-making process, such as standard cost-benefit analyses, where social and environmental issues are often treated as sequential, peripheral distractions (Gao & Bansal, 2013). In stark contrast, sustainability requires an "integrative logic" because it introduces a pluralistic purpose, demanding that managers balance economic, environmental, and social goals simultaneously, rather than trading one off for another (Gao & Bansal, 2013). Furthermore, these decisions diverge significantly in their temporal orientations. Slawinski & Bansal (2015) discussed the difference that standard corporate decisions are heavily weighted toward short-term horizons, driven by quarterly earnings, rapid return-on-investment timelines, and immediate market pressures. Sustainable decisions, however, necessitate an intergenerational time horizon that balances present operational demands with the future needs of society and the biosphere.

2.3 Artificial Intelligence in Decision-Making

2.3.1 Definition and Types

AI represents a continuously evolving frontier of computational advancements that reference human intelligence to address complex decision-making problems (Berente et al., 2021). AI encompasses systems that display intelligent behaviour by analysing their environment and taking actions, with varying degrees of autonomy, to achieve specific objectives (Zhao & Gómez Fariñas, 2022). To support the SDMP, AI can be categorised based on its underlying functionalities and its level of autonomy. Functionally, AI spans several typologies: descriptive analytics (providing operational visibility), predictive AI (estimating future outcomes based on historical patterns), prescriptive AI (recommending optimal decisions under constraints), and generative AI (creating novel text, images, or scenarios) (Hutzschenreuter & Lämmermann, 2025). Based on the degree of autonomy and delegation, agentic systems are classified into four archetypes: reflexive (reacting directly to stimuli), supervisory (monitoring and guiding decisions to maintain norms), anticipatory (proactively anticipating needs without instructions), and prescriptive (taking independent, autonomous actions) (Baird & Maruping, 2021). Furthermore, in an organisational context, AI can assume roles ranging from a passive assistant that automates data processing, to an advisory "augmented AI" that navigates complex options, and ultimately to a fully autonomous decision-maker (Zhao & Gómez Fariñas, 2022).

2.3.2 Mitigating Behavioural Traps of Decision-Making

When integrated effectively, AI offers profound capabilities to mitigate the traditional traps of human decision-making, primarily by relaxing the severe constraints of bounded rationality (Csaszar et al., 2024). Human decision-makers operate with limited cognitive resources and frequently succumb to decision stressors such as information overload, time pressure, complexity, and high uncertainty (Phillips-Wren & Adya, 2020). Under these stressful conditions, managers instinctively rely on cognitive heuristics that save time but introduce systematic biases, riskier behaviours, and "choice overload" (Shrestha et al., 2019; Phillips-Wren & Adya, 2020). AI directly neutralises these constraints by processing massive datasets at near-instantaneous speeds, uniformly evaluating millions of alternatives without suffering from cognitive fatigue or choice paralysis (Shrestha et al., 2019). Furthermore, AI enhances the core

cognitive processes of search, representation, and aggregation, bypassing human emotional biases, such as "sugarcoating" or groupthink, to inject objective, data-driven rationality into top management teams (Csaszar et al., 2024).

2.3.3 New Traps, Algorithmic Biases, and Negative Sides of AI

However, the integration of AI into the SDMP requires close supervision, as it introduces entirely new cognitive traps, vulnerabilities, and algorithmic biases. Firstly, because machine learning models are fundamentally past-oriented, they can inherit and amplify human prejudices present in historical training data, leading to discriminatory or unethical outcomes (Zhao & Gómez Fariñas, 2022). Additionally, AI algorithms are highly susceptible to "input incompleteness" when human agents strategically manipulate or omit information fed into the system, requiring human domain expertise and vintage-specific technical skills to correct the machine's salience bias (Choudhury et al., 2020). A second critical trap is the inherent opacity or "black box" nature of AI. As deep learning models grow in computational complexity, their internal logic becomes increasingly inscrutable to end-users (Berente et al., 2021). This unexplainable nature generates intense epistemic uncertainty for professionals attempting to form critical judgments. When managers cannot decipher how an AI reached its conclusion, they frequently exhibit "algorithm aversion", a conscious or unconscious rejection of the algorithm in favour of their own flawed instincts (Brink et al., 2024). Conversely, without proper "AI interrogation practices" to reconcile divergent viewpoints, users may fall victim to automation bias by blindly accepting opaque recommendations, thereby reducing their critical thinking abilities (Lebovitz et al., 2022). Finally, the socio-technical integration of AI brings forth broader negative externalities and ethical blind spots. A prominent example is the "heuristics gap in AI ethics" and the mentalisation bias, in which humans heuristically treat AI as a weightless, intentional agent (Tamburrini, 2025). This intentionally veils the technology's physical materiality, causing stakeholders to overlook the massive energy consumption and carbon footprint associated with training "Red AI" models (Tamburrini, 2025). Additionally, AI disruption can prompt workers to engage in "symbolic conformity," where they pretend to utilise the AI while actively subverting it, ironically accelerating their own obsolescence and polluting organisational data (Pachidi et al., 2021). Thus, while AI resolves many traditional cognitive

traps, its strategic deployment demands robust governance, critical digital literacy, and deliberate human oversight to navigate its complex new pitfalls.

2.3.4 Managerial Attitude: Usage & Risk

Managerial attitudes toward the integration of AI into the SDMP for sustainability decisions are characterised by a complex interplay between perceived instrumental benefits and systemic risks. Top managers serve as the essential architects of digital strategy, where their cognitive frameworks and background significantly influence the successful adoption of AI for navigating long-term strategic complexity (Bevilacqua et al. 2025). The intention to utilise AI is primarily driven by perceived usefulness, the belief that the technology improves job effectiveness and trust (Nguyen et al. 2026). Trust is foundational for human-AI collaboration, directly influencing the decisional weight (typically 25% to 30%) that managers are willing to attribute to AI agents (Wen et al. 2025). This relationship is mediated by a manager's willingness to collaborate, which translates cognitive trust into the active behavioural act of delegating authority to the system (Wen et al. 2025). To manage these processes effectively, leaders must evolve beyond traditional styles, adopting skills such as agility, data-driven decision-making, and social-emotional intelligence to reconcile AI's analytical speed with human-centred values (Bevilacqua et al. 2025). However, significant perceived risks can dampen these positive attitudes. A major psychological barrier is the "uncanny valley effect," which occurs if a manager perceives an AI agent as possessing "free will" or high autonomy (Wen et al. 2025). Such perceptions can trigger identity threats and concerns regarding human uniqueness, negatively moderating the link between trust and collaboration (Wen et al. 2025). Furthermore, technological risks such as algorithmic bias, opacity in machine learning models, and potential ethical violations necessitate the implementation of robust governance frameworks (Nguyen et al. 2026; Bevilacqua et al. 2025). Ultimately, organisational readiness depends on fostering a digital culture that prioritises transparency and ethical use (Bevilacqua et al. 2025). To ensure sustainability goals are met, firms may undergo structural reorganisation, such as introducing the role of a Chief Digital Officer (CDO) to align AI initiatives with the broader environmental and strategic objectives of the organisation (Bevilacqua et al. 2025).

2.4 Organisational AI Governance

2.4.1 The Need for Governance

The integration of AI into strategic and operational routines fundamentally disrupts traditional organisational structures, necessitating novel and robust governance frameworks. A primary catalyst for this requirement is the "black box" problem and the algorithmic opacity associated with complex machine learning models. Knowledge workers frequently face a paradox wherein they deploy inscrutable algorithms without comprehending their underlying mathematical mechanisms. By concealing these analytical technologies, organisations risk creating inconsistent knowledge outputs and stunting the development of expertise among junior employees (Anthony, 2021). Furthermore, the introduction of AI demands substantial organisational change management by challenging established "regimes of knowing" (Pachidi et al., 2021). Pachidi et al. (2021) notes how algorithms act as new authorities, disrupting established valuation schemes and power arrangements. When technical experts introduce AI, incumbent workers often respond with "symbolic conformity", pretending to adopt the technology while covertly maintaining traditional practices to avoid real change. To navigate this organisational resistance, shifting power dynamics, and the complexities of data valuation, formal governance mechanisms are strictly required to define new procedural structures and ensure AI is adopted effectively (Schneider et al., 2023).

2.4.2 Ensuring Accountability

Governing AI requires a distinct allocation of responsibility that transcends traditional Information Technology (IT) management. As per Baird & Maruping (2021), because some AI systems can learn and act with varying degrees of autonomy, there is a persistent challenge regarding the delegation of rights and the assignment of accountability. Thus, the scholarly consensus of how allocating responsibility within the firm requires a multi-layered, interdisciplinary approach (Schneider et al., 2023; Papagiannidis et al, 2023; Batool et al., 2025). Effective AI governance demands senior leadership involvement, typically operationalised through an overarching AI oversight board or steering committee composed of executives, data scientists, legal experts, and ethicists (Gregory & Sircar, 2025; Schneider et al., 2023). This interdisciplinary structure is essential. It requires business domain experts to interpret outputs,

ethicists to evaluate algorithmic fairness, and legal officers to navigate regulatory compliance (Schneider et al., 2023). By establishing clear decision-making structures, organisations dictate exactly which human actors hold ultimate accountability for the strategic and ethical outcomes of the AI systems, ensuring that responsibility is never fully abdicated to the machine (Kellogg et al., 2020).

2.4.3 The Scope of Governance

AI governance is not singular; it requires the discrete management of multiple, interconnected technological and strategic artefacts. According to the AI Governance Cube, what is governed must be decomposed into three primary components: data, machine learning models, and the overall AI systems (Schneider et al., 2023). Data governance focuses on the quality, security, and valuation of the input data, mitigating the risk of algorithmic bias that stems from skewed historical datasets. Model governance entails overseeing the training, reproducibility, and logic of the specific algorithms, whereas system governance addresses the broader software architecture and human-AI interaction interfaces. Beyond these technical artefacts, governance must comprehensively address strategic alignment and ethical accountability. A holistic enterprise AI governance framework implements multidimensional "guardrails" to ensure that the technology aligns with corporate strategy (Desirable AI), technical robustness (Robust AI), ethical fairness and privacy (Dependable AI), and regulatory compliance (Lawful AI) (Gregory & Sircar, 2025).

2.4.4 Implementation Mechanisms and Policies

Schneider et al. (2023) notes that implementing AI governance is conducted through the AI Governance Cube via defined roles (structural), policies and performance metrics (procedural), and communication and training (relational). This process-oriented view complements the Four-Dimensional Guardrail system (Gregory & Sircar, 2025), which provides the strategic principle topics, such as ethical anti-bias measures and technical risk management, that these procedural mechanisms must enforce. A critical aspect of this implementation is managing the role of human oversight. According to Lebovitz et. al. (2022), to prevent the abdication of moral responsibility, organisations should implement "human-in-the-loop" protocols and cultivate "AI interrogation practices". This requires continuous training of the workforce, ensuring that

employees possess both the technical literacy to understand algorithmic limitations and the psychological safety to challenge opaque recommendations (Heger et al., 2025). Furthermore, organisational governance does not exist in a vacuum; it is heavily influenced by legislative and macro-level rules. AI governance must be categorised across team, organisational, industry, national, and international levels (Batool et al., 2025). While internal policies dictate daily operations, team and organisational artefacts are fundamentally subject to the influence of external legislative governance, such as the EU AI Act, which mandates specific risk-tiered documentation and conformity assessments that organisations must embed into their operational pipelines (Batool et al., 2025; Butt et al., 2025).

2.4.5 Temporal and Maturity Dynamics

The temporal dimension of AI governance must address both the chronological lifecycle of the technology and the evolving systemic maturity of the organisation. If the firm is developing their own AI, governance cannot be a retrospective compliance exercise, but occur continuously across the entire AI development lifecycle, including the pre-development (data collection), during-development (model training), and post-development (monitoring) phases (Batool et al., 2025). Moreover, as the technological capabilities of AI advance, governance structures must dynamically adapt to organisational maturity. When AI functions merely as an assisting tool (early adoption), governance focuses primarily on data accuracy and user compliance. However, as organisations move toward fully integrated governance and AI evolves into an agentic IS artefact, possessing the autonomy to independently develop protocols and select actions, the agency itself shifts away from humans (Baird & Maruping, 2021). This technological shift fundamentally alters the role of human oversight, transforming humans from direct operators into supervisors navigating complex forms of "conjoined agency" (Murray et al., 2021). Consequently, mature governance systems must dynamically calibrate their oversight mechanisms to match the escalating autonomy and inscrutability of the algorithms they deploy (Berente et al., 2021; Heger et al., 2025).

2.5 The Effect of AI in Sustainability Decision-Making

2.5.1 AI Heuristics linked to Sustainability

AI has emerged as a critical enabler for navigating this immense strategic complexity. Yet, deploying AI in this context introduces a profound paradox: while AI possesses the capability to drastically reduce carbon emissions and optimise resource use within companies, the technology itself generates a massive carbon footprint (Nishant et al., 2020; Tamburrini, 2025). The training of complex machine learning models, often shrouded in the weightless metaphor of "the cloud," is highly energy-intensive and heavily reliant on fossil fuel-powered data centres (Tamburrini, 2025). Consequently, stakeholders often suffer from a heuristics gap, overlooking the physical materiality of this "Red AI," which can inadvertently exacerbate the very climate crises it is deployed to mitigate if left unmanaged (Tamburrini, 2025). Despite this paradox, the impacts of applying AI for sustainable decision-making are transformative. AI systems empower firms to accurately track, manage, and achieve specific sustainability goals by analysing vast datasets across global supply chains and operations (Aljohani, 2025; Mustafa et al., 2025). Furthermore, predictive and prescriptive AI models allow both new ventures and established firms to make decentralised, data-driven decisions that actively tackle climate change (Vinuesa et al., 2020). For instance, AI-driven predictive analytics can forecast environmental risks and optimise renewable energy grids, while generative models can assist in designing eco-friendly, circular product life cycles (Aljohani, 2025; Guzmán et al., 2025).

2.5.2 Sustainability Heuristics in SDM

However, when circling back to the human decision-maker, managers utilising AI in the SDMP are still faced with persistent cognitive traps. In fact, these traps can be significantly amplified by the technology. Because the nature of sustainability is inherently integrative and paradoxical, the massive volume of high-speed outputs generated by AI can act as an additional source of information overload, exacerbating decision stressors such as time pressure and complexity rather than alleviating them (Phillips-Wren & Adya, 2020). Moreover, behavioural considerations heavily influence the use of AI, particularly given the unexplainable, "black box" nature of deep learning algorithms. When confronted with opaque AI outputs, managers face severe epistemic uncertainty, leading to polarised behavioural responses, either algorithmic appreciation or aversion, thereby rejecting their moral and strategic responsibility when

interrogating the data (Lebovitz et al., 2022). Additionally, integrating AI into sustainable decision-making fundamentally alters top management governance. Boards of directors and executives must adapt their oversight mechanisms to ensure algorithmic accountability and mitigate automated biases (Antonicic, 2020; Zhao & Gómez Fariñas, 2022).

2.6 Interpretation of Literature

The interpretation of the literature suggests that SDM is structured as a complex organisational process involving the identification, development, and selection of strategic alternatives (Mintzberg et al., 1976; Shrestha et al., 2019). Despite the normative pursuit of rationality, the process is inherently constrained by the cognitive boundaries of human agents and the influence of power dynamics, leading to a persistent reliance on heuristics and mental shortcuts (Elbanna et al., 2020; Powell et al., 2011). A primary interaction identified within the literature is that these inherent heuristics are not merely present but are significantly amplified when SDM is applied to sustainability decisions (da Silva & Filho, 2020; Hodgkinson et al., 2023). The "wicked" nature of sustainability, characterised by extreme multi-dimensionality, non-linearity, and a fast-changing environment, stretches human cognitive load to its limit (Alibašić, 2022; Nishant et al., 2020), often resulting in an "intertemporal paradox" where short-term financial returns are prioritised over long-term ecological resilience due to innate cognitive biases (Korteling et al., 2023; Slawinski & Bansal, 2015).

AI has emerged as a transformative solution capable of relaxing these constraints of bounded rationality by facilitating the processing of vast, multi-layered datasets beyond human perception (Csaszar et al., 2024; Mustafa et al., 2025). However, a critical interaction exists between AI's technical capability and the human element; the efficacy of AI-enabled SDM is strictly contingent upon the manager's attitude, trust, and propensity to use the technology (Brink et al., 2024). The persistence of "algorithm aversion", the reluctance of human decision-makers to rely on superior but imperfect algorithmic outputs, remains a significant barrier, indicating that managerial perception directly dictates the ultimate usefulness of AI within the firm (Burton et al., 2020). This creates a nuanced dynamic where AI is not a standalone solution but is co-constituted by the organisational and psychological context in which it is deployed

(Holmström & Hällgren, 2022). Furthermore, the role of AI in this context presents a profound paradox: while AI serves as a "cognitive fix" to reduce human heuristic distortions, it simultaneously introduces new heuristics and cognitive traps into the sustainable decision-making process (Raisch & Krakowski, 2021). The phenomenon of "black boxing", where the logic of analytical technologies remains opaque to the user, frequently leads to "automation bias," where managers may unthinkingly accept flawed algorithmic recommendations without interrogation (Anthony, 2021; Lebovitz et al., 2022). While current literature points to AI having the capacity to be agentic, acting autonomously to achieve goals, there is a robust consensus that this autonomy must be tempered by "meaningful human control" and "human-in-the-loop" configurations to ensure accountability (Baird & Maruping, 2021; Butt et al., 2025; Zhao & Gómez Fariñas, 2022). AI is thus viewed as an agentic partner that requires careful calibration with evaluative human judgment to avoid introducing unobserved variance into sustainability outcomes (Aljohani, 2025).

Strategic usage of AI therefore requires the robust integration of organisational and data governance frameworks. Data governance acts as the foundational mechanism, ensuring that the information fueling AI systems is high-quality, unbiased, and ethically sourced to prevent the automation of existing biases (Schneider et al., 2023). Simultaneously, organisational governance provides the structural guardrails, strategic, technical, ethical, and legal, necessary to manage the transition toward hybrid human-AI decision-making (Gregory & Sircar, 2025; Papagiannidis et al., 2023). Despite the proliferation of general governance models, the literature review identifies a critical research gap: there is a lack of research specifying how AI governance should be tailored for the unique complexities and ethically charged trade-offs inherent in sustainability strategy (Aljohani, 2025; Mustafa et al., 2025). Consequently, the aim of this research is to discover the implications or develop an approach for how AI governance is applied specifically to this context, harmonising algorithmic efficiency with the nuanced human judgment required for sustainable value creation.

3.0 Methodology

3.1 Research design

To investigate the tension between algorithmic efficiency and human control in SDM, this study employed a qualitative, multi-organisational qualitative study design. Because the research problem centred on the experience of corporate leadership, a qualitative approach was necessary to understand AI not merely as a software tool but as a complex Socio-Technical System (STS). Qualitative research was specifically appropriate here as it allowed for the exploration and understanding of the meaning of individuals or groups assigned to a social or human problem (Creswell & Creswell, 2023). This exploratory approach was vital for addressing topics that were difficult to reduce to a structured survey format (Hair et al., 2020). By reporting multiple perspectives, the research generated a holistic picture of the complex factors interacting within the "real life" strategic environment (Creswell & Creswell, 2023; Magnani & Gioia, 2023). The STS perspective considered how technology fundamentally altered organisational processes, power structures, and strategic logic rather than being an isolated change (Appelbaum, 1997). STS theory suggested that an organisation is a combination of social and technical parts that are interdependent; therefore, changes in the technical system, such as AI implementation, would inevitably affect the social network, including leadership and culture (Appelbaum, 1997). This view was necessary to build a complete overview of all AI implications in the decision-making context, considering its duality in balancing technical and social aspects. As per Narayan (2025), the STS ecosystem of AI involved various perspectives of institutions, users, data, AI systems, culture, regulations, and surrounding infrastructure.

Furthermore, a multi-organisational qualitative study was required to understand the patterns and drivers behind the decision to implement AI for sustainability strategy, capturing the prevailing organisational realities within the sector. According to Creswell and Creswell (2023), a multiple-firm approach allowed researchers to inductively develop findings for description and generalisation, focusing on the recurrent patterns that emerged across diverse organisational settings. This facilitated the discovery of "portable principles", theoretical constructs derived from the collective data set that could be applied to a wider variety of circumstances within the research scope (Magnani & Gioia, 2023). Crucially, while the ultimate objective lay in

identifying these broader theoretical patterns, the research design actively preserved the uniqueness and contextual nuance of individual accounts. Moreover, the inductive empirical analysis allowed for discovery-oriented research where employees were treated as insightful, "knowledgeable agents" capable of explaining their own reality (Magnani & Gioia, 2023; Gioia et al., 2013). This approach enabled a process of abductive analysis, characterised by the recursive movement between emerging data and extant literature to generate the "best explanations" for the observed phenomena (Magnani & Gioia, 2023). This abductive interaction ensured that the resulting approach remained grounded in the authentic lived experience of the participants while attaining the theoretical abstraction required to create a general model applicable to the firms covered within the scope (Magnani & Gioia, 2023; Schwaeye et al., 2025).

3.2 Data Collection

Data was primarily collected through semi-structured interviews. This approach followed the qualitative tradition of using multiple sources of data to build a complex, holistic picture of the phenomenon (Creswell & Creswell, 2023). Unlike quantitative methods focused on "horizontal breadth," this qualitative approach prioritised "vertical breadth" and in-depth collection to capture the subjectivity of the events described by participants (De Sordi, 2024). By gathering information through different perspectives, the study utilised data triangulation, a validity strategy where evidence from various sources was converged to build coherent themes (Creswell & Creswell, 2023; Magnani & Gioia, 2023).

3.2.1 Interview Protocol

The interviews were designed as semi-structured conversations, which allowed for a standardised core of questions while remaining flexible enough to follow the participant's train of thought (Creswell & Creswell, 2023). This study employed an interview protocol, a specialised guide for asking questions and recording answers, to ensure consistency across the five cases, with deviations if three interviews could not be achieved (Creswell & Creswell, 2023). Each firm participated with three stakeholder perspectives, which were addressed using a tailored question guide (Appendix 9.1, 9.2 & 9.3). While the inquiries were standardised, the researchers used

probes to elicit more detail or explanation when an informant's response was vague, to ensure case uniqueness was preserved (Creswell & Creswell, 2023). These probes (e.g., "Tell me more about that") helped the researcher go through the data by identifying the most significant themes during the interview process itself (Creswell & Creswell, 2023). To ensure the depth required for theory building, the researchers practiced active listening to capture as much information about the phenomenon as possible (De Sordi, 2024). The interviewers followed the "law of nondirection" by avoiding leading questions that suggested a preferred answer (Creswell & Creswell, 2023). By minimising the researcher's influence on the informant's account, the study upheld the Gioia assumption that "people at work are knowledgeable" (Magnani & Gioia, 2023).

The interviews were theoretically anchored in the Critical Incident Technique (CIT). CIT is a qualitative method used to collect functional or behavioural descriptions of events that are critical to the success or failure of an activity (Butterfield et al., 2005). Participants were asked to walk through specific, recent strategic initiatives regarding AI and sustainability. In accordance with CIT rigor, an incident was only deemed appropriate for inclusion if the participant provided full, precise details (Butterfield et al., 2005). Specifically, the report had to include: antecedent information (what led up to the event); a detailed description of the experience; and the outcome of the incident (Butterfield et al., 2005). This method grounded the STS perspective in "real-life" happenings rather than hypothetical ideals (Butterfield et al., 2005).

The interview guide was chronologically structured around the SDMP outlined by Mintzberg et al. (1976), utilising the CIT to ground informants in specific, real-world examples of data synthesis and strategic choice (Arnold et al., 2023; Elbanna et al., 2020). To operationalise the theoretical clash of logics, targeted probes were designed to examine how the instrumental, reductionist nature of algorithmic outputs interacted with the pluralistic and intertemporal demands of sustainability mandates (Gao & Bansal, 2013; Slawinski & Bansal, 2015). Finally, questions regarding output validation and algorithmic trust were directly derived from the literature on epistemic uncertainty and unengaged augmentation (Lebovitz et al., 2022), allowing for the systematic capture of the micro-level socio-technical friction experienced by managers.

3.2.3 Data Recording and Transcription

Interviews lasted approximately 30 to 60 minutes and were conducted via video conferencing (Teams or Zoom) to accommodate diverse organisational policies. The use of these digital platforms facilitated the synchronous recording of verbal communication, providing a robust saved record for subsequent analysis (De Sordi, 2024). All meetings began with explicit informed consent, notifying participants that the session would be audio-recorded for transcription (Creswell & Creswell, 2023). Transcription was performed using Sunet Scribe to ensure analytical accuracy through a denaturalised (verbatim) approach (Ahmad & Wilkins, 2025). This was followed by a manual verification by the researchers to correct errors in technical terminology and capture lost meanings that automated software failed to identify, such as contextual nuances, smiles, or significant pauses (De Sordi, 2024). This manual check was a critical reliability strategy, ensuring the transcripts did not contain obvious mistakes before the coding process began (Creswell & Creswell, 2023). During this check the text was cleaned and filler words, for example “um”, “uh” and “hmm”, were removed. Because qualitative research required a high sensitivity to perceive subjective nuances in the field, these recording and verification activities were conducted by the researchers themselves (De Sordi, 2024).

This study was conducted in strict accordance with GDPR standards to ensure the protection of human rights and participant safety (Creswell & Creswell, 2023). To minimise "self-preservation bias" and social factors that could influence responses, all interviewees, their job titles, and their respective companies remained strictly anonymous, utilising fictitious names, codes, or aliases to disassociate identities from specific responses (Creswell & Creswell, 2023; De Sordi, 2024). This anonymity was vital in SDM research, as it allowed "knowledgeable agents" to speak candidly about opaque or sensitive algorithmic processes without fear of professional repercussions (Creswell & Creswell, 2023; Magnani & Gioia, 2023).

3.3 Sampling and Participant Selection

To isolate the impact of AI from broader industry variables, the scope of this study was strictly narrowed to a single sector: manufacturing. This sector was selected because it is resource-intensive, collects vast amounts of data, and faces strong regulatory pressures requiring

detailed sustainability strategies. By focusing on a single industry, the study adopted an idiographic approach, aiming to discover detailed information about a restricted and specific subject rather than seeking broad generalisations (De Sordi, 2024). The study utilised purposive and criterion sampling to select five for-profit companies. Purposive sampling, also referred to as judgment sampling (Hair et al., 2020), was used in qualitative research to select sites or individuals that could best inform the researcher's understanding of the central phenomenon (Creswell & Creswell, 2023). Identifying robust inclusion and exclusion criteria was considered a crucial pillar in determining the rigor of such a sampling strategy (Ahmad & Wilkins, 2025). The selection of five firms aligned with established qualitative standards, which suggest that case studies typically include approximately four to five cases to allow for in-depth analysis (Creswell & Creswell, 2023). All selected firms were founded in Sweden to establish a cultural baseline. This geographic focus was justified by convenience and the region's significance in sustainability; Sweden was a global frontrunner, ranking 2nd in the 2025 Sustainable Development Goal Index with a score of 85.7 (Sachs et al., 2025). Furthermore, Scandinavia has a historical and theoretical link to the STS perspective, where the theory is considered nearly synonymous with work design and employee involvement (Appelbaum, 1997).

Within each firm, the case study aimed to target three distinct perspectives to ensure a holistic STS view of the AI ecosystem (Narayan, 2025). This triad of informants was essential to capture the interplay between technology, task, and people (Appelbaum, 1997):

- Sustainability Manager: Functioned as the primary business user of AI for strategy, representing the strategic intent and goal-setting of the firm.
- Data Expert: A specialist actor (De Sordi, 2024) who possessed a deep understanding of the AI system's technical considerations, data pipelines, and infrastructure (Narayan, 2025).
- Operational User: A business user whose daily operations were affected by AI-driven sustainability decisions, providing insight into the social and structural repercussions of the technology (Appelbaum, 1997).

By selecting these specific participants, the researchers ensured that they were "knowers" of the organisational reality, capable of explaining the "how" and "why" behind AI implementation in sustainability strategy (Magnani & Gioia, 2023).

Table 1: Table of Respondents

Firm	Firm Description		Job Role	Date	Duration	Note
Hydrangea	A vast multinational industrial infrastructure group. The firm manages immensely complex, high-volume global supply chains with rigorous, industry-leading sustainability and circularity mandates.	A	Sustainability Decision-maker	04/05	45min	Interviewed together
		B	Data Expert	04/05		
		C	Operational AI User	15/05	30min	
Lavender	A medium-large European manufacturer of interior building materials. The company manages significant cross-border production facilities and emphasises sustainable, bio-based material sourcing in its supply chain.	A	Sustainability Decision-maker	13/04	30min	
		B	Sustainability Reporting Specialist	20/04	30min	
			Reports	–	–	2024 Sustainability & Annual Report
Orchid	A medium-large international manufacturer of Fast-Moving Consumer Goods (FMCG) packaging. The firm operates globally with a strategic focus on circular economy models and the	A	Sustainability Decision-maker	24/04	30min	
		B	Sustainability Calculations Expert	11/05	30min	
		C	Operational AI User	26/05	30min	

	life-cycle tracking of short shelf-life products.					
Peony	A massive multinational heavy industrial engineering group. The firm leverages extensive global operations and highly advanced, data-driven manufacturing processes for complex tooling and equipment.	A	Sustainability Decision-maker	16/04	60min	
		B	Sustainability Reporting Specialist	21/04	60min	Interviewed together
		C	Sustainability Data Specialist	21/04		
Tulip	A large multinational medical technology company. The firm operates in highly regulated production environments, focusing heavily on quality compliance and sustainable healthcare innovation	A	Sustainability Decision-maker	24/04	40min	
		B	AI Expert	11/05	40min	
			Reports	—	—	2024 Sustainability & Annual Report

3.3.1 Exceptions to the Interview Protocol

As per Table 1 above, practical organisational constraints within the corporate environments necessitated deliberate deviations from the interview protocol to ensure the research represented its full complexity (Creswell & Creswell, 2023). To maintain the integrity of the data, the following alterations were implemented, focusing on enhancing data triangulation and capturing the collective dynamics of algorithmic oversight (Papagiannidis et al., 2023).

For the firm Lavender and Tulip, where the three-stakeholder threshold of strategic decision-maker, business user, and technical expert could not be met, the study integrated qualitative document analysis as a third aspect of triangulation (Creswell & Creswell, 2023). This process involved the collection and evaluation of the firms' 2024 sustainability reports and annual reports to provide an official record of strategic narratives (Maitland & Sammartino, 2015). Treating these documents as primary "text data" enabled the researcher to obtain the

specific "language and words" used by the organisation to frame its sustainability commitments (Creswell & Creswell, 2023). The significant implication of this change was the ability to cross-examine corporate "framing effects", where managers filter information to appear compatible with business interests, against the operational logic described by technical experts (Mustafa et al., 2025; Wright & Nyberg, 2017). By comparing archival data with interview transcripts, the research could identify potential discrepancies between public-facing sustainability goals and the internal realities of "black box" algorithmic technologies (Sakchuenyos et al., 2025; Anthony, 2021). Furthermore, this approach provided a stable, non-retrospective record that mitigated the risk of "recollection bias" or "fundamental attribution bias" often present in individual self-reporting (Maitland & Sammartino, 2015).

Furthermore, combined interviews were conducted for Peoney (participants B and C) and Hydrangea (participants A and B) due to internal accessibility issues. While this diverged from the individual stakeholder format, it provided a unique methodological advantage by allowing for the observation of "sociality" and "intersubjectivity" (Maitlis & Christianson, 2014). Since the implementation of responsible AI was inherently a "collective action" problem requiring "cross-discipline collaboration," combined sessions allowed researchers to witness how different roles mutually interpreted and calibrated algorithmic outputs to fit organisational goals (Heger et al., 2025; Pachidi et al., 2021). The implication of this interaction was the ability to observe "sensemaking" in real-time as actors engaged in "framing contests" to define what was worth knowing (Maitlis & Christianson, 2014). These joint sessions revealed how "shared interpretations" were formed within a firm's "regime of knowing," providing deeper insight into how the balance between AI and human judgment was negotiated within a "team effort" framework (Zhao & Gómez Fariñas, 2022; Pachidi et al., 2021). Consequently, while the individual "law of nondirection" was secondary in these instances, the researcher gained visibility into the "power dynamics" and "collective agency" that ultimately drove the firm's sustainability strategy (Pachidi et al., 2021; Creswell & Creswell, 2023).

3.4 Data Analysis

The collected data was analysed using the Gioia method within an inductive and later abductive coding approach. This method included applying the systematic approach of organising the data to discover 1st-order concepts, 2nd-order themes, and aggregated dimensions. By identifying reappearing ideas and comparing findings with previous algorithmic governance research, underlying patterns were visualised in a comprehensive data structure diagram. The method's purpose was to generate a grounded theoretical approach that bridged the identified empirical gap, translating subjective qualitative data points into an applicable approach for human-centric AI leadership (Gioia et. al., 2013; Magnani & Gioia, 2023). The collected data was analysed using the Gioia methodology within an inductive coding approach and later abductive interpretation. This systematic approach allowed the research to move beyond mere description to provide a deep, grounded explanation of the processes underlying AI-driven sustainability strategy (Gioia et. al., 2013; Magnani & Gioia, 2023). Qualitative analysis was inherently iterative, requiring the researchers to work "back-and-forth" between the data and emerging themes to identify meaningful patterns (Creswell & Creswell, 2023).

The analysis was executed through three key procedural pillars (Gioia et. al., 2013; Magnani & Gioia, 2023):

- 1st-Order Analysis: The process began with open coding across all 13 interviews and 2 reports, where the researchers adhered to informant/report terms to identify major themes and recurrent expressions (Gioia et. al., 2013; Magnani & Gioia, 2023). During this stage, the researchers performed an intensive reading of the transcripts/reports to highlight significant statements, then aggregated these statements into a range of concepts apparent throughout the interviews/reports.
 - The researchers employed the application NVivo throughout this coding procedure to sort and process the interview transcripts, and followed the same procedure in NVivo for the document analysis. While pure Gioia methodology traditionally favors manual, non-software-assisted coding to maximise textual immersion, NVivo was deliberately employed from the outset to enhance systematic data management and auditability across the 15 transcripts/reports.

- 2nd-Order Analysis: The analysis then transitioned to abductive reasoning. By systematically combining empirical data with existing socio-technical theories, the researchers distilled the 1st-order concepts into broader themes and aggregate dimensions. The researchers identified patterns and relationships within the data to formulate themes in theoretically relevant terms (Gioia et. al., 2013; Magnani & Gioia, 2023).
 - Here, the NVivo codebook was exported and manually analysed, taking individual and firm frequency into account, to report coherent underlying explanations, which resulted in a reduced number of distinct themes. Later, this process was repeated to find a few underlying aggregate dimensions categorising the current state of AI use in sustainability organisations. While a traditional Gioia approach allows aggregate dimensions to emerge entirely open-endedly, the final stage was abductively focused on dimensions that specifically categorised the socio-technical state.
- Data Structure & Model Development: The analysis culminated in two visual representations. The researchers constructed the data structure diagram (Table 2) to provide a graphic representation of how the research progressed from raw data to abstract dimensions (Gioia et. al., 2013; Magnani & Gioia, 2023). These static dimensions were then "set in motion" to develop a grounded theoretical approach (Figure 1), which illustrated the dynamic relationships within the AI ecosystem (Gioia et. al., 2013; Magnani & Gioia, 2023).

This method was specifically chosen because it treated organisational members as "knowledgeable agents" capable of explaining their own lived experiences (Gioia et al., 2013). Ultimately, the Gioia method aimed to generate a "best explanation" for how manufacturing firms navigated algorithmic efficiency and governance risks (Gioia et al., 2013; Magnani & Gioia, 2023). By reducing the methodological limits of deductive research, this combined approach between inductive and abductive reasoning allowed the study to push the boundaries of what was known and translated subjective data points into a rigorous approach for human-centric AI leadership (Gioia et al., 2013; Corley et. al., 2021). Furthermore, the methodology facilitated the discovery of "portable principles", theoretical constructs derived from the specific industrial

context that maintained relevance across the broader Swedish manufacturing sector (Magnani & Gioia, 2023). This approach was particularly advantageous given the multidimensional complexity of sustainability-oriented AI adoption, which required an abductive interaction between emerging data and established theoretical approach to explain how managers balanced economic performance with environmental stewardship (Schwaeke et al., 2025). By recursively matching empirical patterns with extant literature, the research identified deep structures such as the "automation–augmentation paradox" that were critical for governing strategic decision-making in transitional environments (Raisch & Krakowski, 2021; Krakowski et al., 2023). This ensured the resulting model was both empirically grounded in participant realities and theoretically robust (Gioia et al., 2013).

Table 2: Gioia Method Chart

1st Order Concepts	2nd Order Themes	Aggregate Dimensions
Human oversight required for AI inputs/outputs for quality and risk considerations	Heavy need human oversight & control	Defensive Personal Attitudes Toward AI
High awareness and efforts to minimise AI hallucinations		
Varying individual awareness of the correlation between the quality of the prompt and the AI output	Learning curve hindering usage	
Slow AI implementation due to knowledge gaps, corporate cultural and reluctance to change		
AI requires education, feedback and training for users		
Claims not to see bias in the AI system	Mistrust in AI outputs	Socio-Technical Friction in Sustainability Strategy
Desire to see AI output traceability		
Mistrust in AI due to low input data quality and hallucination in the output	Sustainability data quality as a bottleneck	
High focus on improving data quality and data structure		
Introduction of internal data centres/data owners to improve data quality	Sustainability nuance & complexity	
Sustainability requires nuance not captured by concise AI outputs		
Seeing AI as a tool for reviewing sustainability assumptions	AI as efficiency tool not strategic sustainability driver	Organizational Risk-Averse Governance Orientation
AI output assisting limited use cases, but not used for decision-making		
A few firms developing AI agents for strategic use cases		
AI as a tool for efficiency for certain tasks, not replacing human expertise	Governance focused AI implementation	
Inconsistently communicating when and how AI is used internally		
Implementing a strategy for AI to improve internal infrastructure and accuracy		
Strategic use of AI for competitiveness through top-down reinforcement		
Demand for top-down sustainability governance for effective AI use		
Setup of AI excellence teams to govern corporate AI use	Focus on safety, compliance & regulation within sustainability	
High focus on cybersecurity, safety, and compliance over strategic value creation		
Large concerns for changing regulations and local variation	Internal sustainability AI systems for control	
Intentionally blocking AI from compliance reporting and social human rights data		
Predominant use external AI providers, typically Copilot for cybersecurity and system compatibility		
Concern for public data or proprietary data impacting outputs	Internal sustainability AI systems for control	
In-house solutions built for cybersecurity concerns and specific requirements.		

3.5 Validity and Reliability

To consider the validity and reliability of this study, it should be noted that the collected data resulted in subjective statements found in the interviews and reports (De Sordi, 2024). Interviewees presented their own interpretations of events and the company's strategic direction, which were inherently tied to human perceptions and feelings attached to an experience (De Sordi, 2024; Creswell & Creswell, 2023). Furthermore, sustainability reports and strategies were often published to boost public opinion and shareholder value, leveraging the "framing effect" or the "halo effect," which could contribute to cognitive biases in the data (De Sordi, 2024; Hair et al., 2020). To mitigate these biases and ensure qualitative accuracy, the researchers utilised data triangulation to verify and compare the strategic narratives of decision-makers against the technical, operational realities described by system providers and specialist actors (Creswell & Creswell, 2023; Magnani & Gioia, 2023). This process involved converging evidence from multiple sources to build coherent themes (Creswell & Creswell, 2023). Furthermore, the use of the CIT anchored the discussion in past realities rather than hypothetical ideals, limiting the opportunity for fabricated narratives (Butterfield et al., 2005; Flanagan, 1954). Within the CIT framework, the credibility of an account was deduced from the level of full, precise details provided by the participant (Butterfield et al., 2005, p. 602; Flanagan, 1954).

To ensure qualitative reliability, all interviews were recorded via digital platforms to provide a robust saved record (De Sordi, 2024). Transcripts were checked for obvious mistakes, and a manual verification process was conducted by the researchers to ensure the text accurately reflected the nuances of the dialogue (Creswell & Creswell, 2023; De Sordi, 2024). Finally, the Gioia methodology provided a systematic research approach that explicitly demonstrated data-to-theory connections, addressing accusations that qualitative research was merely impressionistic (Gioia et. al., 2013; Magnani & Gioia, 2023).

3.6 Ethical Considerations

This study was conducted as a truth-seeking function, adhering to core values of objectivity, honesty, and accountability (Hair et al., 2020). First and foremost, the researchers had an

obligation to respect the rights, needs, and values of the informants (Creswell & Creswell, 2023). All participants were provided with a statement of the research purpose and had to provide explicit informed consent before data collection commenced (Creswell & Creswell, 2023; Hair et al., 2020). To speak candidly about opaque or sensitive algorithmic processes, strict individual and organisational anonymity was guaranteed (Creswell & Creswell, 2023). The researchers utilised fictitious names, codes, or aliases to disassociate specific identities from the responses in the final report (Creswell & Creswell, 2023; De Sordi, 2024). This anonymity was vital to allow "knowledgeable agents" to speak without fear of professional repercussions (Magnani & Gioia, 2023). Furthermore, investigators had an obligation to protect participants from potential professional harm and ensure the research did not expose subjects to unreasonable stress or risk (Creswell & Creswell, 2023; Hair et al., 2020). All collected data was stored securely using appropriate safeguards and used only for the intended purpose (Hair et al., 2020).

3.7 Limitations

A primary limitation of this study was that the qualitative research did not seek statistical generalisability to a broad population, but rather focused on the "vertical breadth" and particulars of a specific context (De Sordi, 2024; Creswell & Creswell, 2023). By narrowing the scope strictly to the Swedish manufacturing sector, the results were idiographic, meaning the "portable principles" derived might not be directly transferable to different industries or cultural settings (De Sordi, 2024; Magnani & Gioia, 2023). The reliance on retrospective self-reports within the CIT framework introduced the risk of selective recall or memory bias (Butterfield et al., 2005). Participants might selectively remember dramatic events while forgetting routine but critical interactions, particularly those that had occurred several months prior (Butterfield et al., 2005; Flanagan, 1954). Another limitation lay in assigning participants to a single category (Sustainability, Data, or Operational) for interview-guide selection, as many individuals occupied overlapping roles that did not fit neatly into one perspective. Furthermore, while the researchers strove for information richness, the sample size was limited to five firms (Ahmad & Wilkins, 2025; Creswell & Creswell, 2023). Additionally, the adoption of combined interviews for Hydrangea and Peony introduced a distinct relational limitation, as the presence of a peer could elicit social desirability bias, prompting participants to conform to perceived organisational

norms rather than providing candid individual accounts (Maitlis & Christianson, 2014). This methodological exception risked the suppression of idiosyncratic dissent or "unique human knowledge," as the interaction was mediated by internal power dynamics and a pursuit of intersubjective agreement (Pachidi et al., 2021). Consequently, joint sessions could emphasise a unified "regime of knowing," potentially masking the "sensemaking" tensions and "framing contests" that individual interviews were better equipped to surface (Maitlis & Christianson, 2014; Pachidi et al., 2021). Furthermore, utilising sustainability and annual reports as a method of triangulation for Lavender and Tulip is limited by the inherent purpose of these documents. Because they are often published to enhance public perception and shareholder value, they frequently leverage the framing effect, potentially introducing cognitive biases into the data.

To mitigate these methodological constraints, several validation strategies were systematically employed to enhance the study's "trustworthiness," "credibility," and "confirmability" (Creswell & Creswell, 2023; Papagiannidis et al., 2023). To address the inherent limitations of idiographic research and the small sample size, the study provided a thorough description of the Swedish manufacturing context, which transported the reader to the setting and established a solid approach for "transferability" to other industrial or cultural environments (Creswell & Creswell, 2023). The potential for "memory bias," "revisionism," and the observed speculative drift toward hypotheticals within the CIT framework was rigorously mitigated through "triangulation" and, for exceptions, with reports (Maitland & Sammartino, 2015). This was an adequate substitute for three interviews due to the effect of cross-checking participant recollections against "official records", which minimised the risk of "distorted retrospective accounts" and "fundamental attribution bias" (Maitland & Sammartino, 2015; Shrestha et al., 2019). This multi-modal approach ensured that the final governance approach remained grounded in "lived experiences" while maintaining the objectivity required for theory building (Schwaeke et al., 2025; Creswell & Creswell, 2023).

4.0 Empirical Data

4.1 1st Order Concept Findings

In accordance with the Gioia methodology, the following empirical findings report 1st-order concepts and interviewees' quotes under the headings of 2nd-order themes. This became the foundation for the 2nd-order themes and aggregate dimensions in chapter 5. The 1st order concepts remain focused on the interviews, adhering closely to the respondents' phrasing and perspectives regarding AI usage in sustainability tasks. Informants frequently noted the heavy organisational focus on cybersecurity, data compliance, and the structural limitations of data quality over strategic long-term planning. There was also a strong consensus around the need for ongoing human oversight, the presence of operational learning curves, and the nuanced complexities of sustainability data that standard AI models fail to capture (Appendix 9.4).

4.1.1 Heavy need human oversight & control

AI is framed as a tool requiring constant human control, where professionals ensure accuracy through oversight and proactive efforts to reduce hallucinations. Across firms, participants expressed a strong conviction that AI systems require continuous human oversight, particularly in sustainability contexts where interpretive nuance, regulatory precision, and ethical accountability are central. Interviewees at Peony articulated this most explicitly. Peony A stated:

“I wouldn't let AI decide anything. I decide [...] it's truly important to weigh things and understand” (Peony A)

This underscores a deeply held belief in human responsibility. Peony C similarly emphasised the need for critical evaluation, noting that they are “very critical when it comes to the answer [...] because it's limited to open data sources.” Orchid participants expressed comparable concerns but framed them in terms of verification practices. Orchid A described a meticulous process:

“Read through, go to the sources, make sure that the AI tool is referring to the right sources [...] it's like a teacher checking the homework of a student.” (Orchid A)

Orchid B added that users must possess foundational knowledge to evaluate outputs, explaining that without such knowledge, “you just have to trust the information,” which they viewed as

unacceptable. Lavender interviewees echoed this emphasis on oversight. Lavender A noted that “you always need to cross-check and be aware that it’s a tool [...] dependent on the information you have inputted,” while Lavender B highlighted the practical implications: “A lot of our manual work is trying to fill in the gaps or correcting mistakes.” Tulip A reinforced this stance by explaining that their internal system was intentionally designed to require human validation: “We wanted a certain amount of human validation of the data.” Hydrangea C added a governance-oriented perspective, describing how their organisation cultivates “confidence but confidence with caution,” with user feedback “closely monitored” to ensure responsible use. Thus this creates a shared emphasis on human oversight and reflects a broader skepticism toward AI’s reliability and a recognition of the high stakes inherent in sustainability decision-making.

4.1.2 Sustainability data quality as a bottleneck

Organisations treat sustainability data quality as a limiting factor, prompting investments in improved data structures and the creation of dedicated internal roles and centres to safeguard accuracy. Data quality emerged as one of the most pervasive constraints on AI adoption.

Interviewees at Hydrangea articulated this challenge most vividly. Hydrangea B explained:

“Because we work with thousands of suppliers [...] the data may not be in the right shape or right format. It requires a lot of manual effort.” (Hydrangea B)

Hydrangea C added that sensitive data cannot be fed into AI systems due to confidentiality risks: “There is a risk [...] confidential data could leak out [...] we would not have wanted to feed the tool with that.” Lavender participants similarly emphasised data limitations. Lavender A noted that “we always find new emission factors [...] it’s still a challenge with some production sites to really get the data,” while Lavender B stressed that “a lot of our challenge [...] is not really anything to do with AI. It’s data architecture and data quality.” Tulip A linked these issues directly to AI limitations, stating that “our data quality [...] is too low to allow an AI to interpret it.” Peony B reinforced this theme, explaining that “data accuracy [...] is one of the main issues why I’m not using it.” Multiple companies discuss the quality, structure, and accessibility of sustainability data significantly limiting the potential of AI to support analysis or decision-making. While the specific data challenges differ, supplier heterogeneity at Hydrangea,

architectural fragmentation at Lavender, accuracy concerns at Peony, the overall pattern is consistent. To mitigate these challenges, some organisations attempt to strengthen internal reporting structures. Tulip Reports described efforts to enforce consistency through “standardised reporting protocols across the entire organisation: uniform reporting guidelines are applied across all Business Areas and Corporate Functions to ensure consistency in data collection and validation.”

4.1.3 AI as efficiency tool not strategic sustainability driver

AI's role in sustainability work remains primarily efficiency-oriented, confined to supportive functions rather than strategic influence, with human expertise still central and only isolated cases exploring more advanced AI agents. Participants consistently framed AI as a tool that enhances efficiency, particularly in summarisation, drafting, and information retrieval, rather than as a strategic driver of sustainability transformation. Hydrangea C emphasised productivity gains, describing how AI reduces repetitive requests to experts:

“People were disturbing an expert many times [...] this was creating a lot of productivity issues” (Hydrangea C)

Lavender interviewees highlighted communication benefits. Lavender A noted that “AI can frame things in the right way [...] I’m a pretty good writer but AI can always help make that a bit more crisp,” while Lavender B described using AI to generate draft narratives that still required manual adjustment. Tulip B emphasised iterative development, explaining, “After using the first MVP version [...] it’s easier to understand what bigger solution we could build.” Orchid participants stressed the limits of AI’s strategic role. Orchid B stated, “You wouldn’t go and ask AI to write you a whole new sustainability strategy”. Orchid C claimed, “Goals or objectives [...] we have decided on ourselves, it has been more using AI to make executive summaries” In comparison, Orchid A described AI as a tool for acceleration rather than strategy.

“AI is a tool to help you speed up drafting text, proposals, generating ideas, building structures [...] but it doesn’t take away the responsibility for you as an individual.” (Orchid A)

A later example from Orchid A states: “We needed to argue the case for the local authority to get the exemption [...] Co-pilot helped me find the right paragraphs in the waste regulation.” Peony A offered a similar perspective, explaining that AI is useful for “gathering data [...] colliding data [...] compare data and do pivot tables,” but that strategic insight still derives from “blood, sweat, tears and toil.” Across firms, AI is positioned as an operational accelerator rather than a strategic actor.

4.1.4 Governance focused AI implementation

Firms anchor AI adoption in governance structures, formalising top-down strategies, establishing AI excellence teams, strengthening internal infrastructure, and demanding clearer sustainability governance, while still communicating AI use inconsistently across the organisation. Governance structures play a central role in shaping how AI is implemented and used. Tulip B emphasised multi-layered governance, explaining that “there is a review process in place [...] it depends on the use case and involves different teams.” Hydrangea participants described centralised oversight. Hydrangea B explained: “We have a team called AI Centre of Excellence [...] they evaluate the need” while Hydrangea C explained that their internal AI tool is restricted to prevent data leakage:

“We are restricting that this tool will not bring information outside of our company premises.” (Hydrangea C)

Lavender interviewees highlighted strict platform controls. Lavender A stated, “We’re not allowed to use other platforms and share information [...] that’s the way it’s governed,” and Lavender B described extensive checks and balances in sustainability reporting. Adding to this Lavender Reports mentions:

“Risk management is built into every process, and various methods are used to assess and limit risks and to ensure that the risks to which [Lavender] is exposed are managed in accordance with established policies, instructions and follow-up routines designed to reduce potential risks and to promote correct accounting, reporting and information”
(Lavender Reports)

Orchid B linked governance to cybersecurity training, noting, “It’s part of the cybersecurity policy [...] I’ve done the e-learning for it.” Peony A framed governance as responsible AI usage, asking, “Was it necessary to generate that?” In summary, governance functions as a protective mechanism that shapes who can use AI, how it can be used, and what data it may access.

4.1.5 Focus on safety, compliance & regulation within sustainability

AI use in sustainability is shaped primarily by safety and regulatory considerations, with firms prioritising secure external tools, avoiding AI in high-risk compliance areas, and grappling with evolving regulatory demands across jurisdictions. Given the regulatory nature of sustainability work, participants expressed a strong emphasis on safety, compliance, and accuracy. Hydrangea B articulated this most clearly, stating:

“We don’t use AI for compliance reporting because compliance reporting has to be very accurate. We cannot have errors and it has to be really accurate for the authorities” (Hydrangea B)

Hydrangea C added that even with encryption, “there is still some kind of threat which looms [...] you cannot control it or put it in a vault.” Lavender participants emphasised compliance-driven design. Lavender Reports explained that “The Sustainability Scorecard [...] ensures products comply with the EU Ecodesign Regulation.” Tulip A described their intentionally constrained internal system:

“We wanted it to be a closed environment [...] to minimise the risk of hallucinations.” (Tulip A)

Orchid interviewees stressed ethical boundaries. Orchid A stated, “I would not delegate the decision to AI [...] especially not for social sustainability topics,” while Orchid B emphasised the need to verify AI-generated references. Peony B raised concerns about the environmental impacts of computing power, noting, “When you’re wasting a lot of energy and resources [...] that doesn’t generate any value.” In comparison, Orchid C presents the opposing perspective noting: “from a data/IT security perspective, there’s a lot of governance in place when it comes to AI usage across the group [...] perhaps it goes a bit too far in certain areas” speaking to a

disagreement with firm policy. However, the firms' safety and compliance concerns continue to significantly shape AI adoption.

4.1.6 Mistrust in AI outputs

Confidence in AI remains constrained, as professionals question outputs affected by data quality issues and hallucinations, reinforcing the need for clearer traceability to validate results. Mistrust in AI outputs was expressed across several firms, particularly regarding hallucinations, misinterpretations, and incomplete reasoning. Lavender B described frustration with AI's inability to admit uncertainty, stating: "It will hardly ever say "I don't know" [...] you get into a loop". Orchid participants expressed similar concerns. Orchid B noted that AI "does have instances of [...] making up the rest of what it does not know," and added that they need to "check [AI] for potential greenwashing claims", underscoring the interpretive complexity of sustainability communication. On the other hand Orchid A emphasised personal accountability:

"If it's pointing to the wrong parts of the regulation or it's interpreting it the wrong way [...] if I want to make the proposal, I need to own it and therefore understand it on a fundamental level" (Orchid A)

They continue by stating "I could not blame AI for having talked me into taking the wrong decision [...] you should admit your mistakes and be prepared to change the decision if that's needed." Arguing that human accountability in AI-supported work, should remain central throughout AI implementation, emphasising that responsibility for decisions cannot be delegated to algorithmic systems. Peony interviewees reinforced this theme. Peony C explained that they are "very critical when it comes to the answer [...] because it's limited to open data sources," and Peony A added, "I wouldn't want AI to represent me [...] I want a person in flesh and blood". Thus, mistrust reinforces the need for human oversight and careful validation. On the other hand, Hydrangea participants A and B oppose this, articulating a technically informed stance toward AI, combining high trust in its utility with a nuanced awareness of its accuracy constraints. This dual orientation suggests a mature, experience-based understanding of AI's epistemic boundaries.

4.1.7 Learning curve hindering usage

AI adoption is slowed by a steep learning curve, marked by uneven understanding of how prompting quality shapes outputs, broader knowledge gaps and cultural resistance, the need for ongoing user training and feedback, and limited awareness of potential bias in the system. Participants noted that AI adoption varies by age, digital literacy, and organisational readiness. Lavender interviewees described the trial-and-error nature of early adoption. Lavender B stated:

“You just get better at prompting as you use the tool more and more”

(Lavender B)

Similarly Orchid C describes their prompting technique as “mak[ing] sure to provide enough context and background [information].” On the other hand, while Lavender A observed that “I’m much more into AI [...] than many other people within the company.” Hydrangea C described extensive training efforts, including live demos and Q&A sessions designed to build “confidence but confidence with caution.” Orchid A linked adoption to generational differences, noting, “Younger people are quicker to jump into AI tools [...] I’m trying to stay abreast.” Peony A described their organisation as being in a “test and try stage [...] let the cows loose in spring,” illustrating a culture of experimentation that nonetheless results in uneven maturity across employees. As a result it becomes prominent that learning curves shape both the pace and quality of AI adoption.

4.1.8 Sustainability nuance & complexity

Sustainability work demands nuance that AI’s concise outputs often fail to capture, positioning AI primarily as a tool for reviewing and stress-testing sustainability assumptions rather than generating them. Interviewees emphasised that sustainability involves complex, context-specific judgments that AI struggles to capture. Tulip A described how AI surfaced unexpected insights, stating:

“The AI had identified something that we wouldn’t have done [...] If you’d asked anybody on factory floors what the what the most common raw material is, they would have picked something else” (Tulip A)

Hydrangea A highlighted regulatory complexity, noting that “regulations are more local now [...] you need to comply everywhere.” Orchid A emphasised multidimensionality, asking “Is it scope three? Is it the products? Is it materials?” Similarly Tulip Reports describes how “insights from key stakeholder groups [...] inform strategic decision-making, ensuring alignment with regulatory requirements, industry best practices, and sustainability objectives.” Peony A illustrated complexity across the value chain, describing how sustainability considerations extend from machinery to material sourcing. Lavender B emphasised the need for manual refinement, noting that sustainability reporting requires “manual adjustment” to ensure nuance and accuracy. Hydrangea C explained that they restrict AI access to curated data lakes to avoid “garbage in, garbage out,” underscoring the need for human judgment in selecting relevant information. Hence, sustainability complexity limits the extent to which AI can automate or interpret sustainability-related decisions. Adding in a deeper contextual understanding, Hydrangea A and Orchid A, independently observed a broader geopolitical recalibration of sustainability priorities, noting that external crises have contributed to a global deprioritisation of environmental agendas. They note: “Everyone is kind of shifting down the focus on sustainability... across the globe due to wars,” (Hydrangea A), while reflecting on how sustainability has “taken a bit of a backseat,” (Orchid A), thereby heightening the strategic imperative to embed it directly into value creation rather than treat it as an auxiliary concern. This contextual framing introduces an important macro-level tension and meaningfully shapes organisational sensemaking. Similarly, interviewees A and B from Peony raised concerns regarding the environmental footprint of AI itself, pointing to an often-overlooked paradox in which sustainability ambitions rely on resource-intensive digital infrastructures.

4.1.9 Internal sustainability AI systems for control

Firms build internal sustainability-focused AI systems to retain control over sensitive data—avoiding risks linked to public or proprietary data exposure—and to meet strict cybersecurity and specialised organisational requirements. Some firms have developed internal AI systems to ensure confidentiality and compliance. Hydrangea A explained: “We have similar solutions to Copilot [...] built ourselves in a secure environment”. Hydrangea B noted that this approach may slow innovation but ensures safety: “We might be delayed [...] but we need to ensure compliance and security.” Tulip B described strict review processes involving “risk

calculations for each specific use case,” and Tulip A explained that their system was intentionally constrained:

“We wanted to primarily shut off functionality [...] we were very keen that it couldn't access the internet; that it couldn't access a lot of documentation. It could only do a very limited amount of sort of responses within quite a closed environment” (Tulip A)

Peony B described how their company selected Copilot as the “official backbone” of their AI stack due to security and integration benefits, emphasising that internal AI use must “give more than it takes.” As a result, internal AI systems reflect a preference for control, security, and compliance over speed or experimentation in sustainability-related AI adoption.

4.2 Summary of 1st Order Concepts

Table 3: 1st Order Concepts Respondent Data

1st Order Concept	Lavender			Peony			Hydrangea			Tulip			Orchid		
	A	B	C	A	B	C	A	B	C	A	B	C	A	B	C
1. Human oversight required for AI inputs/outputs for quality and risk considerations	Y	Y		Y	Y	Y		Y	Y	Y	Y	Y	Y	Y	Y
2. High awareness and efforts to minimise AI hallucinations	Y	Y						Y	Y	Y			Y	Y	Y
3. High focus on improving data quality and data structure	Y	Y	Y	Y	Y		Y	Y	Y	Y	Y	Y			
4. Introduction of internal data centres/data owners to improve data quality							Y	Y	Y		Y				
5. AI output assisting limited use cases, but not used for decision-making	Y	Y		Y	Y		Y			Y	Y		Y	Y	Y
6. A few firms developing AI agents for strategic use cases					Y		Y	Y		Y	Y				
7. AI as a tool for efficiency for certain tasks, not replacing human expertise	Y	Y		Y	Y		Y		Y	Y	Y	Y	Y	Y	Y
8. Inconsistently communicating when and how AI is used internally	Y	Y						Y							Y
9. Implementing a strategy for AI to improve internal infrastructure and accuracy		Y					Y	Y			Y	Y			
10. Strategic use of AI for competitiveness through top-down reinforcement					Y	Y		Y						Y	
11. Demand for top-down sustainability governance for effective AI use		Y	Y	Y		Y		Y			Y				
12. Setup of AI excellence teams to govern corporate AI use							Y	Y	Y	Y	Y				
13. High focus on cybersecurity, safety, and compliance over strategic value creation			Y	Y	Y		Y	Y	Y	Y	Y				Y
14. Large concerns for changing regulations and local variation			Y	Y	Y	Y	Y	Y					Y		Y
15. Intentionally blocking AI from compliance reporting and social human rights data				Y				Y	Y				Y	Y	Y
16. Predominant use external AI providers, typically Copilot for cybersecurity and system compatibility	Y	Y		Y	Y	Y							Y	Y	Y
17. Desire to see AI output traceability	Y				Y	Y								Y	Y
18. Mistrust in AI due to low input data quality and hallucination in the output		Y		Y	Y	Y					Y		Y	Y	
19. Varying individual awareness of the correlation between the quality of the prompt and the AI output	Y	Y		Y	Y	Y							Y	Y	Y
20. Slow AI implementation due to knowledge gaps, corporate cultural and reluctance to change	Y	Y		Y	Y						Y		Y		
21. AI requires education, feedback and training for users	Y			Y			Y		Y		Y		Y	Y	
22. Claims not to see bias in the AI system	Y	Y												Y	Y
23. Sustainability requires nuance not captured by concise AI outputs	Y		Y			Y			Y				Y	Y	Y
24. Seeing AI as a tool for reviewing sustainability assumptions	Y	Y					Y			Y					
25. Concern for public data or proprietary data impacting outputs				Y	Y					Y					Y
26. In-house solutions built for cybersecurity concerns and specific requirements.					Y		Y	Y	Y	Y	Y				

5.0 Analysis and Discussion

The integration of AI into the SDMP for sustainability introduces a profound set of organisational and cognitive paradoxes. As established in the problematisation of this thesis, decision-makers are tasked with navigating inherently complex, multidimensional sustainability topics that demand an integrative logic. While AI is theoretically positioned as a cognitive remedy to alleviate information overload and bounded rationality, the empirical findings of this study demonstrate that the technology significantly exacerbates existing tensions and introduces novel systemic challenges. The initial theoretical friction between human ethical judgment and algorithmic efficiency fails to capture the full extent of the operational reality. Instead, the complexity is far more extensive, rooted in the foundational friction between advanced technical capabilities and immature corporate data infrastructures. Chapter 4 presented the empirical reality observed within the participating firms, detailing how sustainability managers, data experts, and operational users interact with AI. This chapter interprets those findings through the theoretical lenses established in Chapter 2. The discussion is structured to analyse ten 2nd-order themes that manifest into three overarching aggregate dimensions: Defensive Personal Attitudes Toward AI (the micro-level phenomenon), Socio-Technical Friction in Sustainability Strategy (the meso-level phenomenon), and Organisational Risk-Averse Governance Orientation (the macro-level phenomenon). Crucially, these dimensions do not exist in isolation; they continuously interact, creating reinforcing loops and potential path dependencies that dictate the firm's strategic trajectory. By assembling these dimensions in a micro-meso-macro progression and applying abductive reasoning, this chapter identifies the deep theoretical structures defining the current state of AI in corporate sustainability.

5.1 Interpreting 2nd Order Themes into Aggregate Dimensions

The operational reality of AI integration reveals a strong contrast to the normative assumptions of seamless technological adoption. The following sections analyse the 2nd-order themes to explain why firms are currently trapped in a state of structural and behavioural friction.

5.1.1 Defensive Personal Attitudes Towards AI

The micro-level analysis encompasses the themes of Heavy need for human oversight & control, Mistrust in AI outputs, and Learning curve hindering usage.

The literature often positions AI as an autonomous or semi-autonomous agent capable of outperforming human judgment in complex, data-heavy scenarios, theoretically leading to seamless "agentic delegation" (Baird & Maruping, 2021). However, the empirical evidence demonstrates that AI is only conditionally accepted by sustainability professionals, and strictly under intense human supervision. Decision-makers consistently view AI as a supplementary decision-support tool, explicitly refusing to allow it to replace human judgment. This is perfectly encapsulated by Orchid A's assertion that using AI is akin to "checking the homework of a student." This behaviour aligns closely with theoretical prescriptions for "meaningful human control" and "human-in-the-loop" configurations (Lebovitz et al., 2022; Raisch & Krakowski, 2021). Yet, the findings advance this concept by revealing that human supervision is not merely a subsequent verification action for AI output but as the ultimate authorisation routine in decision making process (Mintzberg, 1976). It is an absolute precondition for its adoption. Without the guarantee of manual oversight, strategic non-use is triggered. Instead of relieving the decision-maker's bounded rationality, the AI introduces a new layer of cognitive overload. Managers no longer struggle to find the data; rather, their cognitive capacity is entirely consumed by the need to continuously audit and verify the machine's inscrutable outputs. This overload triggers an immediate reliance on defensive heuristics (Atanasiu et al., 2025).

Consequently, mistrust in AI outputs serves as the default cognitive baseline for sustainability managers. Current behavioural-strategy literature frequently centres on the risks of "automation bias" and "algorithm appreciation", the danger of delegating too much moral and strategic responsibility to the machine (Lebovitz et al., 2022; Brink et al., 2024). In contrast, the informants in this study overwhelmingly exhibit "algorithm aversion" (Burton et al., 2020). This aversion is theoretically characterised by engagement followed by rejection. Managers test the system, recognise its hallucination risks and opacity, and subsequently retract their trust. This default mistrust is heavily compounded by the "black box" nature of Large Language Models (LLMs), which obscures the analytical logic from the user, generating severe epistemic

uncertainty. However, the qualitative data reveals that the primary driver of this algorithmic aversion is not necessarily a fear of the algorithm's mathematics, but deep-seated doubts cast over the quality of the input data itself, complex nuance of sustainability as a domain, a reality that links this micro-level attitude to the meso-level infrastructure of the firm.

Compounding this default mistrust is a severe learning curve that hinders operational usage, primarily manifesting as an asymmetrical skill gap in algorithmic prompting. As detailed in the findings, the cognitive burden on the sustainability manager has not been eliminated by AI but rather shifted from manual sustainability data calculation to the precise articulation of prompts and continuous output verification. This shift represents a profound disruption in established "regimes of knowing" (Pachidi et al., 2021), where traditional heuristics must now be coupled with "vintage-specific technical skills" (Choudhury et al., 2020) to correct the machine's inherent salience biases. However, empirical data reveals that this required digital literacy is highly uneven across organisations, creating a significant operational bottleneck. To prevent overwhelmed employees from engaging in "symbolic conformity", pretending to adopt the technology while covertly maintaining traditional practices (Pachidi et al., 2021), organisations must actively cultivate robust "AI interrogation practices" (Lebovitz et al., 2022). Yet, because firms currently lack comprehensive training protocols and the psychological safety required to challenge opaque outputs (Heger et al., 2025), users feel drastically underequipped. Consequently, this steep learning curve acts as a powerful catalyst for algorithm aversion; when managers lack the technical fluency to effectively prompt the system or interrogate its hidden assumptions, their epistemic uncertainty spikes, forcing them to retreat into defensive posturing and refuse strategic delegation out of a rational fear of mismanaging the firm's complex sustainability logic.

5.1.2 Socio-Technical Friction in Sustainability Strategy

At the meso-level, the intersection of decision-makers, data, and algorithmic systems yields significant friction, surrounding the themes of Sustainability data quality as a bottleneck and Sustainability nuance & complexity.

While IS literature extensively documents the computational advancements and maturity of AI models (Csaszar et al., 2024), empirical observations indicate that sustainability professionals are

far less concerned with the technical sophistication of the AI model than they are with the integrity of their own corporate data. The true bottleneck to strategic AI adoption is data architecture. Informants consistently highlighted the reality of fragmented enterprise systems and thousands of misformatted supplier records. As Tulip A noted, corporate data is often "too low to allow an AI to interpret it." This aligns with the theoretical constructs of "trapped value" (Nishant et al., 2020) and "input incompleteness" (Choudhury et al., 2020), where algorithms are rendered strategically useless by the "garbage in, garbage out" paradigm, as specified by Hydrangea C. Sustainability data is uniquely complex; it is non-linear, intertemporal, and dispersed across the entire global value chain. Therefore, the architectural reality of the firm actively suppresses the AI's capability.

Furthermore, this data friction is compounded by the inherent nuance of sustainability itself. Issues such as Scope 3 emission ambiguity, the necessity of greenwashing checks, and localised regulatory variances make sustainability decisions highly multidimensional, contested, and value-laden. AI operates on an instrumental, reductionist optimisation logic, seeking a single mathematical truth (Gao & Bansal, 2013). However, corporate sustainability demands a pluralistic logic that accommodates paradoxical goals (Hahn et al., 2014). The informants demonstrated a keen awareness that AI cannot capture the nuanced socio-ethical interactions and judgments required to navigate these trade-offs, leading to the intentional ring-fencing of the technology.

5.1.3 Organisational Risk-Averse Governance Orientation

The macro-level phenomenon encompasses AI as an efficiency tool not a strategic driver, Governance-focused implementation, Focus on safety, compliance & regulation, and Internal AI systems for control.

Despite the transformative power AI theoretically holds for strategic decision-making, the participating firms deploy it almost exclusively for administrative, incremental operational efficiency tasks. While AI accelerates the development phase of the SDMP by synthesising vast amounts of data, managers explicitly build firewalls to prevent the technology from entering the selection phase (Mintzberg, 1976). The implementation speed of AI lags significantly behind its technological evolution. Returning to the automation-augmentation paradox (Raisch &

Krakowski, 2021), these firms have not yet reached the threshold of strategic augmentation. Their usage remains confined to "weak AI" administrative support (Hutzschenreuter & Lämmermann, 2025). Even among firms attempting to build proprietary AI solutions, the technology is relegated to the operational level. No firm in the sample currently applies AI to make definitive, strategic-level sustainability decisions.

This strategic hesitation is directly tied to the inherited nature of corporate AI governance. Governance is rarely designed for the sustainability function. Rather, it is inherited from IT departments and centralised AI centres of excellence. Consequently, it carries the restrictive scope of conventional IT governance, prioritising cybersecurity, Intellectual Property (IP) protection, and regulatory compliance. Established theoretical frameworks, such as the AI Governance Cube (Schneider et al., 2023) or 4D guardrails (Gregory & Sircar, 2025), are powerful but domain-neutral. When applied uniformly across the firm, they leave sustainability governed by accident rather than intentional design. Firms operate with a reactive technical bias, addressing cybersecurity risks while failing to proactively design governance that guides the ethical and strategic purpose of the AI. The imposition of domain-neutral IT governance over sustainability decisions is a manifestation of power dynamics within the SDMP (Eisenhardt & Zbaracki, 1992). Centralised IT coalitions prioritise institutional safety and compliance, effectively neutralising the advanced predictive capabilities the sustainability team requires to navigate complex environmental paradoxes. Notably, Hydrangea mitigates this tension through a sustainability team focused on integrating AI into their domain, signalling that the firm recognises the need for governance mechanisms tailored beyond conventional IT priorities.

This risk-averse orientation manifests in the predominant procurement of external "safe" systems, such as Microsoft Copilot, chosen not for their superior analytical capabilities but for their system compatibility and localised data security. Conversely, firms that build in-house AI solutions (like Hydrangea and Tulip) do so explicitly to maintain strict control over proprietary data. By intentionally establishing "closed environments," these firms deliberately trade advanced computational speed and capability for data sovereignty and the mitigation of inscrutability (Berente et al., 2021).

This macro-level risk aversion cannot be fully understood without acknowledging the intense external regulatory environment in which these Swedish manufacturing firms operate. The European Union is currently enforcing the most rigorous ESG reporting mandates in the world, driven primarily by the Corporate Sustainability Reporting Directive (CSRD) and the incoming EU AI Act. Under the CSRD, sustainability data is no longer treated as supplementary marketing material; it is elevated to the same legal status as financial accounting, requiring mandatory third-party auditing. In this high-stakes regulatory context, the penalty for an algorithmic hallucination is exceptionally severe. If an AI system fabricates an emission factor or misinterprets a supplier's human rights data, the firm does not merely suffer a technical glitch; it commits legal and financial fraud. Consequently, the pervasive risk aversion observed among the informants is not an irrational fear of technology, but a highly rational response to existential legal threats. This regulatory pressure triggers a phenomenon best described as Institutional Conformity (DiMaggio & Powell, 1983). When faced with high environmental uncertainty and strict regulations, organisations stop attempting to innovate and instead mimic the behaviours of their peers to appear legitimate and safe.

5.2 Analysis of Aggregate Dimensions

The empirical data, when distilled through the Gioia methodology, culminates in three aggregate dimensions that capture the multidimensional reality of AI integration within corporate sustainability. To understand how firms navigate the introduction of algorithmic technologies, these dimensions must first be defined as distinct theoretical constructs, representing the micro, meso, and macro levels of the organisation, before being set into motion to reveal the dynamic model that ultimately governs firm behaviour.

At the micro-level, Defensive Personal Attitudes Toward AI emerges as the psychological and behavioural baseline of the sustainability decision-maker. This construct is characterised by a profound, default state of epistemic uncertainty and "algorithm aversion" (Burton et al., 2020). It moves beyond a mere lack of technical skills, manifesting instead as a deliberate, protective posture. When confronted with the "black box" opacity of LLMs, human agents instinctively recognise a fundamental mismatch between the machine's reductionist, instrumental logic and

the pluralistic, value-laden logic required for sustainability (Gao & Bansal, 2013). Consequently, managers engage in "defensive conjoined agency" (Murray et al., 2021). They refuse to act as passive operators or succumb to "automation bias"; instead, they actively establish themselves as interpretive buffers, maintaining extreme oversight to ensure that algorithmic hallucinations or missing contextual nuances do not result in the abdication of moral and strategic responsibility (Lebovitz et al., 2022).

At the meso-level, Socio-Technical Friction in Sustainability Strategy defines the structural and material clash between advanced algorithmic capabilities and the immature reality of corporate data architecture. Drawing upon STS theory, this construct encapsulates the "capability void" (Mikalef & Gupta, 2021) that paralyses AI implementation. The friction is deeply sociomaterial. It is the fundamental incompatibility between the fluid, highly qualitative, and localised nuance of sustainability metrics (such as Scope 3 emissions, human rights supply chain audits, and dynamic regulatory variance) and the rigid, structured data ingestion requirements of AI models. This incompatibility creates a pervasive state of "input incompleteness" (Choudhury et al., 2020) and "trapped value" (Nishant et al., 2020). The constraint is not the computational power of the algorithm, but the fragmented, siloed, and inherently unstable nature of the sustainability data itself, rendering the technological asset strategically inert.

At the macro-level, Organisational Risk-Averse Governance Orientation encapsulates the institutional and structural control mechanisms that dictate how AI is deployed across the firm. This construct represents the manifestation of the "pacing problem" (Berente et al., 2021). Executives exert top-down pressure to adopt AI to signal market competitiveness, yet the actual governance frameworks provided are inherited, reactive, and entirely domain-neutral. Because these governance structures are typically designed and enforced by central IT departments, they are dominated by traditional security imperatives: cybersecurity, IP protection, and localised regulatory compliance. This results in a rigid, compliance-heavy architecture that entirely fails to address the unique socio-ethical demands of sustainability. It is a governance model focused exclusively on risk containment and IT safety rather than purpose-led, sustainable value creation, effectively stripping the AI of its strategic autonomy before it reaches the end-user.

While defining these constructs isolates the specific friction points, the theoretical contribution of this research lies in how these dimensions interact. They do not exist in isolation; rather, they form a highly restrictive, reinforcing loop that traps the organisation. The model operates sequentially: the fragmented, unstructured reality of corporate sustainability data (meso-level friction) inevitably generates inaccurate, hallucinated, or contextually shallow algorithmic outputs. When the sustainability manager is exposed to these flawed outputs, it immediately triggers and validates their baseline algorithm aversion and epistemic mistrust (micro-level defense). This causal link is explicitly recognised by the informants; as Peony A highlighted, the absence of a “structured data backbone” directly inhibits algorithmic trust, forcing managers to “sit with the problem” and resolve it independently rather than relying on the tool.

Fearing the moral and professional repercussions of relying on a flawed system, these decision-makers refuse to advocate for deeper algorithmic integration. Instead, they demand blanket human oversight and deliberately restrict the AI to low-risk, administrative tasks. Crucially, this lack of user pushback for advanced AI capabilities aligns perfectly with the central IT department’s default preference for restrictive, highly controlled software environments. The firm subsequently settles into a state of blanket, risk-averse governance (macro-level control). Because the governance structure successfully restricts the AI to low-level synthesis and drafting, the firm extracts very little strategic return on investment from the technology. Consequently, without a proven strategic business case, executive leadership refuses to allocate the massive capital expenditure required to overhaul and fix the foundational data architecture. This financial constraint acts as the final mechanism of the trap, with Hydrangea A confirming that establishing the necessary data infrastructure ultimately “comes down to IT budget”, which acts as a harder limit on the firm than a lack of strategic ideas. The foundational data remains broken, and the cycle begins anew. The friction persists continuously, fed by its own structural inertia.

This reinforcing loop culminates in a phenomenon that this study defines as the Defensive Equilibrium. The most critical insight of the Defensive Equilibrium is that every individual actor within the loop is behaving in a highly logical, locally rational manner. The data expert is entirely rational to lock down the AI systems to prevent cybersecurity breaches; the

sustainability manager is entirely rational to mistrust and buffer an algorithm that consistently hallucinates on sensitive Sustainability data; and the executive board is entirely rational to withhold infrastructure investment when the technology is only being used as a simple text-generation functions. However, while locally rational, the firm becomes collectively locked into a state of severe, systemic under-use. This establishes a profound state of path dependency, which stands as the central theoretical claim of this study. Early organisational decisions to treat AI adoption as a generic IT software procurement exercise, focusing on safe, off-the-shelf tools like Microsoft Copilot, rather than a foundational transformation of sustainability strategy, establish a rigid operational trajectory. Once the Defensive Equilibrium is formed, it becomes self-sustaining and increasingly difficult to dismantle. The organisation becomes trapped in a capability void, fundamentally unable to evolve from basic administrative automation to genuine strategic augmentation (Raisch & Krakowski, 2021). The path dependency ensures that as long as the underlying data architecture remains unfixed and governance remains domain-neutral, the behavioural response will permanently default to algorithmic aversion. The technology is neutralised by the very organisational structures deployed to support it.

This dynamic provides a definitive and explicit answer to the primary research question guiding this thesis: How do decision-makers navigate the tension between algorithmic efficiency and human control within sustainability? The empirical evidence and subsequent theoretical model prove that decision-makers do not truly "balance" these competing forces; rather, they navigate the tension by definitively settling it toward control. When faced with the epistemic uncertainty of the algorithmic black box and the immense ethical complexity of the economical, social, and environmental aspects, managers engage in strategic non-use. They willingly and deliberately sacrifice the much-touted algorithmic efficiency on high-stakes, complex sustainability decisions to maintain absolute human-centric governance. In doing so, they protect the firm from critical reputational, social, and regulatory risks, ensuring that the profound moral responsibility required for sustainable value creation is never abdicated to a machine. Strategic non-use is an illustration of a mental shortcut that allows the manager to immediately eliminate the massive risks of algorithmic bias, prioritising ethical safety over computational optimisation

5.3. Cross-Case patterns and Internal Contradictions

The defensive equilibrium holds across all five participating firms, providing a robust replication logic (Creswell & Creswell, 2023) despite variations in firm size and sector. However, the specific mechanisms of control diverge meaningfully based on industry context. Larger, heavily regulated industrial firms (Hydrangea, Peony, Tulip) treat AI much like physical industrial machinery; they attempt to internalise it, building closed, in-house data fortresses to maintain absolute control over the input-output pipeline. Thus introducing a structured approach and the SDMP to the implementation of AI (Shepherd et. al., 2024; Elbanna, 2006). In contrast, Orchid relies heavily on off-the-shelf, external tools like Copilot, which correlates with their demonstrably higher expressions of mistrust and fear of hallucination.

A critical internal contradiction emerges in the observation of sensemaking and cognitive bias. While participants express profound mistrust of the system regarding data quality, only a couple informants explicitly acknowledged the existence of algorithmic bias. The majority of informants focus their defensive efforts entirely on factual accuracy (hallucinations) while remaining dangerously blind to the systemic, historical biases that "black box" models inevitably reproduce (Zhao & Gómez Fariñas, 2022).

Furthermore, a significant strategic contradiction exists between executive intent and operational reality. There is intense top-management pressure to "use AI for competitiveness," which actively collides with the users' mistrust-by-default. Finally, the most profound paradox identified across the cases is the complete absence of governance regarding AI's own environmental footprint. Sustainability functions are actively deploying highly carbon-intensive "Red AI" technologies to streamline reporting, yet not a single governance instrument observed addresses the massive physical materiality and energy consumption of these models (Tamburrini, 2025).

5.4 Revisiting the Paradox: the Literature-Practice Gap

The empirical data exposes a severe literature-practice gap. The academic literature has built a highly sophisticated theoretical apparatus for mature AI integration, discussing agentic systems, profound automation bias, and post-human strategic delegation. However, the vast majority of firms remain in the early-to-mid adoption phases, facing an entirely different cluster of foundational problems centred on data integrity, intentional non-use, and domain mismatch.

5.4.1 The Transfer of Cognitive Overload and Defensive Heuristics

The foundational promise of Business Intelligence and Analytics (BI&A) in the SDMP is its ability to relax the constraints of bounded rationality. The literature assumes that by processing millions of alternatives at high speeds, AI alleviates the “information overload” that typically forces managers into flawed, fast-and-frugal heuristics (Csaszar et al., 2024; Phillips-Wren & Adya, 2020). The empirical findings directly contradict this assumption. In the context of corporate sustainability, AI does not eliminate cognitive overload; it merely transfers it. Managers are no longer burdened by the effort of finding the data, but their cognitive capacity is now entirely consumed by the necessity to constantly audit, verify, and prompt the machine to prevent hallucinations. This transference creates a new form of epistemic exhaustion. Consequently, rather than achieving the “procedural rationality” predicted by the literature (Dean & Sharfman, 1996), managers react to this new overload by deploying “strategic non-use” as a rational heuristic (Bingham & Eisenhardt, 2011). By deliberately refusing to use AI for high-stakes human rights or compliance decisions, managers utilise a mental shortcut to instantly eliminate the risk of algorithmic bias, prioritising safety over computational optimisation.

5.4.2 Agentic AI is Rare in Sustainability Practice

Prominent IS and management scholars (Baird & Maruping, 2021; Berente et al., 2021; Murray et al., 2021) theorise the rise of agentic AI capable of autonomous function, leading to complex forms of conjoined agency where the act of decision-making shifts away from the human. The reality presented by the interviewed firms challenges this trajectory. In practice, AI is fiercely maintained as a subordinate assisting tool, leveraged almost exclusively for analytical synthesis heavily reliant on human prompting. While there is executive rhetoric regarding the future of agentic AI, actual projects remain perpetually "under development." The literature often

discusses AI in sustainability as a futuristic paradigm of autonomous climate optimisation, while in reality, firms are struggling with the foundational prerequisites: basic AI technical literacy, user prompting skills, unstructured data silos, and primitive risk management.

5.4.3 Algorithmic Appreciation and Aversion versus Non-Use

A substantial strand of the literature (Lebovitz et al., 2022; Brink et al., 2024; Raisch & Krakowski, 2021) frames the central behavioural risk of Industry 5.0 as managers over-trusting AI. The fear is that "automation bias" and "algorithm appreciation" will lead to an abdication of moral responsibility. Conversely, our empirical observations point to a deeply entrenched behavioural pattern of systematic non-use. AI is deliberately excluded from compliance reporting, social sustainability, and high-stakes strategic choices. Mistrust in the system is not an anomaly; it is the default posture. This strategic under-reliance undermines the theoretical potential of AI to resolve paradoxical sustainability decisions (Csaszar et al., 2024). Algorithm aversion (Burton et al., 2020) is not merely a transient cognitive bias to be trained away; it is the deliberate governance equilibrium that risk-averse firms have chosen to settle on to protect their legitimacy.

5.4.4 Neutrality of AI Governance and the Need for Multidimensional Governance

Scholars such as Schneider et al. (2023), Gregory & Sircar (2025), and Batool et al. (2025) have proposed comprehensive AI governance frameworks covering data, models, systems, and 4D guardrails. However, as these frameworks are applied firm-wide, their domain-neutrality becomes a fatal flaw when intersecting with sustainability. In practice, AI governance is established by IT departments, focusing on traditional risk vectors like IP protection and data leakage. Consequently, sustainability AI governance operates in a passive position, inheriting an IT framework by accident rather than intentional, purpose-led design.

This reactive approach exacerbates the inherent tensions of sustainability. Sustainability requires a pluralistic and integrative logic, shifting focus from conventional short-term financial priorities to long-term ecological resilience and multi-stakeholder value (Gao & Bansal, 2013). Sustainability extends beyond the firm's borders, covering the entire upstream and downstream value chain. Therefore, governance in this domain cannot simply be about IT security; it must encompass long-term survival, legitimacy, and environmental ethics. When AI is introduced

under a strictly technical, risk-based governance model, the firm falls short on the capabilities required to resolve the ethical requirements of sustainability. This mismatch leads directly to the observed outcomes: sensitive data is withheld from the AI, risk-based compliance supersedes purpose-led innovation, and AI is ultimately governed as a software liability rather than a strategic sustainability asset. Effective governance requires firms to leverage both traditional "horizontal" IT governance and "vertical," domain-specific sustainability governance.

5.4.5 Data Governance should be Considered Together with AI Governance in Sustainability

Historically, data governance and AI governance have been treated as separate, sequential streams. Data governance concerned itself with storage and architecture, while AI governance focused on algorithmic opacity and output biases (Anthony, 2021; Martin, 2019). The empirical observations strongly highlight that at the current stage of AI adoption, participants are acutely concerned with data quality over algorithmic logic. Due to the extreme multidimensional nature of sustainability metrics, poor data integrity instantly poisons the AI system. Consequently, in the context of sustainability, data governance and AI governance cannot be decoupled. Firms must ensure that foundational data collection, validation, and ownership are completely integrated into the AI oversight approach, as this combined infrastructure forms the absolute prerequisite for future AI maturity.

5.5 Re-Interpreting the SDMP: AI-Sustainability Governance and Human Agency

The integration of the empirical findings into a final theoretical understanding necessitates a recursive re-examination of the SDMP. By mapping the identified aggregate dimensions: Organisational Risk-Averse Governance Orientation (macro), Socio-Technical Friction in Sustainability Strategy (meso), and Defensive Personal Attitudes Toward AI (micro), onto the seminal SDMP model proposed by Mintzberg et al. (1976), this study illustrates exactly how sustainability managers navigate the "automation-augmentation paradox" (Raisch & Krakowski, 2021).

The macro-level dimension establishes the broader organisational context through a horizontal alignment between IT governance and sustainability-specific mandates (Papagiannidis et al.,

2023). Within Mintzberg's Identification Phase, this governance orientation dictates the overarching "rules of engagement", ensuring that AI adoption is not an isolated technological pursuit but is bound by corporate compliance (Gregory & Sircar, 2025). This macro-filter alters the SDMP by institutionalising strategic guardrails that strictly define which sustainability problems the AI is legally permitted to diagnose.

The meso-level dimension, characterised by socio-technical friction, primarily impinges upon the Development Phase of the SDMP. As Mintzberg et al. (1976) note, this phase involves "search" and "design" routines to synthesise potential solutions. Empirical evidence suggests that poor corporate data quality acts as a critical bottleneck here, necessitating robust data governance and localised data ownership (Aljohani, 2025; Schwaeye et al., 2025). Managers navigate this friction by utilising AI for pattern recognition and scenario analysis while simultaneously countering the "inscrutability" of algorithmic logic through internal data centers that ensure input integrity (Berente et al., 2021; Zhao & Gómez Fariñas, 2022).

Finally, the micro-level dimension of defensive personal attitudes manifests most prominently in the selection phase of the SDMP. Managers exhibit "algorithm aversion" due to the perceived lack of traceability and the prevalence of hallucinations (Burton et al., 2020; Gregory & Sircar, 2025). Consequently, the manager acts as a human buffer in the evaluation and choice routines, preventing AI from acting independently (Brink et al., 2024; Shrestha et al., 2019). This "human-in-the-loop" configuration ensures that while AI may augment the identification and development phases, the ultimate selection remains a product of professional judgment and "integrative logic" (Slawinski & Bansal, 2015). In summary, as illustrated in Figure 1, the research re-interprets the SDMP as a hybrid process where AI-driven efficiency is consistently moderated by phase-gated governance filters, ensuring that the delegation of analytical tasks to algorithms never equates to an abdication of moral responsibility.

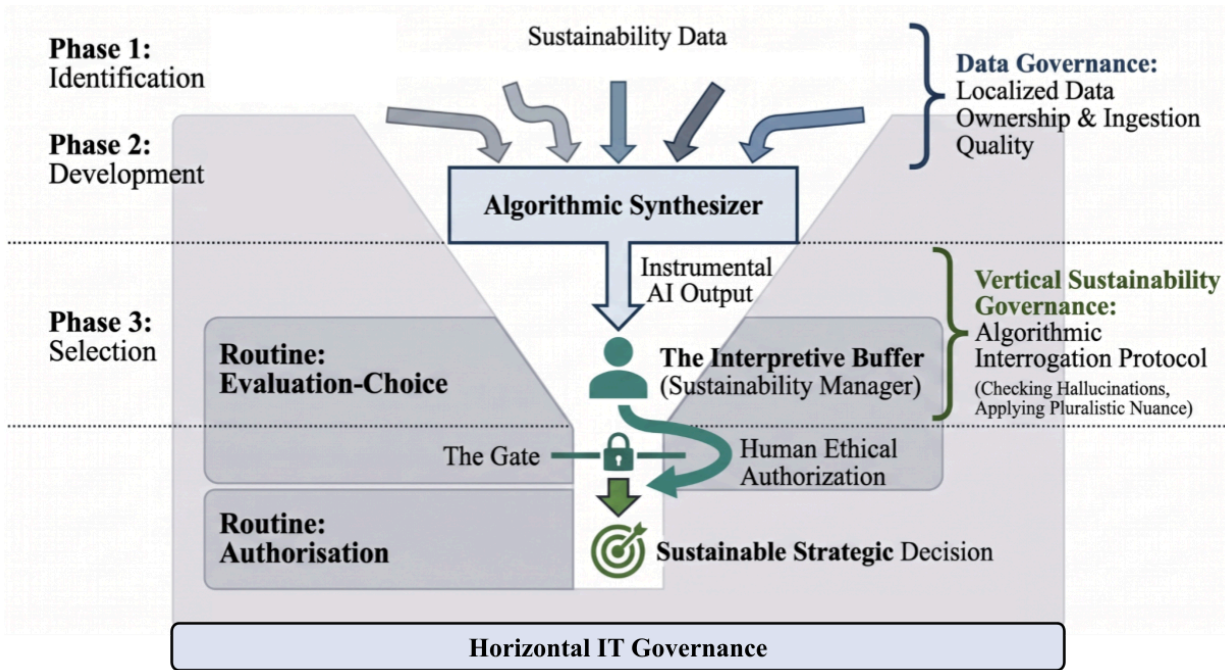


Figure 1: Human-Centric AI Governance for Sustainability Decision-Making

5.5.1 Explanation of Figure 1

The foundational architecture of the approach relies on a Cross-Layered Matrix, merging domain-neutral “Horizontal” IT governance (e.g., cybersecurity, GDPR) with a bespoke “Vertical” governance layer dedicated to the pluralistic ethics of sustainability. Rather than governing the technology as a static software tool, this matrix is applied dynamically across a Phase-Gated Decision Pipeline:

Gate 1: The Development Phase (Algorithmic Augmentation)

In the secondary stages of the SDMP, decision-makers face severe information overload when executing "search routines" across heterogeneous sustainability data. At this gate, the approach encourages maximum algorithmic efficiency to overcome meso-level socio-technical friction. AI is fully authorised to synthesise supply-chain metrics and generate predictive scenarios. This phase is heavily governed by the Data-to-Decision Pipeline, which mandates that localised "Data Owners" certify the integrity of the inputs before algorithmic ingestion, neutralising the "garbage in, garbage out" risk at the source.

Gate 2: The Selection phase - Evaluation-Choice Routine (The Interpretive Buffer)

As the SDMP enters the Selection Phase and transitions toward evaluating strategic alternatives, the approach enforces a cognitive checkpoint to address micro-level algorithm aversion. Because AI operates on a singular, instrumental logic, it cannot autonomously evaluate pluralistic, intertemporal sustainability trade-offs (Gao & Bansal, 2013). Therefore, this gate institutionalises the human manager as an Interpretive Buffer. The approach mandates an Algorithmic Interrogation Protocol, requiring managers to utilise targeted digital literacy to actively challenge hidden assumptions, catch hallucinations, and translate instrumental outputs into integrative strategy. Here, “defensive conjoined agency” (Murray et al., 2021) is formally codified as a governance mechanism to prevent automation bias.

Gate 3: The Selection phase - The Authorisation Routine (Retention of Human Agency)

The pinnacle of the SDMP dictates the final authorisation of the strategic choice. To permanently resolve the epistemic uncertainty and legal risks associated with algorithmic opacity, this final gate explicitly bars AI from participation. It establishes a hard boundary: autonomous agentic delegation is strictly restricted for high-stakes environmental and social decisions. The ultimate locus of moral, ethical, and strategic accountability remains unequivocally human.

By structuring governance as a phase-gated matrix, the firm achieves a secure AI environment that safely accelerates the development of strategy, while legally and ethically guaranteeing that the selection of that strategy aligns with the nuanced demands of sustainability.

5.5.2 Conditions and Limitations

While the Human-Centric AI Governance approach provides a robust blueprint for dismantling the Defensive Equilibrium, an elite theoretical model must explicitly acknowledge its boundary conditions, the specific structural and temporal scenarios where the approach may require significant adaptation or risk breaking down entirely.

Sustainability Data Inflow

A critical failure of generic AI governance is the assumption that sustainability data is a static asset ingested at a single point in time. In reality, data flows continuously, but its function shifts violently across the SDMP (Mintzberg et al., 1976). In the Identification Phase, data acts as a

stimulus trigger; in the Development Phase, it acts as an immense volume of unstructured raw material; and in the Selection Phase, it must manifest as a synthesised, ethically weighted alternative. Moreover, the proposed model acknowledges this dynamic flow. It recognises that the extreme “socio-technical friction” observed in the participating firms occurs because massive, unstructured data flows are misaligned with algorithmic capabilities during the Development Phase. Consequently, this approach states that data governance should be coupled from AI governance. By enforcing localised “Data Owners” at Gate 1 (Development), the approach actively manages the data flow before it reaches the AI, thereby ensuring that the outputs flowing into Gate 2 (Evaluation-Choice) are clear of the hallucinations that trigger managerial algorithm aversion.

The Scale and Resource Boundary

The primary boundary condition of this approach is organisational scale and capital resourcing. The model relies on a sophisticated "Cross-Layered Matrix," which inherently demands that a firm possesses the financial and human capital to maintain dual governance structures, a central IT governance team (Horizontal) and a dedicated sustainability ethics team (Vertical). Furthermore, the mandate for localised "Data Owners" requires a vast, formalised workforce. While the massive multinational corporations in this study, such as Hydrangea and Peony, possess the institutional bandwidth to construct and maintain this architecture, Small and Medium Enterprises (SMEs) do not. For SMEs operating within the manufacturing supply chain, attempting to build a bespoke vertical governance layer for sustainability AI could result in severe administrative bloat and resource exhaustion. To adapt this approach, resource-constrained firms must rely on externalised vertical governance. Instead of building an internal ethics board, SMEs must lean on industry consortia, shared supply-chain data pools, and open-source ESG benchmarking approaches to provide the necessary ethical guardrails, maintaining only the human-interrogation protocol internally.

The Temporal and Autonomy Boundary (The Rise of Agentic AI)

The second critical boundary condition relates to the technological velocity of AI itself. This approach is grounded in the current reality of Industry 5.0, where AI operates primarily as a sophisticated synthesiser that requires active, manual "prompting" from a human user. Consequently, the approach's protective mechanism relies heavily on the "Interpretive Buffer",

the human manager actively catching hallucinations between the prompt and the final strategic decision. However, as the technology inevitably evolves toward truly autonomous "Agentic AI", systems capable of not just answering prompts, but independently formulating goals, executing multi-step environmental strategies, and automatically adjusting supply-chain logistics without manual triggers (Baird & Maruping, 2021), the "human-in-the-loop" mechanism will be severely strained. If AI acts autonomously at speeds beyond human cognitive tracking, the active interpretive buffer becomes an impossible bottleneck. Under the condition of fully Agentic AI, the approach must transition from a protocol of active prompting to a protocol of passive algorithmic auditing. The human role will no longer involve checking the "homework" of the AI prior to execution; instead, human-centric governance will require the construction of automated "circuit breakers." These circuit breakers would allow the AI to act autonomously within predefined, mathematically hard-coded sustainability thresholds, instantly halting the agent's autonomy the moment it attempts to execute a decision that violates human rights or exceeds carbon-emission boundaries. Thus, while the core philosophy of human accountability remains permanent, the mechanical execution of the approach must shift as the technology achieves higher degrees of autonomous agency.

The Decision-Velocity Boundary

This approach is fundamentally designed for strategic decision-making, which is inherently infrequent, unstructured, and long-term (Elbanna, 2006). Because strategic decisions (e.g., shifting to a circular supply chain) afford decision-makers the time to pause and act as an "Interpretive Buffer," the approach thrives. However, if this governance model is misapplied to operational sustainability decisions that require high-velocity, real-time choices (e.g., minute-by-minute autonomous energy grid optimisation or real-time factory emission throttling), the mandatory human-in-the-loop checkpoint will cause catastrophic systemic lag. For high-velocity operational decisions, the approach's human-centric authorisation gate should be removed and replaced with mathematically hard-coded "circuit breakers" that allow the AI to act autonomously within pre-approved environmental thresholds.

6.0 Conclusion

The transition toward Industry 5.0 explicitly demands the harmonisation of advanced technological systems with human-centric sustainability objectives. As corporate sustainability evolves from a retrospective compliance exercise into a core strategic imperative, decision-makers are increasingly burdened by the immense complexity of sustainability. AI has been theoretically positioned as a critical cognitive remedy to navigate this complexity, promising to relieve information overload and optimise sustainable value creation. However, this thesis identified a profound empirical gap: while macro-level technical capabilities are well-documented, the micro-foundational reality of how human managers actually interact with these opaque algorithms when making high-stakes, ethically charged sustainability decisions remained underexplored. To address this gap, this study posed the central research question: How do decision-makers navigate the tension between algorithmic efficiency and human control within sustainability?

Through a qualitative, multi-case study of Swedish manufacturing firms, this research reveals that decision-makers navigate this tension not through seamless human-machine augmentation, but through deliberate, defensive control. When faced with the epistemic uncertainty of the AI traceability and the pluralistic demands of corporate sustainability, managers intentionally prioritise human-centric governance over algorithmic efficiency. They engage in strategic non-use, actively relegating AI to low-risk, administrative synthesis and explicitly blocking the technology from influencing highly sensitive social, ethical, and regulatory decisions. The application of the Gioia methodology distilled this reality into three interacting theoretical dimensions. At the meso-level, firms are trapped in severe Socio-Technical Friction. The primary bottleneck to AI adoption is not the computational limit of the algorithm, but the fragmented, unstructured, and unstable nature of corporate sustainability data. This capability void ensures that AI outputs are frequently flawed. At the micro-level, these flawed outputs trigger Defensive Personal Attitudes. Employees, lacking the targeted digital literacy to confidently prompt or interrogate the machine, default to profound algorithm aversion. Finally, at the macro-level, this hesitation is compounded by an Organisational Risk-Averse Governance Orientation. Executive

mandates to adopt AI are met with inherited, domain-neutral IT governance that focuses strictly on cybersecurity, failing to provide the ethical guardrails required for sustainability.

The primary theoretical contribution of this thesis is the conceptualisation of the Defensive Equilibrium. The findings challenge the prevailing academic narrative that the primary risk of AI integration is "automation bias" or the rapid rise of "agentic" algorithmic delegation. Instead, this study proves that for firms in the early-to-mid phases of AI adoption, the dominant paradigm is path-dependent stagnation. Because the data is poor, managers mistrust the system; because they mistrust the system, they demand restrictive governance; and because the governance restricts AI to basic tasks, the firm extracts no strategic value, resulting in a refusal to invest in the foundational data architecture. The firm becomes collectively locked into a state of algorithmic under-use, despite every actor behaving in a locally rational manner. Practically, this research demonstrates that generic corporate governance is insufficient for the integration of AI into sustainability. To dismantle the Defensive Equilibrium, this study suggested a Human-Centric AI Governance approach to SDMP. This approach asserts that firms should construct a cross-layered matrix, fusing horizontal IT security with vertical, domain-specific sustainability ethics. Furthermore, it highlights that data governance and AI governance can no longer be treated as sequential silos; they should be fully integrated. Most importantly, the model shifts the managerial imperative away from passive operation toward active "algorithmic interrogation," demanding that firms train their sustainability professionals to act as critical interpretive buffers who can safely navigate the machine's hallucinations and hidden biases. Finally, this approach insists that true sustainability governance needs to account for the physical materiality of the technology itself, actively managing the environmental footprint of highly carbon-intensive "Red AI" models.

6.1 Practical Implications

The findings and figure 1 yield several urgent, actionable guidelines for corporate executives and Chief Sustainability Officers (CSOs) navigating the Industry 5.0 transition. First, top management and CDOs should consider resolving the underlying data architecture as soon as possible for effective AI implementation. Pushing for advanced, agentic AI deployment in an

environment with fragmented, unstructured Sustainability data directly collides with managerial mistrust, leading only to algorithm aversion and "symbolic conformity" (Pachidi et al., 2021). Executives must re-align their expectations, viewing current AI as a high-powered text and data synthesiser, and redirect capital investment toward foundational data governance and system integration. Second, firms must dismantle the silos between the IT department and the sustainability function. Relying solely on an IT-dominated "AI Excellence Team" guarantees a governance system that is blind to ethical and ecological nuance. CSOs should secure a permanent, proactive seat in AI-governance design, establishing joint task forces that co-create the domain-specific "Vertical" guardrails necessary for pluralistic decision-making. Finally, firms should actively address the prompting and interrogation skill gap. Organisations should invest heavily in AI literacy programs specifically tailored for sustainability personnel, moving beyond basic software usage to teach critical interrogation of algorithmic outputs.

6.2 Limitations and Future Research

While this study provides deep, micro-foundational insights into algorithmic governance, it is not without limitations. The idiographic nature of the qualitative design, strictly delimited to the Swedish manufacturing sector, means that these findings are inherently contextual. Furthermore, while the use of the CIT and triangulation mitigated retrospective bias, the cross-sectional nature of the study captures only a specific moment in the rapid evolution of AI technology.

These limitations present clear spaces for future research. Scholars should pursue longitudinal studies to observe whether the "Defensive Equilibrium" and its observed self-reinforcement naturally dissolves as corporate sustainability data architectures mature, or if algorithm aversion remains a permanent fixture of human-AI interaction in ethically charged domains. Additionally, quantitative research is required to statistically test the relationship between a firm's data maturity and its tendency to delegate strategic tasks to AI across different global regulatory environments. Similarly, to capture the shifting dynamics of AI integration, future studies should employ quasi-longitudinal or comparative case designs to examine firms actively transitioning between stages of AI maturity. This would render the gradual evolution of managerial trust, the

resolution of socio-technical friction, and the dissolution of defensive governance highly visible to researchers.. Finally, further research should be conducted into the ecological paradox of AI, quantifying the trade-off between the carbon emissions saved by AI-optimised supply chains versus the massive energy consumed by the data centres powering those algorithms.

Ultimately, this thesis concludes that Artificial Intelligence will not autonomously solve the grand challenges of corporate sustainability. AI is a highly capable but inherently flawed instrumental tool. Its strategic value is entirely contingent upon the organisation's willingness to build pristine data architectures and human-centric, purpose-led governance systems. In the pursuit of sustainable value creation, the final locus of moral, ethical, and strategic accountability should, and will, remain profoundly human.

7.0 References

- Ahmad, M. & Wilkins, S. (2025). Purposive sampling in qualitative research: a framework for the entire journey, *Quality & Quantity*, vol. 59, pp. 1461–1479, <https://doi.org/10.1007/s11135-024-02022-5>.
- Aljohani, A. (2025). A decision-support framework for evaluating AI-enabled ESG strategies in the context of sustainable manufacturing systems, *Scientific Reports*, vol. 15, no. 1, p.23864.
- Alibašić, H. (2022). Strategic Resilience and Sustainability Planning: Management Strategies for Sustainable and Climate-Resilient Communities and Organizations, 2nd edn, Cham: *Springer Nature*.
- Anthony, C. (2021). When Knowledge Work and Analytical Technologies Collide: The Practices and Consequences of Black Boxing Algorithmic Technologies, *Administrative Science Quarterly*, vol. 66, no. 4, pp. 1173-1212.
- Antonicic, M. (2020). Incorporating sustainability into corporate governance and strategic decision-making using big data and artificial intelligence, *Journal of Risk Management in Financial Institutions*, vol. 13, no. 4, pp.290-294
- Appelbaum, S. H. (1997). Socio-technical systems theory: An intervention strategy for organizational development. *Management Decision*, 35(6), 452–463. <https://doi.org/10.1108/00251749710173823>
- Atanasiu, R. (2021). The lifecycle of heuristics as managerial proverbs. *Management Decision*, 59(7), 1617–1641. <https://doi.org/10.1108/MD-08-2019-1025>
- Atanasiu, R., Wickert, C., & Khapova, S. N. (2025). Towards a heuristic view of managerial heuristics: Integrating divergent perspectives. *International Journal of Management Reviews*, 27, 58–80. <https://doi.org/10.1111/ijmr.12382>
- Baird, A., & Maruping, L. M. (2021). The next generation of research on IS use: A theoretical framework of delegation to and from agentic IS artifacts. *MIS Quarterly*, 45(1), 315–341.

- Batool, A., Zowghi, D., & Bano, M. (2025). AI governance: a systematic literature review, *AI and Ethics*, vol. 5, pp. 3265-3279.
- Berente, N., Gu, B., Recker, J., & Santhanam, R. (2021). Managing artificial intelligence. *MIS Quarterly*, 45(3), 1433–1450.
- Bevilacqua, S., Masárová, J., Perotti, F. A. & Ferraris, A. (2025). Enhancing top managers' leadership with artificial intelligence: insights from a systematic literature review. *Review of Managerial Science*, vol. 19, no. 9, pp. 2899–2935.
- Bingham, C. B., & Eisenhardt, K. M. (2011). Rational heuristics: The ‘simple rules’ that strategists learn from process experience. *Strategic Management Journal*, 32(13), 1437–1464. <https://doi.org/10.1002/smj.965>
- Brink, A., Benyayer, L.-D., & Kupp, M. (2024). Decision-making in organizations: Should managers use AI? *Journal of Business Strategy*, 45(4), 267–274.
- Butt, T. A., Iqbal, M., & Arshad, N. (2025). From Policy to Pipeline: A Governance Framework for AI Development and Operations Pipelines, *IEEE Access*, vol. 14, pp. 1-15. *(Note: Volume and page numbers estimated based on repository metadata format for standard IEEE continuous publication).*
- Butterfield, L. D., Borgen, W. A., Amundson, N. E., & Maglio, A. T. (2005). Fifty years of the critical incident technique: 1954–2004 and beyond. *Qualitative Research*, 5(4), 475–497. <https://doi.org/10.1177/1468794105056924>
- Chen, H., Chiang, R. H. L., & Storey, V. C. (2012). Business Intelligence and Analytics: From Big Data to Big Impact, *MIS Quarterly*, vol. 36, no. 4, pp.1165-1188.
- Choudhury, P., Starr, E., & Agarwal, R. (2020). Machine learning and human capital complementarities: Experimental evidence on bias mitigation. *Strategic Management Journal*, 41(8), 1381–1411.

Collins, C., Kumar, A., Leiendecker, J., Meyer zum Felde, A., & Pagano, E. (2024). Sustainability Data Is a Big Opportunity in Information Services [pdf], Boston Consulting Group.

Corley, K., Bansal, P., & Yu, H. (2021). An editorial perspective on judging the quality of inductive research when the methodological straightjacket is loosened, *Strategic Organization*, vol. 19, no. 1, pp. 161–175, <https://doi.org/10.1177/1476127020968180>

Creswell, J. W. & Creswell, J. D. (2023). *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*, 6th edn, Thousand Oaks, CA: SAGE.

Csaszar, F. A., Ketkar, H., & Kim, H. (2024). Artificial intelligence and strategic decision-making: Evidence from entrepreneurs and investors. *Strategy Science*, 9(4), 322–345.

da Silva, R. F., & Razzolini Filho, E. (2020). Sustainability in the Decision Making Process: A Systematic Review of Literature, In: *International Conference on Engineering and Technology Management (ICETM)*

Dean, J. W., & Sharfman, M. P. (1996). Does decision process matter? A study of strategic decision making effectiveness. *Academy of Management Journal*, 39(2), 368–396. <https://doi.org/10.2307/256784>

De Sordi, J. O. (2024). *Qualitative Research Methods In Business: Techniques for Data Collection and Analysis*, Springer Nature Switzerland AG.

DiMaggio, P. & Powell, W. (1983). The iron cage revisited: institutional isomorphism and collective rationality in organizational fields, *American Sociological Review*, vol. 48, no. 2, pp. 147–160

Eisenhardt, K. M. (1989). Making fast strategic decisions in high-velocity environments. *Academy of Management Journal*, 32(3), 543–576. <https://doi.org/10.2307/256434>

Eisenhardt, K. M., & Zbaracki, M. J. (1992). Strategic decision making. *Strategic Management Journal*, 13(SI), 17–37. <https://doi.org/10.1002/smj.4250130904>

- Elbanna, S. (2006). Strategic decision-making: Process perspectives. *International Journal of Management Reviews*, 8(1), 1–20. <https://doi.org/10.1111/j.1468-2370.2006.00118.x>
- Elbanna, S., Thanos, I. C., & Jansen, R. J. G. (2020). A literature review of the strategic decision-making context. *M@n@gement*, 23(2), 1–29. <https://doi.org/10.37062/mana.23.2.1>
- Felin, T., Foss, N. J., & Ployhart, R. E. (2015). The Microfoundations Movement in Strategy and Organization Theory, *Academy of Management Annals*, vol. 9, no. 1, pp. 575-632.
- Ferraro, F., Etzion, D., & Gehman, J. (2015). Tackling Grand Challenges Pragmatically: Robust Action Revisited, *Organization Studies*, vol. 36, no. 3, pp. 363-390.
- Flanagan, J. C. (1954). The Critical Incident Technique, *Psychological Bulletin*, vol. 51, no. 4, pp. 327–358
- Fredrickson, J. W., & Mitchell, T. R. (1984). Strategic decision processes: Comprehensiveness and performance in an industry with an unstable environment. *Academy of Management Journal*, 27(2), 399–423. <https://doi.org/10.2307/255932>
- Gao, J., & Bansal, P. (2013). Instrumental and Integrative Logics in Business Sustainability, *Journal of Business Ethics*, vol. 112, no. 2, pp.241-255
- George, G., Merrill, R. K., & Schillebeeckx, S. J. D. (2021). Digital Sustainability and Entrepreneurship: How Digital Innovations Are Helping Tackle Climate Change and Sustainable Development, *Entrepreneurship Theory and Practice*, vol. 45, no. 5, pp. 999-1027.
- Gioia, D. A., Corley, K. G., & Hamilton, A. L. (2013). Seeking Qualitative Rigor in Inductive Research: Notes on the Gioia Methodology, *Organizational Research Methods*, vol. 16, no. 1, pp. 15–31, <https://doi.org/10.1177/1094428112452151>.
- Gregory, S., & Sircar, A. (2025). *AI Governance Handbook: A Practical Guide for Enterprise AI Adoption*, Professional Practice in Governance and Public Organizations, Cham: Springer.
- Guzmán, E., Tobón, S., Andres, B. and Torres, M. (2025). AI-driven circular construction: Integrating social, economic, and environmental dimensions through a holistic triple-bottom-line

model, *Proceedings of the 2025 IEEE International Conference on Engineering, Technology and Innovation (ICE/ITMC)*, IEEE, <https://doi.org/10.1109/ICE/ITMC65658.2025.11106619>

Hahn, T., Preuss, L., Pinkse, J., & Figge, F. (2014). Cognitive frames in corporate sustainability: Managerial sensemaking with paradoxical and business case frames. *Academy of Management Review*, 39(4), 463–487.

Hair, J. F., Page, M., & Brunsveld, N. (2020). *Essentials of Business Research Methods*, 4th edn, New York: Routledge.

Heger, A. K., Passi, S., Dhanorkar, S., Kahn, Z., Wang, R., & Vorvoreanu, M. (2025). Towards a Responsible AI Organizational Maturity Model. *Proceedings of the ACM on Human-Computer Interaction*, vol. 9, no. 7, art. CSCW333, <https://doi.org/10.1145/3757514>

Hodgkinson, G. P., Burkhard, B., Foss, N. J., Grichnik, D., Sarala, R. M., Tang, Y., & Van Essen, M. (2023). The heuristics and biases of top managers. *Journal of Management Studies*, 60(5), 1033–1063.

Holmström, J., & Hällgren, M. (2022). AI management beyond the hype. *AI & Society*, 37, 1575–1585.

Hutzschenreuter, T., & Lämmermann, T. (2025). What is your AI strategy? *Academy of Management Perspectives*, 529–549.

Kellogg, K. C., Valentine, M. A., & Christin, A. (2020). Algorithms at Work: The New Contested Terrain of Control, *Academy of Management Annals*, vol. 14, no. 1, pp. 366-410.

Korteling, J. E., Paradies, G. L. & Sassen-van Meer, J. P. (2023). Cognitive bias and how to improve sustainable decision making, *Frontiers in Psychology*, vol. 14, 1129835

Krakowski, S., Luger, J., & Raisch, S. (2023). Artificial intelligence and the changing sources of competitive advantage. *Strategic Management Journal*, 44(6), 1425–1452.

Lebovitz, S., Lifshitz-Assaf, H., & Levina, N. (2022). To engage or not to engage with AI for critical judgments. *Organization Science*, 33(1), 126–148.

- Leonardi, P. M. (2012). Materiality, Sociomateriality, and Socio-Technical Systems: What Do These Terms Mean? How Are They Different? Do We Need Them?, in P. M. Leonardi, B. A. Nardi, & J. Kallinikos (eds), *Materiality and Organizing: Social Interaction in a Technological World*, Oxford: Oxford University Press, pp. 25-48.
- Magnani, G., & Gioia, D. A. (2023). Using the Gioia methodology in international business and entrepreneurship research. *International Business Review*, 32(2), 102096.
- Maitland, E., & Sammartino, A. (2015). Decision making and uncertainty. *Strategic Management Journal*, 36(10), 1554–1578.
- Maitlis, S., & Christianson, M. (2014). Sensemaking in Organizations: Taking Stock and Moving Forward, *The Academy of Management Annals*, vol. 8, no. 1, pp. 57–125, <https://doi.org/10.1080/19416520.2014.873177>
- Martin, K. (2019). Ethical Implications and Accountability of Algorithms, *Journal of Business Ethics*, vol. 160, no. 4, pp. 835-850.
- Mikalef, P. & Gupta, M. (2021). Artificial intelligence capability: Conceptualization, measurement calibration, and empirical study on its impact on organizational creativity and firm performance, *Information & Management*, vol. 58, no. 3, 103434
- Mintzberg, H., Raisinghani, D., & Théorêt, A. (1976). The structure of “unstructured” decision processes. *Administrative Science Quarterly*, 21(2), 246–275.
- Mustafa, F., Smolarski, J., & Elamer, A. A. (2025). The Convergence of Artificial Intelligence and Sustainability Reporting: A Systematic Review of Applications, Challenges and Future Directions, *Business Strategy and the Environment*, pp. 1-28.
- Murray, A., Rhymer, J., & Sirmon, D. G. (2021). Humans and technology: Forms of conjoined agency in organizations. *Academy of Management Review*, 46(3), 552–571.
- Narayan, S. (2025). The Socio-Technical Systems Perspective, in *The Bridgerton Paradox in Artificial Intelligence*, Springer Nature Switzerland AG, pp.65–75, https://doi.org/10.1007/978-3-031-99493-7_7

- Nguyen, P., Watson, G. P., Barnes, D., Agrawal, S., Schuster, A. M. & Cotten, S. R. (2026). Navigating workplace AI adoption: The influence of perceptions and affective attitudes on employees' intentions to use AI at work. *Journal of Management & Organization*, vol. 32, no. 1, pp. 1-26, <https://doi.org/10.1017/jmo.2026.10090>
- Nidumolu, R., Prahalad, C. K., & Rangaswami, M. R. (2009). Why sustainability is now the key driver of innovation. *Harvard Business Review*, 87(9), 57–64.
- Nishant, R., Kennedy, M., & Corbett, J. (2020). Artificial intelligence for sustainability. *International Journal of Information Management*, 53, 102104.
- Pachidi, S., Berends, H., Faraj, S., & Huysman, M. (2021). Make way for the algorithms. *Organization Science*, 32(1), 18–41.
- Papadakis, V. M., Lioukas, S., & Chambers, D. (1998). Strategic decision-making processes. *Strategic Management Journal*, 19(2), 115–147.
- Papagiannidis, E., Harris, M. A., & Morton, D. (2023). Toward AI Governance: A study identifying best practices, barriers and outcomes, *Information Systems Frontiers*, vol. 25, pp. 123–141, <https://doi.org/10.1007/s10796-022-10251-y>
- Pietrzak, P. (2025). How artificial intelligence can support the sustainable development of organizations? Findings from a literature review, *Issues in Information Systems*, vol. 26, no. 1, pp. 296-307.
- Phillips-Wren, G., & Adya, M. (2020). Decision making under stress. *Journal of Decision Systems*, 29(S1), 213–225.
- Powell, T. C., Lovallo, D., & Fox, C. R. (2011). Behavioral strategy, *Strategic Management Journal*, vol. 32, pp. 1369-1386.
- Raisch, S., & Krakowski, S. (2021). Artificial Intelligence and Management: The Automation-Augmentation Paradox, *Academy of Management Review*, vol. 46, no. 1, pp. 192-210.

Sachs, J. D., Lafortune, G., Fuller, G., & Iablonovski, G. (2025). Sustainable Development Report 2025. United Nations & Dublin University Press.

Sakchuenyos, S., Akapipat, S., & Hoang, H. (2025). The Effect of Greenhushing and ESG Disclosure Assurance on Investors' Willingness to Invest, *Accounting & Finance*, <https://doi.org/10.1111/acfi.70157>

Schneider, J., Abraham, R., Meske, C., & Vom Brocke, J. (2023). Artificial Intelligence Governance For Businesses, *Information Systems Management*, vol. 40, no. 3, pp. 229-249.

Schwaewe, J., Gerlich, C., Nguyen, H. L., Kanbach, D. K., & Gast, J. (2025). Artificial intelligence (AI) for good? Enabling organizational change towards sustainability, *Review of Managerial Science*, vol. 19, pp. 3013-3038.

Shepherd, N. G., Lou, B., & Rudd, J. M. (2024). Going with the gut. *Journal of Business Research*, 181, 114740.

Shrestha, Y. R., Ben-Menahem, S. M., & von Krogh, G. (2019). Organizational decision-making structures in the age of AI. *California Management Review*, 61(4), 66–83.

Slawinski, N., & Bansal, P. (2015). Short on Time: Intertemporal tensions in business sustainability, *Organization Science*, vol. 26, no. 2, pp.531-549, <https://www.jstor.org/stable/43663662>

Tamburrini, G. (2025). The heuristics gap in AI ethics. *Journal of Responsible Technology*, 21, 100104.

Teh, D. & Corbitt, B. (2015). Building Sustainability Strategy in Business, *Journal of Business Strategy*, vol. 36, no. 6, pp. 39–46, <https://doi.org/10.1108/JBS-08-2014-0099>

Vinuesa, R., Azizpour, H., Leite, I., Balaam, M., Dignum, V., Domisch, S., Fellländer, A., Langhans, S. D., Tegmark, M., & Fuso Nerini, F. (2020). The role of AI in achieving the SDGs. *Nature Communications*, 11(1), 233.

Wang, Q., Qi, Y., & Li, R. (2025). Artificial Intelligence and Corporate Sustainability: Shaping the Future of ESG in the Age of Industry 5.0, Sustainable Development.

Wen, Y., Wang, J. & Chen, X. (2025). Trust and AI weight: human-AI collaboration in organizational management decision-making. *Frontiers in Organizational Psychology*, vol. 3, 1419403, <https://doi.org/10.3389/forgp.2025.1419403>

Wright, C., & Nyberg, D. (2017). An Inconvenient Truth: How Organizations Translate Climate Change into Business as Usual, *The Academy of Management Journal*, vol. 60, no. 5, pp. 1633-1661, <https://doi.org/10.5465/amj.2015.0718>

Zhao, J., & Gómez Fariñas, B. (2022). Artificial intelligence and sustainable decisions. *European Business Organization Law Review*, 24, 1–39.

8.0 Declaration of AI Use

Tool	Task(s)
NotebookLM	• Verify underlying logics between journal articles and the thesis • Summarise articles to simplify the literature finding process • Ensure in-text citations are correctly placed and written • Generate references according to LUSEM Harvard referencing style
Microsoft Co-pilot, Google Gemini, Claude, NotebookLM	• Fine tune language (e.g. altering/checking verb tense) • Correct spelling • Search tool to find relevant concepts and journal articles

9.0 Appendices

9.1 Interview Guide: Sustainability Manager

SECTION A: INTRODUCTION & CONTEXT

Thank you for taking the time to participate in this research.

Everything you share will remain confidential. Your individual name and your organisation will be anonymised, including the thesis, publications, and presentations. If it is okay with you, we would like to record this interview and use Lund University's authorised transcription service. The recording will be transcribed and stored in accordance with GDPR. Are you ok with this?

[START RECORDING]

This is a Master's thesis study conducted at Lund University, investigating how companies balance the efficiency of artificial intelligence with human-centred governance when formulating sustainability strategies. We're interested in your genuine experience: how you and your team navigate these tensions in real strategic decisions. There are no right or wrong answers; we're exploring the decision-making process in your firm.

Warm-up Questions:

- Can you briefly walk me through your role and responsibilities at the company?
- How long have you been in this role, and what was your background before this?

SECTION B: SUSTAINABILITY STRATEGY & DATA LANDSCAPE

- When it comes to sustainability, what do you think would be opportunities and challenges your company faces at the moment?
- When you think about making a major sustainability decision, what data or information do you typically need to gather? Do you have any challenges with the data you gather?

SECTION C: CRITICAL INCIDENT: AI IN DECISION-MAKING

"I'd like you to think about a strategic decision or initiative with sustainability goals, in the last 6–12 months, using data or AI tools. Can you walk me through that process?"

Core CIT Questions:

- What was the business context for the decision or initiative?
- What specific sustainability trade-offs did you face? (e.g., cost vs. environmental impact, speed vs. thoroughness, one environmental goal vs. another)
- What data sources or AI tools did your team use to form this decision (during the decision-making process)?
- How did you use this AI output? Did you implement the recommendation directly, modify it, or not use it? Why? Were there concerns about data quality, bias, or hidden assumptions?
- Who else was involved in this decision? How did the AI output get communicated to leadership, and how did they respond?
- Looking back, what worked well in how you used AI or data for this decision? What would you do differently?

SECTION D: GOVERNANCE, OVERSIGHT & CONSTRAINTS

- How does your organisation currently govern data analytics tools/AI use in sustainability decisions? Are there formal frameworks, approvals, or risk assessments?
- Are there types of sustainability decisions where you *would not* use AI, regardless of technical capability? What are the limits?
- What would it take for you to have greater confidence in delegating more sustainability decisions to AI-assisted systems?

CLOSING

"Thank you so much for your time today.

It was incredibly helpful to hear your practical perspective as a sustainability manager. Moreover, we would love to make sure our research captures the full picture. We're hoping to pair your insights with the operational/data expert side of things. Would you be open to introducing us to someone who works directly with the data analytics or AI tools on your team?

We'd love to share our final findings with you once the project is complete as a thank you. In case you have any questions, feel free to reach out to us.

Thank you.

9.2 Interview Guide: Data Expert

SECTION A: INTRODUCTION & CONTEXT

Thank you for taking the time to participate in this research.

Everything you share will remain confidential. Your individual name and your organisation will be anonymised, including the thesis, publications, and presentations. We will record this interview and use Lund University's authorised transcription service. The recording will be transcribed and stored in accordance with GDPR. Are you ok with this?

[START RECORDING]

This is a Master's thesis study conducted at Lund University that investigates how companies balance the efficiency of artificial intelligence with human-centred governance when formulating sustainability strategies. We're interested in understanding technical perspectives, what the AI can and cannot do, where the risks lie, and how you navigate the handoff between AI recommendations and human decision-makers. There are no right or wrong answers; we want to understand the practical reality of building and maintaining these systems.

Warm-up Questions:

- Can you describe your role and responsibilities? What are you in charge of regarding AI and data systems for sustainability?
- How long has your organisation been actively using AI or machine learning for sustainability decisions? Walk me through that evolution. What came first?

SECTION B: AI ARCHITECTURE & CAPABILITY LANDSCAPE

- What are the primary AI tools or systems your company currently uses for sustainability strategy and decision-making?
- For these systems, how accurate are they? Do you have benchmarking data or validation results? Where do they break down or require human intervention?

SECTION C: CRITICAL INCIDENT: TECHNICAL DATA

"I'd like you to walk us through a specific, recent instance where you used an AI or advanced analytics tool to generate an insight, report, or recommendation that was

ultimately used for sustainability decision making. Can you walk me through that process?"

Core CIT Probe Questions:

- What was the business problem you were trying to solve? And what was the strategic driver?
- Walk me through how you used the input data or analytics tools. Describe the data you collected, how did you assess its quality and relevance?
- Walk me through the exact moment the AI gave you its output. Did you accept it as-is, or did you have to manually adjust it?
- How did you communicate the insights and their limitations to the business users and decision-makers? Did the business adopt them as intended?
- Did any unexpected problems emerge, technical failures, accuracy issues, unintended consequences, or ethical concerns?
- How do you currently oversee or govern the use of these analytics? What checks and balances exist?
- If you could redesign the process or the tools you used, what would you change?

SECTION D: DATA QUALITY, BIAS & TRANSPARENCY

- In your experience, how much of the challenge in AI for sustainability decisions is *technical* (building accurate models) vs. *organisational* (governance, trust, data access)?
- AI bias, black box, and incomplete data might be seen as key risks. Have you encountered these in your sustainability systems? How do you manage them?

SECTION E: HUMAN-AI COLLABORATION & GOVERNANCE DESIGN

- What would need to change in your organisation, technically, culturally, governance-wise, for you to have greater confidence in delegating more sustainability choices to AI systems?

CLOSING

"Thank you so much for your time today.

It was incredibly helpful to hear your practical perspective on the technical side.

Moreover, we would love to make sure our research captures the full picture. We're hoping to pair your insights with a governance perspective. Would you be open to introducing us to someone who works with data control or excellence?

We'd love to share our final findings with you once the project is complete as a thank you. In case you have any questions, feel free to reach out to us.

Thank you.

9.3 Interview Guide: Operational User

SECTION A: INTRODUCTION & CONTEXT

Thank you for taking the time to participate in this research.

Everything you share will remain confidential. Your individual name and your organisation will be anonymised, including the thesis, publications, and presentations. If it is okay with you, we would like to record this interview and use Lund University's authorised transcription service. The recording will be transcribed and stored in accordance with GDPR. Are you ok with this?

[START RECORDING]

This is a Master's thesis study conducted at Lund University, investigating how companies balance the efficiency of artificial intelligence with human-centred governance when formulating sustainability strategies. We're interested in your genuine experience: how you and your team navigate these tensions in real strategic decisions. There are no right or wrong answers; we're exploring the decision-making process in your firm.

Warm-up Questions:

- Can you briefly walk me through your role and responsibilities at the company?
- How long have you been in this role, and what was your background before this?

SECTION B: AI ARCHITECTURE & CAPABILITY LANDSCAPE

- What are the primary AI tools or systems your company currently uses for sustainability strategy and decision-making?
- For these systems, how accurate are they? Do you have benchmarking data or validation results? Where do they break down or require human intervention?

SECTION C: CRITICAL INCIDENT: AI FOR OPERATIONAL SUSTAINABILITY

"I'd like you to think about an operational decision or initiative with sustainability goals, in the last 6–12 months, using data analytics or AI tools. Can you walk me through that process?"

Core CIT Questions:

- What was the business context for the decision or initiative?
- What data sources or AI tools did your team use to form this decision (during the decision-making process)?
- How did you use this AI output? Did you implement the recommendation directly, modify it, or not use it? Why? Were there concerns about data quality, bias, or hidden assumptions?
- Who else was involved in this decision? How did the AI output get communicated to leadership, and how did they respond?
- Looking back, what worked well in how you used AI or data for this decision? What would you do differently?

SECTION D: GOVERNANCE, OVERSIGHT & CONSTRAINTS

- How does your organisation currently govern data analytics tools/AI use? Are there formal frameworks, approvals, or risk assessments?
- Are there types of sustainability areas where you *would not* use AI, regardless of technical capability? What are the limits?
- What would it take for you to have greater confidence in delegating more sustainability decisions to AI-assisted systems?

SECTION E: DATA QUALITY, BIAS & TRANSPARENCY

- In your experience, how much of the challenge in AI for sustainability decisions is *technical* (building accurate models) vs. *organisational* (governance, trust, data access)?
- AI bias, black box, and incomplete data might be seen as key risks. Have you encountered these in your systems? How do you manage them?

CLOSING

"Thank you so much for your time today.

It was incredibly helpful to hear your practical perspective as an operational user.

We'd love to share our final findings with you once the project is complete as a thank you. In case you have any questions, feel free to reach out to us.

Thank you.

9.4 1st Order Concepts

1st Order Concepts
High focus on cybersecurity, safety, and compliance over strategic value creation
Concern for public data or proprietary data impacting outputs
AI output assisting limited use cases, but not used for decision-making
Varying individual awareness of the correlation between the quality of the prompt and the AI output
Inconsistently communicating when and how AI is used internally
Sustainability requires nuance not captured by concise AI outputs
High focus on improving data quality and data structure
Desire to see AI output traceability
Large concerns for changing regulations and local variation
Implementing a strategy for AI to improve internal infrastructure and accuracy
Intentionally blocking AI from compliance reporting and social human rights data
Human oversight required for AI inputs/outputs for quality and risk considerations
A few firms developing AI agents for strategic use cases
Slow AI implementation due to knowledge gaps, corporate cultural and reluctance to change
Seeing AI as a tool for reviewing assumptions
Introduction of internal data centres/data owners to improve data quality
Predominant use external AI providers, typically Copilot for cybersecurity and system compatibility
Claims not to see bias in the AI system
Mistrust in AI due to low input data quality and hallucination in the output
In-house solutions built for cybersecurity concerns and specific requirements
Strategic use of AI for competitiveness through top-down reinforcement

High awareness and efforts to minimise AI hallucinations
Demand for top-down sustainability governance for effective AI use
Setup of AI excellence teams to govern corporate AI use
AI as a tool for efficiency for certain tasks, not replacing human expertise
AI requires education, feedback and training for users