Ida Merete Enholm

Responsible AI Governance and Its Effect on Competitive Performance

Master's thesis in Computer Science
Supervisor: Patrick Mikalef
June 2021

Master's thesis

NTNU
Norwegian University of Science and Technology
Faculty of Information Technology and Electrical Engineering
Department of Computer Science

NTNU
Norwegian University of
Science and Technology

Ida Merete Enholm

Responsible AI Governance and Its Effect on Competitive Performance

Master's thesis in Computer Science
Supervisor: Patrick Mikalef
June 2021

Norwegian University of Science and Technology
Faculty of Information Technology and Electrical Engineering
Department of Computer Science

NTNU
Norwegian University of
Science and Technology

Abstract

Artificial Intelligence (AI) promises several benefits for firms, but at the same time, it also introduces new challenges and risks not seen in previous technologies. Organizations should establish responsible AI governance practices to minimize the potential risks of AI while at the same time maximizing the potential benefits. However, there is uncertainty about what exactly makes up responsible AI governance and how it can help organizations attain a competitive advantage. This thesis explores the field of responsible AI governance by employing a mixed methods approach. First, a single case study provides in-depth insight into how a company has successfully managed to control and govern its AI. Building on this case study and previous literature on AI governance and responsible AI, this thesis provides a definition of what responsible AI governance entails and identifies several principles that organizations should govern their AI according to. Moreover, this thesis develops a survey instrument to capture the responsible AI governance of firms. Then, a survey method is employed to examine the effect responsible AI governance has on knowledge management capability, organizational agility, and competitive performance. Survey data from 144 high-level IT executives working in Nordic companies are examined to test the proposed research model. The findings empirically support the proposed research model and prove that firms can increase their KMC and organizational agility by deploying responsible AI governance, which in turn enhances their competitive performance.

Sammendrag

Kunstig intelligens («Artificial Intelligence» – AI) kan gi bedrifter mange fordeler, men det bringer også med seg flere utfordringer og risikoer som ikke er sett ved bruken av teknologi tidligere. Bedrifter bør etablere praksis for ansvarlig styring av AI for å minske disse risikoene, men samtidig utnytte potensialet fra AI. Det er imidlertid usikkerhet rundt hva ansvarlig styring av AI betyr for bedrifter, og hvordan det kan hjelpe bedrifter med å oppnå et konkurransefortrinn. Denne masteroppgaven utforsker feltet for ansvarlig styring av AI ved å bruke en "mixed methods" tilnærming. Først utføres en casestudie for å få innsikt i hvordan en utvalgt bedrift har lykkes med å kontrollere og styre sin AI. Basert på denne casestudien og tidligere litteratur om AI-styring og ansvarlig AI, legger denne masteroppgaven frem en ny definisjon av begrepet ansvarlig AI-styring og hva det innebærer, samt identifiserer flere prinsipper som bedrifter burde styre og kontrollere AI i henhold til. Denne masteroppgaven utvikler også et kartleggingsinstrument for å måle graden av ansvarlig AI-styring i bedrifter. Deretter brukes en spørreundersøkelse for å undersøke effekten som ansvarlig AI-styring har på bedrifters evne til å håndtere kunnskap, organisatorisk smidighet og konkurransedyktighet. Undersøkelsesdata fra 144 nordiske IT-ledere på øverste nivå i sine selskaper blir analysert for å teste forskningsmodellen som er lagt frem. Funnene støtter empirisk den foreslåtte forskningsmodellen og viser at bedrifter kan øke sin evne til å håndtere kunnskap og organisatoriske smidighet ved å ha ansvarlig AI-styring, noe som igjen forbedrer deres konkurransedyktighet.

Preface

This thesis is written during Spring 2021 as part of the course TDT4900 Computer Science, Master's Thesis at the Norwegian University of Science and Technology (NTNU) in Trondheim, Norway. The work of this thesis builds on previous work from the course TDT4501 Computer Science, Specialization Project, completed during Fall 2020. The project was supervised by Associate Professor Patrick Mikalef and conducted within the Information Systems and Software Engineering Group at the Department of Computer Science at NTNU.

I want to give special thanks to my supervisor, Patrick Mikalef, for providing invaluable guidance and support throughout the work of this thesis. Additionally, I would like to thank Ph.D. candidate Emmanouil (Manos) Papagiannidis for collaborating with me throughout the project and being a good discussion partner. Thanks also go to Christian Dremel and John Krogstie for their participation and feedback.

Contents

List of Figures

List of Tables

1 Introduction

1.1 Motivation

Artificial Intelligence (AI) has gained much attention in recent years because of its potential benefits. Organizations implementing AI in their line of work are expected to attain several advantages in terms of added business value, such as increased revenue and cost reduction (Alsheibani et al., 2020). In the search for competitive advantage, many organizations are thus investing in AI technologies. However, despite the growing interest in AI, many companies struggle to realize value from their AI investments (Fountaine et al., 2019). There are several challenges, concerns, and risks associated with adopting AI technologies that should be addressed.

The advancements in AI have raised ethical concerns about how the technology is applied (Butcher & Beridze, 2019). According to KPMG (2021), 94% of IT decision-makers want their organization to focus more on corporate responsibility and ethics while developing AI solutions. AI exhibits many of the same traits as humans, which can result in human jobs being automated away (Ford, 2013). Also, many of the AI models developed today can be seen as black-boxes that are difficult to understand (Adadi & Berrada, 2018; Loyola-González, 2019). Moreover, how can one trust a decision made by a machine without knowing the reasoning behind the decision? Other concerns relate to the risk of algorithmic bias and the ability of AI models to provide fair results (Ntoutsi et al., 2020). There are several real-world examples of companies getting negative attention due to their AI applications being discriminatory, such as Apple Card being accused of being "sexist" against women applying for credit (Vigdor, 2019). These issues should be addressed when developing and deploying AI applications.

1.2 Problem Statement

To realize value from AI investments, there is a need to build trust in AI both inside and outside the organization (Accenture, 2021). Moreover, to build trust, organizations want to minimize the risks and unintended consequences of AI (Siau & Wang, 2018). However, according to Accenture (2019b), only 11% of risk leaders feel capable of assessing the risks that AI brings. Thus there is a need for increased guidance on how to govern AI technologies and manage the potential unintended consequences to help bridge the gap between AI's potential and the risks it brings (KPMG, 2021). Because of the high speed of AI innovation, laws and regulations struggle to keep up with the most up-to-date AI technologies. Therefore, given the scale and transformative impact of AI, businesses should be proactive and develop responsible AI governance practices on their own before regulations are caught up. However, there is a lack of a coherent understanding of what exactly makes up responsible AI governance and how organizations can implement it in practice.

While there is much discussion regarding the advantages and necessities of responsible, ethical, and trustworthy AI, its effects on organizations are still uncertain. There is little empirical work demonstrating the mechanisms through which it affects organizations. More specifically, if and how it can help organizations enhance performance and attain a competitive advantage.

1.2.1 Research Questions

The main goal of this thesis is to explore the field of responsible AI governance. In particular, this thesis aims to investigate what precisely responsible AI governance entails for organizations and how to put it into practice. Also, its effect on organizations is examined. More precisely, if and through what mechanisms organizations can attain a competitive advantage by deploying responsible AI governance. These problems can be expressed by the following research questions:

- **Research question 1:** What does responsible AI governance comprise, and how is it implemented in practice?
- **Research question 2:** What are the effects of deploying responsible AI governance, and through what mechanisms are performance gains realized?

1.3 Research Method

This thesis aims to investigate the research questions by employing a sequential exploratory mixed methods approach. First, an in-depth case study is performed to learn how a company has successfully managed to adopt AI and how they control and govern their AI to act according to the organizational objectives. Then, building on the case study and existing literature, the notion of responsible AI governance is defined, and a theoretical framework for responsible AI governance is developed. This framework presents several dimensions that organizations should govern their AI according to. In addition, a survey instrument is developed to quantify and measure an organization's maturity in terms of responsible AI governance. Next, a research model is proposed. I hypothesize that responsible AI governance will affect an organization's knowledge management capability and organizational agility, which in turn will enhance their competitive performance. These relationships are examined through a survey sent out to high-level IT executives working in Nordic companies.

1.4 Thesis Structure

The rest of this thesis is structured as follows. The next chapter (Chapter 2) introduces AI in an organizational setting by discussing several of its essential concepts. Then, in Chapter 3, the research approach is outlined. In Chapter 4, the in-depth case study is discussed in further detail. Next, in Chapter 5, the responsible AI governance instrument is conceptualized. Chapter 6 introduces a research model proposing hypotheses about the effects of deploying responsible AI governance. Following, Chapter 7 presents the methodology for the survey study, which is used to test the research model. The results from the survey study are presented in Chapter 8. In Chapter 9, findings from both the case study and survey study are discussed, as well as the limitations of this research. Lastly, Chapter 10 provides some concluding remarks to this work.

2 Background

This chapter aims to give an introduction to the domain of AI in organizations. First, the most important concepts related to AI are presented. Second, the challenges organizations face when adopting AI are discussed. Lastly, the concept of AI governance is presented as a set of concepts relevant to managing AI.

2.1 Defining Core Concepts of AI

Even though AI has gained much attention in recent years, there is still ambiguity around the notion. Since the foundation of AI as a scientific field in the 1950s, several definitions of AI have been published in an attempt to differentiate it from other conventional information technologies. However, there is still no universally accepted definition of the term (Wang, 2019). A reason for this is that AI is not a single technology but rather a set of technologies and sub-disciplines that are rapidly evolving (Schmidt et al., 2020; Wamba-Taguimdje et al., 2020). Therefore, it is necessary to draw a clear distinction between the core concepts of AI, specifically: *AI as a scientific discipline, technologies used to realize AI,* and *AI capabilities.* The next subsections differentiate the three concepts.

2.1.1 Artificial Intelligence

In the absence of a universally accepted definition of AI, several definitions of AI are identified in the literature to enable a more holistic understanding of the term. Five of the definitions are presented in Table 1. It is evident from these definitions that there is a consensus that AI refers to algorithms giving the computer human-like capabilities. This means giving the computer the ability to perform activities that usually require human intelligence, such as learning, reasoning, and problem-solving (Afiouni, 2019; Demlehner & Laumer, 2020; Mikalef & Gupta, 2021). More generally, one can say that AI refers to computers that exhibit traits that are associated with human minds. The aim of AI is to reproduce human cognition by emulating how humans learn and process information without being explicitly programmed (Demlehner & Laumer, 2020). This description implies that AI should be able to sense, interpret, plan, learn, comprehend, and act all on its own. In other words, AI should be able to correctly interpret external data, learn from it, and use this learning to achieve predetermined organizational and societal goals (Mikalef & Gupta, 2021).

Table 1: Sample Definitions of Artificial Intelligence

Author(s)	Definition
Kolbjørnsrud et al. (2017)	Computers and applications that sense, comprehend, act, and learn
Afiouni (2019)	The general concept for computer systems able to perform tasks that usually need natural human intelligence
Schmidt et al. (2020)	The endeavor to mimic cognitive and human capabilities on computers
Demlehner and Laumer (2020)	A computer system having the ability to percept, learn, judge, or plan without being explicitly programmed to follow predetermined rules or action sequences throughout the whole process
Mikalef and Gupta (2021)	The ability of a system to identify, interpret, make inferences, and learn from data to achieve predetermined organizational and societal goals

2.1.2 AI Technologies

Several techniques can be used to realize the objectives of AI. For the past years, the focus has been on *machine learning* and *deep learning,* following the increase in data availability and the advances in computational power (Afiouni, 2019). Machine learning is a subset of AI techniques, and is about training a machine to be capable of learning from data, make inferences, predict, and identify associations, which can guide decisions (Afiouni, 2019; Wang et al., 2019). Machine learning algorithms can be sub-divided into four categories: *supervised, semi-supervised, unsupervised,* and *reinforcement learning* (Wang et al., 2019). In supervised learning, the target value is included in the training data, from which the system identifies patterns and infer its own rules (Afiouni, 2019; Schmidt et al., 2020). On the other hand, unsupervised learning does not include the target value in the training data. The structure of the training data and its statistical properties are used to solve the problem (Afiouni, 2019). In semi-supervised learning, both labeled and unlabeled data are used. In contrast, reinforcement learning does not learn from past data. Rather, the system is driven by experiences. The system has an objective and receives rewards based on how well this objective is met (Afiouni, 2019). Learning is then enabled through this feedback.

Conventional (shallow) machine learning techniques are limited in their ability to process raw data, as they require a good feature extractor to transform the raw data into features that can be used by the learner (LeCun et al., 2015). In contrast, deep machine learning, usually referred to as deep learning, can learn good features automatically. Deep learning is based on the use of an artificial neural network architecture, which imitates how the neurons in the human brain works (Afiouni, 2019; Jelonek et al., 2020; Schmidt et al., 2020; Wamba-Taguimdje et al., 2020). These neural networks are composed of multiple layers, where the layer closest to the data vectors learns simple features and the higher layers learn higher-level features (Quinio et al., 2017). Deep learning is producing promising results for various tasks and has thus gained considerable attention in the last years.

Machine learning and deep learning are often used in combination with other AI technologies to provide solutions that can evolve and learn. Examples of such technologies are Natural Language Processing (NLP), machine vision, and speech synthesis systems. NLP refers to the process in which machines can understand and analyze human language (Jarrahi, 2018), and it can be used for a wide range of applications, such as chatbots and classification of emails. Machine vision refers to algorithmic inspection and analysis of images to automatically extract information from an image (Jarrahi, 2018). Possible use cases for machine vision are the detection of objects and patterns in images. Speech synthesis systems refer to solutions that can translate text to speech and speech to text (Lichtenthaler, 2019). Examples of solutions that exploit such technologies are virtual assistants, such as Amazon Alexa and Google Home.

2.1.3 AI Capabilities

AI is increasingly becoming an essential asset for organizations to gain a competitive advantage. However, gaining a competitive advantage from AI requires organizations to leverage not only technological resources, such as the AI technology itself. Organizations should also acquire and leverage other organizational resources, as the technological resources alone are easily acquired by competitors (Mikalef & Gupta, 2021). Therefore, the notion of AI capability has been introduced to extend the view of AI to not only focus on the technological resources but also include all related organizational resources that are of importance to exploit the full potential of AI. In other words, the notion of an AI capability is about how an organization selects, orchestrates, and leverages all its AI-specific resources, both technological (e.g., training data and AI-algorithms) and non-technical (e.g., employee skills), to enable value creation (Mikalef & Gupta, 2021; Schmidt et al., 2020; Wamba-Taguimdje et al., 2020).

2.2 Challenges of Implementing AI in Organizations

For the past years, organizations are increasingly turning to AI in the search for competitive advantage (Ransbotham et al., 2017). However, most AI initiatives fail, even though time and effort are being invested. The introduction of AI brings a new set of barriers and challenges for organizations to overcome to successfully implement AI technologies, and to create an AI capability. An organization's ability to successfully deploy and utilize AI depends on several factors relating to *technological readiness, organizational aspects*, and *environmental factors*. Some of these factors, which can either promote or impede AI deployments, are discussed below.

2.2.1 Technological Readiness

To successfully deploy AI, organizations need to understand the technological resources that are required. A common challenge for organizations wanting to adopt AI technologies is the lack of knowledge regarding technological requirements. Three things are needed when deploying AI: *computing power infrastructure, algorithms*, and *rich data sets* (Wamba-Taguimdje et al., 2020). AI learns to make decisions based on data rather than being explicitly programmed to perform a task. For the AI to obtain this ability, it should be trained on massive data sets. Thus organizations need to produce or have access to large amounts of data (Demlehner & Laumer, 2020; Mikalef & Gupta, 2021). However, it is not enough to merely have large amounts of data. The data must also be of high quality (Baier et al., 2019; Demlehner & Laumer, 2020). "Garbage-in, garbage-out" is a fundamental principle for AI. This principle means that low-quality training data will generate low-quality insights, which are not useful for the organization.

AI algorithms build models based on these data sets. These models are, in turn, used to make predictions. The data sets can be enormous, and the algorithms complex, which could require an infrastructure with massive amounts of computing power (Baier et al., 2019; Wamba-Taguimdje et al., 2020). For many companies, it is not feasible to have these resources on-site. Large companies, like Google and Amazon, have thus started to provide cloud-based solutions, such as Amazon AWS and Google Cloud AI. These solutions allow organizations to choose if they want to have the infrastructure on-site, in the cloud, or a combination of the two.

2.2.2 Organizational Aspects

Besides the technological resources, various organizational resources are needed to successfully adopt AI and build firm-specific and hard-to-imitate AI capabilities. Several studies have pointed out that the lack of leadership to support AI is one of the most critical challenges to overcome to realize value from AI investments (Alsheibani et al., 2020; Demlehner & Laumer, 2020). The top managers play a crucial role in establishing an environment that fosters AI initiatives, as well as allocating the resources needed, such as financial resources (Pumplun et al., 2019). Organizations with an innovative culture that exploits and supports new ideas are better positioned to integrate a transformative technology as AI into their line of work (Mikalef & Gupta, 2021; Pumplun et al., 2019).

Working with AI brings a new set of skill requirements for both technical and managerial personnel. The lack of technical skills is a great challenge for many organizations wanting to adopt AI. Organizations need employees with technical skills to create and deploy AI systems, e.g., to utilize technical AI libraries such as TensorFlow and PyTorch (Pumplun et al., 2019). They also need domain experts who understand the workings of the existing business processes, and understand how AI can improve these processes (Pumplun et al., 2019). Organizations should thus ensure that both technical and managerial staff have an understanding of the potentials of AI, how to utilize AI technologies, and which business areas are appropriate to target (Mikalef & Gupta, 2021).

Another challenge mentioned by several studies is the challenge of integrating AI projects with existing processes and systems (Davenport & Ronanki, 2018). New requirements will arise when integrating AI solutions, and the organization's business processes will need to adapt to these requirements. How this is accomplished should be described in a dedicated AI strategy. The AI strategy should describe how the organization will implement AI by providing a concrete plan to realize the desired objectives, and it should be aligned with the company's existing goals.

2.2.3 Environmental Factors

Organizations operate in dynamic and constantly changing environments that influence the way organizations can and should conduct business. There are several factors related to these environments which can challenge the adoption of AI. As AI can perform tasks previously reserved for humans, several ethical and moral aspects should be considered (Ntoutsi et al., 2020). Transparency, bias, and discrimination are only some of the challenges emerging when developing AI systems (Baier et al., 2019). Organizations need to reflect on the ethical issues of AI to make sure that its use aligns with the organization's values.

Regulations and laws can affect the way AI can be deployed in an organization. An example of this is the General Data Protection Regulation (GDPR) which was enforced in the European Union (EU) and the European Economic Area (EEA) in May 2018. GDPR regulates activities concerning the processing of personal data and can cause issues for organizations wanting to deploy AI solutions that are trained using personal data (Goodman & Flaxman, 2017). Other regulations can be industry-specific and define how companies in that particular industry can interact with their environment.

2.3 AI Governance

Organizations can experience several challenges when adopting, developing, and deploying AI. To mitigate these challenges and exploit the potentials of AI, implementing governance mechanisms are crucial (KPMG, 2021). The notion of AI governance is about how organizations can govern and monitor their AI capabilities through rules, practices, and processes. Organizations employing AI technologies should be able to control their AI systems so that it behaves according to the organizational strategies and objectives. By governing their AI capabilities, organizations help to minimize the potential downsides and risks of AI while at the same time exploiting AI technologies' potentials in the organization.

AI governance can be examined from different perspectives, both of which should be guided by a set of principles. It can be understood as a function describing the different mechanisms of AI governance (Government of Singapore, 2020). Alternatively, it can be understood as a process spanning all stages of AI projects' life cycle (Amershi et al., 2019). These two perspectives on AI governance are of varying interest to employees in organizations, depending on their position and role. AI governance as a function describes the various types of AI governance practices and mechanisms an organization should apply. This perspective is essential, especially to high-level executives, to know what types of AI governance mechanisms to employ. On the other hand, AI governance as a process describes the practices and processes used to govern AI systems in the different stages of the AI life cycle. To employees who are developing or working with AI solutions, this view is vital to know what practices and processes should be executed at what time. The two perspectives of AI governance are further described below.

2.3.1 Principles of AI

With the increasing interest in AI, the potentially negative impacts of AI are also getting more and more attention (Castillo et al., 2020). AI introduces new ethical, legal, and governance challenges, such as the risks of unintended discrimination and bias and issues related to the customers' awareness and knowledge about how AI is involved in making decisions (Government of Singapore, 2020). There are several real-world examples of AI solutions that have had negative consequences. An example is the Twitter chatbot called Tay, released by

Microsoft in 2016 (Wolf et al., 2017). In less than 24 hours, Tay needed to be shut down after posting offensive tweets because users were teaching it politically incorrect phrases. Another more significant example is U.S. courts using AI to predict the likelihood of a criminal committing a new crime in the future (Angwin et al., 2016). The system has proven biased against black people, almost twice as often misclassifying black people as future criminals than white people.

These are only a few of many examples showing that AI has possible downsides and risks that organizations need to take into account when adopting and deploying AI. To avoid these negative consequences, there is a need for a set of principles that can guide organizations deploying AI. Several efforts have been made in defining principles for ethical, trustworthy, and responsible AI. These initiatives are made by various stakeholders, ranging from private companies (Benjamins et al., 2019; Google, 2020), academic research (Clarke, 2019; Kumar et al., 2021; Thiebes et al., 2020), consultancy firms (Accenture, 2018; PwC, 2019), institutions (European Commission, 2019; Government of Singapore, 2020; National New Generation Artificial Intelligence Governance Committee, 2019) to non-profit organizations (IEEE, 2019). For instance, The European Commission has developed guidelines for trustworthy AI (European Commission, 2019). IEEE has addressed ethical considerations that should be taken into account when designing and developing AI (IEEE, 2019). Also, several governments have published national AI strategies, and guiding principles, such as the government of Singapore (Government of Singapore, 2020).

2.3.2 AI Governance as a Function

One perspective of AI governance sees it as a function, consisting of various mechanisms and practices that organizations can employ to govern the deployment and use of the relative technologies. Previous research on IT governance (Peterson, 2004), information governance (Borgman et al., 2016; Tallon et al., 2013), and data governance (Tallon, 2013) have decomposed governance into a range of structural, procedural, and relational practices. These practices have not been analyzed in the context of AI. However, they can be used as a baseline to understand how to build practices to achieve AI governance. The functions of AI governance can thus be divided into three categories of practices: (a) structural, (b) procedural, and (c) relational governance mechanisms, which are further explained below.

Structural Practices

Structural governance practices are about connecting business with AI management and decision-making functions. They comprise reporting structures, governance bodies, and accountabilities (Borgman et al., 2016). The main mechanisms of the structural practices are the formal positions and roles, as well as formal groups and team arrangements (Peterson, 2004). This means identifying the key decision-makers regarding AI and their respective roles and responsibilities. An example is having a Chief AI Officer (CAIO) who is in charge of all the AI-related activities. Formal groups are the structures used to coordinate decision-making across business and AI management functions. This can include specifying committees to oversee compliance with internal policies, principles and requirements of responsible AI.

Procedural Practices

Procedural governance practices concern the policies, processes, standards, and protocols used by organizations to execute AI governance. The goal is to ensure that the AI systems and models operate as expected and according to principles and objectives. Procedural practices comprise the strategic decision-making and monitoring, and to what extent they follow specified rules and standard procedures (Peterson, 2004).

Relational Practices

The relational practices of AI governance cover the aspects of collaboration between all stakeholders. AI governance involves a large group of stakeholders, from top-level managers to the users of the AI solutions. Relational governance practices describe the formalized links among all these stakeholders in terms of how knowledge is shared and how stakeholders are educated and trained in the use of the AI systems.

2.3.3 AI Governance as a Process

AI governance can also be seen as a process that spans all stages of the AI project life cycle. Figure 1 shows a commonly used workflow for machine learning projects presented by researchers at Microsoft (Amershi et al., 2019).

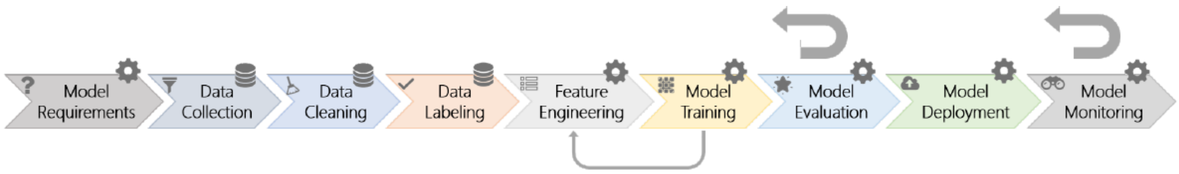

Figure 1: The Stages of the AI Project Life Cycle. Source: Amershi et al. (2019).

The first stage in the AI project life cycle is the *model requirements* stage, where the features to be implemented are decided on, as well as what type of model is appropriate to use. Then, in the *data collection* stage, data used to train the model is acquired, either through already existing datasets or by collecting new data. This data is then going through a *data cleaning* stage, where inaccurate, incomplete, irrelevant, and dirty data are corrected or removed. After being cleansed, the datasets go through a *data labeling* stage where ground truths are assigned to each record in the dataset. Next, *feature engineering* is performed to select the features that the model will work with. For deep learning models, however, features can be automatically learned by the model (LeCun et al., 2015), making the step of manual feature engineering redundant. The selected features are then used to train the model in the *model training* stage. After training, *model evaluation* is performed to evaluate the performance of the model with some pre-defined metrics. If the model performs as wanted, the model is *deployed* in the real world. The deployed model is then continuously *monitored* in the case of errors. Throughout the AI project life cycle, there are several feedback loops to be able to respond to change, making the process highly non-linear (Amershi et al., 2019).

For all stages, there are a set of activities and mechanisms that should be established to ensure that the AI solution behaves as intended and is in line with the principles of AI.

3 Research Approach

This thesis aims at investigating the research questions presented in Chapter 1.2 by employing a mixed methods approach. The design of the thesis is presented in Figure 2 and discussed below.

3.1 Preparation

As a preparation for this thesis, a systematic literature review (SLR) was performed during Fall 2020. The review was a part of a specialization project, resulting in a report and a journal article. The objective of the SLR was to identify inhibitors and enablers of AI adoption, in which ways organizations can deploy AI, and what value-generating mechanisms AI can enable. From the review, several areas for further research were identified, one of which was about the governance of AI projects. This work motivated the creation of the two research questions that guide this study.

3.2 This Thesis

This thesis aims to provide a holistic understanding of responsible AI governance and try to shine a light on how organizations control and govern their AI so that it behaves responsibly. Given the exploratory nature of the research issue, and the need to build theory in a relatively new research area, a mixed methods approach is employed. Mixed methods research uses both quantitative and qualitative research methods to understand a phenomenon and is a powerful method when existing theories do not sufficiently explain the phenomenon of interest (Venkatesh et al., 2013). The design of the approach is *sequential* and *exploratory*. First, qualitative data are collected. Then, quantitative data are collected to test the findings from the qualitative data.

The guidelines of Venkatesh et al. (2013) for conducting mixed methods research guided the research process, which is presented in Figure 2. First, an in-depth case study was performed to gain in-depth insight into the domain. The qualitative data was collected through semi-structured interviews. The goal of the case study is to explore how AI is used and governed in an organization that employs AI for critical parts of their work. The in-depth case study is described in further detail in Chapter 4.

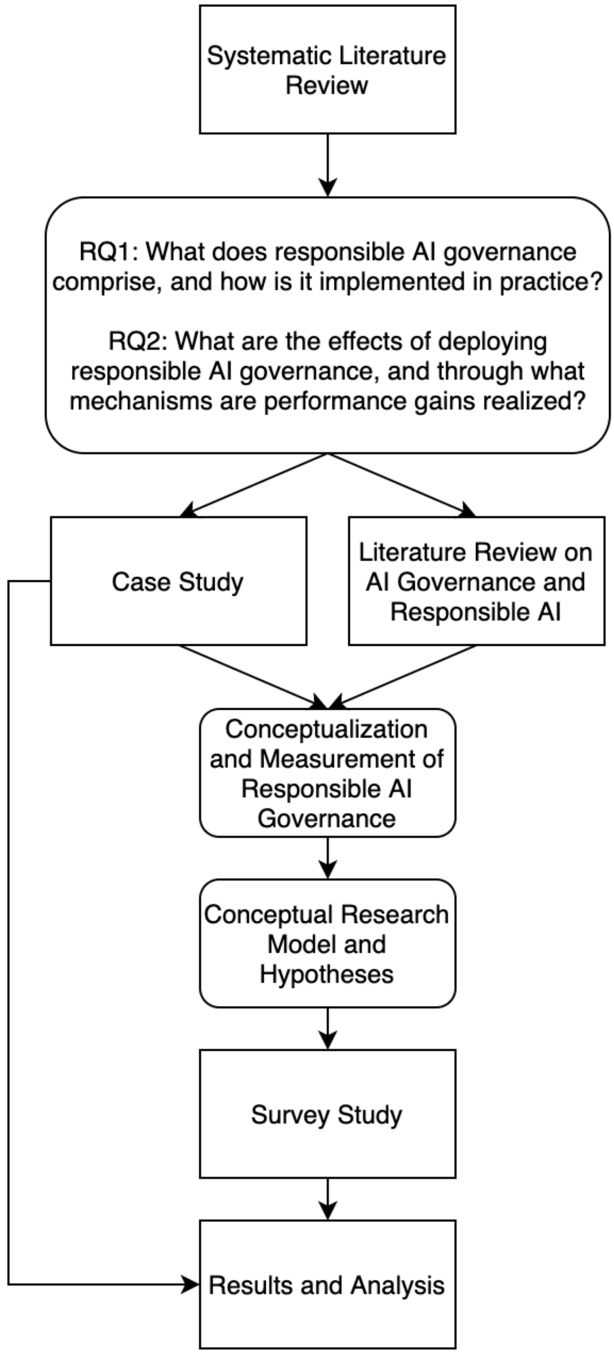

Figure 2: The Research Design and Process

Concurrently, a review of the existing literature on AI governance and trustworthy, ethical, and responsible AI was performed. Building on this review and the case study, the responsible AI governance instrument is created. This work includes conceptualization and dimensionalization of the term responsible AI governance, as well as creating a survey instrument to quantify and measure an organization's maturity in terms of responsible AI governance. Also, a research model is proposed, containing several hypotheses about the impact responsible AI governance has on organizations. Chapter 5 presents the responsible AI governance instrument, while the proposed research model is presented in Chapter 6.

Then, a survey method was used to test the research model empirically. The quantitative data for the survey were collected through a questionnaire sent out to Nordic companies. Chapter 7 further explains the survey method.

Lastly, the results from the case study and the survey study were analyzed. This process included drawing meta-inferences (Venkatesh et al., 2013). In other words, integrate findings from both the qualitative and quantitative studies. By drawing meta-inferences, a holistic explanation of the phenomenon of responsible AI governance can be provided.

4 In-depth Case Study

A qualitative method was employed to gain in-depth insight into the domain of AI use in organizations. A single, in-depth case study is the method of choice. This chapter presents the methodology of the case study and the analysis of the qualitative data.

4.1 Qualitative Method

The qualitative method aims at exploring the phenomenon of AI by investigating how companies are using AI to realize their organizational objectives and create business value (Plastino & Purdy, 2018; Ransbotham et al., 2017; Wamba-Taguimdje et al., 2020). In particular, the study aims to explore the mechanisms of value generation and realization and the specific challenges that AI technologies bring. In addition, the mechanisms used to control the behavior of AI solutions so that it acts upon the goals of the organization are explored.

An exploratory single case study approach is used to explore the field of AI use in organizations. Case studies are helpful in acquiring an in-depth understanding of a phenomenon (Yin, 2003), and a single case study is thus a good approach to gain deep knowledge on how a company can exploit and control its AI capabilities in the real world. Data were collected through semi-structured interviews with multiple respondents within the company. Also, secondary data sources (e.g., reports) are used to triangulate and verify results.

This subchapter presents the case that is studied, as well as describes how data are collected and analyzed.

4.1.1 Selection of Case

The process of selecting the case to study, targeted companies that have successfully adopted AI technologies. In addition, several other factors were considered to select a case that represents the population well. The company chosen to study should currently deploy AI solutions to support operations. Also, the company should utilize machine learning, as that is the AI technology of choice for most companies these days. The selection process was performed in collaboration with my supervisor.

The company chosen to study, from now on denoted as PowGen, was chosen because of their successful experience in using AI in critical parts of their work. Also, the AI solutions they are deploying are based on machine learning techniques. PowGen is further described in Chapter 4.1.4.

4.1.2 Collection of Data

Data were collected over a period of two weeks in February 2021. The data were collected through semi-structured interviews with five employees in PowGen. Semi-structured interviews are a flexible way of conducting interviews. It provides a general structure to keep the direction of the interview on track according to the research objective, but at the same time, it allows the researchers to further examine topics that emerge during the interview (Oates, 2006).

An interview guide (see Appendix A) was developed to provide the structure of the interviews. The interview guide contained guidelines for the interview in the form of open-ended questions that were directly tied to the research objective. Open-ended questions encourage the interviewees to share their opinions and experiences and are useful for gaining in-depth insight into a domain. The interview guide was split into two parts. The first part focused on the organizational effects and challenges of using AI and how it was used to transform existing processes. The second part was more focused on the technical aspects and challenges faced when implementing AI solutions. The interview guide was used as a reference template for all the interviews but was altered slightly depending on the interviewees' position and field of expertise. All the employees were interviewed separately to prevent the answers from one employee from influencing the answers of his or her colleagues.

In total, five employees from PowGen were interviewed. This included people with both technical and managerial skills. The roles of the respondents and the interview length are presented in Table 2. The person responsible for the AI team and the AI solutions, the chief AI officer, was a natural starting point. Concerning technical employees, three machine learning engineers from the AI team were interviewed. One of the machine learning engineers also has a managerial role in the AI team. A power market analyst and project manager that had worked closely with the AI team were also interviewed to get another point of view on the use of AI in the company.

The data collected from the interviews were handled according to laws and regulations. The study was approved by the Norwegian Centre for Research Data (NSD). Before each interview, an information letter was sent to the participants, including a consent form that they had to fill out. The interviews were performed digitally, and a screen recording, including audio, was performed. These recordings were stored using NTNU Box, which satisfies Norwegian laws and regulations for data storage.

Table 2: Respondents' Role and Length of Interview

Respondent ID	Role	Interview Time
1	Chief AI Officer	1 hour and 32 minutes
2	Machine Learning Engineer	55 minutes
3	Machine Learning Engineer	45 minutes
4	Machine Learning Engineer + Managerial Role	43 minutes
5	Project Manager + Power Market Analyst	49 minutes

4.1.3 Analysis of Data

Before analyzing the data, the recordings from the interviews had to be transcribed. The recordings were transcribed in two steps. First, the voice recordings were transcribed using a software called Otter[1]. Otter uses AI to automatically transcribe voice recording. Second, the transcriptions created by Otter were checked and corrected manually.

The transcriptions were then imported into the software NVivo[2], which is a data analysis software for qualitative and mixed methods data. In NVivo, the data were coded. A combination of deductive and inductive reasoning was used when analyzing the data. An inductive approach aims at developing a theory by coding the data in an open-minded way, going from observations to generalizations (Oates, 2006; Thomas, 2006). A deductive approach, however, moves the other way around. The goal of a deductive approach is to test an existing theory by moving from generalizations to observations (Oates, 2006; Thomas, 2006).

First, data were coded based on terms and concepts found in the data, not based on any pre-existing idea. This process is known as *open coding*. I went through the collection of data, looking for any themes emerging, which resulted in a list of 208 codes. Based on the list of codes that emerged, *axial coding* was applied. This means moving to a higher level of analysis, looking for relationships between the different codes (Oates, 2006). In the end, a list of 7 codes remained, including the following themes: AI adoption, AI strategy, human-AI relationship, AI development, challenges, organizational change, and business value.

[1] https://otter.ai/
[2] https://www.qsrinternational.com/nvivo-qualitative-data-analysis-software/home

Second, I went through the data material one more time. This time using a set of *pre-defined codes* to label the data. The set of codes used in this process was related to the concept of AI governance and included the different practices that governance mechanisms can be divided into. More specifically, *structural, procedural,* and *relational* practices.

To triangulate and validate the findings, a number of other sources were examined. These sources included reports, presentations, and public information. By looking at other sources of information, on top of the interview data, the findings from the interviews can be confirmed to be accurate.

4.1.4 PowGen: Power Generation Company

This section gives a brief introduction to the company studied. The real name of the company is withheld for confidentiality reasons. For simplicity, the company is referred to as PowGen.

PowGen is a power generation company that produces renewable energy through hydropower plants and wind farms. Around 10% of the energy is produced by wind. The rest comes from hydropower. In addition to power generation, PowGen is responsible for power trading. The company is also involved in other energy-related businesses, such as the development of future-oriented energy-related services.

The company is situated in Norway and operates in the Nord Pool market. Its competitors are other power generation companies situated in the same part of Norway. Several Norwegian municipalities own the company.

4.2 Analysis of Case Study

PowGen started to develop AI solutions around three years ago and has since deployed many solutions successfully. Table 3 shows a subset of these AI applications. What is clear is that AI is applied in critical parts of their work. For example, all the energy that is produced by wind turbines owned or operated by PowGen is traded using AI algorithms. This is the longest ongoing project, and by several employees described as the most successful one. Wind is a variable power source. It is not possible to control how much energy the wind turbines will produce at any point in time. The amount of electricity produced is highly uncertain. However, wind energy has to be traded in advance, and thus it is important to make good forecasts which can be used when trading. This is at the core of the wind forecasting project. The AI models that make these forecasts are based primarily on weather forecasts and historical production. There are many aspects to take into consideration when using this data. For example, the winds in the wintertime are stronger. Also, weather forecasts themselves are uncertain, which makes it difficult to produce accurate wind production forecasts.

The wind energy is traded in several different markets in Nord Pool[3], using different forecasts for each market. First, the energy is traded in the *spot market*, where the prices for every delivery hour the next day are decided. The bids have to be sent before noon, including prices for every hour for the next day (from 00:00 to 24:00). This trading is based on day-ahead forecasting, meaning that the day ahead, the production for each hour the next day is predicted. Then, during the day of the production, energy is traded in the *intraday market (Elbas)*. The forecasts are delivered one hour ahead of the production hour. Here, the company has a reasonable estimate of how much energy will be produced. Thus it is possible to sell excess energy if needed. Lastly, there is the *regulation market*, where prices are decided after delivery. This market is used when there is a need to buy more energy, when the produced energy is less than the traded energy. The forecasts used in this market are short day forecasts, or in other words, replanning. Here, the price is set by the market, and it is unknown until after it is purchased. In many cases, this can be extremely expensive, resulting in massive losses for the company if they have misjudged their production capacity. Thus, having good quality forecasts is extremely important.

Table 3: Examples of AI Applications in Case

Application Type	**Description of Use**
Trading of wind energy	Algorithmic trading of wind energy in Nord Pool based on forecasts. The energy is traded in several different markets, using different forecasts: spot market, intraday market (Elbas), and regulation market
Planning for hydropower plants	Optimize how much water to have in the reservoirs at all times, and how much energy to make from the water resources.
Predictive maintenance of wind turbines	Get real time data from wind turbines, and predict their condition. Used to better plan maintenance of wind farms and turbines.
Day-ahead nomination of grid losses	Predict grid loss[4] for each hour the next day, for grids from different power grid companies. This is then nominated to Nord Pool.

[3] https://www.nordpoolgroup.com

[4] Grid loss is the difference in produced and sold electricity, i.e. how much electricity is lost in the power grid.

4.2.1 AI Adoption

The leadership of PowGen is described by several employees as very forward-looking. The top management saw that the energy market was changing and that data was becoming more critical. For example, changes in the energy market require organizations to use algorithmic trading in a few years. To stay alive, one has to be able to react fast. The energy system will get faster and faster, leaving it impossible for human workers to trade manually anymore. Respondent 1 state the following:

> *"If you were adaptable and dynamic, and able to use this as a competitive advantage, you would be one of those who ate other companies and not be eaten by those larger than you."*

It was then decided by the leadership that AI is the way to go to survive in the market. As mentioned by several employees, there was a big focus on digitalization and optimization in the organization. There are many innovation initiatives in the company, and adopting AI was considered a good choice for driving the organization further. The process started with two test projects with two different companies before they had a strategic discussion on how to proceed.

Instead of outsourcing the development, PowGen decided to develop the solutions internally. The reason for choosing to develop these solutions internally, rather than outsourcing, was because it was considered a competitive advantage doing so. By outsourcing, a large software company or consultancy firm comes in with the competency of developing software and machine learning, but they miss the data and domain knowledge that are needed to train the models. PowGen thus had to give away access to their data and domain knowledge to use an external company. The problem by doing this is that the models and solutions made based on those data and domain knowledge can be sold to their competitors as well. So, by pairing up with an external company, they give away quite a lot without getting that much in return. In addition, having intelligence that is different from their competitors can be an advantage in itself. Thus it was considered a competitive advantage to develop these AI solutions internally.

Instead of growing the AI department from the inside, PowGen contacted a person outside the company who was experienced with developing AI solutions. This person, now the chief AI officer, was then responsible for growing the AI department and developing the AI capabilities of the company. The chief AI officer had experience with several startups based on AI. Doing it in this way can have several advantages. For example, an experienced person will bring valuable knowledge about how things should be done and thus also have the authority to, possibly, make a larger impact. This can be an important factor in why PowGen managed to deploy two successful projects in less than a year after the AI department was formed.

4.2.2 AI Strategy

The goal of the company has always been to focus on automation and growth at the same time. When growing, the manual workers will have a much higher workload. Thus they need automation and decision support systems to ease parts of the workload. So, by growing and automating at the same time, all employees are able to keep their job. Some tasks will be automated away, but the employees responsible for those tasks are then put to do more meaningful work. As respondent 1 states:

> *"Now you can do the tasks that your competencies are needed for. We need you as a complete person, not someone calculating numbers of moving numbers from one place to another."*

In the beginning, the strategy was to do the easiest projects where AI could show some financial value. This meant starting with projects where they did not make anything new but rather improving and replacing already existing processes and systems. For example, respondent 2 describes how AI applications replaced previous solutions:

> *"When I started, they had finished one project before. And that was a prediction, short term prediction for wind, and that replaced previous predictive system, that was sort of a third party thing. So when I started, I started working on long term prediction for wind. And this, again, it took over for actually an excel sheet."*

Showing value from the start by choosing the easy wins first was important for the company to make employees feel confident in AI as the way forward. Now, however, as they have matured a bit, they have also started developing some completely new solutions, not being a replacement for previous systems.

4.2.3 Human-AI Relationship

All the models that PowGen has deployed up until now are part of an automated process. The human operators are not allowed to make changes or fill in for the models if they disagree with the predictions. For example, for the trading of wind energy, the trading is done directly with the forecasts given by the AI model. The reason behind this decision, the chief AI officer explains, is that if you let an operator make changes to the decisions made by a machine learning model, the performance will get much worse:

> *"So you have a model. It has an understanding of all the concepts that you're classifying or forecasting. And then the operators understand that, okay, I've classified my cats, but poorer at dogs. So often you have to make changes to dogs to compensate for this. But then if you retrain your model the concepts might change. So it might get better at dogs than cats. But still, the operator doesn't understand this. So it will still operate the model as if it's poorer at dogs as it was before. But because you've changed it, it's now good on dogs, and bad on cats, and this is really hard with having a human working together with the model and you're sort of updating the model."*

Even though human operators are not allowed to make changes to the model directly, they are still part of the process. They monitor the performance of the model and give feedback if they think it is not performing well enough. Thus the AI team needs to make models that work in a way that the operators understand. For this reason, the operators are involved in the development process.

4.2.4 AI Development

The AI solutions are developed with the help of a dedicated AI department within the company. This AI department help solving several operational challenges for all the different business units, using machine learning.

AI Department

The AI department consists of around ten employees, none of whom have a background in renewable energy. Their background is more focused on the aspects of AI, including data analysis, computer science, cybernetics, and industrial economics. They all have a Master's Degree, and several of the employees have a Ph.D.

Workflow

The AI solutions are developed through continuous collaboration between the AI department and domain experts. The domain experts are the ones that know the field and how things should work. In addition, the domain experts are the ones that work side by side with the AI solution and have to monitor its behavior. A close collaboration here is thus crucial to develop sustainable and successful solutions.

The human experts are the ones that posit the most knowledge on what is needed of the system. Thus it is essential for them to be included throughout the process, and especially in the beginning when writing requirements. During the development phase, the domain experts are constantly consulted. They give feedback to the AI department on how the system performs and behaves and if any changes should be made. After deployment, domain experts and AI developers still work closely to monitor the behavior of the solution.

There is no clear and structured methodology that all projects should follow. However, there exist guidelines for how things should be done. These guidelines are continuously updated as they gain more experience. Included in the guidelines are, for example, processes for quality assurance, such as code reviews, design reviews, and deployment reviews. Also, the best practices of several technical aspects are gathered in a wiki, working as a collective knowledge base.

All people from the AI department work full-stack, meaning that they all have a deep understanding of the whole technology stack. Also, in addition to the development, the AI department is responsible for the operations. This means operating and maintaining the machine learning systems, as well as developing them.

4.2.5 Challenges

PowGen managed to get AI solutions into production quickly and successfully, which is often not the case when adopting AI. After talking with several employees, it seems like PowGen has experienced very few bumps in the road. A few challenges were mentioned. However, they were quickly resolved.

Organizational Challenges

When introducing AI as a tool for automating processes, some employees started worrying that they might be automated away. To counteract this fear, the chief AI officer made sure to regularly explain why the domain experts are needed, and much effort was put into this before even starting up the AI department. As part of the AI strategy of the company was to scale and adopt AI at the same time, the employees who might lose large parts of their tasks due to automation got assigned new tasks immediately. Thus they were in no need to worry about being automated away, and they have ended up doing more creative and meaningful tasks than before. However, they have to supervision the new, automated process. A challenge mentioned was how to inform the human supervisors without overflowing them with information. This challenge was resolved by creating dashboards in collaboration with the human experts, showing values and information they were already familiar with.

The AI department consists of a group of highly educated people, with many of them having a Ph.D. They are focused on doing the right thing and making it perfect. However, this is not the objective when developing AI for PowGen. Making it "good enough" is the goal, without using too much time on making it perfect, as respondent 1 stated:

> *"[…] because you spent too much time on quality and you optimize too early in your code or in the systems, and often what is needed is something that runs."*

Making solutions "good enough" has been an area of focus for the company so that they spend their time doing the right things. Their focus is on setting up the system first, producing value immediately. Then it can be optimized later on.

Technical Challenges

Some technical challenges were mentioned by the interviewees. For example, how challenging it has been finding a middle ground between flexibility and the use of resources. On one side, there are ready-made platforms, such as managed cloud platforms for AI. These services deal with most things concerning the infrastructure, making it easy to deploy models. However, they are not flexible enough as they, for example, are set up for deploying individual models. PowGen usually deploys several models at once, making these platforms inappropriate for many of their use cases. On the other side, there is the option of making their own platform, which is a really flexible way of working. However, this requires resources which they do not have. Therefore, they have ended up using Kubernetes[5] for several of their projects because it gives them the flexibility they need without requiring too many resources.

Other technical problems were related to the retrieval of data. Data are retrieved from several sources, both internally and externally. Sometimes data can be retrieved late. At other times, the services where data is retrieved can be unavailable. This problem is hard to overcome, as they do not have 100% control over the data sources that are used.

4.2.6 AI Governance

Several practices and mechanisms concerning the governance of AI were employed (Table 4). Essential for the company is controlling the AI behavior so that it acts upon the organizational objectives. PowGen is partly owned by several municipalities in Norway and is thus very focused on being open and transparent.

[5] https://kubernetes.io/

Structural Practices

Decision-making

The decisions on which areas the AI department should focus their attention on are made by the chief AI officer, together with other leaders in the organization. However, the AI department themselves have the opportunity to decide internally precisely how the AI solutions will be implemented. Before a project is started, an analysis is performed to ensure that it is worth investing resources and time into the project.

Responsibility

There are several regulations to follow when trading, which can result in legal repercussions if not followed. When the trading happens automatically by an AI system, the question of who is responsible for its actions arises. This is a point that is still up for discussion in PowGen, according to respondent 5, who said:

> *"I think we are not completely finished by defining the process of when we do algo trading in physical or financial markets. Who is it? Who is responsible in the end?"*

For now, the potential economic losses for PowGen are limited. However, if AI-based trading is going to be used in the financial markets as well, more work has to be done in order to define who is responsible for the actions of the AI models.

Procedural Practices

Data

The AI solutions of PowGen are based on data from different sources. Sometimes this data is incorrect and needs to be corrected. Before, this was a manual process, where human workers had to look at the values to make sure they were okay. Now, however, this process is automatic. Errors in the incoming data are detected automatically, and an alert is sent to human workers if something needs their attention.

The use of personal and sensitive data is non-existing, which makes the process of governing the data less complicated. To ensure the quality of data, data cleansing is performed. Data transformation and feature engineering are also performed. What transformation is being done depends on the AI models. Some require extensive data transformation, while others do not. For example, much of the data they are dealing with are time-series data, where seasonality is vital to consider.

Model

Important for PowGen is to produce robust, rigid, reliable, and flexible AI systems. All of the employees put an emphasis on the importance of building robust systems and models. As respondent 2 states:

> *"[…], but a lot of my time is spent on the deployment process and robustness of the systems. The AI part ends up being sort of rather minor, […]."*

Robustness is a focus for the company, and they have put in place several processes to make sure of this. There are processes in place for quality assurance of the model. For example, to make sure that the model behaves as wanted and that it is robust enough, it is put in "stage" for a few days or weeks. To put it in "stage" means that the model runs as if in production, using real-world data. The reason for this is that real-world data can behave quite different from the training data, as respondent 1 states:

> *"We know that we have real world data and they behave very differently from training data. In the real world, some data is delayed, or it might not come at all, which means that you still have to make a decision, you still have to make a forecast without this data, and it might arrive too late."*

The results from the model when it is in "stage" are not being used, but they are sent to human experts, which can give feedback. If they find values that are not correct, an investigation can be performed to find the reason and fix it before the model is put in production. Working with real-time data can be difficult. Thus PowGen has found that this is a good approach to test if there are some cases that the model does not fully cover before the model is being used. Only when both the human experts and the development team have complete faith in the model is it put in production.

When deciding on which AI models to use, PowGen prefers using simple models. For example, decision trees and CatBoost[6]. The reason for this, they explain, is partly because they do not have big data to feed the deep learning algorithms and because the simple white-box models are the ones that are easiest to explain, as respondent 3 states:

> *"So we are big proponents of having simple systems, let the simple things work because they are easy to explain."*

[6] https://catboost.ai/

In addition, PowGen has found out that using more complex models does not necessarily enhance the performance of their AI systems. Their challenges are often related to the data, e.g., noisy data, and using a more complex model have proven not much better than the simple ones in overcoming these issues.

There are several legal issues PowGen has to address when developing AI models. When predicting, they are constrained to do the best forecast they can. This restriction does not mean the forecast that generates the highest value but the objectively best forecast. It is illegal to do otherwise, and the government checks up on this. In other words, it is illegal to tune the model to make the most profit, e.g., by inserting a certain bias in the model. This restriction is extensively followed by PowGen, who was called "more catholic than the Pope" by external auditors. Other regulations can include restrictions that should be included in the model. E.g., for hydropower plants, there are regulations on how much water can be used from the water sources and how much water the water sources should have at all times. These restrictions should be included in the model so that it does not behave illegally.

Another area of focus is to ensure reproducibility. In other words, making sure that the experiments they perform can be reproduced by anyone else in the AI department. This means that the code should not only run on one person's computer but that it can run by itself, e.g., in the cloud.

Monitoring

To make sure that the AI solutions behave properly, it is monitored by human controllers. If a model encounters a situation that has never happened before, it will probably do something unexpected. Thus it is crucial to have a human employee overseeing its actions to make sure that it acts according to laws and organizational objectives. This is mainly done through different dashboards, which display different metrics related to the behavior of the AI system. An analytics and monitoring solution called Grafana[7] is used for this. For example, on a daily basis, the submitted values are plotted against the actual values to show how far the prediction is from the actual value.

The dashboards are targeted at different groups of people. For example, one can be targeted at the domain experts and another one for the AI department. The dashboard created for domain experts contains different KPIs, plots, and graphs that they are already familiar with looking at.

[7] https://grafana.com/

For example, the amount of money made this year and accumulated profit over the last month. In other words, how the system as a whole performs. This way of doing it makes it easy for them to monitor without putting too much effort into it. Moreover, if they see that something goes wrong or is not behaving as wanted, feedback can be sent to the AI department that can look into the issue.

Dashboards targeted at the AI team, however, contain more technical measures. For example, which model performs best. In other words, everything that is needed from a machine learning point of view, about the performance of individual models rather than the performance of the system as a whole.

Regular reports of the models' performance are generated to higher-level managers. The report shows how the system has performed lately. From this, it can be decided if changes need to be done to the model if it does not perform according to the organizational objectives.

If a model does not act according to laws and regulations, there will be severe repercussions. Even though these requirements are put into the logic of the model, wrongs can still happen (e.g., because of the input data). Human controllers are thus needed to ensure that no rules are violated. As respondent 3 explains about the hydropower regulations:

> *"We do try to put all these restrictions into the models, but sometimes they don't work because of some input or something. So we need a human intervention […] These values don't make sense, and if this happens, our boss is going to jail."*

It is illegal to trade on internal knowledge. Thus, it is vital to ensure that the trading algorithm does not use data that is not allowed to use. Respondent 1 said:

> *"If one of your parks, or your hydro plants cannot produce power, it might affect the market. […] And if you trade on this information, before you sort of tell it to others, then you've done something illegal."*

The problem with this is that an algorithmic trader does not understand by itself what data are legal to use or not. The algorithmic trader will trade using the data it has available. Thus it is essential to have processes in place for handling the case of insider information. In PowGen this is solved by having a "big, red button" that stops the algorithmic trader from trading. When they realize they have inside information, the button is pushed, and algorithmic trading is stopped until the information is released to the rest of the market.

Table 4: AI Governance Mechanisms in Case

<table>
<tr><th>Structural Practices</th><th>Procedural Practices</th><th>Relational Practices</th></tr>
<tr><td>Responsibility
Chief AI Officer
Assess value and risk before doing a project</td><td>Data
Automatic error detection on incoming data
No use of sensitive data
Data cleansing and transformation</td><td>Between Departments
Talk about why domain experts are needed
Educate staff in other departments about AI
Feedback loop
Give continuous reports
Cross-functional teams
Shared vocabulary
Workshops</td></tr>
<tr><td></td><td>Model
Include laws and restrictions in the model
Ensure robust models, e.g., by put it in “stage” to test the behavior
Simple models are easiest to explain
Reproducibility</td><td>Inside AI Department
Knowledge transfer through good documentation</td></tr>
<tr><td></td><td>Monitoring
Different types of dashboards, targeted at domain experts and AI department
Human controller
Performance reports</td><td></td></tr>
<tr><td></td><td>Issue Management
Alert if something needs your attention
Task force to assess what went wrong when failure
Lesson learned report
Backup predictions</td><td></td></tr>
</table>

Issue Management

For all systems in operation, a monitoring system is in place, which is responsible for sending alerts when there are issues that need human workers' attention. The alert is either sent as an email or a message in an appropriate channel to those responsible for maintaining the project. There are checks in place at different stages of the system, which evaluate the output values from the AI systems and send out alerts if they do not look as they should. The alerts are related to different metrics, such as the performance of the system or if the system did not receive any data. Since there are deadlines for sending bids to the market, the alerts are set up to notify the appropriate people much in advance so that they have time to react and make the necessary changes.

When more significant incidents happen, there is an evaluation process to ensure that the same thing will not happen again. A task force is created involving several stakeholders in the organization. The system developers meet with some stakeholders from the rest of the organization, as well as the head of the HR department. Together they try to find out what happened and its consequences. To ensure that the same incident does not happen again, they discuss what they have learned from the incident and if any routines should be changed. The task force concludes with a lesson learned report.

In case of issues with the predictions, backup predictions can be applied. These backup predictions will be of lower quality than the original predictions but are necessary to be able to provide predictions at all times. These backup predictions are especially relevant when there are problems with the cloud-based solution that the original predictions run at. In that case, the backup predictions can run locally, avoiding the problem.

Relational Practices

Between Departments

An important area of focus has been to develop trust and get employees in other departments to feel good about AI. Building employee trust is essential because those employees have to deal with the AI application in their day-to-day work. In the beginning, people were afraid to get automated away. Therefore, the Chief AI Officer used much time to talk about why domain experts are needed to make good AI solutions. There seem to be only minor problems with getting people to trust AI solutions. A reason for this can be that the company has an innovative culture, with experts willing to learn and use new methods and tools. As described by respondent 3:

> *"But I think we're very lucky that the group of experts we have, they are very open to learning new things."*

Another significant reason can be that the AI department has invested much time into making the other employees understand what the AI department is doing. This process includes explaining what AI is and its possibilities and limitations. There are also several other processes and mechanisms in place to include the other departments and employees in the AI journey. Continuous reports are delivered, and there is a feedback loop. Nevertheless, most importantly, the domain experts work together with the AI team in developing AI solutions. Doing it in this way lets them get a better grasp of what AI is. The development of the different AI solutions was performed using cross-functional teams, integrating different parts of the organization. Business developers, AI developers, and domain experts work closely together in developing AI solutions. This way of developing solutions is vital for making the collaboration between the operators and the AI solution as smooth as possible. For example, by defining problematic scenarios, and when the operators would like to get alerts and notifications, it is easier for them to interact with the AI system.

None of the employees in the AI department have worked with power markets before. The terminology used by AI developers when discussing problems and possible solutions is thus very different from the domain experts'. When AI developers and domain experts collaborate closely, coming up with a shared vocabulary has been important to communicate efficiently. As respondent 3 states:

> *"We see things as xy predictions [...] But since we always need experts in the loop as well, we need to find out: What does it mean? What does this prediction mean?"*

In addition, the AI department has put in the effort to make the domain experts able to run the models that they deliver to them. Several workshops have been arranged for the domain experts to be able to utilize the same tools as the AI team. Examples of workshops include Python, Jupyter Notebook, good habits of coding, and an introduction to Github. These efforts make it easier for everyone to discuss using the same vocabulary and make sure they are all on the same page.

Inside AI Department

To not rely too much on one person's competence, it is essential to make it easy for someone to take over the code written by another person. This is done through documentation of the model development process. The goal is to make it easy and not too time-consuming to understand what is being done and why it is being done.

4.2.7 Organizational Change

After introducing AI, and a dedicated AI department, PowGen has experienced several changes happening at the process level.

Change in Culture

AI brings with it an increased focus on data, as data is the fundamental building block of every AI application. After adopting AI, there has been a change in mindset throughout the whole organization to make use of all the data resources that exist, not wasting opportunities. This change has resulted in other departments now initializing new AI projects as they see opportunities in the data they possess.

Before, there was minimal programming in the company. Now, however, other departments are also developing software. There has been a shift in culture, from doing things manually to being able to automate tasks through programming. This change is due to the creation of the AI department and the good developer practices they brought with them. The AI team has been good at including the other departments, arranging workshops, among others, to make them inherit some of the good coding practices. These efforts have resulted in several departments developing and deploying systems to production.

Shorter Time to Production

By having an internal AI department, the threshold for actually doing a project has lowered. It is easier to try out new ideas without investing too many resources, time and money. If using an external provider, it is easy to abandon project ideas that are seen as too risky due to the possible price that needs to be paid. In addition, the time from the idea arises, and until the system is out in production, has shortened. This change can be due to the fact that it is easier to communicate within the company than with an external provider.

Also, the AI department has had a constant focus on building a robust, reliable, and flexible infrastructure. This infrastructure makes it easy to deploy new models quickly and safely. An infrastructure for dashboards and alerts is already in place, making it very easy to deploy new models robustly, rather than starting from scratch every time.

New Business Models

One of the most significant changes caused by introducing an AI department is that the company now also has become a service provider. They offer AI services to other companies. This change is a result of the promising results their AI systems have shown. Since they can outperform what other companies could provide them with, they saw an opportunity to expand their repertoire. This effect surprised several of the employees, who did not anticipate this change before joining the company.

Also, AI allows for scaling. This ability to scale things means that more assets can be optimized, possibly creating new services. In other words, AI creates possibilities for new business models.

More Meaningful Work

Introducing AI has led to changes in the way several employees work. Some employees have experienced a change in what tasks to perform, as many tasks now are automated. The tasks that are automated away are usually repetitive and tedious tasks, which means that the employees now can use more time on creative and more meaningful tasks.

4.2.8 Business Value

The AI department has brought value to the company in several ways, especially related to operational and financial performance.

Operational Performance

The AI solutions developed by the AI department are outperforming the old solutions in terms of accuracy. Also, the predictions are better than what they would have achieved with a ready-made AI solution. A reason for this can be that the AI systems are built to solve *their* problems, meaning that it is optimized primarily for them. The systems are developed in close collaboration with domain experts, helping the AI developers evaluate the performance and making improvements, making a system optimized for PowGen’s needs.

In addition to better predictions, introducing AI has brought robustness and reliability. This is mainly because of the AI department’s constant focus on providing robust solutions. For example, by putting the model in “stage” to test its behavior on real-world data or by introducing alerts and notifications when systems do not work.

Financial Performance

The implemented AI solutions do not make a very high earning today. Rather than focusing on financial gains, the focus of the company has been to set up for future requirements. There is some financial value from the implemented AI solutions, but it is limited. However, there has been a significant reduction in costs. Better predictions result in less penalty from the market, which, in turn, translates to money saved. As respondent 1 states:

> *"So it reduced our balancing costs with 17%, [...]. But then when you look at the numbers for how much it's earned, it's very little, you cannot say that the time we spent was well spent because of this."*

5 Towards an Instrument of Responsible AI Governance

Adopting AI can make a significant impact on an organization. It can end up changing the way people are doing their job and possibly leave people unemployed (Ford, 2013; Makridakis, 2017). Also, it can make a notable financial impact on the organization (PwC, 2018). For instance, the company studied in the case study uses AI to predict wind production and trade energy based on these forecasts. If the models are completely off, it can have a significant negative financial impact on the organization. Also, it can possibly have legal repercussions if the behavior of the AI models is not monitored well enough. Furthermore, deploying AI applications brings the risk of acting biased and discriminatory (Ntoutsi et al., 2020). These examples are only a few of the risks that AI brings about.

Developing and deploying AI applications with no form of control or supervision can be harmful in several ways, as exemplified above. Therefore, organizations want to apply responsible AI governance to reduce this risk and maximize the benefits from its AI applications. However, there is a question of what precisely responsible AI governance entails for organizations. In this section, the notion of responsible AI governance is conceptualized, and a new definition of the term is provided. Also, I propose a set of principles that organizations should govern their AI according to. Building on the definition of responsible AI governance and the proposed principles, the responsible AI governance instrument is defined.

5.1 Conceptualization

Organizations want to conduct ethical business and maintain their corporate social responsibility, meaning they want to be socially accountable and conscious of their impact (Jones, 1980). AI brings potential risks that challenge these intentions, such as the risk of AI applications acting biased (Ntoutsi et al., 2020). Thus there is a need to talk about how organizations can control and govern their AI applications to act responsibly.

Several efforts have been made to describe how AI should behave ethically, trustworthy, and responsibly (Dignum, 2019; European Commission, 2019; Thiebes et al., 2020). However, these points are raised as principles and not frameworks for implementation within an organization. Putting the principles into use and ensuring that they are met can be done through deploying governance mechanisms. Governance deals with the structures and processes used to ensure that organizations meet their strategy and objective (De Haes & Van Grembergen, 2004). These mechanisms are not thoroughly discussed in the AI literature.

The existing work on AI governance is mainly discussed from a societal and regulatory perspective (Butcher & Beridze, 2019). For instance, governments in several countries have released national strategies to address AI's concerns (Norwegian Ministry of Local Government and Modernisation, 2020). Additionally, several governmental bodies are discussing how to impose restrictions to ensure the ethical development and use of AI.

While there are several studies concerning responsible governance of AI applications on a societal and regulatory level (Buiten, 2019; Erdélyi & Goldsmith, 2018), the work on responsible AI governance targeted at organizations is limited. Some work on AI governance targeted at organizations is put forward, such as the *Model artificial intelligence governance framework* published by the government in Singapore (Government of Singapore, 2020). Moreover, some consultancy firms have proposed guidelines for developing and controlling AI (Accenture, 2019a; KPMG, 2019). Others discuss governance of AI within an organization from a technical point of view, such as researchers at Microsoft (Amershi et al., 2019), which focuses on the best practices of software engineering of AI projects. However, there is still no established governance framework for organizations to follow when developing and deploying AI to ensure that it acts ethically, responsibly, and trustworthy.

Thus there is a need to discuss what responsible AI governance for business entails so that organizations can fulfill their corporate social responsibilities, build trust in their AI applications, and gain value from AI while minimizing its potential risks. However, a unanimous and clear definition of responsible AI governance in a business context is lacking. Thus, I present the following definition:

"Responsible AI governance is the collection of rules, practices and processes used by organizations to ensure that the development and use of AI is ethical, transparent, accountable, and complies with laws and regulations."

This definition builds on the definition of "corporate governance", which refers to the system used to direct and control companies (Cadbury, 1992). Additionally, it builds on the established notion of "IT governance", which provides the link between business and IT (De Haes & Van Grembergen, 2004). Furthermore, it draws on the notion of "responsible AI". As Dignum

(2019) states, responsible AI is "about being responsible for the power that AI brings", and it refers to AI that is consistent with the users' expectations and developed with societal laws and norms in mind (Accenture, 2018).

The notion of responsible AI governance concerns how and through which mechanisms organizations control and govern their AI capabilities to act responsibly. These mechanisms go beyond what is stated in laws and regulations and include taking ethical and socio-technical implications into account when developing and deploying AI applications. Responsible AI governance includes creating internal rules, regulations, and processes and applying and following these consistently. Through employing these mechanisms, the goal is to build robust AI applications that act according to the organizational strategies and objectives, and societal values, moral and ethical considerations (Adadi & Berrada, 2018). The goal of responsible AI governance is to maximize the wealth of all the involved stakeholders while at the same time finding a balance between the interest of the different stakeholders. All parties affected (both directly and indirectly) by the AI solution should be considered stakeholders, such as employees, customers, investors, and members of the public affected by its use. Finding a balance between the differing needs of the stakeholders can be difficult, but it is vital not to neglect any stakeholders in the process.

5.2 Dimensionalization

Here, I propose eight principles, consisting of 13 sub-dimensions, that organizations should govern their AI according to in order to have responsible AI (Table 5). These principles have emerged through a process involving my supervisor, a doctoral researcher, and me. First, a systematic review of the existing frameworks and literature on AI governance were performed, such as the AI governance framework (Government of Singapore, 2020) published by the government in Singapore. Included in the review were also literature on ethical, trustworthy, and responsible AI, such as the *Ethics guidelines for trustworthy AI* (European Commission, 2019) published by the European Commission. From the literature, several themes and principles emerged. These principles were then discussed by the group, together with the findings from the case study. To be included, the dimension had to be distinct enough. The final dimensions were decided on through group consensus.

These principles will be of varying importance to companies depending on factors, such as the type of AI solution they are deploying and which sector or industry they are operating in.

Table 5: Dimensions of Responsible AI Governance

Dimension	Description	Sub-dimensions
Transparency	Organizations should be open and transparent with all its stakeholders regarding elements relevant to the AI solution, such as the use of data and the business model (European Commission, 2019).	• Explainability • Traceability • Communication
Fairness	AI systems should enable inclusion and diversity, and not lead to discriminatory outcomes (European Commission, 2019; Government of Singapore, 2020; KPMG, 2021).	• No unfair bias • Accessibility
Accountability	AI actors should be responsible and accountable for the proper functioning of AI systems, and for the respect of AI ethics and principles (Government of Singapore, 2020).	
Robustness and Safety	AI systems should reliably behave as intended without harming the surroundings (European Commission, 2019).	• Resilience and security • Reliability and reproducibility • Accuracy
Data Governance	The process of managing data throughout its useful economic life cycle (Tallon et al., 2013).	• Data privacy • Data quality • Data access
Laws and Regulations	AI should comply with all applicable laws and regulations through the entire life cycle of AI projects (European Commission, 2019).	
Human-centric AI	AI should generate tangible benefits for people, and always stay under human control (Benjamins et al., 2019).	• Human oversight • Human well-being
Environmental and Societal Well-being	AI should be at the service of society and environment, and not cause them any harm (European Commission, 2019).	

5.2.1 Transparency

Organizations should be open and transparent to all their stakeholders about the fact that they are using AI technologies in their products and services. People that interact directly with an AI system should be aware that this is the case. This means being transparent about the data, the system, and the business models used for the AI-powered products and services (European Commission, 2019) and communicate this to the users. Users should be aware of what kind of data, whether personal or impersonal, that is used by the AI system and the purpose of using such data (Benjamins et al., 2019). Organizations should also provide traceability, meaning that the decisions, datasets, and processes that yield the AI's decision are documented in an easy-to-understand manner (Government of Singapore, 2020). In that way, the reasoning of AI can be better understood.

AI systems are often used to assist with or make decisions. In such cases, it is important that the users of the system are provided with an explanation of why the AI reached the conclusion it did (Licht & Licht, 2020). This requires organizations to provide explainability of the technical process of the AI system (European Commission, 2019). When deciding on the level of explainability, there is a trade-off between explainability and the complexity and accuracy of the models (Benjamins et al., 2019; European Commission, 2019; Loyola-González, 2019). More complex models can often provide more accurate solutions, but this comes at the cost of less transparency and explainability. Simple white-box models, such as decision trees, can generate explanations that are easy for domain experts to understand. On the contrary, the more complex black-box models, such as deep learning techniques, are difficult for experts to understand, as they are not able to provide human-understandable explanations for their decisions (Benjamins et al., 2019; Loyola-González, 2019). How much explainability is needed depends on the context it is used in. For instance, AI solutions in health care will have higher requirements for explainability than, for example, a music recommendation engine. Furthermore, the European Union has introduced initiatives to make AI more explainable through the introduction of GDPR, providing users with the right to know how decisions are made and what data it is based on (Goodman & Flaxman, 2017). The field of Explainable Artificial Intelligence (XAI) is concerned with making the results and behavior of AI systems more understandable to humans, thus enabling more transparent AI applications (Arrieta et al., 2020; Došilović et al., 2018).

5.2.2 Fairness

One of the more discussed topics regarding AI is its ability to provide fair results. The use of AI should enable inclusion and diversity and not lead to discriminatory impacts on people based on their personal conditions, such as race, gender, religion, sexual orientation, and disability (Benjamins et al., 2019; Binns, 2018; European Commission, 2019). However, many of the AI solutions that are reported in the news and social media have led to discriminatory outcomes. A fundamental component of AI systems is data. AI learns to make decisions based on data, which means that the results can be biased or discriminatory if the underlying data is (Mehrabi et al., 2019). Organizations should be aware of these issues and put in place processes to make sure that the outcomes of their AI systems are fair.

AI systems should be user-centric and designed so that all people can use the system (European Commission, 2019). This means ensuring that the AI system accommodates a wide range of abilities and preferences. E.g., users of assistive technologies, such as voice commands, should also be able to interact with AI products and services.

5.2.3 Accountability

Companies that use AI should be responsible and accountable for the outcomes of their AI systems (Government of Singapore, 2020; Martin, 2019). This means making sure that the AI systems respect AI ethics and principles, and in the cases where there are violations, the company should have policies and processes in place to deal with it, e.g., redress in case of negative impact. This applies both before, during, and after the development, deployment, and use. In addition, each individual person involved in creating the AI system is accountable for considering the impact of the system on the world (IBM, 2019).

An important part of achieving trustworthy and responsible AI is the auditability of the AI system (Raji et al., 2020). Auditability refers to an AI system that is ready to undergo an assessment of its algorithms, data, and design processes (European Commission, 2019; Government of Singapore, 2020) and is enabled by the traceability of the AI system. By being auditable, third parties can review and understand why the AI behaves as it does. Having both internal and external auditors to evaluate the system is vital to increase the trustworthiness of the AI, and it is especially important in cases that can have a significant impact on human life and society.

5.2.4 Robustness and Safety

AI systems should be developed in a way that promotes robustness and safety. They should reliably behave as intended while at the same time minimizing and preventing harm (European Commission, 2019). This requires the systems to be technically robust so that they are resilient and secure and not vulnerable to tampering or compromising of the data they are trained on (Government of Singapore, 2020; Hamon et al., 2020). In addition to being reliable, the system should also be reproducible, meaning that it exhibits the same behavior when repeated under the same conditions (European Commission, 2019).

The potential safety risks of the AI system should also be addressed (Yampolskiy & Spellchecker, 2016). Its potential negative impacts should be identified and minimized. This means, among others, knowing how the model behaves in unexpected situations and environments (European Commission, 2019). To ensure its behavior in these situations, an appropriate fallback plan should be in place. A fallback plan should also be ensured in the case of attacks.

5.2.5 Data Governance

At the core of AI is data. AI learns to make decisions based on data. Thus it is essential to manage those data assets in a responsible manner. Data governance is about the policies and procedures to effectively manage an organization's data assets (Tallon, 2013). It touches upon areas such as the quality and integrity of the data used, who has access to what data, and the relevance of the data with regard to the domain it will be deployed in (European Commission, 2019). Privacy and data protection are also a part of data governance practices. Organizations must respect people's personal data and their right to privacy (Benjamins et al., 2019). The concept of data governance is well established, and many organizations already have several of these procedures in place. However, it is essential to mention in the light of AI, as data is the fundamental building block of most AI systems (Janssen et al., 2020).

5.2.6 Laws and Regulations

AI systems should follow the required laws and regulations. There are a number of rules at the national, regional, and international levels that are relevant for organizations that utilize AI technologies. Examples of this are regulations such as the General Data Protection Regulation (GDPR), which regulates activities concerning the processing of personal data in the EU and EEA and can cause issues for organizations wanting to deploy AI solutions that are trained using personal data (Goodman & Flaxman, 2017). Other regulations can be industry-specific and affect the way companies in that industry can operate. For example, how much water should be in a water reservoir at all times, as mentioned in the case study.

5.2.7 Human-Centric AI

When developing AI systems, the main priority should be humans' well-being. AI systems should be human-centric, meaning that they should be at the service of humanity and society, with the goal of improving welfare and freedom (Benjamins et al., 2019; European Commission, 2019). This implies that AI systems should respect human autonomy and do not cause harm or affect human beings in a negative way (European Commission, 2019). To make sure of this, organizations should assess to what extent human oversight is needed. The level of autonomy range from humans actively involved to no human intervention at all. The degree to which humans are involved should be assessed in light of the probability and severity of harm (Government of Singapore, 2020). Applications with high severity and probability of harm should not operate alone, without human oversight or involvement. The less oversight and involvement of humans in AI applications, the more extensive testing and stricter governance, are needed (European Commission, 2019).

5.2.8 Environmental and Societal Well-Being

AI systems should not cause any harm to the environment and broader society (European Commission, 2019). Data centers that are used to store data and process algorithms often use much energy. According to Strubell et al. (2019), training a large AI model can produce over five times the amount of CO2 emission produced by a car over its lifetime. On the other side, AI solutions can be used to fight climate change, such as tools to monitor pollution. The environmental footprint caused by AI is thus essential to take into consideration when developing AI solutions. Organizations should assess what impact their AI applications have on the environment and evaluate methods to make it more environmentally friendly.

Applications for AI technologies are found in almost all aspects of human life (e.g., education, work, health, finance, entertainment). A chatbot can be the point of contact when contacting customer support at a bank or an online store. Decisions at work are made by algorithms. When listening to music, recommendations are suggested by an intelligent engine. Many of these systems can enhance the quality of life for humans, but there is also the possibility that it can degrade humans' physical and mental wellbeing. The effects that these systems have on human life and democracy should be carefully evaluated and monitored. Also, its impact on democracy, institutions, and society at large should be taken into account (European

Commission, 2019). AI can, for example, be used in elections. To protect the democratic process, it is important to carefully assess the impact of AI when used in situations that concern politics and elections.

5.3 Measurement of Constructs

The AI governance construct is conceptualized as a multidimensional third-order formative construct comprised of eight dimensions. Three of these dimensions are first-order constructs, while the remaining dimensions are second-order constructs comprising in total 13 first-order constructs. The items used to capture the first-order constructs are presented in Table 6.

These items were created by reviewing the literature on AI governance, and trustworthy, ethical, and responsible AI once more and see if they included any explicit rules, processes, or procedures for organizations to implement. For example, the High-Level Expert Group on AI has created an assessment list for trustworthy AI (European Commission, 2020b). All relevant questions or statements found in the literature were put together in a spreadsheet. Then, a group consisting of my supervisor, a doctoral researcher, and me, went through all questions and grouped them based on the dimensions and sub-dimensions presented in the previous section. In addition, all questions were turned into statements so they could be measured on a 7-point Likert scale (Joshi et al., 2015). After grouping the statements together, several rounds of iterations were performed to refine the set of items. Each person in the group went through the statements independently and provided suggestions for improvements on the formulations. These suggestions were then discussed in the group and decided on through group consensus. Also, an elimination round was performed, removing statements that overlapped, or were considered redundant or not relevant enough. The remaining statements formed the set of items used to measure the constructs. The nomological validity of the responsible AI governance construct is checked by placing it in a research model presented in the next chapter (Chapter 6).

Table 6: Constructs and Measures of Responsible AI Governance

Dimension	Construct	Items
Transparency	Explainability	EX1: When designing and building an AI model, interpretability and explainability are in favor over accuracy
		EX2: We assess to what extent the decisions and hence the outcome made by the AI system can be understood
		EX3: We design AI applications with explainability and interpretability in mind from the start

	Traceability	TR1: Processes and mechanisms for data collection, data labelling, data transformation and data use (i.e., how the data are used inside the company) are well documented
		TR2: Processes and mechanisms for AI development are well documented
	Communication	CO1: We communicate to users that they are interacting with an AI system and not with another human
		CO2: Users can provide feedback for their AI experience
		CO3: Characteristics, limitations and potential shortcomings of the AI system have been identified and informed to the end-users
Fairness	No unfair bias	BI1: The datasets we use for AI applications are assessed in terms of diversity and representativeness of the population
		BI2: Responsible data tagging is used to minimize bias in AI applications
		BI3: We have put in place processes to test and monitor for potential biases during the development, deployment and use phase of the system
	Accessibility	ACCE1: We have involved and consulted different stakeholders (e.g., users of assistive technologies) in the AI system's development and use
		ACCE2: We have ensured that our AI applications accommodate a wide range of individual preferences and abilities
		ACCE3: We have ensured that the information about the AI system is accessible also to users of assistive technologies
Accountability		ACCOU1: We have established an "ethical AI review board" or similar mechanisms to discuss overall accountability and ethics practices, including potential grey areas
		ACCOU2: We made company policies clear and accessible to design and development teams so that no one is confused about issues of responsibility or accountability of AI
		ACCOU3: We established an adequate set of mechanisms that allows for redress in case of the occurrence of any harm or adverse impact from our AI applications

		ACCOU4: We have established processes that facilitate the assessment of algorithms, data and design processes
		ACCOU5: We have established mechanisms that facilitate the system's auditability, such as logging of the AI system's processes and outcomes
Robustness and Safety	Resilience and Security	RS1: We have assessed potential forms of attacks to which the AI system could be vulnerable (e.g., data pollution, physical infrastructure, cyber-attacks)
		RS2: We have measures or systems in place to ensure the integrity and resilience of the AI system against potential attacks
	Reliability and Reproducibility	RR1: We have put in place verification methods to measure and ensure different aspects of the system's reliability
		RR2: We have put in place verification methods to measure and ensure different aspects of the system's reproducibility
		RR3: We have processes in place for describing when an AI system fails in certain types of settings
		RR4: We have tested whether specific contexts or particular conditions need to be taken into account to ensure reproducibility
	Accuracy	AC1: We measure if our AI applications are making unacceptable amount of inaccurate predictions
		AC2: We have processes in place to increase the AI applications' accuracy
Data Governance	Data Privacy	DP1: We have established mechanisms allowing users to flag issues related to privacy or data protection in the AI system's processes of data collection and data processing
		DP2: We have considered ways to develop the AI system or train the model without or with minimal use of potentially sensitive or personal data
		DP3: We have taken measures to enhance privacy (e.g., encryption, anonymization, aggregation)
		DP4: We have ensured that our products and services that uses anonymized data poses no unreasonable risk of re-identification

	Data Quality	DQ1: We have put in place processes to ensure the quality and integrity of our data
		DQ2: We do periodic reviewing and updating of our AI datasets
		DQ3: We have assessed the extent to which we are in control of the quality of the external data sources used
	Data Access	DA1: We ensure that people who access data are qualified and required to access data, and that they have the necessary competence to understand the details of data protection policy
		DA2: We have an oversight mechanism to log when, where, how, by whom and for what purpose data was accessed
Laws and Regulations		LR1: We comply with law, regulations and guidelines that our AI have to work within
		LR2: We understand national and international laws, regulations and guidelines that our AI have to comply with
		LR3: We have established processes to ensure that our AI applications are in alignment with the latest laws and regulations
Human-centric AI	Human Oversight	HO1: We have safeguards to prevent overconfidence in or overreliance on the AI applications
		HO2: We have considered the appropriate level of human control for the particular AI system and use case
		HO3: We have put in place mechanisms and measures to ensure human control or oversight over the AI
	Human Well-being	HWB1: We have assessed whether there is a probable chance that the AI system may cause damage or harm to users or third parties
		HWB2: We have assessed the possible negative impacts of our AI products and services on human rights
		HWB3: We ensure that an AI system does not undermine human autonomy or causes other adverse effects
Environmental and Societal Well-Being		EWB1: We have established mechanisms to measure and reduce the environmental impact of the AI system's development, deployment and use

		EWB2: We monitor and consider the effects of our AI applications on the environment
		EWB3: We have ensured that the social impacts of the AI system are well understood

6 Research Model

Figure 3 presents the research model and hypotheses proposed in this thesis. I propose that responsible AI governance will have significant effects on the competitive performance of a company, mediated by two main paths. I argue that by having responsible AI governance, firms enhances its knowledge management capability (KMC) and the organizational agility. Which, in turn, will influence the competitive performance of the company. The constructs presented in the research model are defined in Table 7.

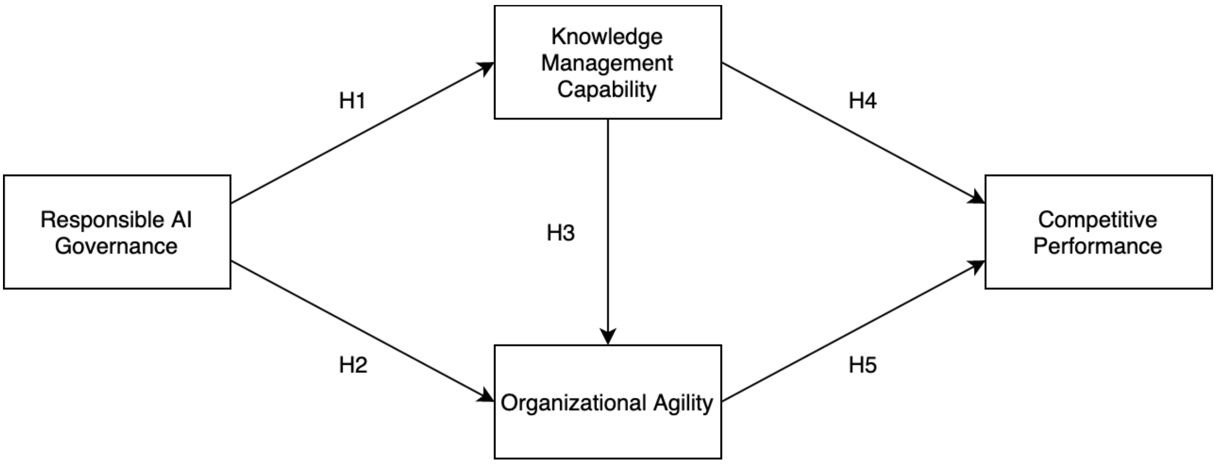

Figure 3: Research Model

6.1 Hypothesis 1

Responsible AI governance refers to policies and procedures used to govern and control AI applications so that it behaves according to a set of principles. Organizations want to ensure that their AI applications act according to the principles and are thus focused on gaining knowledge on all aspects related to the AI life cycle, such as knowing how the solution behaves when in production. This knowledge can then be shared and used to enhance the quality of existing products, services, or processes, and to make sure that the AI application acts according to the principles. For example, when finding out that an AI application acts biased, this knowledge can be used to change existing processes related to AI development to make sure that the application acts fair.

With responsible AI governance in place, organizations are focused on assessing, monitoring, and evaluating the behavior of the AI application, both before and after deployment. These processes facilitate the creation of knowledge about the solutions' behavior. For example, in the case study (Chapter 4), several dashboards are used to monitor the different AI solutions. This way of monitoring the AI models is a way of gaining knowledge, which can be used to enhance the system's performance, both in terms of accuracy and impact on the surroundings and stakeholders. Also, using dashboards is a way of effectively communicating and spreading information about the solutions' behavior throughout the organization, making it accessible to all those who need it. For example, in the case study, several dashboards targeted at different employees and departments are created to spread information about the system's behavior.

Table 7: Constructs and Definitions

Construct	Definition	Source(s)
Responsible AI Governance	Responsible AI governance is the collection of rules, practices and processes used by organizations to ensure that the development and use of AI is ethical, transparent, accountable, and complies with laws and regulations	Self-developed
Knowledge Management Capability (KMC)	The process-based ability of the organization to mobilize and deploy knowledge-based resources to gain competitive advantage	Mao et al. (2016)
Organizational Agility	The ability to detect and respond to opportunities and threats with ease, speed, and dexterity	Tallon and Pinsonneault (2011)
Competitive Performance	The degree to which a firm attains its objective in relation to its main competitors	Rai and Tang (2010)

Explainability of AI models is an area of focus when governing AI applications in a responsible way. Explainability means that the reasoning of the AI models' decisions can be understood. By understanding the reasoning of the AI models, knowledge about the system can be acquired. This knowledge can, in turn, be used to justify decisions, identify and correct errors, and improve the quality of the model, among others (Adadi & Berrada, 2018). For example, in case of discriminatory or biased outcomes, explainable models make it easier to understand why the system acted in that way, which can be used to either justify the results or make changes to make sure that it does not happen again.

Responsible AI governance also involves having clear and concise ways of documenting processes and mechanisms for all aspects of the AI life cycle, including both data and AI development aspects. For example, all steps from data collection to data use should be documented, such as how the data is transformed. By having good documentation of the processes and mechanisms, knowledge can more easily flow within the organization, thus enhancing the KMC. Also, documentation makes the organization less dependent on one person's knowledge, and capabilities as the processes are documented. From the preceding discussion, I hypothesize that:

***H1**: Responsible AI governance will have a positive effect on knowledge management capability*

6.2 Hypothesis 2

Deploying responsible AI governance can enhance an organization's ability to respond to changes. To control their AI applications and make sure that they behave according to the set of principles, organizations should implement specific processes for AI development. For example, to ensure the integrity of the system or that it does not execute bias. Also, these processes and mechanisms are documented in a clear and concise manner. These well-established and documented processes can make it easier and faster for organizations to respond to changes due to several reasons.

Reacting to change can mean that activities have to be coordinated across different business units (Tallon & Pinsonneault, 2011). If there is already a solid infrastructure and established processes in place for developing AI applications, it is easier for organizations to overcome these coordination barriers. An example is from the case study, where there is a continuous collaboration between the AI department and the other departments, making it easy for them to coordinate activities. Also, they established a shared vocabulary, making it easier for them to communicate efficiently. Their internal business processes make it easier to rapidly cope with changes, as they facilitate communication across business units. Also, communication and collaboration within a single department can happen more easily when established processes are in place.

In addition to speeding up development, having established processes and mechanisms can also speed up decision-making. Responsible AI governance mechanisms include defining responsibilities, such as AI and data ownership responsibilities. For example, the company studied in the case study had a dedicated chief AI officer responsible for all the AI-related activities in the organization. When responding to changes, decisions have to be made to know how to respond. By assigning clear roles and responsibilities, decision-making can be executed more efficiently, making the organization able to respond to changes faster. Thus, I propose the following hypothesis:

***H2**: Responsible AI governance will have a positive effect on organizational agility*

6.3 Hypothesis 3

KMC reflects how well an organization is able to identify, develop and leverage its knowledge resources (Liu et al., 2014). Organizations identify and acquire knowledge about its environment (e.g., customers and suppliers), which enhances the organizations ability to sense changes, threats or opportunities that should be responded to, such as change in customer needs. These knowledge resources can then be used to adjust processes, strategies and operational capabilities to respond to the changes with robust solutions (Rafi et al., 2021). Knowledge resources can include knowledge about products and customers, and managerial knowledge (Tanriverdi, 2005). Having good KMC can make firms able to integrate the relevant knowledge more effectively, enhancing its ability to cope with market or demand changes. For example, integrating knowledge about customers can lead to new market insights (Tanriverdi, 2005), which in turn can be used to adjust activities according to the market (e.g., product adjustments). Therefore, I hypothesize that:

***H3**: Knowledge management capability will have a positive effect on organizational agility*

6.4 Hypothesis 4

It is widely known that there is a relationship between KMC and organizational performance (Mao et al., 2016; Tanriverdi, 2005; Tseng & Lee, 2014). Knowledge is seen as one of the most valuable and important resources of a company (Rafi et al., 2021), and companies that have the ability to create, use and manage the right knowledge can experience several benefits (Tseng & Lee, 2014). Having strong KMC can, among other, enhance the quality of product and services, and contribute to new product and service development. KMC helps organizations enhance their processes, which is critical for competitive performance. Several studies have investigated the relationship between KMC and performance. For example, Tanriverdi (2005) shown that KMC has a positive effect on corporate financial performance of multibusiness firms. Thus, I propose the following hypothesis:

***H4**: Knowledge management capability will have a positive effect on competitive performance*

6.5 Hypothesis 5

Agility allows organizations to improve their critical business processes to meet the needs and demands of the market (Rafi et al., 2021). For example, organizations can better adjust its product and service offerings to respond to changes in customer preferences. Agility is associated with enhanced ability of sensing and responding to opportunities for competitive actions. By expanding its repertoire of competitive actions, firms are more likely to experience improved performance (Mikalef & Pateli, 2017; Tallon & Pinsonneault, 2011). Hence, I propose the following hypothesis:

***H5**: Organizational agility will have a positive effect on competitive performance*

7 Survey Method

A quantitative study was performed to test the research model presented in Chapter 6 empirically. The survey method was chosen as a strategy. In a survey study, the same kinds of data are collected from a large group, which can be analyzed for patterns that allow for the generalization of the findings (Oates, 2006; Pinsonneault & Kraemer, 1993). A questionnaire, sent out to Nordic companies, was chosen as the method for generating data. This chapter presents the process of collecting data for the questionnaire-based survey and the construct measures used in the survey.

7.1 Data Collection

To test the research model, an electronic questionnaire-based survey was sent out to Nordic companies. Data were collected from Sweden, Denmark, Finland, and Norway (Figure 4). These countries are considered at the front of global competitiveness, ranking at 8th, 10th, 11th, and 17th place, respectively, according to the 2019 Global Competitiveness Report of the Global Economic Forum (Schwab, 2019). The Nordic countries have high levels of ICT adoption, and the majority of people have advanced digital skills, making them well equipped for digital transformation (Schwab & Zahidi, 2020).

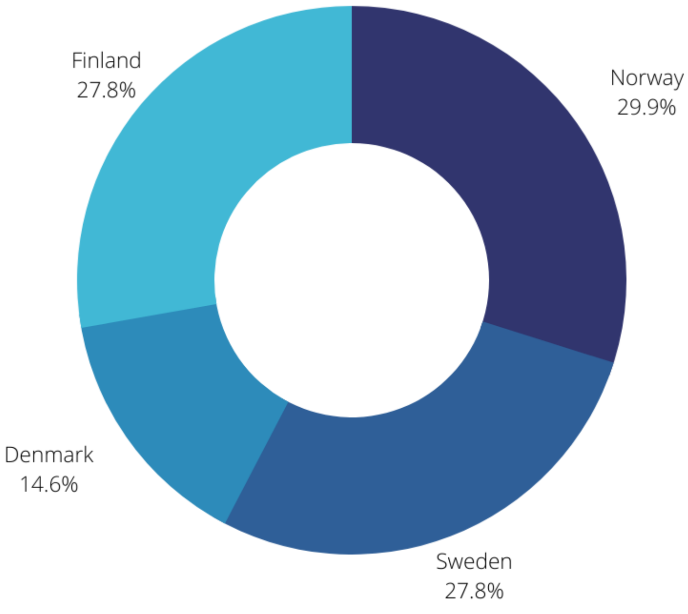

Figure 4: Distribution of Responses by Country

The data collection period lasted approximately four weeks, during April and May 2021. Two different approaches for gathering data for the questionnaire were employed. The first approach targeted senior IS executives through email. The names of senior IS executives were obtained through several sources, personal contacts, and LinkedIn[8], among others. An email invitation was sent to respondents for them to participate in the study. The invitation was followed by one email reminder, sent one week after the initial invitation. A personalized report for each respondent was provided, benchmarking their performance in relation to averages obtained from the survey.

A panel service company (Alchemer[9]) was also contacted to collect data for the questionnaire because of its broader reach in the Nordic countries. The company was asked to target senior IS executives in Nordic companies.

After the data collection period ended, two different datasets were obtained. The first dataset was obtained from the approach targeting respondents through email and consisted of 24 complete responses. The second dataset was obtained from the panel service company and consisted of 120 complete responses. The two datasets were merged into a single dataset, consisting of 144 complete responses to be further analyzed.

A wide range of industries is present in the final data set (Table 8). The largest proportion of responses comes from the technology sector (27.8%), followed by ICT and telecommunications (13.9%), financials (7.6%), and consumer goods (5.6%). The rest of the responses comes from a variety of different industries, such as education, transport, and media. Most of the responses came from relatively large firms, with 69% of the companies having more than 250 employees. The survey was mainly targeted at senior executives having knowledge about both AI use in the company and the companies' key performance indicators. As shown in Table 8, most respondents had a managerial or higher-level role relating to the use of IT.

[8] https://www.linkedin.com

[9] https://www.alchemer.com/

Table 8: Descriptive Statistics of the Sample and Respondents

Factors	**Sample (N = 144)**	**Proportion (%)**
Industry		
Technology	40	27.8%
ICT and Telecommunications	20	13.9%
Financials	11	7.6%
Consumer Goods	8	5.6%
Consumer Services	8	5.6%
Health Care	8	5.6%
Industrials	7	4.9%
Other (Education, Manufacturing, Media etc.)	42	29.2%
Firm Size (Number of Employees)		
1-49	17	11.8%
50-249	27	18.8%
250-499	28	19.4%
500-999	29	20.1%
1000-2499	26	18.1%
2500+	17	11.8%
Total AI Experience of Company (Years)		
<1 year	8	5.6%
1 year	10	6.9%
2 years	41	28.5%
3 years	46	31.9%
4+ years	39	27.1%
Respondent's Position		
CIO/CTO	17	11.8%
Head of IT Department	16	11.1%
IT Project Manager	15	10.4%
IT Director	13	9.0%
CEO	10	6.9%
Operations Manager	10	6.9%
Other (Business Manager, Project Manager etc.)	63	44%

The majority of firms seem to have several years of experience using AI, with 59% having three or more years of experience with AI. From Table 9, it is clear that many companies are employing AI in several applications and also using several different AI technologies. The most common AI technology employed is machine learning (64.6%), followed by robotics (36.1%), natural language processing (30.6%), and planning, scheduling, and optimization techniques (30.6%). AI is used for a variety of different tasks. The most common application is chatbots (47.9%), with almost half of the respondents ticking that alternative. Chatbots are followed by cybersecurity (39.6%), virtual agents (34.7%), and AI for decision management (34.7%) for the most common AI applications.

Table 9: Descriptive Statistics of AI Use in the Sample (Top Answers)

AI Use	**Sample (N = 144)**	**Proportion (%)**
AI Applications		
Chatbots	69	47.9%
Cybersecurity	57	39.6%
Virtual Agents	50	34.7%
AI for Decision Management	50	34.7%
Robotic Process Automation	49	34.0%
Intelligence Supply Chain Management	37	25.7%
Real-Time Translation	35	24.3%
AI Technologies		
Machine Learning	93	64.6%
Robotics	52	36.1%
Natural Language Processing (NLP)	44	30.6%
Planning, Scheduling and Optimization	44	30.6%
Expert Systems	38	26.4%
Speech Analytics	33	22.9%
Machine Vision	30	20.8%
Feed-Forward Networks	28	19.4%
Recurrent Neural Networks	23	16.0%
Reinforcement Learning	20	13.9%
Convolutional Neural Networks	19	13.2%

7.2 Construct Measurement

To quantitatively test the relationships between different constructs, a concrete way of measuring these constructs has to be established. The measures for the constructs KMC, organizational agility, and competitive performance were adopted from prior studies and are presented in Appendix B. These scales have previously been tested in empirical studies. On the other hand, the measures for responsible AI governance have been developed in this thesis (Chapter 5).

7.2.1 Responsible AI Governance

Responsible AI governance is conceptualized as a third-order formative construct. Eight dimensions comprise the notion of responsible AI governance. Three of these dimensions, namely accountability, laws and regulation, and environmental and societal well-being, are first-order reflective constructs. The remaining five dimensions (transparency, fairness, robustness and safety, data governance, and human-centric AI) are second-order formative constructs, comprising 13 first-order reflective constructs. The items used to measure the responsible AI governance construct were developed in this thesis (Chapter 5) and are presented in Table 6. The development of the items was based on previous literature and the case study presented in Chapter 4, following the guidelines of (MacKenzie et al., 2011). All items were created as statements to be able to employ a 7-point Likert scale (1: Do not agree, 7: Agree completely).

7.2.2 Knowledge Management Capability

KMC reflects an organization's ability to create, transfer, integrate and leverage knowledge within the organization (Tanriverdi, 2005). The items used to measure the KMC of firms were adopted from the study of Mao et al. (2016), where they also were empirically confirmed. The respondents were asked questions about the degree to which they are able to manage knowledge within the organization. A 7-point Likert scale was used, where a value of 1 means disagree entirely, and 7 means agree entirely.

7.2.3 Organizational Agility

Organizational agility refers to the ability of firms to detect and respond to opportunities and threats with ease, speed, and dexterity (Tallon & Pinsonneault, 2011). The items used to measure the degree of organizational agility include, among others, responding to customer demand, expanding to new markets, switch suppliers, and adopting new technologies. The measurements were based on scales of agility used in previous research and have been empirically confirmed (Tallon & Pinsonneault, 2011). Respondents were asked to evaluate how easily and quickly their firms can perform a number of eight actions, using a 7-point Likert scale (1: Do not agree, 7: Agree completely).

7.2.4 Competitive Performance

Competitive performance refers to the degree to which a firm performs better than its main competitors (Rai & Tang, 2010). Respondents were asked to evaluate the degree to which they perform better than their key competitors in different aspects, such as market share, delivery cycle time, and customer satisfaction. A 7-point Likert scale was employed, ranging from 1 (totally disagree) to 7 (totally agree).

8 Results

A partial least squares-based structural equation modeling (PLS-SEM) analysis was performed to estimate the hierarchical research model. PLS-SEM is considered appropriate to use in this thesis, as the objective is to develop a theory rather than performing theory confirmation (Hair et al., 2011). The software SmartPLS3[10] was used for all parts of the analysis. The relationships between the constructs (latent variables) and their indicators are presented in the measurement model, while the structural model presents the relationships between the constructs (Sarstedt et al., 2017).

8.1 Measurement Model

The measurement model describes the relationships between the observed data and the latent variables, and its reliability and validity have to be assessed before assessing the structural model (Hair et al., 2019). First, the reliability and validity of the first-order constructs are established, then the higher-order constructs are assessed.

To assess the reliability and validity of the first-order constructs, a number of tests were performed. This process included assessing their reliability, convergent validity, and discriminant validity. Assessing reliability means checking that measures are consistent over time (Golafshani, 2003). The reliability at both the construct and item level was assessed. For the construct level, the Composite Reliability (CR) and Cronbach's Alpha (CA) values were examined. Nunnally (1978) recommends using a threshold of 0.70 for reliability values. However, since this is considered exploratory research, reliability values above 0.60 are considered acceptable (Hair et al., 2019). The CR and CA values are presented in Appendix C, from which it is clear that all reliability values exceed the threshold of 0.60, with most values also exceeding the threshold of 0.70.

For the item level, construct-to-item loadings were examined. The loadings are presented in Appendix D. It is recommended to have values above 0.7, indicating that the construct explains more than 50% of an item's variance (Sarstedt et al., 2017). All the loadings have a higher value than the recommended threshold.

[10] https://www.smartpls.com/

Validity refers to how well a measure corresponds to what it was intended to measure (Golafshani, 2003). Two types of validity are assessed, namely convergent and discriminant validity. Convergent validity reflects the extent to which measures that should be related are actually related, or in other words, how closely measures of the same constructs are related (Carlson & Herdman, 2012). Conversely, discriminant validity reflects the extent to which measures are not correlated to unrelated measures and ensures that a construct's measure is empirically unique (Henseler et al., 2014). Convergent validity was established by examining the average variance extracted (AVE) values (Appendix C). The AVE value for a construct is obtained by squaring each of its items loading and computing the mean value. All values were above the lower limit of 0.5, indicating that the constructs explain more than 50% of the variance of its items (Hair et al., 2019).

Discriminant validity was established using three different methods. First, the Fornell-Larcker criterion was employed by looking at each construct's AVE square root to verify that its value exceeds its highest correlation with any other construct (Fornell & Larcker, 1981). Second, the outer loadings of all constructs were examined to make sure that their values were greater than their cross-loadings with other constructs (Appendix D). Lastly, the Heterotrait-Monotrait ratio (HTMT) was assessed (Appendix E). According to Henseler et al. (2014), the HTMT ratio is better at assessing discriminant validity than the traditional Fornell-Larcker criterion and cross-loadings. Using a threshold of 0.9 (Henseler et al., 2014), discriminant validity is established.

For the higher-order constructs, the weights of the lower-order constructs on the higher-order constructs were assessed (Table 10). Weights close to (+/-) 1 indicate a strong (positive/negative) relationship, while weights close to 0 indicate a weak relationship. To assess the statistical significance of the weights, bootstrapping was performed. This check resulted in all weights found to be significant. To make sure that no multicollinearity exists, variance inflation factor (VIF) values were examined. Generally, VIF values below 5 suggest low multicollinearity, but values close to or lower than 3 are recommended (Hair et al., 2019). From Table 10, it is clear that all VIF values are below the threshold of 5, with a majority of the values also being close to or lower than 3.

Table 10: Higher-Order Construct Validation

Construct	Measures	Weight	Significance	VIF
Transparency	Explainability	0.489	$p<0.001$	1.575
	Traceability	0.298	$p<0.001$	1.517
	Communication	0.435	$p<0.001$	1.407
Fairness	No Unfair Bias	0.584	$p<0.001$	1.542
	Accessibility	0.537	$p<0.001$	1.542
Robustness and Safety	Resilience and Security	0.293	$p<0.001$	1.438
	Reliability and Reproducibility	0.570	$p<0.001$	2.326
	Accuracy	0.300	$p<0.001$	1.995
Data Governance	Data Privacy	0.504	$p<0.001$	2.053
	Data Quality	0.377	$p<0.001$	1.910
	Data Access	0.259	$p<0.001$	2.135
Human-centric AI	Human Oversight	0.538	$p<0.001$	1.896
	Human Well-Being	0.551	$p<0.001$	1.896
Responsible AI Governance	Transparency	0.167	$p<0.001$	2.872
	Fairness	0.149	$p<0.001$	2.383
	Accountability	0.133	$p<0.001$	3.285
	Robustness and Safety	0.180	$p<0.001$	3.146
	Data Governance	0.235	$p<0.001$	3.146
	Laws and Regulations	0.102	$p<0.001$	3.282
	Human-Centric AI	0.175	$p<0.001$	3.988
	Environmental and Societal Well-Being	0.087	$p<0.001$	1.756

8.2 Structural Model

The structural model from the PLS analysis is presented in Figure 5, and shows the standardized path coefficients (β) and the coefficient of determination (R^2). The path coefficients describe the strength of the relationship between two constructs (Sarstedt et al., 2017), where a value close to (+/-) 1 indicates a strong (positive/negative) relationship. The coefficient of determination describes the explained variance of endogenous variables and is a measure of the model's explanatory power (Hair et al., 2019). Its value ranges from 0 to 1, where higher values indicate higher levels of predictive accuracy (Hair et al., 2017). To make sure that the results from the PLS analysis is significant, a bootstrap analysis is performed to obtain the t-statistics.

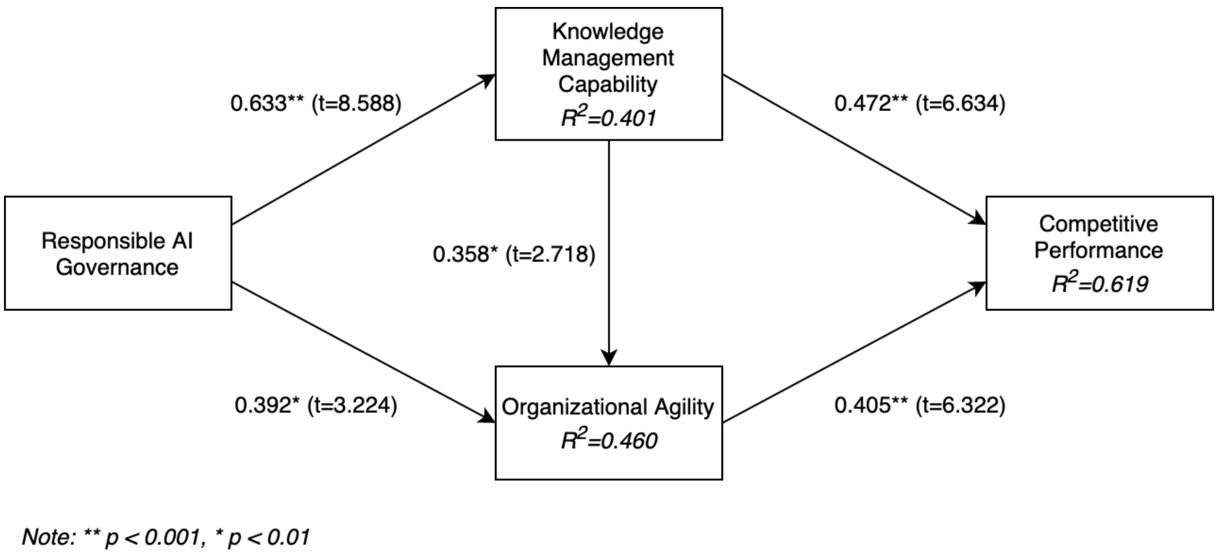

Figure 5: Estimated Relationship of Structural Model

All five hypotheses are empirically supported, as presented in Table 11. Deploying responsible AI governance practices is found to impact the KMC of an organization (β=0.633, t=8.588, p<0.001). Also, it is found to affect the organizational agility (β=0.392, t=3.224, p<0.01). In addition, the KMC of an organization are positively associated with organizational agility (β=0.358, t=2.718, p<0.01) and competitive performance (β=0.472, t=6.634, p<0.001). Additionally, an organization's agility is found to affect the competitive performance (β=0.405, t=6.322, p<0.001).

Table 11: Results of Hypothesis Testing

Hypothesis	**Effect**	**t-value**[11]	**Conclusion**
H1: Responsible AI governance → KMC	0.633	8.588**	Supported
H2: Responsible AI governance → Organizational agility	0.392	3.224*	Supported
H3: KMC → Organizational agility	0.358	2.718*	Supported
H4: KMC → Competitive performance	0.472	6.634**	Supported
H5: Organizational agility → Competitive performance	0.405	6.322**	Supported

[11] ** p < 0.001, * p < 0.01

The structural model explains 40.1% of the variance for KMC (R^2=0.401), 46.0% for organizational agility (R^2=0.460) and 61.9% for competitive performance (R^2=0.619). According to Hair et al. (2011), this can be considered moderate to substantial predictive power.

9 Discussion

This chapter aims to discuss the findings of this thesis in light of the research questions. Moreover, the implications for research, practice, and society are discussed. Lastly, some limitations are discussed together with proposals for future research.

9.1 Discussing Research Questions

In this section, I will discuss the research questions presented in Chapter 1.2 in light of findings from both the case study (Chapter 4), and the survey study (Chapter 5, 6, 7, and 8).

9.1.1 RQ1: What Does Responsible AI Governance Comprise, and How is it Implemented in Practice?

Even though the interest in AI has increased considerably in the past years, there is still no clear and unanimous framework, rules, or laws for organizations to follow when developing their AI capabilities. This issue can be emphasized by the increasing number of cases one read about in the news, about AI applications acting biased, or having some other negative impact. Many organizations see AI as the tool that will solve all of their problems. However, the reality is quite different for most companies. AI brings with it high risks that should be managed and controlled. Deploying adequate AI governance is thus essential for organizations to minimize the risks of AI while at the same time exploiting its potentials.

In this thesis, I have examined the development and deployment of responsible AI governance in Nordic companies. By using a mixed methods approach, valuable insight about what constitutes responsible AI governance in a business context is provided, together with illustrations of how to implement it in practice. Findings have revealed eight dimensions of responsible AI governance: transparency, fairness, accountability, robustness and safety, data governance, laws and regulations, human-centric AI, and environmental and societal well-being. These dimensions, in turn, include 13 sub-dimensions. Together, these dimensions comprise the mechanisms to be incorporated by organizations to ensure ethical, responsible, and trustworthy AI applications. Also, findings from the work of this thesis suggest how to put responsible AI governance into practice. More specifically, the survey instrument highlights processes and mechanisms to incorporate for each of the dimensions. In addition, the case study provides several real-world examples of governance mechanisms that have successfully been deployed.

The principles of responsible AI governance have been thoroughly discussed in Chapter 5. However, how to implement these principles in practice can be exemplified through the case study analyzed in Chapter 4. For instance, mechanisms to provide transparency include choosing simple models because they are the easiest to explain and having a continuous feedback loop with the employees using the AI solutions so that they can provide feedback for their experiences. Moreover, a chief AI officer is in charge of the AI development, bringing accountability practices, such as the use of task forces to learn from failures. To ensure the robustness and safety of their AI applications, the company studied sketch problematic scenarios together with domain experts, as they are the ones that know the domain the application will be deployed in the best. During development, there is also a continuous focus on creating robust solutions, such as testing the solution using real-world data before deploying it.

By avoiding using sensitive data, aspects related to data governance are less complicated. However, several mechanisms to ensure data quality can be illustrated, such as having automatic error detection on incoming data. To ensure that all AI applications comply with laws and regulations at all times, restrictions to follow are included in the logic of the model. In addition, the behavior of the model is monitored, and a "big red stop-button" exists to be able to stop the applications from doing illegal trading. Furthermore, there is an extensive focus on making human-centric AI applications, having the employees who will interact with the solutions in mind at all times. There is a continuous interaction and feedback loop between the AI developers and the domain experts, and they work together in finding out how to inform the employees interacting with the AI applications in the best way. These examples help bridge the gap between the principles and the practical steps to execute it.

9.1.2 RQ2: What are the Effects of Deploying Responsible AI Governance, and Through What Mechanisms are Performance Gains Realized?

While there is a growing interest in how organizations responsibly can govern their AI, there is still uncertainty about its effect on organizations. More specifically, if deploying responsible AI governance can turn into competitive performance, and if so, through which mechanisms this occurs. This thesis employed a questionnaire-based survey study to explore the relationships between responsible AI governance, KMC, organizational agility, and competitive performance. The proposed research model hypothesizes that deploying responsible AI governance will positively affect both the KMC and organizational agility of a firm, both of which will enhance competitive performance. In addition, I hypothesized that KMC will have a positive effect on organizational agility. Survey data were collected from 144 senior IS executives in Nordic companies. To analyze the data, a PLS-SEM analysis was performed. An analysis of the structural model confirmed that all paths are significant, indicating that all the proposed hypotheses are supported.

Findings from the PLS-SEM analysis reveal that deploying responsible AI governance will make a significant positive impact on an organizations' KMC and organizational agility. These findings implies that organizations get better at gaining and distributing knowledge and can respond to changes more easily by deploying responsible AI governance. Furthermore, the results indicate that the impact of responsible AI governance on KMC is more substantial than its (direct) impact on organizational agility. However, organizational agility is also found to be significantly enhanced by the KMC of a firm. This result indicates that deploying responsible AI governance can affect organizational agility by several means, both directly and indirectly, through enhanced KMC.

Both KMC and organizational agility are found to have a positive effect on competitive performance. Their impact on competitive performance is more or less equal, with KMC being slightly more substantial. These findings suggest that deploying responsible AI governance can enhance the competitive performance of organizations through the mediating roles of KMC and organizational agility.

9.2 Research Implications

Several contributions to the AI governance literature can be concluded from this research. First, the in-depth case study was designed to get insight into how a company has successfully adopted AI, and it has revealed several interesting findings. Specifically, the case study has revealed how a company successfully has managed to leverage resources to gain business value from AI investments, which is essential in understanding the value generation from AI. Several mechanisms of AI governance are identified through an analysis building on the IT governance (Peterson, 2004), information governance (Borgman et al., 2016; Tallon et al., 2013), and data governance (Tallon, 2013) literature. More precisely, the mechanisms identified are grouped into structural, procedural, and relational governance practices as done by previous literature. This work opens the discussion of what the notion of AI governance for organizations can entail, and through which mechanisms it is executed. Also, from the case study, several process-level and firm-level changes caused by the adoption of AI is identified.

Second, a definition of responsible AI governance is provided. This definition draws on the already established notions of “corporate governance” (Cadbury, 1992) and “IT governance” (De Haes & Van Grembergen, 2004; Peterson, 2004). Additionally, it builds on work from the field of responsible AI. Several studies have addressed the need to talk about how to practically implement responsible, ethical, and trustworthy principles into the entire life cycle of AI (Chowdhury et al., 2020; Jobin et al., 2019). Thus, by providing a definition of what responsible AI governance includes, it addresses the calls made by several scholars and provides a step in the right direction to define what responsible AI governance entails for organizations.

Third, I present a theoretical framework for responsible AI governance, which consists of different principles that organizations should govern their AI according to. The AI literature draws on the IT literature. However, because of AI’s characteristics, there are possibly other topics that need to be taken into consideration, and thus there is a need to assess the dimensions making up AI governance. For example, adopting AI brings risks not seen in previous technologies, such as the potential for acting biased (Ntoutsi et al., 2020) and the black-box nature of many AI models causing problems with providing explanations (Adadi & Berrada, 2018; Loyola-González, 2019). These risks have to be addressed by organizations employing AI. Several efforts have been made in the field of responsible AI (Benjamins et al., 2019; Dignum, 2019). However, existing literature has pointed out the need to identify what exactly constitutes responsible AI and how to put it into practice (Chowdhury et al., 2020; Jobin et al., 2019). To identify and categorize the dimensions and principles of responsible AI governance, a review of existing literature on topics concerning AI governance was performed, such as AI ethics and trustworthiness of AI. This review resulted in several important themes emerging, which were then grouped to form a hierarchy of principles. The identification of principles for responsible AI governance highlights several areas of AI governance that need further attention.

Fourth, building on the framework mentioned above, this study develops a construct that can be empirically applied to assess the maturity of a firm in terms of responsible AI governance. As mentioned, AI brings challenges and risks not seen in previous IT. Thus I argue that there are considerable differences between the governance of AI and IT, which raises the need to define the responsible AI governance construct. Based on the principles identified and by following the guidelines of MacKenzie et al. (2011), the responsible AI governance instrument is developed. The measurements were developed based on existing literature on trustworthy, ethical, and responsible AI touching upon how to put AI principles into practice. The reliability and validity of the responsible AI governance construct were manifested through the survey study. This effort addresses the challenges raised by several researchers in how to put AI principles into practice (Chowdhury et al., 2020; Jobin et al., 2019).

Fifth, the impact that responsible AI governance practices can have on an organization is demonstrated. More precisely, I have assessed to what extent AI governance impacts competitive performance, mediated through the KMC and the organizational agility of an organization. To the best of my knowledge, there are no empirical studies linking the conceptualization of AI governance with performance indicators. Through this study, I have empirically demonstrated that by deploying responsible AI governance, an organization can realize gains in terms of enhanced KMC and organizational agility, and enhanced competitive performance. These findings indicate that developing and deploying responsible AI governance can be an essential contributing factor for organizations to achieve a competitive advantage, as it positively impacts the competitive performance.

9.3 Practical Implications

Organizations find themselves having to choose between speed to market and taking the time to build comprehensive AI governance capabilities (KPMG, 2021). This study highlights the importance for organizations to take into consideration ethical and socio-technical implications when building their AI capabilities. Organizations should not only focus on monetary values when adopting AI but also compromise between the organizational objectives and what is good for society. These aspects can be implemented by deploying responsible AI governance.

A framework for responsible AI governance is presented, comprising several principles that organizations should govern their AI according to. Organizations can use this framework to know which aspects are essential to consider when deploying AI, as it highlights several risks associated with its use. For example, the principle of transparency, with the sub-dimensions explainability, traceability, and communication, helps organizations to recognize that their AI applications should be understandable and that the users should be aware that they are interacting with an AI-powered product or service.

The survey instrument developed in this thesis provides a valuable starting point for organizations to implement responsible AI governance, as it touches upon governance mechanisms to be incorporated in all phases of the AI life cycle. The instrument can guide organizations that employ AI technologies, as it includes crucial processes and mechanisms that should be incorporated. Organizations can use the list of items as a starting point and evaluate which of these mechanisms to incorporate. All items are not necessarily equally relevant to all organizations. It depends on the context the AI application will be deployed in. For example, it can be more critical for AI applications guiding decisions in health care to be robust, solid, and taking into ethical and socio-technical implications than a simple recommendation engine for products in a webshop. Also, by using the framework, organizations are able to self-assess their maturity in terms of responsible AI governance, possibly highlighting processes and mechanisms they need to implement or enhance to be at a satisfactory level.

Adding to the above, more detailed and in-depth knowledge about the different governance mechanisms can be gained from the case study. More specifically, how to put the principles into practice. The case study has revealed through which mechanisms a company has successfully managed to govern their AI, giving organizations an idea of how to develop and deploy responsible AI governance. Existing literature has emphasized the need for including domain experts when developing AI (Tarafdar et al., 2019). This study highlights the importance of using cross-functional teams during AI development. The domain experts are the ones that know the domain best. Possibly, they are also the ones that are going to work side-by-side with the AI solution or monitor its behavior. Including domain experts in the loop at all times of AI development is crucial to make sustainable and robust solutions. Together the AI developers and human experts should decide on the requirements for the application. In addition, problematic scenarios should be defined in collaboration so that alerts are given at the right times.

For providing information about the AI systems' behavior and outcomes, dashboards can be used as an effective management and monitoring tool. This mechanism is highly relevant for AI solutions deployed within an organization, such as the wind forecasting and algorithmic trading systems mentioned in the case study. The dashboards work as the communication channel between humans and machines, tracking KPIs and other relevant data points. According to Loyola-González (2019), problems with understanding the AI model's output are often related to the visual form of the output and not necessarily the black-box nature of the model. Thus it can be an advantage to show plots, graphs, and metrics in the dashboards that the human experts are already familiar with looking at.

Real-world data is uncertain and behaves differently than the training and test data sets. Sometimes data is missing or retrieved late. To ensure that the AI system is robust and behaves as intended, this study proposes that it could be an advantage to test the behavior of the AI solution using real-world data. For example, let the model run using real-world data for a few days or weeks, depending on the nature of the domain. The output from the system in this phase should not be used but rather be checked by human experts who can provide feedback on the model's performance.

In this thesis, I have shown that implementing responsible AI governance can lead to competitive performance. Specifically through the mediating roles of enhanced KMC and organizational agility. I propose that responsible AI governance mechanisms can enhance the KMC of an organization through several means. By focusing on the explainability of the AI models, organizations can gain knowledge on the reasoning of the system. Through continuous monitoring (e.g., using dashboards), knowledge about the behavior and performance of the solution can be gained. Using dashboards to monitor is also a great way of spreading knowledge about the behavior of the solution throughout the organization to the ones that need it. In addition, documenting processes and mechanisms of data and AI development aspects lets knowledge more easily flow within the organization. By creating and gaining knowledge and spreading it throughout the organization, it is easier to spot opportunities and challenges that should be acted on. In other words, enhancing the KMC through deploying responsible AI governance can also lead to enhanced organizational agility, especially in the form of enhanced market capitalizing agility (Lu & Ramamurthy, 2011). In addition, deploying responsible AI governance includes having assigned clear roles and responsibilities and having established processes for AI development. These mechanisms can speed up AI development and decision-making, thus enhancing organizational agility, especially in the form of operational adjustment agility (Lu & Ramamurthy, 2011). These results can work as motivation for organizations to take the time to develop and deploy responsible AI governance, as they see that they can get rewarded in terms of performance gains.

9.4 Societal Implications

The greater society will benefit from the deployment of responsible, ethical and trustworthy AI (Dignum, 2018). Thus, this research takes a step in the right direction by providing a valuable starting point for organizations to develop and deploy comprehensive responsible AI governance. Furthermore, this research provides citizens with a better knowledge to critically use AI-powered products and services.

This study has shown that the deployment of responsible AI governance can significantly impact the competitive performance of an organization. This result can motivate organizations to invest time and resources to develop and deploy comprehensive responsible AI governance, as they see that it pays off. Also, the responsible AI governance framework equips managers and executives with a better understanding of the mechanisms to implement for organizations to deploy ethical, responsible, and trustworthy AI applications. Thus this study can contribute to more responsible, ethical, and trustworthy AI being deployed, benefiting society.

AI is increasingly becoming a part of people's everyday life. The movies watched are chosen using Netflix's recommendation engine, and phones are opened using facial recognition. As a user of these services, it is difficult to know which services should be trusted, given the possible risks of AI (Hurlburt, 2017). This study presents a survey instrument that can be used to assess the maturity of organizations in terms of responsible AI governance. Using this instrument, it is possible to give a score to organizations deploying AI describing their maturity in terms of responsible AI governance, making it easier for users and citizens to know whom to trust.

Additionally, by identifying dimensions of responsible AI governance, this research makes citizens better equipped to critically use AI-powered products and services as they get more aware of the risks brought by the technology. The findings of this research can help educate citizens on how AI technologies work and aspects relevant for their use. For instance, by knowing what transparency or human-centric AI means, citizens can get better equipped to adopt AI technologies and make well-informed decisions about their use.

9.5 Limitations and Future Research

As with any research, this study has its limitations. First, there may be additional important aspects of responsible AI governance that I did not manage to capture. AI technologies are rapidly evolving, and we are still in the initial stage of understanding its potential impact on the world. There may be breakthroughs that affect how the technology can be utilized and applied. Thus it is difficult to provide an exhaustive list of all practices that should guide the responsible governance of AI.

Second, when performing this research, the focus was mainly on one specific AI technology, namely machine learning. Other AI technologies may pose other challenges that should be taken into consideration. Even though machine learning is the AI technology used by most firms today, several other AI techniques are also employed. Thus it would be interesting for future research to investigate responsible AI governance in the light of other AI technologies.

Third, the survey used respondents that worked in companies based in the Nordic countries. Organizations from different regions would likely respond differently. Nordic companies have a high level of ICT adoption compared to other companies in other countries (Schwab & Zahidi, 2020). Thus it could be that companies in other regions are facing other challenges and implements other mechanisms to control the behavior of their AI applications. As future research, it would be interesting to survey companies situated in other parts of the world and comparing the maturity of responsible AI governance in different regions. Additionally, it would be interesting also to study the effects of deploying responsible AI governance in other regions. The effects in other regions might differ from those in the Nordics, where ethics and responsibility play an essential role in business.

Fourth, the choice of using single respondents for the survey could lead to potential bias in the results. To compensate for the use of single respondents, senior-level technology managers were targeted as respondents. However, there still might exist some bias. For future research, a way to overcome this limitation is to collect data from multiple respondents within companies.

Fifth, the case study is collecting data from only one company. This company is not using sensitive data, and the use of big data is limited, which could mean that an incomplete picture has been painted. More specifically, there may be other challenges organizations have to overcome than those mentioned, which impact the different governance mechanisms employed. It would be interesting for future research to investigate this further. For example, through a multiple case study, interviewing respondents from several companies who have successfully adopted AI.

Finally, this research was performed during the covid-19 pandemic, which could impact the results. The pandemic has made significant impacts on companies in several ways (Bartik et al., 2020). Many companies have been forced to digitalize their processes more quickly than usual (Savić, 2020), thus more companies are reaching to AI to solve their problems (European Commission, 2020a). Some companies have experienced a reduction in income due to closed doors. On the contrary, some companies have seen new opportunities for their business. The data collected for this research were based on only a snapshot in time, and because of covid-19, not necessarily a complete picture of how things are in a "normal" world.

10 Conclusion

In this thesis, the field of responsible AI governance has been examined. Valuable insight about what constitutes responsible AI governance in a business context is provided through creating a framework for responsible AI governance, which highlights principles to guide the governance of AI. In addition, insights on how to turn those principles into practice are presented. Furthermore, this thesis suggests that deploying responsible AI governance will impact competitive performance through the mediating roles of enhanced KMC and organizational agility.

A mixed methods approach was used to explore the field of responsible AI governance. First, a single case study gave insight into how a company has successfully managed to control and govern its AI capabilities, identifying some AI governance mechanisms used in practice. Then, building on the case study and existing literature on responsible, ethical and trustworthy AI, the notion of responsible AI governance is defined, and a theoretical framework for responsible AI governance is developed. The framework highlights eight principles, comprising 13 sub-dimensions that organizations should govern their AI according to. Finally, a survey instrument was developed to measure the maturity of organizations in terms of responsible AI governance, together with a research model hypothesizing about the effects of deploying responsible AI governance. Survey data from 144 high-level IT executives working in Nordic companies were analyzed using PLS-SEM. The survey instrument was empirically validated and demonstrated that organizations deploying responsible AI governance could realize performance gains through enhancing their KMC and organizational agility.

The findings from this thesis have notable implications for both research, practice, and society. It contributes to the AI literature by providing several efforts to bridge the gap between principles and practice. Additionally, it makes a valuable contribution to the AI literature by demonstrating the effects of deploying responsible AI governance. Also, the findings can make organizations better equipped to deploy responsible AI governance, providing a valuable starting point for putting principles to practice. Furthermore, society can benefit through more responsible, ethical and trustworthy AI being deployed.

References

Accenture. (2018). *Responsible AI: A Framework for Building Trust in Your AI Solutions*. Accenture. Retrieved May 29, 2021 from https://www.accenture.com/_acnmedia/PDF-92/Accenture-AFS-Responsible-AI.pdf

Accenture. (2019a). *Responsible AI Governance Guidebook*. Accenture. Retrieved June 1, 2021 from https://view.pagetiger.com/aigovernanceguide

Accenture. (2019b). *The sphere of control: Accenture 2019 Global Risk Management Study Financial Services Report*. Accenture. Retrieved May 29, 2021 from https://www.accenture.com/us-en/insights/financial-services/global-risk-study

Accenture. (2021). *Responsible AI: From principles to practice*. Accenture. Retrieved May 29, 2021 from https://www.accenture.com/bg-en/insights/artificial-intelligence/responsible-ai-principles-practice

Adadi, A., & Berrada, M. (2018). Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI). *IEEE Access*, *6*, 52138-52160.

Afiouni, R. (2019). *Organizational Learning in the Rise of Machine Learning* International Conference on Information Systems (ICIS), Munich, Germany.

Alsheibani, S. A., Cheung, Y., Messom, D. C., & Alhosni, M. (2020). *Winning AI Strategy: Six-Steps to Create Value from Artificial Intelligence* American Conference on Information Systems (AMCIS), Virtual.

Amershi, S., Begel, A., Bird, C., DeLine, R., Gall, H., Kamar, E., Nagappan, N., Nushi, B., & Zimmermann, T. (2019). *Software Engineering for Machine Learning: A Case Study* International Conference on Software Engineering (ICSE), Montréal, Canada.

Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016). *Machine Bias*. ProPublica. Retrieved May 25, 2021 from https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

Arrieta, A. B., Díaz-Rodríguez, N., Ser, J. D., Bennetot, A., Tabik, S., Barbado, A., Garcia, S., Gil-Lopez, S., Molina, D., Benjamins, R., Chatila, R., & Herrera, F. (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion*, *58*, 82-115.

Baier, L., Jöhren, F., & Seebacher, S. (2019). *Challenges in the Deployment and Operation of Machine Learning in Practice* European Conference on Information Systems (ECIS), Stockholm, Sweden.

Bartik, A. W., Bertrand, M., Cullen, Z., Glaeser, E. L., Luca, M., & Stanton, C. (2020). The impact of COVID-19 on small business outcomes and expectations. *PNAS*, *117*(30), 17656-17666.

Benjamins, R., Barbado, A., & Sierra, D. (2019). *Responsible AI by Design in Practice* Human-Centered AI: Trustworthiness of AI Models & Data (HAI), Arlington, USA.

Binns, R. (2018). *Fairness in Machine Learning: Lessons from Political Philosophy* Proceedings of the 1st Conference on Fairness, Accountability and Transparency, New York, USA.

Borgman, H., Heier, H., Bahli, B., & Boekamp, T. (2016). *Dotting the I and Crossing (out) the T in IT Governance: New Challenges for Information Governance* 49th Hawaii International Conference on System Sciences (HICSS), Koloa, USA.

Buiten, M. C. (2019). Towards Intelligent Regulation of Artificial Intelligence. *European Journal of Risk Regulation*, *10*(1), 41-59.

Butcher, J., & Beridze, I. (2019). What is the State of Artificial Intelligence Governance Globally? *The RUSI Journal*, *164*(5-6), 88-96.

Cadbury, A. (1992). *Report of the Committee on the Financial Aspects of Corporate Governance*. London: Gee.

Carlson, K. D., & Herdman, A. O. (2012). Understanding the Impact of Convergent Validity on Research Results. *Organizational Research Methods*, *15*(1), 17-32.

Castillo, D., Canhoto, A. I., & Said, E. (2020). The dark side of AI-powered service interactions: exploring the process of co-destruction from the customer perspective. *The Service Industries Journal*, 1-26. https://doi.org/10.1080/02642069.2020.1787993

Chowdhury, R., Rakova, B., Cramer, H., & Yang, J. (2020). *Putting Responsible AI Into Practice*. MIT Sloan Management Review. Retrieved June 1, 2021 from https://sloanreview.mit.edu/article/putting-responsible-ai-into-practice/

Clarke, R. (2019). Principles and business processes for responsible AI. *Computer Law and Security Review*, *35*(4), 410-422.

Davenport, T. H., & Ronanki, R. (2018). Artificial Intelligence for the Real World. *Harvard Business Review*, *96*(1), 108-116.

De Haes, S., & Van Grembergen, W. (2004). IT Governance and Its Mechanisms. *Information Systems Control Journal*, *1*, 27-33.

Demlehner, Q., & Laumer, S. (2020). *Shall We Use It or Not? Explaining the Adoption of Artificial Intelligence for Car Manufacturing Purposes* European Conference on Information Systems (ECIS), Virtual.

Dignum, V. (2018). Ethics in artificial intelligence: introduction to the special issue. *Ethics and Information Technology*, *20*, 1-3.

Dignum, V. (2019). *Responsible Artificial Intelligence. How to Develop and Use AI in a Responsible Way*. Springer.

Došilović, F. K., Brčić, M., & Hlupić, N. (2018). *Explainable artificial intelligence: A survey* 2018 41st International convention on information and communication technology, electronics and microelectronics (MIPRO), Opatija, Croatia.

Erdélyi, O. J., & Goldsmith, J. (2018). *Regulating Artificial Intelligence: Proposal for a Global Solution* Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society, New Orleans, LA, USA.

European Commission. (2019). *Ethics Guidelines for Trustworthy AI*. European Commission. Retrieved May 25, 2021 from https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai

European Commission. (2020a). *Artificial Intelligence and Digital Transformation: early lessons from the COVID-19 crisis*. (EUR 30306 EN). Luxembourg: Publications Office of the EU Retrieved from https://publications.jrc.ec.europa.eu/repository/handle/JRC121305

European Commission. (2020b). *Assessment List for Trustworthy Artificial Intelligence (ALTAI) for self assessment*. European Commission. Retrieved May 25, 2021 from https://digital-strategy.ec.europa.eu/en/library/assessment-list-trustworthy-artificial-intelligence-altai-self-assessment

Ford, M. (2013). Could artificial intelligence create an unemployment crisis? *Communications of the ACM*, *56*(7), 37-39.

Fornell, C., & Larcker, D. F. (1981). Evaluating Structural Equation Models with Unobservable Variables and Measurement Errors. *Journal of Marketing Research*, *18*(1), 39-50.

Fountaine, T., McCarthy, B., & Saleh, T. (2019). Building the AI-Powered Organization. *Harvard Business Review*, *97*(4).

Golafshani, N. (2003). Understanding Reliability and Validity in Qualitative Reseach. *The Qualitative Report*, *8*(4), 597-607.

Goodman, B., & Flaxman, S. (2017). European Union Regulations on Algorithmic Decision Making and a «Right to Explanation». *AI Magazine*, *38*(3), 50-57.
Google. (2020). *AI Principles 2020 Progress update*. Google. Retrieved May 26, 2021 from https://ai.google/static/documents/ai-principles-2020-progress-update.pdf
Government of Singapore. (2020). *Model Artificial Intelligence Governance Framework. Second Edition*. Singapore Digital (SG:D).
Hair, J. F., Hult, G. T. M., Ringle, C. M., & Sarstedt, M. (2017). *A Primer on Partial Least Squares Structural Equation Modeling (PLS-SEM) Second Edition*. Sage Publications.
Hair, J. F., Ringle, C. M., & Sarstedt, M. (2011). PLS-SEM: Indeed a Silver Bullet. *Journal of Marketing Theory and Practice*, *19*(2), 139-152.
Hair, J. F., Risher, J. J., Sarstedt, M., & Ringle, C. M. (2019). When to use and how to report the results of PLS-SEM. *European Business Review*, *31*(1), 2-24.
Hamon, R., Junklewitz, H., & Sanchez, I. (2020). *Robustness and explainability of artificial intelligence*. Publications Office of the European Union. Retrieved May 29, 2021 from https://publications.jrc.ec.europa.eu/repository/bitstream/JRC119336/dpad_report.pdf
Henseler, J., Ringle, C. M., & Sarstedt, M. (2014). A new criterion for assessing discriminant validity in variance-based structural equation modeling. *Journal of the Academy of Marketing Science*, *43*(115-135).
Hurlburt, G. (2017). How Much to Trust Artificial Intelligence? *IT Professional*, *19*(4), 7-11.
IBM. (2019). *Accountability*. Retrieved April 19, 2021 from https://www.ibm.com/design/ai/ethics/accountability/
IEEE. (2019). *Ethically Aligned Design*. Retrieved May 20, 2021 from https://ethicsinaction.ieee.org/#ead1e
Janssen, M., Brous, P., Estevez, E., Barbosa, L. S., & Janowski, T. (2020). Data governance: Organizing data for trustworthy Artificial Intelligence. *Government Information Quarterly*, *37*(3), 101493.
Jarrahi, M. H. (2018). Artificial intelligence and the future of work: Human-AI symbiosis in organizational decision making. *Business Horizons*, *61*(4), 577-586.
Jelonek, D., Mesjasz-Lech, A., Stępniak, C., Turek, T., Ziora, L., Arai, K., & Bhatia, R. (2020). *The Artificial Intelligence Application in the Management of Contemporary Organization: Theoretical Assumptions, Current Practices and Research Review* Future of Information and Communication Conference (FICC), San Francisco, USA.
Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, *1*, 389-399.
Jones, T. M. (1980). Corporate Social Responsibility Revisited, Redefined. *California Management Review*, *22*(3), 59-67.
Joshi, A., Kale, S., Chandel, S., & Pal, D. K. (2015). Likert Scale: Explored and Explained. *British Journal of Applied Science & Technology*, *7*(4), 396-403.
Kolbjørnsrud, V., Amico, R., & Thomas, R. J. (2017). Partnering with AI: how organizations can win over skeptical managers. *Strategy & Leadership*, *45*(1), 37-43.
KPMG. (2019). *Controlling AI*. KPMG. Retrieved June 1, 2021 from https://advisory.kpmg.us/articles/2019/controlling-ai.html
KPMG. (2021). The shape of AI governance to come. home.kpmg/ShapeofAIGovernance
Kumar, P., Dwivedi, Y. K., & Anand, A. (2021). Responsible Artificial Intelligence (AI) for Value Formation and Market Performance in Healthcare: the Mediating Role of Patient’s Cognitive Engagement. *Information Systems Frontier*, 1-24. https://doi.org/https://doi.org/10.1007/s10796-021-10136-6
LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*, 436-444.
Licht, K. d. F., & Licht, J. d. F. (2020). Artificial intelligence, transparency, and public decicion-making. *AI & SOCIETY*, *35*, 917-926.

Lichtenthaler, U. (2019). An Intelligence-Based View of Firm Performance: Profiting from Artificial Intelligence. *Journal of Innovation Management*, *7*(1), 7-20.

Liu, H., Song, D., & Cai, Z. (2014). *Knowledge Management Capability and Firm Performance: the Mediating Role of Organizational Agility* Pacific Asia Conference on Information Systems (PACIS), Chengdy, China.

Loyola-González, O. (2019). Black-Box vs. White-Box: Understanding Their Advantages and Weaknesses From a Practical Point of View. *IEEE Access*, *7*, 154096-154113.

Lu, Y., & Ramamurthy, K. R. (2011). Understanding the Link Between Information Technology Capability and Organizational Agility: An Empirical Examination. *MIS Quarterly*, *35*(4), 931-954.

MacKenzie, S. B., Podsakoff, P. M., & Podsakoff, N. P. (2011). Construct Measurement and Validation Procedures in MIS and Behavioral Research: Integrating New and Existing Techniques. *MIS Quarterly*, *35*(2), 293-334.

Makridakis, P. S. (2017). The fourthcoming Artificial Intelligence (AI) revolution: Its impact on society and firms. *Futures*, *90*, 46-60.

Mao, H., Liu, S., Zhang, J., & Deng, Z. (2016). Information technology resource, knowledge management capability, and competitive advantage: The moderating role of resource commitment. *International Journal of Information Management*, *36*(6), 1062-1074.

Martin, K. (2019). Ethical Implications and Accountability of Algorithms. *Journal of Business Ethics*, *160*(835-850).

Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2019). A Survey on Bias and Fairness in Machine Learning. *arXiv preprint arXiv:1908.09635*.

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*, *58*(3), 103434.

Mikalef, P., & Pateli, A. (2017). Information technology-enabled dynamic capabilities and their indirect effect on competitive performance: Findings from PLS-SEM and fsQCA. *Journal of Business Research*, *70*, 1-16.

National New Generation Artificial Intelligence Governance Committee. (2019). *Development of responsible artificial intelligence: a new generation of artificial intelligence governance principles released*. Ministry of Science and Technology of the People's Republic of China. Retrieved May 26, 2021 from http://most.gov.cn/kjbgz/201906/t20190617_147107.htm

Norwegian Ministry of Local Government and Modernisation. (2020). *The National Strategy for Artificial Intelligence*. Ministry of Local Government and Modernisation. Retrieved May 29, 2021 from https://www.regjeringen.no/en/dokumenter/nasjonal-strategi-for-kunstig-intelligens/id2685594/

Ntoutsi, E., Fafalios, P., Gadiraju, U., Iosifidis, V., Nejdl, W., Vidal, M. E., Ruggieri, S., Turini, F., Papadopoulos, S., Krasanakis, E., Kompatsiaris, I., Kinder-Kurlanda, K., Wagner, C., Karimi, F., Fernandez, M., Alani, H., Berendt, B., Kruegel, T., Heinze, C., Broelemann, K., Kasneci, G., Tiropanis, T., & Staab, S. (2020). Bias in data-driven artificial intelligence systems - An introductory survey. *WIREs Data Mining Knowl Discovery*, *10*(3), e1356.

Nunnally, J. C. (1978). *Psychometric Theory*. McGraw-Hill.

Oates, B. J. (2006). *Researching Information Systems and Computing*. Sage.

Peterson, R. (2004). Crafting Information Technology Governance. *Information Systems Management*, *21*(4), 7-22.

Pinsonneault, A., & Kraemer, K. L. (1993). Survey Research Methodology in Management Information Systems: An Assessment. *Journal of Management Information Systems*, *10*(2), 75-105.

Plastino, E., & Purdy, M. (2018). Game changing value from Artificial Intelligence: eight strategies. *Strategy & Leadership*, *46*(1), 16-22.

Pumplun, L., Tauchert, C., & Heidt, M. (2019). *A New Organizational Chassis for Artificial Intelligence - Exploring Organizational Readiness Factors* European Conference on Information Systems (ECIS), Stockholm, Sweden.

PwC. (2018). *Opportunities and Risks of Artificial Intelligence in the Financial Services Industry*. PwC Switzerland. Retrieved May 29, 2021 from https://www.pwc.ch/en/insights/fs/opportunities-and-risks-of-artificial-intelligence-in-the-financial-services-industry.html

PwC. (2019). A practical guide to Responsible Artificial Intelligence (AI). Retrieved May 26, 2021, from https://www.pwc.com/gx/en/issues/data-and-analytics/artificial-intelligence/what-is-responsible-ai/responsible-ai-practical-guide.pdf

Quinio, B., Skandrani, S., Marciniak, R., & Harfouche, A. (2017). *A Framework for Artificial Knowledge Creation in Organizations A Framework for Artificial Knowledge Creation in Organizations* International Conferene on Information Systems (ICIS), Seoul, South Korea.

Rafi, N., Ahmed, A., Shafique, I., & Kalyar, M. N. (2021). Knowledge management capabilities and organizational agility as liaisons of business performance. *South Asian Journal of Business Studies*, *Ahead-of-print*.

Rai, A., & Tang, X. (2010). Leveraging IT Capabilities and Competitive Process Capabilities for the Management of Interorganizational Relationship Portfolios. *Information Systems Research*, *21*(3), 516-542.

Raji, I. D., Smart, A., White, R. N., Mitchell, M., Gebru, T., Hutchinson, B., Smith-Loud, J., Theron, D., & Barnes, P. (2020). Closing the AI Accountability Gap: Defining an End-to-End Framework for Internal Algorithmic Auditing.*FAT* ‘20* Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency, Barcelona, Spain.

Ransbotham, S., Kiron, D., Gerbert, P., & Reeves, M. (2017). Reshaping Business with Artificial Intelligence: Closing the Gap Between Ambition and Action. *MIT Sloan Management Review*, *59*(1).

Sarstedt, M., Ringle, C. M., & Hair, J. F. (2017). Partial Least Squares Structural Equation Modeling. In *Handbook of Market Research* (pp. 1-40). Springer.

Savić, D. (2020). COVID-19 and Work from Home: Digital Transformation of the Workforce. *TGJ*, *16*(2), 101-104.

Schmidt, R., Zimmermann, A., Moehring, M., & Keller, B. (2020). *Value Creation in Connectionist Artificial Intelligence–A Research Agenda* Americas Conference on Information Systems (AMCIS), Virtual.

Schwab, K. (2019). The Global Competitiveness Report 2019. *World Economic Forum*.

Schwab, K., & Zahidi, S. (2020). The Global Competitiveness Report Special Edition 2020: How Countries are Performing on the Road to Recovery. *World Economic Forum*.

Siau, K., & Wang, W. (2018). Building Trust in Artificial Intelligence, Machine Learning, and Robotics. *Cutter Business Technology Journal*, *31*(2).

Strubell, E., Ganesh, A., & McCallum, A. (2019). Energy and Policy Considerations for Deep Learning in NLP. 57th Annual Meeting of the Association for Computational Linguistics (ACL), Florence, Italy.

Tallon, P., & Pinsonneault, A. (2011). Competing Perspectives on the Link Between Strategic Information Technology Alignment and Organizational Agility: Insights from a Mediation Model. *MIS Quarterly*, *35*(2), 463-486.

Tallon, P. P. (2013). Corporate Governance of Big Data: Perspectives on Value, Risk and Cost. *Computer*, *46*(6), 32-38.

Tallon, P. P., Ramirez, R. V., & Short, J. E. (2013). The Information Artifact in IT Governance: Toward a Theory of Information Governance. *Journal of Management Information Systems*, *30*(3), 141-178.

Tanriverdi, H. (2005). Information Technology Relatedness, Knowledge Management Capability, and Performance on Multibusiness Firms. *MIS Quarterly*, *29*(2), 311-334.

Tarafdar, M., Beath, C. M., & Ross, J. W. (2019). Using AI to Enhance Business Operations. *MIT Sloan Management Review*, *60*(4), 37-44.

Thiebes, S., Lins, S., & SUnyaev, A. (2020). Trustworthy artificial intelligence. *Electronic Markets*, 1-18. https://doi.org/https://doi.org/10.1007/s12525-020-00441-4

Thomas, D. R. (2006). A General Inductive Approach for Analyzing Qualitative Evaluation Data. *American Journal of Evaluation*, *27*(2), 237-246.

Tseng, S.-M., & Lee, P.-S. (2014). The effect of knowledge management capability and dynamic capability on organizational performance. *Journal of Enterprise Information Management*, *27*(2), 158-179.

Venkatesh, V., Brown, S. A., & Bala, H. (2013). Bridging the Qualitative-Quantitative Divide: Guidelines for Conducting Mixed Methods Research in Information Systems. *MIS Quarterly*, *37*(1), 21-54.

Vigdor, N. (2019). Apple Card Investigated After Gender Discrimination Complaints. *The New York Times*. https://www.nytimes.com/2019/11/10/business/Apple-credit-card-investigation.html

Wamba-Taguimdje, S.-L., Wamba, S. F., Kamdjoug, J. R. K., & Wanko, C. E. T. (2020). Influence of artificial intelligence (AI) on firm performance: the business value of AI-based transformation projects. *Business Process Management Journal*, *26*(7), 1893-1924.

Wang, H., Huang, J., & Zhang, Z. (2019). *The Impact of Deep Learning on Organizational Agility* International Conference on Information Systems (ICIS), Munich, Germany.

Wang, P. (2019). On Defining Artificial Intelligence. *Journal of Artificial General Intelligence*, *10*(2), 1-37.

Wolf, M. J., Miller, K. W., & Grodzinsky, F. S. (2017). Why We Should Have Seen That Coming: Comments on Microsoft's Tay «Experiment», and Wider Implications. *The ORBIT Journal*, *1*(2), 1-12.

Yampolskiy, R. V., & Spellchecker, M. S. (2016). Artificial Intelligence Safety and Cybersecurity: a Timeline of AI Failures. *arXiv preprint arXiv:1610.07997*.

Yin, R. K. (2003). *Case Study Research: Design and Methods*. Sage.

Appendices

Appendix A: Interview Guide

Background of Interview

This research project aims at understanding how companies use AI technologies to realize value for their company. In particular, we try to understand the focal point and mechanisms of value generation and realization (e.g., augmentation or automation of decision-making or processes) and what specific challenges (organizational and technical) are brought about by AI technologies. Beyond others, we are interested in the interrelationship of AI system designers or teams of designers (e.g., data scientist and data engineers) with productive AI systems such that these systems act upon the set goals of the organization (e.g., the business unit or the user of the AI system).

Introduction

What is your current role and background in the company?

Business Value/Organizational Context

1. Could you mention briefly the history behind AI use in your company? How long it took you to adopt AI (timeline)?
2. Why did you decide to implement AI and how did you go forward in order to implement AI in the company?
3. How did (i) the use of AI grow over time, (ii) how did the AI team grow over time (iii) how did the value/effectiveness of AI grow over time?
4. Could you describe how your company transformed because of AI use?
5. What are the expected returns of AI investment? Do AI systems give you a competitive edge to your competition?
6. Are there any changes brought by AI that you did not anticipate?
7. What would be the consequences if AI will give a non-accurate suggestion? Do you have a strategy for that?
8. Do you plan to use AI in other aspects of your company?
9. What organizational challenges have you faced adopting AI? Have you faced any regulatory challenges?

Control and Technical Aspects

10. Do AI algorithms augment or replace traditional organizational decision-making and how?
11. What type of data do you collect? How do you ensure to use data and AI algorithms such that they are in line with your organizational objectives?
12. Do you recombine different models and ways of AI to ensure (i) behavior and (ii) outcomes in line with the organizational objectives and legal regulations?
13. Do you specify, monitor and evaluate the (i) behavior and (ii) outcomes of your AI systems and potentially the combination with human decision makers? Which actions are taken upon this?
14. Which control processes and mechanisms are in place to ensure that AI systems are acting upon your set goals? Does this differ depending on the use cases?
15. Was there a portfolio (mix) of controls in place? Did these change over time? How were these controls evaluated? Were they then adjusted based on these evaluation? Give examples of each.
16. During the development of AI systems are formal controls (i.e., explicit instructions for development, operation, and maintenance) or is the process explorative and informal in nature?
17. What technical challenges have you faced adopting AI?

Appendix B: Performance Measures

Knowledge Management Capability (Mao et al., 2016)

To what extent do you agree or disagree with the following sentences? (1 – Completely disagree, 7 – Completely agree)

- KMC1: My organization has processes to gain knowledge on our suppliers, customers, and partners
- KMC2: My organization can generate new knowledge from existing knowledge
- KMC3: My organization has processes in place to distribute knowledge throughout the organization
- KMC4: In my organization, knowledge is accessible to those who need it
- KMC5: My organization has processes for using knowledge to develop new products or services

Organizational Agility (Tallon & Pinsonneault, 2011)

To what extent is your organization able to achieve the following? (1 – To a very low extent, 7 – to a very high extent)

- AG1: React to new product or services launches by competitors
- AG2: Introduce new pricing schedules in response to changes in competitors' prices
- AG3: Expand into new regional or international markets
- AG4: Change (i.e., expand or reduce) the variety of products/services available for sale
- AG5: Adopt new technologies to produce better, faster and cheaper products and services
- AG6: Switch suppliers to avail of lower costs, better quality or improved delivery times

Competitive Performance (Rai & Tang, 2010)

Compared with your key competitors, please indicate how much you agree or disagree with the following statements regarding the degree to which you perform better than them (1 – Totally disagree, 7 – Totally agree)

- PERF1: Decreasing product or service delivery cycle time
- PERF2: Increasing customer satisfaction
- PERF3: In profit growth rates
- PERF4: Providing better product and service quality
- PERF5: Increasing our market share

Appendix C: Reliability, Convergent and Discriminant Validity

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
(1) Explainability	**0.846**									
(2) Traceability	0.541	**0.851**								
(3) Communication	0.487	0.456	**0.814**							
(4) No unfair bias	0.485	0.477	0.432	**0.868**						
(5) Accessibility	0.488	0.521	0.582	0.593	**0.828**					
(6) Accountability	0.442	0.527	0.673	0.546	0.645	**0.811**				
(7) Resilience and security	0.551	0.408	0.448	0.391	0.466	0.510	**0.919**			
(8) Reliability and reproducibility	0.529	0.443	0.498	0.458	0.521	0.671	0.548	**0.826**		
(9) Accuracy	0.421	0.377	0.540	0.319	0.428	0.579	0.430	0.704	**0.898**	
(10) Data privacy	0.483	0.456	0.418	0.463	0.411	0.582	0.618	0.565	0.524	**0.825**
(11) Data quality	0.450	0.479	0.562	0.423	0.433	0.549	0.412	0.472	0.522	0.621
(12) Data access	0.511	0.405	0.449	0.340	0.455	0.489	0.517	0.426	0.469	0.671
(13) Laws and regulations	0.552	0.437	0.337	0.335	0.264	0.306	0.494	0.511	0.513	0.631
(14) Human oversight	0.570	0.439	0.492	0.278	0.397	0.456	0.546	0.567	0.599	0.621
(15) Human well-being	0.504	0.525	0.448	0.528	0.478	0.595	0.542	0.609	0.537	0.631
(16) Environmental and societal well-being	0.423	0.385	0.403	0.523	0.509	0.571	0.398	0.483	0.367	0.381
(17) KMC	0.519	0.479	0.348	0.474	0.432	0.348	0.375	0.376	0.371	0.520
(18) Organizational Agility	0.375	0.350	0.478	0.508	0.540	0.543	0.278	0.366	0.369	0.491
(19) Competitive Performance	0.560	0.427	0.428	0.463	0.442	0.454	0.393	0.411	0.365	0.591
Mean	4.84	5.00	5.03	4.75	4.84	4.92	4.91	5.05	5.13	5.08
Standard Deviation	1.37	1.41	1.47	1.40	1.37	1.44	1.41	1.31	1.27	1.40
AVE	0.716	0.724	0.663	0.753	0.685	0.658	0.844	0.683	0.807	0.681
Cronbach's Alpha	0.801	0.619	0.746	0.836	0.770	0.869	0.816	0.845	0.761	0.842
Composite Reliability	0.883	0.840	0.855	0.901	0.867	0.906	0.916	0.896	0.893	0.895

	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)
(1) Explainability									
(2) Traceability									
(3) Communication									
(4) No unfair bias									
(5) Accessibility									
(6) Accountability									
(7) Resilience and security									
(8) Reliability and reproducibility									
(9) Accuracy									
(10) Data privacy									
(11) Data quality	**0.852**								
(12) Data access	0.640	**0.869**							
(13) Laws and regulations	0.554	0.606	**0.921**						
(14) Human oversight	0.541	0.606	0.748	**0.865**					
(15) Human well-being	0.555	0.627	0.652	0.687	**0.858**				
(16) Environmental and societal well-being	0.440	0.380	0.287	0.366	0.561	**0.899**			
(17) KMC	0.536	0.530	0.601	0.551	0.560	0.376	**0.804**		
(18) Organizational Agility	0.491	0.467	0.400	0.416	0.512	0.510	0.606	**0.785**	
(19) Competitive Performance	0.605	0.568	0.647	0.583	0.551	0.508	0.718	0.692	**0.775**
Mean	5.12	5.05	5.22	5.07	5.00	4.76	5.06	4.94	5.03
Standard Deviation	1.29	1.46	1.49	1.45	1.45	1.58	1.33	1.38	1.36
AVE	0.725	0.755	0.847	0.748	0.736	0.807	0.646	0.616	0.601
Cronbach's Alpha	0.810	0.676	0.910	0.831	0.821	0.881	0.864	0.875	0.834
Composite Reliability	0.888	0.860	0.943	0.899	0.893	0.926	0.901	0.906	0.883

Appendix D: Cross-Loadings

EX – Explainability, TR – Traceability, CO – Communication, BI – No unfair bias, ACCE – Accessibility, ACCOU – Accountability, RS – Resilience and security, RR – Reliability and Reproducibility, AC – Accuracy, DP – Data privacy, DQ – Data quality, DA – Data access, LR – Laws and regulation, HO – Human oversight, HWB – Human well-being, ESWB – Environmental and societal well-being, KMC – Knowledge management capability, AGIL – Organizational agility, PERF – Competitive Performance

	EX	TR	CO	BI	ACCE	ACCOU	RS	RR	AC	DP
EX1	**0.782**	0.372	0.329	0.373	0.359	0.335	0.460	0.399	0.236	0.303
EX2	**0.860**	0.414	0.410	0.386	0.353	0.322	0.455	0.429	0.340	0.451
EX3	**0.892**	0.567	0.482	0.465	0.513	0.453	0.485	0.505	0.468	0.458
TR1	0.457	**0.838**	0.345	0.410	0.371	0.340	0.315	0.311	0.259	0.330
TR2	0.463	**0.863**	0.428	0.403	0.511	0.548	0.377	0.438	0.378	0.443
CO1	0.440	0.373	**0.805**	0.183	0.375	0.490	0.322	0.428	0.471	0.360
CO2	0.382	0.342	**0.835**	0.347	0.539	0.580	0.403	0.402	0.427	0.361
CO3	0.365	0.397	**0.803**	0.532	0.512	0.576	0.371	0.384	0.421	0.299
BI1	0.449	0.351	0.431	**0.864**	0.477	0.434	0.330	0.368	0.280	0.391
BI2	0.449	0.449	0.345	**0.893**	0.510	0.462	0.318	0.378	0.253	0.406
BI3	0.365	0.440	0.350	**0.846**	0.554	0.525	0.369	0.445	0.297	0.410
ACCE1	0.358	0.451	0.504	0.399	**0.773**	0.508	0.315	0.425	0.408	0.352
ACCE2	0.456	0.432	0.473	0.557	**0.871**	0.560	0.394	0.438	0.380	0.362
ACCE3	0.392	0.417	0.476	0.504	**0.835**	0.534	0.442	0.435	0.283	0.308
ACCOU1	0.349	0.422	0.500	0.497	0.566	**0.761**	0.391	0.443	0.324	0.338
ACCOU2	0.358	0.445	0.555	0.486	0.488	**0.847**	0.397	0.570	0.459	0.487
ACCOU3	0.315	0.389	0.498	0.416	0.533	**0.858**	0.410	0.594	0.487	0.501
ACCOU4	0.351	0.521	0.572	0.438	0.577	**0.844**	0.402	0.535	0.520	0.552
ACCOU5	0.416	0.349	0.595	0.389	0.456	**0.738**	0.467	0.568	0.533	0.456
RS1	0.535	0.381	0.396	0.368	0.372	0.435	**0.920**	0.508	0.401	0.562
RS2	0.477	0.369	0.428	0.350	0.485	0.503	**0.918**	0.499	0.388	0.575
RR1	0.466	0.401	0.394	0.341	0.418	0.544	0.539	**0.829**	0.618	0.492
RR2	0.409	0.354	0.356	0.328	0.357	0.520	0.367	**0.815**	0.590	0.430
RR3	0.449	0.326	0.432	0.404	0.464	0.576	0.494	**0.832**	0.498	0.449
RR4	0.422	0.379	0.463	0.442	0.482	0.578	0.406	**0.828**	0.619	0.494
AC1	0.325	0.353	0.423	0.227	0.370	0.498	0.305	0.627	**0.890**	0.428
AC2	0.428	0.325	0.544	0.342	0.398	0.541	0.461	0.638	**0.907**	0.510
DP1	0.423	0.339	0.453	0.518	0.401	0.583	0.606	0.515	0.487	**0.744**
DP2	0.340	0.363	0.289	0.364	0.279	0.439	0.435	0.484	0.437	**0.845**
DP3	0.427	0.417	0.336	0.333	0.371	0.476	0.516	0.470	0.408	**0.877**
DP4	0.406	0.385	0.309	0.326	0.308	0.431	0.490	0.400	0.401	**0.829**

DQ1	0.416	0.406	0.453	0.278	0.264	0.394	0.313	0.425	0.483	0.511
DQ2	0.353	0.441	0.480	0.396	0.368	0.428	0.363	0.333	0.398	0.542
DQ3	0.381	0.377	0.501	0.405	0.472	0.578	0.375	0.448	0.453	0.533
DA1	0.437	0.366	0.379	0.270	0.350	0.375	0.402	0.327	0.445	0.567
DA2	0.452	0.337	0.402	0.321	0.444	0.478	0.499	0.415	0.368	0.601
LR1	0.512	0.380	0.279	0.224	0.197	0.235	0.487	0.473	0.502	0.591
LR2	0.513	0.415	0.305	0.365	0.244	0.284	0.425	0.460	0.460	0.573
LR3	0.498	0.411	0.344	0.334	0.286	0.325	0.451	0.479	0.455	0.579
HO1	0.522	0.408	0.459	0.256	0.396	0.412	0.494	0.433	0.529	0.527
HO2	0.512	0.360	0.402	0.222	0.283	0.352	0.461	0.512	0.557	0.541
HO3	0.445	0.371	0.414	0.244	0.348	0.415	0.461	0.528	0.471	0.542
HWB1	0.524	0.455	0.349	0.394	0.351	0.448	0.472	0.611	0.516	0.499
HWB2	0.368	0.429	0.346	0.520	0.384	0.544	0.385	0.474	0.380	0.549
HWB3	0.396	0.466	0.453	0.461	0.494	0.551	0.527	0.477	0.476	0.581
ESWB1	0.391	0.347	0.379	0.473	0.463	0.487	0.376	0.377	0.289	0.310
ESWB2	0.308	0.313	0.336	0.490	0.426	0.471	0.338	0.360	0.275	0.254
ESWB3	0.428	0.372	0.368	0.449	0.477	0.569	0.357	0.545	0.410	0.443
KMC1	0.408	0.392	0.335	0.470	0.380	0.359	0.304	0.241	0.327	0.542
KMC2	0.418	0.354	0.251	0.314	0.213	0.256	0.197	0.291	0.329	0.388
KMC3	0.438	0.370	0.266	0.406	0.400	0.287	0.376	0.283	0.295	0.394
KMC4	0.472	0.412	0.255	0.423	0.378	0.285	0.330	0.397	0.262	0.382
KMC5	0.348	0.399	0.286	0.266	0.354	0.191	0.294	0.314	0.275	0.360
AGIL1	0.326	0.298	0.364	0.380	0.423	0.437	0.219	0.331	0.354	0.424
AGIL2	0.239	0.256	0.332	0.350	0.363	0.427	0.141	0.301	0.338	0.385
AGIL3	0.251	0.172	0.269	0.356	0.459	0.397	0.210	0.253	0.202	0.373
AGIL4	0.353	0.326	0.478	0.442	0.534	0.470	0.276	0.326	0.291	0.412
AGIL5	0.297	0.254	0.401	0.430	0.406	0.374	0.256	0.289	0.310	0.401
AGIL6	0.292	0.335	0.404	0.433	0.366	0.454	0.206	0.216	0.229	0.311
PERF1	0.477	0.379	0.244	0.400	0.314	0.243	0.317	0.291	0.280	0.431
PERF2	0.488	0.281	0.377	0.437	0.345	0.380	0.361	0.337	0.347	0.576
PERF3	0.421	0.327	0.261	0.369	0.342	0.381	0.371	0.392	0.255	0.458
PERF4	0.472	0.445	0.375	0.282	0.339	0.364	0.296	0.329	0.321	0.404
PERF5	0.310	0.228	0.384	0.314	0.376	0.396	0.188	0.260	0.203	0.421

	DQ	DA	LR	HO	HWB	ESWB	KMC	AGIL	PERF
EX1	0.253	0.330	0.331	0.360	0.291	0.375	0.300	0.278	0.399
EX2	0.434	0.458	0.488	0.504	0.429	0.277	0.506	0.321	0.490
EX3	0.436	0.495	0.558	0.562	0.534	0.419	0.492	0.347	0.522
TR1	0.344	0.354	0.361	0.367	0.440	0.290	0.399	0.241	0.333
TR2	0.467	0.337	0.382	0.381	0.453	0.363	0.416	0.350	0.392
CO1	0.540	0.441	0.396	0.471	0.381	0.252	0.319	0.394	0.355
CO2	0.395	0.389	0.207	0.381	0.333	0.340	0.224	0.342	0.346
CO3	0.435	0.263	0.214	0.347	0.379	0.394	0.307	0.432	0.343
BI1	0.396	0.330	0.334	0.272	0.464	0.436	0.432	0.438	0.450
BI2	0.297	0.277	0.269	0.218	0.468	0.442	0.414	0.430	0.382
BI3	0.411	0.278	0.271	0.236	0.443	0.483	0.389	0.453	0.375
ACCE1	0.365	0.368	0.239	0.379	0.375	0.446	0.379	0.428	0.346
ACCE2	0.360	0.414	0.276	0.313	0.441	0.378	0.357	0.470	0.417
ACCE3	0.354	0.348	0.140	0.303	0.368	0.449	0.342	0.444	0.331
ACCOU1	0.361	0.251	0.069	0.220	0.382	0.547	0.187	0.395	0.280
ACCOU2	0.422	0.368	0.209	0.298	0.497	0.502	0.271	0.402	0.339
ACCOU3	0.398	0.409	0.226	0.366	0.443	0.409	0.284	0.454	0.368
ACCOU4	0.508	0.446	0.333	0.467	0.569	0.518	0.339	0.508	0.440
ACCOU5	0.515	0.483	0.368	0.465	0.499	0.349	0.311	0.429	0.393
RS1	0.378	0.470	0.470	0.519	0.487	0.371	0.334	0.253	0.351
RS2	0.379	0.481	0.437	0.485	0.510	0.361	0.355	0.258	0.371
RR1	0.387	0.331	0.521	0.540	0.526	0.364	0.410	0.260	0.424
RR2	0.413	0.330	0.313	0.397	0.424	0.412	0.242	0.270	0.286
RR3	0.344	0.355	0.444	0.460	0.524	0.390	0.260	0.325	0.290
RR4	0.416	0.391	0.403	0.470	0.536	0.434	0.323	0.354	0.353
AC1	0.429	0.415	0.382	0.481	0.481	0.294	0.269	0.302	0.266
AC2	0.506	0.427	0.533	0.592	0.484	0.363	0.394	0.359	0.385
DP1	0.502	0.542	0.450	0.458	0.555	0.497	0.415	0.448	0.508
DP2	0.515	0.526	0.488	0.453	0.494	0.252	0.442	0.352	0.464
DP3	0.490	0.570	0.560	0.569	0.533	0.248	0.449	0.429	0.433
DP4	0.543	0.576	0.579	0.563	0.504	0.274	0.410	0.393	0.548
DQ1	**0.865**	0.507	0.489	0.463	0.413	0.271	0.452	0.382	0.482
DQ2	**0.868**	0.533	0.491	0.438	0.437	0.379	0.472	0.434	0.522
DQ3	**0.821**	0.593	0.435	0.482	0.566	0.471	0.444	0.437	0.541
DA1	0.623	**0.876**	0.547	0.535	0.545	0.296	0.498	0.378	0.462
DA2	0.486	**0.862**	0.505	0.518	0.545	0.366	0.422	0.435	0.528
LR1	0.519	0.554	**0.918**	0.751	0.584	0.242	0.525	0.298	0.565
LR2	0.506	0.572	**0.929**	0.659	0.592	0.257	0.625	0.411	0.607
LR3	0.505	0.548	**0.914**	0.657	0.623	0.292	0.512	0.394	0.613
HO1	0.481	0.532	0.641	**0.901**	0.591	0.328	0.503	0.407	0.521
HO2	0.445	0.478	0.644	**0.860**	0.546	0.282	0.503	0.363	0.564

HO3	0.477	0.559	0.654	**0.832**	0.645	0.338	0.425	0.310	0.430
HWB1	0.411	0.498	0.651	0.684	**0.860**	0.456	0.509	0.365	0.454
HWB2	0.455	0.451	0.465	0.428	**0.818**	0.541	0.406	0.537	0.478
HWB3	0.563	0.654	0.548	0.631	**0.894**	0.459	0.518	0.438	0.490
ESWB1	0.382	0.331	0.244	0.348	0.470	**0.902**	0.361	0.434	0.458
ESWB2	0.338	0.330	0.216	0.252	0.466	**0.902**	0.273	0.430	0.400
ESWB3	0.454	0.359	0.303	0.374	0.564	**0.892**	0.371	0.502	0.500
KMC1	0.478	0.486	0.470	0.467	0.470	0.344	**0.805**	0.639	0.613
KMC2	0.442	0.368	0.477	0.392	0.382	0.224	**0.804**	0.481	0.540
KMC3	0.389	0.443	0.488	0.474	0.451	0.320	**0.830**	0.488	0.599
KMC4	0.406	0.369	0.495	0.465	0.504	0.348	**0.806**	0.409	0.571
KMC5	0.434	0.455	0.493	0.410	0.442	0.264	**0.774**	0.388	0.553
AGIL1	0.415	0.453	0.359	0.365	0.422	0.438	0.482	**0.825**	0.629
AGIL2	0.440	0.291	0.332	0.313	0.438	0.384	0.427	**0.776**	0.545
AGIL3	0.284	0.316	0.271	0.316	0.362	0.379	0.417	**0.776**	0.516
AGIL4	0.399	0.411	0.257	0.312	0.394	0.391	0.462	**0.783**	0.499
AGIL5	0.389	0.402	0.435	0.387	0.439	0.414	0.536	**0.797**	0.555
AGIL6	0.377	0.310	0.214	0.256	0.350	0.391	0.526	**0.751**	0.503
PERF1	0.374	0.406	0.487	0.491	0.439	0.407	0.631	0.428	**0.732**
PERF2	0.506	0.511	0.587	0.485	0.474	0.414	0.612	0.588	**0.832**
PERF3	0.450	0.424	0.463	0.378	0.434	0.491	0.461	0.446	**0.730**
PERF4	0.501	0.413	0.587	0.525	0.446	0.348	0.604	0.547	**0.821**
PERF5	0.511	0.448	0.370	0.369	0.343	0.331	0.460	0.656	**0.756**

Appendix E: Heterotrait-Monotrait Ratio (HTMT)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
(1) Explainability									
(2) Traceability	0.758								
(3) Communication	0.622	0.668							
(4) No unfair bias	0.590	0.663	0.552						
(5) Accessibility	0.612	0.754	0.775	0.733					
(6) Accountability	0.525	0.710	0.835	0.644	0.791				
(7) Resilience and security	0.683	0.572	0.576	0.473	0.586	0.606			
(8) Reliability and reproducibility	0.638	0.608	0.626	0.545	0.647	0.781	0.658		
(9) Accuracy	0.525	0.546	0.713	0.397	0.563	0.704	0.541	0.878	
(10) Data privacy	0.583	0.629	0.531	0.557	0.514	0.678	0.750	0.672	0.655
(11) Data quality	0.550	0.673	0.721	0.514	0.549	0.648	0.506	0.571	0.663
(12) Data access	0.687	0.627	0.631	0.454	0.633	0.633	0.698	0.565	0.652
(13) Laws and regulations	0.636	0.582	0.406	0.384	0.315	0.334	0.573	0.580	0.613
(14) Human oversight	0.690	0.612	0.623	0.334	0.500	0.526	0.663	0.675	0.752
(15) Human well-being	0.604	0.735	0.570	0.645	0.600	0.703	0.657	0.727	0.674
(16) Environmental and societal well-being	0.497	0.516	0.497	0.610	0.619	0.651	0.469	0.551	0.439
(17) KMC	0.615	0.655	0.431	0.550	0.529	0.390	0.445	0.442	0.453
(18) Organizational Agility	0.444	0.470	0.591	0.594	0.660	0.621	0.329	0.424	0.448
(19) Competitive Performance	0.681	0.595	0.537	0.558	0.552	0.530	0.480	0.492	0.453

	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)
(1) Explainability										
(2) Traceability										
(3) Communication										
(4) No unfair bias										
(5) Accessibility										
(6) Accountability										
(7) Resilience and security										
(8) Reliability and reproducibility										
(9) Accuracy										
(10) Data privacy										
(11) Data quality	0.753									
(12) Data access	0.892	0.862								
(13) Laws and regulations	0.720	0.645	0.772							
(14) Human oversight	0.741	0.659	0.807	0.860						
(15) Human well-being	0.763	0.680	0.836	0.747	0.819					
(16) Environmental and societal well-being	0.439	0.514	0.491	0.316	0.421	0.660				
(17) KMC	0.604	0.639	0.689	0.680	0.649	0.659	0.423			
(18) Organizational Agility	0.572	0.581	0.604	0.444	0.486	0.612	0.576	0.687		
(19) Competitive Performance	0.708	0.735	0.758	0.739	0.698	0.668	0.595	0.839	0.803	

Ida Merete Enholm

Responsible AI Governance and Its Effect on Competitive Performance

NTNU
Norwegian University of
Science and Technology