Tampere University

Anttoni Tukia

DATA GOVERNANCE FOR SUSTAINABLE ARTIFICIAL INTELLIGENCE

Master of Science Thesis
Faculty of Management and Business
December 2022

ABSTRACT

Anttoni Tukia: Data Governance for Sustainable AI
Master's Thesis
Tampere University
Master's Programme in Information and Knowledge Management
Examiners: Professor Samuli Pekkola and Doctoral Researcher Pasi Raatikainen
December 2022

Rapid growth of big data and processing power in recent years has caused an upsurge in artificial intelligence (AI) in numerous domains. It has brought great benefits to our society, but also caused several sustainability and ethical issues. There is a consensus among researchers that the harms caused by AI must be mitigated, and proper AI governance has been identified as an important part of it. AI uses large quantities of data to work, making some aspects of data governance relevant for doing this. Still, studies on how data governance can help solve these issues remain scarce.

The aim of this thesis is to examine how AI can be positively influenced towards sustainability by the means of data governance. Another goal of this thesis is to understand the role of data in sustainable AI. First, a literature review on the topics of data governance, sustainable AI, and AI governance was conducted, and a theory-based data governance for sustainable AI was formed. This framework was then refined in the empirical part by organising two workshops for the data and AI governance experts from the case company, Solita.

The most significant data-related challenge regarding AI identified in this research is how to ensure data quality. If this is not done, the outcome of the AI algorithms using it may become biased or skewed. Additionally, the correctness and transparency of data acquisition must be in place to prevent the misuse of data, especially when it is personally identifiable, sensitive, private, or confidential. These matters are not only relevant for the input data, but for the output as well. To prevent biased or skewed output data from causing damage, it should be properly managed and governed. Furthermore, the data supply chains around AI systems should be accompanied with clear and continuous chains of accountability and responsibility. However, if a biased or skewed outcome is produced by a flaw in the algorithm itself, such as a programming or design error, the issue may need to be addressed by some other way than ensuring data quality.

In this research, a data governance for sustainable AI framework was formed as the main result. Its objective is to be a high-level abstraction of what needs to be considered when supporting sustainable AI by the means of data governance, and to illustrate how different data governance activities support the goals of different AI governance elements. There are three AI governance elements selected for the framework, *AI System*, *Organisation*, and *Ecosystem*, as well as seven data governance activities, (1) *Objectives & Key Results*, (2) *Decision Rights and Accountabilities*, (3) *Data Policies, Rules, Definitions, and Standards*, (4) *Roles and Responsibilities*, (5) Data Processes, (6) Data Governance Metrics, and (7) *Controls*. The numbers indicate the order in which the data governance activities should be done.

In addition to contributing to the academic discussion in the field of sustainable AI, this research has practical implications for the case company. Thus, strengthening the role of private businesses in guiding AI towards sustainability is another contribution of this research.

Keywords: data governance, AI governance, data quality, sustainable AI, AI ethics

The originality of this thesis has been checked using the Turnitin OriginalityCheck service.

TIIVISTELMÄ

Anttoni Tukia: Data governance kestävässä tekoälyssä
Diplomityö
Tampereen yliopisto
Tietojohtamisen DI-ohjelma
Tarkastajat: professori Samuli Pekkola ja väitöskirjatutkija Pasi Raatikainen
Joulukuu 2022

Data määrän ja prosessointitehon nopea kasvu viime vuosina on vauhdittanut tekoälyn (AI) hyödyntämistä monilla aloilla. Tämä on tuonut sekä suuria hyötyjä, että aiheuttanut useita kestävyyteen ja eettisyyteen liittyviä haasteita. Tutkijoiden keskuudessa vallitsee yksimielisyys siitä, että tekoälyn aiheuttamat haitat tulisi paremmin hallita, ja että yksi tärkeä osa tätä on AI governance. Tekoäly käyttää suuria määriä dataa, ja tästä syystä jotkin data governancen osa-alueet ovat tärkeitä kestävän tekoälyn kannalta. Siitä huolimatta tutkimukset siitä, kuinka data governance auttaa näissä haasteissa ovat todella harvassa.

Tämän diplomityön tavoitteena on tutkia, kuinka tekoälyn kestävyyteen voidaan vaikuttaa positiivisesti data governancen keinoin, sekä ymmärtää datan roolia kestävässä tekoälyssä. Aluksi tehtiin kirjallisuuskatsaus aiheista data governance, kestävä tekoäly ja AI governance, jonka pohjalta muodostettiin teoriapohjainen data governance viitekehys kestävälle tekoälylle. Tämän jälkeen viitekehystä jatkokehitettiin empiirisessä osassa järjestämällä kaksi työpajaa case-yritys Solitan data ja AI governance asiantuntijoiden kanssa.

Merkittävin tutkimuksessa havaittu tekoälyn datan käyttöön liittyvä haaste on datan laadun varmistaminen, sillä huono datan laatu voi johtaa vääristymiin tekoälyalgoritmien tuloksissa. Lisäksi datan keruun tulee olla korrektia ja läpinäkyvää, jotta datan väärinkäyttö voidaan estää. Tämä on erityisen tärkeää silloin, kun kerätään tai käsitellään arkaluonteista, yksityistä tai luottamuksellista dataa. Nämä asiat eivät koske vain tekoälyalgoritmien syötedataa, vaan myös niiden ulostuloa. Jotta voidaan estää huonosta syötedatasta johtuvia vääristymiä aiheuttamasta vahinkoa, tekoälyalgoritmien ulostuloadatan laatu tulee myös varmistaa. Lisäksi tekoälyjärjestelmiä ympäröiviin datatoimitusketjuihin tulisi liittää selkeät ja jatkuvat vastuun ketjut. Mikäli tekoälyalgoritmin vääristynyt ulostulo johtuu virheestä itse algoritmissa, kuten ohjelmointivirheestä, sen korjaaminen tulee todennäköisesti tehdä jollain muulla keinolla kuin datan laadun parantamisella.

Tutkimuksen päätuloksena muodostettiin kestävää tekoälyä tukeva data goverance viitekehys. Sen tarkoituksena on olla korkean tason abstraktio siitä, mitä tulee ottaa huomioon kestävän tekoälyn tukemisessa data governancen keinoin, sekä havainnollistaa, kuinka data governance aktiviteetit tukevat AI governancen eri osa-alueita. Viitekehykseen valittiin kolme AI governance elementtiä, *Tekoälyjärjestelmä*, *Organisaatio* ja *Ekosysteemi*, sekä seitsemän data governance aktiviteettia, (1) *Tavoitteet*, (2) *Päätöksenteko-oikeudet ja vastuuvelvollisuudet*, (3) *Käytännöt, säännöt, määritelmät ja standardit*, (4) *Roolit ja vastuualueet*, (5) *Dataprosessit*, (6) *Mittarit*, ja (7) *Valvonta*. Numerot osoittavat, missä järjestyksessä mikäkin data governance aktiviteetti tulee suorittaa.

Sen lisäksi, että tämä työ edistää akateemista keskustelua kestävästä tekoälystä, sillä on myös käytännön arvoa case-yritykselle. Näin ollen tämä työ on myös kontribuutio yksityisten yritysten roolin kasvattamisessa tekoälyn kestävässä kehityksessä.

Avainsanat: data governance, AI governance, datan laatu, kestävä tekoäly, eettinen tekoälyn

Tämän julkaisun alkuperäisyys on tarkastettu Turnitin OriginalityCheck –ohjelmalla.

PREFACE

During my time at Tampere University, I often heard people saying that writing a diploma thesis feels like a mountain to climb, and that it drains you from energy and motivation. For me, this was not the case. I found both data governance and AI sustainability interesting, and by properly organising my work I had no trouble finishing my thesis.

I want to thank my employer Solita for giving me the opportunity to write my thesis on this topic, and the examiner of this thesis, professor Samuli Pekkola, for guiding me through the process. I also want to give special thanks to my colleagues Anna Metsäranta, Manu Setälä, Jeremias Virtanen, Simo Hokkanen, Minna Kärhä, Juha-Pekka Joutsenlahti, Outi Ainasoja, and Lauri Eneh for participating the workshops conducted during this research, and for giving me support, guidance, and feedback.

Tampere, 3 November 2022

Anttoni Tukia

CONTENTS

1. INTRODUCTION

This study examines how artificial intelligence (later referred to as AI) can be positively influenced by the means of data governance. AI is a technology that incorporates algorithms, data, and computing power. Due to the recent development of computing power and the availability of great amounts of data, AI has become a big part of our everyday life, being embedded in most digital technologies that we use daily, and for instance, influencing what kind of information you see online. It has the potential to change our lives in a positive way by improving healthcare, increasing efficiency processes, contributing to the fight against climate change, etc. (Coeckelbergh 2020; European Commission 2020)

While AI has undoubtably presented outstanding opportunities, it has also managed to create numerous economic and social risks (Wirtz et al. 2022). As a response to the explosion of AI, there is an increasing need to the study of social and ethical issues caused by AI, such as violations of citizen privacy, unfairness in decision-making, untransparent black-box systems, and unclear responsibilities and accountabilities (van Wynsberghe 2021; Zuiderwijk et al. 2021). Modern AI algorithms driven by big data have become more autonomous, having now the potential to systematically introduce bias, reinforce discrimination and unwanted behaviour, as well as favour a political orientation (Janssen & Kuk 2016). Thus, there is an increasing demand for managing risks brought by AI systems and enforcing AI-related ethical principles (Mäntymäki et al. 2022a).

Although the risks and potential negative impacts of AI are widely known, regulating AI, and implementing AI governance is still widely neglected. Technologies alone cannot act as moral agents, making us the people as policy makers, the ones responsible for ensuring the sustainability of AI. Since AI involves gathering and processing data, data governance (the exercise of authority, control and shared decision-making on an organisations data assets) has been identified as one of the requirements for sustainable AI. Most of the AI governance frameworks incorporate data governance as a part of overall AI governance, but there are hardly any good practices or frameworks to successfully adopt and apply data governance for AI. Therefore, this study aims to be a contribution to the quest for creating data governance best practices and frameworks for sustainable

AI. (HLEG 2019; Gasser & Almeida 2017; Janssen et al. 2020; Ladley 2012; Wirtz et al. 2022)

1.1 Research objectives and limitations

The objective of this research is to examine how to positively influence sustainable AI governance by practicing data governance. To do this, one must first understand how AI systems are linked to data, and what kind of challenges arise from this. This papers main research question is the following:

How to support sustainable AI with data governance?

To answer this question, we must find the answer to the preliminary research question:

What is the role of data in sustainable AI?

Whereas most of the data governance -related research concerns data in general, the focus of this research is solely on data in the context of AI. And with AI, the focus is on the sustainability aspects of it. Sustainable AI refers to the ecological, economic, and social sustainability of AI. A wider description of sustainable AI can be found in chapter 2.2. Data governance mostly concerns the technical layer on AI governance. Thus, a limitation of this study is that we cannot cover the entire field of AI governance with data governance since it represents only a part of an AI system. While data governance plays a big part in AI governance, it is not enough to cover AI governance unaccompanied. (Mäntymäki et al. 2022a)

The case company for the empirical part is a Finnish information technology (later referred to as IT) company called Solita. The company offers technology, data and business design solutions for businesses and public administration organisations. As a company, Solita is suitable for this research because it has started investing into AI governance expertise from the beginning of 2022 and aims to significantly invest more in the future. The company already has expertise on data governance, but not that much in AI governance. Solita has been developing competence in the field of AI governance, for instance by participating in the AIGA project led by University of Turku. The company is interested in making AI governance and sustainable AI one of their new business areas, and by participating in the AIGA project and by commissioning theses around the topic, they aim to gain knowledge and competitive advantage. One of the goals of this research is to help Solita achieve this.

1.2 Research methodology

Research methodology examines the choices regarding the philosophy, theory development approach, research methodology, strategy, time horizon, and data collection & analysis methods (Saunders et al. 2019). Table 1 below presents the methodological choices of this research.

Table 1. *Research methodical choices*

Aspect	**Research choices**
Philosophy	Pragmatism
Approach to theory development	Abduction
Methodological choice	Mono method qualitative
Strategy	Case Study
Time horizon	Cross-sectional
Data collection method	Group-interview
Data analysis method	Thematic data analysis

The first aspect of the research methodology table above, philosophy, refers to beliefs and assumptions regarding knowledge development. Whether we are aware of it or not, we tend to understand facts and realities differently, and may approach the same problems differently as well as make dissimilar conclusions from the same facts and realities. Research philosophies help take human subjectivity into account by identifying and keeping them in mind. The goal of this research is to develop a practical data governance framework for sustainable AI for an IT-company, and therefore pragmatism was the choice form research philosophy. Pragmatism aims to develop knowledge that contributes to a practical solution which supports future action (Saunders et al. 2019).

One of the central concepts of the sciences philosophy of reasoning is the approach to theory development. There are three different ways to do this: inductive, deductive, and abductive reasoning. This research uses abductive reasoning as the approach for theory development. Where inductive reasoning aims to create a theory from observations, and deductive reasoning tests a theory with observations, abductive aims to combine these. This research aims to create a theoretical data governance framework for AI governance, get feedback from it, and develop it based on observations. (Anttila 1998; Sanders et al. 2019)

The methodology of this research is qualitative. As a research method, qualitative study focuses on understanding the research phenomenon from the people’s point of view around the phenomenon. This means we are interested in understanding the experiences, thoughts, and feelings of those people. It is undoubtably difficult to get inside people's heads and experience the same things as they do, but qualitative research methods give us the tools to get as close to achieving this as we can. Qualitative research typically produces rich and detailed information about the phenomenon to be researched on by interpreting and understanding the actions and interactions of the people around it. The success of this depends highly on how well the researcher gets inside the social worlds, where the phenomenon happens. (Juuti & Puusa 2020)

The strategy of this research is case study. Case studies examine a single example of a phenomenon, providing hypotheses in the preliminary stages of investigations, which can be tested later more systematically (Abercrombie et al. 1984). The strengths of case studies are that they enable the study of a phenomena in its natural setting gaining understanding through practice, provide a relatively comprehensive understanding of the entire complexity of the phenomenon, and are suitable for an early stage, exploratory investigation (Meredith 1998). AI governance and data governance in sustainable AI are new phenomena, making case study a suitable research strategy. Additionally, these topics are under exploratory investigation in our case company.

An important part of research design is choosing a time horizon. There are two options for this, cross-sectional that examines a phenomenon in a particular time, and longitudinal where the examination happens over a longer period (Saunders et al. 2019). The time horizon of this study was chosen to be cross-sectional, since the plan was to understand a phenomenon in the context of the case company based on interviews conducted within a short timeframe.

This research uses group interviews as the data collection method, workshops to be more specific. The benefit of group-discussions is that it may lead to diverse conversations from many points of view, where participants evaluate points made by the group and challenge each other's views (Saunders et al. 2019). Group interviews are a good way to spark discussion on different views and opinions and, therefore they are a suitable data collection method for this study. A more detailed depiction of the data collection and interview settings can be found in chapter 3.

For data analysis, qualitative thematic analysis method was chosen. Thematic analysis is thought to be the general approach to qualitative data analysis. It offers at the same time a systematic and flexible way to analyse qualitative data and works well for large and small datasets. It also helps to develop and test theories based on relationships and thematic patterns, which makes it befitting for this research. A more detailed depiction of the data analysis of this research can be found in chapter 3. (Saunders et al. 2009)

1.3 Structure of the thesis

The structure of this thesis is as follows. In the first chapter, the topic and the motivation for this research are introduced, and the research objective described by introducing the two research questions this study aims to answer. Additionally, the limitations and the methodology of this research are briefly introduced, which are described in more detail in chapters three and six.

Data governance and sustainable AI theory is explained in chapter 2. This chapter was also meant to be a literature review on these topics, and with the information gathered in the process, a theory-based data governance for sustainable AI -framework was formed. The refining of the framework is done in the empirical part of this research.

The third chapter describes the research methodology and settings for the empirical research in more details. First, the research approach is described. The second part describes the plan on how to collect and analyse empirical data. Lastly, the reliability and validity of this research is discussed.

The findings of the empirical research are presented in chapter 4. First, the two workshops used in data gathering are described. Then, the key findings from them are exhibited. Lastly, the evolution of the data governance for sustainable AI framework is depicted.

Chapter 5 summarises the conclusions synthesised from the literature and the empirical research and comprises of two parts. The first part answers the preliminary research question, what is the role of data in sustainable AI. The main research question, how to support sustainable AI with data governance, is answered in the second part.

The sixth chapter discusses the possible impact of this research on AI, and its limitations. Additionally, further research directions derived from this research collected from literature are suggested.

2. THEORETICAL FRAMEWORK

In this chapter, a theory-based data governance for sustainable AI framework is formed. This chapter has three parts. The theory behind data governance is explained in the first part, aiming to give an understanding of what data governance is and what it consists of. Second, the concepts of AI, sustainability, AI governance, and data challenges in AI are discussed. Lastly, a synthesis of data governance for sustainable AI is made.

2.1 Data governance

In information sciences, it is essential to understand the functional differences between data, information, and knowledge (Vilminko-Heikkinen 2017). DAMA (2010) offers the following definitions:

Data is a representation of facts, that is stored as text, numbers, images, sound, videos etc. It is the raw material for creating information.

Information is data put in context, such as business meaning, format, timeframe, or relevance.

Knowledge is information in perspective. It is awareness, understanding, cognizance, and recognition of a situation in its complexity.

Governance in general refers to the decisions made to ensure successful management and use of a decision domain. It also addresses who is accountable for these decisions. (Khatri & Brown 2010)

The formal definition of data governance is “The exercise of authority, control, and shared decision-making (planning, monitoring, and enforcement) over the management of data assets” (DAMA 2010). Another definition for data governance would be organising and implementing procedures, policies, structures, roles, and responsibilities to enforce and outline rules, decision rights, and responsibilities for effective data management (Ladley 2012). Although different authors seem to have slightly dissimilar definitions for data governance, the common understanding is that data governance in there to ensure that data management happens appropriately. Therefore, data governance should not be mixed with data management. The relationship between data assets, data management and data governance is illustrated in figure 1 as “the governance v”.

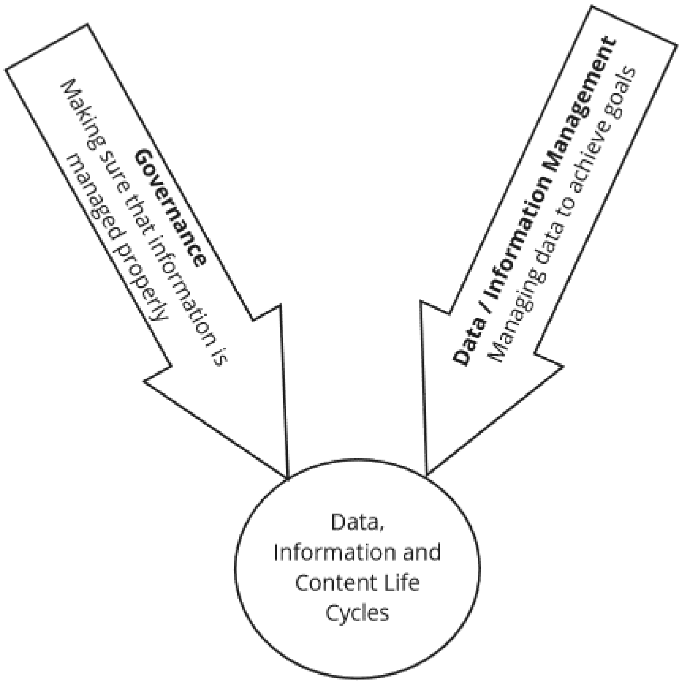

Figure1. *Governance v. (based on Ladley 2012)*

As companies grow, the amount of data they collect grows too. This has sparked a dramatic increase in data utilisation in organizations, making data a critical factor in business operations (Alhassan et al. 2019). When this happens, it is common that they rely more on quick fix solutions, that only servers the individual business units, forming data silos (Sarsfield 2009). As the issues unfold, these companies often find themselves in need of an enterprise-wide data governance program to help them manage data just like any other asset (Ladley 2012; Sarsfield 2009).

An issue regarding the ever-growing amount of data, that heavily concerns this research, are ethical concerns around it (Eryurek et al. 2021). This problem arises when we combine AI and machine learning with data to make data-driven decisions. An example of this, which caused the resignation of Dutch government in 2021 was their child benefit algorithm that wrongfully accused 20 000 families of fraud based on racial profiling (Amnesty International 2021; Henley 2021). These kinds of problems make it necessary for companies to have data governance beyond regulatory requirements (Eryurek et al. 2021).

In the next chapters, the following components of data governance are examined: data management, business processes, compliance and procedure management, and people management.

2.1.1 Data management

Data management is the business function to plan, control and deliver data and information assets. Data management consists of executing, developing, and supervising the plans, practices, policies, programs, projects, processes, and procedures determined by data governance. Its aim is to protect, deliver, enhance, and control the value of an organisation's data and information assets. There are several areas that data management must cover: data quality, data architecture, master data management, and metadata. (DAMA 2010; Ladley 2012)

The first thing that data management must support is good data quality. This means that the data needs to be accurate, complete, timely and consistent (DAMA 2010). Data quality can be assessed by its availability, presentation, relevance, reliability, usability, and quality (Cai & Chu 2015). Another way to assess data quality is to evaluate its accuracy, timeliness, completeness, and credibility (Khatri & Brown 2010). As a summary of all these definitions, data quality can be thought to be good if it is fit for the purpose it was intended for (Ladley 2012). Managing data quality is a critical part of data management, since poor data results into poor information quality, which leads into poor business performance (DAMA 2010). In the US alone, IBM estimated data quality issues to cost businesses $3,1 trillion annually (Redman 2016). Data quality management does not just include correcting data; it is a continuous process throughout a dataset's entire life cycle. This process includes identifying the key metrics for data quality, deploying them, monitoring, and acting to resolve identified issues. Data governance ensures that these steps are put to action. (DAMA 2010; Ladley 2012)

One of the things that data governance councils must approve is data architecture (DAMA 2010). The definition of data architecture is the data models and design approaches aiming to serve strategic business requirements, often at enterprise level (Ladley 2012). It a process of defining and maintaining data requirements, controlling data assets, and aligning business strategies with data investments, which aims to provide standardised business vocabulary, communicate strategic requirements on data, outline data designs to meet them, and align business strategy with business architecture (DAMA 2010). Data architecture incorporate formal names, comprehensive definitions, structures, integrity rules and documentation of and organization's data (DAMA 2010), as well as addresses the aspects of data quality, security, metadata, data processes, and operation concerns such as storage, retrieval, searchability, findability, and accessibility (Martin et al. 2010).

Master data management (later referred to as MDM) aims to provide access to an organisation's most important data, master data, by collecting and matching data into a single source of truth (Rishartati et al. 2019). Done well, MDM ensures that an organisation's data is consistent, up-to-date, connectable between business units and applications, and reduces operating costs of the data model (Hanif et al. 2019). MDM helps to solve data quality-related issues, such as duplications, inaccuracies, and inconsistencies by merging datasets into one dataset, called the single source of truth (Hanif et al. 2019; Vilminko-Heikkinen 2017). These different data sets can be, for instance, sales data, marketing data, and financial data. Master data refers to the entities, attributes and relationships that are critical for an organisations business processes and application systems (Berson & Dubov 2011).

Metadata is used to catalogue an organisation's data and information resources (Haynes 2018). To put it simpler, metadata is data about the data, for instance its name, location, importance, quality etc. (Soares 2014). The different types of metadata are physical metadata, the information regarding the physical data storage; domain-independent metadata, the audit trail of data; domain-specific metadata, the connection between data and business domain; and user metadata, the user specific annotations in data items (Khatri & Brown 2010). The lack of proper metadata means that there is no common data dictionary, which leads to siloes formed by solo applications that handle their own data (Sarsfield 2009).

There are lifecycle-stages that all data moves through and understanding this is central to planning and implementing data governance (Khatri & Brown 2010). A lifecycle of a dataset consists of creation, use, manipulation and eventually disposal (Ladley 2012). When it is properly managed, a datasets lifecycle begins even before creation, as organizations plan for its specifications, capture, delivery, controls, and storage (DAMA 2010).

2.1.2 Business processes

The most important thing when introducing data governance is to understand the organisation's critical business processes and their linkage to data issues (Sarsfield 2009). Introducing data governance often requires changes in business processes (Ladley 2012). An example of this would be when master data management is implemented, and instead of entering data into many different places data is entered and modified from a single place (Ladley 2012).

From an investment point of view, data governance projects must have a business impact by increasing revenue, lowering costs, and reducing compliance risk (Sarsfield 2009). In the end, the purchaser does not care about metadata or duplicates; they just want to make money (Sarsfield 2009). Therefore, it is probable that the return of investment, ROI of data governance projects must be measured (Plotkin 2014).

One of the critical success factors of data governance is to have focused and tangible data strategies to reach short- and long-term goals (Alhassan et al. 2019). Strategy is defined as a set of choices and decisions to draw a high-level roadmap to reach high-level goals (DAMA 2010). The result of a good data strategy should be a high success and low risk rate in system development (Adelman et al. 2005), and competency in data and information management, which makes business strategies more effective (Bhansali 2014). Data strategies may concern areas such as mission and vision for data management, guiding principles, measures for data management success, etc. (DAMA 2010)

To achieve the goals set for data governance, organizations may not only need to consider the data aspect, but also design, deploy, and optimise more efficient business processes and operational workflows (Berson & Dubov 2011). One thing that helps organisations to do this is MDM (Berson & Dubov 2011; Vilminko-Heikkinen 2017). It helps to achieve this is by for instance tackling data quality-related issues by validating data inputs and minimising the shortcoming of the upstream business processes when handling bad data (Berson & Dubov 2011; Vilminko-Heikkinen 2017). MDM also helps by mapping of all entities related to business processes, and creating an understanding of what is important, which helps to rethink business processes (Berson & Dubov 2011).

2.1.3 Compliance and procedure management

In this chapter, the compliance and procedure management side of data governance is discussed. The topics of data policies, standards, security, and regulations are introduced.

Data governance needs cross-functional, enterprise-wide policies to be effective (Bhansali 2014). These data policies are statements and rules considering the creation, acquisition, usage, quality, integrity and security of data and information (DAMA 2010). They are for describing what needs to be done and what not to do with the data (DAMA 2010; Plotkin 2014). It is best practice that there are only a few, briefly and directly stated data policies that are standardised, easily trained, easy-to-follow, and repeatable (Alhassan et al. 2019; DAMA 2010; Ladley 2012). Data management professionals are usually the one to draft data policies, and data stewards and management review and refine them (DAMA 2010). Data policies may concern topics like data modelling, development, architecture, quality, roles, security etc. (DAMA 2010; Plotkin 2014). Data policies must be communicated, enforced, and enforced properly, as well as re-evaluated periodically (DAMA 2010; Bhansali 2014).

Data standards define the organisations data model and its content (Vilminko-Heikkinen & Pekkola 2013). Another definition for data standards is that they are guidelines on how to use, handle, and protect data throughout its entire life cycle (Eryurek et al. 2021). The difference between standards and data policies is that data policies describe what to and what not to do, and standards describe how to do something. Data standards may concern matters like data modelling, metadata, data security, naming conventions, data formats, data quality etc. Just like data policies, data standards also must be communicated, enforced, and enforced properly, as well as re-evaluated periodically. (DAMA 2010; Ladley 2012; Sarsfield 2009)

Data processes dictated by data governance need to support data security and privacy (Ladley 2012). Since data security breaches can be catastrophic for an organisation (e.g., case Vastaamo), most of them have security and privacy policies in place, and data governance just needs to make sure to adopt them (Ladley 2012). Establishing and enforcing information security required participation from various roles, such as business line managers, IT support, data stewards etc. (Sarsfield 2009)

All businesses are impacted by laws and regulations, and data activities are no exception. An important part of data governance is to monitor these laws and regulations, and make sure that they are compiled (DAMA 2010). Access to and use of any classified of restricted data, such as personally identifiable data, needs to be or is regulated (Berson & Dubov 2009). An example of these regulations would be the European General Data Protection Regulation, GDPR, which concerns the previously mentioned personally identifiable data (Wolford 2020).

2.1.4 People management

This chapter introduces the people management -part of data governance. This part consists of roles and responsibilities, change management, communication, ownership, and collaboration.

Data governance includes the formal statement of roles to enforce and outline accountabilities, decision rights, responsibilities, and rules (Ladley 2012). To put it simpler, the designations of roles ensure that data and information is managed properly. Even though an organisation has good data processes, unclear roles and assignments may lead to mistakes in dealing with data, and eventually compromise data governance success

(Alhassan et al. 2019). Typical roles and responsibilities for data governance are executive sponsor (oversight and resourcing), project manager (coordination, communication, change management, risk management), business stakeholders (provide knowledge from business processes and usage of data) and data stewards (support systems and access to data) (Sarsfield 2009). The people who these roles are assigned to are the ones accountable and/or responsible for the data, its proper use and prioritisation. (Plotkin 2014)

Implementing data governance requires the people in the organisation to change their behaviour according to the practices (Ladley 2012). This is not a simple adjustment, since people tend to cling on to the status quo, and thus resist change (Bhansali 2014). Nevertheless, to successfully implement data governance, an organisation must do this. Change management can be done in three steps: 1. planning, 2. doing, and 3. sustaining (Ladley 2012). The resistance can be managed by involving all stakeholders and with clear communication and information sharing (Bhansali 2014). Unfortunately, change management often fails, since it is often overlooked as a “soft” thing, and businesses seldom track the costs of poorly managed change. This is a root cause for failing with data governance and should not be overlooked. (Ladley 2012)

For successful data governance, it is important to have open, honest, frequent enough, communication. This does not only go from top to bottom, but also vice versa: only by understanding how people feel and what they think will management be able to address issues and adjust plans (Ladley 2012). Communication-related problems can be lack of communication, communication channels, awareness, employee competencies and unclear data definitions (Sulanen 2021). Data governance-related communication must reach all its stakeholders, and thus it is important that the message is customised according to the audience, since different people, such as businesspeople and data engineer possess a different level of knowledge about data governance (Sarsfield 2009). A critical success factor for data governance, employee data competencies, is also a part of communication, since training this requires communicating data policies, standards, and procedures to employees (Alhassan et al. 2019). Communication is also too often overlooked as a “soft” thing” or seen as a onetime activity. To be successful, it must be done constantly and systematically. Proper communication leads to a good data culture and with that high-quality data. (Eryurek et al. 2021; Sulanen 2021)

Data ownership means the right to own a dataset, and the responsibility to set business rules, create metadata, and maintain its quality, as well as the designation of accountability and decision-making authority (Plotkin 2014). Unclear data ownership may lead to confusion regarding responsibilities as well as forgotten, lost, and/or mismanaged data

(Eryurek et al. 2021; Sulanen 2021). On the other hand, too strong ownership may form competing data silos owned by business units, which are not willing to share it, even if it would benefit the whole organisation (Sarsfield 2009). Therefore, while data ownership should be established, owners must keep in mind that data should be seen as an organisation-wide asset, not someone's property (Sarsfield 2009). It is also usual for new data elements to pop up from different sources, and when it happens the assignment of responsibility should be done right away, for instance by having frequent assignment meetings. (Plotkin 2014)

Data governance requires business areas, for example, human resources and sales, to collaborate. To properly do this, organisations need to assess and understand their capability for cross functional collaboration. Yet again, collaboration and its assessment are too often seen as a "soft" thing, albeit organizations need the ability to reach over the business units' boundaries to leverage growth. (Ladley 2012)

2.1.5 Data governance metrics

To properly manage data and to prevent a data governance program from fading away, one must measure the impact of data governance programs. Metrics allow organisations to monitor what is happening in their activities, and act quickly if something goes wrong. In the context of data governance, metrics allows organisations to review, monitor, and assess the performance of their assets, which again helps to understand what creates value for the business and what does not. Thus, organisations implementing data governance programs need ways to monitor their effectiveness. (Eryurek et al. 2021; Ladley 2012)

There are four areas in measuring data governance program effectiveness: data quality, data stewardship, business value, and maturity (Eryurek et al. 2012; Ladley 2012; Plotkin 2014). The first, data quality, includes measuring all aspects of data quality: accuracy, completeness, timeliness, and consistency (DAMA 2010; Ladley 2012). The metrics for data quality aims to monitor and maintain data quality standards during its entire life cycle. This consist of enabling data quality monitoring and reporting, root cause analysis, providing action point recommendations, and setting up a baseline for data quality standards. (Eryurek et al. 2012)

The next metric, data stewardship, measures the progress and effectiveness of implemented data stewardship. Progress refers to the number of employees trained for data governance, projects governed, and issues solved or elevated. Effectiveness again refers to the level of data stewardship importance, and the number and impact of data stewardship deliverables. (Ladley 2012; Plotkin 2014)

Business value refers to the financial benefits of data governance programs. The metrics for measuring this are for instance, the return on investment (ROI) of data governance, operation effectiveness, customer satisfaction, revenue growth, number of compliance issues, etc. The exact metrics used to determine the business value of data governance programs depends on the type of business, but examples of these metrics can be the loss of productivity, business cost, and/or compliance costs. Another good way to determine this is to survey data users, for instance, about the level of data quality and understanding. (Eryurek et al. 2012; Ladley 2012; Plotkin 2014; Sarsfield 2009)

The last metric, maturity, refers to the maturity level of information management based on assessment of data governance elements and data management (Ladley 2012). Through maturity assessment, organisations can supervise, evaluate, and upgrade their data governance capabilities. To do this, several data governance domains need to be assessed, the usual ones being data principles, lifecycle, quality, stewardship, master data, and metadata. The maturity of the domains can be assessed with different maturity-level measurement tools, where the lowest level of maturity is that data governance is reactively implemented, and the highest level is that data governance processes are up, running, and continuously improved. (Cheng et al. 2017; Kurniawan et al. 2019; Permana & Suruso 2018)

2.1.6 Data governance framework by Data Governance Institute

The data governance framework by The Data Governance Institute (later referred to as DGI) had the most influence on the data governance for sustainable AI -framework created in this research. The components in their framework are divided into three different categories: Rules & Rules of Engagement, People & Organisational Bodies, and Processes.

The first category, Rules & Rules of Engagement comprises of the first six components of the framework. The first of them, *Mission and Vision*, states a three-part mission for data governance: 1. proactive defining/aligning of rules; 2. the providing of ongoing, boundary-spanning protection and service to data stakeholders; and 3. reacting to and resolving emerging issues. Additionally, this component aims to define a create vision for data governance. The second component, *Goals, Governance Metrics / Success Measures, Funding Strategies*, defines what goal should data governance efforts pursue, how to measure success, and how to fund it. The third component, *Data Rules and Definitions*, covers business rules, compliance requirements, definitions, policies, and standards related to data. The fourth component, *Decision Rights* defines who gets to make decisions and what are the decision-making processes. The fifth, *Accountabilities* answers the question of who should do what and when. The last component, *Controls*, covers the management of data-related risks with preventive and correcting measures.

The second category, People & Organisational Bodies, divides the people relevant to data governance programs into three groups. First of them, *Data Stakeholders*, includes the people who create, use, and set requirements and rules for data. The next group of people is the *Data Governance Office*, who support and facilitate data governance, and collect and report metrics to data stakeholders. The third as last set of people, *Data Stewards*, are a subset of data stakeholders granted the right to make decisions related to data, such as create policies and set standards.

The last category contains only one component, *Proactive, Reactive, and Ongoing Data Governance Processes*. This component describes data governance methods, and comprises of documented, repeatable, and standardised processes.

2.2 Sustainable artificial intelligence

In this chapter, the terms AI, machine learning and algorithm, are defined, and the concepts of sustainable AI and AI governance are discussed. Additionally, the data challenges in AI governance are explained.

2.2.1 Defining artificial intelligence

The definition for AI is a system that behaves intelligently by analysing its environment and acts with some level of autonomy to achieve a specific goal (HLEG 2019). These systems can be based entirely on software or be embedded in hardware devices. When defining AI, it must be considered that “intelligence” is a vague concept both in machines and in humans. In the context of machines, the term “intelligence” is used as notion for its capability to learn, plan, predict and control by processing data and information with some level of autonomy (Zuiderwijk et al. 2021). An example of an AI system is depicted in figure 2. (HLEG 2019)

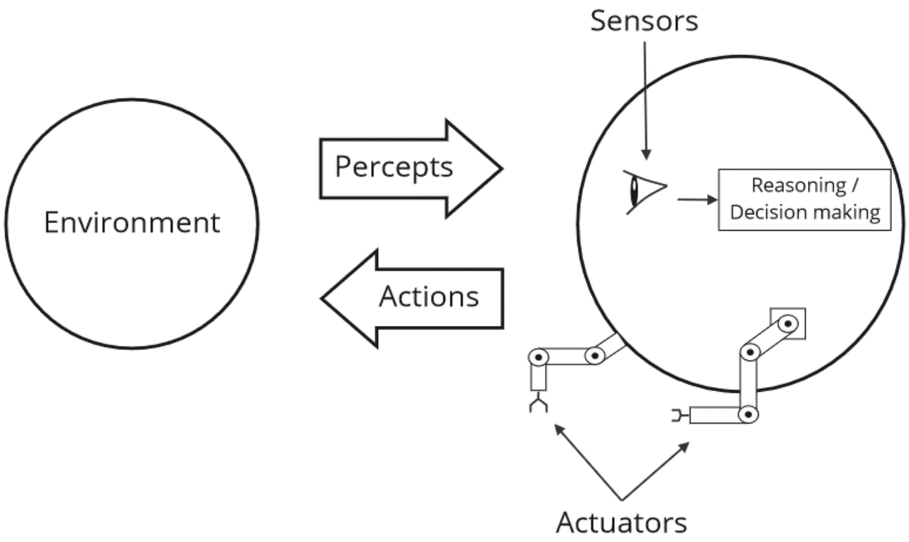

Figure 2. *Illustration of an AI system (based on HLEG 2019)*

As we can see from figure 2 above, AI systems consist roughly of three parts: sensors, reasoning/decision making, and actuators. The first part, sensors, provide the AI system with goal-relevant data present in its environment. These sensors can be for instance cameras, microphones, or sensors of physical variables such as temperature. The second part, reasoning/decision making is the core of the AI system. It takes the data collected by sensors as input and based on it decides which action to take to achieve a given goal. This requires that the input data in processed first into a form, which the reasoning/decision-making module can understand. The third part of the AI system, the actuators, is the part that executes the action determined by the reasoning/decision-making module. (HLEG 2019)

An important concept to understand in the field of AI is machine learning. Machine learning is a learning technique, that allows an AI system to solve imprecisely described problems with unstructured data. These problems can be such as text or language understanding, image processing, or pattern recognition. (HLEG 2019)

Another important concept in the field of AI is algorithms. In all its simplicity, an algorithm is a mathematical procedure, that contains steps to perform calculations (Johnson 2021). In other words, algorithms are sets of instructions on how to perform a task to create a certain output (Doneda & Almeida 2016). Algorithms can be made to perform both simple and highly complex tasks and can be expressed in many kinds of notations, such as natural or computer language. They can be processed by computers as well as humans, but have increasingly become a part of computer programs, especially in the field of AI. (Janssen & Kuk 2016)

2.2.2 Sustainability

The term “sustainable AI” is increasingly referenced in AI governance research (Wilson & Van Der Velden 2022). Sustainable in the context of AI is defined as the development, implementation, and use of AI in a way that minimizes the unwanted ecological, economic, and social impact of its algorithms (Rohde et al. 2021). The goal of sustainability in AI is to foster change towards social justice, economic sustainability, and ecological integrity of an AI application in its entire lifecycle: training, implementation, and use (Kindylidi & Cabral 2021; van Wynsberghe 2021). Sustainable development (of AI) enables to fulfil the needs of today without compromising the ability of future generations to meet theirs (Mensah 2019).

There are three forms of sustainability: social, economic, and ecologic (De Fine Licht & Folland 2019; Wilson & Van Der Velden 2022). The first of them, social sustainability, has no single definition. This is because the main purpose of social sustainability is practical: the desire to produce good and fair processes with good and fair results (Kindylidi & Cabral 2021). A socially sustainable (AI) system distributes opportunities fairly, and provides sufficient support for social services, such as education and political participation (Assefa & Frostell 2007). There are several concerns around social sustainability of AI systems, such as automated decision-making in criminal justice, information biases, the loss of privacy, the loss of humanity in social relationships etc. (Gill 2020), thus making the social aspect of AI sustainability important.

The second form of sustainability, economic sustainability, refers to the ability to satisfy the consumption needs of the present without jeopardising the consumption needs of the future (Mensah 2019). An economically sustainable (AI) system enables the production of goods and services on a continual basis and does not cause sectoral imbalances that harm agricultural or industrial production. (Assefa & Frostell 2007)

The third form of sustainability, ecologic sustainability includes maintaining a stable resource base, avoiding over-exploitation and depletion of resources, and organising business activities in a way which maintains biodiversity and atmospheric stability (Assefa & Frostell 2007). For instance, an ecological issue caused by AI could be the energy usage and the carbon footprint produced by AI hardware (van Wynsberghe 2021). On the other hand, AI may help us to tackle modern environmental issues, such as climate change (Coeckelbergh 2020), for instance by enabling the tracking of real-time carbon footprints and emissions (van Wynsberghe 2021).

To be sustainable, an AI system must foster diversity, be transparent and explainable, subjected to democratic principles, trusted, and the values embedded into it must be aligned with the values held by society (Wilson & Van Der Velden 2022). Additionally, its innovation, training and usage must be guided towards sustainable development goals. Furthermore, sustainable AI must not just be examined from the human rights and ethical point of view, but all of sustainability's perspectives: ecological, economic, and social. (van Wynsberghe 2021)

2.2.3 Artificial intelligence governance

AI governance is the managerial and administrative decisions on how to balance the potential benefits and harms of AI (Wilson & Van Der Velden 2022). Just like data governance, AI governance consists of organising and implementing procedures, policies, structures, roles, and responsibilities to enforce and outline rules, decision rights, and responsibilities for operationalising sustainable AI. The requisites for sustainable AI are appropriate design, implementation, and use of the systems, and AI governance is there to make sure it happens. The pressure to organise and implement AI governance comes from the society and government institutions on organisation and development teams, and not as much from business needs as data governance. (Minkkinen et al. 2022)

AI governance has three constituent elements:

1. *Corporate governance.* AI governance is a subset of and organisation's overall corporate governance. Corporate governance is there to make sure that organisations distribute responsibility and accountability to all stakeholders and implement all their business activities in a socially responsible way. It places organisational AI governance into practices and activities. (Solomon 2020; Mäntymäki et al. 2022a)

2. *IT governance* specifies the decision rights and accountabilities, which ensures that IT related activities are aligned with organisational strategy and objectives, i.e., encourages desired behaviour using IT (Gregory et al. 2018; Weill & Ross 2004). Since AI systems technically are a specific type of IT system, IT governance offers valuable guidelines on how to operationalise AI systems. (Mäntymäki et al. 2022a).

3. *Data governance.* AI systems use data to learn and operate, making data governance critical for AI governance, an utmost important aspect of governing algorithmic AI systems (Doneda & Almeida 2016; Mäntymäki et al. 2022a).

The relationship between corporate, AI and data governance is depicted in figure 3.

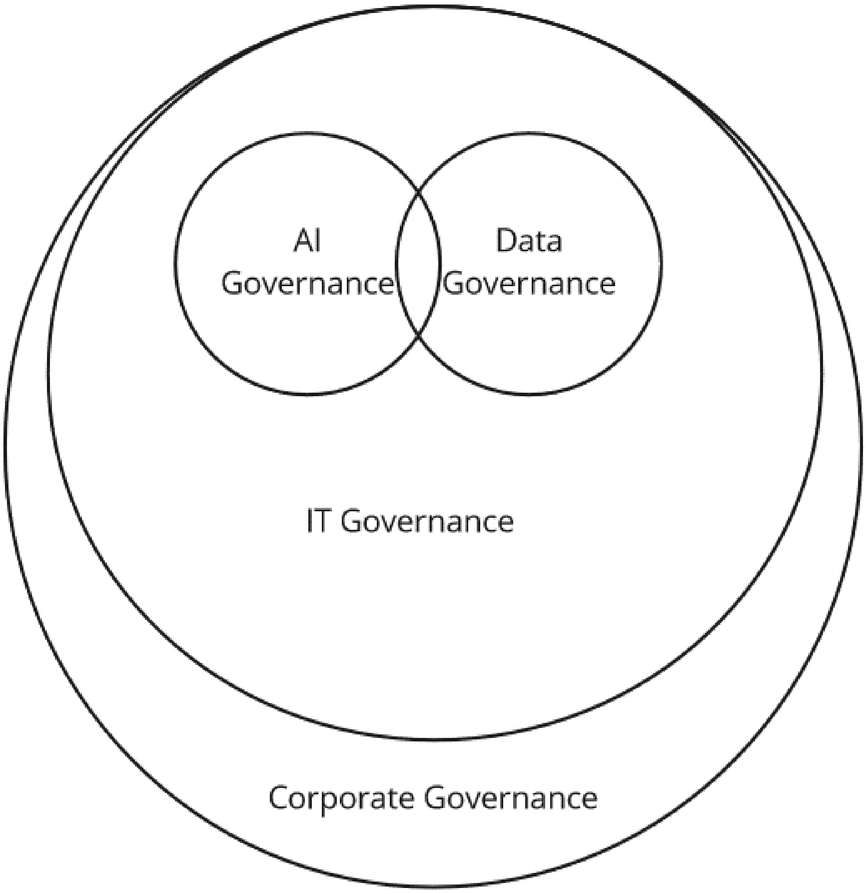

Figure 3. *AI governance as a part of corporate governance (based on Mäntymäki et al. 2022a)*

When examining AI literature, one can conclude that there are roughly six elements; social, ethical, economic, informational, technical, and legal & regulatory layer (Gasser & Almeida 2017, Wirtz et al. 2022; Zuiderwijk et al. 2021). The first element, social, governs the socio-technical infrastructure of AI systems and strives to ensure social equality and compliance, as well as to ensure human well-being (Wirtz et al. 2022). It also tackles the social and societal AI governance challenges of workforce displacement and replacement, acceptance and trust in AI, and the transformation of interaction between human and machine (Wirtz et al. 2019; Zuiderwijk et al. 2019). For this layer to be successful, the sociotechnical system of AI must be trusted and trustworthy (Minkkinen et al 2022).

The next element, ethical, governs the ethical concerns and principles concerning AI systems (Gasser & Almeida 2017). Ethics in AI aims to guarantee the respect of fundamental values and rights by establishing a human-centric, trustworthy AI system (Minkkinen et al. 2022). The ethical challenges this layer aims to solve are such as moral dilemmas, lack of fairness and privacy, threats to human autonomy, unethical use of data, etc. (Wirtz et al. 2019; Zuiderwijk et al. 2019). The guidelines for forming the ethical layer include formulating an ethical code for AI use and development to reflect human culture, norms, and values, and translating it into a programming language for the system. Also, human monitoring, supervision, and quality management must be in place to ensure ethical decision making by the AI system. Additionally, an AI system must not compromise the freedom and autonomy of a human being. (Wirtz et al. 2022)

The economic element addresses the risks of disrupting economic systems, such as unemployment caused by AI systems. Thus, AI governance must ensure that AI is a support tool for human work by training AI competencies, and that way reintegrate individuals unemployed due to AI back into the workforce. Additionally, transparency of AI within business processes must be ensured, as well as fairness in global market competition. (Wirtz et al. 2022; Zuiderwijk et al. 2019)

The informational element of AI governance is heavily linked to social and ethical element. It tackles the AI risks of information manipulation, disinformation, propaganda, censorship, and freedom of speech (Wirtz et al. 2022). The use of algorithms, automation, and AI have given disinformation operations the tools to pose risks to democratic political processes by exploiting the existing tension within societies, manipulating public opinion, and undermining the trust towards institutions and political leadership (Kertysova 2018). Additionally, the mass availability of the internet, smart phones, and digital platforms have become an unprecedented influence machinery by creating the means for malicious actors to broadcast their messages to billions with almost zero cost (Nitzberg & Zysman 2022). This is especially dangerous on a state level, since controlling information flows is an efficient way to weaken democratic institutions and establish an authoritarian regime (Koskelo et al. 2022). When it comes to information, the threats of disruptive technologies, such as AI, are user profiling and segmentation, deep fakes, personalized targeting, and the replacement of human oversight (Kertysova 2018). To prevent these from happening, AI governance must ensure uninfluenced information provision, develop ways to battle disinformation and propaganda, foster freedom of speech, and endure data protection (Wirtz et al. 2022).

The next element, legal and regulatory, addresses the general principles of AI regulation, such as creation of institutions and government authorities, as well as allocation of accountabilities, responsibilities, and supervisory authority for AI regulation (Gasser & Almeida 2017; Wirtz et al. 2022). In addition, it is important to ensure accountability in case of failure related to AI auditing, reporting etc. and evaluate potential impact of AI by developing possible scenarios for future (Wirtz et al. 2022). The challenges regarding this layer can be the lack of accountability, regulation, obedience, and governance (Sun & Medaglia 2019; Wirtz et al. 2022; Zuiderwijk et al. 2019).

The last element, technical, concerns the algorithms and data of an AI system. For socially impactful algorithms, it is necessary to have responsibility, accuracy, auditability, explainability, and fairness principles installed in its decision-making processes (Gasser & Almeida 2017). Additionally, it is important to establish audit and documentation mechanisms for algorithmic decision-making processes, set technological standards for AI, and set safety mechanisms to prevent misuse of AI systems (Wirtz et al. 2022). The prevention of algorithmic unfairness and misuse can be achieved by involving human control, accountability, and values in the decision-making process, and designing the systems in a way that its underlying processes are understandable (Wirtz et al. 2022). The technical element of AI governance is heavily linked to the data governance part of AI governance, since it addresses numerous data challenges, such as data quality, quantity, integration, and so forth (Zuiderwijk et al. 2019). AI governance -related data challenges are discussed in the next chapter.

2.2.4 Data challenges in artificial intelligence

The algorithms in AI systems depend heavily on large quantities of dynamic real-time data to work, and managing such data is challenging (Janssen et al. 2020). Therefore, data governance aspects are relevant for AI governance as well, especially the ones concerning these algorithmic systems (Mäntymäki et al. 2022a).

Although AI algorithms and its data use can be inspected, only a few people can really understand them (Janssen et al. 2020). These hard-to-understand activities are often referred as “black box” activities (Wirtz et al. 2022), and they result to the lack of opportunities to publicly scrutinize, assess risk, audit, sample, validate, control quality, and to implement other inspection mechanisms (Janssen et al. 2020). There are two dimensions in AI complexity: technical and administrative. The technical side covers the technical complexity of an AI system, which is beyond the comprehension of ordinary public members. The other side, administrative, refers to standards, rules, and institutions governing AI. To make these activities more explainable, AI governance must make them not only technically more explainable, but to ensure the opacity of standards and rules, and that the processes and institutions can be influenced by democratic means. Making AI more explainable is a solution to its confidentiality, complexity, and opacity issues (Keller & Drake 2021).

Data quality has been identified as an AI governance challenge by many researchers (Wirtz et al. 2022; Zuiderwijk et al. 2021). Data is the fundamental driver of an AI system as it uses it for learning, and thus keeping data quality high is crucial (Wirtz et al. 2019). The challenges in data quality emerge from lack of standards when collecting, formatting, and storing data (Sun & Medaglia 2019). Since AI systems use large volumes of data from numerous sources, the importance of data quality must be understood in organisations to ensure sustainability (Janssen et al. 2020). Poor data quality leads to inaccurate data, which again may lead to biased or skewed algorithm outcomes and failures (Janssen et al. 2020; Wirtz et al. 2019). Additionally, data acquisition processes must be transparent enough to be viewed and checked for correctness to keep algorithms and AI systems from becoming black box systems (Wirtz & Müller 2019).

An AI system must possess the capability to integrate and manage the independencies between data, processes, and technologies (Wirtz et al. 2019). Therefore, the degree of data integration must be placed. Data integration means the activity of gathering data from multiple sources and combining it into one to make it usable (Gupta 2019). In the context of AI, this could mean the connection between demographic data, for instance, age, education, gender, and longitudinal data, for instance, income development, health, weight loss progress, which an AI system uses to draw conclusions. In addition, an AI system needs the data pool that it uses for learning to be large enough (Sun & Medaglia 2019), for instance to avoid unintended training bias against underrepresented groups (Nitzberg & Zysman 2022).

A perquisite for implementing an AI system is the gathering of data from people and organisations, creating a chance for unethical use of the data, for instance, for commercial purposes (Gupta 2019). The most obvious example of unethical use of data is the violation of privacy (Doneda & Almeida 2016). Securing and guarding data from misuse is especially important with sensitive data, such as address, health status or political preference and the collection of this kind of data should be minimised. Another point of view for unethical use of data is the discrimination caused by AI algorithms. Since datasets are the central part of this, the verification of correctness, legitimacy, and absence of bias must be emphasized. (Doneda & Almeida 2016; Janssen et al. 2020)

Another data governance-related challenge in AI is to decide the level of control over algorithmic data. This depends on the lifecycle in which the data is, and the level of sensitiveness, publicity, and personal identifiability of the data. For instance, if the data is not personally identifiable, open, and not sensitive, then the level of control does not need to be that high. On the contrary, if the data is personally identifiable, sensitive, private, or confidential, the level of control should be high. (NSW Government 2021)

While there are big challenges regarding data and AI, one must not forget that there are benefits in AI data and information processing. A few examples of these would be the improvements in information and big data processing, the capability of machine learning systems to continuously self-improve, the possibility to handle multi-dimensional and -variety data, etc. (Zuiderwijk et al. 2021).

2.2.5 AI governance frameworks

In this chapter, four AI governance frameworks are introduced. The first and second frameworks are layered AI governance models by Gasser & Almeida (2017) and Wirtz et al. (2022), the third is a data governance model for trustworthy AI by Janssen et al. (2020), and the fourth is the AIGA Hourglass Model of AI Governance by Mäntymäki et al. (2022b).

Gasser & Almeida 2017

The model by Gasser & Almeida (2017) aims to capture the complexity of AI governance by layering it. These interacting layers, presented in figure 4 below are social and legal, ethical, and technical layers, which contain most of the AI governance elements discussed in chapter 2.2.3. These can be developed at different timings; for instance, standards and principles concerning AI algorithms can be developed quite quickly, but changing state-level legislation can take some time. Although different layers have different timings, implementing structures to govern AI can happen in multiple layers at a time, and can involve mixed approaches.

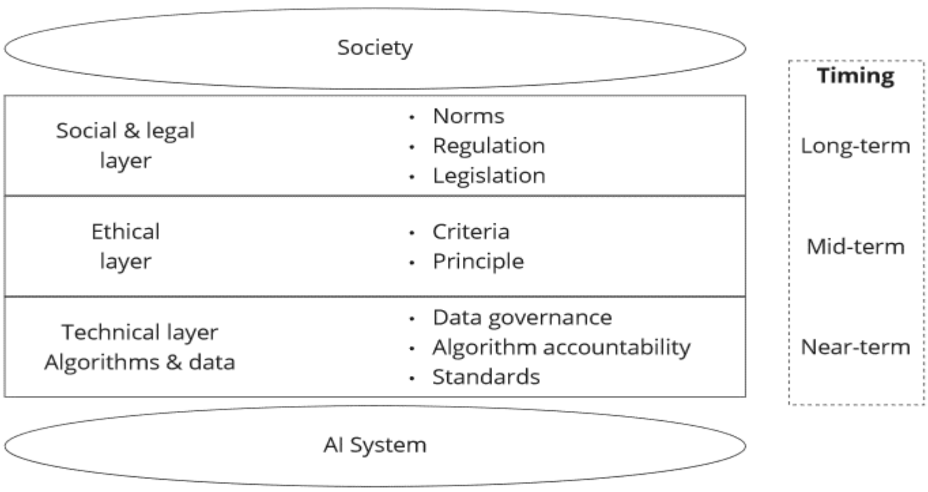

***Figure 4**. AI governance model by Gasser & Almeida (2017)*

Wirtz et al. 2022

The model by Wirtz et al. (2022) is also a layered one, but differently layered and a bit more comprehensive than the one by Gasser & Almeida (2017). The layers in their framework are AI risk layer, AI risk management and guidance process layer, AI guidelines layer, and AI governance layer. With this kind of layered model, Wirtz et al. (2022) aim to connect the risks and corresponding guidelines of AI systems, arguing that the complexity and development pace of AI makes it necessary to create a risk-based, adaptive, flexible, and interactive AI governance model. In addition, AI governance approach should involve all relevant stakeholders, such as the government, public sector, and industry.

The authors divided AI associated risks and guidelines into six categories: 1. technological, data, and analytical, 2. informational and communicational, 3. economic, 4. social, 5. ethical, and 6. legal and regulatory. Wirtz et al. (2022) highlight, that the guidelines to respond to risk concerning analytics, data, and technology are the most important due to their contribution to its implantation and operations level control over AI systems decision making process. Additionally, these guidelines involve data security procedures, and the prevention of unwanted AI behaviour based on poor and inadequate amounts of data.

The risk and guidance layer is linked to the risk management and guidance process layer. The process of linking risks to guidance consist of four stages. Risk and guidance framing is the first stage of the process, defining the regulatory scope and objectives, and further procedures. The next stage, risk and guidance assessment identifies the impact and extent of AI systems risks. The third stage, risk and guidance evaluation aims to evaluate the individual areas affected by the risks, as well as the extend and necessity of countermeasures. The fourth and last stage of the process guidance implications includes the indication supervising and monitoring AI, and legislation aspects. This four-stage-process described above aims to form a basis for formulating the AI guidelines layer.

The last layer, governance layer, implements the AI guidelines by policy-making processes and enabling a continuation cycle for AI governance. There are seven stages in this layer, which aim to turn AI guidelines into regulations and practical governance measures. The staged processes of the AI governance layer proceed as follows. First, AI risk and guideline needs are characterized. Second, AI guidelines are publicly discussed by all stakeholders, while taking public interest and social norms into account. The discussions aim to formulate AI guidelines backed by public acceptance. After this, the means for implementing the regulatory infrastructure and guideline monitoring must be provided. AI governance is not a one-time activity, but rather an on-going process, and thus it needs to be constantly evaluated against environmental changes and adjusted accordingly. The layers of the framework are summarized below in figure 5.

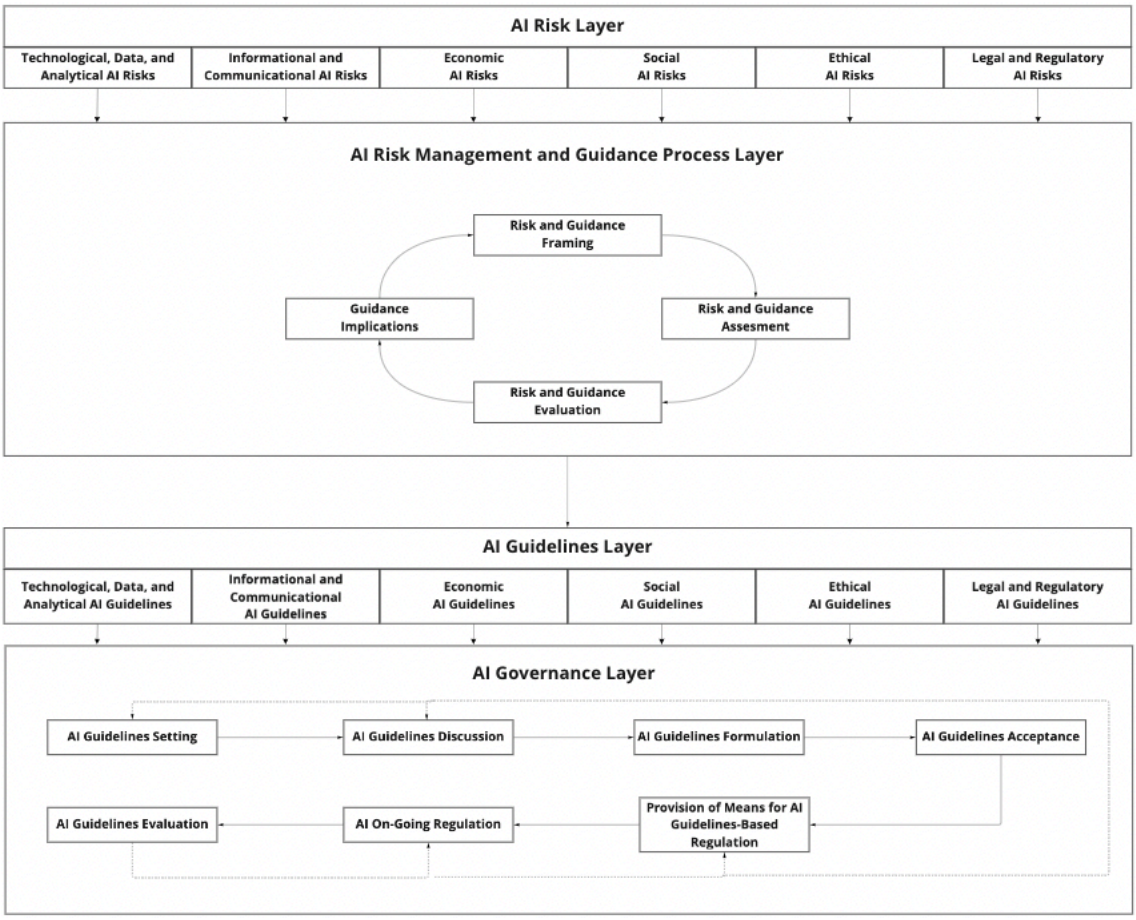

Figure 5. *The layers of AI governance (based on Wirtz et al. 2022)*

Janssen et al. 2020

Unlike the two AI governance frameworks earlier, the framework by Janssen et al. (2020) is technically not an AI governance model, but rather a data governance model for big data algorithmic systems (BDAS). Their framework consists of three elements, 1. system-level governance model for BDAS, 2. data stewardship and based registries, and 3. trusted data-sharing framework based on self-sovereign identities and data-sharing agreements. The first element, system-level governance model for BDAS, aims to design a system-level accountability network to govern the working, usage, data, outputs, and auditing of algorithms. This element consists of many parts; the most important being that BDAS's must be constructed according to laws and regulations. In addition, BDAS-related policies, principles, procedures, and data culture must be designed and implemented in a way that reflects societal values, norms, and expectations. The element also requires the assessment of training and operation data quality, the checking of data changes and learning outcomes, as well as communication of the decisions regarding BDAS. Data governance plays a big part in implementing this element, giving the necessary mechanisms to incentivize wanted behaviours and sanction unwanted.

The second element of this framework, data stewardship and based registries, aims to solve the challenge of establishing data ownership. It provides a foundation for data sharing by assigning data stewardship to formalise data management accountabilities, and creating base registries to create a trusted, re-usable source of information. Data stewards are the ones responsible for data quality, security, and validity, as well as data and risk management.

The last element, trusted data-sharing framework based on self-sovereign identities and data-sharing agreements, tackles the issues of manipulation and misuse of an organisations external data sources. The goal of these trusted data-sharing networks is to ensure reliable and secure sharing of good quality, interoperable, compliance following, and ethically gathered data-assets. To be successful in this, these networks need to provide authentication, authorisation, and identification services. Additionally, the components of a trusted data-sharing network may include services like requirement lists, standards, audit mechanisms, and compliance enforcement tools for data sharing. Lastly, only the minimum amount of data that is necessary should be shared. Janssen et al. (2020) refer to this as the “need to know” principle. Lastly, the suggest applying the following data governance principles to BDAS governance:

1. Data quality and data bias evaluation
2. Detecting and investigations of algorithmic pattern changes
3. “Need to know” principle
4. Encourage error spotting with a bug bounty
5. Inform the people and organizations when sharing data about them
6. Personal and non-personal data should always be separated, as well as sensitive and non-sensitive
7. Give people and organizations the right to check and correct their data
8. Always collect data from its origin to confirm the ethicality of its collection
9. Do not grant data access to parties who do not need it

10. Store data in a distributed way

11. Assign data stewards

12. Divide data responsibilities in a way that no single person can abuse or misuse the data

13. Foster a mindset that recognises data as an asset

Although Janssen et al. (2020) list these principles, they also remind that there are different approaches to data governance.

Mäntymäki et al. 2022b

The last model, the AIGA Hourglass Model of AI Governance had the most influence on the final version of the framework formed in this research. This AI governance model is also a layered one, and it has the *Environmental* layer in the top, the *Organizational* layer in the middle, and the *AI System* layer in the bottom. The proposed hourglass model of organisational AI governance aims to illustrate the flow of governance requirements from top to bottom, and to highlight how the dynamicity of the AI's environment inputs requirements into the organisation and the AI system itself. The model is illustrated below in figure 6.

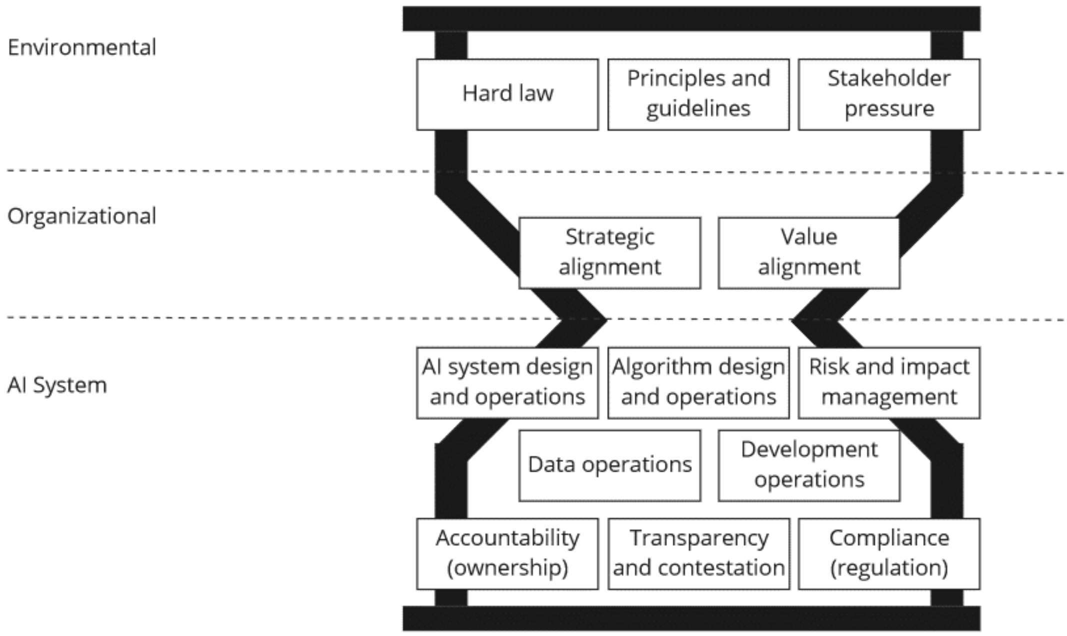

Figure 6. *AIGA Hourglass Model of AI Governance (based on Mäntymäki et al. 2022b)*

The first layer, *Environmental*, covers AI governance requirements coming from outside the organisation, such as law, principles, guidelines, and stakeholder pressure. These requirements are aspects that the organisation cannot influence, at least directly. Examples of these could be GDPR, AI ethics principles, and citizen awareness of privacy issues.

The second AI governance layer, *Organizational*, comprises of two parts, strategic and value alignment. The first part, strategic, requires organisations to align their strategy with AI, by for instance, defining what AI systems are intended to be used, and what business goals are meant to be achieved with this. The second part, value alignment requires alignment between organisational values AI ethics, and ensure this in all its AI systems. In addition, this layer covers the organisation's position on risks, such as regulatory or reputational.

The third layer, *AI System*, refers to the operation of the technical side of AI, where the requirements from the layers on top are implemented. This layer covers the design, development, and operation of AI systems and algorithms, as well as data operations, risk and impact management, designation of accountability, insurance of transparency and contestation, and the embedding of regulatory compliance and monitoring into the AI system.

2.3 Summary

This chapter summarises the connection between data and AI governance. Additionally, a theory-based data governance for sustainable AI framework is constructed based on data and AI governance literature, as well as the existing data and AI governance frameworks.

Already from figure 3 we can see that AI and data governance overlap in some parts, meaning that data governance is a part of AI governance and thus sustainable AI. Data plays a big part of the technical element of AI governance discussed in chapter 2.2.3, which makes yet again data governance an important aspect. The guidelines concerning the technical and data aspect of AI governance are the most important since it has the most influence over AI algorithms (Wirtz et al. 2022).

To ensure the delivery of trustworthy decisions, organisations have turned to data governance to guarantee both legal and ethical compliance and quality of their data. In the context of AI, data governance must take both data and data processing into account in their entire lifecycle. Governing the data of AI systems is not something "soft", since data governance mistakes affecting these systems may cause profound social, legal, and financial issues for businesses, citizens, and overall society at large. These mistakes may cause unlawful decisions, large scale financial losses, or even political crises and loss of lives. Most of data governance principles and guidelines concern the technical element of AI governance. On the other hand, addressing and governing the other elements ultimately comes back to governing the technological toolbox (Nitzberg & Zysman 2022). Since governing the non-technical elements of AI ultimately requires the governance of the technical element, and data governance contains mechanisms to do this, one might think that these data governance mechanism may help to support the governance of the non-technical element via technical solutions. (Janssen et al. 2020)

To explore these possibilities, a theoretical data governance framework for AI governance was formed. This framework consists of three parts: AI governance elements, data governance principles, and supporting activities. Table 2 depicts which AI governance elements were selected for the framework and from which literature they were selected.

Table 2. *AI governance elements chosen for the theory-based framework*

AI Governance Element	Literature
Technical	Gasser & Almeida 2017, Janssen et al. 2020, Mäntymäki et al. 2022b, Wirtz et al. 2022, Zuiderwijk et al. 2019
Environmental	Mäntymäki et al. 2022b
Legal and Regulatory	Gasser & Almeida 2017, Sun & Medaglia 2019, Wirtz et al. 2022, Zuiderwijk et al. 2019
Economic	Wirtz et al. 2022, Zuiderwijk et al. 2019
Informational	Nitzberg & Zysman 2022, Wirtz et al. 2022
Ethical	Gasser & Almeida 2017, Minkkinen et al. 2022, Wirtz et al. 2022, Zuiderwijk et al. 2019
Social	Gasser & Almeida 2017, Minkkinen et al 2022, Wirtz et al. 2022, Zuiderwijk et al. 2019

Whereas relevant AI governance elements could be picked up easily from literature, selecting data governance principles was a bit more difficult task, since there is hardly any research on what aspects of data governance are important from a sustainable AI point of view. On the other hand, the relevant data governance aspects could be identified from AI governance literature, although they were not explicitly listed. Thus, the researcher needed to use his own judgement when choosing them for the theory-based framework. Table 3 explains which data governance principles were selected for the theory-based framework, and from which research they were identified.

Table 3. *Data governance principles chosen for the theory-based framework*

Data Governance Principle	Identified from	Explanation
Data policies, processes, and standards	Janssen et al. 2020, Sun & Medaglia 2019, Wirtz & Müller 2019	Organise day-to-day data activities in a way that supports the elements of AI governance and sustainability
Data strategy	Mäntymäki et al. 2022b	Acknowledge the strategic importance of AI system data, and implement strategies for data and AI systems
Data quality support	Janssen et al. 2020, Wirtz et al. 2022, Zuiderwijk et al. 2021	Create adequate data policies, processes, and standards to ensure good data quality, and assign roles and responsibilities to oversee this
MDM	Janssen et al. 2020	Establish base registries of "single sources of truth"
Metadata	Janssen et al. 2020	Create “data about data”
Roles and responsibilities	Janssen et al. 2020, Mäntymäki et al. 2022b	Assign data stewards, and designate accountabilities, ownerships, roles, and responsibil-ities for data and algorithmic components

Change management, collaboration, and communication were selected as supporting activities for the framework. These activities are identified as success factors for implementing data governance, and thus they are added to this framework as well. The supporting activities selected for this framework are listed in table 4.

Table 4. *Supporting activities for the theory-based framework*

Supporting Activity	Literature
Change management	Bhansali 2014, Ladley 2012
Collaboration	Ladley 2012
Communication	Alhassan et al. 2019, Eryurek et al. 2021, Ladley 2012, Sulanen 2021

The AI governance elements, data governance principles, and supporting activities were then synthesised into a theory-based data governance for sustainable AI framework. The framework is illustrated in table 5 on the next page. The middle section of the framework explains why each of the data governance activities are relevant and how they interact with and affect different AI governance elements. Next, this theory-based data governance framework was discussed and further developed with data and AI governance experts.

Table 5. *Theory-based data governance for sustainable AI -framework*

Supporting Activities		
Change management	**Collaboration**	**Communication**
To help the behavioural changes required to implement data and AI governance to stick	To make data and AI governance activities cross-functional, and to build trusted data-sharing networks	To ensure everyone know why data and AI governance exists, how they are implemented and carried out, and what is their role in it

Data Governance Principles	AI Goverance Elements						
	Technical	**Environmental**	**Legal and regulatory**	**Economic**	**Informational**	**Ethical**	**Social**
Data policies, processes, and standards	To organize data activities in a way that ensures human control over AI system data, prevents technical vulnerabilities, and fosters expertise	To prevent the negative environmental impact via data-driven activities	To ensure legal and regulatory compliance in data activities	To ensure economic sustainability in data activities	To ensure the prevention of informational misuse in data activities	To align data activities with ethical norms and values	To ensure privacy, safety and other dimensions of social sustainability in data activities
Data strategy	To include the prevention of technical risk and vulnerabilities of AI system data in strategies	To include data-driven prevention of negative environmental impact	To include legal and regulatory compliance of AI system data in strategies	To put economic sustainability of AI system data to organization's strategy, and treat data as an asset	To include the prevention of informational misuse of AI system data in strategies	To include ethical collection and use of AI system data to strategies	To include social sustainability of AI system data in strategies
Data quality support	To prevent algorithmic bias due to poor data quality	To keep the quality of the data used for data-driven environmental impact prevention intact	To prevent unintentional legal and regulatory defiance due to poor data quality	To prevent the costs of poor algorithmic decisions due to poor data quality	To ensure informational correctness	To prevent unfair AI decisions, discrimination, and misinterpretation of human values due to poor data quality	To prevent social inequalities caused by AI systems due to poor data quality
MDM	To ensure adequate data quality and quantity processed by algorithms by centralizing data management	To manage the data used for data-driven environmental impact prevention	To make ensuring compliance easier by centralizing data management	To help business process supervision via data process supervision capabilities brought by MDM	To make data and information manipulation harder by centralizing data management	To detect unfair statistical decision and discrimination by centralizing data management	To help ensure privacy, safety and other dimensions of social sustainability by centralizing data management
Metadata	To better understand the technical dimension of AI system data	To better understand the data used for data-driven environmental impact prevention	To help ensure legal and regulatory compliance by better understanding AI system data	To help ensure economic sustainability by better understanding AI systems data	To help prevent informational misuse of AI system data by better understanding it	To help ensure alignment with ethical norms and values by better understanding AI system data	To help ensure social sustainability by better understanding AI system data
Roles and responsibilities	To ensure adequate decision rights and responsibilities to govern data and AI systems	To assign the roles and responsibilities necessary for data-driven negative environmental impact prevention	To ensure decision rights and responsibilities to ensure legal and regulatory compliance	To assign the roles and responsibilities necessary to support economic sustainability	To ensure decision rights and responsibilities to prevent information manipulation and control, as well as spreading of disinformation and propaganda	To ensure adequate decision rights and responsibilities in data and AI governance to align the with ethical norms and values	To assign the roles and responsibilities necessary to support social sustainability

3. RESEARCH METHODS AND SETTINGS

This chapter describes the methods and settings used in this research. This includes describing the research approach, and data collection and analysis. Also, the reliability and validity criteria for this research are explained.

3.1 Research approach

This study aims to answer the research questions with a qualitative case study. The case of this research is to try and build a generalised data governance framework for AI governance and sustainable AI customer cases. The goal is not to create a universal framework for every use case, but rather a starting place for implementing data governance for sustainable AI, which can be later modified according to the specific needs of an organisation. The findings of the research were used to help the case company explore new technological and governance spearheads in the fields of AI and data, and gain knowledge on these topics. Furthermore, this knowledge helps the case company, Solita, find potential business opportunities and niches.

The research began with conducting a literature review on data governance, AI governance, and sustainability. After this, the theories were synthesised into a theory-based data governance for sustainable AI framework. After acquiring adequate knowledge on the research topics and synthesising it, the theory-based framework needed to be refined into a more practical form, to better transfer and translate it into practical solutions. This was done by organising workshops for experts from the area of data governance and AI governance, where more information could be gathered in the form of an unstructured interview. A total of two were organised, where in addition to the researcher, two data governance experts and two AI governance experts were present. Table 6 below describes the participants in more detail.

Table 6. *Workshop participant details*

Participant number	Workshop	Position	Field of expertise in this research	Years of experience from field
1	1	Head of Sustainable AI	Algorithmic Systems, AI & Data Governance, Sustainable AI	10
2	1	Head of Research	AI Governance	> 5
3	1	Head of Data Governance	Data Governance	10
4	1	Data Consultant	Data Governance	3
5	2	Data Strategy Lead	AI & Data Governance, Data Strategy	AI Governance: 4, Data Governance: >10, Data strategy: 7
6	2	Data Engineer	AI Governance	< 1, thesis worker
7	2	Data Advisor	Data Governance	6
8	2	Data Engineer / Data Governance Consultant	Data Governance	5

3.2 Data collection and analysis

The empirical data was collected by conducting group interviews, more specifically workshops. The method of this study is qualitative, and thus interviews were found to be the most fitting data collection methods since they enable the gathering of detailed information and by that gaining of in-depth knowledge about the topic. Interviews are a good way to collect data and gain an understanding of the world view of others but conducting them in not a trivial task. It requires a various set of skills, such as active listing, as well as careful planning and correct preparation. In addition, it is necessary for the interviewer to gain expertise in the relevant topic as possible, to collect interview data that is useful for the research purpose. In this chapter, we elaborate on what kind of interviews were conducted, who and how many did we interview, and how the interview data was analysed. (Qu & Dumay 2011)

The interviews took place in a focus group setting, which aims to utilise flexible and exploratory discussion and emphasise the interaction between participants, leading to diverse conversation and challenging of each other’s views. And since the AI and data governance experts at the case company are busy, the advantage of focus groups being convenient, and timesaving came in handy. Interviewing experts generally require tailored questions and editing the interview frame as the interview progresses (Alastalo et al. 2017). Good expertise is considered a condition for the success of expert interviews, and thus it is especially important for the interviewer to gain expertise on the topics to be discussed. (Alastalo et al. 2017; Sauders et al. 2009; Qu & Dumay 2011)

A qualitative study focuses on a small sample size and aims to analyse them as thoroughly as possible. In qualitative studies, it is not the quantity that makes it scientific, but rather the quality of the participants (Eskola & Suoranta 1998). Therefore, the sample size in this study is rather small: in total, four data governance and four AI governance experts from Solita were interviewed. The participants were divided into focus groups of two AI governance and two data governance experts. An often-raised issue regarding small sample sizes of qualitative studies is the generalisability and transferability of the findings, versus the statistical generalisability of quantitative studies with large sample sizes. However, this does not mean that qualitative studies are less valuable. Rather than providing statistically valid generalisations, qualitative studies are often used to develop theories, which is exactly what this study aims to do. (Saunders et al. 2019)

What comes to the interview type, there are three options: structured, semi-structured, and unstructured interview. Since expert interviews require tailored questions in a flexible frame, the interviews were decided to be conducted as semi-structured theme interviews. In these kinds of interviews, the questions are predetermined, but their order and wording may change depending on the situation (Hirsjärvi & Hurme 2011). The aim was to keep the focus of the discussion on themes of AI and data governance with relatively loose guidance since the intention was not to limit the conversations too much. Most importantly, semi-structured interviews give the participant the freedom to answer in the way they think, using their own language and terms. Additionally, semi-structured interviews enable guiding the conversation to the correct topic and issues if the conversation goes off-track or degenerates into a “chat”. (Qu & Dumay 2011)

Semi-structured interviews require a comprehensive set of questions to keep the participants engaged, and an interviewer that can respond sensitively. There is no single right way of wording questions and interviewing, since the perspectives of the participants and the interviewer create a unique environment every time. Therefore, to get the best possible responses, the interviewer must be responsive and sensitive, as well as possess good interpersonal skills. Additionally, the quality of the interview can be supported by keeping the flow of the interview continuous, maintaining a good relationship with the participants, and avoiding interviewer bias. To support this, a set of interview questions described in appendixes A and B were drafted. (Qu & Dumay 2011)

As mentioned, the interviews resembled more workshops than traditional interviews. To prevent it from turning into a “chat”, the research questions were used to guide the conversation in the right direction when necessary. Both workshops were 1,5 hours long and were organised in a hybrid setting, and in both workshops one participant was online, and the rest were in a meeting room with the interviewer. A Miro-board prepared for the workshops by the researcher, where the theory-based framework depicted in table 2 was presented by the interviewer and commented by the participants. Some of these comments were added as virtual post-it notes on the board. The first ten minutes of the interview were used to establish introductory small talk, explain the purpose of the interview, and provide context knowledge on the topics. Debriefing was done in the last five minutes. The plan was to use a semi-structured approach with the interview questions, and thus the workshops did not entirely follow the planned structure dictated by interview questions. Due to this semi-structured nature, many of the questions asked from the participants were not listed in the question structure.

To create new knowledge from interview data, it must be analysed carefully. The purpose of qualitative data analysis is to summarise and clarify interview data without losing any of it. This is the most problematic part of a qualitative study due to the lack of clear working techniques and the lack of teaching them. In this research, the analytical focus was on the themes and topics of the interview data, and thus thematic analysis was chosen to be the data analysis method. Thematic analysis is recommended for solving pragmatic problems, which sits well with the philosophy of this research. The procedure of thematic analysis consists of four stages: 1. becoming familiar with the data; 2. coding, where data units are categorised and labelled to symbolise their meaning; 3. analysis, where themes and relationships in the data are recognised; and 4. theme refinement and proposition testing, where conclusions and explanatory theories are made. (Eskola & Suoranta 1998; Saunders et al. 2019)

This research utilised Microsoft Teams as a recording tool to capture the audio and visual content of the interviews. Its built-in transcribing feature was also used, but it turned out to be unreliable. The empirical data collected from workshops was collected by watching the recordings and by manually transcribing them. These transcriptions were then carefully synthesised into key findings and themes, which were further refined into conclusions, and the framework was adjusted accordingly. Chapter 4.2 describes the evolution of the framework in more detail.

3.3 Reliability and validity criteria

This chapter lists the reliability and validity criteria of this research by discussing the challenges of case- and quantitative studies and contemplating strategies to tackle them.

Like quantitative research methods, the challenges around case-studies of concern their generalisability. The most common critique of case studies is that 1. theoretical knowledge is more important than practical; 2. they cannot be used for generalisations, and thus cannot contribute to science; 3. they are only useful for developing theories, not testing, or validating them; 4. bias towards verifying the researcher's preconceived ideas; and 5. the difficultness of summarizing and developing general propositions. Although some of the critique is due to the oversimplification of case studies, the issues in case studies are real, and the researcher cannot afford to overlook them. (Flyvberg 2006)

As mentioned before, the challenges of qualitative research methods revolve around generalisation. Quantitative researchers often criticize empirical data, such as interview data for being unobjective, unreliable, and impressionistic (Qu & Dumay 2011). Additionally, often even experts do not identify all the factors influencing the matter and are thus a limited means of producing information (Alastalo et al. 2017). On the other hand, the reliability and validity of qualitative research cannot be measured in the same way that quantitative, but rather from a quality point of view (Eskola & Suoranta 1998). Although there are challenges in qualitative research, there are measures to deal with them, for instance by considering the reliability and validity criteria for qualitative research; credibility, transferability, dependability, and confirmability (Shenton 2004).

The first and the most important criterion, credibility, refers to the alignment of findings and reality (Shenton 2004). To ensure credibility, this study utilises the following methods:

- *Adoption of well-established research methods.* This included the preliminary study of research methodologies, and the design of interview questions and data analysis processes. See chapters 1.3 and 3.2 for details
- *Developing an early familiarity with the culture of participating organisation.* This was an easy task for the researcher since he works in the case company
- *Ensure participant honesty.* Participants were encouraged to speak freely, and the fact that there were no correct or wrong answers was emphasised. Additionally, all participants are given the right to refuse the interview and withdraw from the study at any point

- *Negative case analysis.* Once a hypothesis was formed from the interview data, the empirical data, in this case the workshop recordings, were revisited to ensure correct conclusions
- *Peer scrutiny.* The examiner of this study, and the participants from the case company were given the opportunity to scrutinise this research and give feedback. This was done two times, first after the theory-based framework was formed and after it was refined according to the findings
- *Examination of previous research findings.* The results of the study were compared to previous research findings, such as the one by Janssen et al. (2020), to assess the alignment with them. See chapter 6 for details

The second criterion, transferability concerns the extent to which the findings of this study can be applied into other cases (Shenton 2004). This is only possible under certain conditions, since qualitative studies do not produce statistically valid generalisations in a way that quantitative studies do (Eskola & Suoranta 1998). To understand the boundaries of transferability, one must understand the boundaries of the study, such as the number of organisations and participants, data collection methods, time-period of data collection, restrictions of the people contributing to the data, etc. This study does not aim to create an almighty framework that applies to every AI system, but rather give data governance guidelines for sustainable AI, and a place to start. Additionally, this study aims to contribute to sustainable AI research by addressing the data issues behind it.

The third criterion, dependability refers to whether the same results could be obtained if the research was repeated in the same context with the same methods. To address this criterion, the design and implementation, as well as the data gathering process has been described in detail. Additionally, the effectiveness of the process undertaken was evaluated. See chapter 3.3 for details. (Shenton 2004)

The last criterion, confirmability, refers to whether other researchers have made the same conclusions for the same phenomena (Eskola & Suoranta 1998). To address this, the findings were compared to AI governance frameworks to see if they supported each other. This criterion is linked with the examination of previous research findings done to address the criterion of credibility. Additionally, addressing this issue requires the researcher to admit his own predispositions to reduce investigator bias (Shenton 2004). Therefore, the researcher must admit that he may be biased towards verifying his preconceived ideas, and thus the public scrutinization of this research is more than necessary.

4. FINDINGS

In this chapter, the findings from the workshops are reported. First, the themes discussed in the workshops are described. Second, the evolution of the data governance for sustainable AI framework is depicted.

4.1 Themes

From the empirical data collected from the workshops, nine main themes could be identified. These themes are listed in table 7 below and explained in this chapter.

Table 7. *The themes identified from the empirical data*

Theme number	Theme description	Workshop
1	Data governance activities	1
2	Incommensurate elements	1
3	Why are these elements included and why are the rest left out	1
4	Drawing borders	1
5	Add data processes	2
6	Data strategy	2
7	Order of data governance activities	2
8	Supporting activities	2
9	Algorithmic bias	2
10	What parts of sustainable AI data governance does not cover	1 & 2

Theme 1. *Data governance activities*

The first issue discussed in the first workshop was that “data governance principles” is not a depicting term for the data governance rows listed. “Principle” as a word means the basic rule, truth, or proposition that serves as a foundation for something. None of the data governance rows do not fit to this description. A better term for them would be “activity”, since they are things that need to happen or need to be done.

Theme 2. *Incommensurate elements*

The second theme of discussion was that the AI governance elements, as well as data governance activities listed in the framework are not commensurate. Some of the elements cover much more than others, for instance, *Technical* AI governance element covers much more than *Informational* and *Metadata* as a data governance activity is much smaller than *Data policies, processes, and standards*. Another issue with AI governance elements and data governance activities listed is that some of them are overlapping. For instance, *Ethical* AI governance element overlaps with all other elements. Thus, the AI governance elements and data governance activities needed rethinking and reorganising in a way that they are in the same level of abstraction and have the same level of importance.

Theme 3. *Why are these elements included and why are the rest left out*

Another dilemma with AI governance elements and data governance activities were the reasons why some of them were chosen to the framework and some left out. In this research, AI governance elements and data governance activities listed in the first version of the framework were gathered from literature, and the researcher made choices on what to keep and what to leave out based on his own knowledge and intuition. Therefore, the arguments for why something was included or left out were not good enough. Participants 1, 2, and 3 suggested that instead of gathering AI governance elements and data governance activities from literature, one should choose one AI governance and one data governance and build the solution around them. The justification for going with a ready-made framework would be that they are market best practices, and thus pragmatically proven to work.

If you take a few frameworks and synthesise them, there is a danger that you come up with activities or elements that overlap. Thus, it would be a good idea to choose one "best practice" framework and go with it.

- Participant 1

On the other hand, participant 4 argued that when creating new, one should not be afraid to "draw own lines" and make own conclusions on which are the elements of AI or data governance in the context of sustainable AI. Science is about creating new, and you should not be too worried if it fits perfectly with previous results. This is especially true in the field of AI governance, since the field itself is very immature and thus there is not a very broad consensus on what does in cover and does it not. One must keep in mind that if the outcome of this research differs a lot from previous researcher on this topic, the researcher must be able to justify the results it and explain why they differ.

Theme 4. *Drawing borders*

Participants 2 and 3 stated that in these kinds of situations, where data governance activities are used to solve business problems, one should not focus on whether an activity is categorised into data governance or not. For instance, data governance and data management overlap a lot, and instead of drawing hard lines between them, the more sensible thing is to look at the data supply chain of an AI system, and assign the roles, responsibilities, and accountabilities from the supply chain perspective. Regarding business needs, it is more important that these do not overlap from the supply chain perspective, rather than from the data governance versus other types of governance point of view.

The purpose of data or AI governance is to create business value, and thus first thinking about technology is not smart. Technology must be seen as a mean to reach a goal, which serves a business purpose.

- Participant 2

Theme 5. *Add data processes*

In the first workshop, one participant briefly said that the component Data Processes in DGI data governance framework refers to data governance implementation process, and thus should be left out. Participants 7 and 8 in the second workshop argued against this, saying that data processes are at the core of data governance, and it should be in the framework.

Theme 6. *Data strategy*

In the second workshop, the data governance activity *Data strategy* sparked most discussion. The participants, especially participant 5, did not see it as an activity, pointing out that it is rather something that defines what kind of data is used for what purposes, what are the business areas it is used for, and what are the strategic choices behind this. Although data strategy was not seen as a part of data governance itself, the goals and standards derived from it are a big part of it. Additionally, data strategy as a term was argued by participants 5 and 8 to be misleading, since in this context it overlaps with AI and business strategy. This overlap issues had the same analogy as the overlapping of data governance with other activities, and thus one of the participants suggested that the word "data" should be removed.

You need strategic goals, but in this the term "data strategy" is a bit misleading since it requires aspects of business and AI strategies too.

- Participant 8

Theme 7. *Order of data governance activities*

In addition to the data governance activities being incommensurate, they were not ordered in any way. Participants 7 and 8 stated that listing the order in which the activities should be done is a common approach with data governance frameworks in general, and it would bring more structure to this framework. They recommended a top-to-bottom approach, where executing an activity will affect how the activity below it will be done. Additionally, participant 8 suggested the following logic in ordering them: 1. the goals/objectives; 2. what needs to be done; 3. how they are implemented; and 4. who does and what.

Theme 8. *Supporting activities*

The supporting activities listed on top of the framework were seen as a too detailed thing for the framework, since it aims to be a high-level abstraction and not a detailed instruction for implementation. However, participants 6 and 8 said that when one starts to implement data governance for sustainable AI, then these supporting activities would come in handy as they can be seen as a success factor for everything the framework covers. They are not part of the high-level approach that this framework aims to offer, but rather something that needs to be considered at the operative level. Participant 8 suggested that these could be studied later, if or when organisations start to use this framework.

Theme 9. *Algorithmic bias*

Another big topic of the second workshop was how to prevent algorithmic bias with data governance. Algorithmic bias can be caused by a flaw in the algorithm's design, or the unintended decisions made on the data inputted. All participants agreed that the algorithm's design-related issues are beyond data governance's scope, but the data related biases can be influenced. Additionally, participant 7 pointed out that data governance should not only take the inputted data into account, but also the outputted data. In simple cases of data governance, the aim is to control the quality of data that is inputted into a system, but in the context of AI data, quality related questions become more complex: can the data and information generated by an algorithm be trusted? How do we make sure that it is unbiased? Thus, data quality control brought by the means of data governance should be included in the inputted data, as well as the outputted.

The outcome of an algorithm is also data, and if the outcome of an algorithm is biased, that can also be seen as a data quality problem, but at the other end of the algorithm.

- Participant 7

Theme 10. *What parts of sustainable AI data governance does not cover*

A topic discussed in both workshops was what parts of sustainable AI does data governance cover and what it does not. Participant 1 raised this issue in the first workshop by giving an example of consent management chains starting from input data and stretching out to the analytics products created from output data. The argument was that the gap between data and AI governance is usually that data governance rarely takes the secondary use of data into account, for instance, the data produced by analytics. Matters such as who owns the data, who has the right to use it, what are the identity and access procedures, can it be sold to third parties, how sensitive it is, are something that needs to be tackled with both input and output data. This topic was also discussed in the second workshop alongside theme 9.

4.2 Framework

As mentioned earlier, the adjustment of the framework happened in two phases, after each workshop. After the first workshop, theme1 resulted in changing the umbrella term Data Governance Principles to Data Governance Activities. Themes 2 and 3 had the biggest impact on the framework, due to them the elements of AI governance and data governance activities were almost completely changed. Participants 1 and 2 suggested choosing the AI governance element from AIGA Hourglass Model of AI Governance by Mäntymäki et al (2022b), with one modification: *Environment* layer should be renamed into *Ecosystem*, since the word “environmental” often refers to nature or climate-related issues. As for data governance activities, participants 3 and 4 suggested the DGI Data Governance framework by DGI with some modifications. Components 1 and 2 would be merged into *Objectives and Key Results*, 4 and 5 into *Decision Rights and Accountabilities*, and 7, 8, and 9 into *Roles and Responsibilities*. The justification for modifying these ready-made data governance activities was to make the data governance activities as commensurate as possible. Table 8 in the next page depicts the state of the framework after the first workshop.

Table 8. *The framework after the first workshop*

Supporting Activities		
Change management	**Collaboration**	**Communication**
To help the behavioural changes required to implement data and AI governance to stick	To make data and AI governance activities cross-functional, and to build trusted data-sharing networks	To ensure everyone know why data and AI governance exists, how they are implemented and carried out, and what is their role in it

Data Governance Activities	**AI Goverance Elements**		
	AI System	**Organisation**	**Ecosystem**
Data Strategy	To include the prevention of technical risk and vulnerabilities of AI system data in strategies	To include the alignment of data activities, business goals and organisational values related to AI in strategies	To include legal, regulatory, value, and norm compliance related to AI to strategies
Data Governance Metrics	To measure the impact of data governance activities on AI systems	To measure the impact of data governance activities on organisations and businesses developing or using AI	To measure the impact of data governance activities on AI ecosystem and the level of alignment with laws, regulations, values, and norms
Data policies, rules, definitions, and standards	To have the necessary data policies, rules, definitions, standards to ensure human control over AI system data, prevent technical vulnerabilities, and foster expertise	To have the necessary data policies, rules, definitions, standards to reach business goals and to ensure alignment with organisational values	To have the necessary data policies, rules, definitions, standards to ensure positive impact on AI ecosystems, and alignment with laws, regulations, values, and norms
Decision rights and accountabilities	To ensure adequate decision rights and accountabilities to govern data and AI systems	To ensure adequate decision rights and accountabilities to reach business goals and value alignment related to AI systems	To have the necessary decision right and accountabilities to ensure positive impact on AI ecosystems, and alignment with laws, regulations, values, and norms
Controls	To prevent technical risks and vulnerabilities of AI system data	To prevent business and value comliance related risk regarding AI systems	To prevent legal, regulatory, value, and norm compliance risks related to AI system data
Roles and reposibilities	To ensure adequate roles and responsibilities to govern data and AI systems	To ensure adequate roles and responsibilities to reach business goals and value alignment related to AI systems	To ensure adequate roles and responsibilities to positively impact AI ecosystem, and alignment with laws, regulations, values, and norms

The second workshop did not change the framework as profoundly as the first one, but nevertheless it resulted in a good few improvements. Firstly, the researcher decided to add Data Processes to the framework based on theme 5. The row was also separated from Data Governance Activities since it is not categorically one of them. Second, the term *Data Strategy* was changed to *Objectives and Key Results* according to suggestion by participant 6 regarding theme 6. The third change was to order the data governance activities according to the order suggested in theme 7. Thus, they were ordered and grouped in the following way:

1. What are the goals for data governance?
    a. Objectives and Key Results
2. What needs to be done to achieve them?
    a. Decision Rights and Accountabilities
    b. Data Policies, Rules, Definitions, and Standards
3. The implementation of the thing decided in phase 2
    a. Roles and Responsibilities
    b. Data Processes
4. Monitoring and risk management
    a. Data Governance Metrics
    b. Controls

Lastly, the supporting activities were removed from the framework due to theme 8. The state of the framework after the second workshop is depicted below in table 9.

***Table 9.** The framework after the second workshop*

	AI Goverance Elements		
	AI System (technical)	**Organisation**	**Ecosystem**
Objectives & Key Results	To include the prevention of technical risk and vulnerabilities of AI system data in strategies	To include the alignment of data activities, business goals and organisational values related to AI in strategies	To include AI related legal, regulatory, and societal value and norm compliance to strategies
Data Governance Activities **Decision Rights and Accountabilities**	To ensure adequate decision rights and accountabilities to govern data and AI systems	To ensure adequate decision rights and accountabilities to reach business goals and value alignment related to AI systems	To have the necessary decision right and accountabilities to ensure positive impact on AI ecosystems, and alignment with laws, regulations, values, and norms
Data Policies, Rules, Definitions, and Standards	To have the necessary data policies, rules, definitions, standards to ensure human control over AI system data, prevent technical vulnerabilities, and foster expertise	To have the necessary data policies, rules, definitions, standards to reach business goals and to ensure alignment with organisational values	To have the necessary data policies, rules, definitions, standards to ensure positive impact on AI ecosystems, and alignment with laws, regulations, values, and norms
Roles and Responsibilities	To ensure adequate roles and responsibilities to govern data and AI systems	To ensure adequate roles and responsibilities to reach business goals and value alignment related to AI systems	To ensure adequate roles and responsibilities to positively impact AI ecosystem, and alignment with laws, regulations, values, and norms
Data Processes	To have standardised, documented, and repeatable processes for AI system data	To have the necessary data processes to ensure alignment with organisational values	To have the necessary data processes to ensure positive impact on AI ecosystems, and alignment with laws, regulations, values, and norms
Data Governance Metrics	To measure the impact of data governance activities on AI systems	To measure the impact of data governance activities on organisations and businesses developing or using AI	To measure the impact of data governance activities on AI ecosystem and the level of alignment with laws, regulations, values, and norms
Controls	To prevent technical risks and vulnerabilities of AI system data	To prevent business and value comliance related risk regarding AI systems	To prevent legal, regulatory, value, and norm compliance risks related to AI system data
	Middle section: what is the purpose of each activity from AI governance perspective		

5. DISCUSSION

In the recent decades, AI has gone through a global upsurge, which is not going to stop any time soon. This has been facilitated by rapid technological advancements, and the increase in the availability of data, especially big data. Although the benefits of AI, such as improvements in healthcare and increases in process efficiencies are indisputable, there are many potential risks and issues that need to be addressed. To ensure AI sustainability, tackling these issues is an utmost important mission, and the aim of this research is to contribute to that.

The goal of this research was to examine how to positively influence AI sustainability with the means of data governance. Since an AI algorithm takes data as input and its output is also data, ensuring data quality is an important part of sustainable AI. Data quality has been identified as an AI governance challenge in the literature, as well as in this research. Consequently, previous literature has also recognised data governance as an important part of AI governance and sustainable AI as well.

The AI governance elements of *AI System* and *Ecosystem* on the framework can also be identified from other literature on the topic. For instance, the AI associated risks and guidelines from the framework by Wirtz et al. (2022) can all be mapped into these two elements: the technological, data, and analytical layer is included in *AI System*, and the rest are incorporated in *Ecosystem*. The same mapping could be done with the model by Gasser & Almeida (2017), the social and legal; and ethical are included in *Ecosystem* and technical in *AI System*. What does not occur in only but the AIGA model is the *Organisation* element, which considers AI from a business and organisational value perspective.

Data governance activities listed in the framework do not differ much or conflict with other data governance frameworks or literature. A minor exception for this is the *Objectives & Key Results* activity, although it is not categorically a data governance activity, at least in this framework. From data governance literature one can identify that implementing data governance requires goal setting, or defining a mission and vision, but the notation “objectives and key results” was not used, at least in the literature reviewed in this study.

Although data governance has been in the literature identified as an important part of AI governance, there is little to no research on which aspects of data governance are important from an AI sustainability perspective and how it should be implemented. This is something which has, to the researcher’s knowledge, only been studied in this research and in the one by Janssen et al. (2020). Additionally, the middle part of the framework which explains the importance of each data governance activity for each AI governance element, is a feature not found in AI governance or data governance literature. Thus, this is something new that this research brings in the field of AI sustainability and governance.

When comparing the results of this research to the study by Janssen et al. (2020), one can note that they do differ but do not conflict. On the contrary to this research, Janssen et al. (2020) identified trusted data sharing frameworks as a data governance element for trustworthy AI, whereas it did not appear in the results of this research. Additionally, the majority data governance principles listed by Janssen et al. (2020) do not appear in the results of this research. However, these principles and the data governance activities listed in this study are not entirely comparable. Further comparison between this research and the one conducted by Janssen et al. (2020) is done in chapter 6.3.

6. CONCLUSIONS

Sustainable AI is defined as an AI system that supports all perspectives of sustainability, ecological, economic, and social. To succeed in this, it must foster diversity, be subjected to democratic principles, and be trusted. Additionally, a sustainable AI system is transparent, explainable, and values embedded into it must be aligned with values held by society. In addition to the technical perspective, the sustainability of AI must be examined from a human rights and ethical point of view. (van Wynsberghe 2021; Wilson & Van Der Velden 2022)

The research questions were:

1. What is the role of data in sustainable AI? (preliminary)
2. How to support sustainable AI with data governance?

The first research question was a preliminary one, and its aim was to understand the data supply chains around AI, and its data related challenges. The second one was the core of this research, and this research sought to answer it by creating a data governance framework to support sustainable AI. The answers to these questions and the conclusions derived from the findings are presented in the upcoming chapters 6.1 and 6.2.

6.1 The role of data in sustainable artificial intelligence

The algorithms of an AI system use large quantities of dynamic real-time data to work. Therefore, some aspects of data governance relevant for AI governance, especially the ones related to algorithmic systems (Janssen et al. 2020; Mäntymäki et al. 2022a). One of the biggest data-related challenges regarding AI is how to ensure data quality. Since data is a fundamental driver of an AI system, ensuring its quality is an utmost important task (Wirtz et al. 2019). Data quality-related challenges can be, for instance, the lack of standards for data collection, format, and storing (Sun & Medaglia 2019). If data quality is not ensured, the outcome of the algorithms using it may become biased or skewed (Janssen et al. 2020; Wirtz et al. 2019). Consequently, the correctness and transparency of data acquisition must be in place to prevent the misuse of data, especially when it is personally identifiable, sensitive, private, or confidential (Wirtz & Müller 2019).

Managing data quality is not only relevant for the input data, but for the output as well. To prevent biased or skewed output data from causing damage, it should be properly managed and governed. Furthermore, the data supply chains around AI systems should be accompanied with clear and continuous chains of accountability and responsibility. A practical example of an output data problem could be the breakage of consent management chain or algorithmic bias. Thus, the data governance activities selected for the framework should be applied to both the input and output data. However, if a flaw such as a programming or design error in the algorithm itself is causing biased or skewed outcomes regardless of the quality of the input data, the issue must be addressed by some other way. Figure 7 below illustrates how the outcome of an AI algorithm can be positively influenced by data quality management enabled by data governance.

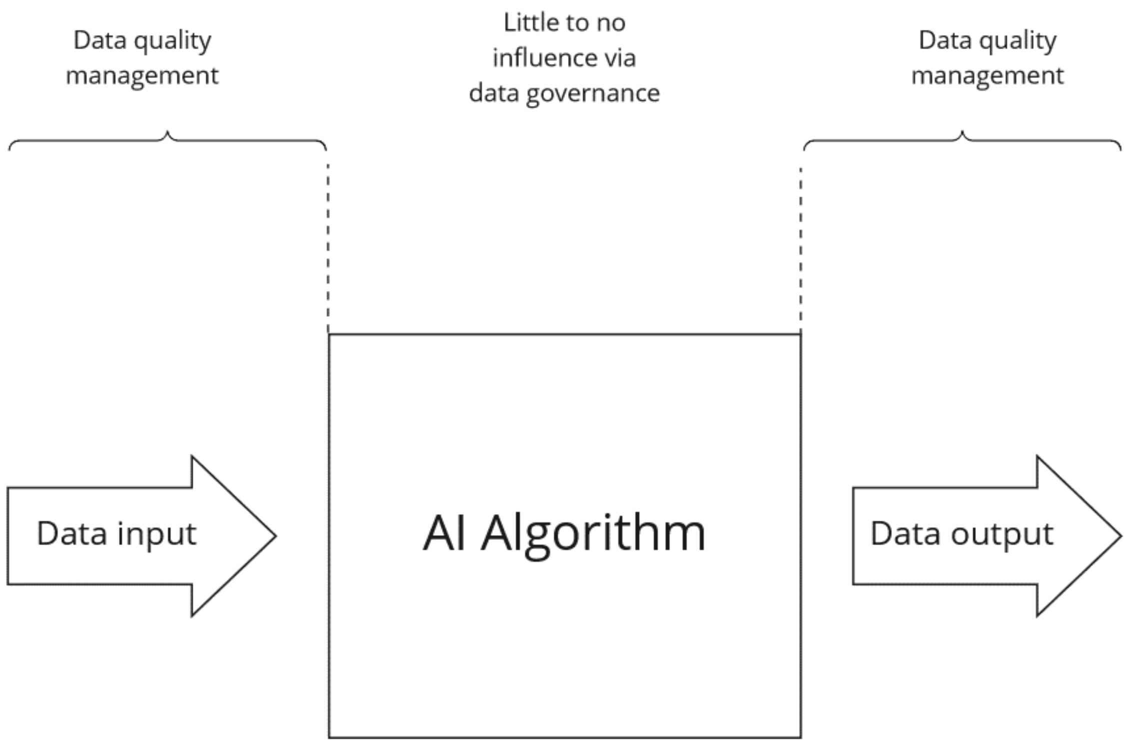

Figure 7. *Data governance influence on AI algorithms*

Although there are big challenges regarding AI system data, one must not forget that there are benefits in data and information processing with AI. A few examples of these would be improvements in information and big data processing, the capability of machine learning systems to continuously self-improve, the possibility to handle multi-dimensional and -variety data, etc. (Zuiderwijk et al. 2021)

6.2 Supporting sustainable artificial intelligence with data governance

Data governance gives the means to manage data quality by exercising authority, control, and shared decision-making over the management of AI system data. It helps by organising and implementing procedures, policies, structures, roles, and responsibilities and by enforcing outlining rules, decision rights, and responsibilities for effective data management (Ladley 2012), aiming to ensure responsible input data collection (Janssen et al. 2020), as well as supervision of output data. With the efforts of data governance, AI systems can be positively influenced towards sustainability.

In this research, a data governance framework for sustainable AI was formed. Its objective is to be a high-level abstraction of what needs to be considered when implementing data governance for sustainable AI, not to be a detailed set of instructions. Furthermore, the framework aims to illustrate how different data governance activities support the goals of different AI governance elements, and in which order the data governance activities should be done.

There are three AI governance elements in the framework, *AI System*, *Organisation*, and *Ecosystem* are selected from the AIGA AI Governance Framework by Mäntymäki et al (2022b), with a small change: the AI Governance element *Environment* was renamed to *Ecosystem*, since the word “environmental” often refers into nature or climate-related matters. First of the elements, *AI System* refers to the operation of the technical side of AI, including the design and development of AI systems, algorithms, and data operations in a way that they are transparent, explainable, contestable, and in line with its regulatory environment. Additionally, this element covers the designation of AI systems accountability and risk management. (Mäntymäki et al. 2022b)

The second AI governance element, *Organisation* concerns the organisation that uses the AI systems. This comprises of two parts, strategic and value alignment. The first part, strategic alignment refers to the alignment of AI and organisational strategy and defines what AI systems are intended to be used to achieve which business goals. The value alignment part requires organisational values to be aligned with the values embedded in their AI systems. Furthermore, this element requires organisations to determine their position on risks, such as regulatory or reputational. (Mäntymäki et al. 2022b)

Laws, regulations, societal values and norms, as well as other AI governance requirements coming from outside the organisation are covered in the third and last element, *Ecosystem*. The many emerging AI regulations, such as the EU AI Act, make the regulatory landscape rapidly evolving, and thus keeping up with them a crucial part of AI governance. Furthermore, many data-related regulations, such as GDPR, also concern these organisations since AI systems consume large amounts of data. In addition to legal and regulatory compliance, the values embedded into an organisation's AI systems must be in line with societal values and norms. Failing to do this can cause massive damage to the organisation's reputation, and thus have a negative impact on business. All things considered, the AI governance element *Ecosystem* not only requires organisations to keep up with the constantly evolving legal and regulatory requirements, but also ensure the ethics of their AI systems. (Mäntymäki et al. 2022b)

The data governance activities in the framework were selected from the components of DGI Data Governance framework by The Data Governance Institute with modifications. The framework contains the following data governance activities:

- *Objectives & Key Results* sets measurable goals for data governance activities. This is separated from data governance activities, since it does not categorically fit into them, but rather guides the implementation of data governance.
- *Decision Rights and Accountabilities* addresses on what kind of decision rights and accountabilities are needed to govern the data supply chain of an AI system.
- *Data Policies, Rules, Definitions, and Standards* includes the creation of policies, rules, definitions, and standards for effective data governance.
- *Roles and Responsibilities* designates the decision right into formal roles, such as data stewards, stakeholders, data governance offices, etc. and assign responsibilities for them.
- *Data Processes* is the implementation and operation of standardised, repeatable, and documented data processes that embody the policies, rules, definitions, and standards created earlier.
- *Data Governance Metrics* measure the value created by data governance, such as increase in revenue, costs cut, risks and vulnerabilities avoided, or data quality improved.
- *Controls* covers the management of data-related risks, including preventive and correcting measures.

These data governance activities can be grouped by the following way: what do we want to achieve (*Objectives & Key Results*), what needs to be done (*Decision Rights and Accountabilities, Data Policies, Rules, Definitions, and Standards*), implementation and operation (*Roles and Responsibilities, Data Processes*), and finally measuring and risk management (*Data Governance Metrics, Controls*). The activities have a top-down approach, meaning that the activities are done from top to bottom. It is to be noted that data governance does not end after the last activity on a framework, but it is rather a continuous process, which needs to be continuously developed by, for instance, re-iterating through the framework. Table 10 below illustrates the framework in its entirety.

Table 10. *Data governance for sustainable AI framework*

	AI Goverance Elements		
	AI System (technical)	**Organisation**	**Ecosystem**
Objectives & Key Results	To include the prevention of technical risk and vulnerabilities of AI system data in strategies	To include the alignment of data activities, business goals and organisational values related to AI in strategies	To include AI related legal, regulatory, and societal value and norm compliance to strategies
Data Governance Activities **Decision Rights and Accountabilities**	To ensure adequate decision rights and accountabilities to govern data and AI systems	To ensure adequate decision rights and accountabilities to reach business goals and value alignment related to AI systems	To have the necessary decision right and accountabilities to ensure positive impact on AI ecosystems, and alignment with laws, regulations, values, and norms
Data Policies, Rules, Definitions, and Standards	To have the necessary data policies, rules, definitions, standards to ensure human control over AI system data, prevent technical vulnerabilities, and foster expertise	To have the necessary data policies, rules, definitions, standards to reach business goals and to ensure alignment with organisational values	To have the necessary data policies, rules, definitions, standards to ensure positive impact on AI ecosystems, and alignment with laws, regulations, values, and norms
Roles and Responsibilities	To ensure adequate roles and responsibilities to govern data and AI systems	To ensure adequate roles and responsibilities to reach business goals and value alignment related to AI systems	To ensure adequate roles and responsibilities to positively impact AI ecosystem, and alignment with laws, regulations, values, and norms
Data Processes	To have standardised, documented, and repeatable processes for AI system data	To have the necessary data processes to ensure alignment with organisational values	To have the necessary data processes to ensure positive impact on AI ecosystems, and alignment with laws, regulations, values, and norms
Data Governance Metrics	To measure the impact of data governance activities on AI systems	To measure the impact of data governance activities on organisations and businesses developing or using AI	To measure the impact of data governance activities on AI ecosystem and the level of alignment with laws, regulations, values, and norms
Controls	To prevent technical risks and vulnerabilities of AI system data	To prevent business and value comliance related risk regarding AI systems	To prevent legal, regulatory, value, and norm compliance risks related to AI system data
	Middle section: what is the purpose of each activity from AI governance perspective		

The middle section of the framework explains why each of the data governance activities are relevant and how they interact with and affect different AI governance elements. This paragraph summarises this relationship. With the *AI System* element, the role of data governance is to create transparent data supply chains and metadata, used to develop AI data more appropriately and manage the output data in a controller manner. It is not enough that AI development itself is transparent if the data development around it is not. Regarding the *Organisation* element, data governance aims to ensure that data management is developed in a way which supports the organisations business strategies, and AI-related data operations are in line with organisational values. Lastly, the role of data governance when crossing over with the *Ecosystem* element is to ensure legal, regulatory, and societal value compliance in data gathering, usage, and distribution.

Although proper data governance can have a positive impact on AI sustainability, many parts of sustainable AI require governance measures outside its scope. An example of this could be when algorithmic bias is caused by a design or programming error in the algorithm itself, which makes it impossible to avoid it by the means of data governance. However, by properly governing the output data, these biases can be detected, and preventive measures can be applied.

6.3 Reliability and validity assessment

Chapter 3.3 listed the reliability and validity criteria for qualitative research to be credibility, transferability, dependability, and confirmability. This chapter reflects the study conducted to those and discusses reliability and validity of this research.

Fulfilling the criterion of credibility requires several measures. First and foremost, research methodologies were studied to adopt well-established research methods, as well as to design questions and data analysis processes. However, the interviews conducted had more resemblance to workshops than traditional interviews, which do not fall into this category. The rest of the credibility criteria, developing an early familiarity with the culture of participating organisation, ensure participant honesty, negative case analysis, peer scrutiny, and examination of previous research findings were done according to the description in chapter 3.3. Nonetheless, the researcher must admit that this is the first time he is doing research of this extent and thus fulfilling these credibility criteria. Although he might see these requirements fulfilled, a more experienced researcher may have another opinion.

The criterion of transferability was a difficult one to fulfil, since the scope of this research was not very broad, and qualitative studies do not produce statistically valid generalisations. Due to the scope of this research and the topic being quite unexplored, transferability of the results cannot be confirmed. To fully understand how transferable the results are, the data governance for sustainable AI framework needs to be pragmatically used and researched.

If this research would be repeated by following the design and implementation, as well as the data gathering process described in chapters 3.1 and 3.2, and the same workshop participants would be used, then the same results would likely be obtained, and thus the criterion of dependability would be fulfilled. However, using different participants may result in different results. All the AI governance experts were working with the AIGA project led by University of Turku, and their ideas about AI governance are heavily influenced by it, whereas other AI governance experts may have different views on the topic. The same issues concern the data governance experts since they are all from the same organisations and share the same perspectives.

At the time this research was conducted, there was hardly any research done around this topic, thus the confirmability of this research is difficult to evaluate. To the researcher's knowledge, there was only one study with a similar topic conducted by Janssen et al. (2020), which examined the data governance challenges and approaches for big data algorithmic systems, as well as lists data governance principles for big data algorithmic systems. When comparing the results of this research with the research by Janssen et al. (2020), one can note that their research lists data governance principles to follow, whereas this research lists data governance activities to be done. These principles and activities are not entirely comparable, but they the results do not conflict. The data governance principle *Evaluate data quality and bias* in their paper is in line with the emphasis on data quality in this research, and the principle *Data stewards* is included in the data governance activity *Roles and Responsibilities*. Additionally, both Janssen et al. (2020) and this research recognised the need for managing AI algorithm output data, as well as taking societal values and norms into account when drafting data policies, principles, procedures.

6.4 Limitations

Although the validity and reliability issues of a qualitative study were considered in this research, there are still limitations. The framework created was based on the literature on data and AI governance and refined in two workshops. This kind of framework would probably need a lot more refinement, but since the scope of this research is limited, the framework could not be further developed. Additionally, the framework has not been pragmatically used or tested.

Another limitation was that the empirical data was collected from members of the same case company, and thus the data does not represent a very broad set of experts. People from different organisations might have a different point of view to data governance AI governance, and with more workshops and involving people outside the case company may have produced different results. Although the number of workshops organised was small, the reliability of the empirical data gathered from them was ensured by considering the reliability and validity criteria of qualitative case studies.

The biggest limitation to this research is that all the AI governance experts and the case company itself participated in the AIGA project. Therefore, their ideas about AI governance are heavily influenced by it, whereas other AI governance experts may have different views on the topic. Since these AI governance experts recommended using the layers of the AIGA model as the AI governance elements for this framework, it raises a suspicion whether this recommendation was tendentious. However, this does not dismiss the results of this research as being biased, but rather something that should be noted when examining them.

6.5 Further research directions

As mentioned in the previous chapter, the framework needs to be further refined as well as practically tested. Therefore, the researcher suggests the development and the use of this framework in an organisation as avenues for further research. Furthermore, data governance in the context of AI needs to be further studied since there is hardly any research on this topic. This research direction could include usage of supporting activities, communication, change management, and collaboration. Additionally, the research on AI governance is relatively scarce when compared to, for instance, the amount of research on data governance, and there are hardly any empirical studies on implementing AI governance by using a conceptual framework, such as the AIGA Hourglass Model of AI Governance.

Other research directions collected from AI governance literature would be studying the interaction between the AI governance elements (Mäntymäki et al. 2022b). Second, the long-term societal risks and challenges of AI and machine learning are poorly understood, and thus need to be studied in order to better govern and regulate AI towards sustainability (Wilson & Van Der Velden 2022). Lastly, tackling the “pacing problem” of government regulation not being able to keep up with the increasing complexity, development, and risks imposed by emerging technologies, such as AI is an important further research avenue (Wirtz et al. 2022).

To summarise, the research on AI governance and sustainability is very scarce, although AI is a big part of our daily lives influencing many parts of it. The technical aspects of AI are constantly developing, leaving its governance and regulation lagging. Ensuring AI sustainability requires constant research and development of governance practices and regulations. A perquisite for this is to understand the societal impact of AI systems. Thus, AI should not be only studied from the technical point of view, but also from other areas of research as well.

REFERENCES

Abercrombie, N., Hill, S., & Turner, B. (1994). *The Penguin Dictionary of Sociology.* Sociology (Oxford). 600–. Print.

Adelman, S., Moss, L., & Abai, M. (2005). *Data Strategy (1st edition)*. Addison Wesley Professional.

Alastalo, M., Åkerman, M., & Vaittinen, T. (2017). *Asiantuntijahaastattelu*. In M. Hyvärinen, P. Nikander, & J. Ruusuvuori (Eds.), Tutkimushaastattelun käsikirja (pp. 214-232). Vastapaino.

Alhassan, I., Sammon, D., & Daly, M. (2019). *Critical success factors for data governance: a theory building approach.* Information Systems Management, 36(2), 98-110.

Amnesty International. (2021). *Dutch childcare benefit scandal an urgent wake-up call to ban racist algorithms*. Accessed 25.8.2022.

Anttila, P. (1998). *Tutkimisen taito ja tiedon hankinta.* Metodix.

Assefa, G., & Frostell, B. (2007). *Social sustainability and social acceptance in technology assessment: A case study of energy technologies.* Technology in society, 29(1), 63-78.

Berson, A., & Dubov, L. (2011). *Master data management and data governance (Second edition.)*. McGraw-Hill.

Bhansali, N. (2014). *Data governance : creating value from information assets (1st edition)*. CRC Press. https://doi.org/10.1201/b15034

Cai, L., & Zhu, Y. (2015). *The challenges of data quality and data quality assessment in the big data era.* Data science journal, 14.

Cheng, G., Li, Y., Gao, Z., & Liu, X. (2017). *Cloud data governance maturity model*. In 2017 8th IEEE International Conference on Software Engineering and Service Science (ICSESS) (pp. 517-520). IEEE.

Coeckelbergh, M. (2019). *Artificial intelligence: some ethical issues and regulatory challenges.* Technology and regulation, 2019, 31-34. https://doi.org/10.26116/techreg.2019.003

Coeckelbergh, M. (2020). *AI for climate: freedom, justice, and other ethical and political challenges*. AI Ethics. https://doi.org/10.1007/s43681-020-00007-2

DAMA International, Books24x7, Inc, Mosley, M., & Brackett, M. (2010). *The DAMA guide to the data management body of knowledge (DAMA-DMBOK guide), first edition.* Technics Publications.

De Fine Licht, K., & Folland, A. (2019). *Defining "social sustainability": towards a sustainable solution to the conceptual confusion.* Etikk i Praksis-Nordic journal of applied ethics, (2), 21-39.

The Data Governance Institute. *DGI Data Governance Framework.* https://datagovernance.com/the-dgi-data-governance-framework/. Accessed 8.11.2022

Doneda, D., & Almeida, V. A. F. (2016). *What Is Algorithm Governance?* IEEE Internet Computing, 20(4), 60–63. https://doi.org/10.1109/MIC.2016.79

Eryurek, E., Gilad, U., Lakshmanan, V., Kibunguchy, A., & Ashdown, J. (2021*). Data Governance: The Definitive Guide (1st edition).* O'Reilly Media, Inc.

Eskola, J., & Suoranta, J. (1998). *Johdatus laadulliseen tutkimukseen*. Vastapaino.

EU AI Act. (2022). *The EU Artificial Intelligence Act -website.* Accessed 18.7.2022. https://artificialintelligenceact.eu/

European Commission (2020). *WHITE PAPER On Artificial Intelligence - A European approach to excellence and trust.* https://ec.europa.eu/info/files/white-paper-artificial-intelligence-european-approach-excellence-and-trust_en. Accessed 15.6.2022.

Flyvbjerg, B. (2006). *Five misunderstandings about case-study research.* Qualitative Inquiry 12(2), 219–245. https://doi.org/10.18261/ISSN1504-2928-2004-02-02

Gasser, U., & Almeida, V. A. F. (2017). *A Layered Model for AI Governance*. IEEE Internet Computing, 21(6), 58–62. https://doi.org/10.1109/MIC.2017.4180835

Gill, K.S. (2020). *AI&Society: editorial volume 35.2: the trappings of AI Agency.* AI & Soc 35, 289–296. https://doi.org/10.1007/s00146-020-00961-9

Gregory, R. W., Kaganer, E., Henfridsson, O., & Ruch, T. J. (2018). *IT consumerization and the transformation of IT governance.* Mis Quarterly, 42(4), 1225-1253.

Gupta, K. (2019). *Artificial intelligence for governance in india: Prioritizing the challenges using analytic hierarchy process (AHP).* International Journal of Recent Technology and Engineering, 8(2), 3756–3762. https://doi.org/10.35940/ijrte.B3392.078219

Hanif, A., Kusumasari, T. F., & Andreswari, R. (2019). *Analysis and Design of Data Synchronization Algorithm for Master Data Management Tools Based on Open Source Platform at PT. XYZ.* In 2019 International Conference on Electrical Engineering and Informatics (ICEEI) (pp. 244-249). IEEE.

Haynes, D. (2018). *Metadata for Information Management and Retrieval: Understanding metadata and its use.* In Metadata for Information Management and Retrieval (2nd ed.). Facet Publishing. https://doi.org/10.29085/9781783302161

Henley, J. (2021). *Dutch government resigns over child benefits scandal.* The Guardian. Accessed: 25.8.2022.

Hirsjärvi, S., & Hurme, H. (2008). *Tutkimushaastattelu : teemahaastattelun teoria ja käytäntö.* Helsinki: Gaudeamus Helsinki University Press.

HLEG. (2019). *A Definition of AI: Main Capabilities and Disciplines: Definition Developed for the Purpose of the AI HLEG's Deliverables.* Brussels: European Commission.

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). https://doi.org/10.1016/j.giq.2020.101493

Janssen, M., & Kuk, G. (2016). *The challenges and limits of big data algorithms in technocratic governance*. Government Information Quarterly, 33(3), 371-377.

Johnson, J. D. (2021). *Algorithm*. In Information : keywords. Columbia University Press. 31-43

Keller, P., & Drake, A. (2021). *Exclusivity and Paternalism in the public governance of explainable AI.* Computer Law & Security Review, Volume 40. https://doi.org/10.1016/j.clsr.2020.105490.

Kertysova, K. (2018). *Artificial Intelligence and Disinformation: How AI Changes the Way Disinformation is Produced, Disseminated, and Can Be Countered.* Security and Human Rights, 29(1-4), 55–81. https://doi.org/10.1163/18750230-02901005

Khatri, V., & Brown, C. V. (2010). *Designing data governance.* Communications of the ACM, 53(1), 148-152

Kindylidi, I., & Cabral, T. S. (2021). *Sustainability of AI: the case of provision of information to consumers.* Sustainability, 13(21). https://doi.org/10.3390/su132112064

Koskelo, P., Lassila, J., Lundstedt, T., Miklóssy, K. (2022*). Diktaattorin käsikirja – eli miten liberaalia demokratiaa puolustetaan.* Ajatuspaja Libera.

Kurniawan, D. H., Ruldeviyani, Y., Adrian, M. R., Handayani, S., & Pohan, M. R. (2019). *Data governance maturity assessment: A case study in IT Bureau of Audit Board.* In 2019 International Conference on Information Management and Technology (ICIMTech) (Vol. 1, pp. 629-634). IEEE.

Ladley, J. (2012). *Data governance: How to design, deploy, and sustain an effective data governance program.* 1st edition. Waltham, Mass: Morgan Kaufmann. Print.

Martin, A., Dmitriev, D., & Akeroyd, J. (2010). *A resurgence of interest in information architecture*. International journal of information management, 30(1), 6-12.

Mensah, J. (2019). *Sustainable development: Meaning, history, principles, pillars, and implications for human action: Literature review.* Cogent social sciences, 5(1). https://doi.org/10.1080/23311886.2019.1653531

Meredith, J. (1998). *Building Operations Management Theory through Case and Field Research*. Journal of operations management. 441–454. Web.

Minkkinen, M., Zimmer, M. P., & Mäntymäki, M. (2022). *Co-Shaping an Ecosystem for Responsible AI: Five Types of Expectation Work in Response to a Technological Frame. Information Systems Frontiers*, 1-19. https://doi.org/10.1007/s10796-022-10269-2

Mäntymäki M., Minkkinen M., Birkstedt T. and VIljanen M. (2022a). *Defining organizational AI governance.* AI Ethics. https://doi.org/10.1007/s43681-022-00143-x

Mäntymäki M., Minkkinen M., Birkstedt T. and VIljanen M. (2022b). *Putting AI Ethics into Practice: The Hourglass Model of Organizational AI Governance.* arXiv.org.

Nitzberg, M., & Zysman, J. (2022). *Algorithms, data, and platforms: the diverse challenges of governing AI*. Journal of European Public Policy, ahead-of-print(ahead-of-print), 1–26. https://doi.org/10.1080/13501763.2022.2096668

NSW Government. (2021). *AI Assurance Framework*. https://www.digital.nsw.gov.au/policy/artificial-intelligence/nsw-ai-assurance-framework. Accessed 4.9.2022

Permana, R. I., & Suroso, J. S. (2018). Data governance maturity assessment at PT. XYZ. Case study: data management division. In 2018 International Conference on Information Management and Technology (ICIMTech) (pp. 15-20). IEEE.

Plotkin. (2014). *Data stewardship : an actionable guide to effective data management and data governance (1st edition).* Elsevier.

Puusa, A. & Juuti P. (editors) (2020). *Laadullisen tutkimuksen näkökulmat ja menetelmät*. Chaper 2.

Qu, S. Q., & Dumay, J. (2011). The qualitative research interview. Qualitative Research in Accounting & Management (3), 238–264.

Redman, T. C. (2016). *Bad Data Costs the U.S. $3 Trillion Per Year*. Harvard Business Review. https://hbr.org/2016/09/bad-data-costs-the-u-s-3-trillion-per-year

Rishartati, P., Rahayuningtyas, N. D., Maulina, J., Adetia, A., & Ruldeviyani, Y. (2019). *Maturity Assessment and Strategy to Improve Master Data Management of Geospatial Data Case Study: Statistics Indonesia.* In 2019 5th International Conference on Science and Technology (ICST) (Vol. 1, pp. 1-6). IEEE.

Rohde, F., Gossen, M., Wagner, J., & Santarius, T. (2021). *Sustainability challenges of Artificial Intelligence and Policy Implications*. Ökologisches Wirtschaften-Fachzeitschrift, 36(O1), 36-40.

Sarsfield. (2009). *The data governance imperative a business strategy for corporate data. (1st edition).* IT Governance Pub.

Saunders, M. N. K., Lewis, P. & Thornhill, A. (2019). *Research Methods for Business Students.* 8. print, Harlow, England, Pearson Education.

Shenton, A.K. (2004). *Strategies for ensuring trustworthiness in qualitative research projects.* Education for Information, 22(2), 63–75. https://doi.org/10.3233/EFI-2004-22201

Soares. (2014). *Data governance tools : evaluation criteria, big data governance, and alignment with enterprise data management.* MC Press Online, LLC.

Solomon, J. (2020). *Corporate governance and accountability.* John Wiley & Sons.

Sulanen, S. (2021). *Improving financial data quality through data governance.* urn.fi/URN:NBN:fi:tuni-202104263618

Sun, T. Q., & Medaglia, R. (2019). *Mapping the challenges of Artificial Intelligence in the public sector: Evidence from public healthcare.* Government Information Quarterly, 36(2), 368-383.

van Wynsberghe, A. (2021). *Sustainable AI: AI for sustainability and the sustainability of AI*. AI and Ethics, 1(3), 213-218.

Vilminko-Heikkinen, R. (2017). *Data, Technology, and People: Demystifying Master Data Management.*

Vilminko-Heikkinen, R., & Pekkola, S. (2013, January). *Establishing an organization's master data management function: a stepwise approach.* In 2013 46th Hawaii international conference on system sciences (pp. 4719-4728). IEEE.

Weill, P., Ross, J. W. (2004). *IT Governance: How Top Performers Manage IT Decision Rights for Superior Results.* Harvard Business School Press.

Wilson, C., & Van Der Velden, M. (2022). *Sustainable AI: An integrated model to guide public sector decision-making*. Technology in Society, 68. https://doi.org/10.1016/j.techsoc.2022.101926

Wirtz, B. W., & Müller, W. M. (2019). An integrated artificial intelligence framework for public management. Public Management Review, 21(7), 1076–1100. https://doi.org/10.1080/14719037.2018.1549268

Wirtz, B. W., Weyerer, J. C., & Geyer, C. (2019). Artificial Intelligence and the Public Sector-Applications and Challenges. International Journal of Public Administration, 42(7), 596–615. https://doi.org/10.1080/01900692.2018.1498103

Wirtz, B. W., Weyerer, J. C., & Kehl, I. (2022). *Governance of artificial intelligence: A risk and guideline-based integrative framework.* Government Information Quarterly. https://doi.org/10.1016/j.giq.2022.101685

Wolford, B. (2020). *What is GDPR, the EU's new data protection law?* https://gdpr.eu/what-is-gdpr/

Zuiderwijk, A., Chen, Y. C., & Salem, F. (2021). *Implications of the use of artificial intelligence in public governance: A systematic literature review and a research agenda.* Government Information Quarterly, 38(3). https://doi.org/10.1016/j.giq.2021.101577

APPENDIX A: INTERVIEW QUESTIONS IN ENGLISH

Introduction

Based on the literature on data and AI governance, I have developed a theory-based data governance for sustainable AI framework. On the horizontal cells, we have the elements of AI governance, and in the vertical cells on the left we have data governance activities. Between them in the middle, we have explained how the data governance activities support the AI governance activities, which contribute to sustainable AI. On top of this matrix, we have listed the activities that support the synergy of data and AI governance. The aim of this workshop is to refine and develop this framework.

There are 6 AI governance elements. The first element, social, governs the socio-technical infrastructure of AI systems and strives to ensure social equality and compliance, as well as human well-being. The ethical element concerns the ethical concerns and principles with AI systems and aims to guarantee the respect of fundamental values and rights, by establishing a human-centric, trustworthy AI system. The economic element addresses the risks of disrupting economic systems, such as unemployment caused by AI. Additionally, transparency of AI within business processes must be ensured, as well as fairness is global market competition. The informational element tackles the AI risks of information manipulation, disinformation, propaganda, censorship, and freedom of speech. The next element, legal and regulatory, addresses the general principles of AI regulation, such as creation of institutions and government authorities, as well as allocation of accountabilities, responsibilities, and supervisory authority for AI system regulation. The last element, technical, concerns the algorithms and data of an AI system.

Introductory

What are your first thoughts on the framework?

Direct

Do you think there are more elements to AI governance than these?

Do you think there are more relevant data governance activities than these?

Do you think there are more relevant supporting activities than these?

Is there any unnecessary AI governance element in the theory-based framework?

Is there any unnecessary data governance element in the theory-based framework?

Are any of the supporting activities unnecessary?

Would you change the structure of the framework?

What changes would you make to the structure of the framework?

Indirect

How do you think data governance supports sustainable AI?

What do you think about the overall structure of the framework?

What do you [*participant name*] think about […]?

Probing questions

What do you mean by […]?

Could you give me a more detailed explanation about […]?

Do you have any examples of […]?

Anything else you would like to add?

Could you explain why do you think that […]?

Structuring questions

Let us talk about […] shall we?

Let us get back to the topic of […], shall we?

Other

[Silence, best of participants break it themself]

[Nodding]

[Repeating what has been said]

Closing remarks

There is only 5 minutes left, so let us start wrapping this up. Next, I will go through the workshop materials and recordings, and further refine framework based on your input. My goal is to finish my thesis by Christmas.

Do you have questions related to the research itself?

Are you interested in hearing about the results of the study?

Thank you for participating!

APPENDIX B: INTERVIEW QUESTIONS IN FINNISH

Johdanto

Perustuen data ja AI governance kirjallisuuteen, olen kehittänyt teoreettisen kestävää tekoälyä tukevan data governance viitekehyksen. Ylähäällä vaakasuorissa sarakkeissa on esitetty AI governacen osa-alueet, ja vasemmalla pystysuorissa sarakkeissa data governance aktiviteettejä. Keskiosassa kerrotaan, kuinka data governance aktiviteetit tukevat AI governance toimintoja, mikä taas tukee kestävää tekoälyä. Matriisin yläpuolella on listattu tätä tukevia aktiviteetteja. Tämä työpajana tavoitteena on kehittää ja jalostaa tätä viitekehystä.

AI governancella on 6 elementtiä. Ensimmäinen, sosiaalinen elementti, koskee tekoälyn sosio-teknistä infrastruktuuria, ja pyrkii varmistamaan sosiaalisen tasa-arvon ja tukemaan ihmisten hyvinvointia. Eettinen elementti taas koskee tekoälyn eettisiä huolia ja periaatteita, ja pyrkii tekemään tekoälystä ihmiskeskeisempiä ja ihmisoikeuksia kunnioittavia. Taloudellinen elementti käsittelee tekoälyn aiheuttamia taloudellisia häiriöitä, kuten työttömyys. Informationaalinen elementti koskee tekoälyn generoimaa disinformaatiota, propagandaa, sekä sen tukemaa sensuuria ja sanavapauden rajoittamista.

Seuraava elementti, laki ja sääntely, käsittelee tekoälyn sääntelyn yleisiä periaatteita, kuten instituutioiden ja valtion viranomaisten luomista sekä tekoälyn säätelyn vastuiden ja valvontavallan jakamista. Viimeinen elementti, tekninen, koskee tekoälyjärjestelmän algoritmeja ja dataa.

Johdattelevat

Mitkä ovat ensimmäiset ajatukset viitekehyksestä?

Suorat

Puuttuuko mielestänne viitekehyksestä jokin AI governancen osa-alue?

Puuttuuko mielestänne viitekehyksestä jokin data governancen osa-alue?

Puuttuuko mielestänne viitekehyksestä jokin tukeva toiminto?

Onko viitekehyksessä joku turha AI governancen osa-alue?

Onko viitekehyksessä joku turha data governancen osa-alue?

Onko viitekehyksessä joku turha tukeva toiminto?

Muuttaisitko viitekehykset rakennetta?

Mitä muutoksia tekisit viitekehyksen rakenteeseen?

Epäsuorat

Olisiko viitekehys hyödyllinen, jos haluat implementoida data ja AI governancea?

Missä järjestyksessä sitä veisit?

Miten mielestänne data governance tukee AI governancea?

[*Osallistujan nimi*] mitä mieltä sinä ole [..]?

Tarkentavat

Mitä tarkoitat […]?

Voitko tarkentaa […]?

Voitko antaa esimerkin […]?

Onko vielä jotain lisättävää?

Miksi olette sitä mieltä että […]?

Rakennetta luovat

Puhutaanko […]?

Palataanko aiheeseen […]?

Muut

[Hiljaisuus, tavoite että osallistujat rikkovat sen itse]

[Nyökkäily]

[Sanojen toistaminen]

Loppupuheet

Aikaa on enää 5 minuuttia jäljellä, joten lopetellaan tähän. Seuraavaksi käyn workshop-materiaalit ja nauhoitteet läpi, ja jatkojalostan sen pohjalta viitekehystä. Tavoitteenani on saada diplomityöni valmiiksi jouluun mennessä.

Onko teillä itse tutkimukseen liittyviä kysymyksiä?

Onko teillä kiinnostusta kuulla tutkimuksen tuloksista?

Kiitos osallistumisesta!