Tampere University

Lauri Eneh

BUSINESS MODELS FOR SUSTAINABLE AI CONSULTING

Conceptualizing interorganizational sustainable AI services through business model design

Faculty of Management and Business
Master of Science Thesis
October 2022

ABSTRACT

Lauri Eneh: Business models for sustainable AI consulting – Conceptualizing interorganizational sustainable AI services through business model design
Master of Science Thesis
Tampere University
Master's Degree Programme in Industrial Engineering and Management
Examiners: Prof. Teemu Laine and Assoc. Prof. Tuomas Korhonen
October 2022

The utilization of Artificial Intelligence (AI) systems is growing rapidly in various domains. Together with the great benefits, the potential for unintended negative effects caused by these systems has been recognized. Indeed, there is an increasing consensus among researchers and practitioners on the need to mitigate the potential harms and risks brought about by AI systems. Consequently, effective AI governance is essential for sustainable AI development and use.

Sustainable AI refers to creating and operating trustworthy AI for the benefit of society, the environment, and businesses. To contribute, AI operators must address the sustainability concerns of their AI systems. However, the fact that most organizations do not have the capabilities to tackle these issues alone, makes it a large-scale challenge requiring cooperation between multiple actors within the AI governance ecosystem. Private companies are expected to play a role in achieving AI sustainability, along with public organizations and international entities. Nevertheless, the role of private consultancies and the nature of their services have not been studied.

The objective of this thesis is to study the interorganizational activities and services provided by private consultancies within the AI governance ecosystem and to investigate business models supporting such service offerings. After reviewing literature on sustainable AI and how to achieve it, the empirical part was conducted as case study action research, using business model design and the Business Model Canvas as tools. The business model was developed for Solita, an IT and design consultancy, willing to expand its involvement in the sustainable AI field. Solita had conducted interviews with 26 organizations across 11 industries shortly before this study, through which the state of sustainable AI had been discovered. In this study, insights from the client interview analysis were used as the primary source to inform the client perspective of the business model, while multiple interviews, co-design, and feedback sessions were conducted with Solita employees to uncover the company point of view.

As a result, a framework was developed mapping out the sustainable AI consulting services provided by private consultancies into three categories: AI system sustainability assessment, organizational sustainable AI capabilities, and technical enablers for sustainable AI. Moreover, it was discovered that the value expected by clients stems from enhancing AI risk management and improving AI operational excellence, which in turn allows for proving a positive impact resulting in competitive advantage. Eventually, working towards the sustainability of AI is expected to result in financial gains for clients and is important in ensuring social responsibility. However, the value-driven business model supporting such an offering did not differ significantly from the existing business model of Solita.

This study contributes to the academic discussion in the emerging field of sustainable AI and AI governance by empirically validating and extending prior literature. Additionally, this study has practical implications for the case company and other actors in the AI governance ecosystem, as it clarifies the role of private consultancies within the developing AI governance ecosystem.

Keywords: Sustainable AI, AI governance ecosystem, business model design, value creation, value proposition, Business Model Canvas

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

TIIVISTELMÄ

Lauri Eneh: Liiketoimintamallit kestävä tekoäly -konsultoinnissa – Organisaatioiden välisten kestävä tekoäly -palveluiden konseptointi liiketoimintamallisuunnittelun avulla
Diplomityö
Tampereen yliopisto
Tuotantotalouden diplomi-insinöörin tutkinto-ohjelma
Tarkastajat: prof. Teemu Laine and Tenure track -prof. Tuomas Korhonen
Lokakuu 2022

Tekoälyjärjestelmien (AI system) käyttö on yleistymässä nopeasti monilla aloilla. Useiden hyötyjen lisäksi tekoälyjärjestelmien potentiaaliset haittavaikutukset ovat nousseet esille. Tutkijoiden ja tieteenharjoittajien keskuudessa kasvaa yhteisymmärrys siitä, että tekoälyjärjestelmien aiheuttamia haittoja ja riskejä on pyrittävä pienentämään. Tehokas tekoälyn johtaminen (AI governance) on olennaista kestävän tekoälyn kehittämisen ja käytön kannalta.

Kestävä tekoäly viittaa luotettavien ja yhteisö, ympäristöä sekä liiketoimintaa hyödyttävien tekoälyjärjestelmien kehittämiseen ja käyttöön. Osallistuakseen tavoitteen saavuttamiseen, tekoälyn käyttäjien on huomioitava käyttämiensä tekoälyjärjestelmien kestävyyshaasteet. Monilla organisaatioilla ei kuitenkaan ole kykyä ottaa kantaan näihin ongelmiin ilman ulkopuolista apua. Tästä syystä ongelma muodostaa suuren haasteen, joka vaatii yhteistyötä useiden tekoälyn johtamisen ekosysteemiin kuuluvien toimijoiden välillä. Sekä yksityisiä yrityksiä että julkisia ja kansainvälisiä organisaatioita tarvitaan tekoälyn kestävyyden saavuttamiseksi. Yksityisten konsultointiyritysten roolia sekä niiden tuottamien palveluiden luonnetta ei ole kuitenkaan vielä tutkittu.

Työn tavoitteena on tutkia yksityisten konsultointiyritysten tarjoamia organisaatioiden välisiä palveluita tekoälyn johtamisen ekosysteemissä, sekä kartoittaa näiden palveluiden mahdollistavan liiketoimintamallin piirteitä. Työn kirjallisuuskatsauksessa selvitettiin ensin mitä kestävä tekoäly tarkoittaa ja miten se voidaan saavuttaa. Tämän jälkeen tapaus- ja toimintatutkimuksena toteutettu empiirinen osa suoritettiin käyttäen apuna liiketoimintamallimuotoilua sekä Business Model Canvasia. Liiketoimintamalli kehitettiin IT- ja muotoilukonsultointiyritys Solitalle, joka on halukas kasvattamaan rooliaan tekoälyn kestävyyden mahdollistamisessa. Hieman ennen tätä työtä Solita selvitti kestävän tekoälyn tilanteen haastattelemalla 26 organisaatiota 11 eri toimialalla. Näiden haastattelujen löydöksiä käytettiin pääasiallisena aineistona asiakkaiden tarpeiden ja näkemysten selvittämisessä. Solitan näkemyksen selvittämiseksi järjestettiin useita haastatteluja ja yhteissuunnittelutilaisuuksia.

Tutkimuksen tuloksena kehitettiin viitekehys, jonka avulla yksityisten konsultointiyritysten tarjooma voidaan jakaa kolmeen kategoriaan, jotka ovat tekoälyjärjestelmien kestävyysarviointi (AI system sustainability assessment), organisaation kestävä tekoäly -kyvykkyydet (organizational sustainable AI capabilities) sekä kestävän tekoälyn tekniset mahdollistajat (technical enablers for sustainable AI). Näiden palveluiden tuottaman arvon tunnistettiin kumpuavan tekoälyn riskienhallinnan parantumisesta sekä tekoälyn operatiivisen huippuosaamisen (operational excellence) kehittymisestä, jotka puolestaan mahdollistavan kilpailuedun saavuttamisen, kun asiakkaan käyttämän tekoälyn positiiviset vaikutukset voidaan todentaa. Tekoälyn kestävyyden tavoitteleminen on tärkeä osa organisaatioiden sosiaalista vastuunkantoa, ja sen uskotaan lopulta johtavat rahallisiin hyötyihin. Kehitetty kestävä tekoäly -konsultoinnin mahdollistava liiketoimintamalli ei kuitenkaan eronnut merkittävästi Solitan nykyisestä liiketoimintamallista.

Tämä tutkimus ottaa osaa akateemiseen keskusteluun tekoälyn ja tekoälyn johtamisen alalla. Se todentaa empiirisesti sekä laajentaa aikaisempaa tutkimusta. Tällä tutkimuksella on lisäksi käytännöllistä hyötyä Solitalle sekä muille tekoälyn johtamisen ekosysteemin toimijoille, sillä se selkeyttää yksityisten konsultointiyritysten roolia orastavassa tekoälyn johtamisen ekosysteemissä.

Avainsanat: Kestävä tekoäly, tekoälyn johtamisen ekosysteemi, liiketoimintamalli, arvonluonti

PREFACE

Some five years ago, I began my university journey unsure of what I wanted to study. After a few changes of direction, I eventually found my cup of tea. This thesis allowed me to bring together my interest in information technology and business management in an exciting and worthwhile way.

First of all, I am grateful to Anna Metsäranta for sharing her deep expertise and providing guidance during the thesis process. Moreover, I would like to thank Anna and Manu Setälä in particular, and Solita Oy in general, for giving me the opportunity to study such an exciting and timely subject.

From the university, I would like to thank Tuomas Korhonen for his advice in setting the direction and eventually refining the thesis, and Teemu Laine for his valuable insights during the thesis project.

Finally, I would like to to express my deepest gratitude to my friends and family, especially Ronja, for supporting and encouraging me throughout these past five years. Balancing life between being a student, a competing athlete, and entering work-life has not always been easy, but you have enabled me to pursue my goals in all aspects of life.

Tampere, 17 October 2022

Lauri Eneh

CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS AND NOTATIONS

AI	Artificial Intelligence
AI Act	Artificial Intelligence Act proposal put forth by the EC
AIGA	AI Governance and Auditing – research project consortium
AI HLEG	High-Level Expert Group on Artificial Intelligence
BMC	Business Model Canvas
BMI	Business model innovation
CEN	European Committee for Standardization
CENELEC	European Committee for Electrotechnical Standardization
CO2e	Carbon dioxide equivalent
EC	European Commission
EU	European Union
IEC	International Electrotechnical Commissions
IEEE	Institute of Electrical and Electronics Engineering
ISO	International Organization for Standardization
IT	Information Technology
JTC	Joint technical committee
ML	Machine Learning
SDG	Sustainable Development Goals
SDO	Standard Developing Organization
VPC	Value Proposition Canvas

1. INTRODUCTION

The utilization of Artificial Intelligence (AI) is growing rapidly in various domains (European Commission, 2020). Private and public sector actors in industries such as healthcare, financial services, manufacturing, and transportation are adopting AI systems to advance their organizational ambitions by streamlining processes and enhancing decision making quality and speed (see Sun & Medagli, 2019; Tomašev, et al., 2020; Lv, et al., 2021; Mäntymäki, et al. 2022b; Viljanen & Parviainen, 2022). In short, AI is defined as “a system’s ability to interpret external data correctly, to learn from such data, and to use those learnings to achieve specific goals and tasks through flexible adaptation” (Kaplan & Haenlein, 2019).

However, there is an increasing consensus among practitioners and researchers on the need to mitigate the potential harms and risks brought about by AI systems (Dignum, 2020). Real-life examples of biased recruitment models, discriminating credit score systems, and other violations of human rights, has led to the discussion on the ethics of AI (see Jobin et al., 2019; Martin, 2019; Asatiani, et al., 2020; Kelly-Lyth, 2021; Fumagalli, et al., 2022; Genovesi & Mönig, 2022; Malik, et al., 2022). Along with the ethical concerns affecting humans, the environmental impact of AI model development is worrying. In addition to their use case dependent impact, the training and tuning process of Machine Learning (ML) models as well as their use in inference require large amounts of computing capacity, resulting in significant energy consumption (Wu, et al., 2022).

If left unaddressed, the challenges threaten to impede the desired advancement of the beneficial applications of AI (Morley, et al., 2020). To establish trust and mitigate risk, carefully monitoring the design, development, and use of AI systems, and assessments of their ethical, legal, and social implications is necessary. Assuring wide trust in AI system development and operation will promote AI’s advancements. (Floridi, et al., 2022) AI systems need to be developed and used sustainably, both from social and environmental perspective, as well from an economic point-of-view. Other than for “doing the right thing”, drivers for tackling these issues can stem from law, stakeholder pressure, or business motivation. (Mäntymäki, et al., 2022b)

Effective governance of AI is essential to ensure the ethical matters are addressed properly when designing and using AI systems (Butcher & Beridze, 2019). Establishing legislative measures and forming official regulatory bodies having the authority to govern AI systems is a vital level for AI governance (Butcher & Beridze, 2019). Indeed, the European Commission (EC) has recognized the need for regulation and has put forth the Artificial Intelligence Act (AI Act) in April 2021, which aims to harmonize the rules to steer organizations towards designing and operating trustworthy AI systems. Indeed, the proposal sets various requirements for organizations developing and using AI systems, depending on the perceived risks of the AI systems (AI Act, 2021). However, responsibility of AI governance lies not only with regulators, but the organizations are responsible as well (Chhillar & Aguilera, 2022). Mäntymäki et al. (2022a) defines organizational AI governance as “a system of rules, practices, processes, and technological tools that are employed to ensure an organization’s use of AI technologies aligns with the organization’s strategies, objectives, and values; fulfills legal requirements; and meets principles of ethical AI followed by the organization”.

However, most organizations do not have the capabilities to tackle the sustainability issues of AI alone (Minkkinen, et al., 2022b). Hence, the promotion of the responsible development of AI is a large-scale challenge requiring cooperation among multiple actors (Seppälä, et al., 2021; Minkkinen, et al., 2022b). Consequently, the European Union (EU) has articulated an ecosystem approach to sustainable AI and positions itself as a key actor in this emerging ecosystem (European Commission, 2020). Both private and public organizations are to play a role in the AI governance ecosystem (Chhillar & Aguilera, 2022), and organizations must understand the elements of AI governance and recognize their own part in the multi-actor ecosystem (Minkkinen, et al., 2022b). Although the roles of various players in the ecosystem have been studies, the academic understandings about the position of consultancies and the services they provide within the AI governance ecosystem is insufficient (Seppälä, et al., 2021; Minkkinen, et al., 2022b; Mäntymäki, et al., 2022b).

1.1 Research objectives, questions, and scope

This study aims to fill in the gap recognized in prior literature by broadening the knowledge around the state of sustainable AI and the AI governance ecosystem, by studying the interorganizational activities within the AI governance ecosystem provided by a private consultancy, and by investigating business models around sustainable AI service offerings. Additionally, the goal of this thesis is to, through business model innovation, conceptualize a business model for Solita, an IT consultancy willing to offer services which enable sustainable AI development and usage for its clients. Indeed, Osterwalder and Pigneur (2010) argue that bringing new services or products to market is a motivation for business model innovation. The Business Model Canvas (BMC) (Osterwalder & Pigneur, 2010), is used as a holistic structuring tool for the business model innovation. Moreover, the BMC has become the de facto standard framework for business model development (França, et al., 2017), and is also relevant in high level academic research (see ur Rehmana, et al., 2016; Metallo, et al., 2018; Sort & Nielsen, 2018; Polydoropoulou, et al., 2020). By using business model innovation as a means, the study aims to answer the following research questions:

1. *What interorganizational activities are needed from consultancies within the AI governance ecosystem, and what value are the activities expected to create?*
2. *What characteristics does a business model supporting the delivering of sustainable AI offering have?*

By designing a business model for the case company Solita, the research questions can be answered. From an academic perspective, insight to the AI governance ecosystem and interorganizational activities within it will be provided. Also, insight to sustainable AI services business models will be gained. Furthermore, possible consistencies and conflicts between academic knowledge and practical experience can be highlighted.

1.2 Background

The company for which the business model is being designed is Solita Oy. Solita is a technology, data, and design company providing several types of consulting and IT services for its clients. Solita creates “impact that lasts by combining tech, data and human insight,” operates in the six countries in northern Europe, and employs around 1500 experts. Furthermore, Solita aims to “create digital services that positively impact business, people, society, and the environment”. Lately, Solita has been active in promoting ethical and sustainable AI and is willing to grow its business in the space. (Solita, 2022) The research was conducted in close collaboration and in an employee relationship with Solita.

This study is conducted as a part of Artificial Intelligence Governance and Auditing (AIGA), a research project aiming to explore putting responsible AI into practice. Particularly, theme four of AIGA (Commercializing AI Transparency and Explainability) aims to “accelerate the market growth around responsible AI”, and this study contributes to this goal. AIGA is a cross-disciplinary innovation project on AI governance and auditing including universities and multiple businesses in Finland, including Solita. The project is coordinated by the University of Turku and partially funded by Business Finland. (AIGA, 2022)

1.3 Research design

The research design describes the general plan of how the research questions are intended to be answered (Wohlin & Aurum, 2015). Moreover, decisions about research philosophy, methodological choices, and data collection, analysis techniques and procedures should be laid out explicitly to bring transparency and authenticity (Saunders, et al., 2019). As per Eriksson and Kovalainen (2008), the aim of the research and the research questions should guide these decisions.

For this study, pragmatism is the philosophical approach, as the practical consequences supporting action determine the relevancy of the research (Goldkuhl, 2012). Furthermore, a research philosophy reveals the assumptions and beliefs about the development of knowledge. Pragmatism, as a value-lead philosophical approach, approaches research as primarily a means to solve a problem and produce a practical solution. Prior beliefs and values are accepted, but they should be stated in public together with the research goals (Friedrichs & Kratochwil, 2009; Morgan, 2014).

To achieve the research objectives, theory creation is approached in an abductive manner. Rather than strictly imposing a conceptual theoretical template (deduction) or straightforwardly inferring propositions from observations (induction), data is collected to explore a phenomenon, and to modify existing theories or to create new ones. To some extent, abduction tries to mimic human practice. Moreover, abduction is a valid reasoning approach for pragmatic research. (Friedrichs & Kratochwil, 2009) In this study, prior knowledge is used to build a theoretical understanding of questions at hand, and the empirical research is used to confirm and evaluate hypotheses, and to gather further knowledge on the subject. Here, the theory of sustainable AI, its benefits, and methods of achieving it is utilized to build the business model for a specific company. Furthermore, theories of designing business models guide the business model innovation process.

To build the theoretical foundation for this study, a literature review is conducted around sustainable AI. In the case of business models, extensive literature reviews were found, omitting the need of conducting one. The process is explained in detail in the beginning of chapter 2.

A multi-method qualitative approach is selected as the methodological choice, as this research aims to explanatorily study interorganizational activities and the business model of a company (Saunders, et al., 2019, p. 179). Moreover, according to Saunders et al. (2019, p. 190), multiple research strategies can be employed simultaneously. Since gaining a deep understanding of a complex interorganizational relationship is of interest, this thesis is performed as a case study. Case studies often use multiple sources of evidence, such as archival research, observation, and interviews. (Yin, 2009, p. 18) Indeed, interviews are the main data source used in this study, as interviews are a practical and effective way to gather unpublished information (Eriksson & Kovalainen, 2008). Within the case study, an action research strategy is adopted, which is an emergent and iterative process, aiming to develop solutions to real problems organizations have, through a collaborative and participatory approach (Oliva, 2019). Indeed, action research case study has been proven to be an effective research strategy (see McManners, 2015).

After the literature review, the empirical part of the research was carried out. It includes analyzing client interview data, observing client meetings, and facilitating business model design sessions. Furthermore, secondary data sources were used to gain further insight into Solita's view. The empirical research strategy, description of the data collection methods as well as the analysis is described later. Also, the empirical research process is laid out in chapter 4.

1.4 Structure of the thesis

Based on the literature review, an understanding of sustainable AI and business models is formed. After this theoretical exercise, data is gathered and analyzed, and design sessions held to formulate a business model for Solita. Figure 1 presents the research plan and used methods.

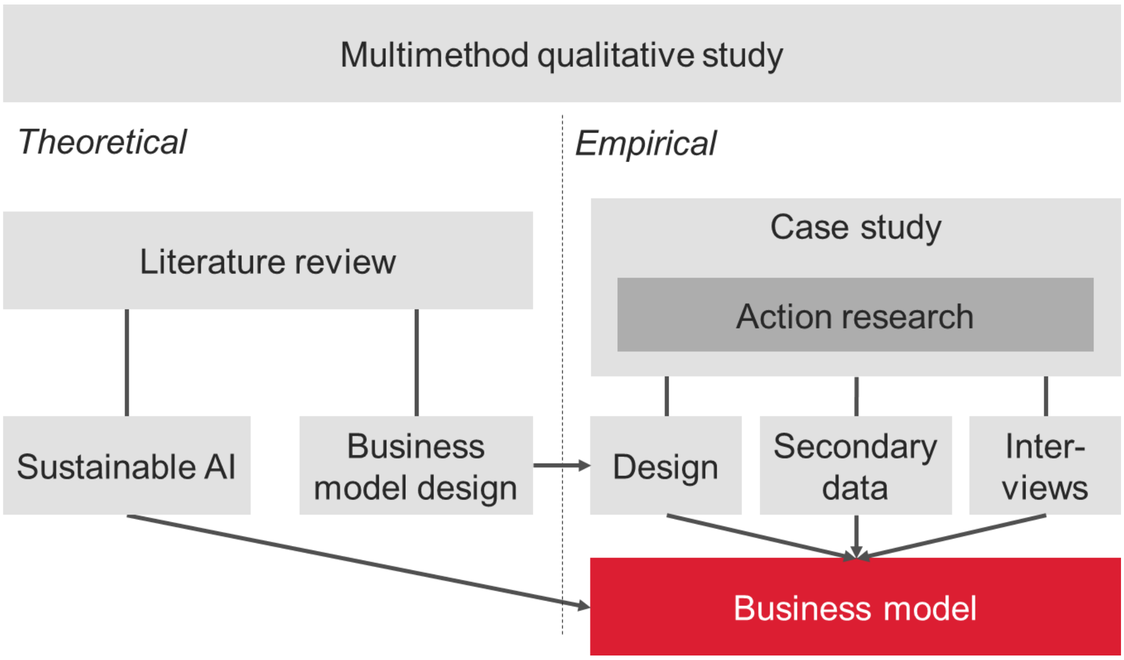


Figure 1. *Diagram of research plan and methods.*

The remainder of this thesis is structured as follows: Chapter 2 covers sustainable AI, methods for achieving it, and the ecosystem around it. Next, business model innovation will be covered in Chapter 3. As mentioned above, the empirical research methods and methodology are brought forth in chapter 4. Results of the empirical study are presented in chapter 5, after which the key findings, implications, limitations, and future research suggestions are discussed in chapter 6. Finally, chapter 7 summarizes and concludes the thesis.

2. ACHIEVING SUSTAINABLE AI

AI systems themselves need to be used sustainably. There is an increasing consensus among practitioners and researchers on the need to mitigate the potential harms and risks brought about by AI systems (Dignum, 2020). If left unaddressed, the social, environmental, and economic challenges threaten to hinder the advancement of the beneficial applications of AI (Morley, et al., 2020). To establish trust and mitigate risk, carefully monitoring the design, development, and use of AI systems, and assessments of their ethical, legal, and social implications is necessary. Assuring wide trust in AI system development and operation will further AI's advancements. (Floridi, et al., 2022)

In this chapter, the sustainability of AI systems is explored through a semi-systematic literature review, the process of which is explained first. Then, sustainable AI and its components are described. Next, methods for ensuring the sustainability of AI systems, namely AI governance and auditing, are presented. Finally, the ecosystems and its actors are elaborated on.

2.1 Literature review process

To build the theoretical foundation for this study, a literature review is conducted around sustainable AI and related terms. According to Snyder (2019), considering relevant prior literature is essential for all research, and allows for justifying hypotheses. However, it is vital to match the literature review method with the goal of the review and the field of study (Snyder, 2019). The goal for the review in this study is to gain knowledge of the current theories and frameworks presented in literature, and to provide an overview of the topics at hand. Semi-systematic review, which is used in this thesis, is a valid method for business research, as well as in circumstances where reviewing all potentially relevant publications is simply impossible (Snyder, 2019).

Reporting of the literature review process transparently is important, since it allows the reader to evaluate the strengths and weaknesses of the review (Liberati, et al., 2009). Adapting the process presented by Fink (2019), a seven-step process is selected for conducting the literature review. The steps included are selecting the goal for the review; choosing databases and sources for the review; choosing search terms to be used when searching from databases; practical screening of the results, for example excluding specific publication types or excluding outdated publications; scientific screening, as in evaluating the quality of the material; conducting the literature review; and finally reporting the results of the review.

Gathering knowledge and theories around sustainable AI and business model design is the goal of the literature reviews, as mentioned earlier. As initial searches on business model design for sustainable AI offering did not yield any relevant results, the two terms are handled separately. First, literature on sustainable AI and its related topics was reviewed. This was followed by studying business model development through prior literacy reviews.

For source databases, Web of Science and Scopus are selected. Also, scientific papers published in the AIGA project are considered, which have been listed on the AIGA website (ai-governance.eu). Furthermore, by using the snowballing technique (Wohlin, et al., 2020), articles referenced in suitable publications were also included in the review. Snowballing was applied using Google Scholar, in addition to the two mentioned databases.

The search terms "*sustainable AI*", *"(AI or ML) AND audit* AND (framework or process)*", and "*(AI or ML) AND governance AND framework*" were used. All searches were conducted in May and June 2022. The search results were limited by language (English language only), and by subject area (Computer Science, Engineering, Social Sciences, Business, Economics, Management and Accounting, Environmental studies).

The results of the searches were ordered by the databases using the "relevance" criteria, which in Web of Science arranges results based on the number of terms matching in the title, abstract and keywords (Clarivate, 2021), and in Scopus also considers the absolute and relative position and uniqueness of the word (Elsevier, 2021). Since the searches resulted in hundreds of hits, only the top 120 articles were considered. Titles and keywords were screened initially to determine if the publication addresses the concept of interest. The abstracts of potential publications were then examined to evaluate the relevance of the study. Finally, the literature review was conducted on the selected material. The results are presented in the rest of this chapter.

2.2 Defining AI, sustainability, and sustainable AI

Artificial intelligence (AI) refers to "a system's ability to interpret external data correctly, to learn from such data, and to use those learnings to achieve specific goals and tasks through flexible adaptation" (Kaplan & Haenlein, 2019). The European Commission's (EC) High-Level Expert Group on AI (AI HLEG) (2019) takes a more inclusive view in defining AI systems: “Artificial intelligence (AI) systems are software (and possibly also hardware) systems designed by humans that, given a complex goal, act in the physical or digital dimension by perceiving their environment through data acquisition, interpreting the collected structured or unstructured data, reasoning on the knowledge, or processing the information, derived from this data and deciding the best action(s) to take to achieve the given goal. AI systems can either use symbolic rules or learn a numeric model, and they can also adapt their behaviour by analysing how the environment is affected by their previous actions. As a scientific discipline, AI includes several approaches and techniques, such as machine learning (of which deep learning and reinforcement learning are specific examples), machine reasoning (which includes planning, scheduling, knowledge representation and reasoning, search, and optimization), and robotics (which includes control, perception, sensors, and actuators, as well as the integration of all other techniques into cyber-physical systems).” Furthermore, the EC (2021) defines AI systems to include machine learning (ML), logic- and knowledge-based, as well as statistical approaches.

As can be seen, AI systems encompass decision models of varying kinds. However, most AI systems share common characteristics and have similar lifecycles. According to the OECD AI expert group (2019), the lifecycle of an AI system typically includes four phases: (1) “design, data and models”; (2) verification and validation; (3) deployment; and (4) operation and monitoring. The first phase is context-dependent and comprises planning and designing the system to fill a business need, gathering of data and prepossessing data to make it usable, and selecting, creating and calibrating models or algorithms. Next, in the verification and validations phase, the performance of the models is assessed across various aspects and considerations. Deployment involves ensuring compatibility with other IT systems, overseeing organizational change and assessing user experience. Finally, once deployed, operation and monitoring must be performed, to continuously monitor the recommendations and impact of the AI system. Problems identified in these phases might have implications to the other phases, or even in retiring the entire system. (OECD, 2019) The AI system lifecycle is depicted in figure 2.

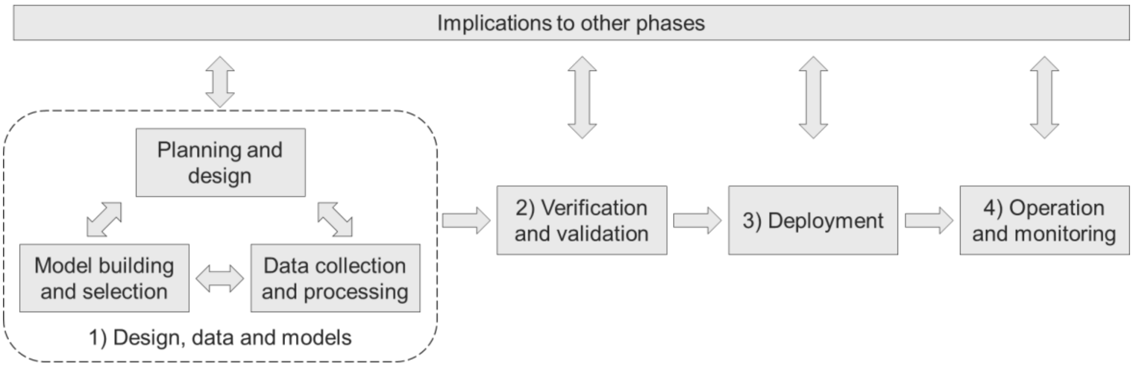


Figure 2. *AI system lifecycle (adapted from OECD, 2019).*

AI systems have been conceptualized as socio-technical systems consisting of an information technology (IT) artifact surrounded by people, institutions, and organizations (Dignum, 2019). Indeed, van Wynsberghe (2021) agrees and claims that sustainable AI refers not only to the AI applications, but rather addresses the entire socio-technical system being impacted by and surrounding the AI system.

Sustainability, defined by Oxford Languages and Google, is “the ability to be maintained at a certain rate or level” and “avoiding the depletion of natural resources to maintain ecological balance” (Oxford Languages, 2022). Although researchers and companies have given the term a superior emphasis on the influence on the physical environment (Vallance, et al., 2011; Bjørlo, et al., 2021), organizations need to consider the impact of their decisions on the social and economic atmosphere (Pfeffer, 2010; Di Vaio, et al., 2020), and take intergenerational justice into account (Halsband, 2022). Indeed, the three-pillar conception of sustainability into its social, environmental, and economic aspects has become ubiquitous (Purvis, et al., 2019).

In line with this view, the 17 sustainable development goals (SDG’s) put forth by the United Nations cover social aspects (e.g., #5 gender equality, #16 justice), environmental aspects (e.g., #6 clean water, #12 responsible consumption and production, and #13 climate action) and economic aspects (e.g., #8 decent work and economic growth and #1 no poverty) of sustainability (United Nations, 2022). In agreement, a systematic literature review by Mensah (2019) found that the issue of sustainability covers environmental, economic, and societal concerns. Moreover, a literature review on the term “sustainable AI” in public sector decision making done by Wilson and der Velden (2022) found that while some papers focus on just either the social or environmental aspects of sustainability, many of the review papers take a holistic approach and include the social, environmental, and economic aspects. Most papers in the review referred to the United Nations SDGs. Furthermore, according to Sætra (2021), the SDG’s serve as a relevant and practical framework for evaluating and categorizing the benefits and harms of AI.

Hence, this study covers all three aspects of sustainability: society, environment, and economy. These aspects are highly intertwined and are not mutually exclusive.

It should be noted that there is a difference between AI *for* Sustainability and sustainability *of* AI. The former views AI as a means for reaching environmental and social goals, such as using AI for reducing emissions by optimizing route planning or energy efficiency. On the other hand, the latter sees AI itself as an object needing to be developed and used sustainably, since using AI systems creates environmental, social, and economic impacts (van Wynsberghe, 2021; Kindylidi & Cabral, 2021). Furthermore, in public sector decision-making literature, the term “sustainable AI” is less frequently used to refer to AI *for* sustainability, and more often comprises aspects of the sustainability *of* AI (Wilson & der Velden, 2022).

This study focuses mainly on the latter lens of sustainable AI, where the social aspects, such as fairness, accountability, and transparency of AI; the environmental impact of training and using AI systems; and the economic impact of placing AI systems on the market is considered. It takes the socio-technical approach to AI and addresses the impacts of operating AI systems. Consequently, sustainable AI in this study refers to *creating and operating trustworthy AI for the benefit of society, the environment, and businesses*.

2.2.1 Dimensions of sustainable AI

Social sustainability of AI

According to Genovesi and Mönig (2022), social sustainability refers to achieving a fair degree of social homogeneity, proper livelihoods, equitable access to resources and social services, as well as promoting equal opportunities, non-discrimination, and autonomy of all citizens. Furthermore, Vallance (2011) proposes social sustainability to include addressing basic needs and justice, as well as the preservation of socio-cultural characteristics of societies.

AI systems can prevent us from reaching those goals by inflicting social harm by systemizing inequality and bias, accelerating harm propagation, and blurring the awareness of harm (Malik, et al., 2022). Interestingly, Vakkuri et al. (2022) found that AI system developers consider the possible harm produced by the system in physical terms, and often ignore systemic effects.

To be socially sustainable, the AI systems must adhere to ethical principles (Seppälä, et al., 2021). In order to be sustainable, AI systems must also be ethical (Kindylidi & Cabral, 2021). A systematic literature review by Jobin et al. (2019) focusing on the ethical aspects of AI, found that out of the 84 documents reviewed, over seven in ten included the principles of transparency, justice and fairness, non-maleficence, and responsibility. Other studies of ethical AI principles have found similar results (Seppälä, et al., 2021). Indeed, similar principles have been stated by the AI HLEG (2019) in defining AI trustworthiness through three closely intertwined components. The systems must be lawful, ethical, and technically robust to be trustworthy. For this study, fairness, accountability, transparency, and nonmaleficence are the most vital principles and themes included in the social aspect of sustainability.

The principle of fairness is strongly linked to prevention of bias and non-discrimination (Fjeld, et al., 2020). Algorithm bias, a common mode of ethical AI failure, refers to prediction systems systematically putting one group of people in a disadvantageous position or even excluding them all together, based on a trait such as skin color, gender, socioeconomic background, or religious beliefs (Floridi, et al., 2022). Even though such failure is referred to as algorithmic social bias (Kordzadeh & Ghasemaghaei, 2021), the algorithm is not the fault, and algorithms cannot be held accountable for their actions (Ryan, 2020). They are inherently agnostic and free holding value. The root causes stem from the algorithm's developer and the used data. Moreover, ML models tune their parameters according to the fed data, thus reflecting and reinforcing the biases present in the real world. (Floridi, et al., 2022)

Ryan (2020) argues that responsibility of AI systems must lie with those who design, develop, deploy, and use the systems, since agnostic systems cannot be held accountable for their recommendations or actions. Indeed, human control and determining accountability are closely linked to responsibility (Fjeld, et al., 2020). However, Vakkuri et al. (2022) found that rather than from an ethical point of view, AI system developers often approached responsibility from a financial, legislative, or customer relations point-of-view pragmatically.

Transparency, as stated by Ryan and Stahl (2020), can be understood as either the transparency of the AI system itself, or as the transparency of the organization designing, developing, and using it. The former refers to explainable AI, i.e., the understanding of the details and reasons a model reaches its decisions (Barredo Arrieta, et al., 2020), while the latter entails awareness of what, by whom, and why were certain decisions made during the design and development process (SIIA, 2017). Opacity in AI systems raises concerns about their quality and their trustworthiness. The question arises when humans are affected negatively by the decisions made by these algorithms. (Floridi, et al., 2022) Moreover, according to Barredo Arrieta (2020), explainable AI tools could help achieve other principles, such as fairness, accountability and safety.

Communication is a key element in social sustainability and ensuring trustworthiness. In their literature review, Laato et al. (2022a) found that there are five high-level goals in AI system communication for end users: understandability, i.e., the end users' ability to form a mental model of the AI systems functioning; trustworthiness, i.e., the end users' perception of the honesty and truthfulness of the AI system; transparency, i.e., the amount of information disclosed about the system; controllability, i.e., the subjective sense of human control over the AI system; and fairness, i.e., the users' perceptions the AI systems recommendations being fair and just.

Environmental sustainability of AI

To be environmentally sustainable, the development and usage of AI systems should recognize the boundaries of Earth's resources and work to prevent undesirable environmental change (Genovesi & Mönig, 2022).

According to Robbins and van Wynsberghe (2022), major processes causing emissions must be factored in, including the cost of the hardware used to run the algorithms, the cost of collecting and transmitting data used and processed by AI systems, computational cost of training and tuning models, the disposal of the hardware network needed by AI, and the cost of ensuring the models are ethically aligned.

This study focuses on the environmental impact of operating AI and the handprint of AI use cases, even though the environmental impact of AI traces back to the supply and procurement chain of AI (Robbins & van Wynsberghe, 2022). Operational emissions include the energy cost and data center overhead but does not take the manufacturing of the machines the models run on into account (Patterson, et al., 2022).

Interestingly, according to Strubell et al., (2019), a single deep learning neural network model's training could lead to around 270 tons of carbon dioxide equivalent (CO_2e) emissions, an equivalent of the emissions produced by 5 cars in their entire lifetime. However, Petterson and colleagues (2022) argue these estimates to be faulty, and by following best practices, operational emissions of ML models could be significantly lower. These best practices include ML model selection, processor choice, and utilizing cloud data centers and selecting locations with the cleanest energy. Then again, according to Genovesi and Mönig (2022), using cloud compute capacity could conceal the fact that physical computers in remote data centers are performing the computational operations and might temp businesses not to address these emissions themselves. Nevertheless, computing in large cloud data centers improves their efficiency (Patterson, et al., 2022). Also, at least the leading cloud providers, Amazon Web Services, Microsoft Azure, and Google Cloud Platform, offset their carbon emissions (Lacoste, et al., 2019), i.e., invest money in removing carbon from the atmosphere or funding renewable energy, energy efficiency or other similar projects, which are believed to eventually reduce carbon emissions (Lovell & Liverman, 2010).

Nonetheless, Strubell et al. (2019) argue that to make the environmental impacts of training ML models transparent, the time used in training should be reported openly. Henderson et al. (2020) point out that energy usage and carbon emission reporting is essential, and that most metrics used in reporting (runtime, compute) can be translated into energy and carbon metrics.

Van Wynsberghe (2021) suggests a proportionality framework to be used to assess whether training or tuning an AI system for a specific task is justified in regard to the carbon footprint, and the general environmental impact. She suggests that not only the environmental impact of computation needed for the solution should be considered, but also the predicted impact of the AI system in its use case must be weighed in.

Economic sustainability of AI

Economic sustainability refers to “practices that support long-term economic growth without negatively impacting social, environmental and cultural aspects of the community” (University of Mary Washington, 2022). According to Lee (2021), academic papers on AI systems often take economic sustainability as a default requirement and focus on the social and environmental aspects.

AI systems themselves are bound to result in economic gains for the companies using them (Sun & Medaglia, 2019; Lv, et al., 2021; Mäntymäki, et al., 2022b). However, as seen in previous chapters, reducing the negative impact on the other sustainability aspects could become costly. Investments mitigating social and environmental harm should not have too negative of an impact on the economic performance. Afterall, public companies must remain profitable to continue operations.

According to Floridi and colleagues (2018), investments in ethical AI yield a dual advantage, in increasing AI adoption by making AI more socially acceptable, along with lessening potential for negative impact by allowing organizations to mitigate or avoid costly mistakes. Ethical AI allows organizations to avoid underuse as well as misuse of AI. However, an environment of public trust and clearly defined responsibilities are required to achieve this dual advantage of ethical AI. (Floridi, et al., 2018)

Initiatives for social and environmental sustainability influences the economic performance of an organization (Gimenez, et al., 2012). Indeed, investments in sustainability itself can lead to economic gains (Willard, 2012). Increasingly, consumers take the environmental impact of their purchasing decisions into account, and businesses are increasingly using green or environmental claims or even greenwashing in marketing (Kindylidi & Cabral, 2021). Moreover, as mentioned earlier, Vakkuri et al. (2022) found financial drivers to be typical for developers of AI systems when addressing responsibility. Furthermore, Seppälä (2021) found similar results with business benefits being a key-driver for ethical AI along with regulatory and stakeholder pressure, risk management, and corporate social responsibility.

The growing adoption of AI systems can have an impact on the entire economy through inequality of resource allocation and effects on the job market. Innovators and redistributors are in positions to reap all profits leaving them in a resource surplus, while forcing others to live in shortage (Korinek & Stiglitz, 2017). However, Dauvergne (2022) argues that in corporate sustainability rhetoric, the benefits of AI are exaggerated, while its costs are obscured. Efficiency and production gains in the middle-sections of supply chains result in increased consumption. This phenomenon benefits large corporations far more than does good to social or environmental sustainability. (Dauvergne, 2022)

According to Korinek and Stiglitz (2017) policy can counter the undesirable economic effects. In fact, the Digital Markers Act (2020) proposed by the EC aims at reducing the power and economic possibilities of gatekeepers, by requiring major players to share the data they have been able to acquire by being innovators in AI and other technology (Bräutigam, et al., 2022).

Moreover, concerns of technological unemployment have also been raised with regards to AI systems advancements (Korinek & Stiglitz, 2017). However, as argued by economists, there is no “lump-of-labor”, i.e., there is no fixed number of jobs in the economy and automating some of them does not result in a forever reduced number of positions (Walker, 2007). The increasing adoption of AI is believed to shift job demand from task-related jobs towards high-skilled jobs (Brendel, et al., 2021). On the other hand, technological unemployment could arise through wages not adjusting fast enough to AI advancements or through transitional phenomena, i.e., AI systems making workers redundant faster than new jobs can be created (Korinek & Stiglitz, 2017).

Since this study takes the point-of-view of a private company, the macro-economic aspects will not be considered. The focus will be on analyzing the economic feasibility of investing into sustainable AI. Indeed, the expected value gained from ensuring AI systems are operated sustainably will be addressed.

It should be noted that the separation of the social, environmental, and economic aspects of sustainable AI is somewhat artificial. These aspects, as mentioned before, are highly interrelated, and not mutually exclusive.

2.2.2 Approaches for ensuring the sustainability of AI

As stressed by the AI HLEG (2019), both technical and non-technical methods are needed for implementing social ethical AI principles in practice. Technical tools include validation, testing, and continuous monitoring, while non-technical methods contain governance frameworks, certification, standardization, education and awareness, stakeholder communication, and ensuring diversity of design and development teams of AI systems (AI HLEG, 2019). Correspondingly, an empirical study done by Seppälä and colleagues (2021) found that organizations implement ethical principles in practice through governance, AI design and development, competence and knowledge development, and stakeholder communication. Still, responsibility and accountability as well as roles related to ensuring sustainable AI was found to be often ambiguous.

However, Vakkuri et al. (2022) found in their multiple case-study, that the academic discussion around AI ethics has yet to be ingrained in the industry. There is a consensus on the potential importance of AI ethics, but different views were held on its practical relevance. Only practitioners in highly regulated fields had a more practical view on AI ethics, since the existing quality requirements demand for it. (Vakkuri, et al., 2022)

To address the environmental issue, van Wynsberghe (2021) introduces two technical tools present in literature being available for evaluating the environmental impacts of ML models: The machine learning emissions calculator (Anthony, et al., 2020) and the experiment-impact-tracker framework (Henderson, et al., 2020). Furthermore, a tool by Lacoste, et al. (2019) also provides a web-based tool, which estimates the carbon footprint of training ML models based on the compute and time used. Additionally, methods such as “Sustainability Budget” have been proposed to quantify AI systems’ environmental impact and to bring awareness to the minds of developers, an allow them to aim for less consuming practices (Raper, et al., 2022). Moreover, tools integrated to codebases, such as CodeCarbon (2020), have been developed to help estimate the CO2e emissions used to execute the code. Furthermore, “Care Label” certification of ML systems, labeling energy consumption among other things are introduced as solutions by Morik et al. (2021).

A study conducted by Foffano et al. (2022) on the national AI investments in EU countries revealed that the most significant method governments are using or planning to use to promote the social benefits of AI and its socially responsible development are AI governance mechanisms, with dialogue among stakeholders and with lifelong learning being mentioned as others.

Furthermore, according to Cihon (2019), standards and standardization offer a key mechanism for developing AI governance on a global scale. Furthermore, standardization is to play a key role in providing technical mechanisms for AI system providers to ensure compliance with regulation (AI Act, 2021).

To address the social, environmental, and economic sustainability aspects brought forth by developing AI systems, effective governance of AI is essential (Butcher & Beridze, 2019). Furthermore, AI auditing has been suggested as a tool to operationalize and measure AI governance effectiveness (Koshiyama, et al., 2021; Minkkinen, et al., 2022a), and to steer the attention of developers and consumers toward these audited areas (Genovesi & Mönig, 2022). In the next subsection, these tools for achieving sustainability of AI are explored further.

2.3 AI governance and auditing

Mäntymäki et al (2022a) defines organizational AI governance as “a system of rules, practices, processes, and technological tools that are employed to ensure an organization’s use of AI technologies aligns with the organization’s strategies, objectives, and values; fulfills legal requirements; and meets principles of ethical AI followed by the organization”. Similarly, Butcher and Beridze (2019) characterize AI governance as consisting of tools, solutions and levers that influence AI systems and their development. Likewise, Floridi et al. (2022) define governance mechanisms as the “the set of activities, structures and controls wielded by various parties to exert influence and achieve normative ends”.

In their extensive literature review on AI governance, Chhillar and Aguilera (2022) found that scholars have emphasized the need to govern AI systems and have approached it from three different conceptual angles. The first angle focuses on the hybridization of AI governance, where responsibility lies with private actors as well as the public sector. The second trend within scholar publications stresses the need for regulation of algorithmic decision-making. Finally, some scholars approach AI governance by emphasizing the need for human supervision (i.e., human-in-the-loop), and stress that humans cannot be entirely replaced by algorithms in important decisions. (Chhillar & Aguilera, 2022)

Furthermore, there is a distinction between formal and informal governance mechanisms (Chhillar & Aguilera, 2022). While the former is officially placed, written, and enforced, the latter comprises beliefs and values which guide the behavior of groups and individuals within an organization. Even though not enforced by law, the latter is relevant, since even technically legal decisions made by AI systems might deserve ethical examination (Floridi, et al., 2022).

In the context of AI systems, auditing refers to a structured process through which assessment of a systems or organizations consistency with relevant requirements, ethical principles, and norms is made (Floridi, et al., 2022). Auditing of AI systems has been proposed as a tool to operationalize and determine the effectiveness of AI governance (Koshiyama, et al., 2021; Minkkinen, et al., 2022a). AI auditing supports the overall goal of enabling sustainable AI development and use, by offering tools to verify the claims made about AI systems (Brundage, et al., 2020). Indeed, audits have been discussed as a tangible action for improving AI system transparency (Vakkuri, et al., 2022). Additionally, AI auditing has been characterized as an emerging industry, providing significant economic opportunities (Koshiyama, et al., 2021).

Not all audits focus on the same aspect of the AI systems. Functional audits focus on the decision-making rationale, code audits inspect the source code of, and data fed into an algorithm, legal audits ensure compliance with law and regulations, and impact audits assess the algorithm's output's forms, seriousness, and prevalence of effect (Mittelstadt, 2016). However, how these audits are performed can vary. An extensive literature review by Myllyaho et al. (2021) found that methods used for validating AI systems themselves can be categorized into trails, simulations, model-centered methods, and expert analysis.

Ethics-based auditing is a governance mechanism that organizations designing and developing AI systems can use to validate claims made about their AI systems, operationalise their ethical commitments, and proactively show regulatory compliance (Mökander & Floridi, 2021). Practically, ethics-based auditing is distinguished from laying out a code-of-conduct, by being a structured process, where an organization's behavior is evaluated against relevant norms and principles (Floridi, et al., 2022).

In line with others, Butcher and Beridze (2019) include establishing legislative measures and forming official regulatory bodies having the authority to govern AI systems as a lever for AI governance. Furthermore, according to Minkkinen et al., (2022a) the forthcoming legislation of the EC, Artificial Intelligence Act in particular, is expected to clarify the rules to the AI governance and auditing landscape.

In the following subsections, tools and aspects of AI governance will be examined. First, the EC's AI Act will be covered, followed by an analysis of AI governance frameworks aimed towards organizations. Next, AI auditing tools and frameworks will be covered, and finally relevant standards will be discussed.

2.3.1 The EC's Artificial Intelligence Act

On April 21st, 2021, the EC published a proposal to regulate AI use and development in the EU. The objective of the Artificial Intelligence Act (AI Act) is to guarantee that any AI system affecting EU citizens or developed within the EU is trustworthy, i.e., lawful, ethical, and technically robust. Most of the document's requirements and assessments obligations are set for AI system operators (defined in the AI Act as organizations that use, develop, or distribute AI systems), and address risk identification and mitigation. (AI Act, 2021)

It should be noted that the AI Act's implications and impacts have a reach outside the EU. It applies to AI system providers (defined in the AI Act as developers of an AI system with an intent of placing it in the market) and users (defined in the AI Act as organizations which operate of AI systems in a professional context) located within in the EU, as well as providers and users whose AI systems are used or have an output in the EU market but are located elsewhere. (ibid.)

The AI Act adopts a broad definitions of AI systems, which entails ML approaches (supervised, unsupervised and reinforcement learning using various methods including deep learning), logic- and knowledge-based approaches (knowledge bases, symbolic reasoning, and expert systems) as well as statistical approaches (Bayesian optimization, optimization and search methods) (AI Act, 2021). However, it takes a narrow view regarding the use case of the system, meaning that it is essentially agnostic to the techniques used but focuses on the risk inherent in the practical application (Floridi, et al., 2022).

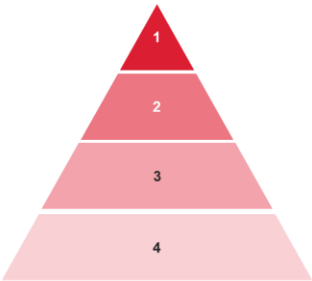
Furthermore, the AI Act is most concerned about the risks of AI systems. It takes a risk-based approach, going as far as to ban the riskiest use cases of AI. On the other end of the spectrum, low-risk systems will be subject to no obligations under the proposed regulation. (AI Act, 2021) A wide range of AI systems fall in between these two extremes with varying obligations.

Indeed, the proposal distinguishes AI systems in four risk-based categories: unacceptable risk, high-risk, low- risk, and minimal-risk. All AI systems posing a clear threat on the civil rights, safety, and livelihood of people are considered to bear a great risk and are prohibited by the AI Act. These unacceptable risk use cases include social scoring, subliminal techniques (stimuli that aim to affect the decisions and thoughts of people without them noticing) resulting in physical or psychological harm, exploiting vulnerabilities (of, for example, children or the elderly), and real-time biometric identification of individuals in public spaces by law enforcement. On the other end of the spectrum, minimal and low risk use cases are applications such as AI in computer games and automatic email spam filters. Some low-risk systems interacting with humans, like impersonation bots or deep fakes, are required to notify the users they are interacting with an AI system but are permitted by the AI Act. Most AI systems in the EU are assumed to belong to the minimal risk category, and the AI Act does not impose requirements on these. (AI Act, 2021)

High-risk systems fall in between these two extremes. These include, but are not limited to, AI systems used in product safety components, critical infrastructure, essential private and public sector services, education, employment, law enforcement, immigration as well as justice and democracy use. (ibid.)

ANNEX IV of the AI Act places the following obligations, among others, on high-risk systems: General description of the AI system including its intended purpose, instructions for use for the users, and the description of all forms the systems is places in the market; detailed AI system element descriptions of the steps and methods used in developing the system and the justification for the choices, validation and testing procedures used, and description of the system architecture; and detailed descriptions of any changes made to the system throughout its lifecycle. (ibid.)

Technology providers of high-risk AI systems will be obligated to demonstrate that AI systems designed or deployed by them adhere to the requirements laid out in the AI Act, before being allowed to deploy their systems in the European market (AI Act, 2021). On top of the pre-deployment conformity assessments, high-risk AI system providers must also establish and document post-deployment monitoring systems, as well as automatic recording of the AI systems operations and decisions, i.e., logging. After providing the relevant conformity assessment, the AI Act requires the system to be registered to an EC managed database to increase transparency. After this, a CE marking is assigned, and the system can be put into service. (Floridi, et al., 2022) Figure 3 shows a summary of the AI Act.



Risk type	Requirements	Examples
1. Unacceptable	Use prohibited	• Subliminal techniques • Exploiting vulnerabilities • Real-time biometric monitoring • Social scoring
2. High	Strict control	• Critical infrastructure (e.g., water, heating) • Essential private and public services • Law enforcement • Education and employment
3. Low	Transparency obligations	• Chat bots • Generated or manipulated content (e.g., deep fakes)
4. Minimal	No action required	• Computer games • Spam filters

Figure 3. *Summary of the Artificial Intelligence Act proposal.*

Importantly, the AI Act does not specify how the assessments are to be conducted, just that they should be. Nevertheless, some instructions are given on when the assessments should be conducted and by whom. The type of conformity assessment depends on the type of high-risk AI system being deployed. (Floridi, et al., 2022)

For example, according to current product safety law, safety component systems, such as medical devices, already need to go through conformity assessments done by a third-party. In such cases when the AI system is a component of a consumer product, the obligations laid out by the AI Act will be integrated into existing domain or sector specific conformity assessment procedures. (Floridi, et al., 2022)

However, not all high-risk systems need to go through such third-party assessments under current regulation. Such systems are called stand-alone AI systems, and include systems used, for example, in employee recruitment or profiling done by law enforcement. (Floridi, et al., 2022) As Mökander et al. (2021) breaks down the AI Act, stand-alone high-risk providers can either demonstrate conformity by conducting an internal ex-ante assessment built on existing internal control or by having their documentation and quality management assessed by an external third-party auditor (i.e., a notified body, see subsection 2.4).

Only stand-alone systems fully compliant with the requirements set out in AI Act Title 3 of Chapter 2 have the first option of showing conformity through internal assessment. These requirements place obligations on an established continuous iterative risk management system, high standards for data quality and data governance, detailed technical documentation and record keeping, transparency and providing information to users, human oversight during use, and accuracy, security, and robustness. (AI Act, 2021) Hence, organizations have the option to either uphold strong internal control and assess conformity based on it, or to involve third-party auditors when willing to deploy AI systems to the EU market.

As can be seen, the current form of the AI Act lacks clear definitions of situations when external audits are required. Indeed, Mökander et al. (2021) points out the to some extent circular reasoning of the proposal. Internal audits are enough if the system complies with the AI Act, but how can compliance with AI Act be verified before external objective audits are conducted? Nevertheless, the AI Act clearly requires external audits in certain situations, for example, in remote biometric identification (AI Act, 2021).

Should these obligations not be met, the AI Act could impose a weighty fine of €30m or 6% of their total world-wide annual turnover of the preceding financial year, whichever is greater (AI Act, 2021).

Shortcomings of the AI Act proposal have been pointed out and suggestions for amendments proposed (see Bräutigam, et al., 2022; Raposo, 2022; Mökander, et al., 2021). However, it should be remembered that the AI Act is still a proposal, and changes to it are expected. Indeed, since its initial publishing, multiple amendments and draft reports have been provided (artificialintelligenceact.eu, 2022). As stated in the webpage following the AI Act's progression (artificialintelligenceact.eu), the proposal will become fully fledged only after the European Council representing EU countries and the European parliament come to an agreement on a version of the text.

Though the AI Act is the most comprehensive set of regulations concerning AI systems (Mökander, et al., 2021; Raposo, 2022), existing regulations and laws guide the use of AI systems (Viljanen & Parviainen, 2022). As Viljanen and Parviainen (2022) demonstrate by conducting two semi-fictional case-studies around the use of AI in recruiting and in a mobile application tracing COVID-19 contacts, that there exist layers of law affecting the use of AI systems to date in Finland. Data rules (e.g., GDPR), general AI rules (e.g., GDPR constrain on automated decision making), application specific non-AI rules (e.g., recruiting), and general non-AI rules (non-discrimination rules or criminal law) must all be taken into consideration when developing and using AI systems (Viljanen & Parviainen, 2022). Furthermore, there are domain specific rules for AI use, for example in the fields of self-driving vehicles, data protection, high-frequency trading, and drones (Ebers, 2021).

As Minkkinen and colleagues (2022a) argue, the AI Act's binding legislation only provides minimum constraints and includes only particular high-risk systems. This leaves a substantial portion of AI systems beyond any requirements.

2.3.2 Organizational AI governance frameworks

Literature on AI governance views it as a layered concept consisting of individual levels. Several frameworks on AI governance are presented in literature. Here, frameworks aimed towards organizations and companies are presented.

According to a framework presented by Gasser & Almeida (2017), AI governance takes effect in the legal, technical, and social layers. Chhillar and Aguilera (2022) present a similar framework but add the market and economics forces as a fourth layer. The legal method refers to explicit mandates imposed by national or regional authorities, while the social method is formed by pressure from surrounding communities and the will to adhere to norms (Gasser & Almeida, 2017). The market indirectly governs AI through economic forces, such as demand and supply, by inflicting financial losses for poor governance. Public failures are bound to result in reduced revenues, which guides organizations to act. (Chhillar & Aguilera, 2022) Finally, technical solutions, such as including code-modules enhancing interpretability (Butcher & Beridze, 2019), setting hard limits for AI system outputs (Myllyaho, et al., 2022), data governance and official standards (Gasser & Almeida, 2017), or constraining programmatic advancement without human intervention, forms the final modality. Even though effective alone, these layers are intertwined, and work best in tandem (Chhillar & Aguilera, 2022).

Moreover, Mäntymäki et al. (2022b) introduces a framework for AI governance, dividing it into three layers: the AI system, organizational and the organization’s contextual environment levels. The framework is intended to be of practical help for organizations designing and deploying AI systems to bridge the gap between principles and practice, aiding them in translating AI ethics principles into practice and aligning with the forthcoming EC’s AI Act. The contextual environment level entails social actors and constructs not in the range of the organizations influence, including regulation and law makers, social principles, and guidelines as well as stakeholders’ demands and pressure. Inputs from the environmental layer are processed by the organizational layer, which includes strategy and value alignment, often performed by managers. External requirements from the environmental layer and internal steering from the organizational layer are practically implemented in the AI system level by designers and developers. The framework divides the operational level into components, including AI and algorithm design and operations, risk and impact management, data and development operations, ensuring transparency and explainability, specifying accountability, and ensuring law compliance. These components require processes during the entire lifecycle of the AI system. Information needs to flow from the AI system level to the organizational level, which in turn communicates with external stakeholders.

Similarly, Shneiderman (2020) presents a framework for human-centered AI governance, consisting of team, organization, and industry layers. Again, this framework is presented to help bridge the gap between principles and practice. The innermost team layer entails applying good technical practices to software engineering teams, such as enabling audit trails, conducting verification and bias testing, ensuring explainability and adopting verified workflows. On the organizational layer of AI governance, leadership commitment, frequent training, open communication, and internal reviews are presented as practices to adopt. Finally, the third governance layer, industry, includes regulation, external auditing, warranties from insurance companies, as well as professional organizations and NGOs developing and promoting sustainable AI guidelines.

Furthermore, Brendel et al. (2021) presents a general framework for AI governance, consisting of interrelated ethical considerations, contextual environmental and managerial decision dimensions. The managerial decision dimensions are further divided into strategic, tactical, and operational level decisions.

As can be seen, the presented frameworks have similarities. They include the technical or internal layer and the contextual environment or external layer. To govern AI adequately, actions must be taken by the organization using AI systems. Furthermore, two also explicitly introduce the organizational layer to bind these two together. As Laato et al. (2022b) highlight, AI governance processes and requirements vary since development projects are unique. Various aspects might be more relevant in different use cases.

Along with frameworks aimed at organizations and companies, multiple broader AI governance frameworks have been presented in literature (see de Almeida, et al., 2021; Wilson & der Velden, 2022; Wirtz, et al., 2020).

In addition to academic publications, private companies such as Boston Consulting Group, PwC, and Google, have published their own AI governance tools and frameworks for achieving sustainable and trustworthy AI (BCG, 2022; PwC, 2020; Google, 2022)). Going even further, Saidot offers AI governance and transparency as Software as a Service (SaaS) (Saidot, 2022).

2.3.3 AI auditing tools and frameworks

According to Floridi et al. (2022), auditing frameworks are protocols that describe how audits and what, and according to which requirements. They are described as a structured procedure and are based on impact assessment. In contrast, auditing tools are technical products of conceptual models, which help measure, assess, or visualize properties of AI systems. (Floridi, et al., 2022)

Mökander et al. (2021) argue that the ethics-based auditing process should be holistic, traceable, accountable, strategic, dialectic, continuous, and driving redesign. Moreover, sustainability should be considered when giving certifications for trustworthy AI, and auditing the social, environmental, and economic sustainability aspects should be a core requirement for ethical assessment (Genovesi & Mönig, 2022).

Assessing and rating ethical requirement fulfillment should start by identifying concrete case and domain specific risks through expert and stakeholder consultation. There should be minimum requirements for all relevant dimensions. Should any of the minimum requirements not be met, the system should be deemed unsustainable, regardless of how the system performs in other features. (Genovesi & Mönig, 2022) Reporting should be punctual, since according to the AI Act (2021), supplying incorrect or misleading information shall be subject to fines up to 2% of the world-wide annual turnover or €10m, whichever is greater.

Audit results can guide consumer valuation and steer developers' attention toward these audit areas (Genovesi & Mönig, 2022). Moreover, proactive assessment of AI systems can help prevent harm in avoiding discrimination, liability issues, and prevent financial and reputational harm for organizations that operate AI systems (Floridi, et al., 2022). However, as Braganza (2022) argues, auditing might not always lead to the desired goal. Audit goals, as accurate as they might be, could be misinterpreted for the actual goals, rather than as proxies of the underlying societal and moral goal. The underlying goal of the audit measures might be missed if excessive focus is given to quantitative measures. (Braganza, 2022)

University of Oxford researchers Floridi and colleagues (2022) have created a procedure for assessing the conformity of AI systems with the EU AI Act, called the capAI (conformity assessment procedure for AI systems). It serves as a practical tool, to demonstrate and ensure trustworthy AI development and operation. Furthermore, it aids organizations in translating high level principles into demonstrable criteria that influence the design, development, deployment, and use of ethical AI.

Moreover, Koshiyama et al. (2021) propose their own AI auditing framework. In the framework, the algorithm, consisting of data, model, and development process, is presented as the centerpiece of AI system auditing. Moreover, the framework outlines development of the algorithm, assessment of relevant auditing criteria (such as explainability, fairness, and robustness), implementing mitigation strategies, and providing assurance as dimensions of auditing.

Similar auditing tools are called for. Minkkinen et al. (2022a) found out that ESG analyses used do not incorporate a sustainable AI perspective but embedding them into ESG pillars could be justifiable, allowing ESG analyzes to be used as ethics-based audit tools. However, Brusseau (2020) argues that standard ESG frameworks are insufficient for evaluating companies with AI as their core competence, but rather specialized and customized metrics from AI ethics principles are required. Moreover, Koshiyama et al. (2021) highlight the need for a tool intended to help organizations self-assess and flag systems into low, medium, high, or unacceptable risk categories.

MLOps procedures and tools can advance AI system auditability. MLOps refers to promoting automation and monitoring during the development and deployment of an AI system, including testing, integration, release, infrastructure management, and deployment. Post-deployment monitoring of ML systems must consider the inherent problems of ML systems, such as bias and model drift that may develop over time. (Granlund, et al., 2021). MLOps pipelines' benefits outweigh its cost when a company no longer develops ML systems as proof-of-concepts but incorporates ML in their normal business activities (Mäkinen, et al., 2021). Furthermore, dashboards for tracking KPIs and other metrics in MLOps pipelines is a good best practice (Papagiannidis, et al., 2022).

To conclude, there are no ubiquitous auditing criteria presented in literature. However, elements of AI governance models and ethics principles serve as a starting point for audit requirements. Additionally, standards have been published addressing sustainable AI. Several standards around AI governance and ethical AI have been proposed. Auditing is a fundamental part of being standardized.

2.3.4 Standards for AI governance and ethics

As mentioned in the AI Act (2021), standardization is to play a key role in providing technical mechanisms for AI system providers to ensure compliance with the proposal. Moreover, Schmitt (2022) claims international non-government lead standard organizations to have a role in the development of AI governance, as they do with most emerging technologies. Johnson and Bowman (2021) agrees and describes standards as providing a means to define regulatory norms providing a common language or basis of measuring the performance of AI governance initiatives.

The International Organization for Standard (ISO), International Electrotechnical Commission (IEC), the European Committee for Standardization (CEN), and the European Committee for Electrotechnical Standardization (CENELEC) share an understanding of the definition of a standard. As defined in ISO/IEC Guide 2:2004, a standard is a "document, established by consensus and approved by a recognized body, that provides, for common and repeated use, rules, guidelines or characteristics for activities or their results, aimed at the achievement of the optimum degree of order in a given context.”

Standards are voluntary, but in order to take part in the modern markets, compliance with at least some standards are often a necessity (Cihon, 2019). Furthermore, standards are inseparably linked to certifications and audits since a certification is used to verify whether the object being certified meets the requirements of the standard. The verification is verified by the means of an audit (Power, 1997).

According to an extensive overview of the current state of AI standardization by Nativi et al (2021), AI standardization is still in a nascent state, with plenty of standards currently in the development process expected to be published in the coming years. However, the most relevant and active international standard developing organizations (SDOs) in the AI system space are the joint technical committee (JTC) ISO/IEC and Institute of Electrical and Electronics Engineering (IEEE) (Cihon, 2019). Also, JTC of CEN and CENELEC (CEN-CLC/JTC 21) is a key SDO in the EU (Ebers, 2021). There exist a couple of useful standards for achieving sustainable AI, which will be presented next.

ISO/IEC JTC 1 SC 42 is a first of its kind to address AI system standardization (Ebers, 2021). While many standardization projects under the JTC1SC 42 have already been published, (e.g., AI governance (ISO/IEC 38507:2022) and AI robustness (ISO/IEC TR 24029-1:2021)), many are still under development (e.g., standards concerning AI systems (ISO/IEC NP 5392), AI trustworthiness (ISO/IEC AWI TS 24462), and AI ethics (ISO/IEC AWI TR 24368)) ((Ebers, 2021); (ISO/IEC, 2022)).

In addition to ISO and IEC, the IEEE launched a program to tackle the ethical issues brought forth by AI systems’ development and distribution (Chatila & Havens, 2019). The P7000 standard series addresses the ethical design principles for AI systems (Ebers, 2021), including considerations on algorithm bias (IEEE 7003), individual and societal ethics during system design (IEEE 7000-2021), transparency of autonomous systems (IEEE 7001-2021) and impact assessment for AI systems on human well-being (IEEE 7010-2020) (IEEE, 2022).

Furthermore, CEN-CENELEC formed its own focus group to address the gaps between the international activities done by other SDOs and the needed European standardization (CEN-CENELEC, 2020). CEN-CENELEC does not produce standards, but rather identifies requirements specific for the EU, and the CEN-CLC/JTC 21 shall work in collaboration with ETSI to adapt those standards into the EU value frame (Nativi, et al., 2021).

Despite the efforts, creating standardizations for the socio-technical AI systems is challenging, since AI technologies change rapidly, ethical subjects concerning AI are still under debate, and standards cannot focus solely on the technology, but must address the social components of the AI systems (Ebers, 2021).

2.4 Ecosystem of AI governance

An ecosystem approach aims to offer a way of thinking about the diverse stakeholder roles and mechanisms needed to enable sustainable AI (Percy, et al., 2021). Nevertheless, Minkkinen et al. (2022b) and Floridi et al. (2022) argue that despite the efforts, a mature multi-actor ecosystem is yet to be established.

According to Tsujimoto et al. (2018), a multi-actor network perspective views an ecosystem as a complex network of multiple actors, which all have diverse backgrounds and characteristics. Conceptualizing the networked interactions of AI actors as ecosystems is well established in policy and strategy papers as well as in the academic literature (Minkkinen, et al., 2022b). As stated by Minkkinen et al. (2022b), organizations must understand the elements of AI governance and recognize their own part in the multi-actor ecosystem. These actors can be, for instance, governments, private firms, universities, consumers, and investors, and variables between these actors include regulations, money, contracts, and knowledge (Tsujimoto, et al., 2018).

In the global space, Schmitt (2022) recognized multiple state-led and non-state-led actors embedded in existing AI governance architecture, of which the most relevant for the context of this study is the EC. The EC has articulated an ecosystem approach and is willing to position itself as a key player in it. This is evident by the large number of events and statements, and especially the AI Act published in April 2021 (Renda, 2020). Moreover, the ecosystem articulated by the EC has the characteristics of a multi-actor ecosystem, formed around the value propositions of sustainable AI development and operation (Minkkinen, et al., 2022b).

As laid out in the AI Act, responsibility of guaranteeing compliance and finding and relieving potential breaches is shared among high-risk AI system providers and users. However, the EC aims at setting up governance entities covering the entire union. Indeed, according to the AI Act, a European Artificial Intelligence Board will be set up on the Union level to gather and share best practices and issue recommendations on uniform administrative praises. Also, to increase public transparency and allow for ex-post supervision, the Commissions will establish and manage a central registry for standalone high-risk AI systems (AI Act, 2021).

Furthermore, the AI Act proposes the establishment of national level authorities to supervise the implementation of the AI Act. This supervisory authority assesses, alerts, and allocates third-party organizations, called conformity assessment bodies, or notified bodies, to conduct the actual conformity assessment. The notified bodies analyze and approve the quality management systems used by the high-risk AI system providers in the design, development, and testing process. Furthermore, examining the technical documentation of each high-risk AI system within the same quality management system is also a responsibility of the notified body. Finally, the notified body shall deliver a proof of conformity via an EU technical documentation. Organizations willing to become a notified body must apply for such a role from the notifying authority. (AI Act, 2021) However, as noted earlier, the AI Act is still at a proposal phase and its implications have yet to take effect.

In addition, Minkkinen, et al. (2022b) try to elaborate on the emerging ecosystem. By analyzing EU documents, they found that trust is assumed as the foundation of the ecosystem. Moreover, ethics and competitiveness are viewed as complementary, not as competing objectives. Furthermore, European values and norms are the basis of the ecosystem. (Minkkinen, et al., 2022b)

According to Floridi et al. (2022), guaranteeing that high-risk AI systems meet the requirements of the AI Act would demand for a well-functioning auditing ecosystem, and none of that sort exists today. This ecosystem would need two components. First, the actors of the ecosystem need well-aligned auditing tools and expertise to conduct the auditing procedures to show compliance. Second, an institution would need to be established to clarify the roles and responsibilities of authorities and private companies. Such an ecosystem is sketched in the AI Act. (Floridi, et al., 2022) Indeed, Percy et al. (2021) argue that to be effective, the ecosystem needs transparency via explainable AI and adequate process and documentation formalization supporting internal auditing, leading up eventually to external authorization processes. Accountability tools function best, when used as part of an ecosystem of accountability, where responsibility is defined clearly. Using some mechanisms without an operating infrastructure could backfire. (Percy, et al., 2021)

As presented earlier, Chhillar and Aguilera (2022) found through an extensive literature review, that scholars have emphasized the hybridization of AI governance, i.e., responsibility being held by private actors as well as the public sector. Moreover, the importance of organizational AI governance has been recognized (see Gasser & Almeida, 2017; Mäntymäki, et al., 2022a; Shneiderman, 2020. There is an emerging consensus that organizations deploying and using AI systems must play a role in the multi-actor AI governance ecosystem in ensuring sustainable AI (AI Act, 2021; Chhillar & Aguilera, 2022; Floridi, et al., 2022; Mäntymäki, et al., 2022a).

However, not all organizations have the resources or capabilities to address the sustainable AI issues on their own, hence, achieving sustainable development and use of AI is a large-scale problem requiring the cooperation of multiple actors (Mäntymäki, et al., 2022a). Moreover, in certain circumstances outlined in the previous subsection, internal audits are not always enough to verify compliance. Therefore, the sustainable AI ecosystem requires interorganizational activities between private and public organizations as well as activities between private companies (Minkkinen, et al., 2022b).

Seppälä et al. (2021) highlights that interorganizational activities addressing AI governance should be deliberately examined in literature. They identified ways organizations deploying AI systems internally implement ethical principles in practice but recognized that the nature of activities transcending organizational boundaries is still ambiguous. Moreover, Mäntymäki et al. (2022b) suggest future research should investigate the dynamic between AI governance actors of all sizes, including private companies.

Interestingly, Minkkinen et al. (2022b) recognized the developing ecosystem to have implications for private company business models. In fact, they recognize a gap in literature, and call for research to examine business models around services and products addressing sustainable AI challenges within the AI governance ecosystem. Butcher and Beridze (2019) recognize activities provided by private companies, such as helping other firms to put AI ethics into practice, assessing bias of AI systems, and publishing governance frameworks to be used by other companies. Nevertheless, the activities between multiple organizations and their nature are not covered in prior literature. Moreover, Minkkinen et al. (2022b) suggest that future research to only focus on developing sustainable AI solutions but could also be conducted on the auditing and consultation business around sustainable AI.

To conclude, the ecosystem around AI governance is still in an emerging state. While the EC is taking a strong stance, the responsibility of ensuring AI sustainability lies also on AI operators and private companies. However, there is a gap in sustainable AI and AI governance literature: the role of consultancies within the AI governance ecosystem is yet to be researched.

2.5 Synthesis of the literature review

Through the literature review, sustainability, AI, and sustainable AI were defined. Furthermore, the point-of-view adopted in this thesis was laid out. In summary, sustainable AI is approached from three interrelated dimensions. From the social point of view, fairness, accountability, and transparency are the most vital principles, while the environmental impact of developing and operating AI software rather than sourcing hardware is of greatest interest in this study. Moreover, macro-economic aspects will not be addressed, but rather the focus will be on analyzing the economic incentives of organizations for investing into initiatives promoting the sustainability of AI.

Next, different forms of AI governance were established as a means to achieve the sustainability of AI. Also, the importance of AI system auditing as a tool for verifying sustainability claims was discussed. The requirements laid out in the AI Act were presented, followed by discussion about frameworks and characteristics of AI auditing and governance. It was found that AI governance frameworks share the understanding that AI governance takes place both internally in an organization as well as externally, steered by regulation and social demands. Established tools or processes for AI auditing are yet to emerge, even though some approaches and frameworks have been introduced. Furthermore, SDO’s have published a handful of standards which address the ethical issues of AI, and more are being prepared at the moment.

Finally, the actors in the AI governance ecosystem were elaborated on. The EU is taking a strong stance and is building an ecosystem around itself. However, activities within and between individual organizations will be needed. Still, the nature of these activities and the role of private companies is yet to be researched. Figure 4 presents the framework, which depicts how sustainable AI is to be supported by the AI governance ecosystem based on the literature review.

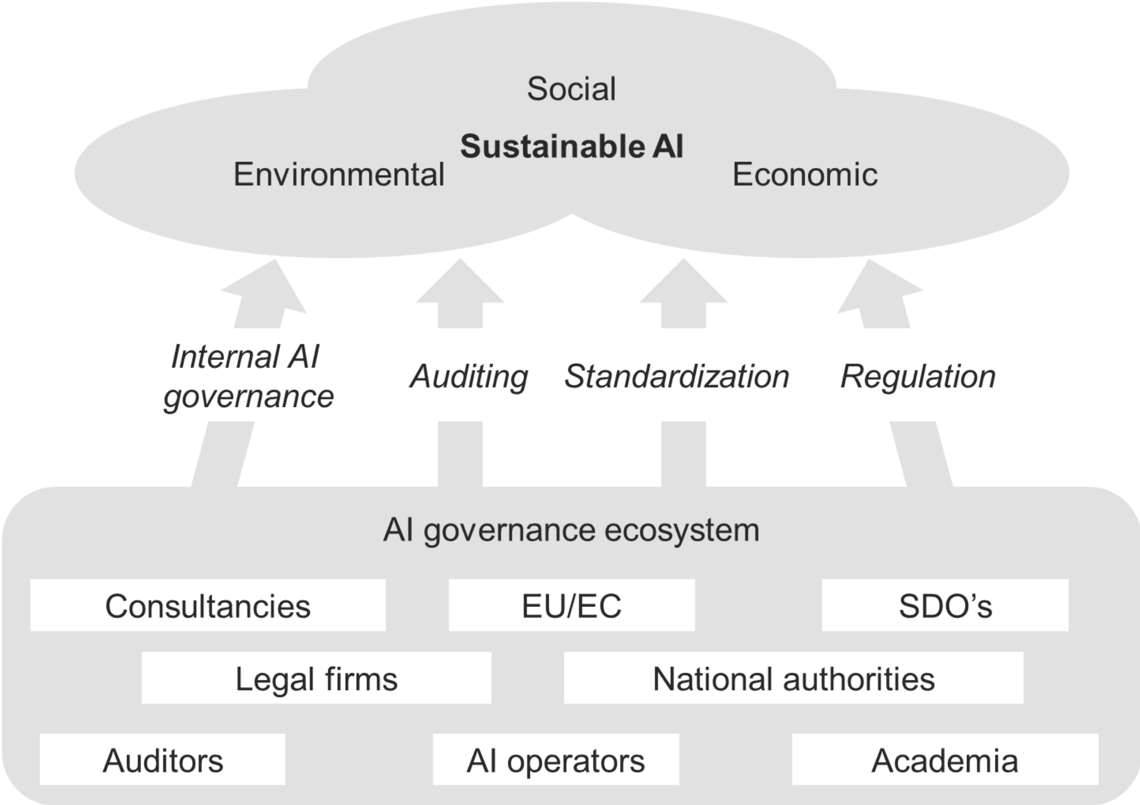


Figure 4: *Sustainable AI enabling framework describing how the AI governance ecosystem supports sustainable AI.*

However, which specific services private consultancies are to offer in the ecosystem, and what value they produce is still unknown. Moreover, the nature of interorganizational activities is yet to be analyzed in detail. In this study, the business model of the case company will be used to study just this: how consultancies create, deliver, or capture value within the AI governance ecosystem. Through designing the business model, insights can be gained on the role of private companies and the nature of the services they offer.

3. BUSINESS MODEL DESIGN

Business models are highly relevant in management practices, and their importance has increased in academia too (Wirtz, et al., 2016). In this chapter, the term business model and business model innovation are explained. First, the concept of the business model is studied, and different approaches introduced. Next, the business model innovation process and tools for it are presented. Finally, the Business Model Canvas and how it is to be used in business model innovation is explored in greater detail.

3.1 Business models and their objective

To gain a comprehensive understanding of the definitions of business models and to find commonalities among them, various authors have conducted systematic literature reviews (see Zott, et al., 2011; Fielt, 2013; Wirtz, et al., 2016; Geissdoerfer, et al., 2018). In their study, Zott et al. (2011) found that scholars do not have a common understanding of the meaning of a business model, and research is developing in different directions guided by the phenomena of interest of the respective researcher. On the other hand, Wirtz et al. (2016) argues that the heterogenous understandings of different scientific fields is converging into a shared business model understanding. Their analysis yielded an understanding of four research focus points: innovation, change and evolution, performance and controlling, and design (Wirtz, et al., 2016).

Moreover, despite the conceptual differences, Zott et al. (2011) recognised four emerging themes: 1) the understanding of the business model as being separate unit of analysis which transcends organizational boundaries; 2) business model being seen as a holistic approach of defining how business is done; 3) cooperation plays an important role; and 4) business models seek to explain value creation and value capture. Similarly, by analyzing the frequency of components presented in literature to be a part of the business model, Wirtz et al. (2016) found value creation or market offering to be present in a clear majority of papers. Geissdoerfer et al. (2018) found equivalent results, with value proposition, value creation and delivery, and value capture as main components of business models.

Osterwalder et al. (2005) distinguishes the concept of business model into three different hierarchically linked categories. The highest conceptual level includes the definitions of what a business model is, and of the conceptualization of its components. This level allows for describing what a business does to stay operational. The second level consists of business model types or generic meta-model types of business models containing common characteristics. Finally, the instance level consists of descriptions of real-life business models. Authors have used this level of the concept to analyze companies.

As argued by Al-Debei and Avison (2010), the business model serves a practical purpose as a conceptual alignment tool between strategy and processes of a business. Additionally, it can be used to measure performance and to support innovation (Osterwalder, 2004).

According to Osterwalder et al. (2005) a business model is "a conceptual tool that contains a set of elements and their relationships and allows expressing the business logic of a specific firm. It is a description of the value a company offers to one or several segments of customers and of the architecture of the firm and its network of partners for creating, marketing, and delivering this value and relationship capital, to generate profitable and sustainable revenue streams". Similarly, Wirtz et al. (2016) define it as "a simplified and aggregated representation of the relevant activities of a company". Likewise, Zott and Amit (2010) conceptualize a business model as a system of dependent activities transcending organizational boundaries, creating value, and appropriating a portion of that value. Moreover, the definition of Osterwalder and Pigneur (2010) is fully in line with the above definitions and concludes a business model to "describe the rationale of how an organization creates, delivers, or captures value". This final definition captures all essential elements for this study and will be used throughout the rest of the thesis as a working definition.

To conclude, business models serve various purposes and have been researched from multiple angles. The core concept of business model lies in an organization's value proposition to the customer, as well as an understanding of through which activities value is delivered and captured.

3.2 Designing business models

With the objectives of a business model clear, focus can be shifted to creating one. New business models or diversifications of the existing models are needed, when to fill a new customer proposition, new activities and processes are required (Johnson, 2018). According to Geissdoerfer et al. (2018), diversification can become topical when new opportunities or challenges appear in the organization's environment. Indeed, Osterwalder and Pigneur (2010) acknowledge bringing new services or products to market to exploit existing knowledge or resources as an objective for designing business models.

Developing a business model is discussed using the concepts of business model design, development, and innovation. The process refers to altering the value proposition to customers by, for example, making changes to the service offering or revenue model, as well as the operating model. (Zott & Amit, 2010; França, et al., 2017) According to Hansson et al. (2022), the distinction between business model development and innovation is subtle, but innovation goes beyond design in the aspect of novelty. However, for example, França and colleagues (2017) throughout their article refer to the overall process of coming up with new or better business models as “business model innovation and design”. Moreover, Geissdoerfer et al. (2018) describes business model innovation to include the development of a new business model as well as the diversification of existing business models into additional ones, for example, as a reaction to new opportunities or challenges in the organization’s environment.

In this study, a new business model is diversified from the existing one, by creating a new value proposition, determining activities, and modifying other business model dimensions, to capitalize on a new business opportunity around sustainable AI. Hence, in line with the above definitions, the process will be referred to as business model innovation.

Amit and Zott (2015) found prerequisites for business model design to include using templates, since templates are proof of successful concepts, and can be used in guiding attention to relevant aspects while designing. Multiple ways have been proposed for designing and communicating business models, referred to as “business model frameworks”. These frameworks do not only define the elements of a business model, but also the relationship between them. (Osterwalder, et al., 2005)

In a review on business models, Fielt (2013) presents well-known frameworks including the Business Model Canvas (Osterwalder & Pigneur, 2010), the Four-Box Business Model (first introduced by Johnson et al. (2008), also presented by Johnson (2018)), the Technology-Market Mediation (Chesbrough & Rosenbloom, 2002) and the Entrepreneur’s Business Model (Morris, et al., 2005). These frameworks have many similarities and differ only in the subtleties and angle of approach (Fielt, 2013).

The Business Model Canvas (BMC) has become a de facto standard framework for business model development (França, et al., 2017). Also, Fielt (2013), already in 2013, stated the BMC to be the most famous and commonly used. Moreover, the BMC has been used in high level academic research (see ur Rehmana, et al., 2016; Metallo, et al., 2018; Sort & Nielsen, 2018; Polydoropoulou, et al., 2020). Hence, the BMC will be used as the framework in this study.

3.2.1 The Business Model Canvas

The Business Model Canvas, presented by Alexander Osterwalder and Yves Pigneur in their book “Business Model Generation” (2010) is a widely adopted framework used in multiple industries to depict and communicate the business model of an organization. It describes the business model through nine related building blocks. Its core-message was first introduced in (Osterwalder, et al., 2005), and is based on Osterwalder’s research (2004).

The nine building blocks of the Business Model Canvas are 1) value proposition, 2) customer segments, 3) channels, 4) customer relationships, 5) key resources, 6) key activities, 7) key partnerships, 8) cost structure, and 9) revenue streams. These blocks cover the offering and value (1), customer interface (2, 3, and 4), infrastructure (5, 6 and 7), and financial aspects (8 and 9), the four main areas of a business (Osterwalder & Pigneur, 2010). The blocks are visually presented in the business model canvas template shown in figure 5.

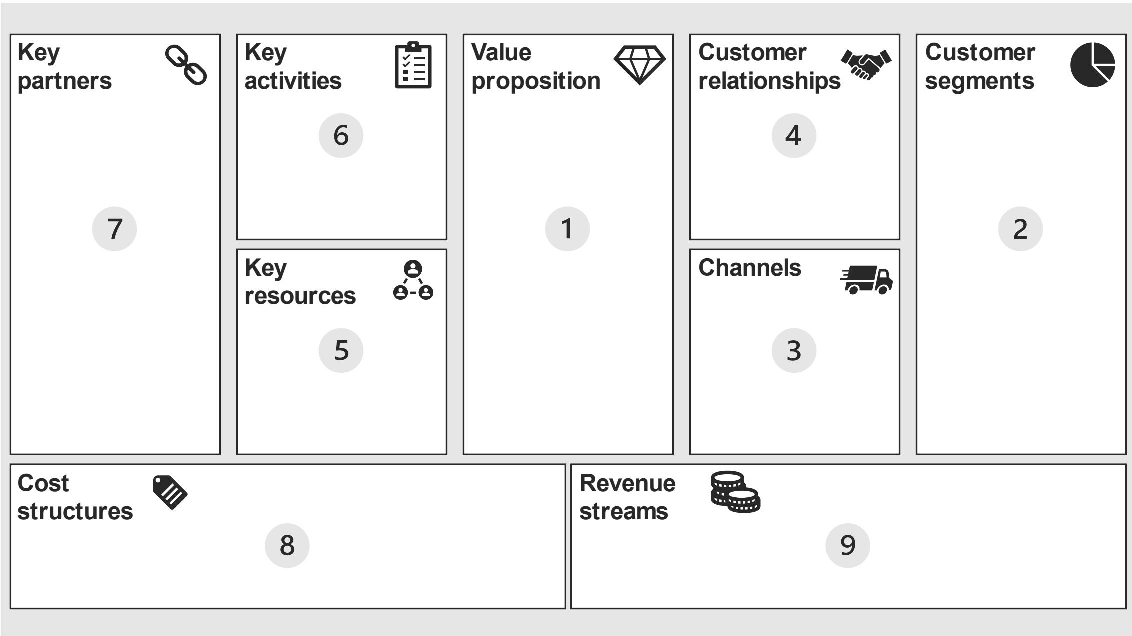


Figure 5: *The Business Model Canvas. Adapted from Osterwalder and Pigneur (2010).*

First, the value proposition block describes what value is created for the customer. Moreover, it covers the offering, i.e., the services, products and/or marketable information that creates that value to customers (Osterwalder & Pigneur, 2010). Value can have multiple forms, including social, economic, and environmental (Zott, et al., 2011). The value proposition is a bundle of benefits the company offers to its customers (Osterwalder & Pigneur, 2010).

Next, the Customer Segments block defines the organizations the company aims to serve. To target its offering, the company might segment its customers into groups with similar behaviors. An understanding of the customer needs is vital. Next, the Channels block describes how a company communicates with its target customers, how it reaches them and how the value proposition is delivered. Channels include communication, distribution, sales, and marketing interfaces towards customers. Moreover, different customer segments and offerings require establishing diverse customer relationships. What kind of a relationship is formed with the customers is addressed in the Customer Relationships block. Types of relationships can range from personal assistance and fully automated and can also include anything in between. (ibid.)

The Key Resources block addresses the most important assets required to create and offer the value proposition, maintain customer relationships, and reach customers. These resources can be physical, intellectual, human, or financial. They can be owned or acquired from partners. Indeed, the Key Partnerships block describes the network of partners that make the business model work. Partners can perform important activities or provide resources. Using the resources and with the help of partners, some activities need to be performed. What is needed to be done to allow the company to deliver the value proposition is addressed in the Key Activities block. It describes the vital activities a company must perform to create and offer a value proposition to its customers. Different offerings and value propositions might require different activities. (ibid.)

Finally, to allow for a continuation of operations, the business model must be financially viable. The Revenue Streams building block describes how and how much customers pay for the value they receive. This includes different pricing mechanisms. To determine profits, costs must be deducted from revenues. The Cost Structure block addresses all costs incurred to deliver the value proposition. (ibid.) Table 1 collects the nine building blocks and offers a set of questions to which the block gives answers.

Table 1: *Business Model Canvas building blocks and guiding questions (Osterwalder & Pigneur, 2010).*

BMC building block	**Questions answered**
Value proposition	What value is delivered to the customer? Which needs are we satisfying? What bundles of products and services are offered?
Customer segments	To whom is value created? How are they segmented? What are their needs?
Channels	Through which channels are communication, value delivery, and awareness managed with the customers?
Customer relationships	What kind of relationships do customers expect? Do they differ depending on the customer?
Key resources	What resources are required to create and deliver the value proposition, and foster customer relationships?
Key activities	What activities must be performed to deliver the value proposition?
Key partnerships	Who are needed as partners and what is acquired from them to deliver the value proposition? Which activities do key partners perform?
Cost structure	What are the inherent costs of the business model? Which activities and resources are most expensive?
Revenue streams	What value are customers willing to pay for? How does the customer pay?

As can be seen, the blocks are closely related to each other, and addressing them alone would be futile. Hence, working on the business model in a holistic manner is essential. However, depending on the situation, some blocks might be more relevant. For example, in bringing new services or products to market, the customer segments and value proposition are of high importance. The Value Proposition Canvas allows for a deep dive into the customer needs and provided value.

3.2.2 The Value Proposition Canvas

The Value Proposition Canvas (VPC) is introduced thoroughly in Osterwalder and colleagues' book Value Proposition Design (2015). It is embedded in the business model canvas, as it provides a zoomed in view and addresses the value proposition and customer segments. The canvas has two sides, customer profile and value map which are both broken down further. First, the customer profile helps in clarifying customer understanding by describing the customer segment in a more granular way. On the other side, the value map describes how value is created for the customers in a structured and detailed way. Fit is achieved when the offering meets the customer needs. (Osterwalder, et al., 2015). The VPC is presented in figure 6.

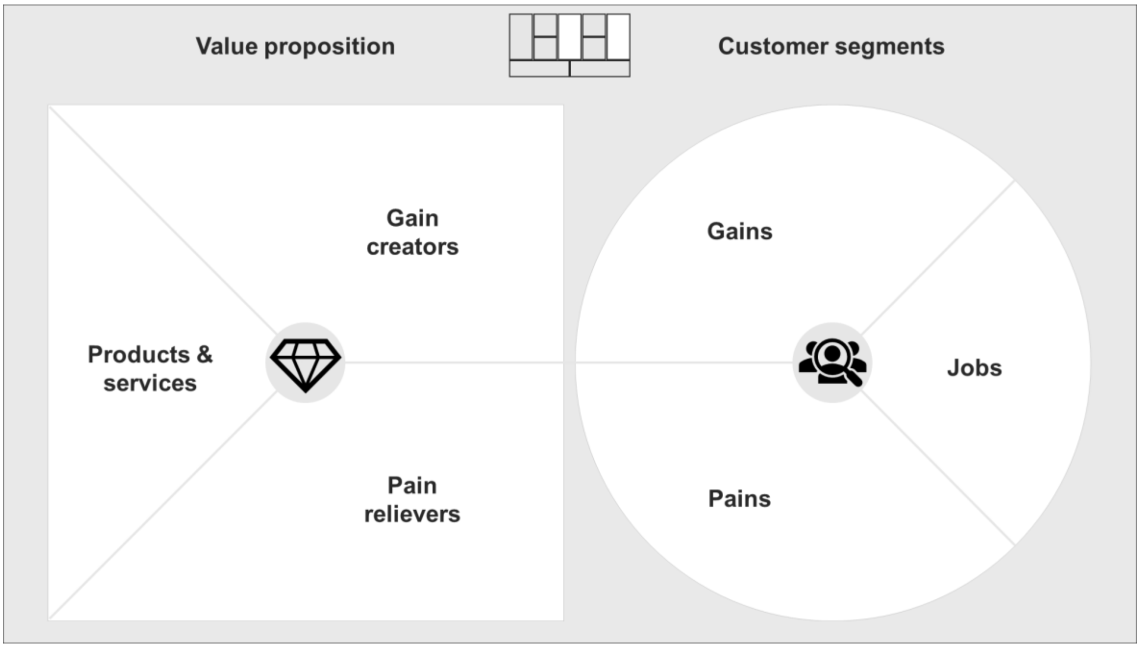


Figure 6. *The Value Proposition Canvas and its relation to the Business Model Canvas. Adapted from Osterwalder et al. (2014).*

Both sides consist of three sections. First on the customer profile side, the customer jobs section describes what the customer is trying to achieve. The style of jobs can vary greatly, as they can be functional, social, or emotional. Secondly, pains describe what is causing challenges for the customer or even preventing the job getting done in the first place. Included in this section are also risks that could occur if the job does not get done or gets done poorly, as well as risk present during getting the job done. Finally, gains describe the benefits customers want. Gains can be required, expected, desired or unexpected and also vary, as they can be functional or social. (ibid.)

On the value map side, products and services simply list the offering, whereas pain relievers describe how the offering alleviates specific customer pain points. Finally, gain creators describe how the offering creates customer gains. When presenting the filled canvas, relevance and importance of each item is signified by their position, with the most important ones at the top. (ibid.) Table 2 gathers the sections described and presents examples of guiding questions needed to be answered in each section.

Table 2: *Value Proposition Canvas sections by and guiding questions (Osterwalder et al. 2014).*

VPC section	Guiding questions
Jobs	What is important to the customer? What are they trying to achieve?
Pains	What challenges are the customers facing?
Gains	What are the customers expecting to gain?
Products and services	What products or services are provided?
Pain relievers	What and how are customer pains relieved?
Gain creators	How are customer gains achieved?

Market fit is achieved when important jobs are addressed, extreme pains are relieved and valuable gains are created. The first level of fit, problem-solution fit is achieved when on paper, a value proposition addresses identified customer jobs, pains, and gains. Product-market fit is attained when the value proposition has gained traction in the market. Finally, business model fit is accomplished when there is evidence that the value proposition can be implemented as a profitable business model. (Bland & Osterwalder, 2019)

Osterwalder and colleagues (2015) present characteristics of designing value propositions in an established organization. They distinguish between invention and improvement, with the former aiming at radically new and different value propositions not constrained by existing business models, while the latter improves or extends an existing value proposition without changing the underlying business model too much. In this study, the VPC is used to diversify and improve a value proposition without introducing radically different approaches.

3.2.3 Using the business model canvas

The Business Model canvas is a hands-on tool designed to encourage discussion, creativity, analysis, and understanding (Osterwalder & Pigneur, 2010). The authors describe the BMC to work best when presented on a large surface allowing multiple people to start sketching and discussing the business model of a company using sticky notes.

The process of business model innovation is described to include gaining understanding, designing, implementing and prototyping, and maintaining. The progression is rarely linear, but rather an iterative process where multiple phases are advanced in tandem informing one another. (Osterwalder & Pigneur, 2010). The BMC focuses on design and innovation and draws influence from the design space (Fielt, 2013). Indeed, Osterwalder and Pigneur (2010) present tools and techniques taken from the world of design, to help design more innovative and better business models.

Ideas for renewing a business model can come from any of the nine building blocks, yet the result will certainly affect multiple blocks. Osterwalder and Pigneur (2010) classify four "epicenters of innovation", i.e., points to begin propagating innovation within the BMC. Offer-driven innovation begins from defining a compelling value proposition and capturing its effects on the other blocks. Resource-driven innovation originates from existing resources or partnerships. In turn, customer-driven innovation takes the needs of target customers as the starting point of innovation propagation. Finally, in finance-driven innovation, new revenue streams or cost structures drive innovations. (ibid.)

Additionally, Amit and Zott (2015) suggest focussing on value creation and value capture, when designing business models. Also, considering stakeholder activities is vital since it promotes an understanding of the business model as a unit within an ecosystem transcending organizational boundaries. (Amit & Zott, 2015). For this study, a value-driven approach is adopted as the starting point. However, resources and customer insight are essential throughout the process.

3.3 Tentative sustainable AI consulting business model

To conclude, a business model describes the logic of how an organization creates, delivers, or captures value. It describes an organization's value proposition to the customer, as well as provides an understanding of through which activities value is delivered and captured. It serves multiple purposes, as it is important, for example, in business strategy and processes alignment and as a tool for innovation.

Business model innovation refers to creating a new or diversifying an existing business model, for example, when faced with a new business opportunity. In this thesis the Business Model Canvas and the Value Proposition Canvas are used as tools for the innovation process. The BMC is a famous framework for presenting a business model, covering the offering, customer interface, infrastructure, and financial aspects. Furthermore, the VPC provides a zoomed in view of the value proposition and customer segments.

By designing a business model for an Solita, an IT- and design consultancy, an understanding of the role of the company in the AI governance ecosystem can be created. Figures 7 present visually how the BMC is used to gain insights into the various elements present in the framework built in the previous chapter.

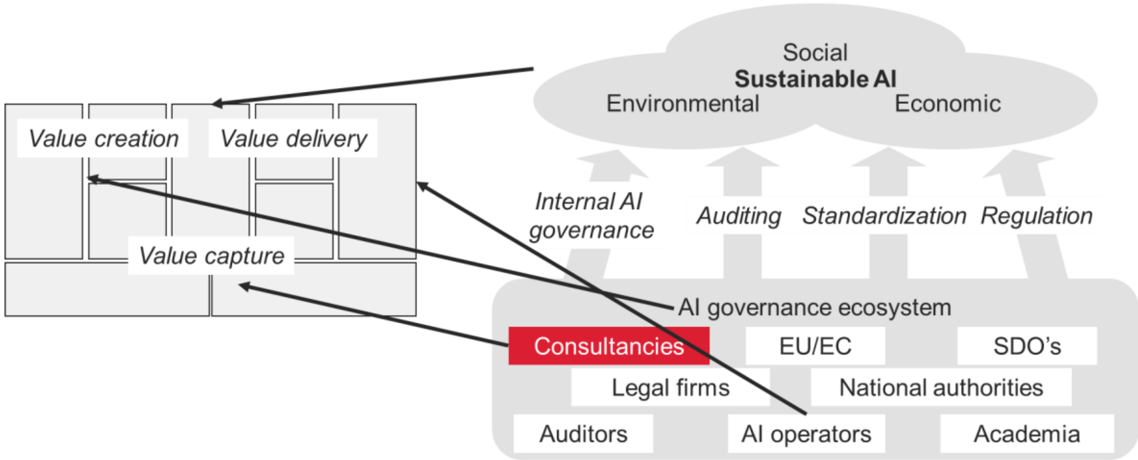


Figure 7: *The BMC is used to examine roles and activities of actors in the sustainable AI enabling framework.*

The BMC is designed from the point-of-view of consultancies within the AI governance ecosystem. Ensuring sustainability of AI developed and used by AI operators is the central value proposition, which is delivered through various activities. Some of the possible activities were recognized in the literature review. However, which of these are performed by consultancies will be discovered in the empirical part of the study. As a hypothesis, advising on and enabling internal AI governance as well as auditing are the most relevant services provided by a private consultancy like Solita. However, standardization and regulation are expected to be provided by other actors in the ecosystem and are likely to be out of the scope of services addressed in this study. Figure 8 depicts how the elements of the framework built based on the literature review are expected to be placed on the BMC.

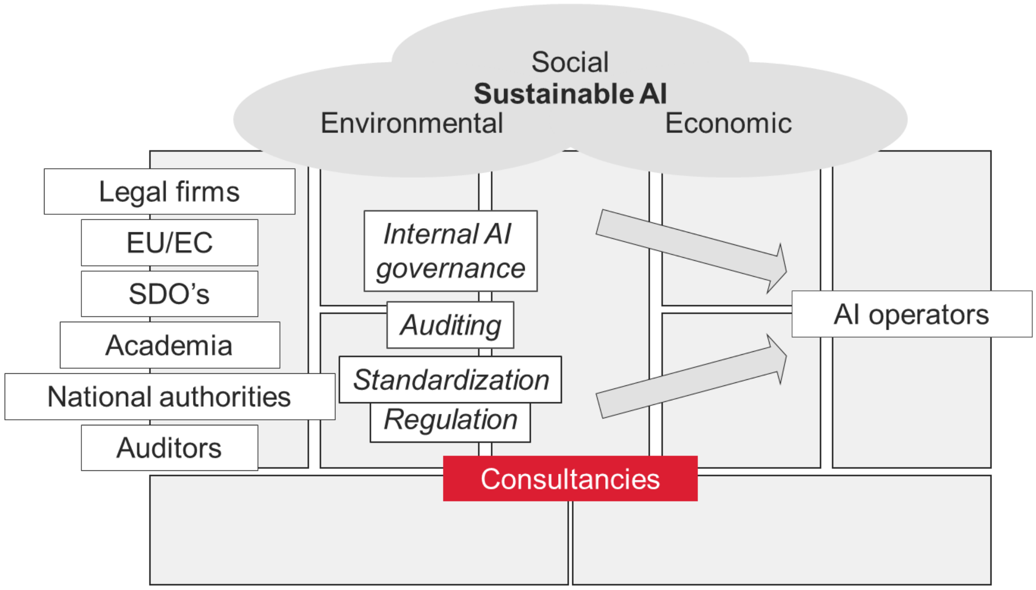


Figure 8: *Elements of the sustainable AI enabling framework placed in respective positions in the BMC. The darker color represents the areas of greater interest, while the objects in lighter color are important but not the focus of the empirical research.*

To conclude, the Business Model Canvas is used in the empirical part of this study to investigates the interorganizational activities performed in the sustainable AI and AI governance ecosystem through the point-of-view of the case company. The empirical process is described in detail in the next chapter.

4. METHODOLOGY

This chapter describes the methodological approach of the empirical part of the study. The empirical part includes designing a business model for Solita around its services enabling sustainable AI. First, the chosen research methods are presented and justified. Next, the data gathering methods are explained. Finally, the process of empirical research and data analysis is described.

Hereafter, the term “client” is used to refer to clients of the case company, i.e., organizations for which the case company, Solita, provides services. On the other hand, “customer” is used to refer to the customers of Solita’s clients.

4.1 Action research case study as research strategy

As mentioned in the introduction, a multi-method qualitative approach is selected as the methodological choice. Furthermore, this thesis employs a research strategy of action research embedded in a case study. A research strategy can be defined as the overall plan of how the research questions are intended to be answered (Johannesson & Perjons, 2014). However, Saunders et al. (2019, p. 190) point out that research strategies are not mutually exclusive but can be intertwined or layered.

Adopting a case study strategy is suitable, since a case study is an in-depth empirical examination of a contemporary phenomenon in its real-life context (Yin, 2009, p. 18), which is among the goals of this thesis. Case studies often use multiple sources of evidence, such as archival research, observation, and interviews to gain a comprehensive understanding of the subject at hand. Moreover, case studies are useful especially when detailed and holistic knowledge is required (Eriksson & Kovalainen, 2008). To make it explicit, the case in this thesis is the sustainable AI consulting business model of Solita, which is studied within its context of the AI governance ecosystem. Business model components are the units of analysis. A common characteristic for case studies is the unclear boundary of the case and its context, which places importance on making it explicit (Yin, 2009).

Moreover, the study adopts an embedded single-case study strategy since the case is believed to be a typical and representative case (Yin, 2009, p. 48) of a sustainable AI consulting business model of a private consultancy within the AI governance ecosystem. The business model components are the unit of analysis, and since there are more than one unit of analysis, the case is embedded (Lee & Saunders, 2017).

This thesis takes an action research approach within the case study. Action research is an emergent and iterative process, aiming to develop solutions to real problems organizations have, through a collaborative and participatory approach (Oliva, 2019). Action research is fitting, as it allows for taking action (i.e., applying business model design tools) in a real-life situation (i.e., a project aiming to conceptualize the business model for an emerging business) (Athanasopoulou & De Reuver, 2020).

More precisely, research through design is the chosen approach, which utilizes design practices to inform research and is deeply rooted in action research. Research through design is an iterative process of generative and evaluative steps taken towards a design objective (here, the business model). During the process, knowledge is gradually gathered and contextualized, which allows for the building of tangible knowledge and solutions simultaneously. (Baldassarre, et al., 2017)

Research, action, and participation must be all present in action research. Moreover, the researchers adopting a facilitative role is a characteristic of action research. It incorporates different forms of knowledge, such as theoretical knowledge, experiential knowledge, and practical knowledge. Furthermore, action research allows for gaining actionable knowledge, useful for the organization, while being academically robust. (Saunders, et al., 2019, p. 201–204).

In action research, researchers and practitioners work closely in iterative cycles to solve a practical problem, and to generate knowledge (Athanasopoulou & De Reuver, 2020). The design objective in this thesis is to conceptualize a business model for Solita's sustainable AI offering, and action is taken in the form applying business model design tools and aligning the designed business model with Solita's capabilities and goals. As is typical to action design research, multiple methods, such as interviews, are used to iteratively construct the design object, which simultaneously produces the research result as co-interpreted by the designer-researcher and the participants who will use the design (Spinuzzi, 2005). Moreover, action research allows for incorporating experiential knowledge and knowledge from practice into academic research (Athanasopoulou & De Reuver, 2020).

4.2 Data gathering methods

4.2.1 Interviews

Interviews are a practical and effective way to gather unpublished information. Furthermore, semi-structured interviews allow for researchers to conversationally gather information, while staying thorough and organized. Open-ended questions are common in such interviews. On the other hand, unstructured interviews are informal in-depth explorations of a general area of interest. Opinions and facts are asked from the interviewee, while allowing them to have room to act as an informant. (Eriksson & Kovalainen, 2008)

In this thesis, unstructured interviews were conducted with workers of Solita, while the client interviews done by Solita in their research project were semi-structured. The former interviews were done as part of design sessions. The unstructured format allowed experienced business designers and other professionals to share knowledge and ideas without constraints.

Solita conducted the semi structured client interviews during spring 2022, to discover the state of sustainable AI (Metsäranta & Rauhala, 2022). 26 organizations across 11 industries were interviewed, including both public and private Finnish and international organizations. Due to the terms of the interviews agreed between Solita and the interviewees, the studied organizations cannot be disclosed. However, the 11 industries covered by the interviews were the following: public services, financial services, health, transportation, media, energy, defense and security, retail, forestry, pharma, and manufacturing. In these interviews, the state of sustainable AI was uncovered through questions regarding the views clients had as well as about actions clients had taken on sustainable AI. Moreover, the tasks clients sought to get done, the goals they were aiming for, and problems they faced were also explored through questions and discussion. The guiding questions used were similar to those presented in table 2 in section 3.2.2.

Client organizations to be interviewed were chosen using purposeful heterogeneous sampling, which is suitable when the aim is to cover a wide range of diverse organizations to gain an overarching understanding of a phenomenon (Patton, 2002). However, selected organizations needed to have some level of AI systems in use or development. The selected organizations ranged in AI maturity from proof-of-concepting to operational AI use. Moreover, the interviewed individuals within the organizations were mainly senior experts or executives, representing the business, analytics, or R&D functions. Permission was granted for the use of the anonymized interview data in further research at Solita.

The author of this thesis was not present in the client interviews but received anonymized, codified, and thematized excerpts from these interviews for analysis. These excerpts were comments from clients which were considered relevant for the qualitative research performed by Metsäranta and Rauhala (2022). A subset of this data, including answers to questions about the tasks, challenges and expected gains, were relevant for this thesis. They were used in this thesis study to inform the customer perspective using the VPC and BMC as tools. Also, qualitative findings from these client interviews made by other Solita employees in the State of Sustainable AI study (Metsäranta & Rauhala, 2022) were used in this thesis. The thesis worker was also able to discuss thoroughly about the interviews with Solita's state of sustainable AI researchers throughout the thesis process.

Internal interviews and design sessions with Solita employees were held during summer 2022 using Microsoft Teams communication platform. In the sessions, data was gathered in the form of interviewing as well as designing and ideating together. Insights were collected on electronic collaboration tools (Miro, PowerPoint) as well as recorded for later analysis. The participants for these sessions were chosen based on their expertise in business models or their knowledge of the sustainable AI business at Solita. Table 3 presents the design and interview sessions, the titles of participants in the session as well as the main topic.

Table 3. *Interviews and design sessions with Solita employees*

Date (2022)	Participant roles	Topic
31.5.	Business lead	Value proposition, different business models, current business model
8.6.	Head of Sustainable AI	Sustainable AI services and products
21.6.	Business lead	BMC
27.6.	Head of Sustainable AI	BMC, offering
7.7.	Head of research	BMC
9.8.	Head of Sustainable AI	Value proposition
10.8.	Head of Sustainable AI	Offering
15.8.	Head of Sustainable AI	BMC, value proposition
18.8.	Data Science Business Lead	Offering, value proposition

During the design interviews where the BMC was of interest, the guiding questions presented in table 1 in section 3.2.1 were used. The length of the sessions varied from 45 minutes to 75 minutes

In addition to the scheduled meetings presented in the above table, the results of design sessions were presented for feedback on multiple occasions to other employees working in the sustainable AI area at the case company. In these sessions, the findings of the thesis were presented, and insight from the emerging discussion was incorporated into the design objects. Nevertheless, these sessions are not listed since they were more spontaneous and shorter.

4.2.2 Supporting sources of data

Observation involves systematically viewing, recording, analyzing, and interpretation of the communication of people. Moreover, observation as a participant refers to the observer informing other participants of the observation, but not taking any part in the activity. Moreover, it can vary in its structuredness, ranging from highly structured and quantitative, with prior goals determined, to unstructured, where no attributes or responses are predetermined. Documenting the observed is essential in all forms of observation. (Saunders, et al., 2019, ch. 9)

In this thesis, client sales meetings were observed, to gain more insight into client needs. Also, internal meetings where offerings were discussed were observed. In all cases, observation as a participant was the chosen approach, i.e., the participants were informed of being observed, but the observer stayed passive. Furthermore, the observations were semi-structured, meaning that objectives and aspects of interest (i.e., gaining insights into client needs) were predetermined, but room was left for observing interesting emerging matters.

Additionally, for data triangulation, secondary data sources were utilized. Secondary data refers to analyzing data that were initially collected or presented for some other purposes. Types of secondary data include surveys, reports, and documentation, including different forms of data such as video, text, pictures, diagrams, and audio. (Yin, 2009, pp. 101-106)

Internal material, such as sales material, workshop notes, project reports and plans, strategy lectures, and company presentations were used to gather information during this study. Moreover, legal reports and third-party submissions were used to validate facts and legal information.

4.3 Research process and data analysis

The empirical research was conducted in phases. First, to form a proposal business model, an initial Business Model Canvas was populated by the thesis worker based on knowledge gained in the theoretical review. Next, findings from Solita's client interviews and insights from observations were used to gain an understanding of the client side of the business model. Secondary data sources were then used to further refine the BMC. Finally, co-design done with Solita employees and feedback received produced the final business model. The canvas was initially presented and edited in Miro, but was later built on PowerPoint slides, where it was modified together with the interviewees.

Qualitative analysis of the client interviews was conducted by Solita using thematic analysis. Also, the analysis performed by the thesis writer was carried out in a similar manner. Thematic analysis is a foundational qualitative analysis method, which aims to search for patterns and themes present in a qualitative data set in order to draw conclusions (Guest, et al., 2012). A process similar to the procedure presented by Saunders et al. (2019, pp. 652-660) was followed to perform the analysis.

First, the anonymized client interview excerpts were read through in detail to become familiar with the data. This was followed by segmentation, i.e., the excerpts were bounded in order to allow for the exploration of thematic elements and comparing them. Guest et al. (2012) describe themes as units of meaning that are observed or noticed in the data by a reader of text data. The excerpts had been coded from the interview transcripts and grouped around emergent themes by Metsäranta and Rauhala (2022) according to the objectives of their state of sustainable AI study. However, during the thesis process, the data was further segmented for the purpose of this study. Indeed, the sections in the value proposition canvas were used to guide the segmentation. For example, challenges mentioned by clients were placed in the pains theme but were further divided into different lower-level themes. Finally, emergent themes were woven together to reduce their number, they were refined, and propositions were formed and tested. The prior analysis performed by Metsäranta and Rauhala (2022) was combined with original analysis done by the thesis worker to form the final understanding of client needs presented in this thesis. Furthermore, the findings of this study were analyzed together with Metsäranta and Rauhala (2022) to ensure they were representative of the original client interviews.

After incorporating the client point-of-view based on interview data analysis, the business model was enriched with insights from client meeting observations as well as internal document research. Client meetings which were observed by the author, were sessions in which a particular service was tailored to the client's needs together with the client. The author attended two of these sessions in May and June 2022. During the sessions, offering proposals were presented and customized based on the client's requests and requirements. Moreover, the value expected by the client was discussed.

Furthermore, internal document search was conducted on multiple types of data, including prior offering meeting notes, client meeting materials, external and internal presentations, and company materials on strategy. Client meeting materials offered valuable insight to the types of services and products proposed to, and requested by, clients. Information of several past client projects was also analyzed. Finally, info-sessions and documents regarding target clients were used to elaborate on the characteristics of most compelling clients, and of the current client strategy of Solita. The Business Model Canvas was filled based on these data, literature review, observation, internal documents, and client interview data, as outlined above.

When designing the sustainable AI offering, current and prior projects, identified client needs as well as possible future needs were listed on an empty canvas. Then, these projects were thematized and organized according to similarities and differences, forming a two-dimensional framework (see figure 13 in section 5.1.2). This process was similar to the thematic analysis done from the client interview excerpts.

Next, during the design process, the BMC proposal was presented to Solita employees. During design sessions, valuable insight and Solita perspective was used to modify the business model. In between design sessions, the author refined the ideas and formulated a following BMC, VPC and offering proposal, which was again addressed in a subsequent design session.

The value proposition and the offering were analyzed on their own in greater detail in later sessions. While the client perspective was left untouched during these sessions, ideation and proposals for various services or products fulfilling the client needs were generated. Furthermore, the desired message and precise wording of each offering category was aligned with the business goals and capabilities of Solita. Moreover, the thesis findings were presented for feedback in internal meetings on multiple occasions. After the design and feedback sessions, the final business model was refined by the author, and was presented to Solita representatives for approval. The entire business model design process is presented in figure 9.

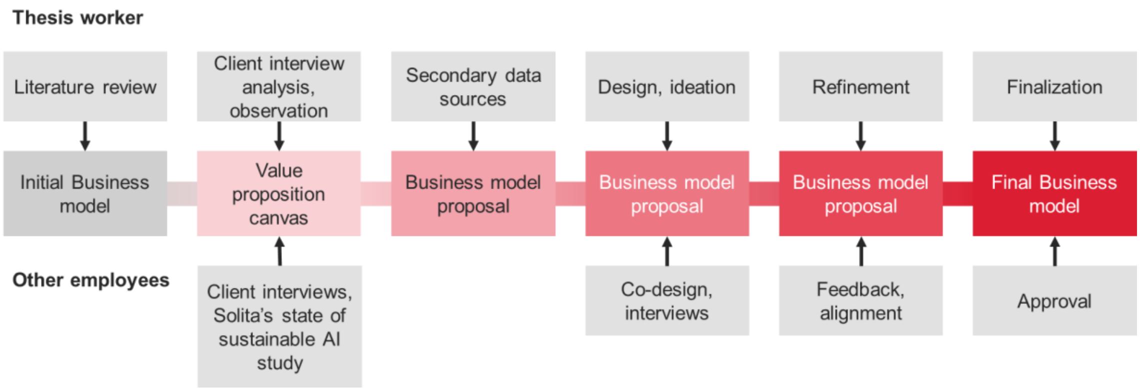


Figure 9: *Business model design process.*

Actions taken by the thesis worker are presented in the top row of figure 9, while input and actions from other Solita employees are presented in the bottom row. The middle row represents the evolution of the business model, which was the focus of the design process. The results of the design process will be presented in the next chapter.

5. RESULTS

This chapter presents the results and findings obtained during the study using the methods described in the previous chapter. The result of defining the business model for Solita's sustainable AI offering using the Business Model Canvas is presented below in figure 10.

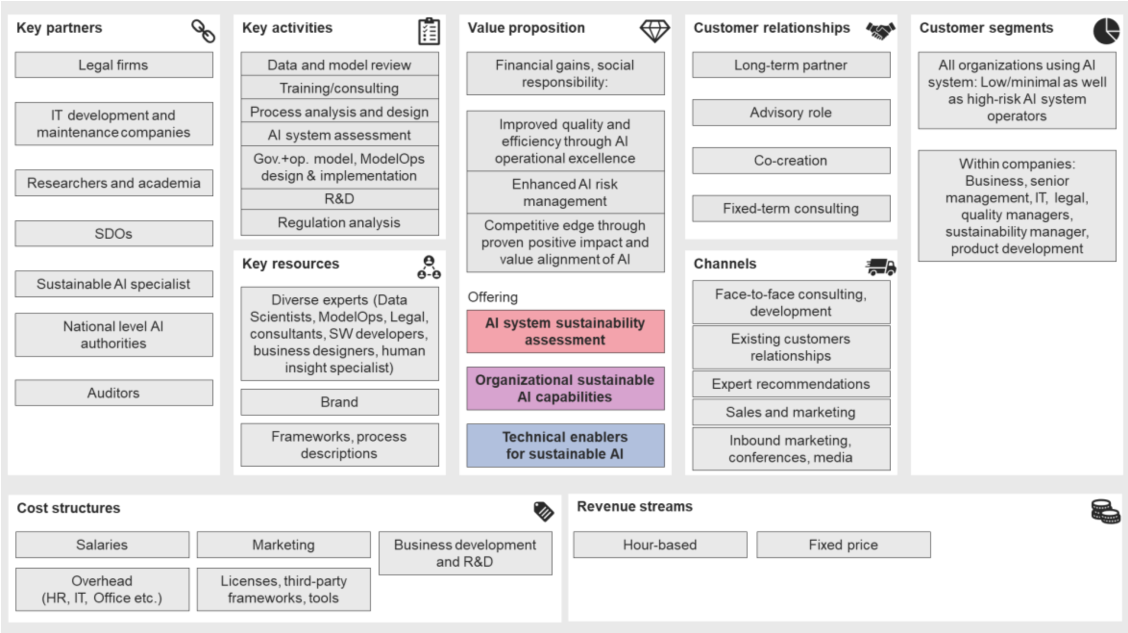


Figure 10. *Solita's Sustainable AI business model depicted using the BMC.*

In the remainder of this chapter, each element of the BMC will be analyzed in detail, and findings regarding those elements will be explained. First, the value proposition and client segments will be discussed. Also, the offering and its value is addressed in greater detail. Next, the characteristics of the client relationships as well as means of reaching clients and delivering value will be presented. After that, the key activities, resources and partners needed to deliver the value proposition will be discussed. Finally, insights of the revenue streams and cost structures inherent in the business model will be reviewed.

5.1 Value proposition and client segments

Solita's sustainable AI offering was arranged into three categories. The categories are 1) AI system sustainability assessment, 2) organizational sustainable AI capabilities, and 3) technical enablers for sustainable AI. Clients of the sustainable AI offering include AI system operators, i.e., all organizations using, developing, and distributing AI systems. The VPC was used as a tool to match the offering with aspects valued by clients. Based on qualitative analysis conducted by Metsäranta and Rauhala (2022) on the client interview data, the client side of the VPC, i.e., jobs, pains, and gains were recognized. This was enriched with analysis of themed excerpts done by the thesis writer, as well as insights collected from client meeting observations. Moreover, Solita's perspective was gained through multiple design and feedback sessions (research process thoroughly described in section 4.2.1). Figure 11 presents the final version of the VPC.

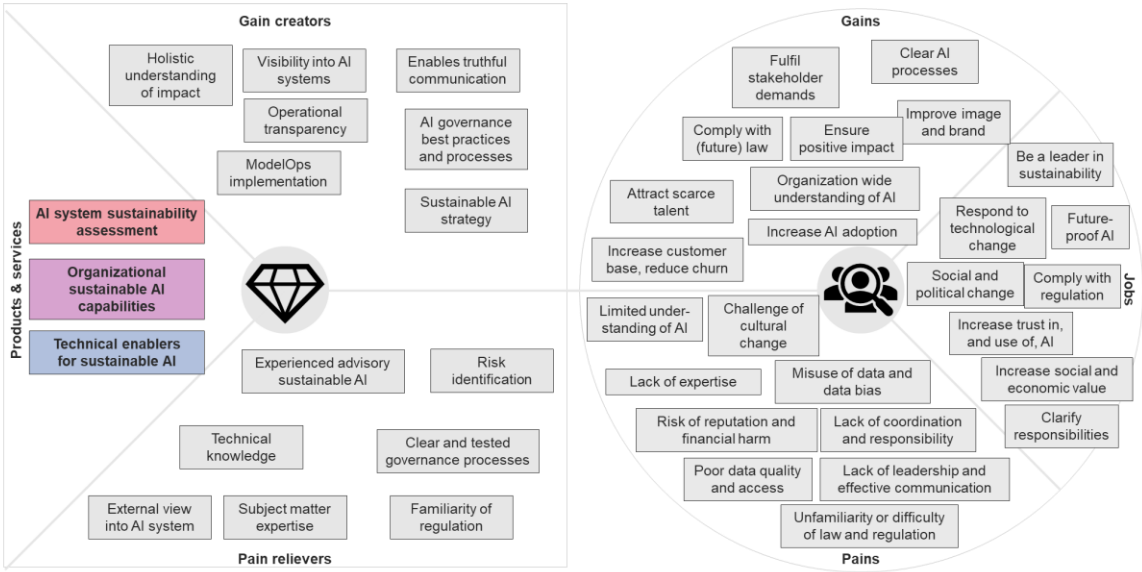


Figure 11. *Clients' needs captured using the VPC.*

The central value proposition includes gaining competitive edge through proven positive impact and value alignment of AI, improving quality and efficiency through AI operational excellence, and enhancing AI risk management. Clients' jobs, pains and desired gains; the designed offering; and the value proposition will be discussed next.

5.1.1 Target clients

In general, private and public organizations developing, using, or distributing AI were recognized as potential clients for the sustainable AI offering. The target client segments were intentionally left broad in the business model. Even though Solita has its strategic domains, it does not shy away from specific industries. Solita's existing approach to clients is adopted in the sustainable AI business model, therefore any AI system operating organization aligned with Solita's values is a potential client.

Within target client organizations, key roles were defined as the main beneficiaries of the designed offering. On a high level, business and senior management benefit most from the offering in taking a top-down approach to addressing the sustainability of AI. Moreover, sustainability managers, product development, and legal personnel benefit from the offering since holistic sustainability, ethical design principles, and legal compliance fall into their responsibility area. Furthermore, IT and quality management value especially the technical dimensions of the offering, e.g., sustainable AI operating model, MLOps implementations, and technical system assessments. These roles were recognized in the business model design sessions.

Based on Solita's client interviews (Metsäranta & Rauhala, 2022), being a leader in sustainability, responding to change in the technological and social environment, as well as ensuring the possibility to use AI systems emerged as the most important jobs client organizations pursued to achieve. Many organizations expressed willingness to be seen as frontrunners in sustainability in general and had noticed their efforts regarding sustainable AI specifically to be lacking. Also, incorporating sustainability into IT strategy was a goal that had been found to be challenging. Furthermore, rapid technological advancements, as well as changes in the social and political landscape put pressure on addressing sustainability issues of AI systems. Moreover, future-proofing AI systems by being proactive and anticipating further change rather than merely reacting to it was found to drive sustainable AI efforts. Also, future-proofing AI systems with regards to forth-coming regulations was of concern. For example, an interviewed client stated a desire to "stay ahead of the curve", as they believed strict regulation was coming in the near future. Another highlighted the inherent desire to be an example in sustainability since the company consults its own customers in holistic sustainability, while many clients stated this desire explicitly. As phrased by an interviewee:

> "*As we want to be a frontrunner [in sustainable AI], we should ask out loud, if someone recognizes a threat so that they can be looked into*."

Though ensuring the possibility to use AI by preparing for and anticipating future regulations came up as jobs to be done, the AI Act and its implications were not familiar to all organizations. Most clients had acknowledged the proposal but had not begun assessing its implications on their business. A few clients had not even heard about the AI Act at all. However, organizations working in already highly regulated industries, like financial services or health care, had analyzed the AI Act and mapped its implications to their own context, while others had familiarized themselves with the proposal and discussed it with external experts. Only a few clients reported to be actively preparing for the AI act. Nevertheless, assuring compliance with current and future law came up as a job to be done.

Furthermore, jobs to be done by client organizations included increasing general trust in AI. As an example, a client organization had begun an AI ethics initiative aiming at increasing trust in their respective sector in general. Also, clarifying responsibilities and processes around AI systems was desired by multiple clients. Indeed, several clients interviewed reported the ethical considerations of AI to be highly personified and in lack of determined processes or roles. Additionally, multiple clients admitted that even though their organizations have determined sustainable AI principles, there was still work to be done in incorporating them into practice.

By getting the recognized jobs done, organizations desire to fulfill the demands of stakeholders and society at large. Clients expect to guarantee positive and reduce negative impact, ensure compliance with current and future law, and improve brand image. Consequently, for example, employee retention and attracting new talent was hoped to become easier. This potential for improving employer image has been seen by clients, while others could not verify the effects but believed the enhanced brand image would attract talent. As a client stated:

> "*Responsible AI is more than a hygiene factor. … I don't know if it is visible externally, but it definitely impacts employee experience. It might later also affect the jobseeker experience*."

Moreover, the customer base is anticipated to increase, eventually driving revenues up. Other gains include gaining an organization wide trust and understanding of AI, which again would allow for increased utilizations of AI systems. Also, clarifying processes and responsibilities around AI systems was expected to reduce operational costs and streamline development and decision making.

Furthermore, pains, i.e., obstacles or challenges in getting these recognized jobs done, were also acknowledged. People and culture, data, and regulatory issues were identified as key challenges. People and culture issues include leadership challenges, lack of resources and expertise, and challenges of cultural change. Lack of established leadership control or commitment was seen as slowing down efforts regarding sustainable AI.

Furthermore, individuals working with AI systems in less AI oriented organizations found convincing leadership of the business benefits of sustainable AI to be challenging. A lack of understanding of AI among people in leadership positions, as well as general lack of AI expertise also played a role in preventing change. For example, a client explained they found it hard to raise AI sustainability issues into discussion, when management in charge did not understand much about AI, while another saw it challenging to convince senior management to invest in the sustainability of AI before it becomes a problem.

Most of the client organizations recognized a lack of expertise and resources as a major challenge. For example, an organization had several teams working on AI, but they still lacked resources to dedicate to investigating and ensuring the sustainability of their AI systems. Another organization was reported to be constantly recruiting but finding experts in sustainable AI had been discovered to be challenging, while a third did not have enough resources for maintaining and ensuring the desired impact of its operational AI systems. Gaining the expertise through organizational partnerships was brought up as a solution in multiple instances.

Moreover, finding the right level and approach to communicating difficult issues was found to potentially slow down sustainable AI adoption. In essence, transitioning towards an organization cherishing the sustainability of AI is a cultural issue. For example, a client worried that tight project schedules did not leave space for the discussion around ethics of the AI systems. As stated by another client emphasizing the need for cultural change:

> "*If we aspire to go beyond risk management, culture is key. It's … about empowering people to ask the right questions at the right time*."

Apart from people related pains, data issues, such as its quality and controlling access to it were recognized as obstacles. Access to biased data in the first place caused issues. Indeed, a client admitted it was too costly to have verified experts validate data, which limited the potential use cases of AI systems. Furthermore, some clients recognized the issue of limited data access in situations, where their customers requested AI systems to be produced without access to the underlying data due to the sensitivity of the data. Ensuring transparency and sustainability in such situations would be impossible.

Moreover, challenges include law and regulation interpretation. Understanding the requirements set by the AI Act was particularly challenging. However, the vague terminology used in the AI Act is expected to get clearer as the proposal gets refined. A limited understanding of law and regulations and not being fully confident on what is allowed was also found to possibly interfere with innovation.

Finally, misuse of data and data bias were recognized as risks needing to be mitigated. Without addressing these risks, the AI systems used by client organizations were feared to cause social harm to their customers. Moreover, the organizations allowing such harm to be done risks damaging their reputation, eventually resulting in financial losses.

5.1.2 Sustainable AI offering

The offering was created based on identified client needs and challenges as well as analyzing Solita's resources and projects together with other Solita employees during design sessions. It was designed to solve the clients' problems while utilizing Solita's capabilities. Three offering categories (AI system sustainability assessment, organizational sustainable AI capabilities, and technical enablers for sustainable AI) emerged as a result of the design.

To help describe the nature of the services, a framework was constructed. In the framework, the services are arranged along two dimensions depending on their technicality and explorative nature. The level of technicality refers to the technicality of the interaction with the client. Activities where models are developed or quantitative data are analyzed hand-on, are considered technical, while, for example, strategic consulting or sustainable AI awareness training are deemed non-technical. The former requires thoroughly investigating IT systems and technical expertise in computer science and related fields, while the latter revolves around higher-level concepts.

Along the other dimension, the exploration-implementation axis, services are distinguished by the level of execution done by the service provider. On the most exploratory end are services such as sustainable AI use-case identification and model assessments since these services do not include implementing recommendations or modifying client systems or processes. These exploratory services are diagnostic in nature, with recommendations or reports being the result of the service. On the other end of the range are services such as MLOps pipeline development, which requires modifying IT systems of clients, and AI operating model introduction, consisting of driving the process and cultural change in the client organization.

Figure 12 depicts the offering framework, with the different offering categories placed into their respective positions. It should be noted that there is overlap between the service categories presented. Moreover, the positioning of the organizational sustainable AI capabilities category block is intentionally detached from the far end of the implementation axis. This is to illustrate the more concrete nature of the technical services compared to the advisory and consultative services.

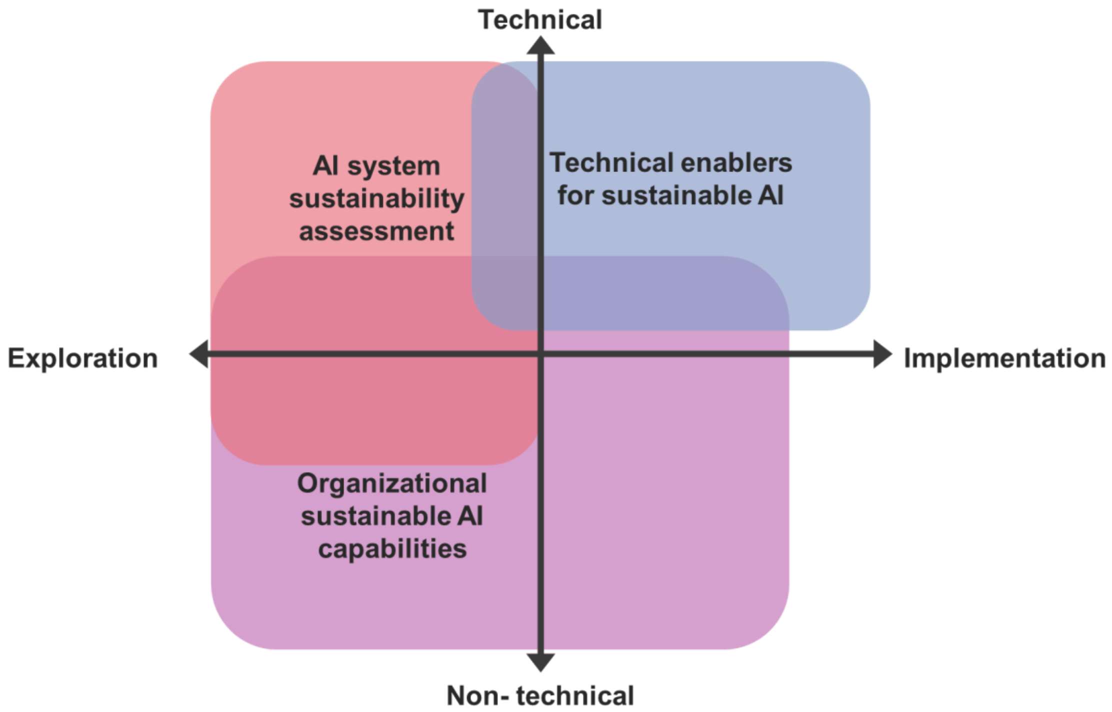


Figure 12: *Sustainable AI service offering framework.*

Next, each category will be investigated in more detail. **AI system sustainability assessment** and AI system assessment in general, is an essential tool to verify AI systems' claims, for example, about fairness and transparency. Indeed, it is believed to be a vital part of being standardized and ensuring regulatory compliance. This service involves investigating the AI system of a client in varying levels of technicality, using methods such as data quality analysis, code reviews and tests, as well as interviews. Also, the processes, documentation, and responsibilities around the AI system are subject to assessment. Typically, as a result, a report of the state of the AI systems is delivered to the client, either for internal or external use. While aiming to be objective and transparent, these services are done in close collaboration with the client.

The most technical assessments are those of algorithms and data, in which sustainability of the system is evaluated through the used models, data, and integration into other systems. Ethics-based auditing, on the other hand, allows for validating claims made about AI systems, verifying their ethical commitments, and proactively showing compliance to regulation. As an example, a transparency assessment aims to holistically assess the transparency of the algorithmic decision-making process, the recommendations of the AI systems, their disclosing, and the openness of the organization providing the AI system aided service or product.

According to Solita's study (Metsäranta & Rauhala, 2022) holistic AI system impact assessment was found to be a challenge even for the most mature organizations. Since the unintended consequences of AI systems can be propagated faster and be more massive, AI systems' impact should be assessed through potential risks to individuals, the society, and the environment. Moreover, impact assessment contributes to risk management and transparency, as it allows potential harms to be recognized ex ante. Indeed, impact assessment was recognized in the design sessions to be a vital component of risk management as it fundamentally includes risk identification. Moreover, impact assessment is an operational service, which requires both technical and conceptual analysis of the organization and AI use cases.

The next offering category is **organizational sustainable AI capabilities**. These services are aimed at improving the AI related capabilities and understanding of the client organization or at providing the capabilities externally. Its less technical nature distinguishes it from AI system sustainability assessment, as the AI systems themselves are not in central focus, but rather the awareness, processes, and structures around AI systems. For example, compliance assessment and ensuring conformity as well as impact assessment and risk identification are services which are located in the overlapping area of these two categories. They require AI systems to be assessed and result in an improved ability of the client to address AI sustainability issues down the line.

In essence, the services in this category aim to build the capabilities of the client organization or to provide expertise as an external consultant. On the explorative end are general AI training and sustainable AI awareness training which are key services in this category. Indeed, education and training are vital since a lack of AI understanding outside AI units was mentioned multiple times in the interviews with clients as an obstacle for achieving sustainability of AI. Furthermore, use case and investment opportunity identification as well as AI operating model definition are examples of these services. Through these services, the capabilities of the client are improved.

Furthermore, towards the implementation end of the spectrum are services such as SDO issued AI standards consultation and implementation, AI governance or operating model definition, and incorporating AI sustainability into data and other strategies. Moreover, aiding in the implementation of, for example, the designed AI operating model or providing project management office services during the change projects are services in this category. These services on the implementation side of the category consist mainly of providing expertise and knowledge to the client organizations as external advisors.

The organizational sustainable AI capabilities services can be further divided into strategic, tactical, and operational points-of-view. On the strategic level, AI sustainability in data strategy, as well as defining sustainable AI principles and guide rails are covered.

By providing experience and knowledge of shaping organizations and principles, a top-down view on sustainable AI can be taken by the client organizations' senior executives. These services are conceptual in nature, and recommendations are the most typical result.

AI governance model design and standardization consultation are important tactical services. Also, applying impact and risk management frameworks as well as sustainable AI awareness and general AI education are classified as tactical services in the organizational AI capabilities category.

Finally, on the operational level there are AI system assessment, impact assessment and ModelOps consulting. Additionally, assisting in adopting AI operating models, i.e., the people, processes, and technology around AI systems, is an essential operational level service. Driving the change brought forth by introducing a new or modified operating model into the day-to-day operations of the client organization is crucial, and such change management projects can be lengthy. Assisting a project management office group is an example of a service recognized as operational services.

Finally, the third category, **technical enablers for sustainable AI,** refers to creating or helping clients create scalable and sustainable AI systems with the help of technical tools. Practically, ModelOps (Artificial intelligence model operationalization), the convergence of AI model development and operations, covers most of this offering. ModelOps, as defined by Gartner (2022), is "a set of capabilities that primarily focuses governance and the full lifecycle management of all AI and decision models".

More granular examples of services in the technical enablers for sustainable AI category include developing AI solutions sustainable by design and MLOps development. AI system development is the most technical service and overlaps with Solita's data science work. Here, an AI based solution, which is sustainable by design, is developed to solve a client need, and delivered to the client. Another technical service is MLOps pipeline development and enhancement. MLOps, a subset of ModelOps, consists of a set of practices and technical solutions aimed at managing the lifecycle of ML models efficiently and reliably. This includes integrating development, deployment, operating, and monitoring of AI systems by establishing processes and using technical tools and automation.

To sum it up, figure 13 shows some of these services placed in the sustainable AI offering framework. These more concrete services can be placed in multiple categories, depending on the detailed nature of the service. This overlapping is presented in the figure by the physical overlap of the objects. The figure does not aim to include all possible services nor all possible, present, or past projects of Solita, but rather bring more context and further elaborate on the categorizations of the offering.

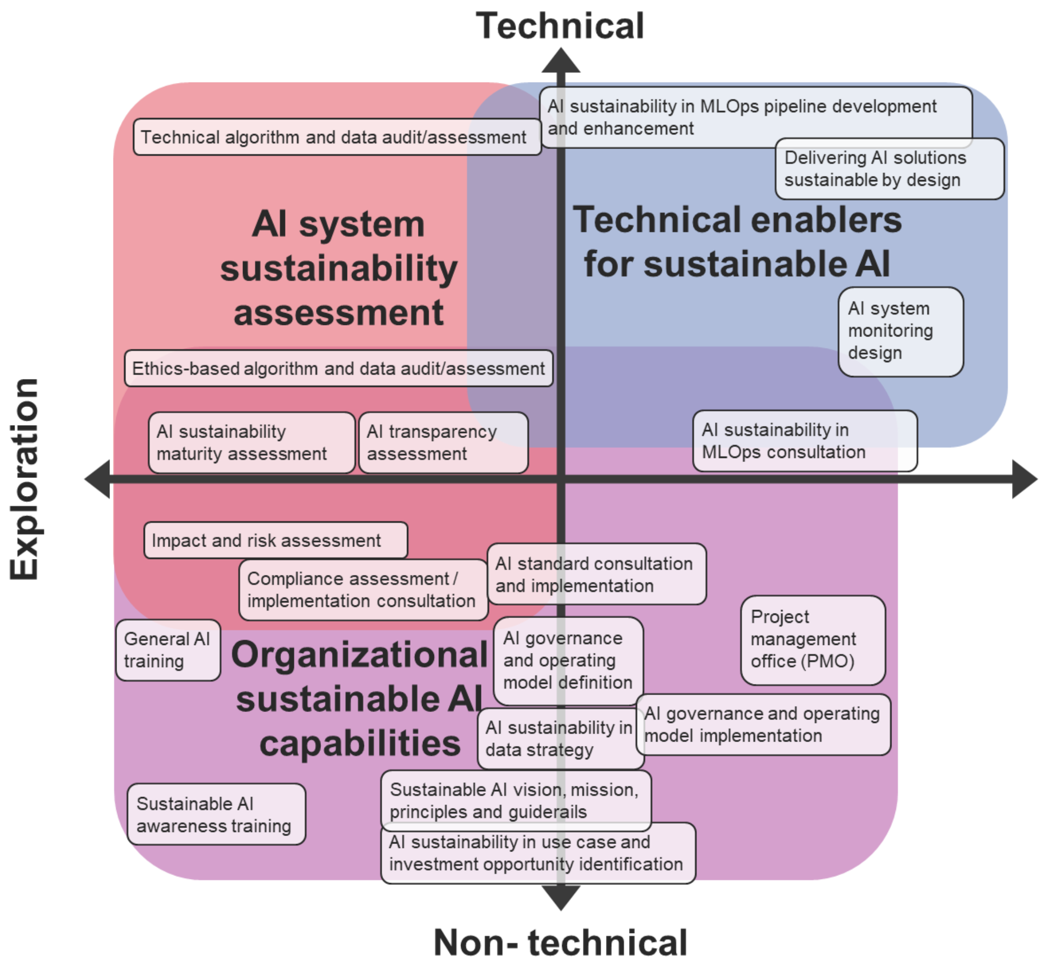


Figure 13: *Sustainable AI service offering framework, with a more granular view of the offering with concrete examples mapped in the technicality and implementation dimensions.*

Many of the highlighted services can be delivered individually, while some must be embedded in other projects. For example, impact assessment, algorithm and data auditing and AI governance model design can be stand-alone projects. However, for instance, AI sustainability in strategy and ModelOps initiatives are not reasonable standalone projects but should rather be incorporated into other services. Such as enabling the sustainability of AI should be taken into consideration in ModelOps pipelines, rather than building a pipeline solely for sustainability goals.

5.1.3 Value proposition

The value of the offering comes from its ability to relieve pains of clients while simultaneously creating gains. In essence, the value created by the offering comes from its ability to enhance AI risk management and improving quality and efficiency through AI operational excellence, which in turn contribute to an organization's ability to ensure and prove the positive impact of its AI systems, resulting in competitive advantage. Eventually, financial gains are expected. The value proposition was a result of the design and feedback sessions conducted with Solita employees.

AI system sustainability assessment and improving organizational sustainable AI capabilities aim at providing a holistic understanding of the impact the client organizations' AI systems have on people, the society, and the environment, and providing the client organizations with means to address these issues. Essentially, this allows for **enhanced AI risk management**. Furthermore, the value is realized through the enhanced ability to recognize pitfalls in an early stage, assure compliance with current and future law, and avoid reputational damage brought by faulty systems.

Gaining a thorough understanding of the AI systems at an early stage allows for better resource allocation, since ideas too risky to be implemented or not beneficious enough can be spotted before resources have been wasted. If any such undesired impacts or faults are spotted later, resources will be lost since such faulty systems are likely to be discarded or large investments are required to fix them. Consequently, the risk of funneling limited resources to initiatives not creating enough value can be avoided. Moreover, early impact assessment allows for tackling problems holistically, once the social, environmental, and economic impacts have been examined. Early-stage impact assessment, as the term suggests, is performed in the early phases of an AI systems lifecycle, before deployment.

For systems already in production, ensuring compliance with law falls into the category of value gained through risk management by enabling a holistic understanding of the AI systems. Ensuring compliance with current law (e.g., GDPR and criminal law) as well as proactively preparing for future law (e.g., the AI Act) reduces the risk of being fined or requiring large investments later on to ensure adherence to these laws. Furthermore, thorough AI system assessments are essential should an organization wish to have its AI systems certified with SDO issued standards.

Production level systems should also be assessed for potential unrealized lingering risks, for example, those of discrimination or exploiting vulnerabilities. If such risks are realized, the reputational damage or legal issues it might cause may turn out to incur massive financial losses. Being publicly accused of causing social harm or excessive environmental damage is certain to cause financial harm and damage to the brand, which could be hard and costly to restore.

All in all, a holistic understanding of the AI systems and their impact allows for enhanced AI risk management and mitigation of the negative effects before they take place. This, in turn, contributes to creating social value, in allowing for action to be taken towards achieving fairness and equal treatment, and ensuring the impact of AI is aligned more with organizational values. Moreover, it allows for taking the holistic environmental and economic impact into account. Additionally, by avoiding fines or reputational risk, clients can avoid financial losses. As stated by a business lead in a design session:

> "*Basically, the value comes from risk management. … Cost savings are bound to occur if you tackle problems early enough. There are plenty of examples of the potential negative, and costly effects of not doing so*."

On the other hand, organizational sustainable AI capabilities services and services for technical enablers for sustainable AI contribute directly to improving quality and efficiency through **AI operational excellence** in the client organizations. AI operational excellence refers to principles and tools adopted by an organization aimed at creating a mindset and culture of continuously improving operations towards sustainability of AI systems. It encompasses technical aspects, such as ModelOps tools, code transparency, and appropriate documentation, as well as ethical principles, AI operating model, and AI governance model.

Adopting ModelOps practices were found to allow enhanced operations efficiency. Automating deployment pipelines and utilizing technical tools in monitoring are to result in meaningful time savings by reducing the need for duplicate work. Also, gaining visibility into the decision-making process of the algorithms through technical tools such as attention mechanisms in ML models and efficient logging improve monitoring capabilities. Moreover, the potential of MLOps tools helping in documentation was recognized by clients. Furthermore, monitoring the energy consumption of AI systems using technical tools allows for making more conscious environmental decisions.

Defining clear processes and assigning accountability will likely further improve productivity. Setting up AI governance structures and an AI operating model forces organizations to determine clear processes and responsibilities. In fact, defining accountability of AI systems was believed to result in more mindful decisions regarding the use of such complex systems. This in turn speeds up decision making and allows for addressing emerging issues faster, since responsibilities have been determined beforehand, limiting the need for ad hoc problem solving.

Indeed, an AI operating model, as well as adopting ModelOps practices were expected to increase visibility into AI systems, which eases addressing issues and improves auditability. Furthermore, some industries or applications must have such processes and practices in use to comply with law. In fact, in some highly regulated fields, the requirements for logging and monitoring are already high and expected to grow with the AI Act. MLOps was believed to aid in adhering to these requirements. Proactively addressing these issues improves the adaptability of the systems to new regulation. Moreover, an increased visibility and determined responsibilities improves the ability to react to changes in the social and regulatory sphere. This is also essential for increasing transparency.

Also, best practices for reducing environmental impact can be enforced through AI governance. These include, for example, deciding which data center locations to use for training and how much computation should be used for the optimization of certain AI systems. Indeed, a client explained that they had processes in place to determine which use cases of AI and data collection efforts were worthwhile from a holistic environmental impact aspect, contributing to environmental sustainability.

In essence, AI operational excellence is believed to streamline lifecycle management, reduce costs of operations, improve quality, and to increase cooperation. All these result in economic value for the organizations. As explained during a feedback session by a Solita senior expert:

> "*AI operational excellence means that we're able to do things faster or cheaper, with higher quality, with less manual work. But what justifies the investment? This might come down to euros saved, or more euros made as we can utilize AI faster and more reliably*."

With the ability to manage and mitigate AI risks and with the structures, responsibilities, tools, and processes in place to streamline operations, an organization can ensure that the impact of its AI systems is positive and aligned with its values. **By being able to prove the positive impact, competitive advantage is achieved** through improved customer and employee satisfaction, as well as a pleased general public.

By taking the potential social issues into account and addressing them can increase customer base. Loyalty of existing customers is expected to increase once the positive impact can be proven. Moreover, ensuring that AI systems are fair and non-discriminating towards all groups of people is bound to result in a better experience for all diverse user groups. Furthermore, this can improve the brand image of the organizations, improving its brand value. Also, taking sustainability into consideration holistically is expected to be valued by some if not most customers. Moreover, this is believed to help attract scarce talent in the highly competitive recruiting landscape. Additionally with the rising popularity of socially responsible investing, investors are becoming more willing to invest in socially and environmentally conscious organizations, increasing funding opportunities.

Proving the positive impact was also found to be meaningful internally. Ensuring that the efforts made by employees contribute to social and environmental good can increase loyalty and satisfaction. Consequently, retention can improve, and the wellbeing of workers increases. As recalled by a researcher who conducted the original client interviews:

> "*Feeling proud [of working for your company] was a big thing that came up in the state of sustainable AI interviews [conducted by Solita*]."

However, the extent to which economic value is the only important value for companies was considered in a feedback session. As discussed by two experienced researchers in a feedback session:

> "*If we're talking about businesses, in the end, there's only one kind of value standing and that's [economic] business value. … That is such a high-level value that only in itself it's not interesting. But what's interesting is … how this fundamental economic value is achieved? What kind of tools there are? This relates to better risk management, avoiding costs, and operational excellence*."

> "*There is [based on the State of sustainable AI study's client interviews] inherent value in risk management. For example, there is value in wanting to do the best to minimize harm and that does not necessarily have to be quantified into financial terms. Operational excellence is, however, typically driven by monetary pressure. There can be other economic benefits than driving down costs, [for example] improved ability to adopt AI, which could increase the [economic] benefit [eventually]*."

To conclude, the services are expected to enhance AI risk management and AI operational excellence as well as increase brand value, resulting in competitive advantage. The client organizations are expected to gain social, environmental, and economic value in different forms. What value each service eventually creates is depicted in figure 14.

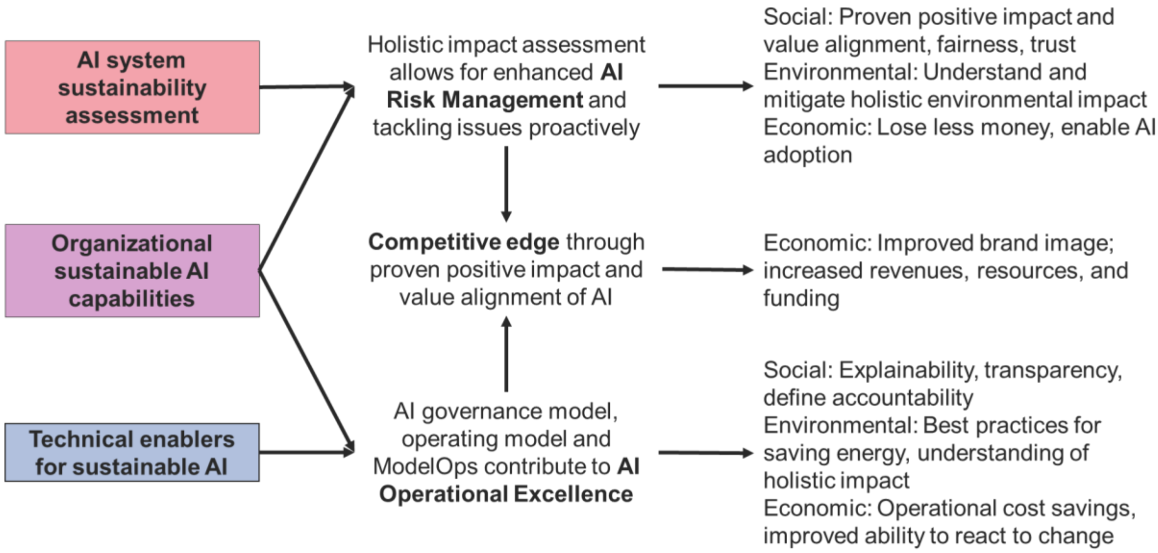


Figure 14. *Value created by each offering category.*

To conclude, all AI system operators are included as potential clients of the designed offering. AI system sustainability assessment, organizational sustainable AI capabilities, and technical enablers for sustainable AI are the categories of the offering. These services create value by enhancing AI risk management and improving AI operational excellence resulting in creating competitive advantage through ethical and sustainable AI as a result of proving positive impact and value alignment of AI systems. This is expected to result in social and economic value, as well as the possibility to create positive, and mitigate negative, environmental impact.

5.2 Client relationships and channels

The mapped services all require a high level of customization depending on the needs of a particular client organization. No client organization nor project is exactly similar to another. Moreover, the nature of all the services is highly collaborative. While taking an advisory role, close collaboration with the client is required to deliver the promised value. Also, the perception clients have is important. Trust is an essential part of being in an advisory role. As phrased in a design session concerning the client relationship in services requiring design work:

> "*In this kind of design work, it is important to create a feeling that the expert knows their thing and can guide the project sturdily over the finish line*."

AI system sustainability assessment and technical enablers for sustainable AI services can be more independent but cannot be done in silos. Especially in the beginning of a service relationship, defining the goals must be done in close collaboration with the client.

Also, since the subject of these services is an existing client system, help in getting to know the tools and processes used is essential.

Furthermore, all categories of the conceptualized service offering are delivered to the clients through close interactions, such as workshops, analysis sessions, development, and training. These value delivery channels are described to be typical in IT consulting businesses. Also, project lengths can vary from short, fixed term definition projects to long lasting partnerships. However, most projects have a clear beginning and end, making the services fixed term consulting. Nevertheless, the aim would be to create a long-lasting advisory relationship with the client. As stated in a design session:

> "*We would generally want to forge long-lasting partnerships, rather than deliver one-off projects*."

Clients for the sustainable AI services are reached in the first place through marketing efforts and existing client relationships. Attending conferences, being present in social and traditional media as well as inbound marketing (blogs, limited guides, online assessment tools) were recognized as vital channels for building awareness. However, promoting sustainable AI services to existing clients, and getting referrals by appreciated experts are believed to be the main source of demand. Moreover, incorporating sustainable AI services into any other client projects is a channel to deliver the outlined value proposition:

> "*It is important to have experts trusted by clients promote and explain the value of sustainable AI. Current [employees] in related projects should try to promote additional services and projects, … if relevant for the client's situation*."

Furthermore, it is evident through various initiatives taken, that Solita wishes to act as a thought leader in sustainable AI, promoting the subject and emphasizing its importance. Taking part in public discourse, not only spreads concepts and ideas valued by Solita, but could also result in increased business.

5.3 Key resources, activities, and partners

Key resources needed to deliver the offering consist mostly of diverse experts. Skills possessed by data scientists and engineers, ModelOps specialists, legal personnel, consultants, software developers, human insight specialists, business designers and project leaders are essential in delivering the value proposition. Other resources include the brand and the trustworthiness of Solita. Additionally, proven and tested frameworks and process descriptions can be considered as key resources enabling an effective delivery of the value proposition. For example, after a sustainability assessment has been conducted, a similar framework can be utilized in a subsequent AI system assessment. The importance of such frameworks was recognized in the design sessions and are considered valuable resources once they have been validated and refined.

Activities carried out by these experts include data and model analysis and review; analyzing the people, processes, and technologies of the client; training; implementation of MLOps solutions; continuously researching the sustainable AI ecosystem; and interpreting laws and regulation in specific contexts. Also, an essential activity is designing and helping implement AI governance and operating models.

Most of these activities can be done by Solita employees. However, some key partners are needed to perform activities needed to deliver the value proposition. For example, legal expertise and regulatory knowledge could be more suitable to be acquired from key partners, compared to developing such capabilities internally. As explained during a design session:

> "*Even though it is needed by customers, I don't believe we'd want to invest in hiring more layers … to provide legal advice. We'll leave that to the legal specialist firms*."

National level authorities proposed by the AI Act are also vital in interpreting regulation in specific situations. However, such authorities are yet to be established. Moreover, in some cases, implementation or maintenance of designed solutions can be assigned to a third-party IT or software company, making them key partners.

Other key partners include researchers and academia and standard developing organizations. These partners are important in research and development. Researching the development of the AI governance ecosystem, social trends, and technical solutions, as well as developing standards and AI governance frameworks is valuable in adapting the business model to the changing environment. Despite key partners being recognized, the exact role of them is still unclear. However, their role is expected to increase as the AI governance ecosystem matures.

5.4 Cost structure and revenue streams

Finally, the financial liability of the business model can be determined. Once key resources, activities and partners have been identified, it is possible to specify the inherent cost structure. Moreover, with the offering, client segments and client relationships identified, the revenues streams and pricing strategies can be elaborated on.

The business model designed is value-driven rather than cost-driven. Therefore, employee salaries generate most costs. When it comes to the characteristic costs of an expert service consulting business model, most costs are fixed salaries paid to employees, including consultants and technical experts.

Moreover, overhead, such as internal IT, human resource management, office costs, etc., increase the cost borne by the consulting company. These overhead costs are often estimated at 1.6 times the gross salary being paid to employees. These salary costs accrue up to 90% of total costs in the business model. These figures came up as examples or as "rules of thumb" by distinct participants in the design sessions. As stated by an experienced business designer during a design session:

> "*The cost structure [in the sustainable AI consulting business model] is that of basic consulting, i.e., primary, if not all costs occur from salaries to consultants, developers, or designers. The salaries paid, plus overhead from IT, offices, finances, HR etc*."

Furthermore, sales and marketing needed to generate demand creates costs. Time spent preparing an offering proposal, as well as time used in preparing a conference speech are included in this category. Moreover, research and development costs incurred add to the overall cost of the business model. To keep employee skills up to date, investments in research must be made. Also, business development and research into market needs and trends must be performed to stay up to date with the changing environment. Other minor cost factors include third-party tools and frameworks, costs of licenses for software, and payments for certification standards. However, these costs are negligible compared to the above-mentioned.

Pricing such an offering is dynamic. Value created is often intangible, thus hard, or even impossible to be measured in absolute terms. Similar activities can yield different amounts of value depending on the situation of the client company. Pricing should be above costs and wanted margin, but less than the perceived value of the offering.

Two major revenue models were identified: hour-based and fixed cost. Hour-based billing (and day- or week-based) is a common practice in consulting business. Clients are billed for the time experts work on their problem in real time. On the other hand, in the fixed cost model, a price is estimated at the beginning of the project, and as much work as needed is performed to fill the requirements. Such a model imposes the risk to the consulting company. If no complications arise, great margins can be made. However, if the task turns out to be more challenging, the consulting company must lower its margins. In both cases, the client can be expected to pay for the value it is estimated to be receiving.

5.5 Summary

The business model canvas presented in the beginning of this chapter in figure 10 captures all essential aspects of Solita's sustainable AI business model. The value proposition of enhanced risk management, operational excellence, and improved brand image is delivered through AI system sustainability assessment, organizational sustainable AI capabilities, and technical enablers for sustainable AI services.

The client interface, required infrastructure, and the finance are similar to other IT and strategy consulting business models. As most building blocks of the designed BMC do not differ from what was in place already, the result is a diversified business model rather than a new one.

6. DISCUSSION

This chapter highlights the most important findings of the study. In the first sub sections, the aim is to generalize the findings and project them to those presented in prior literature. Findings related to the business model as well as the state of sustainable AI in general are presented.

Next, the contribution of the thesis is discussed, and the implications of the thesis for different stakeholders are considered. Finally, limitations are uncovered, and future research suggestions are provided.

6.1 Key findings

6.1.1 Findings on the business model of sustainable AI consulting

Mechanisms presented in prior literature for achieving sustainable AI (see Butcher & Beridze, 2019; Seppälä, et al., 2021; Floridi, et al., 2022) were generally in line with the offering designed in this study. Jobs clients seek to get done were found to be answered with different forms of organizational AI governance, further divided and explained in the previous chapter (see figure 11 in section 5.1.3).

As stated by Mäntymäki et al. (2022a) achieving sustainability of AI requires cooperation between multiple actors, since not all organizations have the resources or capabilities needed to address emerging issues themselves. This was empirically validated through the client interview analysis, as a majority of clients expressed a lack of resources or capabilities to ensure the sustainability of AI. Indeed, the interviewed clients called for assistance in various forms, similar to those which Seppälä et al. (2021) had previously recognized organizations to be implementing internally. Moreover, services designed in this study were in line with activities provided by private companies which Butcher and Beridze (2019) had recognized.

What’s more, the services outlined in the thesis have been proven to have a product-market fit, i.e., the value proposition has gained attraction in the market. Indeed, the categories were formed based on either past projects or needs of clients, proving there is real interest towards them. Figure 15 highlights the essential business model blocks associated with the finding, which are the value proposition and client segments.

Finding 1: *Interorganizational services needed by AI operators provided by private consultancies enabling sustainable AI include AI system sustainability assessment, organizational sustainable AI capabilities, and technical enablers for sustainable AI services*.

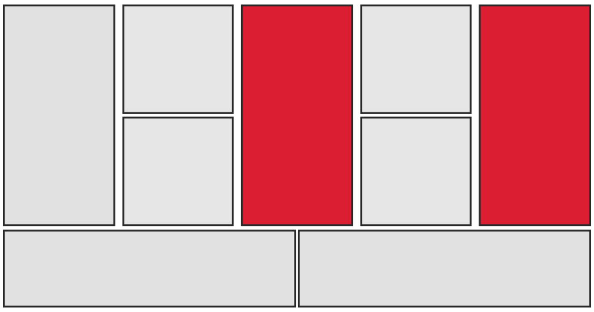


Figure 15: *AI operators' (i.e., clients') needs of services are satisfied by the designed offering.*

Furthermore, by analyzing the nature of these services as well as the key activities and key resources needed to deliver them, the sustainable AI service offering framework was developed (figure 11 in section 5.1.3). The dimensions of the framework emerged from combining theoretical structures presented in literature with data gathered on the experiences of employees of the case company. Indeed, as emphasized by the AI HLEG (2019), both technical and non-technical AI governance methods encompassing all stages of AI systems' lifecycle are needed for achieving sustainable AI development and use. The varying dimensionality of AI governance activities has also been recognized by multiple researchers in their AI governance frameworks (see Chhillar & Aguilera, 2022; Mäntymäki, et al., 2022b; Laato, et al., 2022b; Brendel, et al., 2021; Shneiderman, 2020).

The dimensions of "implementation – exploration" and "technical – non-technical" provide a framework for guiding discussion around the role of consultancies in the AI governance ecosystem. These dimensions capture the essential varying aspects, even though they are not the only possibilities. For example, the positioning of the various services in relation to the lifecycle of the AI system is not explicitly addressed by the presented dimensions. However, the lifecycle component is inherently embedded in the "technical – non-technical" dimensions to some extent, as the technical services become relevant only as an AI system is about to be deployed.

Furthermore, the varying nature of the services has implications for the business model. Services far apart from each other require vastly different types of expertise and activities. For instance, developing or enhancing MLOps pipelines requires deep hands-on technical know-how, while Sustainable AI awareness training requires a holistic understanding and a convincing presence. However, human expertise is central to all forms of the designed offering. Moreover, the channel through which the service is delivered varies from providing suggestions in face-to-face meetings all the way to developing software. Figure 16 highlights the essential business model sections, which reflect the varying nature of the service offering designed.

Finding 2: *Sustainable AI service offering framework: The designed sustainable AI services differ in the technicality of the interaction with the client as well as in the degree of exploration versus implementation*.

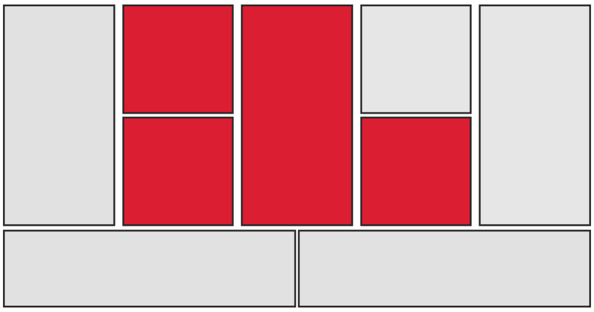


Figure 16: *The varying technicality and explorative nature of the offering requires different types of resources, activities and delivery channels.*

Enhancing AI risk management and AI operational excellence were found to be important practical gains expected or experienced by client organizations. However, business interest was also found to be a driving force for achieving sustainable AI.

Expected benefits can be eventually traced back to improving the profitability of the organization. These findings are in line with the ones reported by Vakkuri et al. (2022), who found that responsibility in the context of AI systems is typically approached from a financial, legislative or customer relations point-of-view, rather than an ethical point-of-view. Also, Floridi et al. (2018) recognized the potential benefits of avoiding risks related to AI systems. Moreover, similar results were discovered by Seppälä (2021), as he found risk management, external pressure as well as corporate social responsibility and business value to be drivers for sustainable AI initiatives in client organizations. The third finding relates strongly to the value proposition in the BMC, which is directed towards the target client. This is highlighted in figure 17.

Finding 3: *The business value of Sustainable AI created to clients stems from enhanced AI risk management and improving AI operational excellence, resulting in competitive advantages through proven positive impact*.

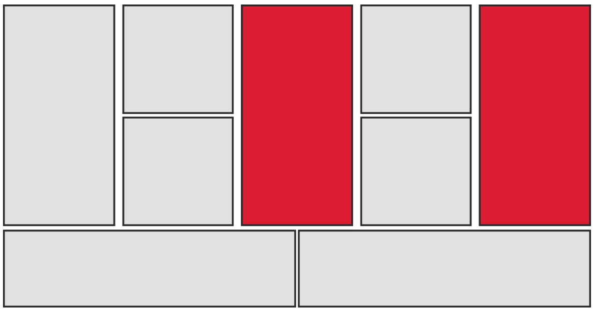


***Figure 17:** The offering is to create value to the identified customers through AI risk management and improving AI operations.*

The business model supporting the offering detailed in chapter 5 shares multiple characteristics with traditional IT and strategy consulting. As discussed in the literature review, Minkkinen et al. (2022b) argued the emerging ecosystem around AI governance has implications for the business models of private companies providing services that address the sustainability issues of AI. However, the implications were not found to be dramatic.

Business model innovation was used according to one of its use cases highlighted in literature, as a tool for diversifying a business model to bring new services to market with the aim to realize on emerging opportunities in an organization's environment (Osterwalder, et al., 2005; Geissdoerfer, et al., 2018). During the process, it became eminent that only the value propositions and offering were clearly altered, while the other aspects of the business model stayed relatively unchanged. Also, resources required to deliver such a value proposition were different. This is driven by the fact that the value driven business model relies heavily on expert knowledge. Figure 18 highlights the modified BMC blocks.

Finding 4: *The business model supporting the sustainable AI offering has most of the characteristics of a traditional IT or strategy consulting business model*.

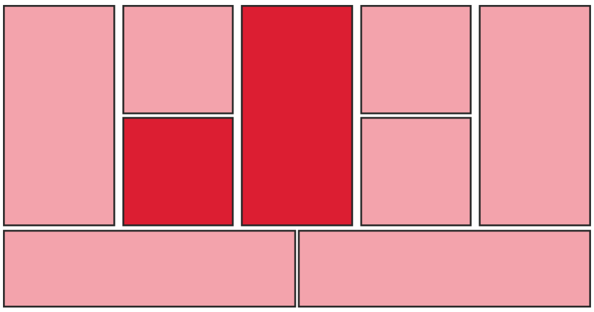


***Figure 18:** The altered value proposition and offering require new types of expertise, while the rest of the business model is left essentially unaltered.*

As can be seen from the business model component highlights, most focus was given to the value proposition and offering within the BMC, since they were recognized early on in the design sessions to be the most vital aspects of the BMC in this case.

Additionally, during the design process, other interesting findings were made not directly related to the business model. These will be discussed in the next subsection.

6.1.2 Findings on the state of Sustainable AI

By analyzing the client interview excerpts and incorporating findings of Solita's state of sustainable AI research (Metsäranta & Rauhala, 2022) a view of the state of sustainable AI and AI governance was formed. Findings relevant in the context of this thesis are presented next.

While the possible negative impact of AI was recognized widely, less than half of the interviewed client companies had AI-specific guidelines or principles, let alone a clear sustainable AI strategy and established processes for it. This is in line with findings presented in prior literature (see Vakkuri, et al., 2022; Floridi, et al., 2022; Mäntymäki, et al., 2022b).

Despite Cihon (2019) arguing that standards and standardization offer a key mechanism for developing AI governance on a global scale, standards have yet to become relevant in achieving sustainable AI. Seppälä (2021) found equivalent results, in that audits and AI standards are not yet implemented by AI operators. This was seen to be due to the nascent state of the standards, and no enforcement for adhering to them. However, during design sessions, the potential importance of standards produced by SDOs was recognized. As more relevant standards are published and the ecosystem around sustainable AI develops, their importance is expected to grow. Moreover, analyzing released standards can provide valuable insight into what aspects of the AI systems should attention be paid to.

Finding 5: *The potential sustainability issues of AI are recognized, yet practical actions taken to mitigate harm are limited*.

Interestingly, while the AI Act was familiar to most interviewed clients, only a handful of organizations had thoroughly assessed it and mapped out its implications. According to Minkkinen et al. (2022a), the AI Act would clarify the rules of AI governance and auditing landscape. However, this has not yet been the case, since both client organizations and the case company are yet to find the rules of AI governance clear. However, organizations already highly regulated (e.g., financial services) were most familiar with the proposal and its implications.

The lack of thorough analysis of the AI Act is to be expected at some level since the proposal has not yet been accepted and modifications to it are expected (see Raposo, 2022; Mökander, et al., 2021; artificialintelligenceact.eu, 2022). Consequently, the AI Act is not the primary driver for implementing AI sustainably at the moment, even though it promotes the ethical issues to be considered and raises awareness.

Though the AI Act will not have implications to all AI systems, it will most likely have an impact in more situations than seems obvious at first glance. AI systems are defined extremely broadly in the AI Act, and essentially all algorithmic systems are in the scope. The use case and the areas of effect of the AI system is the main determinant of the requirements imposed by the AI Act. Hence, thorough impact assessments ought to be needed to verify the extent to which the AI Act applies to certain AI systems.

Finding 6: *The AI Act proposal has not yet caused much action, as only organizations in industries already heavily regulated have actively discussed it*.

Even though in literature a significant importance was given to the environmental impact of sourcing hardware for AI systems and their operational emissions (Strubell, et al., 2019; van Wynsberghe, 2021; Robbins & van Wynsberghe, 2022; Genovesi & Mönig, 2022), clients seldomly associated the environmental impact of AI strongly with sustainable AI. Indeed, in client interviews and observation, the environmental impact of training and operating AI systems was discussed, but social aspects were highlighted much more when talking about sustainable AI.

This seemed to be due to the fact that environmental sustainability was addressed in other sustainability initiatives in most organizations. However, the extent to which organizations pay attention to the environment stressing dimensions of sourcing AI system hardware was not studied in this thesis.

This may also be reflected in the terminology used. The broad terminology around the subject of developing and using AI systems, which do not cause harm and have a positive impact, was visible already in the literature review. The most common terms used in literature are "sustainable" (e.g., van Wynsberghe (2021)), "responsible" (e.g., Dignum (2019)), "ethical" (e.g., Seppälä (2021)), and "trustworthy" (e.g., AI HLEG (2019)). However, different terms used reflect the different emphasis around the subject. Indeed, for example, Seppälä et al. (2021) in using the term "ethical AI" clearly focus on social aspects, while van Wynsberghe (2021) takes a broader view, including the environmental and economic aspects and uses the term "sustainable AI". The choice of terms in academic literature are clearly defined most of the time, and obviously reflects the scope at which the issues around uncontrolled AI are addressed.

In the case of Solita, the term “sustainability AI” is deliberately chosen to reflect the holistic approach taken towards the subject and to be aligned with enterprise communications. Within Solita, the term is defined as “creating and operating trustworthy AI systems that deliver long-term value to business, people, and the environment”.

Interestingly, the clients interviewed seemed to use the above-mentioned terms almost interchangeably. Only in a few cases, the clients had clear definitions in mind for the terms they used. The rest seemed to lack clarity in their choice of terms. However, the reasons for the varying terminology cannot be stated certainly. On the one hand, the choice of terminology used could reflect the nascent state of sustainable AI, in which terminology has yet to be established, and terms are used carelessly. Discussing parties could be referring to nearly the same matter but happen to use differing terminology, due to not having given thought and definitions to the terms used. On the other hand, the varying terminology could be caused by the conscious decision to emphasize particular aspects of the wide field of sustainable AI, similar to the case in literature.

Another possible explanation is that a term is chosen based on the desired state, used in a way as a guiding star or a goal, towards which the organization strives to advance. Multiple intrinsic values could be present in the field, which each term emphasized slightly differently. Indeed, Jobin et al. (2019) recognized multiple principles being important in the ethical AI space. However, as found out in Solita’s study (2022), it was the case that in most interviews the terminology used seemed to be reflecting the thoughts of the individual being interviewed, rather thoughts of the organizations they were representing.

Nevertheless, this study cannot answer the terminology aspects definitely. What was noticed in the discussions is that the terms have enough overlap to allow for different organizations to talk about issues rising from ungoverned use of AI. However, clearly defining what is meant by the used terminology ought to make communication, both internal and external, clearer.

Finding 7: S*ustainable, ethical, responsible and trustworthy are terms used to refer to a desired state of AI. Interpretations of the terms differ and are often not defined by organizations.*

6.1.3 Conclusions of the academic contribution

Table 4 collects the findings of the study and highlights the nature of their contribution, being either new knowledge or empirically validating existing literature. None of the findings made in this thesis contradict existing literature, but rather elaborate and extend on prior literature in important ways.

Table 4: *Empirical findings of the study*

#	Description	Nature of finding
Findings regarding the business model of sustainable AI consulting services		
1	*Interorganizational services needed by AI operators provided by private consultancies enabling sustainable AI include AI system sustainability assessment, organizational sustainable AI capabilities, and technical enablers for sustainable AI services.*	Existing literature empirically validated (see Butcher & Beridze, 2019; Floridi, et al., 2022; Mäntymäki, et al., 2022a; Seppälä, et al., 2021).
2	*Sustainable AI service offering framework: The designed sustainable AI services differ in the technicality of the interaction with the client as well as in the degree of exploration versus implementation.*	New knowledge, extending on prior literature (see AI HLEG, 2019; Brendel, et al., 2021; Chhillar & Aguilera, 2022; Laato, et al., 2022b; Mäntymäki, et al., 2022b; Shneiderman, 2020).
3	*The business value of Sustainable AI created to clients stems from enhanced AI risk management and improving AI operational excellence, resulting in competitive advantages through proven positive impact.*	Existing literature empirically validated (see Floridi, et al., 2018; Seppälä, 2021; Vakkuri, et al., 2022).
4	*The business model supporting the sustainable AI offering has most of the characteristics of a traditional IT or strategy consulting business model.*	New knowledge, extending on prior literature (see Minkkinen, et al., 2022b)

Findings about the state of sustainable AI		
5	*The potential sustainability issues of AI are recognized, yet practical actions taken to mitigate harm are limited*.	Existing literature empirically validated (see Vakkuri, et al., 2022; Floridi, et al., 2022; Mäntymäki, et al., 2022b; Seppälä, et al., 2021).
6	*The AI Act proposal has not yet caused much action, as only organizations in industries already heavily regulated have actively discussed it*.	New knowledge, extending on prior literature (see Minkkinen, et al., 2022a; Mökander, et al., 2021; Raposo, 2022).
7	S*ustainable, ethical, responsible and trustworthy are terms used to refer to a desired state of AI. Interpretations of the terms differ and are often not defined by organizations*.	New knowledge, needs further investigation.

As stated in the introduction of this thesis, the objective of this study is to fill in the gap recognized in prior literature by extending the knowledge around the state of sustainable AI and the AI governance ecosystem, providing insights into the interorganizational activities within the AI governance ecosystem provided by a private consultancy, and by studying business models for sustainable AI offering. The research questions were formulated as follows:

1. *What interorganizational activities are needed from consultancies within the AI governance ecosystem, and what value are the activities expected to create?*
2. *What characteristics does a business model supporting the delivering of sustainable AI offering have?*

Findings 1 and 3 answer the first question directly, while finding 2 further elaborates on the nature of the activities. Furthermore, the second question is answered by finding 4 in particular and findings 1,2 and 3 in general. These findings regarding the business model also provide valuable insights into the state of sustainable AI, which is further expanded on by findings 5, 6 and 7. The implications of the findings and future research directions will be covered in the following subsections.

6.2 Implications

As discussed in the previous chapter, this thesis contributes to the ongoing academic discussion around sustainable AI and AI governance (see Brendel, et al., 2021; Butcher & Beridze, 2019; Chhillar & Aguilera, 2022; Floridi, et al., 2022; Laato, et al., 2022b; Minkkinen, et al., 2022a; Minkkinen, et al., 2022b; Mäntymäki, et al., 2022a; Mäntymäki, et al., 2022b; Mökander, et al., 2021; Raposo, 2022; Seppälä, 2021; Seppälä, et al., 2021; Shneiderman, 2020; Vakkuri, et al., 2022) by providing deeper knowledge on the role of private companies in the AI governance ecosystem. Moreover, insight to the state of sustainable AI was provided, which has been recognized in prior literature to still be in a nascent state.

Overall, this study provides practical insight into the private sector approach to the sustainable AI ecosystem. This practical point-of-view is welcome, since current academic discussion is still highly theoretical (Vakkuri, et al., 2022).

Additionally, this study provides further evidence of the usefulness of business model design, and the utility of the Business Model Canvas (Osterwalder & Pigneur, 2010) and the Value Proposition Canvas (Osterwalder, et al., 2015) in academic research. Indeed, their utility had been recognized also before (see ur Rehmana, et al., 2016; Metallo, et al., 2018; Sort & Nielsen, 2018; Polydoropoulou, et al., 2020), which this thesis confirms.

As practical implications, valuable knowledge is provided for the case company, AI operators, similar consultancies and other players in the AI governance ecosystem.

Importantly, the results of this thesis have practical implications for Solita, in conceptualizing their business model. Action research has implications beyond the research project, as the gained knowledge can be used to inform decisions in an organizational context (Saunders, et al., 2019, p. 204). Indeed, portraying the business model in the famous BMC format allows for clear communications both internally and externally (Osterwalder & Pigneur, 2010; Sort & Nielsen, 2018). Internally, the conceptualized business model can be used to communicate the approach taken by the sustainable AI team. Furthermore, by explicitly laying out the business model and the nature of various offered services, resources and activities needed can be organized to deliver the value proposition. Additionally, the business model can be used to align business strategy with operations (Al-Debei & Avison, 2010). Moreover, the BMC functions as a starting point for constructing service proposals. Externally, the canvas and the mapping of expected value provides a framework utilizable in communicating the value proposition to clients in a structured form.

For AI system operators, this study maps out aspects of sustainable AI, and provides insights into how they can be addressed. It provides a reference point for benchmarking activities addressing AI's sustainability done within an organization. However, the study does not aim to create an exhaustive list of all the aspects of sustainable AI, nor does it claim the presented dimensions of AI governance to be superior to others. Indeed, the objective was not to map out all viable activities and all achievable value, but to conceptualize services which IT consultancies could offer when contributing to the AI governance ecosystem. The ecosystem around sustainable AI and AI governance is still emerging, and new and better methods for addressing the sustainability of AI issues are sure to surface. Nevertheless, the study highlights concrete services and the value they are expected to create. By providing AI operators with a means to govern their AI systems, the potential benefits of AI systems could be realized in a safe way (Morley, et al., 2020). Nevertheless, an environment of public trust is required (Floridi, et al., 2018), which the services designed in this study can support.

What's more, this study compactly gathers the expected business value of sustainable AI, which can be used in various organizations to justify investments. As found during the empirical part of the study, communicating the value of sustainability in AI was challenging for subject matter experts in many organizations. This study can be used as a starting point for discussion about the importance and benefits of investing in the sustainability of AI, as it demonstrates that investments in sustainable AI can increase operational excellence and mitigate AI related risks. Consequently, by investing in AI sustainability, organizations can reap the dual advantage of AI ethics by avoiding misuse (mitigate risk realizations) and underuse (increase AI adoption) of AI systems (Floridi, et al., 2018).

By analyzing the role of private consultancies, actors in the AI governance ecosystem are provided with practical benefits and implications. The contribution towards academia was already extensively discussed, but other actors are to benefit too. For instance, the implications of the AI Act proposal in the AI operator organizations speaks to the extent which the proposal has had impact. Furthermore, as stated by Minkkinen et al. (2022b), organizations must recognize their own part in the multi-actor ecosystem. Legal firms and AI system auditors are provided with knowledge on which services they are needed for, since companies like Solita are not willing to cover all required activities. Moreover, other companies similar to the case company willing to take part in the ecosystem are offered a structured breakdown increasing the understanding of the field and requirements for operating in it.

All in all, this study provides insight into the role of private consultancies within the AI governance ecosystem, in categorizing the services they provide to meet client needs. Moreover, an understanding was created of the value created by investing in sustainable AI. This study also wishes to propagate sustainable AI awareness and hopes to encourage AI system operators to take all aspects of sustainability into consideration with their AI systems.

6.3 Limitations

In this thesis, general characteristics of a sustainable AI business model were provided. Insight from numerous target organizations was used to clarify the client point of view. Furthermore, multiple design sessions were held, interviews conducted, and supporting data analyzed to gain a complete understanding of Solita's view, goals, and capabilities. Nevertheless, there are several limitations associated with the findings and their implications arising from data and methods used as well as the subject of interpretation due to the qualitative nature of the study.

The data used in this study focuses on the sustainable AI landscape in companies working in Finland in this particular time. Had the data been collected in a different location or at a different time, the results might have been different. Moreover, as secondary data sources were utilized, the usefulness and accuracy might be compromised, since the data were originally intended for a different purpose and a different audience (Yin, 2009, p. 106).

Though the client interviews were not conducted for the purpose of this study, they are considered to not pose significant limitations. The client interviews covered a wide range of topics of which a subset is precisely what was relevant for this thesis. Moreover, the researchers who conducted and analyzed the client interviews were consulted throughout the business model development process, to ensure that the inferences drawn for this thesis were consistent with the original perspectives of the clients. They took part in the design sessions and provided feedback during the thesis project.

Looking at the methods used, some limitations can be recognized. According to Eriksson and Kovalainen (2008), case studies have been critiqued as being "anecdotal descriptions, which lack scientific rigor". Indeed, the beliefs and thoughts of individuals in Solita played a major role in the final business model.

Reliability indicates the replicability and consistency of the study, and it addresses how certainly another researcher following the same procedure would arrive at similar findings (Kirk & Miller, 1986). The co-design nature and unstructured interviews might compromise the reliability, since these methods used have high dependence on the individuals designing or being interviewed. However, by explicitly laying out the literature review process as well as the data gathering and analysis processes, the reliability was improved.

Validity, on the other hand, refers to the appropriateness of the measures used, accuracy of the analysis of the results and generalizability of the findings (Kirk & Miller, 1986). Since the offering covered various services with differing levels of technicality and implementation, it was challenging to capture all different requirements in other parts of the business model. Consequently, some distinct requirements of a particular service type might have not been represented. Indeed, the BMC or just the VPC could have been mapped out for each offering categories separately, to allow for all aspects of the services or products to be discovered. However, on the higher-level point-of-view taken by the research, the results are backed up by rigorous and extensive evidence and can hence be regarded as valid (Eriksson & Kovalainen, 2008, ch. 19).

According to Yin (2009), external validity refers to the degree of generalizability of the results into other situations. They point out that results from a single-case study are more challenging to be generalized, but the representative nature of the case in this thesis justifies the approach. Moreover, the external analytical validity can be enhanced by comparing results of the study with previously developed theory presented in literature. (Yin, 2009, p. 41) Indeed, many of the findings regarding the interorganizational activities were in line with prior research.

Furthermore, the generalizability of the results is limited, due to the emphasis on practicality during the research process. While aiming for a real-life business model for Solita, aspects not relevant to Solita's business lines were not investigated thoroughly. Consequently, some relevant interorganizational activities within the AI governance ecosystem, like legal advisory, insurance, and third-party AI system auditing did not get much attention.

6.4 Future research

The findings and limitations of this research provide an opening for future research. Future research should focus on extending on the findings of this study and improve on the limitations by using different methods of research.

By being the first of its kind, this study provides a starting point for future research in business models supporting sustainable AI services. There was no exploratory research done into the services and activities provided by private companies in the AI governance ecosystem. Even though AI governance activities for achieving sustainability had been discussed, the business model supporting such services had not been explored. Indeed, finding 2 especially provides a framework which can guide further academic discussion.

Relating to the findings concerning the offering and value proposition (findings 1, 2 and 3), future research should focus on assessing how the value proposition of sustainable AI services is received by the market. Indeed, as Osterwalder and colleagues (2015) point out, there are three levels of value proposition-market fit. Even though various services have been shown to fulfill the second level of market fit (the product-market fit), the remaining services should be tested to find out if they too are able to gain traction in the market. A thorough business model fit is accomplished once there is evidence the business model can be implemented profitably. Certainly, as Geissdoerfer (2017) points out, a design-implementation gap between conceptualization and actual implementation might be present, which leads to promising ideas not being further considered, concepts not being applied, and applied business models failing in the market. Furthermore, the value created by the outlined services should be quantified through studying delivered sustainable AI services. Concrete cost of projects and their expected values could not be discussed in this thesis, due to the confidentiality of the data.

Regarding the value expected (finding 3), the relative importance of social, environmental and economic value should be further investigated. Indeed, the extent to which economic value drives investments was not thoroughly examined in this study.

Future research should also be conducted to analyze the business models of multiple organizations, for example, by using multiple case studies as a methodology. Moreover, business model design and research for other private companies or public sector organizations should be conducted to offer a comparison point for this research. Furthermore, research could focus on specific services or products, and investigate the value proposition and business model in greater retail. Such study directions would provide valuable reference points for the nature of business models supporting sustainable AI services (finding 4).

Also, similar studies should be conducted later in time once the ecosystem around sustainable AI is more established, for example, after the AI Act is in full effect. Indeed, such research would extend on the findings in this study around the state of sustainable AI (findings 5, 6 and 7). Additionally, the terminology used around sustainable AI (finding 7) would need more in-depth research.

Furthermore, quantitative methods could be used to answer similar research questions using larger interviews sets or surveys as data sets. Such research would supplement the current state of research, which is largely qualitative. Moreover, similar studies in a different location could complement the findings of this study.

In summary, research on the state of sustainable AI and the AI governance ecosystem as well as services within the ecosystem is still in an emerging state. Even though this research provides valuable findings, more research in the space is needed.

7. CONCLUSION

The objective of this thesis is to fill in the recognized gap in sustainable AI and AI governance ecosystem research, by studying interorganizational activities provided by private consultancies within the AI governance ecosystem. Additionally, a practical goal of the thesis was to conceptualize the case company's business model around their sustainable AI consulting business. By using business model design and the Business Model Canvas as a tool, the thesis set forth to answer the following research questions:

1. *What interorganizational activities are needed from consultancies within the AI governance ecosystem, and what value are the activities expected to create?*
2. *What characteristics does a business model supporting the delivering of sustainable AI offering have?*

By utilizing insights from Solita's interviews with 26 organizations in 11 industries, objectives, expected gains, and obstacles for achieving the client goals around the sustainability of AI systems were discovered. Additionally, by designing a business model over multiple design, interview, and feedback sessions, perspectives and goals of the service providing company were clarified.

As a result of the study, providing services for AI system sustainability assessment, organizational sustainable AI capabilities, and technical enablers for sustainable AI were found to be the interorganizational activities performed by private IT consultancies needed in the AI governance ecosystem. Moreover, the sustainable AI service offering framework was developed to elaborate on the nature of these services, distinguishing services along their "implementation – exploration" and "technical – non-technical" dimensions.

Financial gains and ensuring social responsibility are at the center of the value proposition of these services, which stem from enhancing AI risk management and improving AI operational excellence. These, in turn, contribute to creating competitive advantage, as the positive impact of the client organizations can be proven. However, the business model supporting such an offering did not differ much from the existing business model of the case company around IT and strategy consulting. Furthermore, the limited concrete actions taken to ensure AI systems' sustainability, the lack of familiarity of the AI Act, and the scattered terminology used in the space describe the state of sustainable AI.

This study contributes to the academic discussion around sustainable AI and AI governance. Additionally, it has implications for the case company as well as other actors in the AI governance ecosystem, in defining the role of consultancies within the ecosystem, and conceptualizing the value propositions of sustainable AI consulting services. Furthermore, it demonstrates the utility of the Business Model Canvas in academic research.

To conclude, the ecosystem around AI governance is still developing. However, it is apparent that private IT consultancies have a role in enabling sustainable AI, i.e., creating and operating trustworthy AI for the benefit of society, the environment, and businesses.

REFERENCES

AI Act, 2021. European Commission: Artificial Intelligence Act: Proposal for a Regulation laying down harmonised rules on artificial intelligence, Brussels: European Commision.

AI HLEG, 2019. High-Level Expert Group on Artificial Intelligence: Ethics guidelines for trustworthy AI, s.l.: s.n.

AIGA, 2022. Artificial Intelligence governance and auditing. [Online] Available at: https://ai-governance.eu/ [Accessed 10 June 2022].

Al-Debei, M. M. & Avison, D., 2010. Developing a unified framework of the business model concept. European Journal of Information Systems, Volume 19, pp. 359-376.

Amit, R. & Zott, C., 2015. Crafting Business Architecture: the Antecedents of Business Model Design. Strategic Entrepreneurship Journal, 9(4), pp. 331-350.

Anthony, L. F. W., Kanding, B. & Selvan, R., 2020. Carbontracker: Tracking and Predicting the Carbon Footprint of Training Deep Learning Models. arXiv.

artificialintelligenceact.eu, 2022. The AI act: Developments. [Online] Available at: https://artificialintelligenceact.eu/developments/ [Accessed 19 June 2022].

Asatiani, A. et al., 2020. Challenges of Explaining the Behavior of Black-Box AI Systems. MIS Quarterly Executive, 19(4).

Athanasopoulou, A. & De Reuver, M., 2020. How do business model tools facilitate business model exploration? Evidence from action research. Electronic Markets, Volume 30, pp. 495-508.

Baldassarre, B., Calabretta, G., Bocken, N. & Jaskiewicz, T., 2017. Bridging sustainable business model innovation and user-driven innovation: A process for sustainable value proposition design. Journal of Cleaner Production, Volume 147, pp. 175-186.

Barredo Arrieta, A. et al., 2020. Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. Information Fusion, Volume 58, pp. 82-115.

BCG, 2022. Deliver Powerful Business Results with Responsible AI. [Online] Available at: https://www.bcg.com/beyond-consulting/bcg-gamma/responsible-ai [Accessed 20 June 2022].

Bjørlo, L., Moen, Ø. & Pasquine, M., 2021. The Role of Consumer Autonomy in Developing Sustainable. Sustainability, Volume 2332, p. 13.

Bland, D. & Osterwalder, A., 2019. Testing Business Ideas. Newark: John Wiley & Sons, Incorporated.

Braganza, O., 2022. Proxyeconomics, a theory and model of proxy-based competition and cultural evolution. Royel Society Open Science, Volume 9.

Bräutigam, T. et al., 2022. EU regulation builds a fairer data economy: The opportunities of the Big Five proposals for businesses, individuals and the public sector, Helsinki: Sitra.

Brendel, A. B., Mirbabaie, M., Lembcke, T.-B. & Hofeditz, L., 2021. Ethical Management of Artificial Intelligence. Sustainability, 13(4).

Brundage, M. et al., 2020. Toward trustworthy AI development: mechanisms for supporting verifiable claims. arXiv.

Brusseau, J., 2020. AI human impact: toward a model for ethical investing in AI-intensive companies. Journal of Sustainable Finance & Investment, 1(28).

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

CEN-CENELEC, 2020. Road Map on Artificial Intelligence (AI), s.l.: CEN-CENELEC.

Chatila, R. & Havens, J. C., 2019. The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. In: M. Idinhas Ferreira, J. Silva Sequeira, G. Singh Virk & M. E. K. E. Tokhi, eds. Robotics and Well-Being. s.l.:Springler, pp. 11-16.

Chesbrough, H. & Rosenbloom, R. S., 2002. The role of the business model in capturing value from innovation: Evidence from Xerox Corporation's technology spin-off companies. Industrial and Corporate Change,, 11(3), pp. 529-555.

Chhillar, D. & Aguilera, R. V., 2022. An Eye for Artificial Intelligence: Insights Into the Governance of Artificial Intelligence and Vision for Future Research. Business & Society, 61(5).

Cihon, P., 2019. Standards for AI Governance: International Standards to Enable Global Coordination in AI Research & Development, Oxford: Future of Humanity Institute, University of Oxford.

Clarivate, 2021. Web of Science help: search results: sorting search results. [Online] Available at: http://webofscience.help.clarivate.com/en-us/Content/search-results.htm?Highlight=relevance
[Accessed 16 September 2022].

CodeCarbon, 2020. Track and reduce CO2 emissions from your computing. [Online] Available at: https://codecarbon.io/
[Accessed 20 June 2022].

Dauvergne, P., 2022. Is artificial intelligence greening global supply chains? Exposing the political economy of environmental costs. Review of International Political Economy, 29(3), pp. 696-718.

de Almeida, P. G. R., dos Santos, C. D. & Farias, J. S., 2021. Artificial Intelligence Regulation: a framework for governance. Ethics and Information Technology, Volume 23, pp. 505-525.

Di Vaio, A., Palladino, R., Hassan, R. & Escobar, O., 2020. Artificial intelligence and business models in the sustainable development goals perspective: A systematic literature review. Journal of Business Research, Volume 121, pp. 283-314.

Dignum, V., 2019. Responsible artificial intelligence: How to develop and use ai in a responsible way. 1 ed. s.l.:Springer Cham.

Dignum, V., 2020. Responsibility and artificial intelligence. In: M. Dubber, F. Pasquale & S. Das, eds. The Oxford handbook of ethics of AI. Oxford: Oxford university Press, p. 213–231.

Ebers, M., 2021. Standardizing AI - The Case of the European Commission's Proposal for an Artificial Intelligence Act. In: L. A. DiMatteo, M. Cannarsa & C. Poncibò, eds. The Cambridge Handbook of Artificial Intelligence: Global Perspectives on Law and Ethics. s.l.:Cambridge University Press .

Elsevier, 2021. Scopus: Access and use Support Center: What does "Relevance" mean in Scopus?. [Online] Available at: https://service.elsevier.com/app/answers/detail/a_id/14182/supporthub/scopus/kw/relevance/
[Accessed 16 September 2022].

Eriksson, P. & Kovalainen, A., 2008. Qualitative Methods in Business Research. s.l.:SAGE Publications Ltd.

European Commission, 2020. Digital Martkets Act: Proposal for a REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on contestable and fair markets in the digital sector, Brussels: European Commission.

European Commission, 2020. WHITE PAPER On Artificial Intelligence - A European approach to excellence and trust. [Online]

Available at: https://ec.europa.eu/info/files/white-paper-artificial-intelligence-european-approach-excellence-and-trust_en
[Accessed 2022].

Fielt, E., 2013. Conceptualising business models: Definitions, frameworks and classifications. Journal of Business Models, 1(1), pp. 85-105.

Fink, A., 2019. Conducting Research Literature Reviews: From the Internet to Paper. s.l.:Sage publications.

Fjeld, J. et al., 2020. Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-Based Approaches to Principles for AI. Berkman Klein Center for Internet & Society.

Floridi, L. et al., 2018. AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations. Minds and Machines, 28(4), pp. 689-707.

Floridi, L. et al., 2022. capAI - A Procedure for Conducting Conformity Assessment of AI Systems in Line with the EU Artificial Intelligence Act, Oxford: University of Oxford.

Foffano, F., Scantamburlo, T. & Cortés, A., 2022. Investing in AI for social good: an analysis of European national strategies. AI & SOCIETY.

França, C. L. et al., 2017. An approach to business model innovation and design for strategic sustainable development. Journal of Cleaner Production, 140(1), pp. 155-166.

Friedrichs, J. & Kratochwil, F., 2009. On Acting and Knowing: How Pragmatism Can Advance International Relations Research and Methodology. Cambridge University Press, 63(4), pp. 701-731.

Fumagalli, E., Rezaei, S. & Salomons, A., 2022. OK computer: Worker perceptions of algorithmic recruitment. Research Policy, 51(2).

Gartner Glossary, 2022. Artificial Intelligence Model Operationalization (ModelOps). [Online]
Available at: https://www.gartner.com/en/information-technology/glossary/artificial-intelligence-model-operationalization-modelops-
[Accessed 8 July 2022].

Gasser, U. & Almeida, V. A., 2017. A Layered Model for AI Governance. IEEE Internet Governance, 21(6), pp. 58-62.

Geissdoerfer, M., Savaget, P. & Evans, S., 2017. The Cambridge Business Model Innovation Process. Procedia Manufacturing, Volume 8, pp. 262-269.

Geissdoerfer, M., Vladimirova, D. & Evans, S., 2018. Sustainable business model innovation: A review. Journal of Cleaner Production, 198(10), pp. 401-416.

Genovesi, S. & Mönig, J. M., 2022. Acknowledging Sustainability in the Framework of Ethical Certification of AI. Sustainability, Volume 14.

Gimenez, C., Sierra, V. & Rodon, J., 2012. Sustainable operations: Their impact on the triple bottom line. International Journal of Production Economics, 140(1), pp. 149-159.

Goldkuhl, G., 2012. Pragmatism vs interpretivism in qualitative information systems research. European Journal of Information Systems, Volume 21:2, pp. 135-146.

Google, 2022. Responsible AI. [Online] Available at: https://cloud.google.com/responsible-ai [Accessed 20 June 2022].

Granlund, T., Stirbu, V. & Mikkonen, T., 2021. Towards Regulatory-Compliant MLOps: Oravizio's Journey from a Machine Learning Experiment to a Deployed Certified Medical Product. SN Computer Science, Volume 342, p. 2.

Guest, G., MacQueen, K. M. & Namey, E. E., 2012. Applied Thematic Analysis. Thousand Oaks, CA: SAGE Publications.

Halsband, A., 2022. Sustainable AI and Intergenerational Justice. Sustainability, 14(7).

Hansson, A. M., Pedersen, E., Karlsson, N. P. E. & Weisner, S. E. B., 2022. Barriers and drivers for sustainable business model innovation based on a radical farmland change scenario. Environment, Development and Sustainability.

Henderson, P. et al., 2020. Towards the Systematic Reporting of the Energy and Carbon Footprints of Machine Learning. Journal of Machine Learning Research, 21(248), pp. 1-43.

IEEE, 2022. Explore our approved IEEE 7000™ Standards & Projects. [Online] Available at: https://ethicsinaction.ieee.org/p7000/ [Accessed 29 June 2022].

ISO/IEC, 2022. ISO/IEC JTC 1/SC 42: Artificial Intelligence. [Online] Available at: https://www.iso.org/committee/6794475.html [Accessed 29 June 2022].

Jobin, A., Ienca, M. & Vayena, E., 2019. The global landscape of AI ethics guidelines. Nature Machine Intelligence, 1(9), pp. 389-399.

Johannesson, P. & Perjons, E., 2014. Research Strategies and Methods. In: An Introduction to Design Science. Cham: Springer, pp. 39-73.

Johnson, M. W., 2018. The Four-Box Business Model Framework. Boston: Harvard Business Press.

Johnson, M. W., 2018. The Four-Box Business Model Framework. In: Reinvent Your Business Model : How to Seize the White Space for Transformative Growth. Boston: Harvard Business Press, pp. 25-50.

Johnson, M. W., Christensen, C. M. & & Kagermann, H., 2008. Reinventing your business model. Harvard Business Review, 86(12), pp. 50-59.

Johnson, W. G. & Bowman, D. M., 2021. A Survey of Instruments and Institutions Available for the Global Governance of Artificial Intelligence. IEEE Technology and Society Magazine, 40(4), pp. 68-76.

Kaplan, A. & Haenlein, M., 2019. Siri, Siri, in my hand: Who's the fairest in the land? On the interpretations, illustrations, and implications of artificial intelligence. Business Horizons, 62(1), pp. 15-25.

Kelly-Lyth, A., 2021. Challenging Biased Hiring Algorithms. Oxford Journal of Legal Studies, 41(4), pp. 899-928.

Kindylidi, I. & Cabral, T. S., 2021. Sustainability of AI: The Case of Provision of Information to Consumers. Sustainability, 13(21).

Kirk, J. & Miller, M. L., 1986. Reliability and validity in qualitative research. Beverly Hills, CA: SAGE Publications Inc.

Kordzadeh, N. & Ghasemaghaei, M., 2021. Algorithmic bias: review, synthesis, and future research directions. European Journal of Information Systems, 31(3), pp. 388-409.

Korinek, A. & Stiglitz, J. E., 2017. ARTIFICIAL INTELLIGENCE AND ITS IMPLICATIONS FOR INCOME DISTRIBUTION, Cambridge: NATIONAL BUREAU OF ECONOMIC RESEARCH.

Koshiyama, A. et al., 2021. Towards Algorithm Auditing: A Survey on Managing Legal, Ethical and Technological Risks of AI, ML and Associated Algorithms, Londo: Working papaer, University College London.

Laato, S. et al., 2022b. AI Governance in the System Development Life Cycle: Insights on Responsible Machine Learning Engineering. Pittsburg, CAIN'22, IEEE.

Laato, S., Tiainen, M., Islam, A. N. & Mäntymäki, M., 2022a. How to explain AI systems to end users: a systematic literature review and research agenda. Internet Research, 32(7), pp. 1-31.

Lacoste, A., Luccioni, A., Schmidt, V. & Dandres, T., 2019. Quantifying the Carbon Emissions of Machine Learning. arXiv.

Lee, B. & Saunders, M. N., 2017. Introduction to Case Studies. London: SAGE Publications Ltd.

Lee, K., 2021. A Systematic Review on Social Sustainability of Artificial Intelligence in Product Design. Sustainability, 13(5).

Liberati, A. et al., 2009. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. Journal of Clinical Epidemiology, 62(10), pp. e1-e34.

Lovell, H. & Liverman, D., 2010. Understanding Carbon Offset Technologies. New Political Economy, 15(2), pp. 255-273.

Lv, Z., Lou, R. & Singh, A. K., 2021. AI Empowered Communication Systems for Intelligent Transportation Systems. IEEE Transactions on Intelligent Transportation Systems, 22(7), pp. 4579-4587.

Mäkinen, S., Skogström, H., Laaksonen, E. & Mikkonen, T., 2021. Who Needs MLOps: What Data Scientists Seek to Accomplish and How Can MLOps Help?. s.l., WAIN'21 at ICSE'21, IEEE.

Malik, H. M., Viljanen, M., Lepinkäinen, N. & Alvesalo-Kuusi, A., 2022. Dynamics of Social Harms in an Algorithmic Context. International Journal for Crime, Justice and Social Democracy, 10(4).

Malik, H. M., Viljanen, M., Lepinkäinen, N. & Alvesalo-Kuusi, A., 2022. Dynamics of Social Harms in an Algorithmic Context. International Journal for Crime, Justice and Social Democracy, 11(1), pp. 182-195.

Mäntymäki, M., Minkkinen, M., Birkstedt, T. & Viljanen, M., 2022a. Defining organizational AI governance. AI Ethics.

Mäntymäki, M., Minkkinen, M., Birkstedt, T. & Viljanen, M., 2022b. Putting AI Ethics into Practice: The Hourglass Model of Organizational AI Governance. arXiv:2206.00335.

Martin, K., 2019. Designing Ethical Algorithms. MIS Quarterly Executive, 18(2).

McManners, P., 2015. The action research case study approach: A methodology for complex challenges such as sustainability in aviation. Action Research, 14(2), pp. 201-216.

Mensah, j., 2019. Sustainable development: Meaning, history, principles, pillars, and implications for human action: Literature review. Cogent Social Sciences, 5(1).

Metallo, C., Agrifoglio, R., Schiavone, F. & Mueller, J., 2018. Understanding business model in the Internet of Things industry. Technological Forecasting and Social Change, Volume 136, pp. 298-306.

Metsäranta, A. & Rauhala, S., 2022. Solita qualitative research: State of sustainable AI, Helsinki: Solita Oy. Unpublished.

Minkkinen, M., Niukkanen, A. & Mäntymäki, M., 2022a. What about investors? ESG analyses as tools for ethics-based AI auditing. AI & SOCIETY.

Minkkinen, M., Zimmer, M. P. & Mäntymäki, M., 2022b. Co-Shaping an Ecosystem for Responsible AI: Five Types of Expectation. Information Systems Frontiers.

Mittelstadt, B., 2016. Automation, Algorithms, and Politics| Auditing for Transparency in Content Personalization Systems. International Journal of Communication, Volume 10, pp. 4991-5002.

Mökander, J., Axente, M., Casolari, F. & Floridi, L., 2021. Conformity Assessments and Post-market Monitoring: A Guide to the Role of Auditing in the Proposed European AI Regulation. Minds and Machines, pp. 1-27.

Mökander, J. & Floridi, L., 2021. Ethics-Based Auditing to Develop Trustworthy AI. Minds and Machines, Volume 31, pp. 323-327.

Mökander, J., Morley, J., Taddeo, M. & Floridi, L., 2021. Ethics-Based Auditing of Automated Decision-Making Systems: Nature, Scope, and Limitations. Science and Engineering Ethics, 27(44).

Morgan, D. L., 2014. Pragmatism as a Paradigm for Social Research. Qualitative Inquiry, 20(8), pp. 1045-1053.

Morik, K. et al., 2021. Yes We Care!- Certification for Machine Learning Methods through the Care Label Framework. arXiv.

Morley, J., Floridi, L., Kinsey, L. & Elhalal, A., 2020. From What to How: An Initial Review of Publicly Available AI Ethics Tools, Methods and Research to Translate Principles into Practices. Science and Engineering Ethics, Volume 26, pp. 2141-2168.

Morris, M., Schindehutte, M. & Allen, J., 2005. The entrepreneur's business model: Toward a unified perspective. Journal of Business Research, 58(6), pp. 726-735.

Myllyaho, L. et al., 2022. On misbehaviour and fault tolerance in machine learning systems. Journal of Systems and Software, Volume 183.

Myllyaho, L. et al., 2021. Systematic literature review of validation methods for AI systems. Journal of Systems and Software, Volume 181.

Nativi, S., De Nigris, S. & European Commission, J. R. C., 2021. AI Watch: AI standardisation landscape state of play and link to the EC proposal for an AI regulatory framework, s.l.: Publications Office.

OECD, 2019. Scoping the OECD AI principles: Deliberations of the Expert Group on Artificial Intelligence at the OECD (AIGO), Paris: OECD Publishing.

Oliva, R., 2019. Intervention as a research strategy. Journal of Operations Management, 65(7), pp. 710-724.

Osterwalder, A., 2004. The business model ontology: a proposition in a design science approach. s.l.:Doctoral dissertation, Université de Lausanne, Faculté des hautes études commerciales.

Osterwalder, A. & Pigneur, Y., 2010. Business model generation: A handbook for visionaries, game changers, and challengers. s.l.:Wiley.

Osterwalder, A., Pigneur, Y., Bernarda, G. & Smith, A., 2015. Value Proposition Design: How to Create Products and Services Customers Want. 1 ed. s.l.:John Wiley & Sons, Incorporated.

Osterwalder, A., Pigneur, Y. & Tucci, C. L., 2005. Clarifying business models: Origins, present, and future of the concept. Communications of the association for Information Systems, Volume 16.

Oxford Languages, 2022. Oxford Languages and Google. [Online] Available at: https://languages.oup.com/google-dictionary-en/ [Accessed 10 June 2022].

Papagiannidis, E. et al., 2022. Toward AI Governance: Identifying Best Practices and Potential Barriers and Outcomes. Information Systems Frontiers.

Patterson, D. et al., 2022. The Carbon Footprint of Machine Learning Training Will Plateau, Then Shrink. TechRxiv.

Patton, M. Q., 2002. Qualitative research and evaluation methods. 3 ed. Thousand Oaks: Sage Publications.

Percy, C., Dragicevic, S., Sarkar, S. & Sarkar, S., 2021. Accountability in AI: From principles to industry-specific accreditation. AI COMMUNICATIONS, 34(3), pp. 181-196.

Pfeffer, J., 2010. Building Sustainable Organizations: The Human Factor. Academy of Management Perspectives, 24(1).

Polydoropoulou, A. et al., 2020. Prototype business models for Mobility-as-a-Service. Transportation Research Part A: Policy and Practice, Volume 131, pp. 149-162.

Power, M., 1997. The Audit Society: Rituals of Verification. Oxford: New York: Oxford University Press.

Purvis, B., Mao, Y. & Robinson, D., 2019. Three pillars of sustainability: in search of conceptual origins. Sustainability Science, Volume 14, pp. 681-695.

PwC, 2020. PwC Ethical AI Framework. [Online]
Available at: https://www.pwc.com/gx/en/issues/data-and-analytics/artificial-intelligence/what-is-responsible-ai.html
[Accessed 20 June 2022].

Raper, R. et al., 2022. Sustainability Budgets: A Practical Management and. Sustainability, p. 14.

Raposo, V. L., 2022. Ex machina: preliminary critical assessment of the European Draft Act on artificial intelligence. International Journal of Law and Information Technology, 30(1), pp. 88-109.

Renda, A., 2020. Europe: Toward a Policy Framework for Trustworthy AI. In: M. D. Dubber, F. Pasquale & S. Das, eds. The Oxford Handbook of Ethics of AI. Oxford: Oxford University Press., pp. 649-666.

Robbins, S. & van Wynsberghe, A., 2022. Our New Artificial Intelligence Infrastructure: Becoming. Sustainability, 14(8).

Ryan, M., 2020. In AI We Trust: Ethics, Artificial Intelligence, and Reliability. Science and Engineering Ethics, Volume 20, pp. 2749-2767.

Ryan, M. & Stahl, B. C., 2020. Artificial intelligence ethics guidelines for developers and users: clarifying their content and normative implications. Journal of Information, Communication and Ethics in Society, 19(1), pp. 61-86.

Sætra, H. S., 2021. AI in Context and the Sustainable Development Goals: Factoring in the Unsustainability of the Sociotechnical System. Sustainability, 13(4).

Saidot, 2022. Saidot. [Online]
Available at: https://www.saidot.ai/platform
[Accessed 21 June 2022].

Saunders, M., Lewis, P. & Thornhill, A., 2019. Research Methods for Business Students. 8 ed. s.l.:Pearson Education, Limited.

Schmitt, L., 2022. Mapping global AI governance: a nascent regime in a fragmented. AI and Ethics, Volume 2, pp. 303-314.

Seppälä, A., 2021. Implementing Ethical AI : From Principles to AI Governance, Turku: Utupub. University of Turku.

Seppälä, A., Birkstedt, T. & Mäntymäki, M., 2021. From Ethical AI Principles to Governed AI. Austin, texaTexas, ICIS 2021 Proceedings.

Shneiderman, B., 2020. Bridging the Gap Between Ethics and Practice: Guidelines for Reliable, Safe, and Trustworthy Human-centered AI Systems. ACM TRANSACTIONS ON INTERACTIVE INTELLIGENT SYSTEMS, 10(4).

SIIA, Sotware and Information Industry Association;, 2017. Ethical principles for artificial intelligence and data analytics, s.l.: SIIA.

Snyder, H., 2019. Literature review as a research methodology: An overview and guidelines. Journal of Business Research, Volume 104, pp. 333-339.

Solita, 2022. About us. [Online] Available at: https://www.solita.fi/en/company/ [Accessed 1 July 2022].

Sort, J. C. & Nielsen, C., 2018. Using the business model canvas to improve investment processes. Journal of Research in Marketing and Entrepreneurship, 20(1), pp. 10-33.

Spinuzzi, C., 2005. The Methodology of Participatory Design. TECHNICAL COMMUNICATION , 52(2), pp. 163-174.

Strubell, E., Ganesh, A. & McCallum, A., 2019. Energy and Policy Considerations for Deep Learning in NLP. arXiv preprint arXiv:1906.02243.

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), pp. 368-383.

Tomašev, N. et al., 2020. AI for social good: unlocking the opportunity. Nature Communications, 11(2468).

Tsujimoto, M., Kajikawa, Y., Tomita, J. & Matsumoto, Y., 2018. A review of the ecosystem concept — Towards coherent ecosystem design. Technological Forecasting and Social Change, Volume 136, pp. 49-58.

United Nations, 2022. Sustainable Development Goals. [Online] Available at: https://sdgs.un.org/goals [Accessed 10 June 2022].

University of Mary Washington, 2022. Economic Sustainability. [Online] Available at: https://sustainability.umw.edu/areas-of-sustainability/economic-

sustainability/
[Accessed 13 June 2022].

ur Rehman, M. H. u., Chang, V., Batool, A. & Wah, T. Y., 2016. Big data reduction framework for value creation in sustainable enterprises. International Journal of Information Management, Volume 36, pp. 917-928.

Vakkuri, V. et al., 2022. Ethically Aligned Design of Autonomous Systems : Industry Viewpoint and an Empirical Study. Electronic Journal of Business Ethics and Organization Studies, 27(1), pp. 4-15.

Vallance, S., Perkins, H. C. & E.Dixon, J., 2011. What is social sustainability? A clarification of concepts. Geoforum, 42(3), pp. 342-248.

van Wynsberghe, A., 2021. Sustainable AI: AI for sustainability and the sustainability of AI. AI and Ethics, Volume 1, p. 213–218.

Viljanen, M. & Parviainen, H., 2022. AI Applications and Regulation: Mapping the Regulatory Strata. Frontiers in Computer Science, Volume 3.

Walker, T., 2007. Why economists dislike a lump of labor. Review of Social Economy, 65(3), pp. 279-291.

Willard, B., 2012. The New Sustainability Advantage : Seven Business Case Benefits of a Triple Bottom Line - Tenth Anniversary Edition. s.l.:New Society Publishers.

Wilson, C. & der Velden, M., 2022. Sustainable AI: An integrated model to guide public sector decision-making. Technology in Society, Volume 68.

Wirtz, B. W., Pistoia, A., Ullrich, S. & Göttel, V., 2016. Business Models: Origin, Development and Future Research Perspectives. Long Range Planning, 49(1), pp. 36-54.

Wirtz, B. W., Weyerer, J. C. & Sturm, B. J., 2020. The Dark Sides of Artificial Intelligence: An Integrated AI Governance Framework for Public Administration. International Journal of Public Administration, 43(9), pp. 818-829.

Wohlin, C. & Aurum, A., 2015. Towards a decision-making structure for selecting a research design in empirical software engineering. Empirical Software Engineering, Volume 20, pp. 1427-1455.

Wohlin, C., Mendes, E., Felizardo, K. R. & sKalinowski, M., 2020. Guidelines for the search strategy to update systematic literature reviews in software engineering. Information and Software Technology, Volume 127.

Wu, C.-J.et al., 2022. Sustainable ai: Environmental implications, challenges and opportunities. Proceedings of Machine Learning and Systems 4 (2022), pp. 795-813.

Yin, R. K., 2009. Case Study Research : Design and Methods. 4 ed. s.l.:SAGE Publications.

Zott, C. & Amit, R., 2010. Business Model Design: An Activity System Perspective. Long Range Planning, 43(2-3), pp. 216-226.

Zott, C., Amit, R. & Massa, L., 2011. The Business Model: Recent Developments and Future Research. Journal of Management, 37(4), pp. 1019-1042.