UNIVERSITY OF TURKU

ABSTRACT

	Bachelor's thesis
x	Master's thesis
	Licentiate's thesis
	Doctoral dissertation

Subject	Information Systems Science	Date	26.8.2021
Author	Anniina Niukkanen	Number of pages	95+appendix
Title	Taking the responsible use of AI into account in ESG analyses		
Supervisor	Ph.D. Matti Minkkinen		

Abstract

Including nonfinancial information related to environmental, social, and governance (ESG) issues to investment analyses has become mainstream practice among investors during the past decade. However, considering issues with no historical background in the practice can prove challenging. One of these potential issues is related to artificial intelligence (AI), as the increasingly efficient and complex algorithms also make it more difficult to ensure that the decisions they arrive at are ethically sound and do not cause unintended harm to individuals. As more attention has been directed towards the potential ethical issues of AI, organizations have started creating their own sets of principles of responsible AI, with the intention of limiting the potential issues through self-regulation.

This thesis is a qualitative exploratory study, with the purpose of investigating whether companies' responsible use of AI is currently included in ESG investment analyses, and how investors generally perceive questions related to responsible use of AI. The possible role of principles of responsible AI in ESG analyses is also included in the study. Insights from 5 semi-structured interviews with professionals from the field of responsible AI and ESG investing were collected and analyzed using the thematic analysis approach.

The findings indicate that as of now, AI is still seen as a relatively unknown topic to investors. Following this, taking the responsible use of AI into account in ESG analyses is still a novel topic in most cases, although a case-by-case analysis may still be conducted for companies which clearly leverage AI in their operations. AI was still recognized as a potentially material issue for various industries and companies, indicating that incorporating it to ESG evaluations in the future may be justified. While the principles were not considered to currently have a noticeable role in analyzing companies' use of AI, their potential role both in indicating companies' understanding of ethical AI issues and incorporating the topic of AI to ESG evaluations did emerge during the interviews.

Key words	artificial intelligence, AI, ESG, sustainable investing, ethics, principles

TURUN YLIOPISTO

TIIVISTELMÄ

	Kandidaatintutkielma
x	Pro gradu -tutkielma
	Lisensiaatintutkielma
	Väitöskirja

Oppiaine	Tietojärjestelmätiede	Päivämäärä	26.8.2021
Tekijä	Anniina Niukkanen	Sivumäärä	95+liite
Otsikko	Tekoälyn vastuullisen käytön huomioiminen ESG-analyyseissä		
Ohjaaja	FT Matti Minkkinen		

Tiivistelmä

ESG-asioihin, eli ympäristö- ja yhteiskuntavastuuseen sekä yrityksen hallintotapaan liittyvän ei-taloudellisen informaation sisällyttäminen sijoituspäätöksiin on valtavirtaistunut viimeisen vuosikymmenen aikana. Uusien ongelmien huomioon ottaminen voi kuitenkin olla vaikeaa ilman historiallista dataa, jota vasten yritysten vastuullisuutta voisi mitata. Yksi näistä mahdollisista ongelmista on tekoäly, sillä entistä tehokkaampien ja tämän myötä vaikeaselkoisempien algoritmien tekemien päätösten eettisyyden varmistaminen sekä yksilöihin kohdistuvan tahattoman haitan ehkäiseminen on entistä vaikeampaa. Näiden mahdollisten eettisten ongelmien tunnistamisen myötä organisaatiot ovat alkaneet luomaan omia tekoälyn eettisiä periaatteitaan, jotta mahdollisia haittoja voitaisiin ehkäistä itsesäätelyn avulla.

Tämä pro gradu -tutkielma on kvalitatiivinen ja eksploratiivinen tutkimus. Tutkimuksen tarkoitus on selvittää miten yritysten harjoittama tekoälyn vastuullinen käyttö otetaan tällä hetkellä huomioon ESG-sijoitusanalyyseissä, ja miten sijoittajat ymmärtävät tekoälyyn liittyviä kysymyksiä yleisellä tasolla. Tekoälyn eettisten periaatteiden mahdollinen rooli ESG-analyyseissä on myös sisällytetty tutkimukseen. Puolistrukturoiduista haastatteluista kerättyjä viiden vastuullisen tekoälyn ja ESG-sijoittamisen ammattilaisten näkemyksiä käytettiin materiaalina tutkimuksen temaattisessa analyysissä.

Löydösten perusteella tekoäly on vielä tällä hetkellä suhteellisen vieras aihe sijoittajille. Tästä seuraten myös tekoälyn vastuullisuuteen liittyvät kysymykset ja niiden huomioiminen ESG-analyyseissä ovat edelleen useimmissa tapauksissa vieraita, vaikkakin tapauskohtaisia tarkempia analyysejä voidaan tehdä, jos yritys selkeästi hyödyntää tekoälyä toiminnassaan. Tekoäly kuitenkin tunnistettiin mahdolliseksi olennaiseksi ongelmaksi monille toimialoille ja yrityksille, jonka johdosta tekoälyyn liittyvien kysymyksien sisällyttäminen ESG-arvioihin tulevaisuudessa voi olla perusteltua. Vaikka tekoälyn eettisten periaatteiden ei vielä koettu omaavan suurta roolia yritysten tekoälyn käytön arvioinnissa, niiden mahdollinen rooli yritysten tekoälyyn liittyvien ongelmien ymmärtämisen osoittamisessa sekä tekoälyn liittämisestä osaksi ESG-arviointeja nousivat esiin haastatteluissa.

Avainsanat	tekoäly, ESG, vastuullinen sijoittaminen, etiikka, periaatteet

UNIVERSITY OF TURKU
Turku School of Economics

TAKING THE RESPONSIBLE USE OF AI INTO ACCOUNT IN ESG ANALYSES

Master's Thesis
in Information Systems Science

Author:
Anniina Niukkanen

Supervisor:
Ph.D. Matti Minkkinen

26.8.2021
Turku

The originality of this thesis has been checked in accordance with the University of Turku quality assurance system using the Turnitin Originality Check service.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1 INTRODUCTION

1.1 Background

Investors are no longer including only traditional financial measures into their analyses when considering their investments. Nonfinancial information, including topics related to the impact companies have on the environment or societies, as well as whether companies are governed in a responsible manner, are increasingly being looked at by various groups of investors. Ernst & Young found in their latest global institutional investor survey that in 2020, 98 percent of the participated investors stated that they either conducted "a structured, methodical evaluation of nonfinancial disclosures" or at least "evaluate nonfinancial disclosures informally" (EYGM Limited 2020, 8). The results showcase a rapid development, as in the first corresponding study conducted in 2013, more than a third of the respondents reported that they included little or no evaluation of nonfinancial information to their investment analysis. Additionally, the increasing interest towards the formal use of nonfinancial information can be seen when comparing the survey results to the results of the corresponding study conducted two years prior, as the structured evaluations increased from 32 percent in 2018 to 72 percent in 2020 (EYGM Limited 2020, 8). Based on these results, it is not a surprise that the environmental, social, and governance (ESG) investing has become popular during the past decade, as investors are starting to see the benefits of including related matters into their analyses.

Despite ESG gradually turning mainstream, defining what should be considered material for companies and included in their ESG evaluations is not always an easy task. ESG has been criticized for utilizing backwards-looking data, even though the reason for including nonfinancial information to analyses is most often related to assessing the future performance of companies (Esty & Cort 2017, 27). This in turn leads to question whether new issues which have existed for a relatively short time may be overlooked, as their possible consequences are not well known to investors and the general public at large. Artificial intelligence (AI) can be said to belong to this category, since even though the concept was introduced already in the 1950s, it was only during this century when its potential could be better taken into use thanks to the developed computing power (Yang et al. 2018, 7). The development of AI is far from over, as it will continue disrupting and transforming industries. PwC has reported that AI may contribute $15.7 trillion to the global economy by 2030 (PwC, 2017), which does not seem implausible when

considering the possible benefits and opportunities in a wide array of industries that AI has been associated with: healthcare, transportation, financial services, retail, and energy sector, among others, will likely benefit from the introduction of AI (PwC, 2017). It will not only transform industries, but its benefits can reach to our lives by allowing us to do more with our time or enhance our capabilities. Floridi et al. (2018) have argued that AI can liberate us to use our time on more meaningful tasks by handling mundane tasks through automation, support us so that each individual can use their existing capabilities better or more efficiently or so that we can collectively enhance our societies at large, and lastly, provide a way for coordinating complex global challenges among societies. Additionally, researchers have investigated how AI based technologies may be used to achieve the Sustainable Development Goals (SDGs) set by the United Nations in the 2030 Agenda for Sustainable Development, with the conclusion that AI may indeed enable 134 of the 169 of the targets that have been set across 17 SDGs (Vinuesa et al., 2020).

However, even though AI might help us achieve greater heights as individuals as well as on a societal and global scale, the technology does not come without risk – and corresponding with the great benefits it may bring forth, the risks may lead to severe consequences as well. In addition to malicious or illegal activities, even systems which were originally intended for legal and beneficial use may cause negative effects. The power of AI lies especially in the massive learning capability, which allows AI models to better deal with the complex problems which may emerge in the real world (West 2018, 24; Arrieta et al. 2020, 82-83). With certain types of models, however, the increased ability to learn from the provided data may also lead to a situation where the developers who originally created the model cannot verify how the system makes decisions after the system has gone through the training period (Matthias 2004, 181-182). The question then arises, who should, or even could be held accountable for the decisions an AI makes? Even if someone claims the accountability, the increasing complexity of the models in use even today may lead to situations where it is practically impossible to decipher how the models reached certain conclusions. Trusting a system to make decisions even without knowing on what the system has based its decision on can be a source of distrust, but even more importantly, have lasting effects on the surrounding stakeholders. A commonly used example of this is an AI-powered recruitment program which was developed by Amazon starting from 2014 to screen their job applicants' resumes – only for the development project to be eventually discarded in 2018 due to the bias against female applicants, as

the training data contained mostly resumes from male applicants', who the company had been hired in the past (Dastin, 2018).

Given these reasons, it seems justified to say that at least when a company produces or uses an AI system which may have a direct effect on human lives, the possible impact of the system should be accounted for in ESG analyses. Still, deciding how this should be done in practice is not an easy task – all investors have their own preferences on what they consider important, and the multiple ways an AI system can be implemented ranging from relatively transparent linear models to complete black-boxes, as well as the various scenarios AI can be used, makes it difficult to create an all-encompassing method for analyzing its effects (Arrieta et al. 2020; Cort & Esty 2020, 493-494; Du & Xie 2021). Despite the apparent challenges, the possible issues need to be addressed: if not for mitigating potential damage to the environment or societies, the potential damage to company reputation and the possible decrease of company value would justify including AI to ESG considerations.

Whether this will be done by merely adopting the existing ESG rating frameworks, or by creating new indicators specifically for the responsible use of AI remains to be seen. One possibility in this area would be to harness principles of responsible AI to help investors identify trustworthy investment targets. The principles, being essentially sets of guidelines for guiding others on how AI should be used in an ethically sound manner, or a pledge from the releasing organization to take the principles into account in their AI operations, could potentially be useful for both the investor and the investment target. Even though they have emerged only in the past few years, there are already hundreds of principles released by organizations ranging from individual companies to large international organizations – and despite the diverse set of publishers, the lists of principles have already been found to include similar core elements in them (Jobin et al. 2019, 391). The convergence of important themes suggests that asking whether an AI system complies with them or not could be a good starting point for investors to start their AI related analyses: whereas for companies, releasing principles could signal about CSR activities and leadership in the topic, among other things (Schiff et al. 2020, 156).

Even though taking ESG issues into consideration in investment analyses has greatly raised in popularity during the past decade, and the responsible use of AI having been brought to the attention of audiences outside the academic community during the recent years, there seems to be little information on how these two topics are or could be intertwined. Given that AI is expected to continue transforming various industries and human

lives, ensuring that its impact can also be properly taken into account should be of importance – but there is little information whether or how this is currently considered. This thesis will thus explore the conjunction of two large topics of responsible AI and ESG investing. Here ESG investing is considered a subset of the wider whole of sustainable investing (which includes different investment styles which take nonfinancial measures into account in investment analyses) where nonfinancial information is considered material for an asset's future financial performance. As an exploratory study, the purpose is to find out whether companies' responsible use of AI is currently considered in ESG analyses, and whether investors see taking related issues into account as a positive sign for companies. Additionally, the possible role of the principles of responsible AI in ESG analyses will be investigated, as well as how evaluating the impact of AI to company performance could develop in the future.

To gain an understanding of these topics, interviews with professionals from both the realm of AI and ESG investing were considered a natural approach, as these would provide insights of how these topics are depicted by practitioners. All interviews were conducted with Finnish participants in order to keep the scope of the study manageable. However, it should still be recognized that this topic will not be relevant only in Finland in the future, but should be considered by all affected stakeholders globally. Given the novelty of the topic and limited availability of prior research, conducting a qualitative exploratory study was considered an appropriate choice.

1.2 Definition of AI

Despite the massive interest AI has received in the past years and the long history of the field, there is still no clear consensus on what the term "artificial intelligence" means. For this reason, this final section of chapter 1 is used to briefly explain what is meant with AI in this study in order to define the scope of the research subject. The technical details of the related technologies (e.g. machine learning or natural language processing) are however omitted, as the goal of this study is to investigate how investors perceive AI in general, rather than how the technical solution of an AI product affects investor interest.

One widely agreed aspect of the current AI systems is that they all fall under the *narrow intelligence* category of AI, referring to systems which cannot reach the same level of intelligence as humans can (AI HLEG 2019a, 5; Bostrom & Yudkowsky 2014, 318). In essence, while these systems can be highly efficient in specific tasks even to the extent where they can beat their human counterparts, they can only be used in the specific

contexts where they were developed to be used – unlike humans, who can learn various unrelated tasks by observing their surroundings and applying the gathered information in different scenarios (Bostrom & Yudkowsky 2014, 318). Some critics have claimed that because of this limitation, the current systems should not be considered intelligent at all, but rather just highly effective software meant to accomplish certain tasks. In their view, only the next levels of intelligence in AI systems, which would be able to reach the same level of intelligence as humans (*artificial general intelligence*) or surpass our intelligence (*artificial super intelligence*) – should receive the attribute (Bostrom & Yudkowsky 2014, 318)). In this study, narrow intelligence systems are considered to be a subcategory of AI, as it is the only level of AI which has been successfully taken into use thus far and excluding it would thus limit the scope of the thesis to a purely hypothetical basis.

Even within the subcategory of artificial narrow intelligence, there is still great variance on the different types of AI systems in use. One subset that is often associated with AI, and is sometimes even used synonymously with it, is machine learning (ML). This subset of AI is concerned with how machines can improve themselves autonomously through experience, and it has been used widely to improve the other subsets as well (Jordan & Mitchell 2015, 255). The reason why ML in particular has advanced the whole field of AI is the efficient use of computing power with which the systems can train themselves, instead of having a developer manually state outputs for all possible inputs. For example, an ML algorithm can be assigned a learning task (e.g. mark a credit card action as fraudulent or not fraudulent) and a performance metric (e.g. accuracy of detecting fraud), and with a set of training data (e.g. a data set with past credit card actions labeled as fraudulent or not fraudulent), let the algorithm find the optimal way of performing the given task from a large space of possibilities (Jordan & Mitchell 2015, 255).

Giving the algorithms the power to adjust themselves to provide better outcomes has improved the whole realm of AI tremendously, and both researchers and practitioners have put forth considerable efforts for creating increasingly efficient algorithms (Jordan & Mitchell 2015, 255). However, as the accuracy and efficiency have been improved, the models have become decreasingly understandable for humans, eventually leading to so called black-box models. These types of models, as the name implies, arrive at a given outcome in a way that cannot easily be verified by humans, if at all: they are given a set of input, and they provide the output without further explanations on how they actually produced it. A set of ML algorithms which are often associated with this idea are artificial neural networks and deep neural networks, which are groups of models created to loosely

resemble the way a human brain works (AI HLEG 2019a, 4). It should be noted, however, that not all ML algorithms used today are opaque black-boxes. Several types of models, such as linear regression, decision trees, or Bayesian models, can be considered transparent in the sense that their logic can either be understood by humans due to their simplicity, or through the use of mathematical models or visualizations. (Arrieta et al. 2020.) Figure 1 depicts the field of AI and the related terminology.

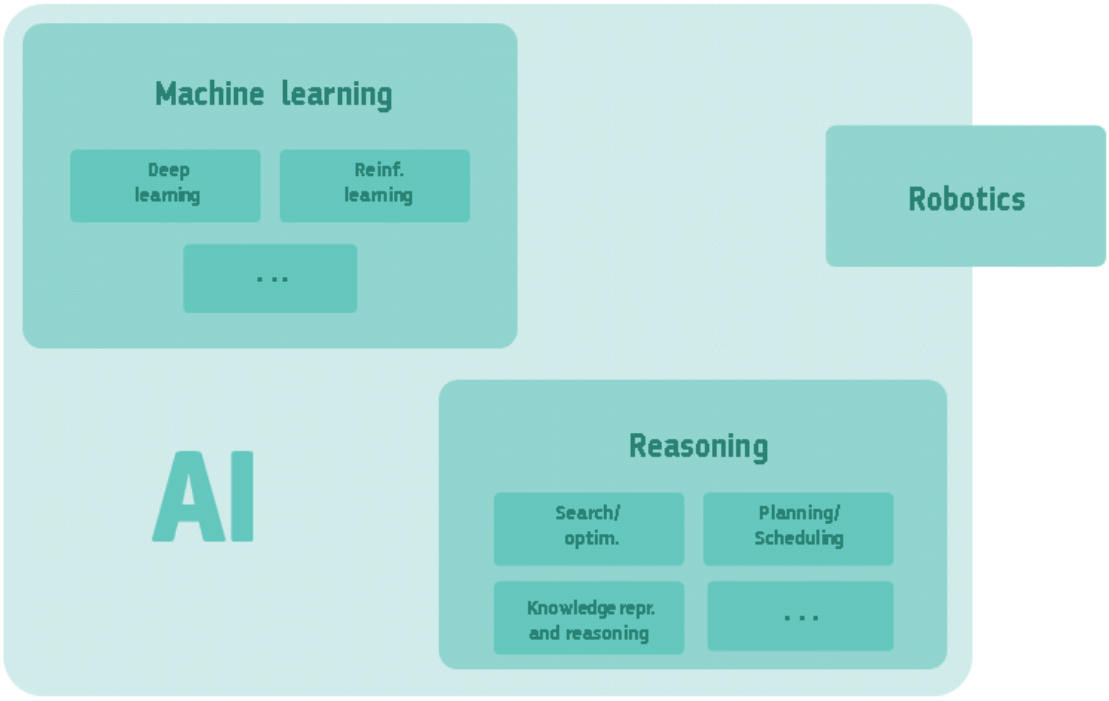

Figure 1 AI and its subsets (AI HLEG 2019a, 5)

As a summary, systems which can be said to belong under any of the three levels of AI intelligence are recognized as AI systems in this study. However, it is natural that especially during the empirical part of this study, emphasis is given on the artificial narrow intelligence systems, as the aim is to explore how the investment world currently perceives questions related to this topic. It is thus to be expected that as only narrow intelligence systems are currently in use, they would have received more attention from the general public compared to the more sophisticated general or super intelligence systems which are not guaranteed to ever be developed. Additionally, no restriction is made regarding which subsets of artificial narrow intelligence are considered, but rather the term AI is used as a common name for the field as a whole (including e.g. machine learning, neural networks, visual recognition or natural language processing).

1.3 Research questions & structure of the thesis

To answer the topics presented in section 1.1, the following research questions will be investigated in this thesis:

- How is the responsible use of AI taken into consideration when an ESG investment analysis is conducted?
- What kind of connections can be found between the existing principles of responsible AI and the criteria in ESG ratings?
- How could the responsible use of AI be considered in ESG analyses in the future?

Here, the responsible use of AI refers to whether companies develop and use AI systems in an ethically sound manner – meaning that a system does not cause deliberate harm to anyone, but also that potential unintended harm is accounted for throughout the system's lifecycle. Looking at each separate step of the lifecycle and how companies take the responsible use into account in each step is outside the scope of this thesis (the lifecycle is still shortly introduced in chapter 2), but instead the responsible use is considered from a wide perspective covering the whole lifecycle.

This thesis will be structured as follows. After the introduction, chapter 2 will present the first main theme of the thesis, the responsible use of AI. In addition to introducing the subject overall and explaining why organizations who use AI should take it into consideration, the principles of responsible AI are also introduced, and some of the core principles will be discussed in more detail. Chapter 3 will introduce the other main theme of the thesis, which is ESG investing. The beginning of the chapter will provide a description of what ESG investing is and what type of issues are included in it, as well as shortly introduce two other major sustainable investment styles (socially responsible investing and impact investing) which take ESG issues into account. Related issues, as well as how certain topics turn material for companies are also considered in this chapter. The last section of chapter 3 will also combine the gathered information from the literature review. Chapter 4 is then used to describe the research methodology, for example, why certain approaches to conducting this study were selected and how the research process was conducted. A trustworthiness evaluation is also included in this chapter. Chapter 5 presents the empirical findings of the study, and in chapter 6, these findings are discussed and reflected to the previous theories. Finally, chapter 7 concludes the study.

2 RESPONSIBLE AI

2.1 Definition & reasoning

The possible use cases of AI are manifold, and the possible consequences AI may have on our individual lives and societies at large are equally widespread. From significantly increased efficiency and accuracy in decision-making to enhancing the productivity of us individuals, AI has been and will be a major force in transforming our societies along with the other technologies leading the fourth industrial revolution (Schwab 2016). In addition to supporting the efforts towards heavy use of automation in societies, many initiatives for using AI to promote socially beneficial targets have been proposed. Some have argued that AI could be used for "fostering human dignity and promoting human flourishing" (Floridi et al. 2018, 690), and researchers have also recognized that AI may even support the path to reaching the United Nations' 2030 Sustainable Development Goals (SDGs) (Truby 2020; Vinuesa et al. 2020). Notions such as "AI for Good" (AI for Good Foundation, 2021), AI for People (Floridi et al. 2018), "AI for Social Good" (Floridi et al. 2020) have been presented to emphasize the possibilities of using AI to advance societies and diminish the existing societal inequalities, as well as to fight the threats of climate change and other severe environmental issues.

However, harnessing the immense power to transform societies towards automation and the world for the good can also lead to unwanted outcomes. Among others, concerns have been raised towards biased results which may lead to discrimination of minority groups, user privacy being compromised either through the use of AI in surveillance or through the need for massive training data sets for the AI systems (AI HLEG 2019b, 11, 18; Price & Cohen 2019), and how moral dilemmas of autonomous vehicles should be solved. For example, should a vehicle protect a passenger or a pedestrian in an unavoidable collision? It would seem irrational to not utilize AI for the undeniable benefits it may bring forth and for the fact that it may even help us solve long-standing global issues: however, the possible downfalls of AI are not only unsettling, but can lead to severe ill-advised consequences – even the loss of human lives, as the recent semi-autonomous vehicle crash in Texas unfortunately implies (Singh 2021).

One might argue that at their core, at least the current AI systems are merely pieces of technology created to fulfil some particular purpose, and should thus be treated as any other software. However, they are not just isolated entities: they are part of complex

sociotechnical systems with ethical challenges on product, consumer and society levels (Dignum 2020, 2; Du & Xie 2021, 965-969). Bostrom & Yudkowsky (2014, 317) remind, however, that implementing AI will not present entirely new ethical challenges that have never existed before. They state that the problem lies instead within the use cases AI will be utilized for, as many of these use cases may be infused with societal expectations over the outcomes the machine makes, but which it cannot take responsibility for. The AI systems' ability to act autonomously, and especially the ML algorithms' learning capabilities can make controlling their impact difficult, leading to questions of how the consequences can be controlled and who should bear the responsibility over an AI system throughout its lifecycle (Dignum, 2020).

The more complex ML algorithms, such as neural networks, have made the question of responsible AI and having humans in control increasingly difficult. According to Fischer & Ravizza (1998), moral responsibility generally indicates that an individual has the needed information of their surroundings to make an informed decision over something, and that they understand the possible consequences of their actions. Put in other words, the individual must have a reasonable amount of control over their own actions (Fischer & Ravizza, 1998; Matthias 2004, 175). To fulfil this requirement, users of an AI system must receive a reasonable amount of information about the situation in which they use it, and what their interaction with it may lead to. However, the development of complex black-box models has led to a situation where gathering information for educated decision-making may be impossible, or at least unreasonably difficult. While the user may initially be aware of the way the developer intended the model to work, neither the user nor the developer might be able to accurately determine how it makes decisions after the training period. The sheer complexity which makes the model efficient is also why its reasoning is difficult to explain – it receives an input and produces an output, but no one has an exact idea why or how it reached the conclusion it did. (Matthias, 2004, 175-176.) This does not mean that the conclusion is wrong, but as Matthias has described it, "there is an increasing class of machine actions, where the traditional ways of responsibility ascription are not compatible with our sense of justice and the moral framework of society because nobody has enough control over the machine's actions to be able to assume the responsibility for them." (2004, 177).

Since some systems which utilize AI technologies can be used in environments where they may impact members of society on a large scale, the questions above need to be considered by a wider audience than just their developers or direct users – for example,

autonomous vehicles do not only affect the lives of their owners, but all other vehicle owners and pedestrians as well. Consulting different stakeholders throughout the system's lifecycle is important for ensuring that the benefit of one stakeholder group is not achieved with unjust detriment to another. Related to this, Rahwan (2018) has proposed that stakeholder groups with conflicting interests should form a so-called social contract among themselves, and through this contract communicate their shared values and preferences over which tradeoffs they deem acceptable (e.g. can the safety of pedestrians be increased, if it leads to decreased safety of autonomous vehicle passengers?). Their viewpoints would then be included in the system's operation by a human controller, essentially keeping "society in the loop" over how AI affects our lives (Rahwan 2018, 9). However, forming such a contract in an increasingly globalized world complicates the matter even further, as different societal values must be considered when discussing how AI systems should be used, and for what purposes they should be deployed (Dignum 2020; 2-3).

The responsible use of AI thus implies that an AI system is used in an ethically sound manner, where the development and use of AI is not only guided by applicable laws, but also the prevalent societal values. Using AI responsibly can mitigate the possible negative consequences, but also help ensure that the decision made by an AI are justifiable and that the possibilities of AI are recognized and leveraged to the fullest. (Floridi et al. 2018; Dignum 2020.) Floridi et al. (2018, 694) call this the "dual advantage of an ethical approach to AI", stating that organizations have the possibility to simultaneously identify new socially acceptable use cases for AI and discover possible downfalls and mitigate them beforehand, even in situations where the downfall would have been legally acceptable. They also add that by clearly conducting AI related business in an ethical way, a company may increase public trust towards itself by being open about their practices, engaging with the public, as well as providing the possibility for redress (Floridi et al. 2018, 694-695, 702).

2.2 Recognizing harm

It should be noted that despite the matters addressed above, not all AI applications will have a negative effect on individuals or the wider society: for example, the high-level expert group on artificial intelligence set up by the European Commission (referred to as AI HLEG from hereafter) recognizes that some ethical considerations may not have much significance in industrial settings (AI HLEG 2019b, 15). Similarly, not all ethical issues need to be addressed at the same level of concern even among AI applications which

affect human lives. Each application's risks should be considered carefully and suitable risk mitigation processes for the identified risk level should be put in place, but classifying and accurately measuring the impact of different types of AI can be difficult. Du & Xie (2021, 963-965) have suggested a three-dimensional classification system of consumer AI products consisting of the products' multi-functionality (how many distinct tasks can the system perform), interactivity (how and how often users engage with the system) and intelligence (from less to more powerful AI technologies). The first two dimensions are more related to how consumers depict the usefulness of the product, whereas intelligence is a more technical aspect considered by companies themselves. Classifying a product as being multi-functional or highly interactive does not automatically lead to it posing more threats to its users, but rather the nature of the systems' risks changes. As an example, the authors present that products with low interactivity which also rank low on the multi-functionality dimension may be more prone to issues related to AI bias due to the limited number of data points for the system to use, whereas increasing the functionality pushes the focus more towards possible unemployment issues stemming from automating processes with AI (Du & Xie 2021, 964).

Even though the work of Du & Xie (2021) is limited to consumer products, it still highlights that even the class of potentially harmful products based on AI contains varying key issues depending on the product features. Actual risks of causing harm to individuals are not however the only issue that AI companies must face, as fear, misconceptions, and ignorance over what AI is and what it can currently be used for already pose a risk for the use of AI systems (Floridi et al. 2018, 691). It has been speculated that any damage caused by an AI system, whether due to system failure, unintended harm or some other cause, may produce excessive reactions and lead to overly strict regulations or incite negative public sentiment towards AI. Both could potentially hinder the use of AI to its full capacity, delaying us from actualizing the benefits of AI and thus causing opportunity costs for the wider societies in addition to the individual companies producing AI. (Imperial College 2017; Floridi et al. 2018, 691.) Ensuring that this will not happen should thus be of interest to all parties either developing or using AI systems, as well as governmental bodies seeking answers for reducing societal issues or developing new ways of serving their citizens with the use of AI.

Finally, it should be noted that taking appropriate steps to ensure the ethical use of AI systems is not just limited to monitoring the end product. Instead, the ethical viewpoint should already be included in all phases of an AI system's lifecycle. Morley et al. (2020,

2149) have divided the lifecycle in six phases, which include defining the business case for an AI system (*business and use-case development*), designing the future system (*design*), training the system with test data (*training and test data procurement*), building the system (*building*) and testing it (*testing*), and finally deploying the created system (*deployment*) and continuing to monitor it once it has been successfully deployed (*monitoring*) (2020, 2149). While this is not the only way to divide the phases, it gives an appropriate description of the diverse settings that should be included into considerations of using AI responsibly. For example, only considering ethics during the initial phases could lead to unwanted bias being introduced later on if the initial plans are not enforced properly. Similarly, only considering the related questions at the final phases could unveil severe ethical issues in the decisions the system makes, which could have been avoided if appropriate measures would have been taken throughout the lifecycle.

Summarizing the previous sections, AI has great potential for enhancing individuals and societies at large, and it may even help us solve complex global issues. Nevertheless, the possible severe harm AI may cause cannot be left unnoticed, and all affected stakeholders from individual citizens to various organizations and governments are needed to ensure that the technologies will be used in a just manner. Despite the increasing difficulties in controlling the complex systems, they will be assigned to make decisions in situations where humans would normally have to bear moral responsibility for the consequences. However, as the current narrow intelligence systems cannot be given the same responsibility over their decisions as humans, societal values and norms must be implemented in all stages of their lifecycle so that the systems will act according to the same standards that would be expected from their human counterparts.

2.3 Principles of responsible AI

As the understanding towards the possible negative consequences AI may cause has increased, various organizations across the globe have started to consider how these consequences could be governed. This has led to the emergence of ethical guidelines and sets of principles created by both the public and private sector, as well as by the non-governmental organizations (NGOs), both on the national and international levels (Jobin et al. 2019; Schiff et al. 2021). As Jobin et al. (2019, 389) remind, unlike legal regulations set by legislatures, these ethical guidelines are merely persuasive in nature, with no legal authority or mandate over the target audiences. Instead, as the researchers further state, the guidelines provide assistance for decision-making where ethical consideration or

guidance based on societal values is needed, by describing how ethics *should* ideally be integrated in the systems.

2.3.1 Motivations

Despite a consensus of the need for governing and mitigating the risks of AI gradually emerging globally, not all organizations have the same underlying motives for creating their own AI ethics documents. According to Schiff et al. (2020, 155-156), the motivations for creating AI ethics documents can be divided to three pairs of two based on their use cases, namely pursued end results, target audiences, and using the documents as communication methods.

End results

1. *Social responsibility* considers that AI should be used for generating social benefits to a wider audience and that risks of AI should be mitigated. Schiff et al. (2020, 155) consider that being driven by this motivation is seen to be most strongly related to NGOs and other groups where the wellbeing of all affected stakeholders is considered a priority.
2. *Competitive advantage* implies that AI ethics documents are created in an attempt to gain economic or political gains with AI. The issuers may include companies, or governments which want to emphasize their advances in utilizing AI. It should be noted that pursuing one of the first two motivations does not exclude an organization from pursuing the other as well.

Target audiences

3. *Strategic planning* implies that organizations use their AI ethics documents to either drive or support change internally. The goal of this change may be to better take ethical considerations into account in their operations or alter their internal practices so that they can better serve their clients.
4. *Strategic intervention* is similar to strategic planning, but the change or prevention of change has an external focus. Organizations may want to adopt principles of responsible AI to signal that no legislative actions are needed to ensure acceptable conduct of AI business, or otherwise impact their legal or political environments to suit their motives.

Communication

5. *Signaling social responsibility* refers to actions which try to convince stakeholders that the organization takes ethics of AI into consideration, or that they use AI according to the motivation of social responsibility. This can be done to achieve greater trust among stakeholders and thus improve the organization's position compared to their competitors, or as a method of strategic intervention. Additionally, some organizations may claim to be socially responsible even when it is not true to gain some benefit for themselves.

6. *Signaling leadership* can be done to elevate an organization's position among their peers, either to improve their public image or market share, or to be included in the previous leaders' discussions over AI ethics. Similarly to the previous motivation, signaling leadership can also be done whether the organization actually is advanced in the topic or not.

These motivations are not mutually exclusive, and a single organization may have several motivations for releasing their own AI ethics guidelines. For example, even the two seemingly contrary motivations of social responsibility and competitive advantage can be pursued simultaneously, though the emphasis between the two may vary between organizations. (Schiff et al. 2020, 155-156.)

As can be noticed from the work of Schiff et al. (2020), few motivations for creating ethics guidelines are driven by altruistic purposes. Some may create their principles in fears of appearing incompetent among their peers, or even claim to consider ethics of AI by creating their own principles merely to gain new clientele or avoid negative backlash from conducting questionable business (Floridi 2019; Schiff et al 2020). It has been recognized that ethical guidelines may be used to affect how the common rules of utilizing AI are developed, with each invested party having their own priorities which they want to advance. Especially the private sector has been criticized of developing their own guidelines primarily to avoid stricter laws being put in force, as the laws could limit their use of AI and thus lead to fewer financial returns (Benkler 2019; Floridi 2019, 188-189; Hagendorff 2020, 100). Wanting to appear more responsible by releasing ethics guidelines, which Floridi (2019, 187) has referred to as ethics bluewashing (stemming from greenwashing, or claiming to be environmentally friendly despite causing environmental damage), is closely related to the communication efforts mentioned by Schiff et al. (2020, 155-156), who also recognized that companies may signal implications which are not true.

2.3.2 Principles

Morley et al. (2020, 2145) have stated that even with organizations possibly advancing their self-interests with their ethics documents, several researchers and research groups have noticed regarding some principles, a significant level of mutual understanding of what should be included in the documents has emerged over the years. In their work, Jobin et al. (2019) analyzed a set of 84 documents where academic and legal documents had been excluded and discovered 11 themes of principles. Of these 11 themes, transparency, justice and fairness, non-maleficence, responsibility and privacy emerged in over

half of the documents. Floridi & Cowls (2019), in their assessment of six documents containing AI principles, argue that the 47 discovered principles could be summarized in five core principles: beneficence, non-maleficence, autonomy, and justice adapted from the realm of bioethics, as well as an additional principle of explicability. Similar themes were also found by Hagendorff (2020), whose analysis of 22 documents resulted in a high emergence of principles related to privacy, fairness, accountability, transparency, safety, common good, human oversight and solidarity, all of which emerged in at least half of the analyzed documents. It would thus seem that all three analyses agree on the most common principles, even though they use slightly different terms for themes of principles with the same underlying content. The similarities are presented in table 1, where the principles which appear at least in half of the analyzed documents in Jobin et al. (2019) and Hagendorff (2020) are compared to the aggregate core principles of Floridi & Cowls (2019). The numbers which are included after the principles in the columns for Jobin et al. (2019) and Hagendorff (2020) indicate the number of analyzed documents where the specific principle was found.

Table 1 Emergence of principles

Floridi & Cowls (2019)	Jobin et al. (2019, 395)	Hagendorff (2020, 102)
Beneficence		Common good, sustainability, well-being (16/22)
Non-maleficence	Non-maleficence (60/84) Privacy (47/84)	Privacy protection (18/22) Safety, cybersecurity (16/22)
Autonomy		Human oversight, control, auditing (12/22)
Justice	Justice and fairness (68/84)	Fairness, non-discrimination, justice (18/22) Solidarity, inclusion, social cohesion (11/22)
Explicability	Transparency (73/84) Responsibility (60/84)	Accountability (17/22) Transparency, openness (16/22)

It should be noted that despite the apparent consensus over which ethical aspects should be included in the sets of principles, the ethics guidelines cannot be said to be uniform or generally encompassing. Differences are common especially when comparing

documents made by different organization types, and some ethical aspects (such as the potential threats of artificial general intelligence, psychological impacts of using AI or cultural sensitivity) have largely been omitted thus far. (Schiff et al. 2021.) For example, by studying a set of 112 AI ethics documents from the private sector, public sector and NGOs, Schiff et al. identified that the private sector tends to publish documents with less ethical breath and which "emphasize ethical issues with ostensible technical fixes" (2021, 32), while the public sector and NGOs tend to have a wider scope of ethical issues and consider aspects related to regulation more. Hagendorff (2020, 103) similarly recognized in his comparison of the 22 documents that the common principles included in the guidelines are generally more prone to have technical fixes to them. Schiff et al. (2021, 31) additionally remind that the apparent consensus might hide disagreement over how the principles should be applied in practice, or whether different organizations prioritize the same principles among the most common ones. As it is not within the scope of this work to consider the underlying reasons behind the apparent consensus or the reasons for the differences and omissions, these aspects will not be investigated further. However, it should be noted that the guidelines are not definitive, and the field of AI ethics will likely develop in the future rapidly as new topics and problems emerge.

With this notion, central principles will be introduced as a conclusion to this section. The selected principles are based on the work of Jobin et al. (2019), Floridi & Cowls (2019), Hagendorff (2020) and Schiff et al. (2021), all of whom have studied several sets of AI ethics documents and guidelines (the number of studied documents ranging from six in Floridi & Cowls (2019) to 112 in Schiff et al. (2021)). A separate analysis of individual guidelines was deemed unnecessary for the purposes of this study, as the previous research done in this area has mostly agreed over which core principles are included in the guidelines. The selected principles thus contain the five principles which appear in more than half of the documents reviewed by Jobin et al. (2019), namely *transparency*, *justice*, *non-maleficence*, *responsibility* and *privacy*, as these themes are commonly found in the other studies as well. Additionally, *beneficence* will be included in the principles presented in this study, as Schiff et al. (2021, 37) found *social responsibility*, which they consider to be similar to beneficence presented in Floridi & Cowls (2019), among the most often mentioned principles. The work of Hagendorff (2020, 102) provides further support for including this theme, as he found the issue of *common good, sustainability, well-being* being addressed in 16 of the 22 analyzed documents.

Transparency is arguably one of the most commonly referred principles in the guidelines, being present in 73 of the 84 documents analyzed by Jobin et al. (2019, 391). In addition to being listed on its own, transparency is sometimes categorized as one aspect of the broader principle of explicability, as in Floridi & Cowls (2019) or the Ethics Guidelines for Trustworthy AI by the AI HLEG (2019b). Despite the apparent agreement on the importance of the principle, what it actually entails is more ambiguous: as Jobin et al. found in their analysis, there is "significant variation in relation to the interpretation, justification, domain of application and mode of achievement" (2019, 391) when it comes to explaining transparency. According to AI HLEG (2019b, 18), transparency is a combination of traceability, explainability, and communication, which should be applied not only to the AI system itself, but also to the data and the business models related to the system. Traceability means that people should be able to examine how the system arrived at a given decision from the data sets and processes. Explainability refers to being able to explain how the system works technically and how it is used in the environment where it is operationalized. Finally, the users should be informed when they are in contact with an AI system, and given the opportunity to decline interaction with a machine if they want to. (AI HLEG 2019b, 18.) Special attention should be given to the transparency of AI systems which handle human data or may otherwise impact human lives (Dignum 2020, 4). Despite its common occurrence among AI principles, Turilli & Floridi have argued that in computer ethics, transparency is not an "ethical principle in itself but a pro-ethical condition for enabling or impairing other ethical practices or principles" (2009, 105). In their view, ensuring the compliance of some other principles (e.g. accountability or informed consent) is dependent on disclosing a suitable amount of information through higher transparency. Nevertheless, full transparency could be harmful to some other principles or practices (e.g. privacy or copyright), implying that a suitable amount of control over data should be retained. This does not mean that the system cannot be transparent, as disclosing what information is being constrained can also enable other principles. (Turilli & Floridi 2009, 107.)

Justice is related to ensuring that the decisions made by an AI are fair and do not encourage discrimination towards any group of people. This sort of discrimination could result from a biased training data set, through which some bias that already exists in a society is transferred to algorithmic decision-making. The notions that the benefits of AI should be available to everyone on a global scale, and that AI should encourage diversity, can also be included within the principle of justice. (Floridi & Cowls 2019, 7.) As an

example, the document by AI HLEG (2019b, 11) specifically includes a section for ensuring that the "rights of persons at risk of exclusion" are considered in AI systems. Jobin et al. also found in their analysis that questions related to "the labour market, and the need to address democratic or societal issues" (2019, 394) are included under this principle from the public sector's viewpoint. This finding is in line with the work of Schiff et al. (2021, 38), who found that the public sector indeed emphasizes these types of issues more frequently than NGOs or especially the private sector.

Non-maleficence indicates that an AI system should be secure and technically robust, and that it should not cause any form of harm towards its users, such as discrimination, issues with privacy, or bodily harm. Both the intentional misuse of AI and unintended harm, which may be caused by overusing the system, are included within the principle. (Floridi & Cowls 2019, 6; Jobin et al. 2019, 394.) Fjeld et al. (2020, 38) remind that especially ML systems should be regularly tested even after deployment, as the algorithm continuously enhances itself, possibly leading to maleficent acting which was not originally observed. Also, in addition to an AI system itself being safe for the affected stakeholders, it should also be adequately protected against external threats, such as cyberattacks aiming to use the system for malicious use (AI HLEG 2019b, 12).

Responsibility and the closely related concept *accountability* are sometimes used interchangeably, and both are included under the principle of responsibility in Jobin et al. (2019). This does not seem surprising, as within the context of AI ethics, accountability is often said to refer to the question of who is responsible for the way the system works or for its decisions (Floridi & Cowls 2019, 8). However, the two do have their differences. For example, in Dignum's (2020) ART (accountability, responsibility, and transparency) model, accountability contains the idea that the system itself must be explainable and its decisions justifiable considering the prevailing societal norms and values. Responsibility is related to how the stakeholders themselves affect the way the increasingly complex system works, and must be assumed already before a certain action has been completed (Dignum 2020, 5). There is wide disparity regarding who should be responsible or accountable for an AI system, or who ensures that the systems themselves conform to these principles. Suggestions for this problem include individual developers to industries at large. (Jobin et al. 2019, 395.) It should be noted at this stage that despite the decision over who is responsible or accountable over a certain AI system, the question of what is deemed as acceptable behavior from the system and what should thus be integrated in the system's design is still a matter of all affected stakeholders, as was discussed in section

2.2. This principle is thus related to who ensures that the societal values from these stakeholders are considered throughout a system's lifecycle from design to having it in use.

Privacy has been a central issue of information technology due to the risks it poses in surveillance mechanisms and fast data processing and distribution capabilities, among others. Having already been an issue in the 1960s, the increasing efficiency of processing power and larger number of data collection points have only made the possibility of privacy violations more severe. (Nissenbaum 2009.) The proliferation of AI has largely been the result of the same advances, and given the tremendous data processing capabilities that certain AI technologies possess and that AI is being integrated into our lives in many levels, the issues have become even more severe. Generally, privacy can be considered as the right to control one's own personal information, including knowing where and how the information is collected, how it is processed and stored, and what consequences this process may have on the individual – although definitions of privacy have also been criticized of vagueness and inconsistency (Nissenbaum 2009, 2, 4; Brusseau 2021, 9). Nevertheless, privacy has been included as a fundamental right in the European legislation since the General Data Protection Regulation (GDPR) was put to force in 2018 (Regulation 2016/679).

Beneficence is related to utilizing AI for "promoting well-being, preserving dignity, and sustaining the planet" (Floridi & Cowls 2019, 6). Not to be confused with non-maleficence, which indicates that no harm should be caused to humanity or the environment in which the AI is utilized, beneficence contains the idea that AI should be inherently beneficial for its users or for the Earth, making it a distinct principle of its own (Floridi et al. 2018, 697; Floridi & Cowls 2019, 6). For example, Vinuesa et al. (2020) studied how AI can enable the United Nations' Sustainable Development Goals and found that AI may have a positive impact on 134 (79 %) of the 169 targets. To whom AI should be beneficial is not always considered unanimously, as Jobin et al. (2019, 395) found in their analysis that the private sector may primarily consider their customers' benefit: they did however find that most of the analyzed guidelines had a wider perspective on the matter.

Despite the principles being listed here as their individual entities, they are by no means separate from one another. In fact, many of the principles overlap heavily, and many of the principles are preconditions for the implementation of others. As an example, Vakkuri et al. (2020, 51) have presented the relations between the ART principles (along with few related principles) by stating that transparency is an enabler of accountability, while both transparency and accountability motivate responsibility – which in turn leads

to the fulfillment of other principles. As another example, the communication aspect of transparency, which includes the idea that users of AI systems should be presented the choice to refuse interaction with an AI is also connected to the principle of *human autonomy*, which is related to letting humans choose how they want to utilize the systems (AI HLEG 2019b, 12, 18). Trying to fulfill one principle may thus enable the organization to realize others as well.

2.3.3 Issues

The list of principles presented in section 2.3.2 certainly does not cover all principles or even themes of principles (see e.g. Jobin et al. (2019) or Schiff et al. (2021) for comprehensive analyses), but aims to describe the most often occurring themes for the purposes of understanding which aspects stakeholders, including investors, may need to consider in the future. Additionally, depending on the source, many of the omitted principles can be said to belong under some of the presented principles: for example, the principle of sustainability in Jobin et al. (2019, 395) is consider by Floridi & Cowls (2019, 6) to belong under the larger theme of beneficence. These differences in the analyzed documents are not necessarily a matter of organizations being negligent or not taking a certain principle into account due to ignorance. As was presented at the beginning of section 2.2, AI already has a wide array of use cases, and the risks in a certain industry might not have as much significance in another, making the creation of a generic principled (or regulatory) approach to responsible AI difficult (Dignum 2020).

The industry where an AI system is used and the system's specific use case affect which principles should be focused on or how important certain principles are individually or in relation to other principles. Still, the general direction of AI ethics has also received criticism and raised concerns. Jobin et al. (2019, 296) recognized in their analysis that the vast majority of the created AI ethics guidelines are produced in economically developed countries, leaving the rest of the world largely outside the discourse on AI ethics. The requirement presented by Dignum (2020) regarding taking local societal values into consideration in an AI's lifecycle is thus not guaranteed globally, as a small number of countries may steer the global discourse towards a direction which is suitable for their prevailing societal values. A similar imbalance has also been argued to be found in sizes of private companies who produce the guidelines, as most documents are produced by large corporations, leaving their smaller competitors with less influence over the course where AI ethics discourse is headed (Schiff et al. 2020, 154).

Additionally, Mittelstadt (2019) has raised concerns over the principled approach to AI ethics in general. He states in his work that unlike the field of medicine with its established principles, AI development lacks the needed common understanding and norms on how AI should be developed and used, as well as methods for operationalizing the principles and assigning accountability for the systems. As for the principles themselves, there seems to be a divergence on how they are interpreted and implemented, in addition to the previously mentioned question of where and when they should be applied (Jobin et al. 2019, 396). Ethical tradeoffs between the principles have also raised concerns: Jobin et al. (2019, 396) present an example, where large and diverse data sets are needed to ensure that an AI does not turn out biased, but collecting such data set might threaten the privacy and autonomy of individuals. Finally, the work of Morley et al. (2020) highlights that merely creating sets of principles is not enough, but companies must also find solutions for operationalizing them in practice. However, at the current stage where including ethical considerations into AI use may still induce more counterproductive implications, it may still be tempting for organizations to disregard this need in order to gain short-term benefits from the most efficient way of utilizing AI (Morley et al. 2020, 2161).

2.4 Concluding remarks

Even though there are clearly many unsolved issues regarding the ethical use of AI, the importance of preventing the possible harm caused by AI systems is equally undeniable. Moreover, as the work of Schiff et al. (2020) presented in section 2.3.1 suggests, altruistic purposes are not the sole reason for creating – and adhering to – principles of responsible AI. Signaling either social responsibility or leadership could be a much-needed indicator of understanding the types of issues organizations may face by taking AI into use. Sanders (2020) argues in his paper that as the issues and risks of AI become more emergent to large institutional investors, they may start paying closer attention to ensuring that the companies they invest in take ethical aspects of AI into consideration – given that clients, regulators and other stakeholders emphasize these issues as well, making them material for the investment's performance. Indications of regulatory actions potentially being taken in the future regarding use of AI have already emerged, with the European Commission's proposal for "laying down harmonised rules on artificial intelligence" (COM/2021/206 final), or commonly known as the EU AI Act, being a recent example. Indeed, AI related ethical issues could be one aspect of the CSR activities which affect the company's performance in the future, as Du & Xie (2021) also suggest.

Summarizing the covered topics of this chapter, a significant number of ethical guidelines for AI have been released in the recent years. From the ethical viewpoint this is a positive sign towards mitigating risks towards people and ensuring that AI will be used responsibly in its various use cases. In addition to communicating their efforts of being responsible members of societies, organizations who produce these documents can also have other underlying reasons for doing so, ranging from driving a positive impact to seeking competitive or other benefits to the organization itself. Despite the organizations being driven by different motivations, a significant consensus over the importance of certain themes of principles has emerged on a global scale. Transparency, justice, non-maleficence, responsibility, privacy and beneficence have been found to be included in the majority of the released guidelines. Each AI system is different, and the level in which each of these themes needs to be taken into account in a given system varies: for example, black-box algorithms may be more acceptable in use cases where no direct harm to humans can be caused, and privacy may not be an issue to begin with in systems which are meant for enhancing manufacturing processes. Still, these themes have been proposed to ensure that an AI system is developed in a manner where it does not cause damage to related parties, and that accountability over the system's decisions can be ensured. These aspects should thus be evaluated and integrated where necessary in AI systems' design from the beginning in accordance with the prevailing societal values, and implemented in a manner where they will be adhered to even as the systems enhance themselves over time.

3 ESG INVESTING

3.1 Terminology and definitions

As paying attention to environmental, social and governance (ESG) related matters has started gaining popularity especially during the past decade, a vast array of terminology and perhaps even broader selection of conflicting descriptions has also emerged. This has led to a situation where trying to comprehend what the concepts related to ESG investing entail can lead to even more confusion. Some of the central terms will thus be addressed here first, in addition to investigating why ESG issues are increasingly considered by different investor groups.

Depending on the source, ESG investing may be used as a common term for different investment styles which take ESG issues into account or be considered as a distinct style in itself (Hill 2020, 13-14). This has been caused by there being numerous styles where nonfinancial information related to environmental, social, and governance dimensions is integrated to an investment analysis and decision-making process (Hyrske et al. 2020, 18-19). To differentiate the set of different investment styles from the specific ESG investing approach, terms like *sustainable investing* used by the Global Sustainable Investment Alliance (GSIA 2018) have been taken to use for clarification purposes. As there is great variance in the underlying reasons for integrating nonfinancial information to investment decisions, in this study sustainable investing is used to refer to all the different investment styles which take ESG issues into consideration. Following this, ESG investing will be considered as its own distinct investment style, where ESG issues are integrated into investment analysis to mitigate risk with the aim of securing financial returns (Boffo & Patalano 2020, 14; Hyrske et al. 2020, 22-23).

As a short introduction, it is first necessary to consider the types of topics related to the three dimensions of ESG. The environmental dimension covers issues related to the use of natural resources, the effect companies have on the environment both on a local and global scale, and how companies work on reducing their emissions in their own operations and throughout their supply chains. The social dimension is concerned with how a company treats their own workforce or how a fair treatment of workforce in supply chains is managed, and how their operations and products affect their other stakeholders, such as customers. Lastly, the governance dimension is related to enabling and enhancing ethical conduct of business within a company and ensuring that good corporate

governance is practiced in all aspects of its operations. (Boffo & Patalano 2020; Finsif 2020.) Some key issues of each dimension are listed in table 2 for illustrative purposes, but it should still be noted that many of the issues mentioned cannot be assigned to a single category unambiguously, as they can impact several of the dimensions simultaneously (CFA Institute 2015, 4).

Table 2 ESG pillars with example issues

Environmental	**Social**	**Governance**
Climate change	Human rights	Corruption
Waste	Modern slavery	Board diversity
Pollution	Child labor	Executive compensation
Biodiversity	Product responsibility	Tax strategy

Source: PRI Association, n.d., Finsif 2020

It should also be noted that taking ESG issues into account does not have a standard method for doing so, but instead each investor can leverage different strategies both before making an investment and with existing assets. A full *ESG integration* means including ESG issues in the investment analysis along with traditional financial measures, either by evaluating companies in isolation or by comparing how different companies performed compared to each other regarding larger global or sector specific issues. (Silvola & Landau 2019, 38, Hyrske et al. 2020, 141-142). As this strategy has been associated with higher implementation costs compared to traditional investing (Kempf & Osthof 2008, 1279; van Duuren et al. 2016, 526), using less arduous strategies are also commonly practiced. For example, *negative screening* refers to simply excluding companies from investment portfolios, either for ethical reasons or to align a portfolio with an investor's personal values or preferences (PRI Association n.d.). Through *active ownership*, investors can also guide the investees' engagement in ESG issues, both in mitigating ESG risks and guiding towards sustainable operations. Investors can either engage with the companies individually or in collaboration with other investors with similar goals, and they may do so through discussions with company management or formal proxy voting. (Silvola & Landau 2019, 38; Hyrske et al. 2020, 124.)

In ESG investing, ESG issues are considered material for an asset's future financial performance, thus making their integration to the investment analysis necessary for capturing greater benefits from the investment. Even though ethical reasons or values are not

the main reason for engaging in a certain investment, they may also be included in the investment decisions: for example, large institutional investors, such as pension funds or insurance companies, may face societal pressure to refrain from investing in sin stocks and thus exclude them from their portfolios (Hong & Kacperczyk 2009, 16, 24). However, ethical reasons are not the main driver of the investment decisions, unlike in socially responsible investing (SRI) or impact investing. In SRI (also known as ethical investing, especially in the United Kingdom (Sandberg et al. 2009, 524)), investors construct their investment portfolios by integrating their personal or societal values into them. A common method for doing so is excluding certain types of investment assets from portfolios either by refraining from purchasing them in the first place or by divesting already owned assets so that a portfolio can be realigned to suit the values the investor wants to support. Commonly excluded targets include tobacco, firearms, alcohol, and gambling products, among others. (Hyrske et al. 2020, 20.) As a third form of sustainable investing, impact investing is an investment style which aims to achieve some positive environmental or social return with the investment. According to Hill (2020, 18), examples of the areas where impact may be sought include financial inclusion, education, housing or renewable energy. Investors who engage in either SRI or impact investing may have to accept lower returns from their investments to ensure value alignment. This is true especially with SRI, as a simple exclusion may lead to lower returns. In their study, Hong & Kacperczyk (2009) found that sin stocks (including e.g. the commonly excluded targets mentioned previously) have higher expected returns than their comparable stocks (e.g. soft drinks being comparable to alcohol products). While impact investors may also accept slightly lower returns when pursuing their goal, they do still have a stronger emphasis on making a profit on their investment compared to ethical investors. Lastly, it should be mentioned that while ethical and impact investors may focus on a particular ESG dimension, ESG investors consider the possible risks and opportunities that an asset may impose on all three dimensions (Hyrske et al. 2020, 21-22). Figure 2 shortly what the three sustainable investment styles generally focus on.

Sustainable investing		
Socially responsible investing	Impact investing	ESG investing
Personal or societal values have a significant effect on investment decisions, possibly leading to lower financial returns (e.g. due to exclusion of certain assets)	Focus on reaching positive social and/or environmental goal, while still reaching for maximum financial gains.	Focus on gaining greater financial benefits. ESG issues seen material for an asset's performance, making them necessary to account for in analyses.

Figure 2 Comparison of sustainable investment styles (c.f. Boffo & Patalano 2020, 15)

ESG investing has sometimes been understood to be a synonym for either of the two other presented investment styles, and past studies on poorer financial gains from SRI funds where stocks have been excluded due to ethical reasons have falsely been used to claim that all sustainable investment styles would provide lower returns (Hill 2020, 14). Friede et al. (2015) have conducted a second-order meta-analysis of over 2000 empirical studies which had measured the connection between ESG performance and corporate financial performance (CFP), and their findings indicate that ESG integration may in fact be beneficial for financial gains. In their study, the majority of the analyzed studies had found a positive relation between ESG and CFP, or at least that those companies which received higher ESG ratings produced comparable financial returns to lower rating companies.

Financial gains are not the only potential reason for ESG gaining attention. Boffo & Patalano suggest in their work that aspects such as "growing societal attention to the risks from climate change, the benefits of globally-accepted standards of responsible business conduct, [and] the need for diversity in the workplace and on boards" (2020, 6) will have an impact on consumer choices and thus the performance of companies, as well as directly on investors' decision-making. For portfolio and asset managers, the views of their own customers may also drive them towards more sustainable investment choices (Amel-Zadeh & Serafeim 2018, 91-92, 97, 101). The growing interest in environmental and social issues will likely increase in the future, as millennials and younger generations have been found to be more active in terms of incorporating their values in their investment decisions, to the extent where they have been considered driving sustainable investing forward (Boffo & Patalano 2020, 17; Hill 2020, 3).

As a third possible reason for ESG gaining attention, Boffo & Patalano (2020) describe that both companies and financial institutions are seeking a more long-term view

on their operations and risk and return evaluations, so that sustainable financial returns can be achieved. Integrating ESG issues into investment analysis has indeed been considered to affect the long-term risk and financial performance of investment portfolios (MSCI 2020, 2). A partial explanation for previous reluctance to incorporate ESG evaluation in investment analysis might have been partly related to this, as the findings of a late 2016 global survey of institutional investors reported that most asset managers and asset owners used shorter time frames to evaluate their portfolio performance than was needed to realize the benefits from including ESG to their analyses (Eccles et al. 2017, 128-129).

3.2 Measuring ESG compliance

Including ESG information to investment analyses has evolved from a practice of ethical investors to being popular among the mainstream investors as well (van Duuren et al. 2016, 531; EYGM Limited 2020, 8). Companies are under pressure for providing reliable and versatile ESG data for their various stakeholder groups with different interests, and sifting through all of the available data in the digital world can be a challenge for investors with limited resources. Given this, it does not seem surprising that the selection of ESG data providers has likewise increased during the past decade – especially larger investors have favored company-specific ESG ratings due to resource constraints (van Duuren et al. 2016, 529-530). The selection of ESG rating agencies has undergone notable changes since the financial crisis of 2008, after which the focus has gradually shifted from using only traditional financial measures to a more comprehensive investment analysis with nonfinancial information (Lopatta & Kaspereit 2014; Escrig-Olmedo et al. 2019, 3-5, 9). Agencies that are often used in academic literature regarding ESG scores include MSCI, Sustainalytics, Refinitiv, Vigeo Eiris, RobecoSAM, and Bloomberg ESG (see e.g. Escrig-Olmedo et al. 2019; Berg et al. 2020; Gibson et al. 2021), with MSCI and Sustainalytics also being the most favored ones among investors due to their wide coverage of companies (Wong & Petroy 2020, 14, 33-35). How the ratings from these agencies are used varies across investors. It is often not, however, viewed as a single truth of a company's ESG compliance and directly included to the investment analysis – most investors use the ratings as an additional source of information for their own investigation, or as a starting point from where their own research will start. (Wong & Petroy 2020, 13.)

While each of the rating agencies measure and compare how companies consider ESG issues in their business practices, their methods for doing so and consequently their

results can vary greatly, and the divergence of ESG ratings has been confirmed in multiple studies (see e.g. Dorfleitner et al. 2015; Chatterji et al. 2016; Berg et al. 2020). It has been suggested that some of the lack of convergence can be ascribed to regional differences in culture and ideology (Sandberg et al. 2009, 527), which may also be intentionally embraced by the rating agencies in order to differentiate their products from their competitors (Sandberg et a. 2009, 527; Daugaard 2020, 1512).

Berg et al. (2020) identified three elements which contribute to divergence in ESG rating agencies' results when evaluating the same companies. *Scope divergence* means that the rating agencies include different issues in their ratings. Even though all agencies generally include issues related to all three ESG pillars, what they measure within those pillars can vary greatly. For example, only three of the six studied agencies considered companies' toxic spills in their ratings, and perhaps surprisingly, not all agencies included issues related to tax compliance in their evaluations. Next, *measurement divergence* is caused by rating agencies measuring the same issues with different indicators. This may lead to a company being evaluated as sustainable in a certain category by one agency, whereas another using a different indicator would consider the same company as harmful for the same category. The third recognized element is *weight divergence*, which refers to agencies weighing the same issues as more or less important than their competitors, which may be caused by the agencies' intentional focus on certain ESG issues. (Berg et al. 2020.)

While Berg et al. found that rating divergence could be observed across sectors and regions (2020, 11), of the three elements presented above, measurement divergence was considered to have the most influence over the divergence, with scope divergence following closely. Looking at measurement divergence more closely, the most notable differences were observed in categories related to human rights and product safety (Berg et al. 2020, 30), indicating that rating agencies are not unanimous in how these two significant topics should be measured. The cultural or ideological differences highlighted by Sandberg et al. (2009, 527) may have a large impact on how these and other topics with higher divergence are measured, as each rating agency must decide what kinds of issues they consider material for a given category.

Following the presented findings thus far, the three dimensions of ESG also receive varying results from agencies. Interestingly, while the findings of Tamimi & Sebastianelli (2017) indicate that companies tend to be most transparent about the governance dimension, several studies both from academics and practitioners have also found that there is

most divergence related to governance ratings among the rating providers (see e.g. La-Bella et al. (2019); Gibson et al. (2021)). Company characteristics also seem to affect the results, as large companies, which were found to be more transparent in the study by Tamimi & Sebastianelli (2017), have received more divergent ratings compared to smaller companies (Gibson et al. 2021, 15, 20). It would thus seem that the more ESG data companies are able to produce to satisfy the needs of their increasing stakeholder groups, the more possibilities rating agencies and investors have for interpreting this data from their individual perspective. As Cort & Esty (2020, 493-494) have stated, each investor has different expectations for how companies should take sustainability issues into account and reasons for doing so: impact investors seek information about how their social or environmental goals can be reached, while ESG investors strive for decreased risks or greater opportunities from nonfinancial information, and catering for these different needs may well lead to further divergence in the ratings.

3.3 Reporting material issues

In addition to each investor focusing on matters they deem important, companies themselves can contribute to the confusion over what should be included in ESG evaluations, as there are different views over which ESG issues are material for their performance. Even companies within the same industries have been found to report on different issues and use incompatible reporting styles, making it difficult for investors to compare companies within industries (Cardoni et al. 2019). Furthermore, a large portion of the issues in companies' sustainability reports are not a part of their legal risk filings, indicating that companies do not consider the excluded issues as something that would greatly affect their performance. These deficiencies possibly lead to investors and other stakeholders question which ESG issues are truly financially material for them. (Cort & Esty 2020, 499.).

Many organizations have engaged in providing a solution for the situation by publishing guidelines on how companies should report their sustainability issues and risk mitigation efforts, as well as on which issues they should include in their reports to begin with. The use of such guidelines and standards has become common among companies, as 84 percent of the 250 largest companies in the world utilize some form of external framework in their reporting (KPMG International 2020). The Global Reporting Initiative (GRI) has had a large impact on sustainability reporting since their first G1 guidelines in 2000 to the latest version referred to as GRI standards published in 2016, which are still

in place today (GRI 2021). Their documents have for several years been the most widely utilized guidelines for sustainability reporting, with 73 percent of the 250 largest companies globally using either GRI guidelines or standards in their reporting in 2020 (KPMG International 2020). The purpose of the GRI standards is to promote standardized reporting of economic, environmental and social impacts of companies, thus creating a common language for communicating about the impact that companies may have on the three dimensions (GSSB 2016, 3). Despite their widespread use, both the previous GRI guidelines and the current GRI standards have received an array of criticism: even companies in the same sector have been found to produce noncomparable sustainability reports, despite the GRI guidelines being intended to act as a standardizing force (Boiral & Henri 2017).

Another increasingly utilized set of reporting standards have been published by the Sustainability Accounting Standards Board (SASB), which is an independent nonprofit organization founded in 2011 with standards for material issues covering 77 industries (SASB 2021). The SASB standards are among the most utilized ones after the GRI framework (KPMG International 2020, 25), and they seek to provide guidance on ESG issues that companies within the same industries should report on, as in SASB's view, the issues that may impact a company's performance are similar to those of their industry competitors. In contrast, the GRI standards state that material issues which should be included in sustainability reporting are "those that can reasonably be considered important for reflecting the organization's economic, environmental, and social impacts, or influencing the decisions of stakeholders" (GSSB 2016, 10). Unlike SASB standards, however, the GRI standards do not take a stance on which specific issues should be considered material, but instead companies themselves are responsible for recognizing issues which may be relevant for them. The organization suggests that materiality assessments should be done both through internal evaluations and by engaging with external stakeholders, with societal norms and international standards and agreements also being included in the evaluations. (GSSB 2016, 10.)

Leaving the matters that will be included in the reports to the consideration of companies may have its downfalls, as Tamimi & Sebastianelli (2017) have argued that companies tend to be less transparent about issues which they are not required to report on. In their study, companies were found to be the most transparent about issues related to the governance pillar, which the researchers deemed to result from requirements to report on financial and governance metrics set by the Securities and Exchange Commission (SEC)

in the United States. At the same time, deficiencies were found in reporting both social and environmental issues, leading to the researchers speculating that companies may still adopt a more reactive approach to ESG issues instead of proactively identifying and managing them. Interestingly, companies from industries which have been traditionally associated with heavy polluting or considerable societal harm (such as gas and oil, alcohol, or tobacco) were found to have higher social disclosure scores. (Tamimi & Sebastianelli 2017.) This finding supports previous studies where controversial industries have been argued to benefit from taking environmental and social issues voluntarily into consideration to meet the demanding expectations of their stakeholders more so than companies in non-controversial industries with lower perceived risk (see e.g. Cai et al. 2012; Jo & Na 2012).

Whereas the GRI documents are clearly focused on a more holistic view of materiality with impact over environmental and social themes covered, the SASB positions itself more clearly towards considering financially material issues and their effect on companies' financial performance (SASB 2021.) While these two organizations have distinct viewpoints over materiality, it should be mentioned that these two viewpoints do not necessarily have to be mutually exclusive: both could be adopted simultaneously, as is done in the notion of double materiality presented by the European Commission (2019). Nevertheless, there is still uncertainty regarding which issues should be considered material, no matter which standpoint over materiality is adopted. Looking to answer this problem, Rogers & Serafeim (2019) have conducted pioneering work in this area with their working paper, where they aim to unravel how and why ESG issues turn material over time and how stakeholders affect this "pathway to materiality", as they have framed it. As a result of their research, they propose a framework consisting of five stages: status quo, catalyst, stakeholder response, company response, and regulatory response.

According to Rogers & Serafeim (2019), initially companies may cause negative societal impact regarding an ESG issue which is still considered immaterial, as this impact is either not considered problematic under the prevailing societal norms or because the level of negative impact is not properly understood by companies themselves or the society at large. Even though there are negative effects, none of the industry members try to actively gain advantage over the others and cause further damage, essentially leaving the possible issue unnoticed. Within this framework, there are two ways (or *catalysts*) which may initiate the discussion for an ESG issue turning material. First, some companies may seek to gain excessive profits over their competitors and subsequently cause further

negative impact towards society or the environment. The excessive financial gains may entice the company or their competitors to further exploitations, eventually leading to the general public becoming aware of the issue. Alternately, the societal norms against which the acceptability of companies' operations is measured change, even when the companies do not change their operations themselves. This may be caused by the general public gaining access to information that either reveals questionable actions within an industry, or discloses how the previously accepted behavior is harmful to their surroundings. Despite the issue being recognized, it is still not considered material in this catalyst stage. In the next stage of stakeholder response, however, the issue may already turn material for companies which have gained excessive amount of negative publicity in the eyes of their stakeholders through their exploiting activities. Following this, even though the issue may be material only for certain actors, the whole industry may engage in attempts of self-regulation in order to limit the possibility of regulators taking further interest in the issue. If successful, the stakeholders may find the newly created practices acceptable, leading to a new balance in taking the issue into consideration. However, if the actions of companies are not seen adequate, regulatory bodies may start enforcing new laws to mitigate the negative impact, leading to the issue becoming financially material for the whole industry. (Rogers & Serafeim 2019.)

Summarizing the chapter thus far, ESG issues are related to companies' impact on environmental or social matters, as well as good corporate governance practices. Currently there is no widely accepted uniform standard for sustainability reporting which all companies within an industry are required to follow. While being widely utilized, the GRI standards have been criticized to provide excessive freedom for companies to apply the standard according to their judgment, whereas those by SASB are sometimes considered too narrow. With the digital age providing investors an overwhelming amount of data to analyze, ESG rating agencies have tried to provide assistance in bringing the data to understandable form for investors – but not without their problems, as the same companies may be rated as a leader in managing ESG issues by one agency, whereas another classifies them in a significantly lower rank. These differences highlight the current state of evaluating companies ESG performance, with each investor having their own emphasis on issues they find material (either from an economic, environmental or social viewpoint), and ESG rating agencies trying to cater for these different needs. Furthermore, as the framework by Rogers & Serafeim (2019) indicates, ESG issues are not an immutable

cluster: topics which are not considered to affect companies' performance today may be found material even on a short notice.

3.4 Corporate social responsibility

As companies have a significant role in making sustainability issues material through their operations, it is necessary to consider some theories related to corporate social responsibility (CSR) in addition to focusing on ESG issues. Notably, the framework by Rogers & Serafeim (2019) seems to incorporate notions especially from the stakeholder theory. As another example, Du & Xie (2021) have incorporated both stakeholder theory and institutional theory in the framework they presented in their recent work on AI ethics in consumer markets. Both will thus be shortly presented in the subsequent sections to clarify what drives companies to be good corporate citizens.

3.4.1 Stakeholder theory

Brought to the knowledge of wider audience by Edward Freeman in his notable book *Strategic Management: A Stakeholder Approach* in 1984, the stakeholder theory has been a central notion in business ethics and management literature ever since. Drawing from the work of several scholars of finance and business management studies, Freeman (2010) argued that in addition to the changing requirements of a company's internal stakeholders (e.g. owners and employees), the external changes by actors in the surrounding environment force companies to readjust their operations so that they can continue operating in the new unknown playing field. The theory is often compared to Friedman's statement about the company's sole social responsibility being maximizing the profits for its shareholders (Friedman 1970; Hill 2020, 29), with some stating that stakeholder theory was a response towards the shareholder centricity (Phillips 1997, 52). Freeman considered in his work that a company should aim to generate value to all of its stakeholders, consisting of "any group or individual who can affect or is affected by the achievement of the firm's objectives" (2010, 25), including those internal and external to a company. A simplified figure of stakeholder groups is presented in figure 3.

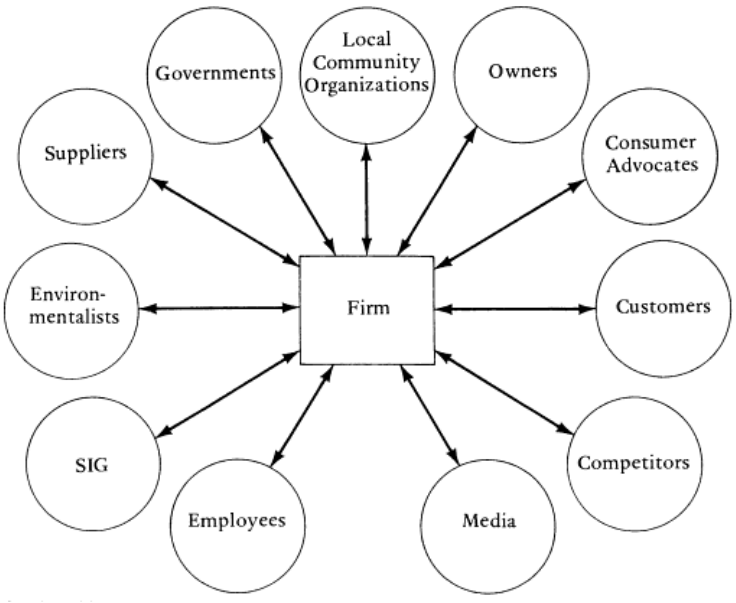

Figure 3 Stakeholder view of firm by Freeman (2010, 25)

Following the work of Freeman, several other scholars have continued after his lead, leading to a vast selection of academic contribution related to the subject. The growing interest has led to various terms surrounding stakeholder theory being defined and used in different ways, in addition to the stakeholder theory itself having developed to various directions, as Donaldson & Preston (1995, 65-66) and Hörisch et al. (2014, 329-330) have noted in their works. According to Donaldson & Preston (1995), stakeholder theory has been developed towards three directions, namely descriptive/empirical, instrumental, and normative theories, each taking a different aspect on what the theory means and how it should be utilized. The descriptive aspect considers stakeholder theory to contain managerial aspects, or recognizing which stakeholders are relevant for a company. The proponents of the instrumental aspect consider that by managing stakeholder relationships, companies may achieve some other benefits they seek. Lastly, the normative aspect looks at the moral standpoint and considers the ends of managing shareholder relations and means to achieve them. (Donaldson & Preston 1995, 70-71; Hörisch et al. 2014, 329-331.) Different normative cores from feminist ethics to Kantianism or stakeholder fairness, among others, have been adopted in the normative aspect (Phillips et al. 2003, 481). As the work of Freeman (2010) and Freeman et al. (2010, as cited in Hörisch et al. (2014)) aim to take an integrative view of the theory, from hereafter, the same viewpoint is adopted in this study as well.

An essential notion of the stakeholder theory is that it does not aim to explain how a company operates in solitude, but instead considers how a company interacts with its stakeholders. Hörisch et al. (2014, 330) use the notion "managing stakeholder relationships" in this context, but emphasize that unlike some have criticized (Gioia 1999), all stakeholders do not have to have an equal standing vis-à-vis the company they have a relationship with. They further state that the top management should instead evaluate the relevant stakeholder groups who are likely affected by certain operations of the company, in order to gain their insight and evaluate the possible consequences to their well-being. Stakeholder theory has also been subject to various other points of criticism, but as the theory is not the main subject of this study, they will not be considered here – for those interested, the work by Phillips et al. (2003) titled *What stakeholder theory is not* addresses these issues in more detail.

Building on stakeholder theory, companies benefit from greater interaction with their stakeholders as they gain a more thorough understanding of the surrounding norms against which their actions are measures against. Including these insights from various stakeholders and utilizing them in taking a proactive instead of a reactive approach to social responsibility issues has been found to be a source of competitive advantage for companies (Chang 2015). Despite this, the findings of Tamimi & Sebastianelli (2017) in their study of ESG disclosure transparency indicate that most companies still adopt a reactive standpoint to sustainability reporting, as the S&P 500 companies included in the study were found to be most transparent about governance issues, the only one that is regulated within the three ESG dimensions. Several other variables in addition to regulation have been argued to affect the transparency of a company, including the industry of a company and its board structure (Tamimi & Sebastianelli 2017), as well as its size (Tamimi & Sebastianelli 2017; Minutolo et al. 2019).

Considering the notions of stakeholder theory, the larger the company is, the more affected stakeholders it will have. Each stakeholder group will also have varying areas of interest, leading to a situation where the company is expected to publish more comprehensive ESG data for each of the dimensions. (Minutolo et al. 2019.) This is further highlighted with companies who operate in a global market, as stakeholders from different countries and legislations have different backgrounds against which they consider issues related to corporate responsibility (Cardoni et al. 2019, 16).

3.4.2 Institutional theory

Institutional theory presents that organizations do not act in isolation, but are a part of a regulatory field where they are expected to operate according to some generally accepted rules. Failing to do so would warrant distrust from the surrounding environment and thus lower the legitimacy for a given organization and subsequently worsen their position in their operating environment. Because of this possibility, institutional theorists state that it is in the best interest of organizations to follow the rules as is expected from them. (DiMaggio & Powell 1983; Scott 1987.) Suchman (1995, 589) adds that instead of conforming to the rules of the environment, organizations also have the option to choose an alternate environment, if they do not want to change their processes to fit the expectations which they are given.

The rules which companies are expected to follow are not limited to mandatory regulation only, but instead societal expectations and prevalent norms also play a part in forming an acceptable way of operating in a given environment. Because of this expectation that organizations should act in a certain way to ensure their legitimacy, it is inevitable that organizations eventually start to near each other, as all actors who are part of the same environment must follow the same set of rules. DiMaggio & Powell (1983, 150-154) describe that there are three sources which may motivate change in the environment, or lead to isomorphic change, as is described in their work. *Coercive isomorphism* is related to the regulatory and legitimization activities organizations are expected to follow. This pressure may be induced both by the surrounding society at large, the regulatory bodies in form of new regulation, or from other organizations who force others to follow the same practices before they accept them as their partners. *Mimetic isomorphism* refers to practices where organizations try to follow what others are doing, because they experience uncertainty in their environment and try to regain their balance by following a seemingly successful organization. Lastly, *normative isomorphism* is related to how professionals within organizations work together to form a generally agreed code of conduct for their group and thus a way to further legitimize themselves – although this might not be possible to be achieved fully, as different groups within and across organizations may have different needs that overlap with each other. (DiMaggio & Powell, 1983, 150-154.)

3.5 Taking responsible use of AI into account in ESG analysis

This section will bind together the topics that have been covered thus far in chapters 2 and 3. Chapter 2 introduced issues related to the ethical use of AI, including why and how the proliferation of AI technologies has led to new possibilities of ethical issues emerging, while in chapter 3, the differences between investing styles which take ESG issues into account in their investment decisions were introduced. Additionally, problems related to both the reporting of ESG issues and analyzing the published ESG data were covered. The concept of materiality both from a wider perspective (e.g. the GRI approach) and from a purely financial standpoint (e.g. the SASB approach) was also introduced in order to explain why companies and investors should be interested in said issues. Stakeholder theory and institutional theory were also examined to further explain the companies' motivations for taking ESG into consideration.

According to Freeman's stakeholder theory, both internal and external stakeholders can affect how organizations take business ethics into account in their operations. Stakeholders can support organizations which they believe to act in an acceptable manner based on the prevailing societal values, but also avoid being engaged with ethically questionable businesses, or even directly judge unethical business conduct. (Freeman 2010.) The materiality framework proposed by Rogers & Serafeim (2019) relies greatly on this theory, as stakeholders are argued to have a significant impact over which issues turn financially material for individual companies or even whole industries. Even if initially the most active stakeholders include merely activist groups or conscious customers, their judgment alone may cause a previously disregarded issue turning material in a short timeframe, if suitable amount of evidence against the alleged perpetrator is provided and wider audience gains access to the information. Stakeholders with regulatory power, such as governments, may also force an issue to become material for the whole industry at once through regulation, even if only a few companies cause damage through their actions. To avoid the possibility of financial losses due to overly strict regulation, companies may take a self-regulatory view towards an issue in order to assure that they act in an ethically sound manner even without such restrictions (Rogers & Serafeim 2019).

AI might bring a new type of issue to be addressed in ESG evaluations, as its disruptive decision-making power exceeds other digital technologies which have been in use for a longer period of time. With this companies may face new types of issues they have not had to consider previously, such as ensuring that decisions made by an AI are justified

or that the fair decisions do not come at a cost of privacy violations in excessive data gathering. Some of the use cases for AI include operating in environments with possibly detrimental consequences – a self-driving vehicle is not only a risk to its passenger, but to all surrounding vehicles or pedestrians as well. Furthermore, even companies who use an AI system which cannot cause damage may still be subject to fear or distrust from their stakeholders, making it necessary for them to provide assurance of their benevolent practices (Floridi et al. 2018, 691). Du & Xie (2021) have stated that companies which take AI ethics into account in their CSR activities may in fact gain from it through enhanced stakeholder relationships. Their framework suggests that by engaging in these activities companies generate better outcomes for their customers (e.g. ensuring privacy or user autonomy) or even societies at large (e.g. ensuring fair outcomes from their products) and are thus subject to increased support from their stakeholders. This in turn leads to positive outcomes for the companies themselves, e.g. in forms of improved brand images and increased company values. (Du & Xie 2021, 971).

In addition to stakeholder theory, Du & Xie also draw upon institutional theory, which proposes that companies seek legitimacy by following the prevalent regulation, societal norms, the expectations of their peers, or other expectations which are presented to them (DiMaggio & Powell 1983; Suchman 1985; Scott 1987). Du & Xie (2019, 969) thus argue that companies engaged with AI must take their stakeholders' expectations into account, but also conform to the rules set by their institutional environment: by following these practices, companies can legitimize their actions in the eyes of their stakeholders and increase the trust of their customers toward their AI products. However, the suitable level of reporting also depends largely on the type of AI product that the company utilizes, as the level of risk varies. (Du & Xie 2021, 963-965.)

In addition to the need of companies reporting on their use of AI, the question of how investors will be able to evaluate these attempts from the ESG standpoint arises. There seems to be little to no research on how the use of AI should be taken into account in ESG evaluations, which in itself highlights the novelty of the topic. Questions like are the current ESG rating frameworks of ratings agencies suitable for evaluating effects of AI, or how investors can or if they even should take responsible use of AI into consideration in their investment analyses remain largely unanswered. Among the first academic contributions to this topic, Brusseau (2021) has recently criticized the use of the current ESG rating methods for evaluating the effects of AI. He forms his statement on the basis that traditionally, ESG issues have been related to larger targets like ensuring that all

employees are treated in a fair manner. Instead, Brusseau argues that the main issue of AI is related to data ownership, or how companies use individuals' data – for example, whether the use of AI leads to our greater benefit or limits our self-determination (Brusseau 2021, 1-2). Based on this setting, Brusseau proposes an alternate *AI human impact* model for evaluating AI companies. Instead of adapting the existing ESG frameworks to AI, this model utilizes a set of AI principles to emphasize issues of AI, and assigns scores from 0 to 2 to each principle based on how well a company takes the related issues into consideration. (Brusseau 2021.)

Whether AI will be included in ESG related evaluations by applying the existing rating methods to it or by creating a completely new method for this particular topic like Brusseau (2021) suggests, the potential negative effects of AI must be accounted for in the future. Leaving it solely to companies' discretion to take these issues into account could enable malpractice, as the findings of Tamimi & Sebastianelli (2017) indicate that companies are more prone to take a reactive approach to sustainability issues, rather than a proactive one. Principles of responsible AI, which have increased in numbers rather rapidly during the last years – which has already led to the emergence of analyses of the principles, like those by Jobin et al. (2019), Hagendorff (2020) or Schiff et al. (2021) – could prove useful for evaluating the impact of AI in investment analyses as well. Despite the principles being a new addition to the discussion over AI in general, a consensus of the most central topics to be included in them can already be seen (Jobin et al. 2019, 391). While the material issues for each AI system may differ, such a core set of principles could act as baseline of matters which companies should take into account. While it is not within the scope of this study to validate a decisive set of such principles, perhaps the currently most often mentioned principles, such as transparency, justice, non-maleficence, responsibility, privacy and beneficence presented earlier in this study, could be the ones which investors will start incorporating in their analyzes in the future in one way or another.

4 METHODOLOGY

4.1 Overview

This study set out to investigate how investors currently understand questions related to the responsible use of AI, as well as how this area might develop in the future. In order to gain answers to these questions, it was considered necessary to collect insights from professionals who are ideally knowledgeable in both of the major topics of this study (responsible AI and ESG investing), as the topic is still in early stages at least in Finland and existing materials on how investors take these issues into account in their work is not readily available. Furthermore, as this area is still lacking in the academic literature, conducting an exploratory study to unravel how the professionals perceive the related questions was considered a suitable approach.

Research methodology is a way to solve a specified research problem in a systematic manner, and it goes beyond choosing suitable research methods to understanding why certain methods have been chosen for the research context (Kothari 2004, 8). According to Tan (2017, 4), methodology can be roughly said to consist of philosophical implications, research design, and chosen methods for conducting the research. The first section of the chapter will focus on the philosophical underpinnings of this study as a basis for the other aspects of methodology, which will be introduced in the following sections in more detail.

Tan (2017, 4) claims that there are two types of philosophies of science, both of which can be further divided to different variants. As the name implies, *causal science* is concerned with causes and effects of events or actions and aims to explain why a certain consequence follows from a certain incident. In addition to studying the simple relation of two events, causal scientists are often interested in how the relation actually affects the two sides. Finding accurate causality may however prove to be difficult, as complex environments may cause two events to be seemingly correlated, while they are in fact both affected by some external effect. The emphasis of causal science is often on testing hypothesis, and causal scientists see that there is an objective truth to be discovered through logical reasoning. (Tan 2017, 5-7.)

Interpretive science forms the other major strand of research philosophy. Interpretive scientists consider that individuals form their own perceptions of the world, effectively rendering one true description of an event impossible, as each participant has their own

view of how some event has unfolded (Collingwood 1946 and Taylor 1971, as cited in Tan 2017, 8). Unlike causal science, where researchers see that there is an objective reality to be found, interpretive science sees reality as a subjective matter which changes according to who describes it. Interpretive science is also considered to emphasize discovery of new findings through explorative frameworks, as opposed to discovering relationships through hypothesis in causal science. The researcher analyzes the data with the support of the framework to discover underlying reasons for the informants' claims, while staying aware that each informant can only tell a story from his or her own perspective, with possible underlying motives for relaying the story in a certain way. (Tan 2017, 8-9.) Because this study is conducted with the purpose of exploring a subject which to the author's knowledge is not yet widely researched, which makes it impossible to form a hypothesis and test for possible relations e.g. between the two main topics, assuming an interpretative approach to it is a natural choice.

4.2 Research design

Ethics of AI, the principles of responsible AI, as well as ESG investing have all been receiving an increasing amount of attention in the academic world during the last decade. Despite attention being turned towards these topics, very little research has been conducted regarding their intersection (as an exception to this, see e.g. Brusseau 2021). The purpose of this study is thus to investigate whether there are some preexisting connections between these two topics – for example, whether the principles of responsible AI are incorporated in ESG analyses in some way – and how investors and other experts who are knowledgeable in ESG investing generally consider the possible issues AI may cause to various stakeholders currently in their work. Because of the largely unexamined nature of the topic, it was natural to assume an exploratory research design for this topic.

An exploratory design is considered a suitable approach when the researched topic is not understood well, even to a point where the nature of the problem may well be clarified as the researcher conducts her studies, leading the research to a different direction than originally anticipated (Ghauri & Grønhaug 2010, 56). Due to the nature of exploratory studies, it may even be possible that the outcome of the research is that there are no meaningful research possibilities in the area which the researcher initially focused on (Jaeger & Halliday 1998, 64). To take this possibility into account, calls for flexibility and open-mindedness in an exploratory data gathering process have been made (Stebbins 2011, 5). In his widely cited work *Exploratory Research in the Social Sciences*, Stebbins has

provided an explanation for why researchers engage in exploratory work, stating that "[r]esearchers explore when they have little or no scientific knowledge about the group, process, activity, or situation they want to examine but nevertheless *have reason to believe it contains elements worth discovering*" (Stebbins 2011, 5; emphasis added by the author of this study). The increasing awareness towards the possible negative impact of ungoverned AI and the need for taking ESG issues into account in investment decisions (which has already led to the investment style becoming mainstream among the largest investors in the world) provide a reason to believe that investors could benefit from taking ethics of AI into account in their work. Furthermore, the principles of AI can possibly provide a clear framework for companies to develop their use of AI to a more responsible state, which could also make them a useful topic to consider in ESG evaluations. Even if this is not evident yet, as both main topics of this study can be assumed to become more important in the following years, this topic was indeed considered worth investigating, even with the lack of prior research to prove its relevance.

4.3 Methods

When considering the overall research methodology, an aspect which will have a large influence over the overall design of a study is the choice between using qualitative or quantitative research methods, or a mixture of both. Both Ghauri & Grønhaug (2010, 104-105) and Silverman (2010, 8-9) emphasize that choosing between the two is not a matter of choosing the superior methods of the two alternatives, but rather adopting an approach which best suits the current research problem, allowing the researcher to arrive at meaningful conclusions. As the name indicates, quantitative research is related to drawing conclusions usually from a broader dataset through measurement. The researcher is generally considered to look at the data from an outside perspective, typically with the motive of testing a hypothesis based on prior knowledge. Contrary to this, qualitative methods are considered to be a better fit for research where new information is sought, and where the viewpoint of informants is emphasized. Additionally, the researcher has a deeper engagement with the data though an insider perspective. (Reichardt & Cook 1979, as referred to in Ghauri & Grønhaug 2010, 105; Ghauri & Grønhaug 2010, 104-105).

Stebbins (2011, 5-6) has described that exploratory studies can benefit from both qualitative and quantitative research methods, depending on the level of prior information available for the phenomenon being focused on. Areas with little or partial information generally benefit more from qualitative methods, whereas quantitative methods may

become more beneficial as initial research has been conducted, and some level of knowledge has been achieved. For example, quantitative methods may prove useful in larger exploratory research projects where researchers want to confirm the findings of the initial stages of their studies. (Stebbins 2011, 5-6.) However, because the scope of this study is limited to only forming the initial picture of how the responsible use of AI and ESG investing may fit together, adopting a purely qualitative approach can be considered an acceptable choice. Indeed, it has been stated that "[i]t is generally accepted that, for inductive and exploratory research, qualitative methods are most useful, as they can lead us to hypothesis building and explanations" (Ghauri & Grønhaug 2010, 106), providing further support for choosing a qualitative approach.

4.3.1 Data collection

Interviews were chosen as the data collection method for this study. While the quality of interview data may be heavily reliant on the interviewer's skills in gathering relevant information during the interviews, and analyzing data from interviews may also be subject to the researcher's subjectivity in interpretation, it is possible to gain new insights from informants who are experts in the studied subject (Ghauri & Grønhaug 2010, 126-127). There are three types of interviews the interviewer may rely to. First, structured interviews consist of a predefined list of questions which is the same for all informants, and is presented in exactly the same structure to allow the use of statistical methods in the analysis stage. Surveys are a common method of conducting a structured interview. On the other end of the spectrum are unstructured interviews, which provide little to no guidance for the interview beforehand, but instead the informant can freely talk about matters related to the overall topic that the interviewer provides. The role of the interviewer may be more demanding here during the interview itself, as she must recognize which statements from the interviewee need further questioning. This does not mean that structured interviews are easier overall, as they require the researcher to define a list of relevant questions beforehand, which usually needs to be piloted to ensure fit for the actual interviews, as the questions cannot be altered afterwards. The third type of interview is a mixture of both of the previous and is accordingly called a semi-structured interview. This type is conducted with a list of predefined questions, but the interview does not have to follow the exact structure: the questions do not have to be answered in a certain order, and some question can even be omitted during the interview. This gives the interviewer a certain degree of freedom, but also responsibility for ensuring that the collected data

will still contain relevant insights despite the possibly varying questions in interviews. (Ghauri & Grønhaug 2010, 125-127; Tan 2017, 84-85)

Semi-structured interviews were selected as the data collection method for this study, as the possibility to create questions beforehand allows the researcher to ensure that certain relevant topics will be covered during the interviews. At the same time, it does not limit the interview to a set of predefined questions that the researcher has formulated. This was considered especially important, as this study is explorative in nature: by utilizing a strict list of questions, interesting viewpoints which the researcher has not thought of could end up missing from the data, thus limiting the exploratory aspect of how the investment world currently perceives the responsible use of AI. It was still necessary to include a few more specific questions related to the principles of responsible AI, mainly to ensure that they would be covered in the interviews even if they are not currently considered by investors, thus rendering the adoption of unstructured approach difficult.

As this study set out to find how investors currently consider the responsible use of AI in their investment analysis, and whether the principles of responsible AI have or could have an impact on the results of the analyses, collecting insights from professionals of these topics was deemed a suitable approach. It would be ideal if the informants had comprehensive knowledge of both topics, especially for the purposes of finding out how the principles could benefit the investment analysis in the future (as it was expected that they would not be utilized yet, due to the relatively recent emergence of the principles). However, gaining access to such informants was expected to be very difficult, as there was no certainty whether such professionals could even be currently found. To factor in this possibility, a wider selection of possible candidates was formed. The limited timeframe for conducting the study was also recognized as a potential issue, further justifying having a larger scope for the potential sample. A total of 28 potential informants were contacted, resulting in five interviews. Furthermore, six additional replies stating that the contacted person would not participate in the study was received. It is worthwhile to note that only one of these replies was ascribed to busy schedule, as the others stated that the related topics had not been considered in their organizations yet. Summary of the informants can be seen in table 3.

Table 3 Informants of the study

Participant	Job title/focus	Organization focus	Interview date	Interview length
P1	CEO	AI products	30.3.2021	52 min
P2	CEO	AI products	19.4.2021	47 min
P3	Responsible investment	Banking	5.5.2021	52 min
P4	Responsible investment	Pension insurance	12.5.2021	48 min
P5	Responsible investment	Asset management	11.6.2021	34 min

All informants have some level of knowledge related to ESG, with the three informants from the investor side being specialists of the topic in their respective organizations, having gained years of experience related to the topic through their work. The specific job titles were removed for the three investor side participants to help ensure that their identities could not be derived from the presented information. While all interviewees were aware of the potential risks that AI could contain, the two first interviewees were more familiar with the topic of responsible AI and its principles, as these topics have emerged in their own work in the field of AI. All interviewees were given an overall description of the study through email along with the invitation to participate to the study. Apart from the first interview, which was partially considered as a test round for the interview questions, all interviewees were also sent the interview structure beforehand, so that they could familiarize themselves with the questions if they so wished. In addition to these questions, follow-up questions were also presented to gain further understanding of the interviewees' statements, as is common with semi-structured interviews. All interviews were conducted in Finnish, which was the shared native language of all of the participants, as this allows participants to express themselves efficiently with their primary language and counter possible misunderstandings due to language barriers. All quotes from the participants which are presented in chapter 5 were translated afterwards from Finnish to English by the researcher. The translated set of initial questions can be found in appendix 1. The interviews were conducted via Microsoft Teams, and all were also recorded with the interviewees' consent. The recordings were transcribed soon after the interviews (i.e., within one day) by the researcher to allow forming initial understanding of the data already during the data collection process, and the transcriptions were later used to further analyze the data set for the purposes of this study.

4.3.2 Data analysis

The selected method for data analysis has a decisive influence over the results that will be formulated. In this study, thematic analysis was selected for analyzing the data from the semi-structured interviews. According to Braun & Clarke (2006, 77-78), thematic analysis is a widely used term for several different types of qualitative analysis methods but may still often be disregarded as its own distinct form of analysis. They see it as a foundational method for qualitative analysis which benefits especially from its flexibility to suit a range of different methodological positions, providing a specified frame for qualitative analysis and for reporting the results, thus enhancing the possibility to evaluate the research findings later on. Furthermore, its purpose as a form of analysis is to identify patterns or themes within the data, organize them and clearly report them to the readers of the study report (Braun & Clarke 2006, 79-80).

Braun & Clarke (2006, 82) have stated that while ideally a possible theme would be emergent in multiple places of the data (e.g. have multiple similar codes which could be arranged to a theme), this is not a requirement in thematic analysis. As they have phrased it, "[a] theme might be given considerable space in some data items, and little or none in others, or it might appear in relatively little of the data set" (Braun & Clarke 2006, 82). The researcher should thus consider which aspects of the data provide suitable answers to the research questions of the study, rather than simply lean on quantitative measures. This provides a clear distinction to other commonly used qualitative methods for analysis, such as content analysis, which relies more clearly on findings themes or categories with a high number of occurrences (Vaismoradi et al. 2013, 401). However, simply considering the number of occurrences has been criticized for possibly taking the data out of context, as the same type of codes may be emergent in the data for very different reasons. Additionally, some themes may be more emergent simply because informants are more comfortable with certain topics, thus leading them to be willing to provide more information about them in interviews (Shields & Twycross 2008, 38; Vaismoradi et al. 2013, 401). As this study touches two relatively young fields which are still formulating towards mainstream adoption (especially in the case of responsible use of AI) and different informants may have varying viewpoints and levels of knowledge of the topics, considering the context of certain comments is necessary, supporting the adoption of thematic analysis instead of e.g. content analysis.

In addition to considering which themes are suitable for answering a particular research question, the researcher must decide whether the goal of the analysis is to arrive at a broad description of all of the relevant themes in the data set, or to focus on a smaller sample of particularly interesting themes (Braun & Clarke 2006, 83). While Braun & Clarke (2006, 83) have stated that the latter option would generally provide a more detailed description of certain topics and that the researcher is bound to lose some depth in the findings with the former approach, for this study, taking a broader viewpoint to the data analysis is a suitable option. This can be mainly justified with the exploratory nature of the study, as analyzing which issues or aspects are emergent in a new topic overall serves the purposes of this study the best: considering only a limited set of themes, albeit in more detail, could lead to important aspects being ignored.

Furthermore, researchers must choose whether they will take an inductive, abductive or deductive approach to building theories in research projects. These approaches are related to how the researcher eventually arrives at some conclusions, or what do they base their claims on. As Ghauri & Grønhaug (2010, 15) have presented, induction is primarily based on empirical observations, which are augmented with prior findings as the research advances. The findings act as support for the observations, and together they will be used to develop new theories. It should be noted that inductive conclusions can never be considered absolutely certain, as they rely heavily on empirical data and the abilities of the researcher to draw meaningful insights together. Rather than arriving at certain conclusions, the researcher can state that they are probable, given all the elements which were gathered during the research process. (Ghauri & Grønhaug 2010, 15) Contrary to induction, deduction uses prior findings in existing literature as their base, forming hypotheses which they then methodologically aim to prove through testing. Once established, the existing base thus affects the rest of the research process heavily. Whereas inductive reasoning may be beneficial for forming developing new theories, deduction is often considered to be better suited for confirming the validity of theories. (Ghauri & Grønhaug 2010, 15-16.) Lastly, building theories through abductive reasoning is based on identifying some surprising evidence from the data, for which new concepts can be formed with the support of the previous theories. Abduction is thus notably different from both induction and deduction, as instead of generalizing a new theory from observations or being tied to previous findings, the theorizing begins from the gathered evidence and trying to provide an explanation for said evidence. (Timmermans & Tavory 2012, 167-169, 180.) In this study, inductive reasoning was largely used during the initial data analysis and building

of themes. As the analysis proceeded, the research moved towards a more abductive stance, since at this stage it was possible to connect and compare the discovered findings to previous literature (Alvesson & Kärreman, 2007; Gioia et al., 2012, 21).

Thematic analysis consists of six stages, which were mostly followed as described in this study as well. First, the researcher familiarizes herself with the data set by transcribing the data when needed, and forms initial ideas of the content by reading through it several times. This initial stage forms a baseline for the coding process, which will be conducted in the following stages. In the second stage, the researcher forms initial codes from the data from content that seem interesting for the research process. (Braun & Clarke 2006, 87-88.) In the third stage, the researcher will start searching for themes among the codes which were formed in the previous stage. At this stage, it is not required to gather decisive themes yet, but instead check how the codes could be organized in suitable groups. At stage four, the researcher reviews the initial themes and checks whether they seem suitable for both the coded extracts and the larger data set, essentially requiring analyzing the themes on two levels. As the result of this stage, a thematic map of the data set is formed. In the fifth stage the themes are further defined and named appropriately. The researcher should ensure that each theme serves a purpose for the analysis, and whether its content is suitable for answering the research question(s). Lastly, stage six consists of selecting relevant data extracts for the final report and ensuring one more time that all themes and their contents suit the purpose of the study well. Additionally, the researcher will write a report of their analysis. (Braun & Clarke 2006, 87, 89-93.) One exception to the described process was made regarding the data extracts: many of the extracts in chapter 5 were already initially chosen during the coding process, as they seemed to convey a viewpoint which could potentially answer the research questions of this study. Their fit for the overall study was still confirmed during the last stage.

Summarizing the methodological choices, this study adopts interpretive underpinnings and a qualitative approach to conducting research. As it can be expected that this intersection might still be relatively new or even unknown in the investment world, comments from each expert are considered to be their viewpoints of the matter, with no underlying general truth to it. As an exploratory study, it aims to formulate the previously uncharted topic of the intersection of responsible AI and ESG investing by utilizing insights from semi-structured interviews with experts of the two fields. The collected data set is analyzed through a thematic analysis following an abductive approach, with the aim

of collecting a broad overview of the emergent themes related to the research questions of this study.

4.4 Evaluation of trustworthiness and research ethics

4.4.1 Trustworthiness

To ensure high quality in research, methods for evaluating both qualitative and quantitative studies have been developed. In quantitative studies, the two main perspectives to be evaluated are rigor and validity, whereas in the case of qualitative studies, trustworthiness and credibility are considered in order to recognize the strengths and weaknesses of studies and communicate both truthfully to the reader of the study report. (Cope 2014, 89.) The work of Lincoln & Guba (1985) is perhaps the most widely recognized set of criteria for evaluating qualitative research and will thus be used as the guideline for evaluating the trustworthiness of this study as well. Their criteria draw from the preceding work of Guba (1981), who considered the trustworthiness of qualitative research by identifying four criterion which should be considered in both qualitative and quantitative studies. They include truth value, applicability, consistency, and neutrality, which are commonly titled as credibility, transferability, dependability, and confirmability in qualitative studies. (Guba 1981, as cited in Krefting 1991, 215; Lincoln & Guba 1985, 294-301)

Credibility refers to whether the research findings are reliably drawn from the data set of the study, and whether the findings can thus be considered a truthful depiction of reality (Lincoln & Guba 1985, 296). As qualitative studies often consider there to be multiple realities based on the different viewpoints of informants, the challenge of the researcher is to report the findings in a manner which accurately reflects the viewpoints (Krefting 1991, 215-126). Lincoln & Guba (1985, 301-307) have suggested several methods for researchers to increase the credibility of their studies, such as prolonged engagement with the studied phenomenon or the informants during data gathering, persistent observation of the aspects which were considered prominent during the prolonged engagement, and triangulation of data, or cross-checking the findings either from various sources or with another researcher. In this study, prolonged engagement was sought by conducting lengthy interviews with the informants. Each interview contained questions related to how investors perceive AI as an investment target in general, but more detailed questions related to how responsible use of AI should be visible in ESG analyses in the future (e.g. will the principles of responsible AI have a role in them) were also asked to

gain a thorough understanding of how the investment world might perceive questions related to the responsible use of AI. The collected data from the interviews was transcribed personally so that familiarization with the data could be started already during the data collection process, and the transcriptions were read several times to allow initial ideas to be formed from the data set. The coding process enabled detailed observations to be drawn from the data set, and forming themes based on the created codes helped to form a coherent entity for the purposes of this study. Lastly, triangulation was included by utilizing both theoretical and empirical findings when forming the results, as well as including experts with different viewpoints through their different professional positions to ensure that the phenomenon could be examined as thoroughly as possible within the limited time frame of the study.

Transferability refers to whether the research findings could be applied to other settings as well and is generally enhanced by providing a description of the research setting, e.g. details regarding the interviewed informants. Qualitative research is often associated with poor generalizability due to the small samples utilized in studies, leading to questions regarding whether the results are bound to only the one specific setting in which they have been discovered. However, providing a detailed description of the research context allows the readers of the study to make their own judgments about whether the findings of the study can be utilized in other research settings as well – for example, as a comparison to findings of a similar study. (Lincoln & Guba 1985, 297-298; Shenton 2004, 69-70.) In this study, transferability is ensured by providing a description of the entire empirical data collection and data analysis processes, including information regarding the informants and the data collection settings.

Dependability is related to the reliability of research and is thus closely connected to the credibility criterion. It generally refers to providing sufficient information regarding the research design and strategy to demonstrate that potential unreliability has been taken into account during the research process. (Lincoln & Guba 1985, 298-299, 318; Krefting 1991, 216) Compared to quantitative research, which is intrinsically conducted in a more controlled environment and is thus more easily repeated afterwards, qualitative research often includes variability due to the human element that is heavily present in the studies, as both the researcher's own judgment and the views of the informants are subject to change. While it is not desirable in qualitative research to control the research design in a rigorous manner to achieve consistency, dependability can be increased by explaining the sources of variability which have been encountered during the research. (Krefting

1991, 216.) A thorough description of the research process from data gathering to interpretation has been provided to facilitate this. Additionally, both the research plan and the final results from the study have been presented to the supervisor of this thesis who has not been a part of conducting the actual research, which according to Krefting (1991, 221) can also increase dependability.

Confirmability refers to how well the research findings can be considered to be drawn from the dataset. It generally contains the idea that the findings should not result from the researcher's biased views, but be a true representation of the informants' statements – or whether the results can indeed be confirmed from the underlying data (Lincoln & Guba 1985, 299-300). Using the previous work of Halpern as a baseline, Lincoln & Guba list six categories and five audit trail steps for providing a comprehensive audit trail, which could be used to increase the confirmability of a study (Halpern 1983, as cited in Lincoln & Guba 1985, 319-320; Lincoln & Guba 1985, 319-320). Due to the time constraints and a full audit performed by an external party being considered as excessive for a thesis study, confirmability was sought mainly with triangulation of theory and collected data, which can be used to provide multiple perspectives to a topic (Krefting 1991, 221).

4.4.2 Research ethics

Including ethical considerations to the research process is an integral part of conducting research. A researcher has the responsibility of reporting their methods and processes accurately, so that potential weaknesses of the research design and subsequently the findings and conclusions can be recognized. Ghauri & Grønhaug (2010, 20) describe that this is particularly important due to the possibility that the readers can take what the researcher has told as a given, without giving much consideration of how the results may be faulty. The researchers' own bias, sponsor of a research project, peer pressure, and several other factors may well influence how the researcher presents their findings, making it also the researcher's responsibility to account for these possible issues in order to ensure that the presented results will be an accurate and objective representation of the gathered data. Furthermore, it is the researcher's responsibility to also guarantee that possible participants of a study are not mislead about the purposes of the study, that the data they provide is handled according to the agreement made with the participant, and that the participants are not forced to provide information they are not willing to give. Researchers may struggle with some of these notions, as for example, it may be tempting to leave some details

off from an interview invitation in order to ensure a higher positive response rate. (Ghauri & Grønhaug 2010, 20-24.)

While it was acknowledged from the beginning that the novelty of the topic may lead to some potential informants being unwilling to take part in the study due to lack of knowledge related to the topic, the purpose of the study was openly described already in the initial emails sent to the potential participants, so that they could freely choose to take part in the study. The informants were also encouraged to ask further details related to the study whenever needed during the research process, and all informants were given the option to withdraw from the study at any given time without having to provide an explanation for making this choice. Furthermore, a permission for recording the interviews was asked at the beginning of each session. All recordings as well as transcripts derived from them were permanently deleted after the study had been completed.

One important ethical consideration for this study stems from the topic being given to the researcher as an assignment from a research project centered around AI governance. The interest of the project could thus have posed a risk for affecting the outcome of this study, although objectivity was strived for at every step of the process. The informants were also informed of the study being an assignment for the project. The researcher was not an official member of the project and there was no connection between the researcher and the informants prior to sending them invitations for the study. There was thus no personal interest involved to produce results which could be considered favorable to any party involved in the process.

5 FINDINGS

5.1 Overview

Analyzing the data from the conducted interviews resulted in two major themes being discovered. *Understanding AI* contains views related to how the market currently understands the role of AI both in companies, but also in the investors' own work, and what AI even is on a basic level. *Measuring impact of AI* contains findings related to how the impact of AI is currently considered, or what kinds of elements would need to be taken into account in the future when it comes to analyzing the impact of AI a company uses in their operations. A depiction of the findings is provided in figure 4, where the discovered themes belonging under the two main themes have been presented. To provide further clarification, some of the central topics which were discussed during the interviews regarding the themes under *Understanding AI* have been included. It should be noted that despite the themes having been presented separately, it does not mean that they would be isolated from each other – for example, the lack of knowledge related to AI likely affects how the impact of AI is currently being considered in ESG evaluations.

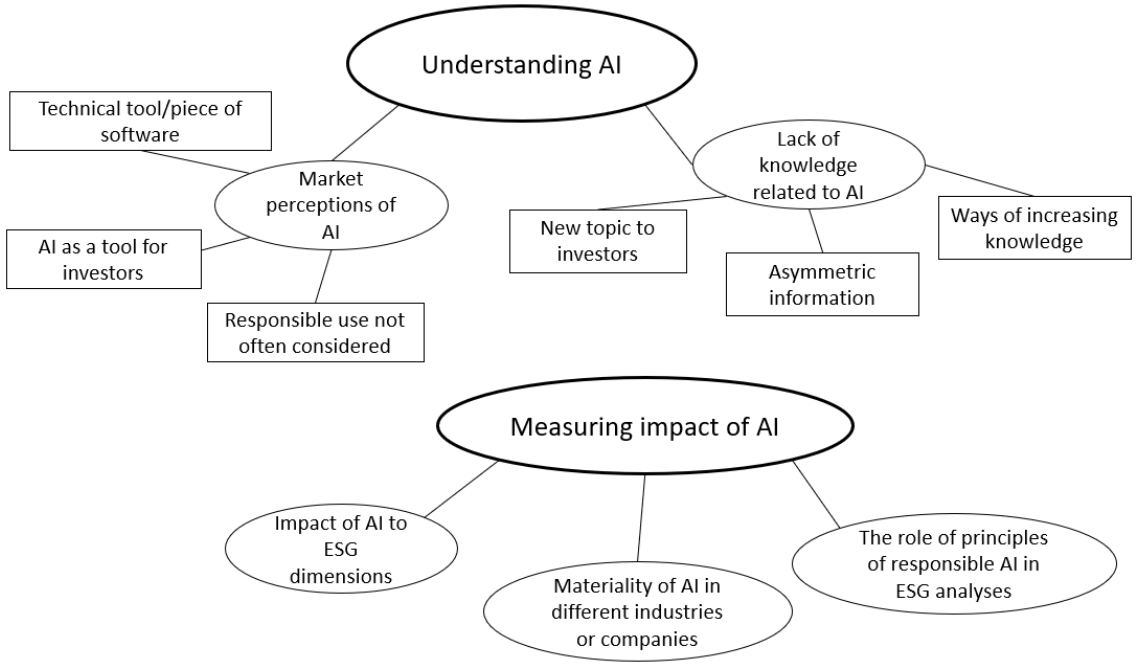

Figure 4 Thematic map of the findings

These two main themes will be presented in more detail in the subsequent sections by presenting the five themes which belong under them. Since this is a qualitative study,

extractions from the conducted interviews will be used to present the professionals' statements of the topic so that the insight they have provided can be properly transmitted. It should still be noted that while all extractions have been translated from Finnish to English by the researcher with the intention of preserving the original meaning as accurately as possible, translating data may still lead to some of the underlying meaning being altered.

5.2 Understanding AI

This theme is related to how the interviewed professionals considered the market to currently understand AI as a technology (e.g. what is AI, or what is it used for), as well as the apparent lack of knowledge related to AI in general on the investor side. It seems that as of now, there is not only little information available which would help investors understand the technology better, but also an insufficient amount of driving forces which could bring the topic of responsible AI into mainstream awareness.

5.2.1 Market perceptions related to AI

When asked how the interviewees saw the market deal with questions related to AI in general, it became evident that at least at this stage, there is a heavy focus on the potential efficiency gains that could be achieved with AI. This was stated to be the case for both companies considering whether they should invest in related technologies, as well as for investors. The general consensus of AI was thus seen to be related to the benefits that it can help achieve, and automated data processing and overall efficiency gains on operations were considered as central topics at the moment. However, another common theme among the interviewees was that they also considered that there is a lack of understanding related to AI in general: while the potential benefits may be easier to understand, understanding the risks related to the use of AI was considered difficult. Additionally, it was brought up that there is still a lot of misunderstandings related to what AI actually is or what it can be used for on a basic level.

> *I think there is a lot of mysticism related to AI. On a general level, the market might be on alert in a way, and if you have a general discussion related to AI as an investment, people do not really understand what it is really about. (P1)*

Understanding the risks is surely more challenging compared to [understanding] the possibilities – – This is not typical and understanding what we are talking about is indeed a challenge. (P4)

It was also stated that for investors, AI is currently considered more as a tool for their own work or for companies in general rather than something that should be taken into account in investment decisions for its possible risks. AI was found especially useful for collecting and analyzing scattered ESG data, which could be difficult for an individual analyst or portfolio manager to process efficiently – although this emphasis could be explained by the interviewees from the investor side being ESG specialists within their organizations. Furthermore, AI was considered an asset for both investors themselves as well as ESG data providers, who use it for gathering and analyzing large quantities of ESG data. AI tools were found to help individual investors internalize the relevant data points better and thus make more justified investment decisions, as unnecessary noise from the larger data set had already been filtered out by AI and compressed to a more easily manageable amount of data to look through.

Instead of the investor reading a hundred pages and finding a few relevant data points, they can get these data points very fast, so that their time is freed to the actual analysis – – I think this is what the investor side is currently talking about [related to AI], instead of thinking the AI that companies use or its impact on responsibility. (P3)

All interviewees found that the responsible use of AI and ethics of AI are still not everyday topics in the work of investors, and the risks of AI were not considered as topics which would be commonly included in ESG evaluations. This was considered to hold true especially when evaluating companies where AI is not the main product, and its use is thus not always clearly brought up by companies. However, the interviewees stated that for AI intensive companies, the potential AI related risks could be evaluated on a case-by-case basis.

If you have a company where it [AI] is a core business, then of course you'd think more about what they are doing, what are the risks and the possibilities – – when it is clear that this is what we are dealing with, then it is at least somehow manageable, but if it [AI] is just one way of doing things [within a company's operations], then it is very challenging to grasp it from the outside. (P4)

> *And of course where AI is in a central role for business, it has been raised as the most important risk from the responsibility aspect, and it also has a large impact for the score that a company received, how the company has tried to mitigate the risk and how they have succeeded. (P5)*

As for the responsibility aspect of using AI, the interviewees provided several insights on what investors could consider as being responsible. Several interviewees mentioned the traditional view of ESG or CSR, stating that being responsible does not equal to merely following the law, but instead responsible actors go above the bare minimum that is required of them. Likewise, multiple interviewees indicated that the type of AI system in use would have an effect on this question, as the material issues would be different based on how or on what purposes a company uses it. P3 also brought up the idea that AI may in fact require us to look at this question from a different angle, as there might be new types of impacts from using AI which we cannot yet specify, as this is a new type of tool which we do not necessarily fully understand yet.

> *All of the current consideration is related to our current way of working and to our current worldview. And if we consider AI, then maybe we should be able to look at it from a different point of view, like what AI is used for and what using AI has impact on. What causes the positive and negative consequences, and with that we could try thinking it through the current ESG frameworks and think what kind of social impact, what kind of governance impact it has, but I also feel like it may also have different kinds of impact which we are not able to evaluate or think about at this stage. And when we get to know this topic more closely, when we start to see that the financial impact is also increasing and its [AI's] significance increases in business, we start to understand the effects in a way which we can also pay attention to. (P3)*

With regulation being considered being too far behind for leading the topic of responsible AI relevant in the market overall, other stakeholders were considered to be the driving force for bringing responsible use of AI emergent to wider audiences. During the first interview, stakeholders who were mentioned as having a possible effect in bringing these topics to companies' awareness included both consumers and investors, as well as other companies who act as customers or partners within company networks. Here it was stated that as especially companies who compete in different markets with their products and may face intense pressure due to their ethical AI considerations, trusting that

consumers or other customers will choose the more responsible option from the available alternatives was seen as important incentive for companies to take these issues properly into account.

> *But it is more about how responsible the consumer is, whether you can trust that [the consumer will make the responsible choice], through that we can make these principles transparent. – – So the choices of consumers, the choices of companies, and the choices of investors will perhaps lead the pace. (P1)*

In addition to acting as an incentive for enticing potential customers, both partner companies and competitors were also seen as actors who can encourage or even force others to join them in taking the responsible use of AI into account.

> *I think that the development will also be company driven, that the understanding rises as the actions of competitors will be seen and it will be noticed that these topics are something which should be included in CSR reports, or that at least the competitors are doing so. And in a way they should through this be able to form a clear link between these issues and corporate responsibility. (P2)*

> *The more business is done in different ecosystems, the more actors there are who will clearly show more effort than others emerge. But they also set criteria, like "if you intend to play with us, you must also be transparent, compatible [with us]". So in a way, if someone puts on an effort in an ecosystem and is a competent, this will pull others with these principles as well. (P1)*

5.2.2 Lack of knowledge related to AI

In addition to the level of knowledge related to responsible use of AI being considered to vary largely between individual investors, the interviewees also perceived that overall the level of understanding AI as a technology is still lacking. Understanding AI was considered to require a new type of expertise from the investors, as the topic is still relatively young in a commercial setting. One interviewee also highlighted that the varying terminology surrounding the field of AI had made internalizing the topic more difficult, while another also pointed out that it can be difficult to comprehend what kind of technology is even used in a given situation, as the term AI could be coined to different systems ranging from relatively simple algorithms to complex black-box machine learning models.

Related to this, questioning what type of AI the investment asset has in use was said to emerge mainly if a certain type of AI has caused issues and this has become public.

> *I think people pay attention to it when an issue comes out in publicity, when something has not gone according to plan you wonder "if things went like that in that company, I wonder how it is in my investment, how have they considered this", and I think that is because we do not have established practices or understanding about what this whole thing is actually about – – (P3)*

It was also mentioned that there is an imbalance between investors and companies who utilize AI, with the latter being considered to have a significantly better understanding of the topic overall. Due to this, it was considered difficult to identify whether a company brings forth relevant information related to their use of AI, if they even give out such information in the first place.

> *I think there is kind of asymmetrical knowledge here, it is kind of like when you are at the dentist and the dentist says that you need to have some cavities filled, it is not like you [oppose to the idea]. I think that with AI it is emphasized when you compare it to a grocery store or a retailer which you have the ability to analyze, but then again this [AI] demands more understanding and very specific type of investing. (P4)*

The question of how investors' understanding of AI could be increased to enable including related matters to ESG analyses became emergent during the interviews. One interviewee stated that it would be necessary to first unravel what AI is actually about or what kind of technologies belong under the term, before evaluation of individual applications could take place. Regulation related to the use of AI in general was seen to be lacking, and it was considered insufficient to support investors in figuring out what sort of issues related to the topic should be addressed. One interviewee did however state that if an overarching, multinational regulation was created, it could potentially act as a guideline for investors as well, in a similar manner as the OECD guidelines for multinational enterprises or the UN Global Compact. Adding to this, the interviewee reminded that creating a regulation that would not be overly strict, which would limit its usability for different use cases, or too broad, allowing questionable acts to be tolerated, would not be easy.

> *And if we want to find these sort of universally accepted ways, then who defines what it is, and is there a risk that it is too broad, or that it dilutes the meaning if it kind of lets all sorts of flowers bloom. – – But if we had an international set of norms for AI which would be extensive enough without being overly strict either, something like that could be thought as the minimum standard which companies should reach. – – From the investor point of view it would be simple, or at least as simple as possible way of seeing a multifaceted topic. (P3)*

More likely options for increasing AI related knowledge level were considered to come from companies which use AI, or from investor experience with handling investments were AI may influence the asset's performance. Both companies which are forerunners in taking the responsible use of AI into account, as well as companies which do not take the matter properly into account were found to be possible teachers for investors in this topic. As for the former option, companies who proactively include ethics of AI into their operations and release their own guidelines were considered to provide learning material for investors of things that they should perhaps consider with other similar investments as well, and to overall spread awareness to the responsibility of AI.

> *Well of course it (releasing principles of responsible AI) helps investors in the sense that once you start reading the principles, your own understanding and expertise will also increase, in a way you start seeing and taking this viewpoint into account as well, since the topic is so new. (P3)*

Companies which do not consider ethics of AI and perhaps eventually lead to some form of harm towards their stakeholders, and concurrently to the company value due to reputation damage, were still said by P3 to provide learning opportunities for the investors on the other end of the spectrum – i.e., what kinds of risks may materialize and should be taken into account in the future or with other investments.

When asked if the interviewees thought that companies might take advantage of the situation where they hold more information than the investor side, it was stated that it is a possibility and that some companies might deliberately try to gain short-term profits from it. Still, the interviewees did also believe that those companies which pay attention to ESG issues – including the responsible use of AI, when applicable – will gain more from it in the long run.

5.3 Measuring the impact of AI

As the second major theme, issues related to including impact of AI to ESG analyses was discussed during the interviews. Based on the interviews, this topic is still relatively unknown in the investors' work, leading to believe that not many would include it in their analyses unprompted (e.g. without a prior AI related issue, which would have been communicated to stakeholders and thus seen as a potential threat to the investment). How AI could influence the three ESG dimensions, and when would AI be a material issue for an investment were seen as important aspects to be considered. Lastly, while the principles of responsible AI were not seen as a major topic in investment analyses at least for now, their potential future use was discussed during the interviews. For example, it was considered possible that they would be included as ESG data points in the future, although there is still uncertainty regarding in what way this would be achieved.

5.3.1 Impact of AI to ESG dimensions

All three ESG dimensions were mentioned to be affected by the use of AI in some way. The social and governance dimensions were mentioned more often than the environmental dimension, which was the primary focus only in the first interview. P1 described the need for efficient computing power for enabling increasingly powerful AI, which in turn leads to higher requirements for electricity to power such systems. This was thus deemed as a potential disbenefit for the environmental dimension. While the environmental pillar received less attention during the other interviews overall, the possibility of using AI to combat issues in this dimension, such as climate change, was also mentioned.

As for the social dimension, while both positive and negative effects of AI were discussed during the interviews, the possible risks received significantly more attention overall. Multiple groups of people were mentioned as possible targets of these risks, such as employees through the increased use of automation which could lead to layoffs or consumers as unwilling data sources for the increasing need of high-quality training data for AI systems as well as victims of misbehaving AI. Bystanders were also recognized as possible victims in scenarios where an AI system would make decisions that could affect their lives as well, with the risk that autonomous vehicles could pose for pedestrians as well being presented as an example. While the potential risks were clearly adduced more for this dimension, some use cases for bringing forth positive change were also given as examples especially during the first interview, where the possibility of using AI for

optimizing work shifts while ensuring employee wellbeing was mentioned, among others. Overall, however, the interviewees seemed to be more concerned about the possible negative effects that could be caused to humans. P4 reminded that this viewpoint was to be expected, as consumers are generally in a vulnerable position compared to companies and would thus need to more protection than business customers.

> *The idea is that companies are able to negotiate such terms with each other that their own safety, cyber security, or anything else would not be jeopardized, but consumers are not in a similar position to negotiate, and more regulation, principles and policies are thus needed for these situations. (P4)*

The governance dimension was considered a central topic as well: however, unlike with the other two dimensions, P2 reminded that the topics within this dimension would not be adaptable to AI specific companies alone. Rather, the interviewees considered this dimension to be overarching to all types of companies by default, as the issues within it are related to good corporate governance practices in general and how the company is managed. It was still stated in the fifth interview, that as the amount related processes increases within a company, the governing bodies should also understand the implications this would pose to ensuring the good practices even with the new operation environment.

> *And of course the more automation exists, the governing bodies should understand this and take this into account in their governance practices, that is a big part of being responsible. And it may be difficult to understand everything it [AI] influences. (P5)*

Even though all of the interviewees could identify especially potential risks that AI could pose to the ESG dimensions and it was generally agreed that a large number of industries have been or will be introduced to AI in the future, the general consensus seemed to be that the impact of AI is not commonly included in ESG evaluations yet, especially if the company is not clearly involved in AI (e.g. companies who sell AI products or heavily rely on AI in their operations). However, even if the need for taking related issues into account could be identified, the lack of guidelines and knowledge of AI in general were considered as obstacles for making reliable evaluations regarding the use of AI.

> *Well the matter is that the topic is so new, that at least to my knowledge there is no standard for investors, or a "look at these things and you can rest*

assured that your investment is a responsible user of AI" sort of guidance available. (P3)

Conflicting opinions related to the suitability of current ESG rating mechanisms for evaluating impact of AI emerged across the interviews. Some stated that in their view the current ratings are or would be suitable for AI products as well, or believed that the rating agencies can develop their ratings to better take impact of AI into account in the future, although the difficulty of doing so was also recognized.

I though this from the basis that if we could just adapt the criteria in these pillars – – in a similar manner as they are adapted fluently to other areas of business as well – – and I actually found a lot of suitable indicators which could be transferred for discussions related to the responsible use of AI. Take whistleblowing practices, it is vital to consider a company's ethical culture and how it is internally supported if a developer notices that someone works unethically or that an algorithm is biased, what are the means to forward this issue anonymously. (P2)

The knowledge they [rating agencies] have increases all the time – – and surely they have good resources for it [evaluating impact of AI] – – but it is likely difficult to take [the responsible use of AI] comprehensively into account as a whole. (P5)

Deviating from the previous opinions, the other interviewees saw the current evaluation methods as a poor fit for evaluating AI, or stated that they did not have sufficient information related to either AI or the current ESG ratings to properly comment on the matter. This was ascribed to the topic being only in its early stages and (investment) professionals not taking it properly into account.

I have to say that I cannot think of any [rating frameworks suitable for evaluating AI], but it may be that there are some and I just do not understand it. And this is caused by me not having a sufficient level of knowledge related to this topic [of evaluating AI]. (P3)

P3 also added to this comment that when ESG was not yet a mainstream topic among investors, some portfolio managers would claim that they did not consider ESG meaningful, even though the topics which they already included in their own analyses could also be found in the governance dimension of ESG. By mirroring to this example, P3 stated that a similar case could now be emergent with AI as well, that there could be elements

in the current ratings which would be suitable for evaluating AI as well, but that they are difficult to see as AI is still an unknown subject. Another interviewee shared similar views related to this matter.

> *[The ratings fit AI] Pretty poorly. I would say that it does not really come through yet, there has only recently been attempts to include topics related to human rights or cyber security into this as well. So, not really in any way [can AI currently be evaluated with the rating methods]. I do not see that it would have been considered much, we are just taking the early steps, and it probably is caused by the fact that even the overall understanding [of the topic] needs to be increased as well. (P4)*

It is notable that the current ratings were considered suitable mainly in the second interview with a professional who works closely with AI, while the interviewed ESG professionals were generally more cautious regarding the topic. However, both excerpts from the interviews with P3 and P4 highlight that this could be merely caused by a lack of understanding of AI on a deeper level, rather than the ratings not being suitable for the related technologies altogether. Furthermore, the earlier excerpt from P5, who also represents the investor side, also describes that the analysts at ratings agencies who specialize in related industries constantly gain new insights through their engagement with companies, and that they could apply the knowledge they gain to their ratings.

The question whether the existing indicators could be applied to evaluating impact of AI or whether it should be its own theme, similar to climate change or protecting biodiversity in many ESG ratings, was also brough up during the interviews. In the fourth interview, the interviewee mentioned that the role which AI will have in the future would influence this. Using Skynet (the artificial super intelligence antagonist of the Terminator movie franchise) as an allegory, the interviewee stated that if we expect AI to develop in this type of direction where it could cause severe damage to societies, a designated theme would be justified. With the current knowledge, however, it was deemed unnecessary to dedicate a separate topic to AI, but evaluating it in the same manner as other operations a company performs was seen appropriate.

In addition to considering AI as its own theme, one interviewee also reminded that some have suggested that technology or digitality should be added as its own dimension in ESG evaluations. However, following the previous statements, this was not considered

as necessary either, but instead adapting the current frameworks was seen as a better option.

> *There are two schools of thought regarding this, with some thinking that AI responsibility should be included in a digital responsibility, or a similar new pillar designated to these matters. When I went through this, I personally thought that this could be made with a smaller effort, I did not necessarily find grounds for why digitality should be made its own pillar. (P2)*

Evaluating the responsibility of using AI was considered to be easier for companies who are clearly in the AI business, as it is easier to question whether there are related risks, but also because companies themselves are more open about AI related issues. Discovering all possible risks which companies do not themselves report on was seen as a difficult task, and it was reminded that for companies who purchase AI products from others, the evaluation scope would be entirely different.

> *If you have an industrial company, for example, then the question turns into how they manage their supply code of conduct – – so in a way there is scope one and scope two, meaning that when the company itself is in the industry, it's an easier case [to evaluate] and the information is more open, but when it [the potential issue] is in the supply chain it turns in to a question of how the supply chain is managed instead. (P4)*

5.3.2 Materiality

As with ESG topics in general, the notion that different companies use AI in different ways, and thus need to consider different ethical issues to ensure conducting business in a responsible way, was clearly emergent during the discussions. One interviewee exemplified that for media companies, the risks may be related to user privacy issues or exposing their customers to filter bubbles through recommendation systems – issues which would not likely be emergent for industrial companies' core operations, for example. Because of this, not all issues of ethical AI would need to be considered with all companies, but rather the material ones should be recognized for each one. Still, it was recognized that even though the specific issues may vary by company, it is likely that AI will have an influence on a number of industries in the future. Material issues were recognized as a starting point for evaluating how responsible companies are with their AI use.

> *I would start with the material issues. [I would look at] Which effects of using AI are the most material ones, and I'd first pay attention to those. And it wouldn't be necessarily about thinking every detail at this stage, or not necessarily just the financially material issues either, but generally think of which issues are material for business continuity. – – And then, when we would know what the most material effects are, then we could start thinking about how those should be evaluated, how could we compare the effects between different companies or industries, or what are the best practice solutions for a given industry. (P3)*

> *But surely it has been recognized for various industries that this sort of thing [AI] is being developed all around and that it will be everywhere someday, in all sorts of machines or services that we use and in which companies invest in product development, and that the material risk exists for these companies [in traditional industries]. (P5)*

Because there is great variance in the type of ESG issues companies face, their own responsibility in recognizing their own material issues was also agreed on during the interviews. P2 stated that companies should be able to provide information regarding their AI use in a similar manner than other topics which are included in their sustainability reports. Instead of investors or rating agencies having a detailed list of potential issues related to e.g. human rights issues caused by the use of AI, the interviewee considered recognizing these granular issues as the responsibility of companies who use AI. Similar statements could be gathered from other interviewees as well, although the role of ESG rating agencies in bringing AI related issues to ESG evaluations was also suggested.

Even though it was generally agreed that companies should be responsible for reporting on their specific AI related issues, it was also stated that even companies themselves may not be able to recognize all of the areas where their AI system could have an influence.

> *A company might not even be able to realize what kinds of threats and other things might be related to these functions. And from the investment point of view, they might of course affect the level of company risk, because the company may suddenly face ridiculously high charges, penalties, and so on. (P5)*

It was also stated that as AI is still a relatively young technology in its current scale, there might not be enough information available for recognizing all of its potential issues

beforehand, leading to the conclusion that at least some issues may turn material only through controversies as they become apparent. Additionally, it was reminded that due to AI being a highly scalable and widely spread technology, controversies in one part of the world quickly turn an issue material elsewhere as well.

> *I think that on some level this topic is brought to analysts' attention through controversies as well, as these kinds of special cases are looked through on a more detailed level with some companies [when controversies arise]. But when there are enough of these special cases, these things will probably rise as a topic which should be included in the ratings and which should receive more attention overall. (P2)*

> *Well, those [controversies] are naturally the first step, and perhaps one should remember that as these are pretty scalable solutions – – which can be implemented globally, their impact is significant, it isn't limited to – – only Finland or Europe, instead when it turns into an issue it happens everywhere at the same time and is material immediately (P5)*

In addition to the global environment being mentioned in the above excerpt, it was also mentioned as a source of contradiction for what types of topics are considered material around the world. It was stated that a company which operates in multiple countries and continents needs to take local norms from each location into consideration, and what AI should be used for was given as an example of viewpoints which may differ greatly in different parts of the world. For example, the question of whether AI should be allowed to replace human workers was brought up by different interviewees, one believing that there would be other tasks that the replaced humans would be more suitable for, while another reminded that some cultures would be heavily opposed to the possibility of increased unemployment.

5.3.3 The role principles of responsible AI in ESG analyses

It was widely agreed during the interviews that the principles of AI are still a somewhat unfamiliar concept within the investment world, which was in line with the previous statements of AI in general being a new topic for investors. Following this, how AI can or why it should be used in an ethically sound manner was also deemed as a topic which is not yet understood well. As the current state of taking the effects of AI or the responsible use of AI are still in their infancy, a company lacking their own set of principles was not

considered as a major factor in ESG analyses or investment decisions, even if the company utilizes AI in some way.

Still, one interviewed expert on the investor side stated that the principles can be included in investment analyses, if the company is in AI business and they have created a set of principles. Indicating that AI related issues are understood and that possible risk control or mitigation efforts were performed was seen as a positive sign among the interviewees. Additionally, the principles of responsible AI were seen as a possible way of communicating that material issues were recognized and considered. It was also described that as it is impossible for all investors to be experts in every possible risk and issue a company may face, even providing evidence that AI related issues are recognized is a positive sign in itself, even if the content of the principles would not be that clear to the investor initially.

> *If a company has a set of – – accepted principles, everyone believes that they are followed until they are proven otherwise. – – And of course, if there are no principles related to the topic, it is already a good indication that perhaps not everything is taken into account. But if there are principles – – that is already a precondition for the matter being considered [in a company]. – – And when it [set of principles] is written and published, someone has to monitor it, and for instance someone from within the company can intervene if things do not go as they should have. (P5)*

Another interviewee took a more cautious approach to how the principles should be considered in ESG analyses, if their role in assessing companies would be increased in the future. In this interview, it was stated that merely verifying that a company has a set of principles available would only be a starting point in assessing the company, and that companies should also be able to explain how they intend to operationalize the principles.

> *Well it is probably a bad situation if we end up just putting a checkmark on a list for having a set of published principles, but with no one being interested whether they are operationalized. That is probably the worst possible outcome, or that is something I would avert. – – Naturally communicating about the principles is the first step and it indicates that the issue has been recognized and that the principles have been defined as part of the strategy and value acts, but in no means can the evaluation stop there. Instead, the core should be targeted towards how they are able to communicate about the mechanisms with which the principles are carried out. (P2)*

Overall, the principles were still considered to be a viable possibility for being used as data points for AI related ESG evaluation. P1 did however remind that following the principles too strictly could also lead to dissatisfaction towards companies.

> *I think they [principles of responsible AI] are a very good fit [to ESG evaluations], but we need to retain some reason and balance – – since the [global] market pressure is quite high after all. If we go overboard with establishing these matters, then there could be pressure from the market that this is too ethical, overly ethical. (P1)*

The possibility of using the principles of responsible AI as "ethical greenwashing" was also brought up, meaning that companies could claim to be responsible in their AI operations and exaggerate the benefits while keeping the investors' attention away from potential issues. This was seen as a possible risk especially in the current stage, when the investor side knowledge related to the matter is still generally low, which makes identifying these types of cases more difficult. However, while this was seen as a possibility, it was also considered as a learning opportunity for the investors.

> *There's a massive amount of data available and no single investor can be perfect in managing all that data, but instead – – we should strive to continuously learn new things and strive to be better at understanding the data and finding the material topics. And in order to find the material topics, you must also see some of the bad versions. (P3)*

As for the individual principles, P2 stated that as some principles have been found to be central in many of the released guidelines, those could be considered as universal, generally accepted principles for a variety of industries, with transparency and accountability being mentioned here as examples. Even though it was stated that many different industries could take these core principles to use, it was not claimed that the issues stemming from AI are similar in all industries or competitors within industries. Rather, the narrow set of principles was seen to be encompassing enough that they could be used as one potential governance mechanism, even if the types of AI applications and subsequently the related risks companies aim to tackle differ.

Risks related to biased results or threats to user privacy were considered as central to AI during the interviews, making principles of privacy and justice also important to be considered. It was also mentioned that privacy and cyber security for some AI products could already be considered as central topics, but not necessarily because they have been

recognized relevant for AI specifically, but because they are already a part of many prominent ESG rating frameworks. Overall, however, it was considered that there is still a long way before AI responsibility would be properly considered on a wider scale.

> *Privacy protection is taken into account [in ESG ratings] – – but in reality it [responsible use of AI] really is not a mainstream topic yet, there is still a lot of work to be done for it to become clearer or something that would always be taken into account in the evaluations, or even considered per se on any level. (P2)*

The first interviewee also thought that even though there should be common principles which would act as general guidelines for all companies within industries, providing them with a fair ruleset and limiting irresponsible actors from reaping excess benefits by disregarding ethical or moral considerations. However, the interviewee also believed that there will inevitably be sector specific or market specific interpretations of the same principles.

> *But then again there will likely form inevitable sector specific or market specific interpretations and pressures due to financial reasons, because there money simply talks. And there will be more responsible companies and I believe they will eventually be the winners, but they must be able lean on the moral principles and on the interpretations of the rules and solutions which have been agreed on. But then there will be those who simply operate on the borderlands of these rules, who could not care less about the sustainable principles, very financially oriented actors who can provide cheap products to a wide audience by targeting their irresponsible marketing, and so on. (P1)*

6 DISCUSSION

6.1 Results

This study set out to investigate how investors currently perceive questions related to the responsible use of AI, and whether there are some preexisting connections between the topics of responsible AI and ESG investing – for example, have some of the principles of responsible AI already been considered in ESG analyses. Based on this, the following research questions were formed:

- How is the responsible use of AI taken into consideration when an ESG investment analysis is conducted?
- What kind of connections can be found between the existing principles of responsible AI and the criteria in ESG ratings?
- How could the responsible use of AI be considered in ESG analyses in the future?

Based on the findings, the responsible use of AI is not currently a topic which would be widely considered when conducting ESG analyses. If a company specifically states that they are in the AI business or otherwise heavily rely on AI, an investor can try to include it in an analysis on a case-by-case basis, but doing so may be difficult since the topic is still novel for investors on a basic level. As for the potentially existing connections between the principles of responsible AI and ESG criteria, the principle of privacy is the only one which can currently be seen in both sides, as privacy is widely included in ESG evaluations already. The principles could still end up having a more central role as a way of signaling leadership or awareness of AI related issues, at least in the near future when it is still not mandatory to report on AI usage. In the long term, the principles could potentially be incorporated to ESG evaluations for those industries or companies where the issues are considered material. Still, the question of who brings the topic to mainstream investors' awareness remains unanswered. The following sections will discuss these findings in more detail.

6.1.1 Including AI in ESG ratings

It became evident during the interviews that for the most part, the interviewees agreed that questions related to the responsible use of AI are still relatively unknown to the investor side. Following this, the interviewed experts on the investor side generally agreed that as of now, the possible AI related ESG evaluations are done on a case-by-case basis,

if AI is recognized as a significant risk for a particular asset's performance. The significance is mainly recognized if a company itself clearly indicates that AI is a central tool in their business – as would the case for companies who sell AI products or otherwise heavily utilize some sort of AI, and highlight this in their reporting. Furthermore, AI in itself was also considered a difficult concept to understand, due to the term being used for a large selection of different technologies, each of which have their own characteristics, as well as different use cases and risks.

Because of this, there may be confusion over what investors should even be evaluating when making investment decisions, and the interviewees did indeed mention that there is a need for a standard or some other form of guideline for evaluating AI systems reliably and comparatively. To provide assistance with this problem, ESG rating agencies were seen as a possible support in providing insights about this topic in the future, which is in line with previous findings of investors valuing rating agencies' work due to resource constraints (van Duuren et al. 2016, 529-530). However, ESG rating agencies have already been found to rate the same companies differently due to the agencies' scope, measurement and weight divergences (Berg et al. 2020). The agencies' different rating methodologies have in turn been explained with reasons related to prevalent societal norms or competitive reasons, as some of the agencies' want to differentiate their offering from their competitors (Sandberg et a. 2009, 527; Daugaard 2020, 1512), among others.

Since these issues related to rating divergence have already been widely discovered (see e.g. Dorfleitner et al. 2015; Chatterji et al. 2016; Berg et al. 2020), it would seem expected that if impact of AI would be included in ESG ratings in the future, companies may be evaluated very differently on their AI responsibility. Furthermore, how AI should be included in the ratings in the first place is another question to be addressed. P2 reminded that discussions over whether technology should be its own dimensions in addition to the environmental, social and governance dimensions had already surfaced, but the interviewees who commented on this felt that with the current knowledge, impact of AI could be evaluated with the existing evaluation methodologies. Still, the rating divergence should be taken into account with this topic as well: Berg et al. (2020) found that the disagreement between rating agencies was highest for the categories titled human rights and product safety, both of which can be considered to belong under the social pillar of ESG. This finding is meaningful in the context of this study, since AI would seem like a suitable candidate to be considered in both categories. Privacy issues from excessive gathering of training data or humans being urged to utilize AI in increasing

amounts, possibly limiting their autonomy, could lead to human rights violations (Floridi et al. 2018), whereas autonomous vehicles powered by AI, among others, have serious safety issues to be addressed.

6.1.2 Role of principles of responsible AI

Brusseau's (2021) *AI Human impact* model gives an example of how the principles of responsible AI could be utilized in the future for evaluating the impact of AI systems. It was generally considered during the interviews that the individual principles do not yet have a major role in defining company responsibility, but having a list of principles was still considered a positive sign for the company: as an example, P5 described having the list of principles as an indication of recognizing the related issues. Of the individual principles, privacy and cyber security, as well as issues related to them (e.g. collecting excessive amounts of user data, or how the data is handled afterwards) were largely agreed on being important to be included in ESG evaluations during the interviews. This was to be expected, as both have established regulation and are widely included in ESG rating agencies' evaluations in some way, likely making both emergent topics in investors' work already even when AI is not necessarily considered as its own topic. Overall, topics related to ensuring the safety of humans were highlighted during the interviews. Interestingly, little attention was given directly to topics related to responsibility or accountability, even though they are among the most important ones in academic literature (Floridi et al. 2018; Jobin et al. 2019; Dignum 2020). This could be due to the company who uses or develops an AI system being automatically seen as the one who should be held accountable for its actions, even though from a moral standpoint this can be difficult with highly automated systems (Matthias 2004). This raises the question of whether companies are expected to only develop and use AI systems for which they can always guarantee the ability to take full moral responsibility for, even though this may limit the possible AI applications being taken to use.

While the potential benefits of the principles in evaluations were agreed on, P2 reminded that companies would also need to be able to operationalize the principles, so that they will not be just a checkmark item on ESG evaluations. How organizations can move from principles to practice has received interest in the academics as well. Morley et al. (2020, 2161) have phrased that failing to operationalize responsible use of AI in search of short-term benefits from disregarding the issues may lead to unfavorable consequences to the field of AI, but also recognize that forerunners may initially face competitive

disadvantage. This was also recognized by P1, who felt that taking ethics more heavily into consideration can lead to increased market pressure, and wished that companies would be given an equal standing in form of guiding principles, or mandating regulation. Due to the market pressure, many organizations may thus still opt for not taking ethics of AI largely into consideration, as they do not see the competitive advantage in doing so. Another potential downfall of the principles which was recognized during the interviews was the possibility of ethics bluewashing (Floridi, 2019, 187), or using the principles as a way of appearing responsible to gain stakeholder trust. Perhaps at this point, when the ethics of AI and AI governance mechanisms are in their early stages, even principles which are published for questionable reasons (which is mentioned as a possibility e.g. in Schiff et al. (2020, 155-156)) can still prove useful for spreading awareness – still, whether there will be a way for investors to ensure that the companies they invest in adhere to their own principles would be a worthwhile effort to study in the future, and perhaps it will emerge as a new topic if the principles are adopted by wider audiences in the future.

6.1.3 Bringing responsible use of AI to mainstream awareness

Even though AI will likely disrupt many industries and thus be a material topic for many companies in the future, the current lack of standards and overall knowledge of AI indicates that there is a possibility that as of now, using AI in an irresponsible manner will likely go unnoticed. This would be especially true for companies which do not include AI in their official reporting or other forms of communication, as it was mentioned during the interviews that investors likely do not specifically investigate whether a company utilizes AI or not. Not communicating about the use of AI could simply be a question of not finding it a financially material risk to company performance, which in turn would lead to the topic not being considered necessary or mandatory to report on. This would be in line with the findings of Tamimi & Sebastianelli (2017, 1674), as they have discovered that companies tend to have a reactive standpoint to CSR practices: while initiatives like the GDPR in the EU have made limited aspects of responsible AI mandatory for companies to consider, on a bigger picture, related regulation was still considered to be lacking during the interviews. This may however change in the future, with propositions such as the EU Artificial Intelligence Act (COM/2021/206 final) being introduced recently.

Voluntary and proactive stance to responsible use of AI could however prove beneficial for companies. As is proposed by Du & Xie (2021) in their recent work, having suitable CSR activities in place to mitigate a recognized level of risk caused by an AI

system can help companies with building a positive brand image and subsequently increase company value. It is still good to remember that as AI is only one topic ESG professionals and investors need to consider, not everyone can be specialists in all possible issues which should be considered with AI, even if the general level of knowledge related to the topic would increase in the future. This was also recognized during the interviews – for example, P5 stated that principles which companies publish are believed to hold true for said company until proven otherwise. Still, P5 added that companies failing to prove that they have somehow taken risks which are relevant to their business into account (for AI companies, publishing principles of responsible AI being a possible example) could also indicate that some material issues are overlooked. Here the current state of AI not being a topic which all companies report on (even when they use it) still presumably affects investors' possibilities in recognizing that a company overlooks some issues, but this could well change in the future.

The previous literature has highlighted the role of stakeholders in bringing new ESG issues to the attention of companies and regulators, as well as raising the importance of ESG and sustainability overall (see e.g. Rogers & Serafeim 2019; Boffo & Patalano 2020). This can be expected to continue in the future as well, as millennials and younger generations have been found to be more engaged in ESG and impact investing compared to previous generations (Boffo & Patalano 2020, 17). Similar findings could also be observed during the interviews, as both investors, but especially consumers and business customers were stated as possible inspiration for bringing awareness to the potential issues related to AI and the importance of responsible use of AI. Furthermore, companies themselves were also considered as having a role in advancing the topic, as for example P1 described that forerunners of responsible AI could start mandating their partners to also consider and mitigate potential risks of AI accordingly. This statement was largely in line with institutional theory, and more precisely with coercive isomorphism presented by DiMaggio & Powell (1983, 150), which states that organizations not only seek legitimacy by following mandatory regulations but are also forced to act in a manner which is accepted by the other organizations within the same environment. One possible hindrance for increasing the level of knowledge related to responsible AI can thus be related to companies not being able to recognize the necessity of including ethical considerations throughout the lifecycle of their AI products, and not mandating their vendors or partners doing the same.

Based on the academic literature (or lack thereof) and the interviews conducted in this study, AI is not currently seen as a material topic for larger groups, such as whole sectors or industries. In addition to the adoption of AI perhaps simply not being at a level where it would have become material, one possible explanation for this could also be that the commercial focus on AI is still on AI systems which are used in industrial settings to enhance the efficiency of data processing. These may largely be settings where AI simply does not have the possibility to impact humans or at least cause direct negative effects on us, perhaps limiting the potential negative effects to inefficient decision-making in company processes or an AI system being a large financial investment with little return on investment to show for it. Comments from the interviewees would at least indicate towards this being the current stage, as efficiency gains were indeed considered as the most prominent image that the market may have of AI.

Still, the potential issues have been identified widely and individual AI intensive companies may already be subject for scrutinizing from investors. This leads to believe that on the Rogers & Serafeim (2019) *Pathway to materiality* model, AI has already departed from the initial status quo stage and is becoming or is already material for certain companies – even to the point where companies are perhaps trying to use the principles of responsible AI as a form of self-regulation in order to avoid restrictive regulation from being enforced. Such a regulation may still be enforced (e.g. the EU AI act), especially if some controversy with wide-scale impact emerges, as Floridi et al. (2018, 691) have speculated. As these controversies are brought to stakeholder attention, companies will likely be forced to readjust their operations so that they can regain their stakeholders trust and be allowed to continue in their working environment, as the stakeholder theory by Freeman (2010) implies. Controversies were also considered a likely possibility for bringing the topic of responsible AI to the attention of wider audiences. This would also include investors who may not currently consider or even be aware of the related issues, even though some of the companies they have already invested in may be subject to them.

There is no decisive answer for how AI will turn material for a wider group of companies, or whether this will even happen in the future. Based on the conclusions discussed thus far, it would seem that this could either happen if companies themselves started leading the chance by mandating others to follow their lead, or through controversies, as stakeholders are informed of possible AI related problems in companies' operations. Through these types of events, investors see how they affect certain players on the market, and start questioning whether similar matters could impact their own investments as well.

If this happens often enough and is considered to have a large enough impact, perhaps even on a variety of industries, some common issues may end up to the investors radar of topics which they should consider before making an investment. Creating an evaluation standard which could be used to reliably compare how companies perform related to AI responsibility may be challenging, as the types of use cases for AI are manifold, as are the underlying technologies which belong under the umbrella term of artificial intelligence. Still, as the AI field has already started considering the possible issues which their applications may cause, and thus recognized some material topics themselves, utilizing this existing work for ESG analyses does not seem implausible. As was indicated during the interviews, certain core principles can well be adapted universally to different industries – as long as the specific AI related risks and mitigation efforts for them can be recognized within industries or by individual companies.

6.2 Contribution

This study set out to explore the current state of taking the responsible use of AI into consideration in ESG analyses and investment decisions. With the recent work of Brusseau (2021) being the only discovered academic source which combines both worlds of ethical use of AI as well as ESG investing, the academic contribution lies largely in bringing these two topics together and suggesting that there is indeed a new area for researchers to explore. For the scholars of AI ethics, this brings new viewpoints on where their area of expertise can potentially be relevant in the future. As for the scholars focusing on ESG matters, this study introduces an issue which is not yet widely considered in ESG evaluations, but which may well be material in the future.

Similar contributions can be found on the practical side as well. For investors, the study presents a topic which may be material for their investments in the future, and highlights matters which the responsible AI scholars have deemed important when considering what “responsible AI” entails. The study also gives idea to companies which use AI already or plan to do so in the future about issues which they should perhaps take into account to ensure ethical use of AI, issues which may be material for their financial success, or what their potential investors may be interested about in the future.

6.3 Limitations and future research

The limited scope of the empirical data poses an obvious limitation to this study. With only five interviewees, it cannot be said to portray a comprehensive view of the current stage of investors taking the responsible use of AI into account. Because of this, some relevant aspects may have been lost, especially if there already are experts on the ESG side who have taken these issues into consideration on a wider scale. Additionally, the scope of the study was limited to the Finnish context only – it is thus possible that investors in other countries have already started integrating this topic into their own work.

As for the interviewed experts, including views from representatives of other organization types on the investor side (e.g. venture capital or impact investing) would have given an additional viewpoint on how they might consider ethical AI in their business. For example, views from the venture capitalist would have been interesting, since they may be among the first financiers for small AI startups.

This study provides follow-up research topics for the academic world. Due to the exploratory nature of the study, the scope for the principles of responsible AI was kept at a rather high level, meaning that it was not possible to gain a detailed view of how important the investment world considers each of the individual principles. For example, there is a wide consensus of the most central principles of responsible AI, but the previous studies do not necessarily take the investor view on this into account. Investigating whether investors would consider the same set of core principles (accountability, transparency, non-maleficence, etc.) as the most important ones would be an interesting topic to study, although it may still take some time before such a detailed level of knowledge is achieved for this study to be possible.

Another possibility for further research is related to the more practical side of evaluating AI from the ESG perspective. This study merely recognized that there will likely be a need for a standardized method for evaluating impact of AI, but how that could be done is still largely an unknown area, although the first suggestions for providing solutions to this topic have already been made (see Brusseau 2021). Using the principles of responsible AI as a base for creating a new evaluation method for investors, like Brusseau (2021) has suggested, is only one option which can be studied further. Other alternatives include e.g. investigating how the current ESG rating methodologies can be applied to evaluating the impact of AI, or how ESG rating agencies could better take this topic better into

account in the future, especially since this would be a new type of issue to be addressed in ESG ratings.

As this study has been conducted in the Finnish context only, this also opens a possibility for conducting similar studies in other countries to see if the topic has already gained more attention elsewhere. Conducting a study with a global group of participants, where their viewpoints on the question related to the responsible use of AI could be compared, could also provide valuable insights. Analyzing viewpoints from countries with different cultural backgrounds, ethical viewpoints and legislation could provide important material in the increasingly global AI market.

As AI seems to still be a relatively unknown topic with the myriad of different related technologies, use cases, etc. making it hard for experts outside the AI landscape to understand, conducting research on how this knowledge gap could be narrowed could also improve the understanding related to the responsible use of AI. Possibilities in this area include e.g. conducting interviews with investors or other groups of professionals who do not directly work with AI, and investigate which aspects are seen the most difficult and thus require more clarification. These findings could potentially be further used to develop more generic training material for larger audiences, or targeted material for specific professional groups related to AI as a technology or the responsible use of AI specifically.

Lastly, the interviewees reminded that there is a distinct difference regarding how a company operates and how they see that their supply chain also acts in a responsible manner. Since AI can be expected to continue transforming different industries, it seems likely that more attention should be given in the future on how companies ensure that their AI suppliers have taken the ethical viewpoints into account. Conducting research on whether companies who purchase AI products from others have already addressed related issues, or if there is need for improvement, could thus be justifiable.

7 CONCLUSION

This thesis was conducted as an exploratory study set out to investigate how the investment world currently considers questions related to the ethical use of AI. Even though including nonfinancial information related to the ESG dimensions has become a mainstream practice during the last decade, finding material topics which may influence investment performance may still be a difficult task for investors. As AI may still be considered as a novel topic for investors on a basic level, trying to evaluate or measure its potential impact is still not practiced widely. A case-by-case evaluation may be conducted if a company is clearly in the AI business, or if the company heavily relies on AI in their business and clearly indicates this in their external communication. These findings can also be considered as answers to the first research question of *"How is the responsible use of AI taken into consideration when an ESG investment analysis is conducted?*" – individual, likely non-comparable evaluations may be conducted related to the potential risks AI may impose on an investment, as no standards or definitions of responsible AI exists for the investors.

"*What kind of connections can be found between the existing principles of responsible AI and the criteria in ESG ratings?*" was investigated as the second main research question in this study. As of now, the principles are not widely recognized in the investor world, leading to believe that no clear connections between the principles and the ESG criteria is to be found at large. The only clear exception to this at this point seems to be the principle of privacy, which has already been included in ESG evaluations, as the topic is central to other types of digital products as well. However, as P2 mention during the interview, there are ESG indicators which could possibly be adapted to evaluating the impact of AI as well, with whistleblowing practices being given as an example – which in turn can be considered to belong under the principle of transparency.

As for the third and final question, "*How could the responsible use of AI be considered in ESG analyses in the future?*", having a set of principles was already seen as a good indication of being able to recognize potentially material issues which AI may cause for a company. Some interviewees did also consider that the principles of responsible AI could well be suited for being included in ESG analyses as well. Additionally, while principles of privacy and justice were mentioned as being potentially more central in the future, it was generally agreed that finding which issues are material for a given company or industry will be important. Identifying these material issues could then also help

investors recognize which related principles they should take into account in their analyses for a given company. Still, the principles for which a wide consensus has already emerged were also seen as being suitable universally for different industries.

Given the exploratory nature of this study, the aim was not to provide definitive answers for how AI should be evaluated or how ESG analyses should be altered to better include the potential issues of AI. Rather, the purpose was to explore whether the topic is currently emergent in the daily work for investors, and how it could develop in the future. With the research questions being answered above, this intent has thus been fulfilled. Based on the comments from the informants, the topic is still likely in very early stages for most investors, although the findings of this study are naturally limited to the Finnish context only. As one interviewee formulated, there is still surely a lot of work to be done before the wider audiences will be introduced to the responsible use of AI.

REFERENCES

AI for Good Foundation (2021). *AI for Good*. https://ai4good.org/, retrieved 29.4.2021.

AI HLEG (2019a). *A definition of AI: Main capabilities and disciplines.* Independent high-level expert group on artificial intelligence set by the European Commission. https://digital-strategy.ec.europa.eu/en/library/definition-artificial-intelligence-main-capabilities-and-scientific-disciplines, retrieved 15.2.2021.

AI HLEG (2019b). *Ethics guidelines for trustworthy AI*. Independent high-level expert group on artificial intelligence set by the European Commission. https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai, retrieved 10.2.2021.

Alvesson, M., & Kärreman, D. (2007). Constructing mystery: Empirical matters in theory development. *Academy of Management Review*, 32(4), 1265-1281.

Amel-Zadeh, A., & Serafeim, G. (2018). Why and how investors use ESG information: Evidence from a global survey. *Financial Analysts Journal, 74*(3), 87-103.

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

Benkler, Y. (2019). Don't let industry write the rules for AI. *Nature, 569*(7754), 161-162.

Berg, F., Koelbel, J. F., & Rigobon, R. (2019). *Aggregate confusion: The divergence of ESG ratings*. MIT Sloan School of Management.

Boffo, R. & Patalano, R. (2020). *ESG investing: practices, progress and challenges*. Technical report, OECD Paris.

Boiral, O., & Henri, J. F. (2017). Is sustainability performance comparable? A study of GRI reports of mining organizations. *Business & Society, 56*(2), 283-317.

Bostrom, N., & Yudkowsky, E. (2014). The ethics of artificial intelligence. *The Cambridge handbook of artificial intelligence, 1*, 316-334.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative research in psychology, 3*(2), 77-101.

Brusseau, J. (2021). AI human impact: toward a model for ethical investing in AI-intensive companie. *Journal of Sustainable Finance & Investment*

Cai, Y., Jo, H., & Pan, C. (2012). Doing well while doing bad? CSR in controversial industry sectors. *Journal of Business Ethics, 108*(4), 467-480.

Cardoni, A., Kiseleva, E., & Terzani, S. (2019). Evaluating the intra-industry comparability of sustainability reports: The Case of the oil and gas industry. *Sustainability, 11*(4), 1093.

CFA Institute (2015). *Environmental, social, and governance issues in investing: A guide for investment professionals*. CFA Institute. https://www.cfainstitute.org/-/media/documents/article/position-paper/esg-issues-in-investing-a-guide-for-investment-professionals.ashx, retrieved 25.5.2021.

Chang, C. H. (2015). Proactive and reactive corporate social responsibility: antecedent and consequence. *Management Decision, 53*(2), 451-468.

Chatterji, A. K., Durand, R., Levine, D. I., & Touboul, S. (2016). Do ratings of firms converge? Implications for managers, investors and strategy researchers. *Strategic Management Journal, 37*(8), 1597-1614.

Collingwood, R. (1946). *The idea of history*. Oxford University Press.

COM/2021/206 final. *Proposal for a regulation of the European Parliament and of the Council laying down harmonised rules of artificial intelligence (artificial intelligence act) and amending certain union legislative acts*. European Commission. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52021PC0206.

Cope, D. G. (2014). Methods and meanings: Credibility and trustworthiness of qualitative research. *Oncology nursing forum, 41*(1), 89-91.

Cort, T., & Esty, D. (2020). ESG standards: Looming challenges and pathways forward. *Organization & Environment, 33*(4), 491-510.

Dastin, J. (2018). *Amazon scraps secret AI recruiting tool that showed bias against women*. Reuters. https://www.reuters.com/article/us-amazon-com-jobs-automation-insight-idUSKCN1MK08G, retrieved 22.3.2021.

Daugaard, D. (2020). Emerging new themes in environmental, social and governance investing: a systematic literature review. *Accounting & Finance, 60*(2), 1501-1530.

Dignum, V. (2020). Responsibility and Artificial Intelligence. *The Oxford Handbook of Ethics of AI, 4698*, 215.

DiMaggio, P. J., & Powell, W. W. (1983). The iron cage revisited: Institutional isomorphism and collective rationality in organizational fields. *American sociological review*, 147-160.

Donaldson, T., & Preston, L. E. (1995). The stakeholder theory of the corporation: Concepts, evidence, and implications. *Academy of management Review, 20*(1), 65-91.

Dorfleitner, G., Halbritter, G., & Nguyen, M. (2015). Measuring the level and risk of corporate responsibility–An empirical comparison of different ESG rating approaches. *Journal of Asset Management, 16*(7), 450-466.

Du, S., & Xie, C. (2021). Paradoxes of artificial intelligence in consumer markets: Ethical challenges and opportunities. *Journal of Business Research, 129*, 961-974.

Eccles, R. G., Kastrapeli, M. D., & Potter, S. J. (2017). How to integrate ESG into investment decision‐making: Results of a global survey of institutional investors. *Journal of Applied Corporate Finance, 29*(4), 125-133.

Escrig-Olmedo, E., Fernández-Izquierdo, M. Á., Ferrero-Ferrero, I., Rivera-Lirio, J. M., & Muñoz-Torres, M. J. (2019). Rating the raters: Evaluating how ESG rating agencies integrate sustainability principles. *Sustainability, 11*(3), 915.

Esty D. & Cort, T. (2017). Corporate sustainability metrics: What investors need and don't get. *The Journal of Environmental Investing, 8*(1), 11-53.

European Commission (2019). *Guidelines on reporting climate-related information*. European Union. https://ec.europa.eu/finance/docs/policy/190618-climate-related-information-reporting-guidelines_en.pdf, retrieved 5.4.2021.

EYGM Limited (2020). *How will ESG performance shape your future?.* EY. https://assets.ey.com/content/dam/ey-sites/ey-com/en_gl/topics/assurance/assurance-pdfs/ey-global-institutional-investor-survey-2020.pdf, retrieved 21.3.2021.

Finsif (2020). Mitä vastuullinen sijoittaminen tarkoittaa?. https://www.finsif.fi/mita-se-on/, retrieved 14.3.2021.

Fischer, J. M., & Ravizza, M. (1998). *Responsibility and control: A theory of moral responsibility*. Cambridge University Press.

Fjeld, J., Achten, N., Hilligoss, H., Nagy, A., & Srikumar, M. (2020). Principled artificial intelligence: Mapping consensus in ethical and rights-based approaches to principles for AI. *Berkman Klein Center Research Publication*, (2020-1).

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valcke, P. & Vayena, E. (2018). AI4People—an ethical framework for a good AI society: opportunities, risks, principles, and recommendations. *Minds and Machines, 28*(4), 689-707.

Floridi, L. (2019). Translating principles into practices of digital ethics: Five risks of being unethical. *Philosophy & Technology, 32*(2), 185-193.

Floridi, L., & Cowls, J. (2019). A unified framework of five principles for AI in society. *Harvard Data Science Review, 1*(1).

Floridi, L., Cowls, J., King, T. C., & Taddeo, M. (2020). How to design AI for social good: Seven essential factors. *Science and Engineering Ethics, 26*(3), 1771-1796.

Freeman, R. E. (2010). *Strategic management: A stakeholder approach*. Cambridge university press.

Freeman, R. E., Harrison, J. S., Wicks, A. C., Parmar, B. L., & de Colle, S. (2010). *Stakeholder theory: The state of the art*. Cambridge University Press.

Friedman, M. (1970). *A Friedman doctrine ‐ The social responsibility of business is to increase its profits*. The New York Times. https://www.nytimes.com/1970/09/13/archives/a-friedman-doctrine-the-social-responsibility-of-business-is-to.html, retrieved 18.4.2021.

Friede, G., Busch, T., & Bassen, A. (2015). ESG and financial performance: aggregated evidence from more than 2000 empirical studies. *Journal of Sustainable Finance & Investment, 5*(4), 210-233.

Gallo, P. J., & Christensen, L. J. (2011). Firm size matters: An empirical investigation of organizational size and ownership on sustainability-related behaviors. *Business & Society, 50*(2), 315-349.

Ghauri, P., & Grønhaug, K. (2010). *Research Methods in Business Studies*. Pearson Educated Limited.

Gibson, R., Krueger, P., & Schmidt, P. S. (2019). ESG rating disagreement and stock returns. *Swiss Finance Institute Research Paper*, (19-67).

Gioia, D. A. (1999). Practicability, paradigms, and problems in stakeholder theorizing. *Academy of Management Review, 24*(2), 228-232.

Gioia, D. A., Corley, K. G., & Hamilton, A. L. (2012). Seeking qualitative rigor in inductive research: Notes on the Gioia methodology. *Organizational research methods*, 16(1), 15-31.

GRI (2021). *Our mission and history*. https://www.globalreporting.org/about-gri/mission-history/, retrieved 4.4.2021.

GSIA (2018). *2018 Global Sustainable Investment Review*. GSIA. http://www.gsi-alliance.org/wp-content/uploads/2019/03/GSIR_Review2018.3.28.pdf, retrieved 28.2.2021.

GSSB (2016). *GRI 101: Foundation*. https://www.globalreporting.org/media/55yhvety/gri-101-foundation-2016.pdf, retrieved 4.4.2021.

Guba, E. G. (1981). Criteria for assessing the trustworthiness of naturalistic inquiries. *ECTJ, 29*(2), 75-91.

Hagendorff, T. (2020). The ethics of AI ethics: An evaluation of guidelines. *Minds and Machines, 30*(1), 99-120.

Halpern, E. S. (1983). *Auditing naturalistic inquiries: The development and application of a model* (Doctoral dissertation, Indiana University).

Hill, J. (2020). *Environmental, social, and governance (ESG) investing: A balanced analysis of the theory and practice of a sustainable portfolio.* Academic Press.

Hong, H., & Kacperczyk, M. (2009). The price of sin: The effects of social norms on markets. *Journal of financial economics, 93*(1), 15-36.

Hyrske, A., Lönnroth, M., Savilaakso, A. & Sievänen, R. (2020). *Vastuullinen sijoittaja*. Helsingin seudun kauppakamari.

Hörisch, J., Freeman, R. E., & Schaltegger, S. (2014). Applying stakeholder theory in sustainability management: Links, similarities, dissimilarities, and a conceptual framework. *Organization & Environment, 27*(4), 328-346.

Imperial College London (2017). *Written Evidence to Select Committee on Artificial Intelligence* (AIC0214). http://data.parliament.uk/writtenevidence/committeeevidence.svc/evidencedocument/artificial-intelligence-committee/artificial-intelligence/written/70509.html, retrieved 25.2.2021.

Jaeger, R. G., & Halliday, T. R. (1998). On confirmatory versus exploratory research. *Herpetologica*, 54(Suppl.), S64-S66.

Jo, H., & Na, H. (2012). Does CSR reduce firm risk? Evidence from controversial industry sectors. *Journal of business ethics, 110*(4), 441-456.

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

Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science, 349*(6245), 255-260.

Kempf, A & Osthoff, P. (2008). SRI funds: Nomen est omen. *Journal of Business Finance & Accounting, 35*(9-10), 1276-1294.

KPMG International (2020). *The time has come: The KPMG Survey of Sustainability Reporting 2020*. KPMG International. https://assets.kpmg/content/dam/kpmg/xx/pdf/2020/11/the-time-has-come.pdf, retrieved 4.4.2021.

Kothari, C. (2004). *Research methodology: Methods and techniques*. ProQuest Ebook Central https://ebookcentral.proquest.com

Krefting, L. (1991). Rigor in qualitative research: The assessment of trustworthiness. *American journal of occupational therapy, 45*(3), 214-222.

LaBella, M. J., Sullivan, L., Russell, J., & Novikov, D. (2019). The devil is in the details: the divergence in ESG data and implications for responsible investing. *New York: QS Investors*.

Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. Sage.

Lopatta, K., & Kaspereit, T. (2014). The world capital markets' perception of sustainability and the impact of the financial crisis. *Journal of Business Ethics, 122*(3), 475-500.

Matthias, A. (2004). The responsibility gap: Ascribing responsibility for the actions of learning automata. *Ethics and information technology, 6*(3), 175-183.

Minutolo, M. C., Kristjanpoller, W. D., & Stakeley, J. (2019). Exploring environmental, social, and governance disclosure effects on the S&P 500 financial performance. *Business strategy and the Environment, 28*(6), 1083-1095.

Mittelstadt, B. (2019). Ai ethics–too principled to fail?. *arXiv preprint arXiv:1906.06668.*

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, 26*(4), 2141-2168.

MSCI (2020). *MSCI ESG rating methodology: Executive summary*. MSCI Inc. https://www.msci.com/documents/1296102/4769829/MSCI+ESG+Ratings+Methodology+-+Exec+Summary+Dec+2020.pdf/15e36bed-bba2-1038-6fa0-2cf52a0c04d6?t=1608110671584, retrieved 19.3.2021.

Nissenbaum, H. (2009). *Privacy in context*. Stanford University Press.

Phillips, R. A. (1997). Stakeholder theory and a principle of fairness. *Business Ethics Quarterly*, 51-66.

Phillips, R., Freeman, R. E., & Wicks, A. C. (2003). What stakeholder theory is not. *Business ethics quarterly*, 479-502.

PRI Association (n.d.). *What is responsible investment?*. https://www.unpri.org/an-introduction-to-responsible-investment/what-is-responsible-investment/4780.article, retrieved 18.4.2021.

PwC (2017). *Sizing the prize: What's the real value of AI for your business and how can you capitalise?*. PwC. https://www.pwc.com/gx/en/issues/analytics/assets/pwc-ai-analysis-sizing-the-prize-report.pdf, retrieved 22.3.2021.

Rahwan, I. (2018). Society-in-the-loop: programming the algorithmic social contract. *Ethics and Information Technology, 20*(1), 5-14.

Regulation 2016/679. *The protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC*. European Parliament, Council of the European Union. http://data.europa.eu/eli/reg/2016/679/oj

Reichardt, C. S. & Cook, T. D. (1979). Beyond qualitative versus quantitative methods in Cook, T. D. & Reichardt, C. S. (eds) *Qualitative and Quantitative Methods in Evaluation Research*. Sage.

Rogers, J., & Serafeim, G. (2019). Pathways to materiality: How sustainability issues become financially material to corporations and their investors. *Harvard Business School Accounting & Management Unit Working Paper*, (20-056).

Sandberg, J., Juravle, C., Hedesström, T. M., & Hamilton, I. (2009). The heterogeneity of socially responsible investment. *Journal of Business Ethics, 87*(4), 519-533.

Sanders, T. (2020). Testing the Black Box: Institutional Investors, Risk Disclosure, and Ethical AI. *Philosophy & Technology*, 1-5.

SASB (2021). *About us*. https://www.sasb.org/about/, retrieved 4.4.2021.

Schwab, K. (2016). The Fourth Industrial Revolution: what it means, how to respond. World Economic Forum. https://www.weforum.org/agenda/2016/01/the-fourth-industrial-revolution-what-it-means-and-how-to-respond/, retrieved 24.4.2021.

Schiff, D., Biddle, J., Borenstein, J., & Laas, K. (2020, February). What's next for AI ethics, policy, and governance? A global overview. *In Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society* (pp. 153-158).

Schiff, D., Borenstein, J., Biddle, J., & Laas, K. (2021). AI ethics in the public, private, and NGO Sectors: A review of a global document collection. *IEEE Transactions on Technology and Society*.

Scott, W. R. (1987). The adolescence of institutional theory. *Administrative science quarterly*, 493-511.

Shenton, A. K. (2004). Strategies for ensuring trustworthiness in qualitative research projects. *Education for information, 22*(2), 63-75.

Shields, L., & Twycross, A. (2008). Content analysis. *Paediatric nursing, 20*(6), 38.

Silverman, D. (2010). *Doing qualitative research*. SAGE Publications.

Silvola, H. & Landau, T. (2019). *Vastuullisuudesta ylituottoa sijoituksiin*. Alma Talent.

Singh, K. (2021*). Two dead in Tesla crash in Texas that was believed to be driverless*. Reuters. https://www.reuters.com/business/autos-transportation/two-dead-tesla-

crash-texas-that-was-believed-be-driverless-wsj-2021-04-18/, retrieved 29.4.2021.

Stebbins, R. A. (2011). What is exploration?. In *Exploratory research in the social sciences* (pp. 2-17). SAGE Publications, Inc.

Suchman, M. C. (1995). Managing legitimacy: Strategic and institutional approaches. *Academy of management review, 20*(3), 571-610.

Tamimi, N., & Sebastianelli, R. (2017). Transparency among S&P 500 companies: An analysis of ESG disclosure scores. *Management Decision*.

Tan, W. C. K. (2017). Research methods: A practical guide for students and researchers. ProQuest Ebook Central https://ebookcentral.proquest.com

Taylor, C. (1971). Interpretation and the sciences of man. *Review of Metaphysics*, 25, 3-51.

Timmermans, S., & Tavory, I. (2012). Theory construction in qualitative research: From grounded theory to abductive analysis. *Sociological theory, 30*(3), 167-186.

Truby, J. (2020). Governing Artificial Intelligence to benefit the UN Sustainable Development Goals. *Sustainable Development, 28*(4), 946-959.

Turilli, M., & Floridi, L. (2009). The ethics of information transparency. *Ethics and Information Technology, 11*(2), 105-112.

Vaismoradi, M., Turunen, H., & Bondas, T. (2013). Content analysis and thematic analysis: Implications for conducting a qualitative descriptive study. *Nursing & health sciences, 15*(3), 398-405.

Vakkuri, V., Kemell, K. K., Kultanen, J., & Abrahamsson, P. (2020). The current state of industrial practice in artificial intelligence ethics. *IEEE Software, 37*(4), 50-57.

Van Duuren, E., Plantinga, A., & Scholtens, B. (2016). ESG integration and the investment management process: Fundamental investing reinvented. *Journal of Business Ethics, 138*(3), 525-533.

Vinuesa, R., Azizpour, H., Leite, I., Balaam, M., Dignum, V., Domisch, S., Felländer, A., Langhans, S., Tegmark., M. & Nerini, F. F. (2020). The role of artificial intelligence in achieving the Sustainable Development Goals. *Nature communications, 11*(1), 1-10.

West, D. M. (2018). *The future of work: Robots, AI, and automation*. Brookings Institution Press.

Wong, C., & Petroy, E. (2020). *Rate the raters 2020: Investor survey and interview results*. SustainAbility.

https://www.sustainability.com/globalassets/sustainability.com/thinking/pdfs/sustainability-ratetheraters2020-report.pdf, retrieved 11.3.2021.

Yang, G. Z., Bellingham, J., Dupont, P. E., Fischer, P., Floridi, L., Full, R., Jacobstein, N., Kumar, V., McNutt, M., Merrifield, R., Nelson, B. J., Scassellati, B., Taddeo, M., Taylor, R., Veloso, M., Wang, Z. & Wood, R. (2018). The grand challenges of Science Robotics. *Science robotics, 3*(14), eaar7650.

APPENDIX

Translated interview structure

1. Could you tell a bit of your background with matters related to ESG?
2. Could you tell a bit of your background with matters related to AI, or the responsible use of AI in specific?
3. What kind of material ESG risks or opportunities do you associate with AI? To which of the three ESG pillars do you associate these risks or opportunities?
4. In your opinion, how well do the current ESG evaluation methodologies and scoring methods suit evaluating the (responsible) use of AI?
5. How is companies' responsible use of AI considered when conducting ESG analyses?
    a. Is there a difference between companies whose business revolves around AI (e.g. manufacturers of AI products) and companies who merely utilize it in their operations (e.g. purchase AI products to enhance their operations)?
    b. If the responsible use of AI is considered in ESG analyses, is it done only when companies report about utilizing AI in some way, or do investors/analysts look into whether companies utilize AI or not?
    c. Do the principles of responsible AI have a role in ESG analyses? If yes, how are they utilized (e.g. do investors check whether a company has their own set of principles, do companies need to provide evidence of operationalizing the principles, etc.)?
6. Would you say that some of the principles of responsible AI are more important than others from the investors' viewpoint? If yes, which one(s) and why?
    a. Is there, or should there be a larger emphasis on complying with some principles? E.g. should some principles be prioritized if there is a conflict between them?
7. How would you say the responsible use of AI or the principles of responsible AI will be visible in ESG analyses in the future?
8. Should certain principles of responsible AI be universal for all AI applications, or should applications used in different industries have different sets of principles for them?