AI & SOCIETY (2024) 39:329–343
https://doi.org/10.1007/s00146-022-01415-0

OPEN FORUM

What about investors? ESG analyses as tools for ethics-based AI auditing

Matti Minkkinen[1] · **Anniina Niukkanen**[1] · **Matti Mäntymäki**[1]

Received: 2 December 2021 / Accepted: 17 February 2022 / Published online: 9 March 2022

Abstract
Artificial intelligence (AI) governance and auditing promise to bridge the gap between AI ethics principles and the responsible use of AI systems, but they require assessment mechanisms and metrics. Effective AI governance is not only about legal compliance; organizations can strive to go beyond legal requirements by proactively considering the risks inherent in their AI systems. In the past decade, investors have become increasingly active in advancing corporate social responsibility and sustainability practices. Including nonfinancial information related to environmental, social, and governance (ESG) issues in investment analyses has become mainstream practice among investors. However, the AI auditing literature is mostly silent on the role of investors. The current study addresses two research questions: (1) how companies' responsible use of AI is included in ESG investment analyses and (2) what connections can be found between principles of responsible AI and ESG ranking criteria. We conducted a series of expert interviews and analyzed the data using thematic analysis. Awareness of AI issues, measuring AI impacts, and governing AI processes emerged as the three main themes in the analysis. The findings indicate that AI is still a relatively unknown topic for investors, and taking the responsible use of AI into account in ESG analyses is not an established practice. However, AI is recognized as a potentially material issue for various industries and companies, indicating that its incorporation into ESG evaluations may be justified. There is a need for standardized metrics for AI responsibility, while critical bottlenecks and asymmetrical knowledge relations must be tackled.

Keywords Artificial intelligence · AI · Auditing · ESG investing · Responsible investment · Ethics

1 Introduction

As the use of artificial intelligence (AI) spreads across industries and societies, it has become evident that AI-based systems need to be effectively governed to avoid, or at least mitigate, potential risks and harms (Gasser and Almeida 2017; Butcher and Beridze 2019). AI governance requires not only ethical principles and guidelines but also and especially actionable tools and mechanisms (Brundage et al. 2020; Shneiderman 2020; Mäntymäki et al. 2022). Translation from abstract principles to practical tools for AI development and governance (Morley et al. 2020; Seppälä et al. 2021; Mäntymäki et al. 2022) is a critical challenge that hinders the widespread adoption and beneficial use of AI (Cowls et al. 2021).

To address this issue, AI auditing, also termed algorithmic auditing (Raji et al. 2020), has been proposed as a tool for operationalizing and assessing AI governance (Sandvig et al. 2014; Koshiyama et al. 2021). However, AI auditing practices are still emerging, and AI auditing suffers from critical shortcomings, such as loose definitions (ForHumanity 2021), limited scope (Cath 2018; Kroll 2018), difficulty quantifying externalities (Rahwan 2018), and lack of information needed to evaluate AI systems (Mökander and Floridi 2021). There are numerous sets of AI ethics principles (Jobin et al. 2019; Schiff et al. 2020), and there is a near consensus on key principles, such as transparency, fairness, non-maleficence, responsibility, and privacy (Jobin et al. 2019; Dignum 2020). However, AI auditing is hindered by a lack of standardized metrics for ascertaining levels of fairness, transparency, and other desirable system characteristics. This means that AI auditing is facing criticisms for its ambiguity, as well as the loose and casual use of the term "audit" by AI consultancies (ForHumanity 2021).

✉ Matti Minkkinen
matti.minkkinen@utu.fi

[1] Turku School of Economics, University of Turku, 20014 Turku, Finland

Impending legislation, particularly the European Union's Artificial Intelligence Act (AIA), is expected to bring clear rules to the AI auditing landscape. However, at the time of writing, the AIA is under development; its finalized form remains to be seen. More importantly, binding legislation provides only the minimum requirements and is likely to cover only particular high-risk AI systems. This leaves significant scope for AI governance beyond the minimally acceptable level.

Organizations that aspire to go beyond minimal legal compliance need to pay particular attention to corporate governance and corporate social responsibility (CSR), as well as stakeholder pressure coming from investors, among other groups. Over the past decade, the use of environmental, social, and governance (ESG) criteria among investors has risen sharply partly as an outgrowth of CSR and, more recently, corporate sustainability discussions (Boffo and Patalano 2020). The acronym ESG thus denotes environmental, social, and governance criteria for evaluating firm performance and screening potential investments. Perhaps the best known among the ESG domains is environmental sustainability, whereby ESG criteria investigate, for example, the climate change impacts of companies' operations. Because AI governance also seeks to mitigate potentially widespread harm, the question of how AI impacts and AI governance could be included in ESG evaluations is warranted. In current discussions on AI risks (e.g., Floridi et al. 2018; Dignum 2020), the social and governance dimensions of ESG are the most prominent.

The current AI governance and auditing literature is mostly silent on the role of investors as stakeholders in the AI governance landscape (for exceptions, see Shneiderman 2020; Brusseau 2021). Investors should be brought into the picture in AI governance scholarship because they have significant financial power to influence companies' operations and governance practices.

Against this backdrop, we conducted an exploratory study on the connections between ethics-based AI auditing and ESG criteria. Using semi-structured interviews, our study seeks to answer the following research questions:

1. How is the responsible use of AI taken into consideration in ESG investment analysis?
2. What connections can be found between the existing principles of responsible AI and the ESG ranking criteria?

This paper proceeds as follows. First, we introduce ethics-based AI auditing and ESG investing in the background section. Then, we present our material and methods, consisting of thematic analyses of interview material. Next, we present the key findings grouped under three themes. Finally, in the discussion and conclusion, we draw out the implications of our study by considering ESG analyses as tools for AI auditing, concluding with limitations and future research directions.

This study contributes to two bodies of knowledge. First, we contribute to the ethics-based AI auditing literature (Sandvig et al. 2014; Koshiyama et al. 2021; Mökander et al. 2021) by exploring the role of ESG criteria in providing metrics for AI auditing and highlighting the role of investors as a source of stakeholder pressure. Second, we contribute to the ESG literature (van Duuren et al. 2016; Sætra 2021) by advancing the understanding of AI impacts and AI governance as potential domains in future ESG analyses.

2 Background

2.1 AI auditing and responsible AI principles

While AI promises efficiency gains and numerous organizational and societal benefits, the opaqueness of algorithmic systems and the risk of unintended consequences remain key challenges (Dignum 2020). AI auditing, also referred to as algorithmic auditing, has been proposed as a means of assessing and governing inscrutable algorithms (Sandvig et al. 2014; Raji et al. 2020; Koshiyama et al. 2021). AI auditing provides procedures to assess claims about algorithmic systems—for instance, with regard to fairness and lawfulness—and thus supports the overall goal of enabling verifiability of claims about AI development and use (Brundage et al. 2020). Thus, AI auditing grants control and oversight over AI systems to strengthen their accountability. In addition to its benefits for AI user organizations, AI auditing has been envisaged as an emerging industry (Koshiyama et al. 2021), which may spur significant economic activity.

In the broad spectrum of AI auditing, which encompasses, for example, technical audits (Brundage et al. 2020; Koshiyama et al. 2021), we focus on ethics-based auditing. Mökander et al. (2021, p. 4) define ethics-based AI auditing as a "structured process whereby an entity's present or past behavior is assessed for consistency with relevant principles or norms." Ethics-based auditing is meant to spark ethical deliberation among developers and managers and identify the values embedded in a system. As such, it does not seek to codify ethics nor provide legal auditing (Mökander et al. 2021).

For ethics-based AI auditing, numerous frameworks and technical tools have been developed (Mökander et al. 2021). However, the emerging AI auditing industry and practitioner community lack metrics and standards, as well as mechanisms to incentivize organizations to adopt robust AI auditing practices (Floridi et al. 2018). Starting points for metrics are provided by the numerous sets of responsible AI principles that have been published in recent years. At the

time of writing, there are at least 173 AI ethics guidelines,[1] as well as several published meta-level overviews of ethical AI principles (Clarke 2019; Jobin et al. 2019; Schiff et al. 2020; Hagendorff 2020). A comprehensive overview of AI ethics guidelines found global convergence on five ethical principles: transparency, justice and fairness, non-maleficence, responsibility, and privacy (Jobin et al. 2019). This list closely mirrors other synthesizing lists, such as beneficence, non-maleficence, autonomy, justice, and explicability (Floridi et al. 2018), as well as accountability, responsibility, and transparency (Dignum 2020).

The proliferation of abstract AI ethics principles has exacerbated the so-called "translation problem"—that is, how these high-level principles can be practically implemented in AI development processes and organizations' day-to-day operations (Morley et al. 2020). Binding legislation is a key mechanism for implementing ethics principles. In this domain, the proposed European Union (EU) AI Act (European Commission 2021) envisions a system of actors, such as national supervisory authorities and notified bodies, involved in certifying the compliance of high-risk AI systems with the new regulation (Stix 2022).

While the legal compliance model brings strong incentives for organizations through binding legislation, regulatory requirements cannot cover all aspects of responsibility for algorithmic systems. This is particularly true for the large set of systems that are not categorized as high-risk in the EU scheme and are thus not as comprehensively regulated.

Because regulation cannot cover everything, effective AI governance is not only about legal compliance; organizations should also strive to go beyond legal requirements in the development of fair, transparent, and accountable AI systems (Dignum 2019; Trocin et al. 2021). Binding legislation leaves space for what Floridi (2018, p. 5) calls soft ethics—that is, considering "what ought and ought not to be done *over and above* the existing regulation."

Within the field of corporate governance, these voluntary commitments have been discussed in terms of CSR (e.g., Maon et al. 2009) and corporate sustainability (e.g., Hahn et al. 2015; van der Waal and Thijssens 2020). This literature is rooted in stakeholder theory (Freeman 2010), which holds that companies should create value for various stakeholder groups, such as customers, employees, and communities, rather than solely maximizing shareholder profits. Stakeholder theory argues 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 (Freeman 2010). Especially since the publication of the UN Sustainable Development Goals in 2015, the CSR field has shifted to corporate sustainability actions and reporting (van der Waal and Thijssens 2020; Sætra 2021). In parallel, over the past decade, scholars and practitioners have paid increasing attention to so-called ESG analyses (Cort and Esty 2020), the topic of the next section, as a tool for incentivizing and promoting responsible corporate practices.

2.2 ESG investing

Concurrently with the growing societal emphasis on sustainability, companies are increasingly expected to report on their sustainability-related performance (Sætra 2021). Investors have also turned to sustainability criteria and investment styles such as sustainable investing (GSIA 2018). ESG investing, the topic of this paper, can be distinguished from this broader sustainability movement by its stronger focus on risk mitigation to secure financial returns (Boffo and Patalano 2020). As paying attention to ESG-related matters has started gaining popularity, especially in the past decade, a vast array of terminology has emerged (Cort and Esty 2020). This section introduces central aspects of ESG investing and then investigates why ESG issues are increasingly considered by different investor groups.

ESG investing complements current AI auditing approaches by focusing on company-level processes that cut across algorithmic systems. AI auditing, as it is currently understood, is often conducted at the level of algorithmic systems—that is, investigating each system as a separate entity and assessing its conformity with particular requirements (Sandvig et al. 2014). From the investor perspective, the appropriate unit of analysis is a company that may develop and use several algorithmic systems. Even though private companies are discussed here, the organizational analysis of AI use and development applies equally to public and third-sector organizations.

Depending on the source, ESG investing may be used as a common term for different investment styles that take ESG issues into account, or it can be considered as a distinct style (Hill 2020, pp. 13–14). There are numerous styles of integrating nonfinancial information related to ESG dimensions into investment analyses and decision-making processes. To distinguish the set of different investment styles from the specific ESG investing approach, terms such as sustainable investing, used by the Global Sustainable Investment Alliance (GSIA 2018), have been utilized. In this study, ESG investing' is considered its own distinct investment style, whereby ESG issues are integrated into investment analysis to mitigate risk with the aim of securing financial returns (Boffo and Patalano 2020, p. 14).

ESG consists of three dimensions: environmental, social, and governance. The environmental dimension covers issues

[1] For an updated inventory of ethics guidelines, see https://algorithmwatch.org/en/ai-ethics-guidelines-global-inventory/.

Table 1 ESG dimensions with example issues

Dimension	Example issues
Environmental	Climate change, waste, pollution, biodiversity
Social	Human rights, modern slavery, child labor, product responsibility
Governance	Corruption, board diversity, executive compensation, tax strategy

related to the use of natural resources, the effect companies have on the environment on both the local and global scale, and how companies work on reducing their emissions from their own operations and throughout their supply chains. The social dimension is concerned with how a company treats its own workforce or how fair treatment of the workforce in supply chains is managed, as well as how its operations and products affect its other stakeholders, including customers. The governance dimension, in turn, is related to enabling and enhancing the ethical conduct of business within a company and ensuring that good corporate governance is practiced in all aspects of its operations (Boffo and Patalano 2020; Hill 2020). Some key issues related to each dimension are listed in Table 1 for illustrative purposes.

No single, standard method exists for taking ESG issues into account, and investors can leverage different strategies both before making an investment and with existing assets (van Duuren et al. 2016; Amel-Zadeh and Serafeim 2018; Cort and Esty 2020). *Full ESG integration* means the inclusion of ESG issues in the investment analysis, along with traditional financial measures, either by evaluating companies in isolation or by comparing how different companies perform compared to each other regarding larger global or sector-specific issues (Amel-Zadeh and Serafeim 2018). As this strategy has been associated with higher implementation costs compared to traditional investing (Kempf and Osthoff 2008; van Duuren et al. 2016), the use of less arduous strategies is also common. For example, *negative screening* refers to simply excluding companies from investment portfolios either for ethical reasons or for aligning a portfolio with an investor's personal values or preferences (PRI Association 2019). Through *active ownership*, investors can also guide the investees' engagement in ESG issues, both in mitigating ESG risks and guiding companies toward sustainable operations (PRI Association 2019).

In ESG investing, ESG issues are considered *material*—that is, relevant to an asset's future financial performance—thus making their integration into the investment analysis necessary for capturing greater benefits from the investment (Cort and Esty 2020). Even though ethical reasons or values are not the main motivation for engaging in certain investments, they may 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 so-called sin stocks and may thus exclude them from their portfolios (Hong and Kacperczyk 2009).

ESG investing resembles socially responsible investing (SRI) and impact investing. However, ESG investing differs from SRI because in ESG investing, ethical reasons and personal values are not the main drivers of the investment decisions (Sandberg et al. 2009). ESG investing also differs from impact investing, which aims to achieve a particular positive environmental or social return with the investment, such as financial inclusion, education, or the promotion of renewable energy (Hill 2020, p. 18).

Four key drivers of ESG issues have gained attention over roughly the past decade. First, research indicates that considering ESG issues may, in fact, benefit financial performance. The findings of a meta-analysis of over 2,000 empirical studies that measured the connection between ESG performance and corporate financial performance indicated that ESG integration might, in fact, be beneficial for financial gains (Friede et al. 2015). The majority of the analyzed studies found a positive correlation between ESG and financial performance, as the companies that received higher ESG ratings produced comparable financial returns compared to the companies with lower ESG ratings. Second, the growing attention to issues such as climate change, standards of responsible business conduct, and diversity in the workplace and on boards will impact consumer choices and, thus, company performance (Boffo and Patalano 2020). For portfolio and asset managers, the views of their own customers may also drive them toward more sustainable investment choices (Amel-Zadeh and Serafeim 2018). The growing interest in ESG issues will likely increase as younger generations are more active in terms of incorporating their values into their investment decisions (Boffo and Patalano 2020, p. 17; Hill 2020, p. 3). Third, 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 (Boffo and Patalano 2020). The long-term view is based on the notion that integrating ESG issues into investment analysis affects the long-term risk and financial performance of investment portfolios (MSCI 2020, p. 2). Fourth, new ESG regulations, such as the proposed EU Corporate Sustainability Reporting Directive, will set new requirements for company reporting on ESG issues (Santoro et al. 2021).

2.3 Measuring and reporting ESG compliance

Including ESG information in investment analyses has evolved from being a practice of ethical investors to becoming popular among mainstream investors as well (van Duuren et al. 2016). Companies are under increasing pressure to provide reliable ESG data for their various stakeholder groups (Cort and Esty 2020). As a result, the number of ESG data providers has increased in the past decade (van Duuren et al. 2016). Agencies that are often mentioned in the academic literature include MSCI, Sustainalytics, Refinitiv, Vigeo Eiris, RobecoSAM, and Bloomberg ESG (e.g., Escrig-Olmedo et al. 2019; Berg et al. 2020). MSCI and Sustainalytics are also the most favored rating agencies among investors due to their wide coverage of companies (Wong and Petroy 2020, pp. 14, 33–35). While the rating agencies measure and compare how companies consider similar ESG issues in their business practices, their methods and results can vary greatly. The divergence within ESG ratings, due to differences in scope, measurement, and weighting, has been confirmed in multiple studies (e.g., Berg et al. 2020; Chatterji et al. 2016; Dorfleitner et al. 2015). Thus, a company may be evaluated as sustainable in a certain category by one agency while being deemed unsustainable by another.

In addition to each investor focusing on matters that he or she deems 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 to their performance. Even companies within the same industries have been found to report on different issues and use incompatible reporting styles, thereby making it difficult for investors to compare companies within industries (Cardoni et al. 2019). These differences may lead to investors and other stakeholders questioning which ESG issues are truly financially material to them (Cort and Esty 2020).

In response, many organizations, such as the Global Reporting Initiative (GRI) and the Sustainability Accounting Standards Board (SASB), have published guidelines on how companies should report their sustainability issues and risk-mitigation efforts. The use of such guidelines has become common among companies, as 84% of the 250 largest companies in the world utilize some form of external framework in their reporting (KPMG International 2020). However, guidelines leave space for companies to consider which issues to report. Because reporting on governance issues is required in the US by the Securities and Exchange Commission, companies have been less transparent regarding social and environmental issues than governance issues (Tamimi and Sebastianelli 2017).

Seeking to clarify which issues should be considered material, Rogers and Serafeim (2019) investigate how and why ESG issues turn material over time and how stakeholders affect this pathway to materiality. They further propose a framework consisting of five stages: status quo, catalyst, stakeholder response, company response, and regulatory response (Rogers and Serafeim 2019). According to the framework, companies may initially have a negative societal impact regarding an ESG issue that is still considered immaterial. At this point, this impact is not considered problematic, or the level of negative impact is not properly understood ("status quo"). According to Rogers and Serafeim (2019), there are two catalysts that may initiate the process of an ESG issue turning material: companies gaining excessive profits and causing negative impacts, leading to public awareness, or a change in societal norms against which the acceptability of companies' operations is measured due to increased information ("catalyst"). In the next stage of the stakeholder response, the issue may become material to companies that have gained an excessive amount of negative publicity in the eyes of their stakeholders ("stakeholder response"). The whole industry may thus engage in attempts at self-regulation to limit the possibility of regulators taking further interest in the issue ("company response"). If the actions of companies are not seen as adequate, regulatory bodies may start enforcing new laws ("regulatory response") to mitigate the negative impact, leading to the issue becoming financially material to the whole industry (Rogers and Serafeim 2019).

2.4 ESG and AI

While separate literature streams on AI auditing and ESG investing have emerged over the past decade, we were able to find minimal research explicitly connecting the two topics discussed in this paper: ESG analyses and AI. It is important to distinguish the problem area of ESG analyses and AI from discussions on using AI to make more robust ESG assessments (e.g., Selim 2020). While the latter is also an important area of research, it is not the focus of this paper. In contrast to using AI to conduct ESG analyses, we seek ways to use ESG methods to audit AI.

While searching academic databases, we found only a small number of relevant papers, and only from the year 2021.[2] This indicates that the connection between ESG and AI is an emerging area of study. Among the first academic contributions to this topic, Brusseau (2021) criticizes the use of the current ESG rating methods for evaluating the effects of AI, starting his paper with the subtitle "ESG does not work for AI." He makes this statement on the basis that ESG issues have traditionally been related to larger targets, such as ensuring that all employees are treated fairly.

[2] We consulted two major academic databases: Scopus and Web of Science Core Collection.

Table 2 Literature on ESG and AI

Source	Focus
Brusseau (2021)	Inappropriateness of collective ESG evaluation in the context of risks to individuals, hence the need for an individual-based model
Sætra (2021)	Sustainability-related impacts at the micro, meso, and macro levels
Du and Xie (2021)	Framework for ethical challenges and CSR issues at the product, company, and society levels

Instead, Brusseau (2021) 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. Thus, AI-intensive companies should be targeted with an evaluation that "begins with unique persons, not demographic segments or collectives" (Brusseau 2021, p. 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 AI issues, and it assigns scores from 0 to 2 to each principle based on how well a company takes the related issues into consideration. The list of principles is like many other lists of ethical AI principles (cf. Floridi et al. 2018; Jobin et al. 2019).

Sætra (2021), in turn, proposes a framework for evaluating the ESG-related impacts of AI using the United Nations Sustainable Development Goals (SDGs). In this framework, the negative and positive micro-, meso-, and macro-level impacts, as well as the ripple effects between the different impacts, are considered for each SDG, promoting a holistic perspective (Sætra 2021). However, this approach considers only impacts—specifically, sustainability-related impacts—while AI governance is generally considered to also include the governance of organizational processes (Schneider et al. 2020; Eitel-Porter 2021).

In addition, Du and Xie (2021) touch on ESG investing and AI indirectly. Within the domain of consumer products, they develop a framework for considering ethical challenges and CSR issues at the product, consumer, and society levels (Du and Xie 2021). However, their framework does not include auditing, the investor perspective, nor ESG analyses. Moreover, the framework focuses on AI-enabled consumer products, although the domain of AI also encompasses products and processes within and between organizations that do not necessarily interact directly with consumers. Table 2 draws together the key foci from the literature on ESG and AI.

Given the early stage of the academic research on ESG and AI, key questions remain largely unanswered. For example, are the current ESG rating frameworks of rating agencies suitable for evaluating the effects of AI? How can investors take the responsible use of AI into consideration in their investment analyses? Thus, there is a need for explorative research on how best to combine these two rising topics. We present our explorative research process in the next section.

3 Materials and methods

This study set out to investigate how investors currently understand questions related to the responsible use of AI and the potential connections between ESG analyses and responsible AI principles. To this end, semi-structured interviews were conducted with a purposefully sampled (Palinkas et al. 2015) set of Finnish senior-level experts in ESG investing, responsible AI, or both areas. Furthermore, as this area is still emerging in the academic literature, an exploratory study to unravel how professionals perceive the related questions was considered suitable. Five interviews were conducted via Microsoft Teams, and the interviews were recorded and transcribed into text. Information about the informants is summarized in Table 3.

The interview protocol can be found in Appendix 1. In addition to open thematic questions, the semi-structured interviews also included more specific questions related to the principles of responsible AI to ensure that they would be covered in the interviews even if they are not currently considered by investors. In addition to these questions, follow-up questions were presented to gain further understanding of the interviewees' statements.

Thematic analysis was used to analyze the interview data. Braun and Clarke (2006) see thematic analysis as a foundational method for qualitative analysis. Owing to its flexibility, it can suit a range of different methodological

Table 3 The informants' profiles

Participant	Job title/focus	Organization focus	Interview length (min)
P1	CEO	AI products	52
P2	CEO	AI products	47
P3	Responsible investment	Banking	52
P4	Responsible investment	Pension insurance	48
P5	Responsible investment	Asset management	34

positions and provide a specified frame for qualitative analysis and for reporting results, thus enhancing the possibility of evaluating the research findings later. The purpose of thematic analysis is to identify patterns or themes within the data, organize them, and report them (Braun and Clarke 2006).

The thematic analysis proceeded through the stages of familiarization with the data, forming initial codes by reading through the material several times, searching for broader themes among the codes, and checking and further defining the themes (Braun and Clarke 2006). The appearance of a theme numerous times in multiple places is not a requirement in thematic analysis (Braun and Clarke 2006). This is a clear distinction from other commonly used qualitative methods, such as content analysis, which relies more clearly on finding themes or categories with a high number of occurrences (Vaismoradi et al. 2013). Simply considering the number of occurrences has been criticized for possibly taking the data out of context, as the same type of codes may appear in the data for different reasons, such as informants being more comfortable with certain topics (Twycross and Shields 2008; Vaismoradi et al. 2013). As this study touches on the two relatively young fields of ESG and ethics-based AI auditing, considering the context of comments is necessary; this supports the adoption of thematic analysis instead of, for example, content analysis.

4 Findings

4.1 Overview of findings

The analysis of the interview data resulted in three major themes: (1) awareness of AI issues, (2) measuring AI impacts, and (3) governing AI processes. *Awareness of AI issues* contains views related to how investors currently understand AI-related issues. *Measuring AI impacts* contains findings related to how the impact of AI is currently considered, as well as the kinds of elements that would need to be taken into account when analyzing the impact of AI that a company uses in its operations. *Governing AI processes* contains views on how AI-related processes should be governed and the possibility of using AI ethics principles as the evaluation criteria. A thematic map of the findings is presented in Fig. 1.

The themes are logically connected to one another, as shown in Fig. 1, because measuring AI impacts and governing AI processes require awareness and identification of relevant issues. A summary of the findings is provided in Table 4, as well as example interview excerpts, and the themes are further discussed in the following subsections.

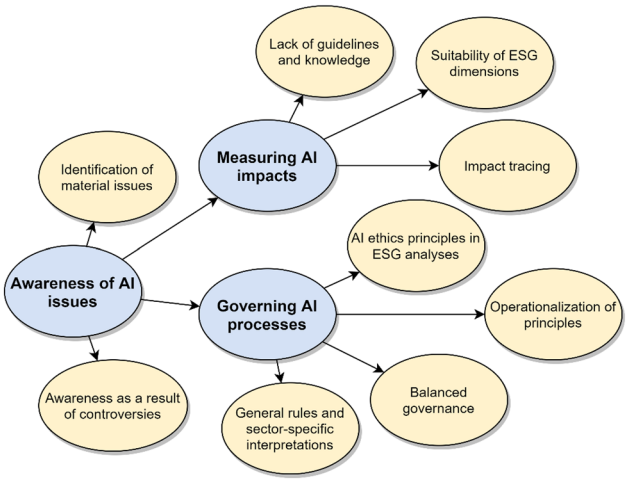

Fig. 1 Thematic map of the findings

4.2 Awareness of AI issues

Awareness of AI issues was the first main theme that emerged from the interview material. In general, awareness of AI-related issues is still comparatively low among investors. All interviewees found that the responsible use of AI and the ethics of AI are still not everyday topics in the work of investors, and the risks of AI were not considered topics that would be commonly included in ESG evaluations.

The interviewees highlighted the necessity of contextual understanding: Different companies use AI in various ways; thus, they need to consider different ethical issues to ensure that they are conducting business responsibly. One interviewee stated that for media companies, the risks might be related to user privacy issues or to exposing their customers to filter bubbles through recommendation systems; these issues would likely not be relevant to industrial companies' core operations. Consequently, not all issues related to ethical AI would need to be considered with regard to all companies; rather, the *material* issues—that is, the issues with financial, societal, and/or business continuity relevance—should be recognized for each one. Material issues were recognized as a starting point for evaluating how responsible companies are regarding their AI use: "I would start with the material issues" (P3).

Because there is considerable variance in the type of ESG issues that companies face, companies' responsibilities with regard to recognizing their own material issues were also agreed on during the interviews. For example, P2 stated that companies should be able to provide information regarding their AI use in a similar manner to other topics that are included in their sustainability reports. Instead of investors or rating agencies having a detailed list of potential issues related to, for example, human rights risks caused by the use of AI, the interviewee considered recognizing these granular

Table 4 Themes and subthemes

Themes and subthemes	Example interview excerpts
Awareness of AI issues	
Identification of material issues	"I would start with the material issues—which effects of using AI are the most material ones?—and I'd first pay attention to those. […] And then, when we would know what the most material effects are, then we could start thinking about how those should be evaluated, how we could compare the effects between different companies or industries, or what the best practice solutions are for a given industry." (P3)
Awareness as a result of controversies	"I think people pay attention to it when an issue comes out in public. 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)
Measuring AI impacts	
Lack of guidelines and knowledge	"The topic is so new that, at least to my knowledge, there is no standard for investors." (P3)
Suitability of ESG dimensions for measuring to measure AI impacts	"If we could just adapt the criteria in these pillars—similarly to how they are adapted easily to other areas of business—and I actually found a lot of suitable indicators that could be transferred for discussions related to the responsible use of AI." (P2)
Impact tracing	"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 [the potential issue] is in the supply chain, it turns into a question of how the supply chain is managed instead." (P4)
Governing AI processes	
AI ethics principles in ESG analyses	"If a company has a set of […] accepted principles, everyone believes that they are followed until 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." (P5)
Operationalizing principles	"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 in whether they are operationalized. That is probably the worst possible outcome." (P2)
Balanced governance	"If we go overboard with establishing these matters, then there could be pressure from the market that this is 'too ethical." (P1)
General rules and sector-specific interpretations	"But then again inevitable sector-specific or market-specific interpretations and pressures due to financial reasons will likely form because this is where money simply talks. And there will be more responsible companies, and I believe they will eventually be the winners, but they must be able to lean on the moral principles and on the interpretations of the rules and solutions that have been agreed on." (P1)

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 raised. While it was generally agreed that companies should be responsible for reporting on their specific AI-related issues, companies may be unable to recognize all the areas in which 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).

Because AI at its current scale of deployment is still a relatively young technology, the informants considered that there might not be enough information to recognize potential issues beforehand. Therefore, at least some issues may become material only reactively through controversies as they become apparent and as "these kinds of special cases are scrutinized on a more detailed level with some companies" (P2). Due to AI being a highly scalable and widespread technology, controversies in one part of the world may quickly cause an issue to become material elsewhere as well (P5). As enough controversies accumulate, "these things will probably come up as a topic that should be included in the ratings" (P2). Questions regarding what type of AI the investment asset has in use and how the particular company uses AI mainly emerge 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 public when something has not gone according to plan" (P3).

The global scale of AI deployment may also lead to various issues being considered material in different parts of the world. The informants stated that a company that operates in multiple countries and continents needs to take local norms from each location into consideration. For example, the

question of whether AI should be allowed to replace human workers, thus risking increased unemployment, was brought up by different interviewees, one of whom reminded us that some cultures would be heavily opposed to the possibility of increased unemployment.

4.3 Measuring AI impacts

Issues related to measuring the impact of AI in ESG analyses constituted the second main theme that emerged during the interviews. Based on the interviews, this topic is still relatively unknown in the investors' work. This indicates that few investors would include it in their analyses unprompted—for example, without a prior AI-related issue, which would have been communicated to stakeholders and thus seen as a potential threat to the investment.

The environmental and social dimensions of ESG were mentioned as being affected by the use of AI. The environmental dimension was the primary focus of impact considerations in the first interview only. P1 described the need for efficient computing power to enable increasingly powerful AI, which in turn leads to higher requirements for electricity to power such systems. This was thus deemed a potential disadvantage 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.

Regarding the social dimension, while both the positive and negative effects of AI were discussed during the interviews, the possible risks received significantly more attention. Various groups of people, such as employees, were mentioned as possible targets of these risks; for example, the increased use of automation could lead to layoffs. It could also make consumers unwilling sources of high-quality training data for AI systems, as there is an increasing need for such data. In addition, consumers could become victims of misbehaving AI. Bystanders were also recognized as possible victims in scenarios in which an AI system would make decisions that could affect their lives, with the risk that autonomous vehicles could pose for pedestrians also being presented as an example. P4 reminded us of the relatively weak bargaining power of consumers, which leads to the need to protect them: "Companies are able to negotiate such terms with each other to ensure 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 consensus among the informants was that the impact of AI is not commonly included in ESG evaluations. However, all the interviewees could identify potential risks that AI could pose to the ESG dimensions. Moreover, the interviewees identified the need to take related issues into account, but the lack of guidelines and knowledge of AI in general was considered an obstacle to reliable evaluations of the use of AI. As one informant put it, "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).

With respect to the suitability of current ESG rating mechanisms for evaluating the impacts of AI, the informants expressed divergent views. Some stated that in their view, the current ratings either are or would be suitable for AI products as well and that the rating agencies can develop their ratings to take the impact of AI into account in the future. One interviewee found "a lot of suitable indicators which could be transferred for discussions related to the responsible use of AI," giving whistleblowing practices as an example (P2). However, the difficulty of developing ratings was also recognized: "It is likely difficult to take [the responsible use of AI] comprehensively into account as a whole" (P5).

In contrast, 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 into account properly: "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 them. And this is caused by my not having a sufficient level of knowledge related to this topic" (P3).

However, the perceived novelty of AI topics may be partly illusory. P3 also stated that in the past, when ESG was not yet a mainstream topic among investors, some portfolio managers would claim that they did not consider ESG meaningful. Nevertheless, topics in the ESG governance dimension were already de facto included in their own analyses. Drawing on this example, P3 stated that there could be elements in the current ratings that would be suitable for evaluating AI as well. 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; […] it is probably caused by the fact that even the overall understanding [of the topic] needs to be increased as well" (P4).

The question of whether existing indicators could be applied to the evaluation of the impact of AI or whether it should be its own theme, similar to climate change or protecting biodiversity in many ESG ratings, was also raised by the informants. P4 mentioned that the role that AI will play in the future would influence whether a new theme is needed for AI. If we expect AI to develop in a direction wherein it could cause severe damage to societies, a designated theme

would be justified, but based on current knowledge, the informant (P4) deemed it unnecessary to dedicate a separate topic to AI. One interviewee also reminded us that some have suggested technology or digitality as an added dimension in ESG evaluations: "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" (P2). However, following the previous statements, this was not considered necessary either; rather, adapting the current frameworks was seen as a better option.

To complicate matters, AI-related impacts may be as yet unforeseen because technologies are advancing at a rapid pace: "All current consideration is related to our current way of working and to our current worldview […], but I also feel like it may have different kinds of impact that we are not able to evaluate or think about at this stage" (P3). The difficulty of foreseeing potential impacts is particularly the case in supply chains whereby companies purchase AI products from others, thus expanding the evaluation scope: "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 [the potential issue] is in the supply chain, it turns into a question of how the supply chain is managed instead" (P4).

4.4 Governing AI processes

Governing AI processes was the third main theme in the interview material. Considering the governance dimension of ESG, AI ethics principles provide a starting point, but their translation into feasible governance requirements needs work. According to the interviewees, AI ethics principles are still a somewhat unfamiliar concept within the investment world. Thus, how and why AI should be used in an ethically sound manner was also deemed a topic that is not yet well understood. As a result, a company lacking its own set of principles was not considered a major factor in ESG analyses or investment decisions. Still, one interviewed expert on the investor side stated that the principles can be included in investment analyses if the company is in the AI business and has 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: "When [a set of principles] is written and published, someone has to monitor it" (P5). Additionally, the principles of responsible AI were seen as a possible way of communicating that material issues were recognized and considered.

However, another interviewee stated that companies should be able to explain how they intend to operationalize the principles. Verifying that a company has a set of principles available would merely be a starting point for assessing the company. As this interviewee put it, "Communicating about the principles is the first step, and it indicates that the issue has been recognized," and then the scrutiny should turn to "how they are able to communicate about the mechanisms with which the principles are carried out" (P2).

Overall, the AI ethics principles were considered viable data points for AI-related ESG evaluation. P1 did, however, remind us that following the principles too strictly could lead to dissatisfaction toward companies: "I think they [principles of responsible AI] are a very good fit [for 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'" (P1). Conversely, the possibility of using the principles of responsible AI as "ethical greenwashing" (ethics washing) was also raised; companies could claim to be responsible in their AI operations and exaggerate the benefits while keeping investors' attention away from potential issues. This was seen as a possible risk, especially in the current stage, when investors' knowledge related to the matter is still generally low, thereby making the identification of these types of cases difficult. However, this was also considered a learning opportunity for the investors: "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, an expert in responsible AI, stated that as some principles have been found to be central in many of the released guidelines, those could be considered generally accepted universal principles for a variety of industries, with transparency and accountability being mentioned here as examples. Risks related to biased results or threats to user privacy were considered central to AI, according to the interviewees. The fact that privacy and cyber security for some AI products could already be considered central topics, primarily because they are already a part of many prominent ESG rating frameworks, was also discussed. Overall, however, the consensus was that there is still a long way to go before AI responsibility will be properly considered on a wider scale: "Privacy protection is taken into account [in ESG ratings], but in reality, it [the 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).

P1 argued that there should be common principles that act as general guidelines for all companies within industries, providing them with fair rules and limiting irresponsible actors from reaping excess benefits by disregarding ethical considerations. However, the interviewee also believed that there would inevitably be sector-specific or market-specific interpretations of the same principles:

Table 5 Themes from the thematic analysis divided into ESG dimensions

Dimension	Relevant themes and subthemes
Environmental	Awareness of AI issues – Identification of material issues – Awareness as a result of controversies Measuring AI impacts – Lack of guidelines and knowledge – Suitability of ESG dimensions for measuring AI impacts – Impact tracing
Social	Awareness of AI issues – Identification of material issues – Awareness as a result of controversies Measuring AI impacts – Lack of guidelines and knowledge – Suitability of ESG dimensions for measuring AI impacts – Impact tracing
Governance	Governing AI processes – AI ethics principles in ESG analyses – Operationalization of principles – Balanced governance – General rules and sector-specific interpretations

> *"But then again, inevitable sector-specific or market-specific interpretations and pressures due to financial reasons will likely form because this is where money simply talks. And there will be more responsible companies, and I believe they will eventually be the winners, but they must be able to lean on the moral principles and on the interpretations of the rules and solutions that have been agreed on" (P1).*

In contrast, according to the informant, "There will be those who simply operate on the borderlands of these rules," providing cheap products and marketing irresponsibly (P1).

5 Discussion and conclusion

5.1 Key themes for ESG dimensions

The interview material highlights the fact that robust AI auditing criteria for investors' ESG analyses are still more of an emerging issue than a current reality. Nevertheless, the rigorous assessment of AI developer and user organizations was seen as rising in importance. As a starting point for considering AI use in ESG analyses, researchers and practitioners would also benefit from identifying which types of AI issues are relevant to each ESG dimension. Table 5 positions the relevant themes and subthemes from the thematic analysis (see Fig. 1 and Table 4) under the three ESG dimensions.

The environmental and social dimensions of ESG are linked to awareness of AI issues and, following from this, the capacity and procedures used to measure AI impacts in the environmental and social domains. Thus, for these dimensions, companies and investors need guidance on how to identify material issues and the metrics for measuring these issues. Guidelines similar to the ones provided by independent international organizations, such as the GRI and the SASB, could fill this gap. The governance dimension, in turn, relates to how AI development and use processes are governed in organizations. Robust governance operates "upstream," at the process level, to mitigate the potential risks for the environmental and social dimensions. AI ethics principles that are suitably operationalized, as well as balanced and sector-specific approaches, provide guidance for achieving effective AI governance, but operationalization remains challenging. Therefore, as is the case for the environmental and social dimensions, standardized guidelines could also be beneficial for the governance dimension. However, the operationalization of AI governance and the translation from high-level principles to practicable governance mechanisms requires further research.

In the following section, we highlight further implications of our study, followed by limitations and future research directions.

5.2 Implications: ESG analyses as tools for ethics-based AI auditing

This paper began with the question of using ESG analyses as tools for ethics-based AI auditing in partial answer to the translation problem—that is, progressing from abstract AI ethics principles to workable auditing of AI systems and their organizational use. Due to the explorative nature of the research, we primarily raise questions rather than providing ready answers. We highlight three key points from our study considering this central question.

First, there is simultaneously a need for *standardized metrics for AI responsibility and contextual variations* according to industry and culture. Because the ESG literature identifies divergence in ESG ratings (Dorfleitner et al. 2015; Chatterji et al. 2016; Berg et al. 2020), it is likely that AI-related ESG ratings will also differ eventually. The crucial issue, then, is when this divergence is simply a question of differently weighted aspects, which are made transparent, and when the divergence becomes problematic. While legislation provides the common minimum baseline, is there still space for cross-sectoral ESG issues, or are ESG issues above the regulatory baseline predominantly sector-specific? Moreover, there is a need for further understanding on the feasibility and value of quantitative metrics on the ESG dimensions compared to qualitative descriptions of how AI-related guidelines are adhered to. What are the relative merits of guidelines and quantitative metrics in the case of AI? While the environmental and social ESG pillars are likely to be amenable to standardized metrics, the operationalization of the governance pillar requires further research.

As a second implication, turning attention to the *critical bottlenecks* of AI-related ESG evaluations helps to find pathways toward ESG analyses as practical tools for ethics-based AI auditing. Our explorative analysis reveals four critical bottlenecks:

1. Reactive awareness of AI ethics issues. Issues may become visible and material to stakeholders only through controversies, and by then, significant harm may have occurred to individuals and groups (cf. Raji et al. 2020). This reactive approach comes with high risks, as algorithmic systems are used in increasingly critical application areas, such as health care, and failures may cause widespread and irreversible damage.
2. Lack of usable metrics at the level of organizations. While there are metrics for characteristics of algorithmic systems, such as fairness (cf. Benjamins et al. 2019), the derivation of organizational metrics for responsible AI performance is still at an early stage. As noted above, the comparative merits of metrics and guidelines require further research.
3. Lack of tools for tracing impacts of AI systems across supply chains. Are established forms of supply chain management suitable for AI supply chains, or are new approaches needed? How are measurable impacts constructed as proxies for real-world harms (Metcalf et al. 2021)?
4. Lack of guidance on trade-offs and tensions in AI governance. Navigating the inevitable trade-offs and tensions of AI governance (Whittlestone et al. 2019) while exceeding the minimum level of legal compliance requires balancing efforts from organizations. Thus, organizations need guidance not only in individual metrics but also in managing trade-offs between metrics, such as privacy and transparency, wherein it is infeasible or impossible to achieve high levels simultaneously.

The third implication is the need to consider *asymmetrical relations of knowledge and influence* among ESG evaluators and evaluated companies. The core actors in the ESG actor network include investors; investment targets; rating agencies; and, considering the network broadly, regulators. Investment targets—that is, AI developer and user organizations—are more knowledgeable about AI than investors. This raises the question of the required level of AI competence for investors to rigorously assess investment targets, as well as the extent to which this competence can be outsourced to ESG rating agencies. Is it possible that special AI-focused rating agencies will emerge in the future? Regarding regulation, the regulatory landscape is currently changing, at least in Europe. A future EU AI agency may be in the making (Stix forthcoming), and both an AI Act and a Corporate Sustainability Reporting Directive have been proposed (Santoro et al. 2021). Even though the focus here is on ESG evaluations that go beyond legal compliance, the regulators' role as the "second-order oversight mechanism" that defines and assesses acceptable practices in AI auditing should not be underestimated.

5.3 Limitations and future research directions

This paper is based on an exploratory interview study with senior-level Finnish professionals. Therefore, at best, it can provide initial theoretical abstractions (Lee and Baskerville 2003) rather than making attempts at generalizations. The findings present key considerations that may be elaborated upon and critically scrutinized in future research.

Considering future research, the key problem is achieving an overview of the state of play in the field of ESG and AI auditing—that is, complementing our explorative analysis with more extensive cross-sectional research and in-depth case studies. Mapping this field serves both theoretical and practical goals. On the theoretical side, research is needed on AI governance and auditing as aspects of corporate governance (cf. Schneider et al. 2020), as well as on their links to the CSR literature (Maon et al. 2009) and the algorithmic impact assessment literature (Metcalf et al. 2021; Selbst 2021). On the practical side, an overarching finding of the current study was that basic awareness and knowledge of AI-related issues is needed among sustainable investment professionals. The provision of this knowledge through research thus enables investors and asset managers to assess investment targets more effectively, thereby promoting the responsible development and use of AI more broadly.

Moreover, the emerging landscape of AI auditing and ESG actors provides a rich object of study. For example,

Shneiderman (2020) envisions that financial audit firms could develop review strategies for corporate AI projects to guide investors. The networked configuration of actors could be studied, for example, to investigate tendencies of centralization or decentralization in professional AI auditing, thus complementing emerging studies on responsible AI ecosystems (Minkkinen et al. 2021; Stahl 2021).

One crucial question for future research and practice is how AI governance can be structurally integrated into sets of ESG criteria. At least three possible options exist. First, AI governance can be integrated into the social and/or governance pillars of ESG criteria. Second, AI governance aspects can form an independent pillar that is used when evaluating organizations that develop or use AI systems. Third, AI governance can be incorporated into a pillar that deals with technology governance more generally. These three options may coexist in different variations, particularly in the near future when AI governance aspects are still finding their place in ESG evaluations. In addition, AI impacts may find a different home compared to processual AI governance questions.

New practical tools for operationalizing AI ethics are urgently needed. While the importance of AI ethics is broadly accepted, the methods, guidelines, and metrics for ascertaining effective AI governance are in early development. In this situation, investors possess a powerful lever: ESG evaluations that assess whether companies perform adequately according to environmental, social, and governance criteria. However, investors and ESG are not the silver bullet that ensures responsible AI. As an analogy, investors are paying increasing attention to environmental problems such as climate change, but this has occurred relatively late, and these issues are far from solved. Excessive faith in the power of investors alone may thus be problematic. Despite these reservations, AI governance and auditing scholarship and practice should increasingly turn to investors and ESG criteria—as parts of a broader AI governance system—to promote actionable AI auditing and, ultimately, the socially responsible development and use of AI.

Appendix 1: Interview protocol

1. Can you tell us a bit about your background regarding matters related to ESG?
2. Can you tell us a bit about your background regarding matters related to AI or, more specifically, the responsible use of AI?
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 the evaluation of 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., those that 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 on the utilization of 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? For example, do investors check whether a company has its own set of principles, and do companies need to provide evidence of the operationalization of the principles?
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, greater emphasis on compliance with some principles? For example, 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?

Author contributions All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by AN. The first drafts of the sections on ESG investing and the first draft of the findings were written by AN. The first drafts of the remaining sections were written by MM(Matti Minkkinen). MM(Matti Mäntymäki) commented on the manuscript. All authors read and approved the final manuscript.

Funding Open Access funding provided by University of Turku (UTU) including Turku University Central Hospital. The research leading to these results is part of the project Artificial Intelligence Governance and Auditing, supported by Business Finland.

Availability of data and material A more detailed coding sheet may be available upon request from the corresponding author.

Declarations

Conflict of interest The authors have no competing interests to declare that are relevant to the content of this article.

References

Amel-Zadeh A, Serafeim G (2018) Why and how investors use ESG information: evidence from a global survey. Financ Anal J 74:87–103. https://doi.org/10.2469/faj.v74.n3.2

Benjamins R, Barbado A, Sierra D (2019) Responsible AI by design in practice. https://doi.org/10.48550/arXiv.1909.12838

Berg F, Kölbel JF, Rigobon R (2020) Aggregate confusion: the divergence of ESG ratings. Social Science Research Network, Rochester

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

Braun V, Clarke V (2006) Using thematic analysis in psychology. Qual Res Psychol 3:77–101. https://doi.org/10.1191/1478088706qp063oa

Brundage M, Avin S, Wang J et al (2020) Toward trustworthy AI development: mechanisms for supporting verifiable claims. http://arxiv.org/abs/2004.07213 [cs]

Brusseau J (2021) AI human impact: toward a model for ethical investing in AI-intensive companies. J Sustain Finance Invest. https://doi.org/10.1080/20430795.2021.1874212

Butcher J, Beridze I (2019) What is the state of artificial intelligence governance globally? RUSI J 164:88–96. https://doi.org/10.1080/03071847.2019.1694260

Cardoni A, Kiseleva E, Terzani S (2019) Evaluating the intra-industry comparability of sustainability reports: the case of the oil and gas industry. Sustainability 11:1093. https://doi.org/10.3390/su11041093

Cath C (2018) Governing artificial intelligence: ethical, legal and technical opportunities and challenges. Philos Trans R Soc A Math Phys Eng Sci. https://doi.org/10.1098/rsta.2018.0080

Chatterji AK, Durand R, Levine DI, Touboul S (2016) Do ratings of firms converge? Implications for managers, investors and strategy researchers. Strateg Manag J 37:1597–1614. https://doi.org/10.1002/smj.2407

Clarke R (2019) Principles and business processes for responsible AI. Comput Law Secur Rev 35:410–422. https://doi.org/10.1016/j.clsr.2019.04.007

Cort T, Esty D (2020) ESG standards: looming challenges and pathways forward. Organ Environ 33:491–510. https://doi.org/10.1177/1086026620945342

Cowls J, Tsamados A, Taddeo M, Floridi L (2021) A definition, benchmark and database of AI for social good initiatives. Nat Mach Intell 3:111–115. https://doi.org/10.1038/s42256-021-00296-0

Dignum V (2019) Responsible artificial intelligence: how to develop and use AI in a responsible way. Springer International Publishing, Cham

Dignum V (2020) Responsibility and artificial intelligence. In: Dubber MD, Pasquale F, Das S (eds) The oxford handbook of ethics of AI. Oxford University Press, pp 213–231

Dorfleitner G, Halbritter G, Nguyen M (2015) Measuring the level and risk of corporate responsibility—an empirical comparison of different ESG rating approaches. J Asset Manag 16:450–466. https://doi.org/10.1057/jam.2015.31

Du S, Xie C (2021) Paradoxes of artificial intelligence in consumer markets: ethical challenges and opportunities. J Bus Res 129:961–974. https://doi.org/10.1016/j.jbusres.2020.08.024

Eitel-Porter R (2021) Beyond the promise: implementing ethical AI. AI Ethics 1:73–80. https://doi.org/10.1007/s43681-020-00011-6

Escrig-Olmedo E, Fernández-Izquierdo MÁ, Ferrero-Ferrero I et al (2019) Rating the raters: evaluating how ESG rating agencies integrate sustainability principles. Sustainability 11:915. https://doi.org/10.3390/su11030915

European Commission (2021) Proposal for a regulation of the European parliament and of the council laying down harmonised rules on artificial intelligence (artificial intelligence act) and amending certain union legislative ACTS COM/2021/206 final

Floridi L (2018) Soft ethics: its application to the general data protection regulation and its dual advantage. Philos Technol 31:163–167. https://doi.org/10.1007/s13347-018-0315-5

Floridi L, Cowls J, Beltrametti M et al (2018) AI4People—an ethical framework for a good AI society: opportunities, risks, principles, and recommendations. Mind Mach 28:689–707. https://doi.org/10.1007/s11023-018-9482-5

ForHumanity (2021) Taxonomy: AI audit, assurance & assessment. https://forhumanity.center/blog/taxonomy-ai-audit-assurance-assessment/

Freeman RE (2010) Strategic management: a stakeholder approach. Cambridge University Press, Cambridge

Friede G, Busch T, Bassen A (2015) ESG and financial performance: aggregated evidence from more than 2000 empirical studies. J Sustain Finance Invest 5:210–233. https://doi.org/10.1080/20430795.2015.1118917

Gasser U, Almeida VAF (2017) A layered model for AI governance. IEEE Internet Comput 21:58–62. https://doi.org/10.1109/MIC.2017.4180835

GSIA (2018) Global sustainable investment review. http://www.gsi-alliance.org/wp-content/uploads/2019/03/GSIR_Review2018.3.28.pdf

Hagendorff T (2020) The ethics of AI ethics—an evaluation of guidelines. Mind Mach 30:99–120. https://doi.org/10.1007/s11023-020-09517-8

Hahn T, Pinkse J, Preuss L, Figge F (2015) Tensions in corporate sustainability: towards an integrative framework. J Bus Ethics 127:297–316. https://doi.org/10.1007/s10551-014-2047-5

Hill J (2020) Environmental, social, and governance (ESG) investing: a balanced review of theoretical backgrounds and practical implications, 1st edn. Academic Press, San Diego

Hong H, Kacperczyk M (2009) The price of sin: the effects of social norms on markets. J Financ Econ 93:15–36. https://doi.org/10.1016/j.jfineco.2008.09.001

Jobin A, Ienca M, Vayena E (2019) The global landscape of AI ethics guidelines. Nat Mach Intell 1:389–399. https://doi.org/10.1038/s42256-019-0088-2

Kempf A, Osthoff P (2008) SRI Funds: Nomen est Omen. J Bus Financ Acc 35:1276–1294. https://doi.org/10.1111/j.1468-5957.2008.02107.x

Koshiyama A, Kazim E, Treleaven P et al (2021) Towards algorithm auditing: a survey on managing legal, ethical and technological risks of AI, ML and Associated Algorithms. Social Science Research Network, Rochester

KPMG International (2020) The time has come: the KPMG survey of sustainability reporting 2020. https://assets.kpmg/content/dam/kpmg/xx/pdf/2020/11/the-time-has-come.pdf

Kroll JA (2018) The fallacy of inscrutability. Philos Trans R Soc A Math Phys Eng Sci 376:20180084. https://doi.org/10.1098/rsta.2018.0084

Lee AS, Baskerville RL (2003) Generalizing generalizability in information systems research. Inf Syst Res 14:221–243. https://doi.org/10.1287/isre.14.3.221.16560

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

Maon F, Lindgreen A, Swaen V (2009) Designing and implementing corporate social responsibility: an integrative framework grounded in theory and practice. J Bus Ethics 87:71–89. https://doi.org/10.1007/s10551-008-9804-2

Metcalf J, Moss E, Watkins EA et al (2021) Algorithmic impact assessments and accountability: the co-construction of impacts. In: Proceedings of the 2021 ACM conference on fairness, accountability, and transparency. Association for computing machinery, New York, pp 735–746

Minkkinen M, Zimmer MP, Mäntymäki M (2021) Towards ecosystems for responsible AI: expectations on sociotechnical systems, agendas, and networks in EU documents. In: Dennehy D, Griva A, Pouloudi N et al (eds) Responsible AI and analytics for an ethical and inclusive digitized society. Springer International Publishing, Cham, pp 220–232

Mökander J, Floridi L (2021) Ethics-based auditing to develop trustworthy AI. Mind Mach 31:323–327. https://doi.org/10.1007/s11023-021-09557-8

Mökander J, Morley J, Taddeo M, Floridi L (2021) Ethics-based auditing of automated decision-making systems: nature, scope, and limitations. Sci Eng Ethics 27:44. https://doi.org/10.1007/s11948-021-00319-4

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. Sci Eng Ethics 26:2141–2168. https://doi.org/10.1007/s11948-019-00165-5

MSCI (2020) MSCI ESG rating methodology: executive summary. https://www.msci.com/documents/1296102/4769829/MSCI+ESG+Ratings+Methodology++Exec+Summary+Dec+2020.pdf/15e36bedbba2-1038-6fa02cf52a0c04d6?t=1608110671584

Palinkas LA, Horwitz SM, Green CA et al (2015) Purposeful sampling for qualitative data collection and analysis in mixed method implementation research. Adm Policy Ment Health 42:533–544. https://doi.org/10.1007/s10488-013-0528-y

PRI Association (2019) What is responsible investment? In: PRI. https://www.unpri.org/an-introduction-to-responsible-investment/what-is-responsible-investment/4780.article. Accessed 29 Oct 2021

Rahwan I (2018) Society-in-the-loop: programming the algorithmic social contract. Ethics Inf Technol 20:5–14. https://doi.org/10.1007/s10676-017-9430-8

Raji ID, Smart A, White RN et al (2020) Closing the AI accountability gap: defining an end-to-end framework for internal algorithmic auditing. In: Proceedings of the 2020 conference on fairness, accountability, and transparency. Association for computing machinery, New York, pp 33–44

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. Harvard Business School. https://hbswk.hbs.edu/item/pathways-to-materiality-how-sustainability-issues-become-financially-material-to-corporations-and-their-investors

Sætra HS (2021) A framework for evaluating and disclosing the ESG related impacts of AI with the SDGs. Sustainability 13:8503. https://doi.org/10.3390/su13158503

Sandberg J, Juravle C, Hedesström TM, Hamilton I (2009) The heterogeneity of socially responsible investment. J Bus Ethics 87:519. https://doi.org/10.1007/s10551-008-9956-0

Sandvig C, Hamilton K, Karahalios K, Langbort C (2014) Auditing algorithms: research methods for detecting discrimination on internet platforms, Seattle

Santoro J, Vaessen M, Chapman M et al (2021) Proposed EU directive on ESG reporting would impact US companies. In: The Harvard law school forum on corporate governance. https://corpgov.law.harvard.edu/2021/06/07/proposed-eu-directive-on-esg-reporting-would-impact-us-companies/. Accessed 16 Nov 2021

Schiff D, Biddle J, Borenstein J, Laas K (2020) What's next for AI ethics, policy, and governance? A global overview. In: Proceedings of the AAAI/ACM conference on AI, ethics, and society. Association for Computing Machinery, New York, pp 153–158

Schneider J, Abraham R, Meske C (2020) AI governance for businesses. http://arxiv.org/abs/2011.10672 [cs]

Selbst AD (2021) An institutional view of algorithmic impact assessments. Social Science Research Network, Rochester

Selim O (2020) ESG and AI: the beauty and the beast of sustainable investing. In: Bril H, Kell G, Rasche A (eds) Sustainable investing sustainable investing a path to a new horizon. Routledge

Seppälä A, Birkstedt T, Mäntymäki M (2021) From ethical AI principles to governed AI. In: Proceedings of the 42nd international conference on information systems (ICIS2021). Austin

Shneiderman B (2020) Bridging the gap between ethics and practice: guidelines for reliable, safe, and trustworthy human-centered AI systems. ACM Trans Interact Intell Syst 10:26. https://doi.org/10.1145/3419764

Stahl BC (2021) Artificial intelligence for a better future: an ecosystem perspective on the ethics of AI and emerging digital technologies. Springer International Publishing, Cham

Stix C (2022) The ghost of AI governance past, present and future: AI governance in the European Union. In: Bullock J, Hudson V (eds) Oxford University press handbook on AI governance. Oxford University Press, Oxford **(forthcoming)**

Tamimi N, Sebastianelli R (2017) Transparency among S&P 500 companies: an analysis of ESG disclosure scores. Manag Decis 55:1660–1680. https://doi.org/10.1108/MD-01-2017-0018

Trocin C, Mikalef P, Papamitsiou Z, Conboy K (2021) Responsible AI for digital health: a synthesis and a research agenda. Inf Syst Front. https://doi.org/10.1007/s10796-021-10146-4

Twycross A, Shields L (2008) Content analysis. Paediatr Nurs 20:38–38. https://doi.org/10.7748/paed.20.6.38.s27

Vaismoradi M, Turunen H, Bondas T (2013) Content analysis and thematic analysis: implications for conducting a qualitative descriptive study. Nurs Health Sci 15:398–405. https://doi.org/10.1111/nhs.12048

van der Waal JWH, Thijssens T (2020) Corporate involvement in sustainable development goals: exploring the territory. J Clean Prod. https://doi.org/10.1016/j.jclepro.2019.119625

van Duuren E, Plantinga A, Scholtens B (2016) ESG integration and the investment management process: fundamental investing reinvented. J Bus Ethics 138:525–533. https://doi.org/10.1007/s10551-015-2610-8

Whittlestone J, Nyrup R, Alexandrova A, Cave S (2019) The role and limits of principles in AI ethics: towards a focus on tensions. In: Proceedings of the 2019 AAAI/ACM conference on AI, ethics, and society. ACM, Honolulu, pp 195–200

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