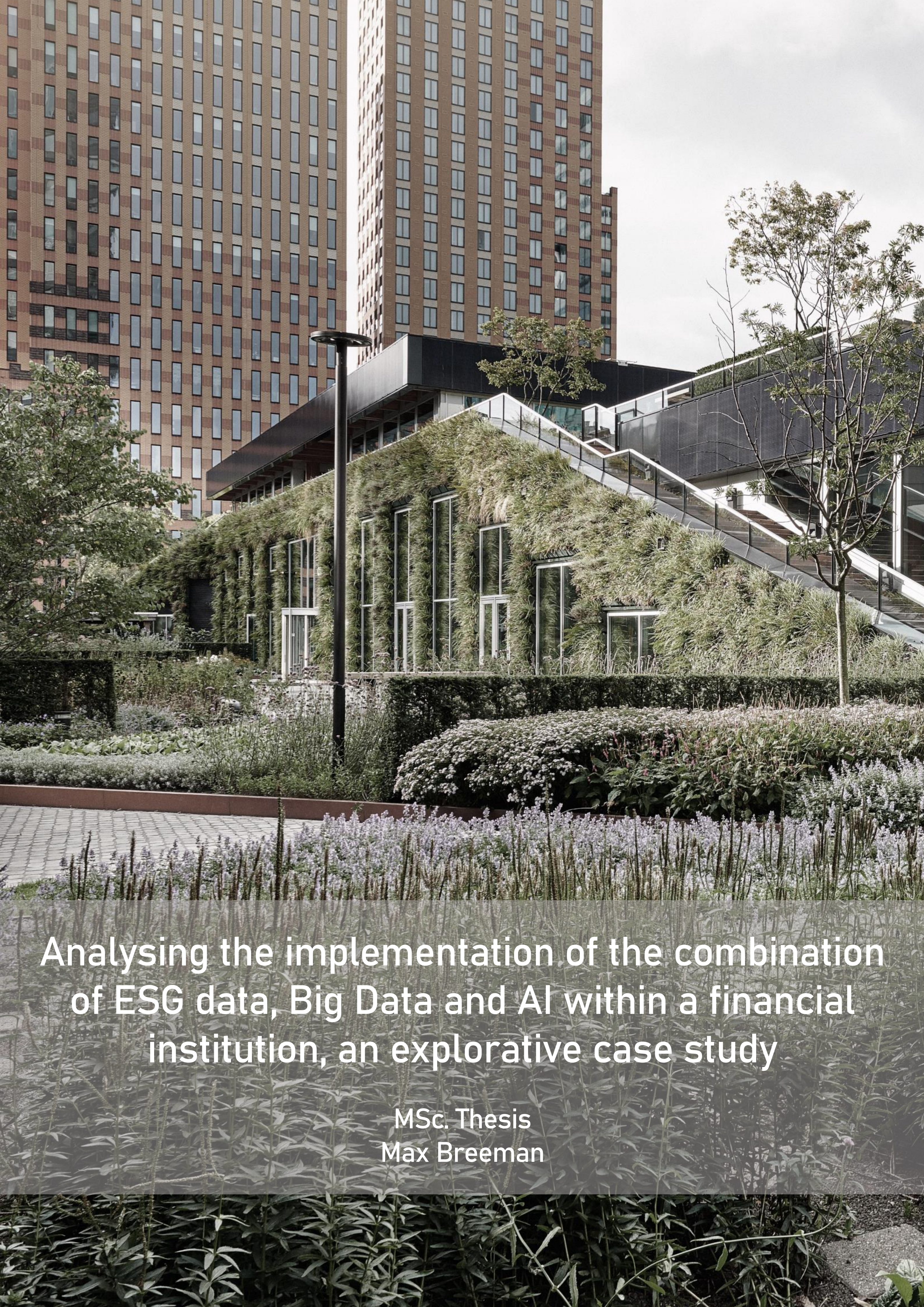

Analysing the implementation of the combination of ESG data, Big Data and AI within a financial institution, an explorative case study

MSc. Thesis

Max Breeman

A special note on the cover image; This is Circl, a circular platform on the Amsterdam Zuidas where stakeholders from society, business and finance come together for a sustainable world.

Analysing the implementation of the combination of ESG data, Big Data and AI within a financial institution, an explorative case study

Master thesis submitted to Delft University of Technology

in partial fulfilment of the requirements for the degree of

MASTER OF SCIENCE

in **Management of Technology**

Faculty of Technology, Policy and Management

by

Maximiliaan Pieter Breeman

Student number: 4748700

To be defended in public on December 17th 2021

Graduation committee

Chairperson	: Dr. J. Rezaei, Section Transport and Logistics
First Supervisor	: Dr. Mr. N. Mouter MSc., Section Transport and Logistics
Second Supervisor	: Dr. Ir. Z. Roosenboom-Kwee, Section Economics of Technology and Innovation

Preface

Finishing my MSc. Management of Technology has been quite the enterprise. As I started at the beginning of 2019, a couple of months after this start I encountered a physical health-related setback that rooted up my life, even to the extent that it delayed the beginning of my graduation. While having recovered, even nearly finishing this thesis, another one occurred. I had to take a break for a couple of months due to a severe concussion. I can say, with confidence, trying to write anything remotely academic with a concussion gives interesting, yet useless results! However, those who persevere and show resilience triumph! With great pleasure, albeit with a slight delay, I present to you my MSc. thesis.

As it was extremely difficult to find an internship during the height of the COVID-19 pandemic, I want to thank the bank, which I am not allowed to call by name due to an NDA, but am allowed to describe some of its characteristics, which is rather silly, for the internship and research opportunity. During my time at the bank, Andre Jakobs provided me with the supervision and insightful discussions needed for my thesis, which was much appreciated!

I could not have finished this thesis without the help and motivation of numerous people. I want to thank my graduation committee for their effort and patience, especially Niek Mouter for his feedback and supervision during the project. I want to thank Tim Koning for his insightful perspectives during the study sessions at TPM. And I want to thank all my friends for the support and jokes along the way.

As having to recover, twice, from an illness during my masters, I want to thank my family for helping me recover during these periods. And I want to convey a special thanks to Carlijn, who I annoyed continuously with banking slang and technical descriptions of obscure facets of my thesis.

Max Breeman

Delft, December 2021

Abstract

Over the last few years, regulations, changes in governance, and societal pressure have led to a push to rethink a firms' approach to sustainability. This push created a need to place sustainability and numerous relevant technologies and approaches at the centre of the firms' decision-making process. Within the financial industry, the combination of novel data technologies such as Big Data and Artificial Intelligence (AI) with the inclusion of sustainability, or so-called "ESG", Environment, Social, and governance data spearpoint this new 'sustainable' frontier. This inclusion of ESG data originates from the field of investing, where it is being applied to maximize risk-adjusted long-term investment returns, also known as 'Alpha'. Thus, using it as an inside-out perspective of assessing investment decisions. However, this combination of data and technology can also be applied to provide an outside-in perspective. Meaning that a firm can use it as a corporate resource to base their management information on, thus for management steering purposes.

Literature shows that the implementation of the combination of ESG data, Big Data and AI, within a financial institution and as a corporate resource is a rather novel subject. Most literature is related to investors implementing such combinations for a better Alpha. However, some authors state that a theoretical foundation regarding ESG data has yet to rise and there is still no well-defined approach to integrating ESG data. Even more, there is a need to expand on the formal and integrated process of implementing Big Data and AI strategies in financial institutions. Even the barriers to adoption are a subject not often addressed within this stream of literature. The lack of clear-cut, to the point, and even just available literature illustrates that this field of study is full of potential development, ready to be explored.

This thesis will address the topic of implementing ESG data, Big Data and AI within a financial institution. This topic came to be through a research opportunity provided by a Dutch financial institution and the aforementioned knowledge gap found in the available literature. The main research question answered within this thesis is set out to explore this combination of sustainability data, also known as ESG data, Big Data and AI. This study aims to explore this subject and provide a starting point to fill this knowledge gap by creating novel theoretical propositions to be tested in future research. The following research question has been devised to address this research problem statement.

What observations can be extracted from assessing the introduction of a Big Data and AI toolset applying ESG data within a procedure?

To answer this research question a single holistic case study has been designed to assess the integration of Big Data and AI, and ESG data consisting of the design and introduction of a procedure. This procedure takes place in the setting of a Dutch financial institution, which provided the research opportunity. Initial propositions have been defined through internal discussion and a literature study, providing scope and direction regarding the case study. These initial propositions are refined through the application of the case study. The case study itself consisted of three parts where data was gathered; assessing the current procedure, designing the procedure, and assessing this procedure. Within these parts, data was gathered through documents and memos, participant observation, and semi-structured interviews. This data, through discussion, lay the foundation for novel theoretical propositions. Within this case study, a total of 43 themes have been observed. These themes have been categorised into six categories. These categories are Current situation, (Big) Data, ESG, Learning & adoption, New situation, and Perception & social. These observations have been assessed, discussed, and novel theoretical propositions have been constructed. These novel theoretical propositions illustrate key observations made during the case study and are used to answer the main research question. For each proposition, future research directions are given. These propositions, thus, the answer to the main research question are:

- The perception within a firm of using Big Data and AI within a process could affect the learning rate and the learning approach taken by the user. This affects the acceptance of the technology. Thus, the perception could affect the adoption rate of Big Data and AI within a firm.

This proposition illustrates the effect the perception of the combination of these technologies have on the learning rate and learning approach. It, in the end, could affect the acceptance of the combination and thus could affect the adoption of such technologies. The proposed future research could illustrate this causal effect and could, perhaps, address other novel relevant factors.

- If Big Data and AI are used within a process, people tend to be convinced by Big Data and AI used within the process, thus Big Data and AI can be used to convince people of the validity of the results of the process.

This proposition illustrates the convincing power of Big Data and AI as observed within the case study. It focuses on the affirmation of the validity of the results of the process, with the focus on convincing people of its validity through the application of these technologies. By creating this proposition and analysing its related observations, a distinct bias became prevalent. This bias is defined as a 'prophet bias'. The following analogy regarding this prophet bias can be made with respect to AI. The output of AI can be seen as the AI telling prophecies without being able to fully comprehend and address the approach to these "prophecies". Causing people to believe the prophet, as it is outside of one's comprehension.

- If conferred management information is substantiated by an information process using Big Data and AI, then people do not have the tendency to acknowledge the inherent biases in such processes.

This proposition shows the general acceptance regarding the usage of these technologies and their relationship to bias. This proposition is continuous on the aforementioned one. It illustrates the convincing power of Big Data and AI. It is interesting to see that this proposition illustrates two things, one of ignoring bias due to the methodology, and the second of accepting and being convinced by the methodology, turning a blind eye to the negative aspects.

- If Big Data and AI are used within a process, data quality and source are perceived as of less importance.

The aforementioned proposition is one of focus. Within the case study, it was observed that within the assessment of a process that uses Big Data and AI, the focus is mainly on the method and the sources are deemed of less importance.

- There could be causation between one's knowledge of Big Data and AI, and the perception of bias when assessing a process that uses Big Data and AI.

Within the case study, it is observed that when Big Data and AI are introduced within a process, one should acknowledge the inherent bias to the process. When these biases are acknowledged, it could positively affect the acceptance of the process, as one perceives more validity regarding the process. Personal knowledge has been observed as an influencing factor regarding this factor of addressing bias, thus, assessing a process using Big Data and AI.

- ESG data is context-dependent, illustrating that a structured or unstructured approach to ESG data depends on the application of ESG data.

This proposition aims to further the field of ESG data. It provides a starting ground for future research to gain insights regarding the currently lacking theoretical foundation of ESG data. The context-dependent aspect of this novel theoretical proposition was already addressed within the literature. The link of this dependency to a structured or unstructured data approach could be considered novel. This proposition could provide additional insights into a potential fundamental framework regarding ESG data.

Further recommendations are made regarding the results found within the case study. The recommendations are discussed in the perspective of where data is lacking, hunches have been found but not materialized, and current technological trends not fully ready to be explored yet. One interesting direction not addressed in the results is the notion of XAI, which is a fully explainable AI. This recent development is still in its early stages, and it would be recommended to readdress this subject in the near future with a focus on the adoption and bias as it could affect the core notion of the prophet bias. Another direction is that of company culture. Company culture could play a role within the usage of Big Data, AI, and ESG data within a firm. Only a hint regarding this causation was observed, thus, it is recommended to address this in future research. Regarding ESG data, upcoming (non-mandatory) regulations regarding the reporting and classification of ESG factors, could affect the way ESG related information is reported. Within an estimate of three years, this subject should be reassessed. As it gives companies a decent time frame to process, increase the quality of, and publish ESG data as they comply with these new regulations.

List of abbreviations and clarifications

AI	Artificial Intelligence, is a broad-ranging branch of computer science concerned with IT tools capable of performing tasks that typically require human intelligence.
Alternative data	"Alternative data refers to data from factors that are not conventionally used for investment decision making, yet these unconventional factors have a special role in corporate profits and sustainability" (In, Rook, & Monk, 2019).
Bias	Within the context of this thesis, bias illustrates the need to keep to the way things have always been, inherent inclination or prejudice for or against a certain technology, idea, or implementation. Often not based on fair judgement or relevant knowledge.
Big Data	Large sets of data are almost impossible to manage, process, and analyse through conventional business intelligence tools. Adhering to the three V's: Volume, that a lot of data is collected through a variety of sources, Velocity, that data streams are included in the dataset near real-time, and Variety, that the data comes in different types and formats, structured and unstructured.
Buzzword	A word or phrase from a particular subject area, often an item of jargon, has become fashionable by being used a lot at a particular time or in a particular context.
Corporate resource	These are assets a company has available, which are controlled by the company, and can be used to achieve its goals.
ESG	Environment, Social, and Governance, an approach to classify and factor sustainability.
ESG data	ESG data is unstructured, multi-faceted, context-dependent, ESG related and based in alternative data.
Expert opinion	Advice, data, or knowledge given by experts, based on their previous experience, knowledge, and discussion with relevant stakeholders.
Fintech	Fintech, an abbreviation of financial technology. Novel technological solutions seek to improve and automate financial services by utilizing specialized software and algorithms.

Perception	How something is regarded, understood, or interpreted. Thus, how something is experienced.
Procedure	A procedure as described within this thesis is the process of using ESG data, Big Data and AI
SaaS	Software as a Service, software delivered through an online platform. The customer does not have to purchase the software but licenses it. Also known as ‘software on demand’.
SASB	Sustainability Accounting Standards Board
Sustainability	Development that meets the needs of the present without compromising the ability of future generations to meet their own needs. Within this thesis, sustainability is applied and categorized through ESG.
Theme	A theme represents a concept or abstraction of potential interests (Yin, 2018). It is used within this thesis to code the results gained from documentation such as company documents, memos, and interviews.
Unstructured data	Unstructured data is a collection of data entailing anything, such as media, imaging, sensor data, or text data. Unstructured means that its datasets are not structured in a database format. It has an internal structure but is not predefined through data models.
XAI	Explainable Artificial Intelligence is AI where the results of the decision and solution process can be understood by humans. It contrasts with the concept of the "black box" in machine learning where even its designers cannot explain why an AI arrived at a specific decision.

List of figures

List of tables

Contents

Chapter 1. Introduction

The new European regulation such as the introduction of the EU Taxonomy, the increase of shareholder activism, and the COVID-19 pandemic have provided a push to rethink the firms' approach to sustainability. This created a push to place sustainability and relevant technologies at the centre of the firms' decision-making process (Macpherson, Gasperini, & Bosco, 2021). This centre proves to be the new frontier on which firms can innovate. The combination of novel data technologies such as Big Data and Artificial Intelligence (AI) spearpoint this frontier.

The financial industry is one of the main industries ready to embrace these novel data technologies, intending to enhance their financial performance. Numerous fin-tech firms have emerged integrating these data technologies, Big Data and AI, in combination with alternative sources of data. Alternative data is data such as satellite imagery, digital sensors, mobile phone data, data from media outlets, or even non-conventional financial data. However, the focus of this alternative regarding sustainability is focussed under so-called ESG data. Sustainability data is classified under E, S, or G, Environment, Social, or Governance. This increase in possibilities for investors of using alternative data sources drove their interest in it, as it is believed that the usage of these ESG criteria for investment decisions minimizes risk and maximizes 'long-term risk-adjusted investment returns' (In, Rook, & Monk, 2019). This rise of ESG data availability and processing capabilities bring down costs to investors due to the ever-expanding usage and scale, as the effects of economies of scale jump into action.

As investors aimed to let their investment decisions be steered upon ESG data, firms began to integrate ESG data to be steered upon. It was, in combination with novel data technologies such as Big Data and AI, being used as a corporate resource. As Antoncic (2020) states, the integration of this aforementioned combination within a firms' business model or even decision-making process proves to be generating a significant competitive advantage. However, there is no well-defined approach to integrate ESG data and novel technological solutions based on Big Data and AI within a firm (Kotsantonis & Serafeim, 2019). Even more, trying to integrate sustainability into business practices such as reporting on and integrating ESG criteria has only led to more "aggregated confusion" among companies alike (Macpherson, Gasperini, & Bosco, 2021). What, in the end, can be concluded is that the social and technological aspects of such an integration of sustainability and novel data technologies as a corporate resource are yet to be explored properly.

The topic of implementing ESG data, Big Data and AI came to be due to a research opportunity provided by a Dutch financial institution. This thesis is set out to explore this combination of sustainability data also known as ESG data, Big Data and AI. To provide an initial overview of potential novel theoretical propositions on which future research could be based.

1.1. Reading guide

This thesis consists of six chapters, the introduction, methodology, case study characteristics, results, discussion, and conclusion and recommendations.

Within the first chapter, an introductory background will be given regarding sustainability, ESG, ESG data, Big Data and AI. This background is required as many aspects as discussed in this thesis have different interpretations and facets. Furthermore, the results of the literature study will be addressed in the literature background, reflecting the available literature regarding the topic. Furthermore, within this sub-chapter, the knowledge gap is defined, as it provides the basis for the problem statement. On the problem statement, the main research question and its sub-questions are devised.

The second chapter will portray the methodology used to answer the aforementioned research question. This thesis will make use of a case study approach to answer the research question. Within this chapter, the research framework will be addressed. The approach will make use of initial propositions to define structure to the case study, these will be addressed after the framework itself. Then, the case study research design is stated, giving information regarding the acquisition of data from the case study itself. Then, a deep dive will be given regarding the data collection procedures. As data is collected, it has to be codified and analysed. This method will be discussed in the following sub-chapter. At the end of the methodology chapter, the validity of this case study will be addressed.

The third chapter illustrates the case study characteristics. These characteristics are unique to this case, as it was based on a research opportunity provided by the firm. Here, the link will be made between the research topic, the topics as illustrated in literature, and as used within the case. It is structured around the approach as stated within the methodology chapter.

The fourth chapter will portray the results gained from the study. A summary will be given regarding the results as portrayed in Appendix A – Results and Appendix B – Clarification results. After the summary, some highlights will be given regarding the results.

The fifth chapter contains a discussion of the results. First, the findings will be discussed. Here, data will be analysed and compared to literature to create a discussion regarding the potential theoretical propositions. After which, additional findings and limitations will be discussed. Then, reflections will be given regarding this study. The discussion chapter will end with a reflection on the validity of this study.

The sixth and last chapter illustrates the conclusions and recommendations. Here, the main research question and its sub-questions will be answered. It furthermore gives recommendations regarding the novel theoretical propositions defined as discussed and proposes future research directions based on the discussion of the results. At the end of this chapter, the contributions of this study are discussed.

1.2. Background

The background is split into four main topics concerning this thesis, Sustainability and ESG, ESG Data, Big Data, and AI. For each subject, an introductory background will be given.

1.2.1. Sustainability and ESG

Sustainability in every sense of the term is commonly perceived as extremely broad. To quantify this broadness, the notion of ESG is introduced. The term ESG was first coined in 2005 by a study called Who Cares Wins (Kell, 2018). It was introduced as an approach for including non-financial factors into the investment process and decision-making. ESG covers facets not included in conventional financial analysis while maintaining financial relevance for the investor or manager. It includes facets regarding the companies response to sustainability practices such as biodiversity, carbon emissions, or supply chain policies. The roots of ESG can be traced down to the notion of Socially Responsible Investing, where an investor does not want to invest its money into a company that engages in environmentally or socially irresponsible practices (Townsend, 2020). Currently, the notion of ESG is more widely adopted and used in investing circles, where it is being used as an investment criterion on which an asset can be assessed. As ESG is adopted within investing circles, it becomes a guiding managerial principle on which guidance to a firms’ practices and governance can be given.

As mentioned within the introduction, ESG stands for Environment, Social, and Governance, which are non-financial categories used to assess sustainability risks and opportunities (Laermann, 2016). Each category has its characteristics revolving around all aspects related to sustainability.

Environment delineates how a firm uses its environmental resources. Examples of this are metrics related to produced waste, pollution, carbon footprint, and the firms’ impact on biodiversity. It depicts the way a firm evaluates its environmental risks.

Social portrays the firms' relationships. It illustrates the interaction of the firm with, for example, its stakeholders, the engagement of suppliers, the working conditions within the firm, and how it treats its employees and customers.

Governance relates to the internal governance of the firm. Such as its anti-corruption policies, risk management, or its tax transparency. Furthermore, it entails aspects like the diversity of the board and its compensation.

1.2.2. ESG Data

The notion of ESG data is rooted in the integration of ESG performance of companies in investment decisions. Integrating sustainability topics and data in the investment decision process. The United Nations-backed Principles for Responsible Investment (PRI) defines ESG data as akin to alternative data (UNPRI, 2021). As mentioned in the introduction, alternative data is data such as satellite imagery, digital sensors, mobile phone data, data from media outlets, or even non-conventional financial data. Thus, data that, in the end, does not show up in a company's financial report.

The main difference between alternative data and ESG data is that ESG data emerges from sustainability initiatives. There are several definitions regarding ESG data, within this thesis the following definition will be adhered to, as defined by In, Rook, & Monk (2019). ESG data is unstructured, multi-faceted, context-dependent, ESG related and based in alternative data. A deep-dive on ESG data can be found in the sub-chapter ESG data, of the Literature background chapter.

1.2.3. Big Data

Big Data is a broad term, indicating large sets of data almost impossible to manage, process, and analyse through conventional business intelligence tools. As this term is being used throughout this thesis, the background will be given regarding this notion.

The notion of Big Data dates back to the early 1990s. as the name already implies, describes a large volume of data. The amount and collection of it is an important aspect of this, but the key benefit of using Big Data is that it can be analysed and insights can be gained impossible to attain through traditional methods. The mainstream definition of Big Data is seen through the three V's: Volume, that a lot of data is collected through a variety of sources, Velocity, that data streams are included in the dataset near real-time, and Variety, that the data comes in different types and formats, structured and unstructured (SAS, 2021).

As stated by Elish and Boyd (2018), the term of big data was a neologism. As the collection of data and implementation of this data to be used for statistics for population measurements date back centuries (Hacking, 1982). There have been numerous definitions and supposed implementations of Big Data addressed within literature over the years. An example given by Press (2014) is that after listing 10 different definitions, two definitions were leading; "the belief that the more data you have the more insights and answers will rise automatically from the pool of ones and zeros" and "a new attitude by businesses, non-profits, government agencies, and individuals that combining data from multiple sources could lead to better decisions.". Both illustrate the near-mythical possibilities of what big data could achieve.

During the same era, companies and enterprises began to focus on big data as a novel business opportunity (Manyika, et al., 2011). While consulting firms emerged to support companies in their digital data transformation, and the need for data science solutions grew during that time (Lohr, 2015), less chivalrous companies preyed on the need of unknowing firms by selling 'Big Data solutions that were little more than vaporware' (Elish & Boyd, 2018). This led to the question of the value and purpose of such analytics as Big Data. It gave rise to critics regarding Big Data. An example of it is the report published regarding the "algorithmic systems, opportunity, and civil rights" (Barocas, Rosenblat, Boyd, & Gangadharan, 2014), where it painted a "concerning portrait about the potential of data discrimination" (Elish & Boyd, 2018). The hype and mystics regarding Big Data flowed towards the novel forefront of technology and melted together in a combination often used, 'Big Data and AI' (Elish & Boyd, 2018).

1.2.4. AI

The previous sub-chapter defined the characterization and history of Big Data, here, a background will be given regarding the notion of AI. This as AI is currently a term slung around by numerous companies trying to hype the technology, and many researchers are trying to look beyond this hype to see what it actually entails (Toh, Dondelinger, & Wang, 2019).

Artificial Intelligence, AI, is a broad term for a process using logic and analytical methods to perform tasks. Meaning that, in essence, it is a computerized approach to simulate human intelligence. It can be used to let algorithms make decisions, carry out tasks, and learn from these processes. Like Big Data, it is decades old. The original concept of AI dates from the '50s but was crystallized by McCorduck (2004). It has always been a technology which capabilities were elusive for the greater public and illustrated as something having 'greater-than-human capabilities (Dreyfus, 1972).

This perception held on, while research continued, the perception and popularity faded during the mid-1990s. The perception even degraded as far as the future capabilities of AI being "brittle" and working only in limited context with lacking results (Suchman, 2007). However, interest and research in the topic picked up in the rebranding of AI under machine learning and deep learning (Elish & Hwang, 2016). The combination of these technologies with large datasets, Big Data, the combination of AI and Big Data reaches new heights. Factors related to this growth are the availability of vast datasets, the increase in computer power, and the industry commitment for the adoption of this combination (Elish & Boyd, 2018).

As seen in this and the previous sub-chapter, there have been different epistemologies regarding Big Data and AI over the time they were developed and the research interests in these topics grew, cooled down, and grew again. So did the perception of what it is able of did. This perception cumulates in the following quote regarding AI; "It does not matter if a system thinks like a human – as long as it appears to be as knowledgeable as a human" (Elish & Boyd, 2018).

Currently, AI is a broad term within many subsets, each subset based on the same connotation of AI while performing different tasks such as image recognition, speech recognition, predictive modelling, machine learning, and Natural Language Programming.

The following subset will be adhered to when referring to AI throughout this thesis. This definition of AI is based in the subset of Natural Language Processing (NLP). NLP is a way for an algorithm to 'talk' to humans. It gives the possibility to interpret, manipulate, and 'understand' human language. NLP is widely considered a subset of machine learning, the subset of AI where algorithms learn and improve through usage.

NLP, as to be used throughout this thesis, adheres to the following scope. NLP is a "*theoretically motivated range of computational techniques for analysing and representing naturally occurring texts at one or more levels of linguistic analysis for the purpose of achieving human-like language processing for a range of tasks or applications.*" (Liddy, 2001) This definition of NLP is widely adopted and can be further dissected to clarify the terms mentioned.

The first one is 'range of computational techniques', which is a broad statement regarding the techniques used. Within the case, NLP is an Artificial Intelligence technique used to gain relevant information from unstructured data. The quote 'naturally occurring texts' is rooted in the fact that it analyses texts, human language, thus unstructured for the purpose of analysis (Liddy, 2001). Within this case, it provides an approach to interpret, understand, and manipulate the English language. An easy understanding regarding this kind of technology is that it makes it possible for humans to talk to machines. This is also seen in the part of 'human-like language processing'. As Liddy (2001) also states, this is the part that separates it from other analysis tools and can be classified as artificial intelligence. It, in the end, strives for a human-like understanding of language.

1.3. Literature background

The following sub-chapter aims to provide a reflection on the available literature. It provides background regarding the integration of the combination of ESG data, Big Data and AI, as an operational procedure within a firm and what this combination entails. This combination will be split up into the following topics to be addressed, ESG data, novel data technologies, Big Data and AI operationalized within a firm and the perception of Big Data and AI.

This literature background furthermore aims to illustrate where a knowledge gap is in respect of the combination of ESG data, Big Data and AI. This knowledge gap will be summarized in the concluding remarks. Furthermore, as the knowledge gap could be perceived as extremely broad, key literature is summarized within a table comprising of source, summary, and key research implications at the end of this sub-chapter. The literature research methodology can be found in Appendix D – Literature research methodology. At the end of this sub-chapter, concluding remarks will be given regarding the findings, on which the research problems statement is based.

1.3.1. ESG data

The first part to understand regarding the intended research topic is ESG data. ESG data is data with its basis in the aforementioned ESG themes. As mentioned in the previous sub-chapter, this notion is rooted in the classification of alternative data. Alternative data is data beyond the typical filings a firm does, such as company reporting, earnings calls, or other fundamental datasets. The alternative data sources concise of datasets such as web-scraped data like social media output, news articles, legislative propositions, but also things such as satellite imagery, geolocations from cell phones, consumer sentiment data, or even credit card transactions (In, Rook, & Monk, 2019). Within the context of the financial sector, this kind of data is providing a non-conventional, novel source of information for an investor to give them a competitive edge (Antoncic, 2020).

In, Rook & Monk (2019) state concerning the adoption of ESG data that "the PRI [Principles for Responsible Investment] has suggested that the key to successful and sustainable implementation of ESG is ultimately a function of whether ESG performance can be linked with business, finance, and investment performance". Stating that, like Antoncic (2020) that there is a connection between ESG data and a potential competitive edge. However, this link is still not fully identified.

ESG data can be compared to alternative data, it entails data focussing on ESG aspects. In, Rook, Monk & Rajagopal (2019) even make a case of ESG data being alternative data, meaning that much of alternative data is ESG data and that much of ESG data is likewise alternative data. While nearly similar, the expected benefits for a firm are also more a matter of perspective, stating that "where ESG data has been seen as a potential drag on returns, alternative data is often perceived as a way of driving higher risk-adjusted performance" (In, Rook, Monk, & Rajagopal, 2019). Showing that the way this kind of data is perceived and how it is applied is just as important as what the actual data entails. However, no further details are given in this direction. An interesting look at ESG data as provided by Macpherson, Gasperini, & Bosco (2021) is that ESG could have inherent information bias in itself. This, however, has not been addressed in other literature yet.

One of the main subjects to be addressed is the integration of ESG data. New data technologies have improved the accessibility, availability, and transparency of ESG related data, but an agreed theoretical framework to evaluate data and its quality is still lacking (In, Rook, & Monk, 2019). Even more, the sheer variety and inconsistency of the measured data and how companies report relevant ESG data could lead to data gaps, as approaches differ on the variety of data used (Kotsantonis & Serafeim, 2019). Thus, one can state that there are no well-defined approaches to integrating ESG data. Macpherson, Gasperinin, & Bosco (2021) even recommend using multiple ESG data providers in the context of decision making within a firm. This shows that there is still uncertainty on the application of ESG data within decision-making in the context of a firm.

Another aspect of ESG data is that there is currently no theoretical grounding for discussing it, however, dimensions have been defined to comprise a theory-grounded starting point for determining ESG data and its properties (In, Rook, & Monk, 2019; Monk, Prins, & Rook, 2019). Assessing the value proposition ESG data brings with it regarding how it will be applied. Emphasizing that these are case-specific and driven by variables an investor emphasises within its investing process. These dimensions regarding ESG data and its quality are: are reliability, granularity, freshness, comprehensiveness, actionability, and scarcity. As this approach is still case-specific, a good theoretical foundation regarding ESG data is lacking.

1.3.2. Novel data technologies

Regarding novel data technologies, Antoncic (2020) states that the integration of Big Data and AI technologies could be used to integrate sustainability data in a firms' business model and decision-making for a possible competitive advantage. However, he also states that the board of a firm needs to be 'sufficiently fluent' in the latest sustainability technology to adopt such a way of operating, giving an arbitrary approach to the adoption of such technologies. Antoncic does however only state that it 'could' be done, giving no further tangible proof or application to his claims. The introduction of such datasets and related technologies can be seen as the next step in his process, as he is already convinced that the sustainability data is beneficial. This is further substantiated by Macpherson, Gasperini, & Bosco (2021), also stating that, if properly developed and integrated, these new technologies and the alternative data sets could provide firms and decision-makers a significant competitive advantage. What was noticed missing within the context of both papers is the implementation side and related results of these technologies, as both papers state the possibility as a fact that it could be used in such a way.

Furthermore, the scope of the research interest, the combination of ESG data with Big Data and AI within the financial sector, usually do not align with the research available. For example, in the case of Macpherson, Gasperini, & Bosco (2021), the authors take the viewpoint of the investor when looking at trends, risks, behaviours, sentiment and other criteria. Taking the viewpoint of the investor is one of predominance within the literature stream of applying big data, AI, and ESG data, referring back to the outside-in view. What looks like to be lacking is an example of implementing such data and technology into a process within a firm, for example for conveying management information. Meaning that 'outside' ESG data will be used to steer upon, taking an inside-out view or perspective. Whereas the look of the investor into the ESG data of a company can be seen as an outside-in view.

Hughes, Urban, & Wójcik (2021) wrote that alternative ESG ratings aim to give the same perspective. Their approach is to rate a firm on their ESG performance through similar outside-in means. They concluded that ESG ratings differ regarding the inside-out view or outside-in view, meaning that it differs regarding the dataset used in these ratings. Implying that the research perspective of the investor might not fit the approach taken within the research direction taken within this thesis. Which is one of inside-out, one of conveying information through the combination of ESG data, Big Data, and AI.

These authors are not the only ones researching this direction. Bala, Bartel, Hawley, & Lee (2015) aimed to include ESG data and AI through NLP to assess firms on their ESG performance. However, their analysis is based on TruValueLabs data, gathered around the implementation of SASB, an ESG rating approach by the Sustainability Accounting Standards Board (SASB). Meaning that their approach was structured around a set amount of ESG topics gathered through the annual consultation of stakeholders by SASB, giving another outside-in view.

Even more, Antoncic (2020) illustrates the use of Big Data, AI, NLP, and machine learning to assess data for sustainable investing, which nearly touches upon the research direction as adhered to within this thesis. However, the viewpoint was taken of the investor, and nothing much was said about anything other than this viewpoint within their paper. Showing that in the end it's the inside-out view all over again and does not faithfully contribute to the literature of introducing such technologies within a firm and what it all entails.

It is not that the implementation of AI and Big Data is novel within the financial sector, financial institutions are deeply involved in the calculation of Big Data events (Hasan, Popp, & Oláh, 2020). Hasan, Popp, & Oláh (2020) even state that "…, there is a need to expand the formal and integrated process of implementing big data strategies in financial institutions.", showing the research setting, but not the right context when compared to the research direction of this thesis.

1.3.3. Big Data and AI operationalized within a firm

Another tangential is the difficulty of integrating Big Data and AI into a firms' processes. As stated by In, Rook, Monk & Rajagopal (2019), "misunderstanding and distrust of the newer methods for gathering and analysing data have been cited as major impediments to businesses taking fuller advantage of the capabilities of data science". This is regarding the adoption of these data toolsets, as they could be used to steer the transformation towards sounder ESG practices. Even more, "The financial industry doesn't strongly feel the need to innovate" (Corea, 2019), showing even more that there is distrust within the industry which could hamper the adoption of such technologies. But that raises the question of what else can be said about the inclusion of such systems, as In, Rook, Monk & Rajagopal focus their paper mostly on the scope of investment decisions. Does their scepticism hold up within the proposed research direction?

Pascheck, Luminosu, & Negrut (2020) state that in their research regarding the implementation of AI in decision making, data available is one of the most important aspects. Big Data in combination with AI is, however, just a small aspect of their study. Their study shows that "…, the thesis that data and its nature are the important prerequisites for AI and decision-making in the business environment could be confirmed.", showing that the application of AI into decision-making is beneficial when based on a solid dataset and strategy. It furthermore shows that incorporating these technologies provide clear advantages. However, it does not take it a step further, it laid the foundation that data and its nature are important for AI and decision-making in a firm context. But their approach is too broad and focuses on a broad notion of data, for example, the connection between a small firm-related dataset and the relevance for the firm is a near given. The approach of then analysing this dataset through AI has high relevance for a firm the dataset comes from. The focus was furthermore on the application of AI on a broad aggregation of different types of datasets in a decision support system. The research direction and intention taken within this thesis, however, does not fully match the approach taken in this paper, as it is more informative and focuses on a particular (ESG) Big Dataset. Big Data & ESG data have a different approach, e.g. looking at the volume, velocity, variety, and versatility of the data and ESG data as an untypical data source. Thus, the question is raised, when looking at alternative datasets, do these propositions as defined by Pascheck, Luminos, & Negrut (2020) of the nature, importance, and relevancy hold up too?

Good to know is that within the literature, the assumption is made that most applications for AI and alternative data are in capital markets for financial services (Corea, 2019). Showing that within the financial service industry other applications are glossed over or downplayed, making this an interesting research direction.

1.3.4. The perception of Big Data and AI

Papers like 'Situating methods in the magic of Big Data and AI' (Elish & Boyd, 2018), gave a socio-technical perspective regarding Big Data and AI. One concept proposed regarding any algorithmic system is "once deployed for public use, recommendations, predictions, and classifications produced by technical systems are often accepted as uncontroversial until a result challenges socially constructed assumptions.". However, within their paper, the lens was used of the choices that inform these systems and the challenges in interpreting their results reveal the limitations when we construct technological tools to solve inherently social problems. Furthermore, Elish and Boyd (2018) take the perspective of AI and Big Data within a socioeconomic system, as a machine learning and data science approach to create methodological tools to understand cultural practices more generally. Their perspective is thus one of aiming to use these technologies as a mode to know the world, while the perspective of the 'man on the ground' is deemed irrelevant. Their approach illustrates the magic Big Data and AI can create, but substantiated research regarding the practice and rhetoric of Big Data and AI is not there yet (Elish & Boyd, 2018).

Regarding the adoption and perception of Big Data, scholars argue that the actors embracing these technologies should fully understand the potential (Shindelar, 2014). This again is a more general remark regarding adoption and their paper does not go deeper into these aspects. This is a common occurrence, as seen in the paper "Big Data and AI – A transformational shift for government: So, what next for research?" (Pencheva, Esteve, & Mikhaylov, 2020), general remarks have been addressed concerning outlining key elements to consider, ethics, technology, process, organisational and institutional change, and analytics. However, the social aspects were near non-existent within their paper. Not going any further than stating that "there are wide-ranging, often dystopian speculations around the ethical use of Big Data" (Pencheva, Esteve, & Mikhaylov, 2020). Referring to the implementation of Big Data in government. Regarding adoption, stating that "a few barriers at the individual level are noted in the literature, but relatively little attention is paid to these" (Pencheva, Esteve, & Mikhaylov, 2020). Even going as far as stating that the mindset of the individual could be of significant importance. But further details are not given.

1.3.5. Concluding remarks

Thus, to summarize relevant literature, most papers address certain aspects of the concept of implementing ESG data through a Big Data and AI toolset as a corporate resource. First, the ESG data aspects have been addressed. Then, literature shows that the combination could be used to integrate sustainability data in a firms' business model and decision-making for a possible competitive advantage, but no further tangibility has been attached to this statement.

By integrating Big Data and AI into a firms' strategy or even the decision-making process, it takes a different approach to implement such data and technologies. This as most literature is about investors implementing it for a better Alpha (financial returns), while not a lot is found on it being implemented from within the financial industry as a way to support decision making. This perspective is mostly lacking within the literature. This could be due to several reasons, the conservativeness of the industry, the novelty of the topic at hand, and the chosen perspective of the application of this combination of technology. It cumulates in the statement by Hasan, Popp, & Oláh (2020) "…, there is a need to expand the formal and integrated process of implementing big data strategies in financial institutions.", illustrating the knowledge gap of information regarding strategies of applying Big Data and AI within a firm.

However, there is literature found regarding the implementation of AI through established datasets. But it was shown that these datasets do not specify the current case of Big Data and AI, and ESG data. The question then raised was that these propositions would still hold up under the scope of combining the facets of Big Data and AI, ESG data, and the setting of the financial sector. Furthermore, the perception of Big Data and AI, the so-called "human" aspect, is mostly neglected within every paper. This includes it being seen as a corporate resource, as it mostly is seen as a technology or combination of technologies as a system on their own. While it was addressed that these aspects are of importance, as seen in the perception of Big Data and AI sub-chapter, no further details were given. This, the knowledge gap again cumulates in an observation which was substantiated by In, Rook, & Monk (2019) and Monk, Prins & Rook (2019). There is no theoretical grounding for the application of ESG data yet.

1.3.6. Summary key literature

As mentioned within the previous sub-chapter, Concluding remarks, several papers are illustrating that there is a need for further research regarding the combination of sustainability, ESG data, Big Data, and AI. This knowledge gap is touched upon directly by several papers. The following table illustrates these papers and is created to give a descript overview of key sources and their relevant key research implications.

Source	Summary	Key research implications
(Monk, Prins, & Rook, 2019).	There is no theoretical grounding for the application of ESG data.	One of the more important papers found regarding ESG data. It illustrates that there is still a long way to go to provide substantiated theory regarding ESG data.
(In, Rook, Monk, & Rajagopal, 2019)	This paper illustrates that ESG data is rooted in alternative data. Furthermore illustrating that the way this kind of data is perceived and how it is applied is just as important as what the actual data entails.	This paper provides an approach for how ESG data is used for investment decision making. The relevant aspects are how ESG data is defined. This definition provides a basis for the definition of ESG data throughout the thesis.
(Kotsantonis & Serafeim, 2019).	This paper addresses ESG data, illustrating that the sheer variety and inconsistency of the measured data and how companies address relevant ESG data leads to data gaps. Further illustrating that there are no well-defined approaches to integrating ESG data.	This paper discusses multiple aspects of ESG data. It illustrates one of the key issues regarding ESG data which add to the aforementioned knowledge gap. That there is still no well-defined approach to integrating ESG data, thus, that the implementation side is still lacking within the literature.
(Antoncic, 2020)	The integration of Big Data and AI technologies could be used to integrate sustainability data in a firms' business model and decision-making for a possible competitive advantage. Furthermore, the board of a firm needs to be 'sufficiently fluent' in the latest sustainability technology to adopt such a way of operating, giving an arbitrary approach to the adoption of such technologies. This paper states regarding the combination of sustainability, Big Data, and AI that it 'could' be done, giving no further tangible proof or application to his claims.	This paper superficially addresses the integration of sustainability data and Big Data and AI. It aims to illustrate how Big Data and AI are used for strategic decision making. Antoncic's claim is one of feasibility, as the author discusses the need of integrating sustainability in the boardroom through Big Data and AI. The author of this paper, however, only provides a superficial basis regarding these claims. Not diving deeper into the subject. The research implications of this paper are of sequential nature. Meaning that it provides a starting point for future research. For example a case study of what this combination of sustainability, Big Data, and AI actually entails and what influencing factors regarding such processes could come to light.
(Hasan, Popp, & Oláh, 2020)	This paper illustrates the current landscape and influence of Big Data on the finance sector. Providing future research directions. One significant example from this paper is the statement of "..., there is a need to expand the formal and	This paper illustrates the research setting as provided by the research opportunity. However, it mostly illustrates, as seen in the quotation, the need to dive deeper into the process of Big Data strategies within financial institutions. It defines a

	integrated process of implementing big data strategies in financial institutions.”.	need for future research regarding this topic. Its implications are of
(Pencheva, Esteve, & Mikhaylov, 2020).	This paper illustrates the adoption of Big Data and AI within governmental institutions. Stating that “a few barriers at the individual level are noted in the literature, but relatively little attention is paid to these”. Even going as far as stating that the mindset of the individual could be of significant importance. But further details are not given.	Here, implications regarding the adoption of Big Data and AI are illustrated. However, not to the extent of the individual level. The quotation provides a significant starting point regarding what to expect when a Big Data and AI solution is integrated. This starting point, however, is only that, a starting point. It does not illustrate what these barriers actually are and only states that little attention is paid to these. Thus, the research implications are that of a potential future research direction, the more personal, or “human” aspects of such integration.

Table 1 Key literature

1.4. Research problem statement

The research problem statement is based on the concluding remarks of the aforementioned research background. As illustrated, the introduction of ESG data, Big Data and AI as a combination and within a firm in the financial is rather unexplored within the literature.

However, what was observed in several papers is that some ESG data aspects have been addressed, but these do not fully match the combination within such a toolset and setting. Literature furthermore shows that the combination could be used to integrate sustainability data in a firms’ business model and decision-making for a possible competitive advantage, but no further tangibility can be found regarding this statement. Integrating Big Data and AI into a firms’ strategy setting, as a corporate resource, or even the decision-making process, is a perspective not often taken within the literature.

Most literature is related to investors implementing such combinations for a better ‘Alpha’, which are financial returns for their investment portfolio (Madhavan, Sobczyk, & Ang, 2021). While the observation was made that there was a negligible amount of literature found on it being implemented from within the setting of the financial industry in other aspects such as a way to support decision making or even the human aspect of implementing such a combination.

However, there is literature found regarding the generic implementation of AI through established datasets. But it was shown that these datasets do not specify the current case of Big Data and ESG data. The fundamental difference between these datasets raises the question to which extent these propositions within literature would still hold up under the scope of combining the facets of Big Data and AI, ESG data, and the setting of the financial sector.

This accumulates in the observation that throughout the literature, published information regarding this topic is quite novel, most papers are from 2019, 2020, and 2021. Another observation is that this field of study has a lot of potential development, according to the researchers publishing in their respective fields. Most state broad statements in their conclusion regarding future research, with the nomenclature used of showing the potential benefits of adopting such technology, or how certain datasets could be beneficial for a firm. Even then, most research is focussed on the potential of ESG data, or Big Data and AI, for investors, and less attention is being paid to the inclusion of these aspects in the operations of a firm or the context of the financial sector.

Thus, the research problem statement is one where there is a broad gap within the literature, as the inclusion of novel digital solutions such as Big Data and AI regarding the analysis of ESG data within a firm, as a corporate resource, and its "human" aspects are yet to be explored. Some theoretical foundation has yet to arise to further continue on this topic. This study aims to provide a starting point to fill this knowledge gap by showing observations and from this define novel theoretical propositions to be tested in future research.

1.5. Research question

To address the aforementioned research problem, the following research question has been established.

Main research question: What observations can be extracted from assessing the introduction of a Big Data and AI toolset applying ESG data within a procedure?

To try and answer the main research question, an explorative case study is defined. This case study aims to extract these observations through a process of gaining qualitative data through direct observations and interviews. These are made during the case, introducing the combination of Big Data and AI with ESG data within a single firm. The decision to go for such an approach lies in the research opportunity provided by the firm. The reasoning behind the choice of a case study will be further elaborated on in the upcoming sub-chapters.

To support the main research question, sub-questions are defined. These aim to steer the research direction and define the operational scope of the case study. These sub-questions are formatted to provide structure supporting the main research questions. This process is further elaborated on in the Methodology chapter.

The sub-questions are defined to mimic the process of introducing a novel procedure using an ESG data, Big Data and AI toolset within a firm. This will be cut into three sub-questions, assessing the current approach of using such data and technology within the firm, assessing the design process and introducing the procedure, and in the end, assessing the procedure.

The first sub-question aims to extract observations by assessing the current approach to these datasets and technologies. It is structured to explore the current situation within the firm, to connect to the second and third sub-question. It aims to bring forward the current state of thought regarding ESG data, Big Data and AI. Potentially illustrating a difference between the current situation, the process, and the results. Sub-question one is defined as follows:

Sub-question 1: What observations can be extracted from observing the current state of the application of ESG data, Big Data and AI within a firm?

As mentioned previous to the first sub-question, the comparison of the current situation to the new one could provide data to structure and substantiate the observations, the approach to the designing and introduction of the procedure using ESG data, Big Data and AI in itself could also entail additional insights. By shining a light on this process and related constraints, information could be encountered that could lead to evidence regarding substantiating the observations, thus, supporting the main research question. The second sub-question focuses on the designing and operationalization of the data and toolset. It aims to gain data through observing the design and introducing the procedure using these data- and toolsets. The second sub-question is defined as follows:

Sub-question 2: What observations can be extracted from observing the development process of a procedure using ESG data, Big Data and AI within a firm?

The third sub-question is one of assessing the case, aiming to create observations to give a backwards-looking approach to the implementation of Big Data AI, and ESG data. Meaning that this sub-question aims to acquire data through assessing the previous steps within the case. This approach gives an overview of how it was perceived within the firm and tries to assess the designed procedure. The third sub-question is defined as follows:

Sub-question 3: What observations can be extracted from looking back at the process of creating a procedure using ESG data, Big Data and AI within a firm?

1.5.1. Scientific relevance

This case study will contribute to and explore the literature stream of integrating Big Data, AI, and ESG data in a firm. While a lot is known about each separate aspect of this combination, the combination itself is rather novel. By observing a case where ESG data is introduced, and used in a Big Data setting, where AI is used to assess the data, perhaps novel propositions can be created for future research. Furthermore, there is a lot known about ESG data and the integration of it within investing, but not using it as a corporate resource. This research could shine a light on this subject.

Current research mostly states that certain ways of implementing a combination of Big Data or the introduction of ESG data, or even the introduction of AI to a dataset could be beneficial to a firm. However, the combination of these subjects is not yet explored within the literature. As ESG data becomes more prevalent due to novel (non-)mandatory regulations such as the EU Taxonomy (European Commission, 2021), and so does the availability of Big Data and AI as information technologies, this field of research interest could become more relevant and prevalent in the near future. This study aims not to completely fill the research gap but aims to shed more light on this subject. It aims to explore this topic by providing observations and from these observations novel theoretical propositions will be created. It, therefore, aims to fill the knowledge gap as stated in the sub-chapter Literature background, as it would provide a basis on which future research could be conducted.

1.5.2. Societal relevance

The societal relevance of this case study is in the application of ESG data, and the assessment of Big Data and AI from a more 'human' aspect. ESG data becomes more prevalent in the near future, as seen in the upcoming (non-)mandatory regulation of firms having to report on their ESG performance. An example of this is the EU Taxonomy. This creates vast amounts of ESG data on which future analysis could be done relevant for management information. The research could show how ESG data could be used within different kinds of settings within a firm. Furthermore, as mentioned in the background sub-chapter, Big Data and AI have a certain tendency of meaning different things and different applications over time. By exploring the introduction of the latest iteration of these technologies, some light will be shed on how (future) decision-makers could introduce, use, and apply these technologies and related results. Thus, it assesses it as a corporate resource, a perspective not often taken.

Chapter 2. Methodology

The methodology chapter will describe the approach taken to answer the main research question. First, the research methodology will be discussed. It includes the defined research framework, initial propositions to give define the research direction, the research design, and how data is collected and coded throughout the case study. After which the data analysis method will be discussed. In the end, the validity of this approach will be discussed.

2.1. Research methodology

The research methodology will be illustrated as follows. First, a conceptual framework is given regarding this research, after which the reasoning behind the single-case study approach will be discussed. Then, the initial propositions are stated which will be used throughout the case study. Then, the case study research design, thus, the workings of the case study, will be discussed. The approach taken throughout the case study to collect data is then defined. From this data, observations and propositions should arise, thus, the coding and data analysis methods are defined in the last sub-chapter of the research methodology.

2.1.1. Conceptual framework

The conceptual framework adhered to within this research consists of three stages, the design and define stage, the prepare, collect, compile, and analyse stage, and the analyse modify, conclude and report stage.

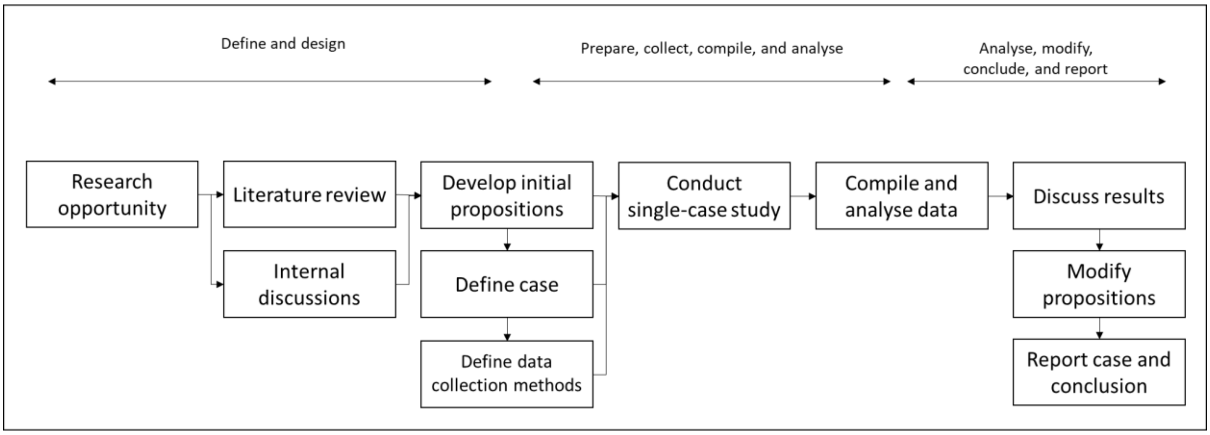

Figure 1 Conceptual framework

The conceptual framework is developed by making use of "Case Study Research and Applications, Design and Methods" by Yin (2018). The idea behind the conceptual framework is to show the steps required to take to successfully conduct the case study and ensure future replicability. This process is defined into three phases, where the longitudinal steps taken are linearly and steps taken with overlap are done simultaneously.

Define and design

The first stage, define and design, illustrates the following. The research direction, thus the starting point for this research, had to come from something. An opening perspective. This perspective comes from the research opportunity provided within the firm. This research direction has led to a literature review and internal discussion within the firm. From the literature review, a knowledge gap has been identified and simultaneously internal discussions have been held regarding this knowledge gap. This has directly led to the main research question this thesis aims to answer. From the main research question, the conceptual framework was created. The research opportunity provided a basis for a literature review and internal discussions, which gave rise to initial theoretical propositions. These theoretical propositions are used to provide scope and directions to the research, with the aim of refining these through the case study.

The results of the literature review are illustrated in the sub-chapter Literature background. The approach taken regarding the literature review is illustrated in Appendix D – Literature research methodology. The results of the define and design step are providing the basis for the case study and are illustrated within the current chapter.

Prepare, collect, compile, and analyse

The second stage, prepare, collect, compile, and analyse, illustrate the beginning of the case study. This case study is set in place to refine the initial propositions. The case to be studied is one of adding the combination of ESG data, Big Data and AI, within a firm, and the purpose is to explore what the addition of this combination entails with the focus on how it is being used as a corporate resource. What is furthermore assessed within this stage are the data collection and analysis methods. These methods are defined to match the intended goal of this research. More can be found, regarding these collection methods and analysis, in the upcoming sub-chapters of this Methodology chapter.

Analyse, modify, conclude, and report

The last step in the conceptual framework is the analyse, modify, conclude, and report stage. Here, the data gathered from the case study will be analysed according to the methods described in this chapter. The results will be discussed and the initial propositions will be modified according to the discussed results. In the end, the results are reported and a conclusion to the main research question will be given.

The notion of "propositions"

As there are many ways to illustrate observations, ranging from visual to descript and elaborately written, the decision has been made to make use of theoretical propositions to portray the results of the main research question. Throughout this thesis two kinds of propositions will be used, initial theoretical propositions and novel theoretical propositions. The initial theoretical propositions are derived from literature and discussion within the case study environment. These are used to provide a scope and basis on which the case study can be conducted. The case study aims to dive deeper into the subject, refining the initial theoretical propositions and through the discussion of the results, novel theoretical propositions are proposed. Novel theoretical propositions will illustrate the answer to the main research question and are used to illustrate the relevancy of the observations made.

2.1.2. Single-case study approach

The approach taken within this thesis is one of a single-case study design. The reasoning behind the decision of going for a case study is one based on the main research question. Information to answer the main research question, to explore the knowledge gap related to this question, can only be done through the actual implementation of the combination of ESG data, Big Data and AI. By aiming to understand a real-world case, several theoretical propositions for further inquiries are developed. Case studies are an ideal approach to gather a broad range of results where novel propositions, theories, can be built upon. It is an ideal way to dive deep into unexplored bodies of knowledge. Or, as Walsham (1995) stated, "Single case study designs are considered to be the most appropriate technique for conducting detailed in-depth studies". One similar example within literature is the case study done by Zhang & Poole (2010). Where a case study is done to analyse Virtual team identity construction and boundary maintenance.

This study aims to extract observations from assessing the introduction of a Big Data and AI toolset using ESG data within a procedure. Thus, it aims to explore this topic, which defines this case study as an explorative case study. The results of these observations are novel theoretical propositions. These propositions could lead to novel theories to be tested, a basis for future research, or nothing at all. However, to understand such a real-world case, the understanding that such a case involves significant contextual conditions have to be taken into account. Meaning that throughout the case study the main focus will be on acquiring and processing data with the goal of abstract theory-building in the form of defining novel theoretical propositions. This goal of a single case study has been characterized perfectly by Murale and Preetha (2014), who stated "(a single case study) may lead to the initial foundations of a theory formation or it may act as a base for broader research".

The proposed case study aims to not focus on aspects solely adhering to the unique context a case study setting brings to the table, as this is one of the more common pitfalls of such case studies. Thus, such a case like the defined case study focuses on contemporary events, where it does not require control over behavioural events. Meaning that the case focuses on the events at hand while maintaining the level of abstraction where empirical light could be shed on the concepts as described in the problem statement. This, to reduce one of the major pitfalls of a case study, the lack of focus (Yin, 2018).

The case itself is of a single-case study format. The decision to go for a single-case study instead of a multiple-case study is both theoretical and practical in nature. The theoretical nature is one that within a case study, the study aims to shed empirical light on novel theoretical concepts and principles (Yin, 2018). Within this case, a single novel process will be introduced within a firm. This causes the case to be, as Yin (2018) describes it, a critical case, where the study is almost analogous to a single experiment.

Additionally, the decision to do a single-case study is also one of a practical nature. As this is a graduation thesis research project, there are only a limited amount of resources available. Only one researcher is available, having only one semester to accomplish this research. The proposition of a multiple-case study within this context would become unfeasible as it would increase the workload to the extent that it is not realistic to finish the study within the proposed time. The core of the case, the designing of the procedure is part of a research opportunity as provided through an internship within a Dutch financial institution. This internship allows research in this field of study and provides necessary resources.

Good to know is that the single-case study, is one of a holistic approach. Meaning that it involves only units of analysis on a single level. This differs from an embedded case study, where there are many subparts to be analysed. The decision to go for a holistic case study is similar to the one going for a single-case study. Meaning that practicality and the specific focus and scope of the results reduce the need for multiple analyses.

Other methods have been assessed to answer the research questions. One significant research approach taken into account was the application of an experiment. The benefits are that of a significantly higher internal validity, however, the external validity is significantly lower as such an experiment could never show the real-world context of procedures, the true application of tools by employees within a firm, and all its relevant (human) interaction and opinions that could lead to a proper answer to the main research question. Furthermore, the lack of initial literature to build upon also hampered the decision to not go for an experiment. The practical nature of such an experiment also weighed in, as setting up an experiment of introducing new technology within this setting would be one of replicating a full business unit of a company, effectively making it a case study.

Another option of addressing the knowledge gap and solely answering the main research question is a desk research approach. Desk research has been considered, however, the decision was made not to go for one. The reasoning behind this is the prevalence of the research opportunity as provided through an internship by the firm. Furthermore, the results gained from desk research would, in the authors' opinion, not capture the broad range of possibilities the introduction of such a combination of technology within a financial institution entails as it just reviews what other authors have done.

There are criteria by which the results of the case study will be judged successful (or not). These criteria are; that the case and the process itself is conducted in such a way that validity has been taken into account, see the sub-chapter Validity regarding this case study for additional information. That, in the end, there are novel theoretical propositions created. Which are substantiated and discussed through results and potentially relevant literature. Concluding, that the main research question and its related sub-questions can be answered by going through the process as defined by the research framework.

There, of course, are also negative connotations surrounding the adoption of a case study, one of the major connotations of such a case study is that one has to identify and delineate the uniqueness of the case itself, that the artefactual conditions surrounding the case may take the overhand within the case itself and the validity of the decision to go for a case study as a whole is challenged. More on this in the sub-chapter 'Validity regarding this single-case study'. The results of the case study will also be compared and discussed according to the aforementioned aspects in the discussion chapter of this thesis.

2.1.3. Stating initial propositions

Before addressing the actual case, one has to acknowledge that such a case does not come falling out of the blue. There is some implicit theoretical orientation, an opening perspective, to address beforehand. An exploratory case study such as this one has the tendency to take an extremely broad. As one does not expect to find the Spanish Inquisition (Chapman, et al., 1970), not having this theoretical orientation delineated could lead to this never-ending rabbit hole of non-relevant tangents addressed in the preliminary steps alone. Even more, ignoring them could lead to an inherent bias throughout the case study (Yin, 2018).

For this purpose, the approach of defining initial propositions are used within this thesis. Initial propositions are a tool within qualitative research to counter such aforementioned challenges. One of the more prominent cases using such a method was the case of Wilson and Wilson (1988), who tested the level of agreement between judges of qualitative case data. This case showed that defining initial theoretical propositions before the data-gathering commences provided a valid and substantiated method for an explorative case study (Hyde, 2000).

These initial propositions are based on the following two factors; initial literature and initial observations through discussion with actors within the case study environment. There is no precise way for setting criteria for which these propositions, and therefore the related results and conclusion based on these propositions, can be 'scored against' (Hyde, 2000). However, as Hyde (2000) states "Cases which confirm the propositions enhance confidence in the validity of the concepts and their relationships; cases which disconfirm the relationships can provide an opportunity to refine the theory.". Showing that stating initial propositions prove to be a substantiated approach to gathering data. Thus, providing a valuable addition to the methodology of this case study as a theoretical orientation to be adhered to throughout the case study.

This theoretical orientation is formed through the creation of several initial propositions, theoretical statements representing issues found in the literature and/or practical matters. The following initial propositions have been identified.

- The introduction of Big Data and AI into a process is more of an approach to carry a big stick to silence sceptics of the subject of ESG than to give results not anticipated by relevant stakeholders using the tool- and datasets.

This proposition is based on the initial meetings within the firm and is based on the following observation. While certain sustainability topics were discussed, the importance of these topics was usually one of the main points up for discussion. Numerous arguments are made about "we talk to the relevant stakeholders", where the other party in the meeting countered with "Yes, but we talk to more stakeholders" etc. The notion of a data-driven approach could silence sceptics, as it could provide a fundamental approach to a subjective step in the process of defining the importance of, in this example, sustainability topics.

- It, however, is a very big stick and has the required effect, showing that the social dimension of implementing such abstract tools- and datasets should not be overlooked.

As the notion of Big Data and AI is not new, this notion has been used within the firm the case takes place. The initial observation was made that this had a positive effect on the credibility of the arguments made within the discussion. This could be perhaps in line with the 'distrust' as addressed by In, Rook, Monk & Rajagopal (2019).

- The financial sector is progressive when it comes to the adoption of novel technologies.

This proposition is based on the paper of Corea (2019), which concluded a contradictory point. While this paper was rather new, 2019, as time progresses and the financial sector also seems to be renowned for its innovation regarding novel information technologies. One only has to look at companies using algorithmic trading to further their profits.

- The source of data is not questioned when the method of implementing data is solid in the eyes of the receivers of the results of the method.

This proposition is in line with the social aspect of integrating new technologies within a firm. It lies in the perception of the employee. Pascheck, Luminosu, & Negrut (2020) state that data is the most important aspect when it comes to this context. I believe that, while their paper had a focus on the more technical side of things, the human context should not be overlooked and this proposition could be the result of this human context. This is further supported by the initial observation made through expert meetings with the firm, where people were more focused on the process and procedures than who collected data.

- Throughout the introduction and utilization of the Big Data and AI toolset, there are always human interactions and interpretations of the data required. As an ontology has to be created, human bias is still inherent to the procedure itself. This shows that, while the assumption within a firm is that such a toolset solves the bias, it is inherently not true.

An ontology is a list of topics to be assessed through the NLP software. It's a way of classifying things. Here, I think that the subjectivity of the process goes down the moment one has to define it themselves. It is in the eyes of the beholder how an ontology is created. However, I think that the perception of such a tool is more important than the input data. Thus, it only pretends to solve bias, and this pretending is accepted by the users of this tool.

- The combination of ESG data analysed through a Big Data and AI toolset creates similar results to any other dataset analysed through the same method, of which conformation is found in the already existing literature.

This statement would illustrate that similar results can be attained such as in the study of Pascheck, Luminosu, & Negrut (2020). Showing that there could be a chance that the different kinds of datasets used could lead to similar results. Building on the statement of Pascheck, Luminosu, & Negrut (2020) that in essence, the nature of data is again important to the process.

- Most data sources and methodologies used to lead back to, or are based in, 'expert opinion', even if the data and/or methodology claim to be objective, disproving that most data-backed or data-driven decisions are unbiased.

One initial observation made was that technologies such as Big Data and AI are implemented to give an unbiased, near straight-forward view of certain processes. At least, they tend to have to possibility to do so. Even the Harvard Business Review dedicated an article to such decision-making processes (Colson, 2019). However, one other observation made during the initial meetings within the firm is that the structuring and creation of datasets are usually founded on human expertise. Even the classification of parts of the dataset or inclusion or exclusion of certain datasets could alter the results, even if the same procedure is being used. Perhaps this stream of thought comes forward through the case study.

2.1.4. Case study research design

The **Error! Reference source not found.** and Stating initial propositions sub-chapters gave a holistic overview of how the case study is set up. This sub-chapter, Case study research design, shows the link between the way data is gathered and the main research questions.

The case is defined based on the main research question. The main research question is supported by three sub-questions. These sub-questions are used to give structure to the case study itself. The sub-questions are defined to gather data from each step within the case study. The results of this data are then used to assess the aforementioned initial theoretical propositions, through which the sub-questions and main research question can be answered. Each sub-question is linked to a step in the process of designing, defining and introducing, and assessing the procedure. In these steps, different methods for data gathering are applied. This structure is illustrated in the following diagram.

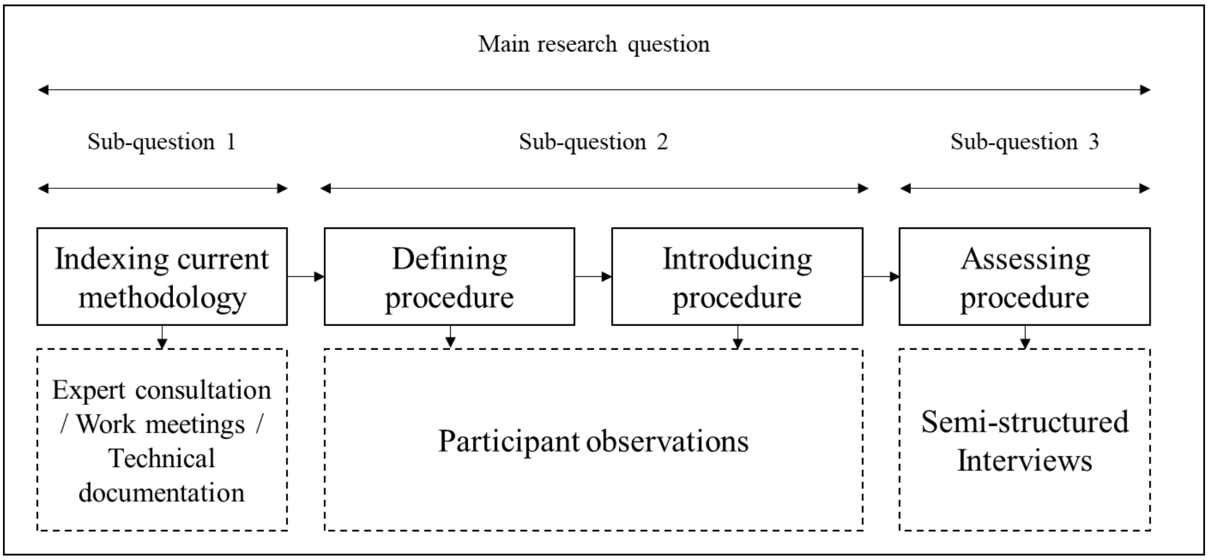

Figure 2 Case study and research question diagram

The link between the case study and research design, and the main research question including sub-questions and related data methods is as follows.

Indexing the current situation, sub-question one

The first sub-question is "What observations can be extracted from assessing the current approach of applying ESG data, Big Data and AI within the firm?".

The current situation will be indexed through data gathering in the form of consulting internal experts within the firm, giving their objective assessment of how processes of acquiring ESG topics and data is accomplished. It also focuses on the current approach of Big Data and AI in combination with ESG data. Furthermore, documentation will be indexed and current processes will be catalogued. Stakeholders within these processes will be consulted to acquire information about the current methodology. The goal is to gather as much information as possible on the subject of the perception of Big Data and AI, and ESG data, how these are taken into account, it's implemented and perceived within the firm. The expected results of this process are that it will lead to a rabbit hole throughout the firm, where a lot of information will be gathered, not all of it relevant. As the research setting is within a firm, there are numerous meetings regarding this subject. Minutes of these meetings, related memos, and observations made in so-called 'analytical episodes' will also be used similarly for data input.

Other approaches to assess the first sub-question have been thought of. However, the first sub-question is one of indexing the current situation to create a comparison between an old and new situation and to assess the current stance of the topic at hand.

The creation and introduction of the procedure, sub-question two

The second sub-question is "What observations can be extracted from observing the design and introduction of a procedure using a Big Data and AI toolset applying ESG data?".

After the current situation has been assessed, the procedure will be defined by incorporating a Big Data and AI toolset in combination with ESG data. Defining the process will be done by finding the right datasets, defining the ESG dataset of ESG subjects, the acquisition of the right Big Data and AI toolset, and consulting internal stakeholders of possibilities and limitations. The mode of case study evidence collection is participant observations. Meaning that by participating in the process analytical episodes can be deduced, learning and collecting evidence by doing.

The method within the second sub-question is thus one of participant observations. As the case is one of defining a procedure and making the procedure operational to learn from. Other methods have been assessed, such as the inclusion of interviews, but this method is more appropriate for the last sub-question. Furthermore, using interviews for this step would create a higher chance of generating similar data. As the aim is to explore the knowledge gap, the abundance of similar data is not preferable. The participant observations approach was chosen to give a more birds-eye view of the case, matching the second sub-question. Another approach possible was the one of direct observation, however, this approach is one with many limiting factors. Such as the availability of the participants to be observed and the timeframe in which these observations have to fit.

Assessing the procedure, sub-question three

The third sub-question is "What observations can be extracted from assessing the introduction of a Big Data and AI toolset using ESG data within a procedure?".

This last sub-question relates to the process after the introduction of the procedure. Here, the procedure and the adoption of it will be assessed. This step will be done by conducting semi-structured interviews, depending on the data gathered in previous steps. Thus, this last step in the process is one of reflection on the procedure, what the perception of it is, and how it is operationalized within the firm.

The method of addressing the third sub-question, semi-structured interviews, is one of the best ways to internalize and assess a procedure. Through this method, information can be attained that otherwise would not be available through other methods, as information gained based on the conclusions and opinions of the employees working with the procedure might add to the creation of novel theoretical propositions. Other methods to gain this information have been assessed, such as observations. However, as these were already conducted for the second sub-question, the data gained would be double. Interviews further support or disprove, the gained results. It is even possible to encounter novel results which have not previously been observed, providing information to further explore the knowledge gap at hand.

2.1.5. Data collection procedures

The previous sub-chapter linked the main research question and its sub-questions to the case study and illustrated where data is collected. Within this sub-chapter, we dive deeper into the data collection procedures.

The research design aims to collect data, with the main approach in mind to organize "many ideas from analysis of data" (Glaser & Strauss, 1967). To gather data on which theories are formed. This is further substantiated by the introduction of the 'maximum variation sampling' rationale used on which the data collection approach is tailored (Hammarberg & De Lacey, 2016). This rationale consists of gathering the maximum amount of variation within the data. Meaning that within this case study, the focus is to gather data through as many valid approaches as possible within the boundaries of the case study.

This rationale comes forward throughout the case study, as many forms of case study evidence collection have been used. This with an eye on the analysis method of triangulation. If multiple sources point to a certain direction, it strengthens the proposition that this direction is a logical and perhaps valid one. Then substantiated proposition-building can take place. In the end, the collected case study evidence is a close-up and in-depth coverage of the case itself. The sources of evidence used are:

- Expert work meetings
    - Observations/minutes in the form of memos of expert consultations
    - Observations/minutes in the form of memos of expert work meetings
- Currently available information
    - Technical documentation
    - (Archival) records
- Field notes in the form of memos
    - Participant observations in the form of weekly memos
- Interviews
    - Semi-structured interviews

Regarding the structure of the collection of evidence, the following steps will be adhered to. First, a moment of observation is apparent, where observations are made. This can be during the expert meeting, interview, etc. Here, minutes will be written down. These minutes illustrate the raw evidence of what is apparent and consist of what has been said, discussed, and where possible summarized. This will be summarized in memo form, short sentences illustrating what has been discussed and distilled down to its core. From here results will be coded according to the next sub-chapter, Coding and data analysis methods.

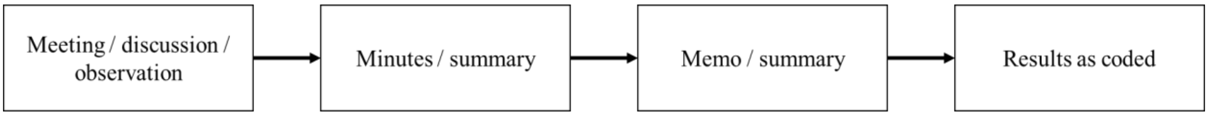

Figure 3 Approach of registering evidence

The evidence, as coded, will be collected in a so-called 'evidence register', a location in the form of an Excel sheet where this data is summarized. From here observations can be made regarding the gained evidence. Overall, documentation of each step within the case plays a prominent role throughout the data collection process. The metadata in itself will remain at the firm the case study takes place. However, the processed data in the form of the evidence register will be included in Appendix A – Results and the results are substantiated in Appendix B – Clarification results.

Good to know is that any observation has its goals, meaning that one observes something with a specific goal in mind. Within this case, the observations are made with the main research question and its sub-questions in mind, assessing the designing and introduction procedure through the lens of what can be learned from it. However, participant observations provide some major challenges (Yin, 2018). Biases are produced as one observes the process, not as an external observer, but having a position to advocate certain practices contrary to the interests of good social science practices (Becker, 1958). However, participant observations provide opportunities such as the ability to observe otherwise inaccessible events, and the ability to perceive reality from the viewpoint of someone "inside" the case (Yin, 2018). Within the circumstances of this case, the participant-observation fieldwork approach has been considered and due to the nature of this case, deemed suitable. As one can fully observe the designing and introduction of such a toolset by participating.

Furthermore, employees directly and indirectly associated with the procedure will be interviewed about their experiences. As stated by Belanger and Watson-Manheim case (2006), one of the more important aspects of a case study is the semi-structured interviews for collecting data. For these interviews, the interview setup has been created according to the previous steps. It is used for assessing and analysing the effects of the introduction of the toolset and the gaining of feedback regarding the previous steps of the case. This approach to interviews is further defined in Appendix C – Interview procedure.

The data acquired and stored throughout this thesis adheres to the Data Management Plan according to TU Delft standards. This plan can be found online at https://dmponline.tudelft.nl/. The results and clarification of results are fully anonymised. The rationale behind this is one of practicality. If an employee within the firm the case study is happening sees the opinions of colleagues and backtracks this, the interviews are not objective anymore due to the confirmation bias. Furthermore, there could be internal consequences if one employee disagrees with another one.

2.1.6. Coding and data analysis methods

As Denzin and Lincoln (2018) state, "the process of qualitative analysis is a creative process, not a mechanical one". There is no single definitive way to accomplish qualitative research, the creation of such an analysis can be seen as a form of intellectual craftsmanship (Denzin & Lincoln, 2018). Thus, the analysis method used within this thesis is curated to fit the data gathered to, at the end of the process, answer the main research questions. The data analysis method will consist of two parts, coding and analysing the data.

Coding, categories

The first step to coding data is that all data is to be categorised. As Byrant & Charmaz (2007) state "We can think of categories as forming the theoretical bones of the analysis, later fleshed out by identifying and analysing in detail their various properties and relations.". This categorisation will be the first step to be illustrated within the results. These categories play the role of 'conceptual elements of a theory' (Glaser & Strauss, 1967). They emerge from data and achieve a higher level of abstraction, which will be used to further assess the data to be coded. Within the Results chapter, the first things to be addressed are these categories.

Coding, coding method and "themes"

The coding approach taken is to support the analytical pathway taken. Meaning that on the coding approach the analysis strategy to be described in the following paragraphs can be applied. The aim of coding is that symbolic codes, "themes", emerge from qualitative data (Aldiabat & Le Navenec, 2011). The notion of "themes" will be used to define each code. These themes help to understand the "Phenomenal world of individuals" (Aldiabat & Le Navenec, 2011), as the integration of a toolset using Big Data and AI is done in a firm setting, where it is being used by employees of the firm.

To dive deeper into the fundamentals of the coding process, the initial approach taken to coding is an 'exploratory' method. It first explores categories, after which it preliminary assigns codes to the data before more refined coding systems are defined and applied (Saldaña, 2009; Bryant & Charmaz, 2007). A coding system for this case study has the requirement to be adaptive to deal with the broad and novel themes. This to support the explorative part of the study, one has to be able to include themes and subjects not previously encountered or initially associated with the main theme of this study. Therefore, a certain amount of flexibility is required. Thus, in a sense constant comparative analysis is used to assess and define codes, as older codes are compared to novel codes, increasing more abstract concepts and theories along the way (Chun Tie, Birks, & Francis, 2019). Through this method of continuous analysis comparative and novel codes can be found within the data gathered and addressed.

To further relate this method to current literature, the arrangement of coding will be done through a Holistic Coding method, as it gives a single code to each large unit of data (Saldaña, 2009). Holistic coding will be used to grasp basic themes of the data rather than assessing the data line by line. Thus, within the evidence register themes encountered within the data will be described.

One of the benefits of holistic coding is that the approach can be tailored to fit between first and second-order coding approach, a 'middle-order approach' (Saldaña, 2009). This has been done within this case study due to the following aspects. Mainly, the approach taken is one of finding general propositions to explore the topic of this study. Thus, granular details through excessive coding do not significantly contribute to the results. Furthermore, the amount of data and time constraint is prevalent. The coding method, a middle-order holistic approach, is one that saves time while delivering functional results for this study. This coding scheme shows in essence several themes encountered throughout the case study.

The complete resulting coding scheme regarding the results is found in Appendix A – Results. Appendix A comprises a table, on the vertical axes the moment of observation is illustrated, on the horizontal axes the coding itself. Where these two axes cross and an 'x' is shown, which illustrates an observation.

Theory-building

On this classification method of data gathered, a theory-building structure will be applied within the Discussion chapter to illustrate the results gained from the case study (Yin, 2018). With this structure, the results will be illustrated through theoretical arguments being made. This will be used to unfold key ideas explored by this case study.

This theory-building approach of illustrating results is structured as follows: First, the main theory and story will be built upon the themes found within the evidence register. Themes will be analysed and interpreted, where the existence and frequency of concepts will be assessed and matched (Kolbe & Burnett, 1991). In essence, this is a form of thematic analysis as it is "a method for identifying, analysing and reporting patterns (themes) within data" (Braun & Clarke, 2006). Within the evidence, the concept of 'triangulation' will be applied. Which is a theory stating that to use evidence within theory-building, it has to come from different sources and assessments to support the proposed theory (Yin, 2018). This will be done as follows, within the sheet of Appendix A exhibiting the results, for each observation a check will be made vertically, over the moment of observation. This is to see if a link can be made with other themes, aiming to find a novel correlation between themes. Then, it will be checked vertically, over multiple observations, to see if a pattern perhaps emerges or aiming to find commonalities between these observations.

Then, this theory will be linked to an initial proposition when possible, assessing the relevance and feasibility. A comparison will be made between the theory built based on evidence as previously mentioned and the development of the proposition. The result is one where this initial proposition could be continued or adapted to a novel proposition. However, it is also the possibility that this combination of theory and initial propositions lead to nothing. These propositions will be illustrated within the Discussion chapter.

From this process, novel propositions are built and discussed. These novel propositions are used to answer the main research question. As stated at the beginning of this sub-chapter, there is no clear-cut approach to qualitative data analysis. While other data analysis methods could be applicable, this approach to qualitative data analysis aims to assess data from a holistic single-case study to provide relationships strengthening or disproving the initial propositions or even defining new ones. By taking this approach, structure will be applied to an otherwise notoriously broad approach to data analysis explorative studies are known for.

Results from a holistic single-case study

Nevertheless, such a proposition is just one idea that emerged from a holistic single-case study. Meaning that rivalling theories could also be prevalent. Through tangential literature and similar theory-building upon found evidence rivalling theories and perspectives will be discussed after the novel proposition is stated (Gibson & Webb, 2012).

The portrayal of the results of the analysis

The following image illustrates the structure of how the results, discussion of the results, and the conclusion tie together.

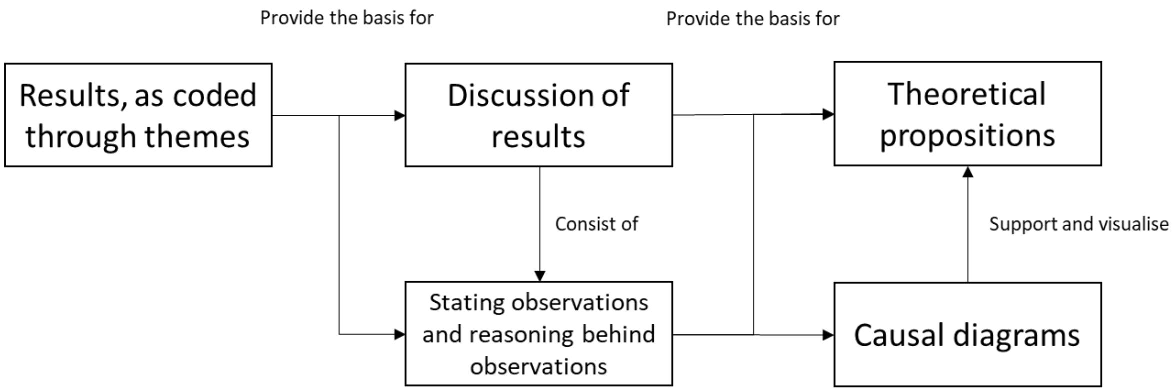

Figure 4 The portrayal of the results

As previously mentioned, the results of the case study are coded through the notion of themes. These provide the basis for the discussion of the results. Within the discussion of the results, the observations are discussed and the reasoning behind the observations are given. The latter illustrates one important aspect of a case study, where one aims to learn and conclude from the phenomena observed, not only reporting the phenomena themselves. This provides the basis for the theoretical propositions. These propositions are used to answer the main research question of illustrating the extracted observations.

These theoretical propositions are supported by causal diagrams. This is an approach to illustrate causal relationships within a conceptual model. It will be used, where possible, to illustrate the causation of the theoretical proposition. This approach of visualization has been chosen as it helps future research by providing clear-cut statements to be tested. Meaning that it provides a model in which the causal relationship of a proposition stated from this study can be falsifiable (Barlas & Carpenter, 1990).

How these causal diagrams can be read are as follows, if X positively affects Y, it is illustrated with an arrow between factor X to Y and a + above the causation. If X negatively affects Y, it is marked with a – above the causation. If X affects Y, both within a positive or a negative sense, or causation is observed but its effects are not fully clear, it is denoted with a +/- above the causation. An example of this is illustrated below, in black the causal diagram, in grey the related explanation.

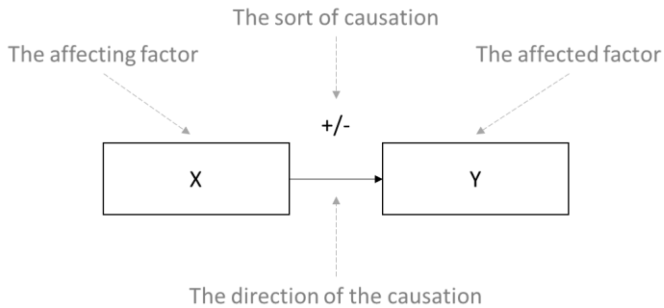

Figure 5 Example of a causal diagram

2.2. Validity regarding this case study

While a case study is a valid way of exploring a novel subject, there are limitations and caveats regarding the design of such a case study. One has to take the construct-, internal-, and external validity into account, and the reliability has to be assessed too. Other ways of exploring this topic have been assessed but were deemed unviable due to the presented research opportunity.

One of the major limitations within this single-case study is that the artefactual conditions surrounding the case could create a unique setting where no valid evidence can be collected. Valid in the sense that deductions can be made from the case study research, not from the case itself. Thus, when all evidence leads to the case itself, it can be concluded that the validity of the case itself is not there. Even more, these artefactual conditions could influence the findings. By having assessed these artefactual conditions within the case through the form of the first sub-question, the method of explanation building is used to mitigate the potential adverse effects of these artefactual conditions.

Thus, to provide strength and validity to this case study, the following perspectives on this validity will be discussed: construct validity, internal validity, external validity, and last of all, reliability.

Construct validity

Regarding construct validity within this case, the case itself consists of illustrating the effect of introducing a new procedure consisting of ESG data, Big Data and AI. Construct validity means that the research measures what it means to measure. The objective of the case study is to find data for the creation of novel propositions to be used to fill the knowledge gap as described in the sub-chapter Literature background. Within the case itself, construct validity is assessed by the operationalization of the case. The focus of the case is on the addition of ESG data, and a Big Data and AI toolset, and what this addition entails within a firm. The 'measurement' will be done through, as mentioned within the 'Case study research design', multiple sources of evidence within the case itself. The design of the case and modes of evidence collection improve the construct validity.

Internal validity

Internal validity is mainly a concern for other types of case studies such as explanatory case studies. E.g. if A leads to B, a causal relationship could be defined by the researcher while ignoring an event C which could have led to the alteration of b. Even more, if a deduction is made about a certain event, this deduction has to be airtight. Within this exploratory single-case study research, internal validity is a non-pressing matter as exploratory research merely brings observations to light, which could be quantified in future research with studies of higher internal validity.

External validity

Whilst internal validity looks at the internal assessment of the case, what is subjectively more interesting is the outward generalization, the external validity. The common approach of statistical generalization of samples and population with a $p < 0.5$ would be a misguided one, this case study strives for a more analytical generalization. The propositions as an outcome of the case study could lead to indirect generalization. Meaning that, as mentioned in the Stating initial propositions sub-chapter, these propositions are the basis for external validity. What is of importance within this case is the evaluation criterion of applicability regarding external validity. Within qualitative research, the transferability of the research findings show applicability outside of the context of the research setting (Hammarberg & De Lacey, 2016).

Reliability

Reliability is the final aspect of a study that has to be taken into account. Reliability within the case study is defined through the case study design (Yin, 2018). Through the defined process, the procedures throughout the case study are made as explicit as possible. With this and the research setting of the case study in mind, a good analogy regarding reliability to be made is one of auditing. The aim is to create an auditable case study, meaning that an auditor could go through the same process as described, and come to the same results. However, as this case captures real-life experiences and aspects of personal and social meaning, it is generally accepted that these are not identical from one person to the next, so might be the replicability of this case study (Hammarberg & De Lacey, 2016).

Chapter 3. Case study characteristics

The characteristics of the case study will describe what the case of 'introducing Big Data and AI' was all about, the people involved, the details of the software involved and the technology behind it.

To answer the main research question through the case study, the case study characteristics have to be delineated. The case itself was conducted within a Dutch bank, thus within the financial sector. The characteristics of this bank are that it is operational in North-Western Europe, is in the top five regarding its size within the Netherlands, and provides multiple financial products, being asset management, commercial banking, investment banking, private banking, and retail banking.

As previously mentioned, the case study itself was done within the settings of a Dutch bank. It takes place within the business unit related to strategy and sustainability. The broad aim of this business unit is to help and steer the firm into a more sustainable direction. To do this, several approaches are part of their toolkit. Approaches like reporting, one-on-one conversations with internal stakeholders, but also providing information in all kinds of formats such as presentations or news bulletins. The case study is intertwined with the latter form of steering, providing information. This team is the direct 'owner' of the proposed procedure.

This exploratory research study aims to assess the integration of ESG data, Big Data and AI into a firms' processes. For this case study, a procedure is defined for the process of assessing ESG topics to transform the current expert-opinion based process to a process rooted in ESG data, Big Data and AI. This procedure is to be run every quarter by one team member of Group Sustainability. It provides a list of sustainability topics relevant to the firm and with it also a list of upcoming sustainability topics which could become relevant to the firm. These results could be used for informative and steering purposes in the form of management information.

The procedure of the case study and the gathering of results follow a similar path. The following steps of introducing a novel procedure are within the scope of the case study:

- Assessing the current situation
- Designing the procedure
- Introducing the procedure
- Assessing the procedure

Within the assessment of the current situation, documents will be assessed looking into the current use of similar systems. Furthermore, as this was the preliminary stage of the procedure, expert meetings have been held to assess how the procedure should be created. These expert meetings are with people from within the bank who have a stake in the procedure. Meaning that they will use the procedure themselves, or the results thereof.

3.1. Assessing the current situation

The first step in the design of the procedure is to assess what to design and how. As the aim of the procedure is to assess ESG topics, one has to assess what is an ESG topic and if it would be relevant to be assessed. This list of ESG topics is the first dataset to be included in the procedure. It consists of around 750+ topics, topics will be added every quarter. Within this part, numerous meetings will be taken part to assess the current situation, to see how a procedure using Big Data, AI, and ESG data can be implemented.

3.2. Designing the procedure

The second step was to define by which method this ESG dataset is to be assessed. The decision was made to go for a third-party SAAS supplier, providing a Natural Language Processing (NLP) based analytics method in combination with an automated machine learning (AutoML) taxonomy feature and real-time public news and blog posts datasets. This decision is based on scope, time, and resource constraints, as the acquisition of such datasets, programming of one's own NLP model, and AutoML features is too big of a task. Furthermore, this case aims to observe the integration of such technology within a firms' processes, meaning that programming is out of the scope of this research.

3.2.1. NLP

Within this case, NLP is being used to assess the provided Big Data, dataset. This dataset will be analysed regarding published articles within media sources. It uses NLP to assess what is being written regarding ESG factors, what the sentiment is regarding a specific ESG factor, and in what context is being written.

3.2.2. AutoML

AutoML falls under the nomenclature of machine learning and data mining. Within this case, AutoML is used as a feature provided by the third party within their platform to define the taxonomy of the analysis. Meaning that it provides a way for people without expertise in the field of machine learning to create a taxonomy to be used within the analysis. It is using Name Entity Recognition to find and define multiple variations of the entities on which the analysis is done. In this case, the ESG dataset is to be analysed over the entity of the geographic area the bank operates in. Furthermore, it uses Latent Dirichlet Allocation (LDA) to assess the ESG dataset to be analysed. It automates ESG related topics and themes from this dataset with the combination of adaptive learning models generating new themes to be relevant within the case.

3.2.3. ESG

The procedure aims to analyse ESG related topics, thus, topics classified under Environment, Social, or Governance. Within the procedure itself, however, no direct distinction is made within the procedure between the classification of E, S, or G. The reasoning behind this lack of distinction is due to the way the toolset approaches topics. The NLP and AutoML features approach each topic in a similar fashion. Thus, a topic such as 'Biodiversity' which is categorised under Environment, is being approached similarly as 'Human rights', which is classified under Social.

3.2.4. ESG Data

As the goal of this case study is to assess the combination of ESG data, Big Data and AI, it is interwoven in the designing of the procedure as follows. As illustrated in the previous chapter, ESG data has to adhere to the following conditions (In, Rook, & Monk, 2019).

- Unstructured
- Multi-faceted
- Context-dependent
- ESG related
- ESG data is alternative data

These conditions are met when the aforementioned ESG topics are assessed through a Big Data set. The conditions for ESG data are met as the moment these ESG topics are assessed through the Big Data set, the data regarding these topics are essentially ESG data, thus, only using a certain portion of the dataset adhering to the initial conditions. Meaning that only ESG data within the Big Data dataset is being used.

3.3. Introducing the procedure

The moment the procedure was designed, it is run within the firm. The available resources will be used to complete the designed procedure. This also will be done through participant observations. The operationalization, input, and output will be assessed. The aim of assessing such an ESG dataset is of finding the relevance of novel ESG topics. It shows the trend over time, how many stories have been published regarding these topics, and the topic sentiment over time.

3.4. Assessing the procedure

As the effects of the introduction of such a procedure are of interest to both the researcher and the firm alike, the procedure will be assessed throughout meetings with internal stakeholders. Interviews are set up with people from within the firm. Interviewees are those who have used the procedure, those who have experience with such software, those who have experience with ESG, those who have an interest in the results, and a combination of the aforementioned factors.

3.5. Interview set-up

Interviews are set up to collect data regarding the case study and to assess certain themes found in the aforementioned steps. These themes found in the observations are directly and indirectly assessed through the interview questions.

The interview questions are created through the lens of the semi-structured method of conducting interviews. As the main goal of the interview is to explore the perception of the interviewee regarding Big Data and AI, the procedure, and the integration, perception, and operability within the firm, and only one interview is held, a semi-structured method is preferred (Barriball & While, 1994).

This semi-structured interview involves several predetermined questions. These questions again are based on the aforementioned themes found. The semi-structured approach gives the interviewer the ability to defer from the predetermined questions to seek clarification regarding answers given by the interviewee (Doody & Noonan, 2013).

3.5.1. Interview recording

The recording of the interviews was done through a hand-held recording device. The initial goal was to conduct the interviews face to face, however, due to COVID-19 measures, this was not possible. Therefore, the interviews were held through Microsoft Teams. Initial permission was sought before the interview was conducted. This was done through the informed consent form, and by verbally asking the interviewee before the interview if it was allowed to record the audio of the interview. The use of audio recordings ensures the replication and reliability of the transcription and reduces potential errors within the results gained through these interviews (Barriball & While, 1994).

3.5.2. Interviewees criteria

The interviewees are categorised into their roles within the firm and their experience. Multiple different kinds of backgrounds were sought to give a broader overview while adhering to the criteria set beforehand. To improve the willingness of interviewees to participate the main language used was Dutch, as all interviewees were of Dutch origin. Furthermore, the transcriptions of the interviews are fully anonymised.

Interviewees approached for the interviews adhere to the following criteria. The interviewee

- must be available within the timeframe
- has a role within the firm where the case study takes place
- has one or a combination of the following factors:
    - a stake in In the design process, the rollout of the procedure, or results of the procedure as used in the case study
    - experience with ESG
    - experience with Big Data and AI, with a preference in NLP

3.5.3. Characteristics of the interviewees

A total of 3 interviews were conducted. The following table illustrates their characteristics.

#	Role within the firm	Work experience	Experience
1	Programme lead SFR	6+ years as an IT product owner, it project leader.	Has in-depth knowledge of the implementation of IT solutions, especially in Big Data and AI.
2	Sustainability Reporting Specialist	2+ years of working with ESG, 3+ years of ESG advisory	Has in-depth knowledge of ESG, ESG reporting, and portraying ESG data.
3	Program manager strategy & sustainability	20+ years of experience in sustainability, IT management, and management consulting	Has in-depth knowledge about ESG, ESG data, and the implementation of IT solutions.

Table 2 List of interviewees

Chapter 4. Results

This chapter illustrates the case study results. The results are themes found throughout the case study, categorised under six categories. For each stage of the case study, as mentioned in the previous chapter, themes were found through thematic coding. These themes will be used to substantiate, debunk, or construct novel propositions. A total of 43 themes were observed and divided among six amount of categories.

As stated within the Methodology chapter, the categorization within the evidence register is the first step to be illustrated. The following categories have been identified by which the themes found were categorized.

Coding - Categories	**Explanation**
Current situation	Within the 'Current situation' category themes are included that play into the current situation of the firm. Meaning that this category is being used to depict the pre-procedure situation of the firm. The results illustrate the situation before the application of the technology. The aim of illustrating these themes within the results is to compare the 'old situation' with the 'new situation'.
(Big) Data	This category depicts everything related and tangential to data. As there are a number of Big Data datasets included, observations are made and therefore themes are coded regarding data. Thus, every theme data related will be categorized within this category.
ESG	As ESG data and the aspects of this kind of specific data is assessed, all themes observed which are related to ESG and ESG data are categorized here.
Learning & adoption	Learning and adoption is a category consisting of two topics, learning, and adoption. This is as learning regarding new technologies and adopting technologies are simultaneous processes that are effectively intertwined. Observed themes regarding the process of integrating a new process using ESG data, Big Data and AI are included in this category.
New situation	Within this category, themes are categorized which provide contrast for the 'Current situation' theme. Here, themes are illustrating observations made while the procedure was introduced and what the output of it was.
Perception & social	The perception and social aspects are categorized under this theme, with a similar rationale as 'Learning & adoption'. The perception and social are aspects are intertwined with each other, and both are therefore assessed under the same category. Here, all observations regarding how people perceive Big Data and AI, and observations regarding how they react to it are categorized.

Table 3 Coding - Categories and Explanation

The results have been structured as follows. Each episode of gathering data, Results of indexing current methodology, Results of defining and introducing procedure, and Results of assessing the procedure, are addressed accordingly. This is as the outcome of the results within the same theme differs in some aspects. Meaning that when compared, an observation might show a different background categorised under the same theme. A theme might come forward during the expert work meetings, where it is discussed by experts, or during an observation by the researcher himself, or during an interview. Three different ways of leading to a similar theme. These are all coded under the same theme, as it illustrates the same sort of observation, concept, idea, or notion, while the data source and means of observation differ. Furthermore, this distinction is done to provide a more holistic overview regarding themes found and to support the triangulation approach taken within the discussion.

Within each sub-chapter, a table structures the results as follows; first, the category will be addressed, after which the theme is assessed, and a short summary of what the observation entails and where it comes from. Each theme, as reported within the results of each observation type, is further elaborated on in Appendix B – Clarification results. Within this appendix, further elaboration is given regarding the evidence and scope of each theme.

4.1. Results of indexing current methodology

The first step to introduce and replace a process with a procedure based on Big Data and AI is to assess how the current process is defined. A total of 9 expert work meetings have been held assessing the current situation and discussing the procedure.

Categories	**Themes**	**Summary of observation**
Current situation	Expert opinion is prevalent/used	It was discussed by experts that within the current situation, ESG data is currently assessed through expert opinion.
Current situation	Expert opinion has bias	It was discussed with experts how expert opinion in the processes, as mentioned in 'Expert opinion is prevalent/used', always has bias.
(Big) Data	Importance of data sources	It was discussed with experts that sources of ESG data are important. Data sources have to be ESG & firm relevant. The sources themselves have to be assessed regarding their applicability within the Big Data dataset.
(Big) Data	Importance of data quality	It was discussed with experts regarding ESG data quality when creating the ESG dataset, one has to 'define the size and relevancy of the ESG topic'. Addressing and further discussing the quality of the data input, regarding the variety and quality of the data itself.
(Big) Data	Importance of how ESG data is formatted	It was discussed with experts that while Big Data is often unstructured and in great quantities, the importance of how data is formatted is of importance. The focus of this theme is regarding ESG data as an input and output for the procedure. Both have to be accessible and usable, which is a prerequisite for the usability of the procedure.
(Big) Data	ESG data origin	The origin of ESG data was discussed with experts. It was discussed that the origin of ESG data could have been

		initially biased. It illustrates that the origin of data matters, which was confirmed by a meeting assessing the potential of ESG data and the output.
(Big) Data	Expert opinion as source of data	It was discussed by experts that ESG related risks are sourced in the opinion and expertise of experts. Thus, expert opinion regarding ESG data is currently used in processes as a source of data.
(Big) Data	Data pre-processing	Experts state regarding ESG data and analysing processes; "garbage in, garbage out". Meaning that data should be pre-processed before being used in a procedure.
(Big) Data	Time focus of data	It was discussed with experts that ESG data has a time element in it. The time relevance of a process using ESG data is prevalent. Illustrating that certain data points within ESG data should have a 'timestamp' to be relevant and usable.
ESG	ESG Data is culturally bound	It was argued by an expert that ESG data is culturally bound. Meaning that different cultures value different ESG aspects more, thus biasing ESG data coming from the respective entity based in that culture. An example given was that an ESG topic such as deforestation was of more importance to the inhabitants of a country with native forests than one without.
New situation	Big Data and AI to reduce manual labour	It was discussed by experts that they had the perception that Big Data and AI could be used to reduce manual labour.
New situation	Big Data and AI to define speed of trends	It was discussed with an expert that a procedure using Big Data and AI could be used to define the speed of ESG trends.
Perception & Social	ESG is shown as business risk data	Experts within the firm state that the current perspective on ESG is that it is shown as business risk data. It is used to mitigate risk within (financial) processes. Illustrating the impact such data has "further down the line".
Perception & Social	Big Data and AI as nonbiased view	It was discussed with experts that the procedure aims to use Big Data and AI to generate a nonbiased view on ESG data. This theme illustrates the perception within the firm that such a software solution can provide such a nonbiased view.
Perception & Social	Big Data and AI as a spearpoint to push information	An expert discussion was observed that the information regarding ESG subjects is usually disputed. It was discussed that Big Data and AI can help to substantiate ESG information. Illustrating the technology being used to spearpoint "data-driven information" to stakeholders within the firm.

Perception & Social	Big Data and AI just for convincing people	An expert discussed that the use of Big Data and AI regarding ESG data is only to convince other stakeholders.
Perception & Social	The approach of reporting on the Big Data and AI tool	It was discussed by experts that the communications approach is of importance. By including Big Data and AI in reporting on ESG subjects, the data-driven approach could provide legitimacy to the published communications.
Perception & Social	People looking only at method, not data quality and source	This was discussed with experts, where the focus of the end-user lies. It was discussed that their main focus was the method of the procedure.
Perception & Social	Not invented/purchased here syndrome	The "not invented here syndrome" was discussed by experts regarding the adoption of Big Data and AI technology within the firm. This illustrates the tendency to avoid (technical) solutions not invented within the firm.
Perception & Social	AI is received with scepticism	An expert perceived the AI, NLP part, of the provided software with scepticism. This, a significant amount of human interaction was still needed within the procedure. Stating that it currently is not that useful, but it could be in the future.
Perception & Social	Results of Big Data and AI, not for strategic choices	An expert stated how this combination of ESG data, Big Data and AI can be used. As it was perceived as backwards-looking due to the provided information it could not be used for strategic choices.
Perception & Social	Results of Big Data and AI, for tactical choices (timing and handling)	An expert stated how this combination of ESG data, Big Data and AI can be used. It was discussed that the procedure and indirectly the technology could be used for tactical choices. The timing factor of ESG data plays a role in these tactical choices.
Perception & Social	Even in a Big Data and AI process, expert opinion is needed	It was discussed by experts that within the procedure, expert opinion is needed. As the procedure has to be reviewed and assessed. Thus, someone who has knowledge and brings their own opinions and bias into the process.

Table 4 Results of indexing current methodology

Highlights of these results are the following two themes, both in the Perception & Social category.

'Big Data and AI as nonbiased view'

It was discussed with experts that the procedure aims to use Big Data and AI to generate a nonbiased view on ESG data. This theme illustrates the perception within the firm that such a software solution can provide such a nonbiased view.

'Even in a Big Data and AI process, expert opinion is needed'

It was discussed by experts that within the procedure, expert opinion is needed. As the procedure has to be reviewed and assessed. Thus, someone who has knowledge and brings their own opinions and bias into the process.

4.2. Results of defining and introducing the procedure

Within this part, the observations are illustrated which were done during the process of defining and introducing the procedure of assessing ESG data through a Big Data & AI NLP tool. A total of 10 field notes each depicting a week of observations have been included. The following themes were discussed.

Categories	Themes	Summary of observation
Current situation	Expert opinion is prevalent/used	It was observed that the source of data is rooted in expert opinion, regarding the current use and application of ESG data.
Current situation	Expert opinion has bias	It was observed that expert opinion, as mentioned in the 'Expert opinion is prevalent/used' theme, is used. It was observed that every expert brings their own experience and opinion, thus bias, to the current process.
(Big) Data	Importance of data sources	It was observed that an important source of discussion is the source of the data by which the ESG datasets are analysed. It illustrates the importance of the source, relevance of the source, and sources used which could affect the potential outcome of the procedure.
(Big) Data	Importance of data quality	The importance of ESG data quality was observed during discussions. It was discussed with a software provider that the Big Data part as provided is unstructured and of 'certain quality'. Illustrating the in- and output of the analysis should be checked regarding the data quality and that one of the success criteria as stated by the firm.
(Big) Data	Importance of how ESG data is formatted	It was observed that regarding the procedure, ESG data shows ESG topics and their relevancy. This is used to scope the Big Data dataset regarding ESG data. Thus, the output of the procedure relies on the input metrics, ESG data. Thus, the way ESG data is formatted is part of the internal validity of the analysis, validating that one analysis what one wants to analyse.

(Big) Data	ESG data origin	It was observed that the origin of ESG data is of importance, as this was discussed often during meetings regarding the design of the procedure, with the focus on the in- and output.
(Big) Data	Expert opinion as source of data	The notion that expert opinion is used as a source for data was observed when discussing data sources with a supplier. This as their provided processes applying Big Data & AI is based on the 'expert opinion' of the company. As that company defined the way the AI was trained, what sources such as Twitter, news websites, and company reports, were included in their Big Data solutions, and how their taxonomy by which the analysis is done. As there is no freedom within this solution to define one's own process, one has to rely on the company's expert opinion of these data sources.
(Big) Data	(No) difference between the ESG aspects of datasets and other datasets	During the analysis of how data is interpreted through NLP, it was observed that little to no difference between other kinds of information and ESG related information. This was also observed within the Big Data and AI software used within the procedure, as ESG data could be just like any other dataset that has the potential to be analysed through their system.
ESG	ESG Data is culturally bound	The cultural relevance of ESG data came to light during the design and discussion of the procedure, as ESG data could have cultural geographic boundaries. This was observed by applying ESG data and discussing this application with experts.
ESG	Using ESG data as an umbrella term for all data	This was observed during a meeting with an ESG data and software supplier. It was observed that they used ESG data as an umbrella term to promote their way of using data. Their approach is different from the one taken in the procedure.
Learning & adoption	Knowledge of capabilities of AI and Big Data differs	It was observed that different parties, such as IT software providers, employees, the provider of the Big Data and AI toolset used within the procedure, and the employees potentially using the procedure all had a different knowledge regarding the capabilities of Big Data and AI.
Learning & adoption	Usability of the software tool	It was observed that the usability of the software as used in the procedure is an essential part of the procedure.
Learning & adoption	Knowledge of Big Data and AI is lacking	While discussing the requirements for the procedure, it was observed that that Big Data and AI was perceived as an abstract concept to analyse data with. This came to be due to the lack of knowledge regarding Big Data and AI.

Learning & adoption	Perceived barriers to use the technology	The perceived barriers to using the technology have been observed in context with the adoption of data analytics technology as used within the firm. It was observed that there is a need for learning as seen in the previous theme 'Knowledge of Big Data and AI is lacking'. It was observed that this theme was partly caused by barriers perceived by users of such technologies. Thus, not knowing the technology was perceived as a barrier to adoption.
Learning & adoption	Learning and adoption perspectives	This theme entails the observation of a willingness to learn regarding ESG Data, Big Data and AI as used in the procedure. It was observed that this learning might be correlated to the notion of adoption of the previous theme 'Perceived barriers to use the technology'.
New situation	Big Data and AI to reduce manual labour	It is observed that this is the core reason for the introduction of Big Data and AI regarding the analysis of ESG data as done within the procedure.
New situation	Big Data and AI to define speed of trends	It was observed that this is a criterion of what the procedure should entail. Defining the speed of trends, as observed, is one of the capabilities of the software & data as structured and used.
New situation	Data driven approach	This theme was observed within the same observation as 'Big Data and AI to define speed of trends'. It was observed that a data-driven approach was perceived as the essence of the procedure.
New situation	Review correct output software (AI)	This theme was prevalent during the design and review of the procedure. The assessment had to be made if the NLP software was assessing the right kind of words, in the right context. Furthermore, the sentiment of these 'hits' within the NLP software has to be assessed. This, as the software, has to be 'trained' to give the correct output.
Perception & Social	ESG is shown as business risk data	This was shown during a discussion with experts regarding the capability of the procedure and its related output. The output of such analysis regarding ESG topics could be used within other departments of the firm, as interest was shown in the outcome due to it also being a business risk.
Perception & Social	Big Data and AI as nonbiased view	During the assessment of the Big Data and AI software, it was shown that using this technology to create a nonbiased view on ESG data was one of the aims of the procedure.
Perception & Social	Big Data and AI as a spearpoint to push information	It was observed that the notions of Big Data and AI are used to spearpoint the sale of software packages. Within this observation, it was perceived that the actual Big Data and AI aspects were negligible, thus, illustrating the usage of this kind of terminology to spearpoint information.

Perception & Social	Big Data and AI just for convincing people	It was observed that an external party used the terms Big Data and AI with the sole intent of convincing people.
Perception & Social	The approach of reporting on the Big Data and AI tool	Reporting regarding the procedure is essential. It was observed that other reports were addressing these technologies. Similarities were made to other procedures where it was stated that a 'state of the art' tool was used, referring to a similar Big Data and AI software used.
Perception & Social	People looking only at method, not data quality and source	It was observed that people look mostly at the method as used within the procedure. That the inclusion of Big Data and AI regarding ESG brings a certain "gravitas" to the discussion.
Perception & Social	AI is received with scepticism	As stated by an expert: "The AI of today will not be classified as what is AI in two years." Here, some scepticism was observed regarding how AI is received and perceived within the firm.
Perception & Social	Results of Big Data and AI, for tactical choices (timing and handling)	This was observed as an essential part of the procedure. This theme shows the main goal of the output of the procedure and how the output data is being applied.
Perception & Social	Big Data and AI are used as buzzwords	It was observed that other parties used Big Data and AI as buzzwords. This was shown in the way the used software within the procedure portrays itself, mostly through buzzwords.
Perception & Social	Perception of Big Data and AI is that it is near mystical	It was observed that the perception of Big Data and AI was of something "unknown". As something near-mystical to someone who does not know what it entails.
Perception & Social	The perception of Big Data & AI is 'state of the art'	It was observed that the perception of Big Data and AI is "state of the art", novel.
Perception & Social	Company culture	This observation came from the notion that the acceptance of software could be assessed through the lens of company culture. It was observed that the tendency of the discussion by experts was that company culture directly affects the adoption of new technology.
Perception & Social	Overwhelmed due to hype	This was observed through a discussion with experts. Novel technologies, e.g. cloud computing, AI, could sometimes overwhelm people and push them into a dichotomous camp, accepting or rejecting, without even properly assessing the technology.

Perception & Social	Generational aspects of acceptance of Big Data and AI	It was discussed that each generation of people has a different approach to technology. Meaning that the adoption rate could also be correlated to the age of the people who are the decision-makers within the firm.
Perception & Social	Big Data and AI used as a black box	It was observed that Big Data and AI was perceived as something where data goes in and results come out. Here, the black box is designed by a person, and people put data in and collect results as output, illustrating the perception that it is something where the actual method might be of lesser importance. Otherwise, the perception was not something as observed.

Table 5 Results of defining and introducing procedure

Highlights of these results are the following two themes, the first comes from the Learning & adoption category and the second comes from the Perception & Social category.

Perceived barriers to use the technology

The perceived barriers to using the technology have been observed once. This is in context with the adoption of such technologies within the firm. It was observed that there is a need for learning as seen in the previous theme. It was observed that this theme was also partly caused by barriers perceived by users of such technologies.

Big Data and AI used as a black box

It was observed that Big Data and AI was perceived as something where data goes in and results come out. Here, the black box is designed by a person, and people put data in and collect results as output, illustrating the perception that it is something where the actual method might be of lesser importance. Otherwise, the perception was not something as observed.

4.3. Results of assessing the procedure

The following themes came across the assessment of the procedure and Big Data and AI tool used. Three interviews were conducted, according to the method as stated within the Methodology chapter. A total of 29 themes were discussed throughout the interviews assessing the procedure, categorised under six categories.

Categories	Themes	Summary of observation
Current situation	Expert opinion is prevalent/used	It relates to the way currently expert input is used regarding the way ESG data is gathered and formatted.
Current situation	Expert opinion has bias	The bias aspects within expert opinion related to the design and operations of the procedure have been discussed by experts.
(Big) Data	Importance of data sources	The importance of data sources was addressed by the interviewees through addressing success criteria for the procedure. The common denominator was that one has to

		think of data input, thus data sources, when assessing the output.
(Big) Data	Importance of data quality	Illustrates the discussion by experts regarding data quality through the lens of ESG data. One of the things addressed was the amount of data is of more importance, mitigating the reduced quality.
(Big) Data	ESG data origin	The origin of ESG data was discussed in the context of ESG rating agencies such as DJSI and GRI, whose requirements do not change very often. The reflection regarding the procedure was that such input was good to take into account.
(Big) Data	Expert opinion as source of data	This theme was discussed by experts in the context of normal data input and ESG data input. It was discussed that the source of, for example, ESG data is down the line based on expert opinion, what is taken into account and what is not.
(Big) Data	(No) difference between the ESG aspects of datasets and other datasets	The observed results to this theme differ, however one point of view illustrates that regarding the input of such a procedure, there is nearly no difference between ESG data and other kinds of data. Illustrating that the method, thus AI aspects and related Big Data data sets through which the analysis is done, do not care about the data input. However, another expert stated that ESG data is bound to future regulation and reporting directives. Illustrating the different perspectives one could take regarding ESG data.
Learning & adoption	Knowledge of capabilities of AI and Big Data differs	It was observed that there is a knowledge gap in the knowledge of applying Big Data and AI. This has been observed by talking to different experts stating different experiences with similar systems.
Learning & adoption	Usability of the software tool	The usability of the software tool was discussed with experts. This was regarding the application of the tool and how results are portrayed.
Learning & adoption	Learning and adoption perspectives	Experts agree that there is a link between knowledge sharing, how technology is being presented, and the adoption of such a technology.
New situation	Big Data and AI to reduce manual labour	As discussed by experts, such technologies could have the capability of making one's life easier. Stating that "the use of big data is the future, we have to go there, we can't do this (the data retrieval process) manually".
New situation	Data driven approach	It was discussed by experts that a data-driven approach might be used to convince people, that you, as an expert, are not an expert unless you can provide a solid background

		of information. Thus, that a data-driven approach is a way to go to present results besides one's expertise.
Perception & Social	Big Data and AI as nonbiased view	This was discussed by multiple experts. The common denominator within these interviews is that Big Data and AI are software solutions, and part of a process. These software solutions and the process as a whole might include bias through the designing of a process.
Perception & Social	Big Data and AI as a spearpoint to push information	As stated by one expert, it is used as "we see it as a stick to hit the dog, we are already planning to hit the dog, but now we got a stick. It is necessary to hit the dog". This is interpreted in two ways. That it can be used as a spearpoint to push information, like Big Data and AI can create a lot of useful information, and it also has a convincing factor.
Perception & Social	Big Data and AI just for convincing people	As discussed by an expert, it is an approach to shake people up with information. If one uses Big Data and AI, the notion of Big Data and AI alone can convince people that the information provided by such a methodology using Big Data and AI is correct.
Perception & Social	The approach of reporting on the Big Data and AI tool	As stated by an expert: "I've experienced over the years is that it's extremely big, but you are restricted by the people that surround it, you see it also here, you get stuck on definitions, how definitions are formatted, instead of the data definition. The data is not the problem, its available, but how you use it and how you do it, and how to instruct people to use it is difficult". Thus, showing that how one reports is of importance regarding such technologies.
Perception & Social	People looking only at method, not data quality and source	This theme was addressed by multiple experts. The common denominator was the "Garbage in garbage out" principle. Furthermore, one interviewee stated that if one of the two is garbage, you cannot get a correct answer. When reflected on the firm the data quality is of more importance, however, another expert stated that the method is more important.
Perception & Social	Not invented/purchased here syndrome	It was discussed by experts that people only like things in which they had a say. The experts discussed that people within the firm are more likely to reject software and software solutions if it was not invented here.
Perception & Social	AI is received with scepticism	It was observed that some experts on the "receiving end" of the procedure and regarding the design perceived the AI aspect with scepticism.
Perception & Social	Even in a Big Data and AI process, expert opinion is needed	This was discussed with experts, as every process, in the end, is designed by people, putting their bias into the process.

Perception & Social	Big Data and AI are used as buzzwords	It was discussed with experts that these terms were used as buzzwords for convincing people.
Perception & Social	Perception of Big Data and AI is that it is near mystical	It was observed that Big Data and AI are perceived as near mystical, as no further details were given and it was still implemented.
Perception & Social	The perception of Big Data & AI is 'state of the art'	It was discussed by experts that it could be perceived as something new and impressive, state of the art so do speak. It was perceived as something that could reduce the time worked on some tasks.
Perception & Social	Company culture	As one expert stated "Culture is everything!", a strong company culture affects the adoption of technologies such as Big Data and AI.
Perception & Social	Generational aspects of acceptance of Big Data and AI	This theme entails more general aspects of the acceptance of Big Data and AI not illustrated by any specific theme. An example of this is that when a firm uses external sources for such a process of Big Data and AI, it is generally more accepted. Or the conservativeness of the average employee in a financial institution, that clashes with the introduction of technologies such as Big Data and AI.
Perception & Social	Combination of business implementation and technology	It aims to illustrate the business implementation of technology and aims to catch relevant information within this theme. An observation was that this kind of software can be used for predictions. Furthermore, it could be used as an addition to the in-house data a financial institution already has. Seeing this as an addition of AI as an analysis method. Another observation regarding this theme is that it could be implemented through current other ESG reporting methods such as GRI and SASB.
Perception & Social	Big Data and AI is perceived as (not) valuable	This theme aims to illustrate the perception of Big Data and AI as something (not) valuable. Experts had a different opinion on the software as used within the procedure.
(Big) Data	ESG data as a precursor for financial data	An expert stated that there might be a correlation between ESG data and financial data, as it could be a precursor to financial data.
Perception & Social	Future of the combination ESG, Big Data, and AI	This theme discusses the observations regarding the future of the combination of ESG data, Big Data and AI and illustrates different backgrounds, such as upcoming regulations, how they could be implemented within the firm, and how external data such as ESG data could be applied.
ESG	ESG Data is culturally bound	The observation was made by an expert that within a similar process, input data was curated by a company of foreign origin to the firm. Thus, the input data was according to their

		standards. This firm gave its cultural relevance to the initial assessment of the data categorisation.
ESG	Using ESG data as an umbrella term for all data	It was discussed by experts how and what ESG data might portray. They stated that sometimes input data could be interchangeable, and therefore ESG data could be an umbrella term.

Table 6 Results of assessing the procedure

Highlights of these results are the following two themes, the first one comes from the category Perception and social, the second theme comes from the (Big) Data category.

Big Data and AI as a spearpoint to push information

This theme has been observed three times. One of the best analogies given to how such technologies are currently used is given by one interviewee. It is used as "we see it as a stick to hit the dog, we are already planning to hit the dog, but now we got a stick. It is necessary to hit the dog". This is interpreted in two ways. That it can be used as a spearpoint to push information, as Big Data and AI can create a lot of useful information, while it also has a convincing factor. Meaning that it can be used to convince people the provided information is correct due to the method of using Big Data and AI. Furthermore, it has been addressed during one interview that even as one pushes information such as management information, not substantiated in Big Data and AI, this information will be challenged.

(No) difference between the ESG aspects of datasets and other datasets

This has been discussed within every interview. The results to this theme differ, however one point of view illustrates that regarding the input of such a procedure, there is nearly no difference between ESG data and other kinds of data. Illustrating that the method, thus AI aspects and related Big Data data sets through which the analysis is done, do not care about the data input. However, another interviewee stated that ESG data is bound to future regulation and reporting directives. Illustrating the different perspectives one could take regarding ESG data. Furthermore, during one interview a point of view was taken of current coding of sustainability data within the firm. There are certain specifics to it. Due to the regulations, a bank has to adhere to, this coding of certain ESG data has to adhere to this (future) regulation imposed by the EU Commission regarding ESG reporting.

Chapter 5. Discussion

This chapter will discuss the research findings. It will assess the results of the exploration of the research topic and these results will be reflected against the initial propositions, as stated in the Methodology chapter, respectively. This assessment of findings will include a personal and literary reflection to provide a substantiated discussion regarding the 'could and could not be' when looking at the results. Providing propositions and rivalling theories, to provide a basis for answering the main research questions.

5.1. Discussion of findings

Within the discussion of findings, the results of the previous chapter will be discussed according to each category found within the data. The discussion done will be by referring to the themes found. The clarification of the results can be found in Appendix B – Clarification results. The evidence register, the location where these themes are stored and where they come from, can be found in Appendix A – Results.

5.1.1. Learning and adoption

The learning and adoption category focuses on the way the software is implemented. Within this category, there have been several themes observed ranging from different kinds of data inputs. This category cumulated into the following novel theoretical propositions and their related discussions.

Adoption through the lens of learning and perception

The notion of learning and the approach to adoption have been prevalent throughout the case but was not initially expected. It, however, is a tangential topic when compared to the second proposition as mentioned in the methodology, "It, however, is a very big stick and has the required effect, showing that the social dimension of implementing such abstract tool- and datasets should not be overlooked.". The social dimension of implementing Big Data and AI can be found in the learning and adoption aspects. The proposition to be built here is one of the correlations between Big Data and AI, perception, and learning rate and adaptability. The following observations have been made supporting this proposition.

Within the case it was observed that the party providing the Big Data and AI toolset emphasized their "ease of use" of the software, fully supporting their customer to adopt this novel technology. They aimed to mitigate the perceived barriers to using the software. These barriers have also been observed. After the introduction to the software, learning gave the user an understanding of the software, and their perception has been changed. These came forward through the following themes:

- **'Knowledge of Big Data and AI is lacking'**, where it was observed during the design and introduction of the procedure that knowledge was lacking within the group of employees to be using this software. Here, it was observed that the technology was perceived as something unknown, as no knowledge was available.
- **'Perceived barriers to use the software'**, which was observed during the design of the procedure in the context of the adoption of these technologies within the firm. It was observed during the design and introduction of the procedure that there is a need for learning, and not knowing the technology was perceived as a barrier to adoption.
- **'Learning and adoption perspectives'**, which was observed during the interviews with experts during the assessment of the procedure phase. The experts agree that there is a link between knowledge sharing, how technology is being presented, and the adoption of such a technology. These experts stated that the size of this procedure (which uses the combination of Big Data, AI, and ESG data) brings different challenges with it regarding learning approach and adoption. The "adoption is bigger, it's getting more complex". If more people are involved

and the procedures to be changed are bigger, so does the complexity of how such technologies should be taught, affecting the learning rate.

By observing the aforementioned themes, it raised the following question, could be a link between the lack of knowledge regarding technologies such as Big Data and AI, and the acceptance rate of such technologies?

This question has also been observed before in the case of Avgar, Tambe, and Hitt (2018). A comparable case where the adoption of an IT system was observed. The similarity is found in the adoption and introduction of a novel IT technology within a firm, of which the employees of the firm were previously unaware. One interesting result of this case was that Avgar, Tambe, and Hitt (2018) concluded that discretion as a factor was more important than training. Meaning that users required a discrete approach to learn, experiment, and adapt the novel technology system. Within the case study of this thesis, however, the opposite has been observed in the following theme:

- **'Big Data and AI just for convincing people'**, which is discussed by experts in meetings, observed during the design and introduction of the procedure, and during the assessment phase. It was observed to "shake people up" with information, illustrating the lack of discretion surrounding the application of the topic.

The aforementioned theme illustrates that there is no discretion observed regarding any learning processes. Discretion is the opposite of boasting with technology to convince people, as to be mentioned in the upcoming sub-chapter The convincing power of Big Data and AI within the realm of ESG analytics. There could be a correlation between the perception of complex technologies and their learning approach, as this pattern has been observed. While the initial proposition could not be confirmed or rejected, the observations made converged into the following proposition:

- The perception within a firm of using Big Data and AI within a process could affect the learning rate and the learning approach taken by the user. This affects the acceptance of the technology. Thus, the perception could affect the adoption rate of Big Data and AI within a firm.

This proposition is illustrated in the following causal diagram.

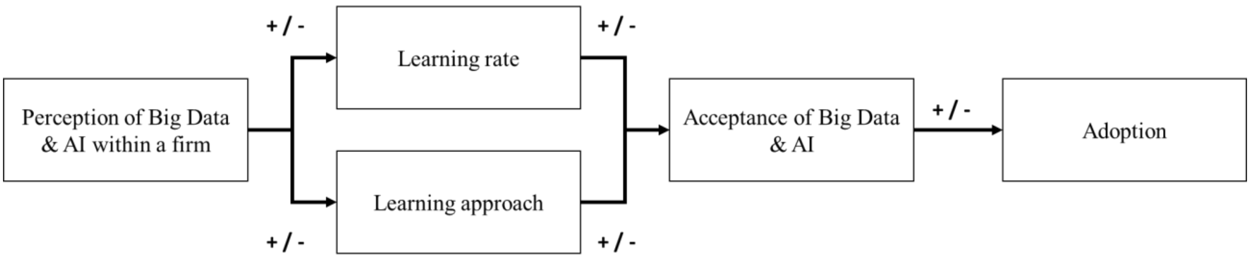

Figure 6 Diagram illustrating a potential causal effect of perception

This causal diagram is created through the aforementioned proposition. It is observed within the case study that the way Big Data and AI within a firm is perceived could affect, positively or negatively, the learning rate and the learning approach. This came forward during the case study in themes such as 'Knowledge of Big Data and AI is lacking', 'Perceived barriers to use the software', and 'Learning and adoption perspectives'. It could furthermore indirectly be seen in the theme 'Knowledge of Big Data and AI is lacking'. The learning rate and the learning approach could positively affect the acceptance of Big Data and AI. Meaning that the knowledge within the firm is influenced by the perception of these technologies, which again could affect the acceptance of these technologies. The acceptance, in the end, could positively or negatively affect the adoption of these technologies.

Reflecting on this causal diagram, one could see the potential future implications. If a firm wants to successfully implement Big Data and AI, one should first address the perception of these technologies. The notion of "not invented here" mostly comes to mind, which was observed within the theme of "Not invented/purchased here syndrome". Then, as such technologies are implemented, the learning approach and rate should be addressed simultaneously, as they could form an important factor regarding the acceptance of Big Data and AI if this proposition holds true. There could be other factors influencing the learning rate and approach, and even the acceptance of Big Data and AI. Furthermore, the author speculates that there are factors leading up to the perception of Big Data and AI within a firm. As the perception of such technologies come from something and are usually not created out of thin air. This, however, has not been observed within the case study. It could provide an interesting base for future research.

Rivalling theories could be considered as follows: First, the discretion factor Avgar, Tambe, and Hitt (2018) could perhaps also be found within this case, as the first user to adopt this technology has the opportunity to discreetly learn, experiment, and adapt this technology. This is a part of corporate culture and the firms approach to corporate knowledge management. A firm could create a culture of fostering novel technologies. One example is given by Sánchez, Sánchez, Collado-Ruiz, and Cebrián-Tarrasóna (2013), who proposed a framework approach to openly create and share knowledge within a company through processes. Stating that by lowering these barriers and making these processes explicit, productivity is increased but mainly knowledge creation and sharing culture and innovation are increased. This last part could counter the proposition, as it is not the technology that creates barriers but corporate culture is. This could be found in the theme 'Knowledge of capabilities of AI and Big Data differs', which, however, is not observed that often. Therefore, the previously stated proposition leans the other way.

Another rivalling theory could be regarding technological aspects such as the user interface, which could be of importance as it could affect the adaptability and perception of the technology. This could be supported by the theme 'Usability of the software tool', which was sparsely observed. It is not often required for the user of such technology to fully understand the technology itself. After all, who nowadays looks under the hood of their car and fully comprehends all technological aspects that make it drive. This could mitigate the perception of technology such as Big Data and AI as complex and could build on the previously mentioned rivalling theory.

In the end, the question still remains how inherent this proposition is to just a Big Data and AI toolset. While there was literature available from Sam & Chatwin (2018), who stated that the attitude toward Big Data analytics is an influencing factor regarding Big Data readiness. Perhaps their conclusion holds true when, within the respective technological system, AI aspects are added to substantiate Big Data analytics.

The concept of Big Data and AI as a black box

The continuous notion of Big Data and AI being a black box could be inherent to the knowledge available within the firm about these technologies. When we look at the themes observed, one major one was:

- **'Big Data and AI used as a black box'**, observed during the defining and introducing phase, through interaction with relevant stakeholders. It was observed that Big Data and AI was perceived as something where data goes in and results come out. Here, the black box is designed by a person, and people put data in and collect results as output, illustrating the perception that it is something where the actual method might be of lesser importance.

This theme was categorized in the Perception and Social category. This, however, does not fully explain the link between the perception of it being a black box and learning and adoption as a factor. What has been observed is the notion of it being a black box within the following theme:

- **'Knowledge of Big Data and AI is lacking'**, observed during the defining and introducing phase, through interaction a discussion with experts. The observation was made that there was the perception that Big Data and AI was a sort of black box to analyse stuff with. That the procedure analyses ESG data according to a big stream of data and AI, thus it creates a nice overview of information. Where the experts perceived it as showing information in, information out, and not how it's done. And that the actual knowledge regarding the software is lacking.

This theme was observed during the Designing and Introducing phase. One could argue that when knowledge is lacking, one perceives such technology as a black box with data coming in and results coming out. Thus, to mitigate this perception of it being a black box, knowledge has to be shared regarding this technology. When knowledge is shared, the adoption rate goes up. This is a near-open-door, relating well to the point made within the rivalling theories section of the previous sub-chapter. However, conflicts again with the allegory made of 'one not looking underneath the bonnet of one's car'. Thus, perhaps a certain level of knowledge is required to accept technologies perceived as inherently complex.

This aforementioned statement could open up a whole level of philosophical discussion within the realm of AI. It is common ground within AI that the neural network, the way AI 'learns', will remain a black box. Thus, a certain boundary is affirmed within this notion of one's full comprehension of what the respective AI software entails. How can one, for example, accept an outcome of which one does not fully know how it has been formed. Furthermore, this acceptance is also based on the person that accepts such an outcome, as a less critical or less knowledgeable person could accept the use and outcome of AI without serious critical thought. This, for example, was observed within the theme 'Generational aspects of acceptance of Big Data and AI', within the 'Assessing the procedure' phase. Where the interviewees discussed general aspects such as that most people don't care about the process, as long as it delivers results.

Thus, one could state that the notion of AI being a black box within the application of Big Data and AI is inherent to the technology. Many papers have been published regarding this notion of AI and its black box (Pedreschi, et al., 2019; Adadi & Berrada, 2018; Asatiani, et al., 2020). However, a rivalling perspective should be considered to give perspective regarding the future of AI. One perspective is the allegory of turning the 'black box' into a 'glass box' (Rai, 2020). Within this paper, Rai (2020) argues for a class of systems called Explainable AI (XAI), where one can take a look under the bonnet of the metaphorical AI car. Stating that this process makes the AI more trustworthy, as XAI provides "the rationale for the decision-making process, surfaces the strengths and weaknesses of the process, and provides a sense of how the system will behave in the future" (Rai, 2020).

This notion of available knowledge regarding AI, the learning aspects, and AI is perceived as a black box opens up a philosophical debate to what is inherent to AI, as one could mitigate the 'black box' by transforming it to XAI. It could be debated to what extend it is 'intelligent', as one could ask to what extent such XAI could sufficiently provide that rationale behind its decision-making process and future behaviour. It, in the end, all depends on what the user finds an acceptable rationale of explanation of the AI, which again is inherent to the users learned knowledge of the respective AI system.

However, to reflect the aforementioned discussion onto the main research question is somewhat futile. The observations made within the case study merely opened up this discussion, not providing any substantiated themes. It is unfortunate that no more observations were made within this direction to substantiate a proper novel theoretical proposition. However, as (X)AI is in continuous development, reassessing this subject in the near future with a focus on the adoption through the lens of available knowledge and the accessibility of this knowledge (meaning AI not being a black box anymore), could provide novel propositions.

5.1.2. Perception and Social

One of the major categories within the results gained throughout the case study is related to the perception of the technology and data combination. Perception entails how the procedure was perceived within the firm, how it was operationalized, how people reacted to it. Furthermore, it entails how it was looked at through the lens of the previous situation. What has been illustrated within the Literature background sub-chapter in chapter 1, is that the human perception of technologies such as Big Data and AI, or even the integration of ESG data for that matter, is not well documented. Within this sub-chapter, the results regarding the perception and social aspects will be discussed.

The convincing power of Big Data and AI within the realm of ESG analytics

One of the more interesting observations made throughout the case study is the way people tend to perceive Big Data and AI when it comes to assessing ESG data. As illustrated through the following theme:

- **'Big Data and AI are used as buzzwords'**, which was observed during the defining and introducing phase and during the assessing the procedure phase. It was observed that it was most prevalent during the meetings with external parties aiming to sell Big Data and AI solutions regarding the application of ESG data.

This theme illustrated the novelty of the combination. It was shown that people tend to use these technologies as something novel, exciting. Within the following theme, it was observed that there is something mystical about using these technologies. This was observed in the following theme:

- **'Perception of Big Data and AI is that it is near mystical'**, observed during the defining and introducing phase and during the assessing the procedure phase. It was observed that the perception of Big Data and AI was of something unknown, and was perceived as something near-mystical to someone who does not know what it entails. This was observed while discussing and showing the procedure including software, as is, with a group of colleagues within the firm.

This is further substantiated by the fact that it's necessary to publish literature regarding the demystification of current AI usage (Brock & Von Wangenheim, 2019). However, as observed within this case, the current knowledge regarding the usage of Big Data and AI with ESG data is still lacking, illustrated by the following themes:

- **'Knowledge of capabilities of AI and Big Data differs'**, observed during the defining and introducing phase and the assessing the procedure phase. This has been discussed with an expert who has experience with similar systems. The expert stated that there was "not a perfect fit" between software and firm, rooted in how such IT systems create value for the firm. Illustrating that there is a difference between what IT people think such software can do and the potential business case relevant for the adoption of such technology.
- **'Using ESG data as an umbrella term for all data'**, observed during the defining and introducing phase and the assessing the procedure phase. This was observed during a meeting with another firm that could provide software. It was observed that they used ESG data as an umbrella term to promote their way of using data. The other observation took place during one interview, where it was discussed with an expert that sometimes input data could be interchangeable, and therefore ESG data could be an umbrella term.

This combination of the four aforementioned themes shows the discrepancy of what is known, the internal view of what such a combination can achieve, and it still being something mystical. This cumulates in the concept of using Big Data and AI as a spearpoint for pushing information, for convincing the unknowing employees of a firm that what one has to say is right, and it is being accepted. This came back in the following themes:

- **'Big Data and AI as a spearpoint to push information'**, which was observed in all three stages. It was discussed to use Big Data and AI technology as a spearpoint to push 'data-driven information' to stakeholders within the firm. It was furthermore observed that an employee immediately assessed the potential of the output and the procedure (with the emphasis on Big Data and AI), and stated that it could be used to push information that other colleagues would accept as true. The application of Big Data and AI was furthermore discussed with experts, that it adds a convincing factor to a procedure to push information.
- **'Big Data and AI just for convincing people'**, which was observed in all three stages. It was discussed during expert work meetings that if one uses Big Data and AI, the notion of Big

Data and AI alone can convince people that the information provided by such a methodology using Big Data and AI is correct. This was also observed externally, it was observed that an external firm used these terms (probably) solely for convincing potential customers their software is useful and should be bought.

This even comes from the data side, as illustrated in the theme

- **'People looking only at method, not data quality and source'**, which was observed in all three stages. It was observed during the designing of the procedure that there was interest in the method, bringing a certain 'gravitas' with it. Big Data and AI were observed during a discussion with experts as something big and illustrious. It was mentioned during these discussions that people would be looking more at the method than the data quality.

Here it is shown that if one uses Big Data and AI, the data quality and source are often neglected. This aggregation of themes points into a certain direction, fitting the following initial proposition.

- The introduction of Big Data and AI into a process is more of an approach to carry a big stick to silence sceptics of the subject of ESG than to give results not anticipated by relevant stakeholders using the tool- and datasets.
- It, however, is a very big stick and has the required effect, showing that the social dimension of implementing such abstract tools- and datasets should not be overlooked.
- The source of data is not questioned when the method of implementing data is solid in the eyes of the receivers of the results of the method.

These initial propositions have to a certain extent been observed, as the combination of Big Data and AI assessing ESG subjects and then portraying this information is seen as effective and even more, believable. These propositions, however, do not fully address the themes observed or are a combination of possible propositions. These propositions will now be addressed through the lens of Big Data and AI, as the ESG aspects will be addressed in the 'ESG' sub-chapter. Thus, the following novel theoretical propositions have been defined.

- If Big Data and AI are used within a process, people tend to be convinced by Big Data and AI used within the process, thus Big Data and AI can be used to convince people of the validity of the results of the process.
- If Big Data and AI are used within a process, data quality and source are perceived as of less importance.

The first proposition illustrates the convincing power of Big Data and AI as observed within the case study. A rivalling theory is that it is not caused by Big Data and AI but through its surrounding hype. Elish & Boyd (2018) state that this convincing rhetoric is due to the hype surrounding Big Data and AI. However, within the case, the hype regarding these kinds of systems was observed, but this link has not been observed. Thus, found evidence points towards the first proposition.

To come to the first propositions, what has been observed is a certain bias coming with AI. It is the inclination to believe, beforehand and afterwards, everything the software tells them to. This believability might come from several factors as observed throughout the case study. People being impressed with the methodology, people not being able to understand the full reasoning behind the output, the lack of ability to comprehend the way to results are derived from data, the overwhelming amount of processed information and therefore assuming the output is correct, and the novelty of the technology. One could argue that even a couple of the aforementioned factors could lead to an inclination to accept the results without any further critical thought. Even more, if there is critical thought within the receiver there is no possibility to, within the current state of the technology, fully assess the process by which an AI comes to its answers.

What a receiver could do is perform a validating check of the output comparing it to already-known results, addressing a singular part of the results. However, as this is already known, the chance that an AI comes to similar results is high. This has been observed within the case of the topic "biodiversity", which was one of the topics analysed within the case study. For a validating check, the results have been compared to another analysis done within the firm, and similar results were achieved. This leads to a couple of things, first, if the analysis is done on a subject on which information is already known, nothing new comes out of it, as the data is already processed conventionally. Furthermore, these kinds of analyses are often not done to validate other conventional analyses. An argument could even be made that conventional analyses are not up to par with a Big Data and AI approach due to the sheer volume of data, thus, are non-comparable. Thus, there is an inclination to accept the results, due to a lack of validating possibilities and insights in the process.

These aforementioned arguments regarding the first propositions all lead to a certain bias. From the perspective of the user towards results gained through an analysis done by AI software.

- It is a bias rooted in being unable to comprehend the AI analysis process.
- It is a bias to accept the results of a process, due to one proclaiming AI has been used.
- It is a bias rooted in the novelty of the term AI, which has the perception of something near-mystical and extremely novel.
- It is a bias rooted in the way people perceive a process, as the focus is mostly on the process when results are assessed. With AI, the process is a "black box", which stumps the first inclination towards further critical review.

These aforementioned directions of bias cumulate in a novel kind of bias, a 'prophet bias'. The following analogy regarding prophet bias can be made regarding AI. The output of AI can be seen as the AI telling prophecies without being able to fully comprehend and address the approach to these "prophecies". Causing people to believe the prophet, as it is outside of one's comprehension.

This prophet bias can be confirmed when observing past cases. One of the more prominent cases of observing this prophet bias was found in a small aspect of the Wirecard fraud. The Wirecard fraud was a series of accounting scandals based on fraudulent statements. It was one of Europe's leading FinTech's, as it proclaimed to use novel data technology for their banking operations. Wirecard proclaimed that novel FinTech solutions such as machine learning and AI were used to analyse data. However, in reality, mere Excel spreadsheets were used to organise information as stated by the Financial Times (Financial Times, 2020). The CEO of Wirecard, Markus Braun, had the preference of using FinTech terminology such as "machine learning", often boasting about it (Financial Times, 2020). This combination of a technology frontier and a "foggy atmosphere" surrounding FinTechs provided an opportunity for white-collar fraudsters to come up with opportunities to abuse FinTech related companies in the capital market, as illustrated within the case of Wirecard (Zeranski & Sancak, 2020). Here, it can be argued that the receivers of information from Wirecard & Wirecard's CEO were affected by the prophet bias of AI. They received information fueled with notions of "AI" and "machine learning", while in the backend nothing of such technology was true. Showing the acceptance and believability of such processes rooted in these technologies.

The first proposition has further implications when Big Data and AI is used within a process of conferring management information. If, for example, the person who provides information uses these terms within the information provided, it could become more believable and therefore the information would be accepted without any complications. An interesting thing is that this approach to bias regarding AI is one not taken within the literature, as the most focus is on how human bias could affect the process of creating AI, or how there are different input biases. It is the authors' opinion that there are numerous factors involved regarding this prophet bias not yet uncovered by this case study. The full reasoning behind this observed bias has not yet been unearthed. Thus, this would make an interesting future research direction. More on this in the Recommendations and future research chapter.

The second proposition is one of focus, where due to the introduction of Big Data and AI, the focus of the process is on the method and not the sources. A rivalling theory is that when using Big Data and AI, data sources become more important within a process. Illustrating that the focus within such a process is more on data. As Pascheck, Luminosu, & Negrut (2020) state "…, the thesis that data and its nature are the important prerequisites for AI and decision-making in the business environment could be confirmed.". However, their paper had a focus on data, which gives an inherent tendency to illustrate this bond. When such a process is assessed, data is a significant and non-neglectable aspect of such a process. What this case study has illustrated however is that the perception of the importance of data could be overshadowed by the methodology, thus the usage of Big Data and AI. Thus, the second proposition takes on a different perspective. This proposition implies that emphasis has to be put on the data aspects of a process using Big Data and AI. That one could not alter the results of such a process by biasing data.

Both propositions, in the end, influence the acceptance of the results of the Big Data and AI toolset, thus the acceptance of the results of the process. What has been observed within the case study is that when such a process of Big Data and AI is introduced, the perception of the importance of data quality and sources is affected, as the focus mostly goes to the process of AI itself. Due to this focus on the process, and less focus on the importance of data quality and sources, a positive effect is seen for accepting the results. Furthermore, by including Big Data and AI within a process, the perception of the validity of the process is furthermore increased. This was observed to positively affect the acceptance of the results of such a procedure. Good to know is that the positive and negative causal effects come from the observations within the case. Other effects could also be prevalent, however, these have not been observed. Thus, the potential effects of the propositions are illustrated in the following structural causal model.

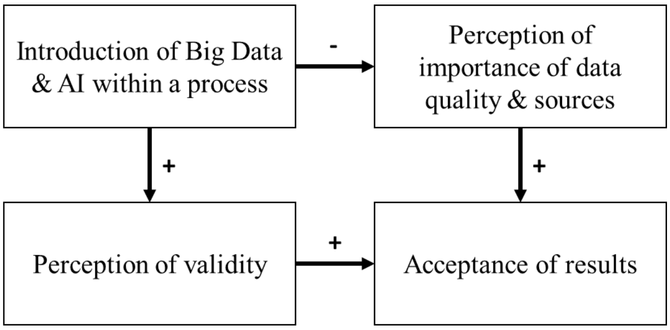

Figure 7 Diagram illustrating the potential causal effects related to the introduction regarding perception

The perception of bias in an 'unbiased' process

Throughout the observations, numerous themes came forward related to the notion of 'bias'. This bias, an initial inclination or prejudice for or against a certain result, is inherent to IT systems such as the ones applied in the case study. However, as discussed in the Case study characteristics chapter, the aim of such inclusion of Big Data and AI within a process were to substantiate the position one takes when passing on management information. As one expert stated during an interview, this illustrates that Big Data and AI can be used to 'wield a big stick' to convince people. However, the basis for this is found in the perception that a Big Data approach is all encumbering, that it is the pinnacle of information gathering, as seen in the following theme:

- **'Big Data and AI just for convincing people'**, which was observed in all three stages. It was discussed during expert work meetings that if one uses Big Data and AI, the notion of Big Data and AI alone can convince people that the information provided by such a methodology using Big Data and AI is correct. This was also observed externally, it was observed that an external firm used these terms (probably) solely for convincing potential customers their software is useful and should be bought.
- **'Big Data and AI as nonbiased view'**, which was observed in all three stages. The aim of using Big Data and AI to assess ESG data is to generate an unbiased view, as discussed during the expert work meetings. However, within this theme, the reflection was also made during an interview with an expert from the perspective of the receiver of such information. Where one expert stated that it has the 'gravitas' to convince people that this is sort of an unbiased way to approach information sharing.

and information found related to

- **'Expert opinion is prevalent/used'**, which was observed in all three stages. It was discussed with experts in the context of how a process was designed and set up. An example was given that with every IT process, there are certain human aspects to it, as one has to initially define the expectations of the output of the IT process.
- **'Expert opinion has bias'**, which was observed in all three stages. The bias aspects within expert opinion related to the design and operations of a procedure, like the one of the case study, have been discussed with experts. The tendency was to group this with the previous theme, as expert opinion is prevalent, thus, bias might creep into one's IT procedure. However, this bias was not deemed as unacceptable, as long as such a system is designed to show how it operates. An example of this was given by an expert of a big online retailer, who used AI to assess their own job listing with success criteria required for such jobs and matched it with future candidates. This system had a significant bias towards the male applicants, as the current jobs were mainly filled by men. As there were more men whose job performance were analysed, the results and success criteria were more applicable for men. Thus, if the people who designed such a system made sure that their bias was properly addressed, it is not a big problem. This example of this retailer illustrated the contrary.
- '**Even in a Big Data and AI process, expert opinion is needed'**, which was observed in the indexing current methodology and results of assessing the procedure stages. It was discussed with an expert, that when one creates such a procedure using Big Data and AI, it has to be done by a so-called 'expert'. Thus, someone who has knowledge and brings their own opinions and bias into the process.

The first themes illustrate how the combination of Big Data and AI regarding ESG data is being used and what the perception is, one of providing an unbiased view. The latter showing that the actual process and technology is based on expert input, including their bias.

This contradiction could show several things. First and foremost, the common denominator within these themes, as observed, is one's knowledge of the process. This comes forward as different things have been observed by talking to people having a different base of knowledge. First, the perception came forward that it could lead to an unbiased process, after which it was discussed with experts that it always has bias inherent to the IT system. The question to be asked when looking at this combination is one of the knowledge regarding, and knowledge of bias in, such a procedure using Big Data and AI.

An interesting theme to be combined with the aforementioned discussion regarding bias is one related to conveying (management) information, which, in the end, was the goal of the procedure. This theme is:

- **'Big Data and AI as a spearpoint to push information'**, which was observed in all three stages. It was discussed to use Big Data and AI technology as a spearpoint to push 'data-driven information' to stakeholders within the firm. It was furthermore observed that an employee immediately assessed the potential of the output and the procedure (with the emphasis on Big Data and AI), and stated that it could be used to push information that other colleagues would accept as true. The application of Big Data and AI was furthermore discussed with experts, that it adds a convincing factor to a procedure to push information. Within this theme, it has been addressed that nowadays, even as an expert in their respective field, one has to substantiate their management information, opinion, and advice on data.

This "convincing factor" as stated in this theme could cause one to turn a blind eye to this inherent bias. This, again, has been indirectly observed in the theme of 'Big Data and AI just for convincing people', as mentioned at the beginning of this sub-chapter. What can be furthermore said about this combination is that, due to the inclusion of Big Data and AI to spearpoint a push of information, and this convincing factor inherent to this technology, the tendency to acknowledge bias could be neglected in such a process.

Thus, let us look at the initial propositions and their relation to the aforementioned themes and their interconnecting links. These are:

- Throughout the introduction and utilization of the Big Data and AI toolset, there are always human interactions and interpretations of the data required. As an ontology has to be created, human bias is still inherent to the procedure itself. This shows that, while the assumption within a firm is that such a toolset solves the bias, it is inherently not true.
- Most data sources and methodologies used to lead back to, or are based in, 'expert opinion', even if the data and/or methodology claim to be objective, disproving that most data-backed or data-driven decisions are unbiased.

Both propositions, when aligned to the results, can be considered plausible. However, these propositions do not fully paint the picture of bias inherent to the application of Big Data and AI within a firm. The disparity of what is perceived and what is fundamentally true to such a process might be ingrained in this combination of Big Data and AI, which also has to be addressed. Thus, a proposition has to show that there is a discrepancy between the bias inherent in a Big Data and AI system and the perception and expectations of such a system. With this in mind, the following propositions were defined.

- There could be causation between one's knowledge of Big Data and AI, and the perception of bias when assessing a process that uses Big Data and AI.
- If conferred management information is substantiated by an information process using Big Data and AI, then people do not have the tendency to acknowledge the inherent biases in such processes.

The first proposition illustrates that when there is more knowledge regarding these technologies, people tend to assess the process better, thus are more aware of biases inherent to these systems. When reflecting this on Big Data and AI, one has to ask themselves if this is inherent to Big Data and/or AI. Or if other technologies face similar challenges. Within this case, I would argue that due to the untransparent nature of AI, and the near-impossible to check the amount of data inflow of Big Data, it is impossible to fully grasp and assess the process of Big Data and AI. Thus, one can only be aware of a bias within such a system if one can comprehend how to assess it.

This notion of bias inherent to such systems is not something novel (Ntoutsi, et al., 2020; Raub, 2018; Akter, et al., 2021). One interesting development during the writing of this discussion regarding bias is the following. The proposition as formatted here is a similar proposition as discussed by Akter, et al. (2021), they published similar results regarding the social aspects of bias in Big Data and AI systems in early July 2021, during the writing of this proposition. While Akter, et al. (2021) extended their focus on three cases of bias, data bias, method bias, and societal bias, the first proposition had significant overlap with the respective societal bias. As this proposition comes from a holistic single-case study, it is interesting to see that another study, which uses a systematic literature review, thematic analysis and a case study on the Robo-Debt scheme in Australia, produces comparable results. However, the difference lies in the factor of perception, inherent to social aspects. This is an additional factor found within this case study, perhaps of importance regarding bias.

Potential rivalling theories regarding the first propositions are mostly mitigated due to the paper of Aker, et al. (2021). With stating the novel proposition, it is a futile exercise the moment they are not 'novel' anymore, solely confirming previous studies. However, different perspectives can be taken regarding AI, if one changes the inherent perception of bias by altering the fundamental characteristics of AI, this proposition will not hold up. As this discussion matches the 'Learning and adoption' category, falling under the sub-chapter 'The concept of Big Data and AI as a black box', this perspective will be elaborated on within this upcoming sub-chapter.

The second proposition shows the general acceptance regarding the usage of these technologies and their relationship to bias. This proposition continuous on the proposition stated within the previous sub-chapter. It, again, illustrates the convincing power of Big Data and AI. Here, with the focus solely on the Big Data and AI aspects, leaving out most ESG aspects. What is interesting about this proposition is that it illustrates two things, one of ignoring bias due to the methodology, and the second of accepting and being convinced by the methodology, turning a blind eye to the negative aspects.

To discuss rivalling theories, the approach taken with this proposition is to address influencing factors that could prove the contrary. One major influencing factor could be company culture, as is observed in the following theme:

- **'Company culture'**, which was observed during the defining and introducing and assessing the procedure phases. Company culture was observed while discussing the procedure with experts. This observation came from the idea that the acceptance of software could be assessed through the lens of company culture. It was observed that the tendency of the discussion was that company culture directly affects the adoption of new technology. One expert stated during an interview that on the front end a lot has to be done to make sure it [the combination of Big Data and AI for assessing ESG data] is accepted within the firm, due to the culture within the firm. Another interviewee stated "Culture is everything!", that with a strong culture the adoption of technologies such as Big Data and AI might be hampered, stating "they only going to do it when they have to".

If a firm has an open culture where it is supported to question information, this proposition could be less prevalent. The question is more about how prevalent company culture is when compared to the tendency to acknowledge the inherent bias in such processes. If company culture is less relevant than the aforementioned tendency, the proposition still holds up. Within this context the facets of company culture are not observed in that manner, thus, the current proposition holds up. However, future research might be able to shed light on the aforementioned proposition, perhaps discovering more influencing factors not observed during this case study. The theme 'Company culture' was, however, meagrely observed within the case study. What this theme entails is discussed in the next sub-chapter.

One thing to address within this proposition is the perspective taken on what causes the neglection/ignorance regarding the bias inherent to the information. The proposition states that the neglect is caused by using Big Data and AI within the process. A case could be made that the moment (management) information is conferred, one does not acknowledge the inherent bias to the complete process in which the management information is grounded. Meaning that in general, information is received with scepticism. This could be true, however, this is out of the scope of this study.

The aforementioned propositions hold some causal effects on the acceptance of the results. It is observed that when Big Data and AI are introduced within a process, one should acknowledge the inherent bias to the process. When these biases are acknowledged, it could positively affect the acceptance of the process, as one perceives more validity regarding the process. Furthermore, personal knowledge has been observed as an influencing factor regarding the possibility of assessing a process using Big Data and AI. This could positively or negatively affect the assessment process, and therefore, the acceptance of the results. As these factors, in the end, influence the acceptance of the results, these causalities coming from the aforementioned propositions are added to the previously made causal diagram. Thus, the following causal diagram illustrates the aforementioned factors, which are based on the observations and tied together through the aforementioned theoretical proposition.

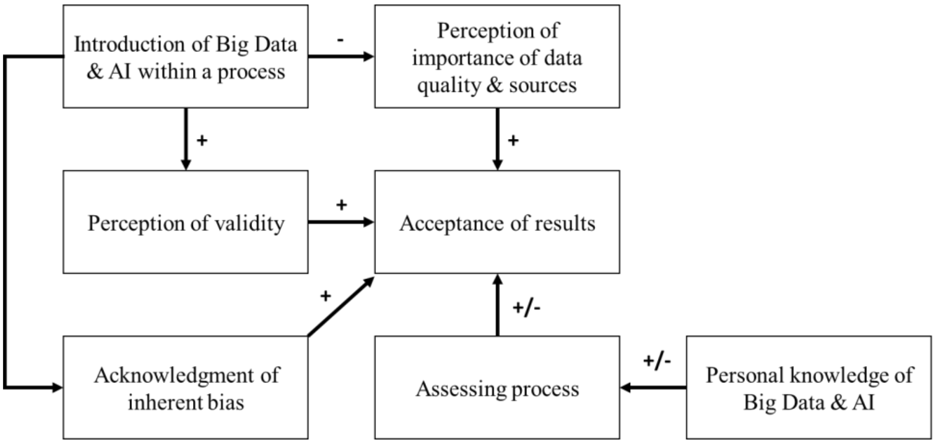

Figure 8 Diagram illustrating additional effects of bias and knowledge

Interesting to see is that it mostly revolves around the acceptance of the results, as previous steps regarding the process could be deemed of less importance. However, the goal of the procedure was the creation of these 'results'. Thus, were deemed of central importance and therefore extremely prevalent within, for example, the third stage of the case study, where the process and results were discussed with experts.

Company culture

One interesting theme observed once during the direct observation phase, is that of company culture.

- **'Company culture'**, which was observed during the defining and introducing and assessing the procedure phases. Company culture was observed while discussing the procedure with experts. This observation came from the idea that the acceptance of software could be assessed through the lens of company culture. It was observed that the tendency of the discussion was that company culture directly affects the adoption of new technology. One expert stated during an interview that on the front end a lot has to be done to make sure it [the combination of Big Data and AI for assessing ESG data] is accepted within the firm, due to the culture within the firm. Another interviewee stated "Culture is everything!", that with a strong culture the adoption of technologies such as Big Data and AI might be hampered, stating "they only going to do it when they have to".

This theme, addressing company culture as a potential influencing factor, was initially thought of as promising. The stream of thought was that company culture is one of the essential building blocks of adopting, assessing, and overall influencing the use of ESG data, Big Data, and AI. Making it particularly interesting concerning Learning and adoption. However, besides the one observation during the direct observation phase and then the re-addressing of this theme during the interview phase, no significant amount of useful data came forward.

The general gist of what came forward was that it definitely plays a role, as one interviewee states "Culture is everything!". Even more, ESG and ESG data are novel subjects within the financial industry, by which it is heavily opinionated. Nevertheless, within this study, one could only hint at the causal relationships between company culture, learning and adoption as the category under which it was classified, and ESG data, Big Data, and AI. This would hint at the following, with a strong culture the adoption of technologies such as Big Data and AI might be hampered. However, this was not observed a lot and came mostly from one interview.

The cause of the lack of useful data is further addressed in the Current situation & sub-chapter, as this lack of useful data regarding company culture and the lack of useful data within the Current situation & New situation category is derived from a similar cause.

5.1.3. Current situation & new situation

This case study gave an ideal opportunity to also assess the introduction of ESG data, Big Data and AI, within a process. Within the case study, it was aimed to assess this aspect through the lens of creating a new situation, departing from an old situation without these technologies. Therefore, two categories were created within the coding to categorise themes regarding this comparison. These are 'Current situation' and 'Output'.

The following two themes found within this category, the current situation, illustrate how a current process is defined.

- **'Expert opinion is prevalent/used'**, which was observed in all three stages. It was discussed with experts in the context of how a process was designed and set up. An example was given that with every IT process, there are certain human aspects to it, as one has to initially define the expectations of the output of the IT process.
- **'Expert opinion has bias'**, which was observed in all three stages. The bias aspects within expert opinion related to the design and operations of a procedure, like the one of the case study, have been discussed with experts. The tendency was to group this together with the previous theme, as expert opinion is prevalent, thus, bias might creep into one's IT procedure. However, this bias was not deemed as unacceptable, as long as such a system is designed to show how it operates. An example of this was given by an expert of a big online retailer, who used AI to assess their own job listing with success criteria required for such jobs and matched it with future candidates. This system had a significant bias towards the male applicants, as the current jobs were mainly filled by men. As there were more men whose job performance were analysed, the results and success criteria were more applicable for men. Thus, if the people who designed such a system made sure that their bias was properly addressed, it is not a big problem. This example of this retailer illustrated the contrary.

It illustrates that regarding ESG, a significant amount of expert opinion is used. This, however, was already addressed in the previous sub-chapter. The following themes were observed in the old and new situations, where expectations of the process are voiced and illustrated.

- **'Big Data and AI to reduce manual labour'**, observed at all three stages. This theme came from the observation that it was perceived that Big Data and AI could reduce manual data analysis labour. It was discussed with experts that such technologies have the capability of making one's life easier. Stating that "the use of big data is the future, we have to go there, we can't do this (the data retrieval process) manually". Even illustrating that certain data retrieval processes could become so big that manual labour is out of the question.
- **'Big Data and AI to define speed of trends'**, was observed during the indexing current methodology and defining and introducing phase. This combination of Big Data and AI could assess trends, as it analyses societal data. It was discussed with experts that it can be used as an instrument to have a short-term feel for the market. It was also observed that this discussion was a bit of speculation on what this kind of software could achieve.
- **'Data-driven approach'**, which was observed during the defining and introducing, and assessing the procedure phase. It was observed that "a data-driven approach" is the core perception of the procedure, meaning that the aim of this procedure, by using Big Data and AI, provides a data-driven approach to analyse ESG data. It was discussed by an expert in an interview that a data-driven approach is a way to go to present results besides one's expertise.
- **'Review correct output software (AI)'**, was observed during the defining and introducing phase. This observation was about how the software works within the procedure. Meaning that it had to be assessed if the NLP software was assessing the right kind of words, in the right context. Furthermore, the sentiment of these 'hits' within the NLP software has to be assessed. This, as the software, has to be 'trained' to give the correct output.

The first three themes illustrate that the process of using Big Data and AI regarding ESG data reduces the manual labour of sifting through ESG data, and even has the possibility to define trends within the ESG data. This, however, is not novel, as mentioned by Macpherson, Gasperini, & Bosco (2021). Their paper illustrates this. The last theme was found regarding the new situation. It mostly illustrates that the AI part of the process could and/or should be reviewable.

There are not a lot of themes to compare the old situation to the new one. Due to this, this comparison is hard to make. Thus, what can be stated regarding the 'Current situation & New situation' category is the following; There is insufficient data to find and illustrate potential novel theoretical propositions. This lack of useful data might be caused by the data gathering method. The participant observations approach might not be the suited approach to gather data, as one is fully emerged in the process. This gives significant difficulty in creating an outsider point of view necessary for such an 'outside in' category such as this one.

This outcome is unfortunate, as the expected result was a clear overview of how ESG data was used before and after, and even how Big Data and AI could play a role before and afterwards. Another expectation that was not observed was the one linking the theme 'Company culture' to the Current & New situation. This as there was a clear link between company culture and AI (Behl, et al., 2021), and within this theme, it could illustrate a parallel link between company culture and Big Data or ESG data.

5.1.4. (Big) Data

Within this theme, the results regarding (Big) Data will be discussed.

Nature of ESG data

The nature of data used, unstructured ESG data was shown of importance. This comes forward in the following themes:

- **'ESG data origin'**, came forward during every phase of the case study. The ESG aspects were discussed by experts, reflecting it regarding the procedure, and origin was assessed through their knowledge regarding frameworks and ESG rating agencies. It was furthermore discussed with experts what the current approach to formulating and creating ESG data entails.
- **'Importance of how ESG data is formatted'**, which was observed in the indexing current methodology, and results of defining and introducing phase. It was observed and discussed by experts that the ESG data format is a requirement for usage that the data output has to be accessible and usable, which again is a prerequisite for the usability of the procedure. It was observed that it was most relevant to the output, as it creates validity regarding the analysis of the procedure.

These themes were observed during the same observation of the following more general data-related themes:

- **'Importance of data sources'**, which was observed in all three stages of the case study. It was observed during expert interviews, through the creation of the procedure, and discussed by experts during the interviews. An interesting view came from these interviews; It was usually assessed through the question of what is more important, data or method within such a procedure. The interviewees all stated that this question is a certain predicament, as both aspects are important. However, it also showed that data and methods have different success criteria. The common denominator was that one has to think of data input, thus data sources, when assessing the output.
- **'(No) difference between the ESG aspects of datasets and other datasets'**, was observed during the defining and introducing, and assessing the procedure phase. Different results and a discussion came from this theme, as discussed during the interviews; one point of view illustrates that regarding the input of such a procedure, there is nearly no difference

between ESG data and other kinds of data. Illustrating that the method, thus AI aspects and related Big Data data sets through which the analysis is done, do not care about the data input. However, another interviewee stated that ESG data is bound to future regulation and reporting directives. Illustrating the different perspectives one could take regarding ESG data. Furthermore, during one interview a point of view was taken of current coding of sustainability data within the firm. There are certain specifics to it. Due to the regulations, a bank has to adhere to, this coding of certain ESG data has to adhere to this (future) regulation imposed by the EU Commission regarding ESG reporting.

- **'Importance of data quality'**, was observed in all three stages of the case study. It was mostly addressed during expert work meetings. The notion of 'garbage in, garbage out' was prevalent during these discussions. This came forward during the discussion regarding what the success criteria are regarding the software used for the procedure. Here, it was discussed with the Big Data and AI software provider that the Big Data part they provide is unstructured and of 'certain quality'. Showing that the output of the analysis can be checked regarding the data quality and that one of the success criteria is that the firm has the ability to do so.

This combination of different data-related themes points towards a similar direction, the direction of how ESG data is applied and its precondition. This was initially guided through the following initial proposition.

- The combination of ESG data analysed through a Big Data and AI toolset creates similar results to any other dataset analysed through the same method, of which conformation is found in the already existing literature.

This proposition, however, could not be confirmed, modified, or rejected. The data collected was too shallow to illustrate the difference between ESG data and other datasets. Meaning that there were no other relevant themes observed due to there not being another dataset to compare it with, as ESG data was, in essence, a subset of the Big Data dataset used within the case study as illustrated in the theme of '(No) difference between the ESG aspects of datasets and other datasets.

The themes did guide towards some more trivial aspects of ESG data. These aspects are similar to the literature found regarding a holistic view of the usage of AI and its related datasets as described by Pascheck, Luminosu, & Negrut (2020). Who stated that "…, the thesis that data and its nature are the important prerequisites for AI and decision-making in the business environment could be confirmed.". One could conclude through the themes mentioned at the beginning of this sub-chapter that this statement is also valid for ESG data. It illustrates the nature of the data, thus, the ESG aspects of the data. However, by taking the view of ESG data as just a subset of the abstract concept of data such as proposed by Pascheck, Luminosu, & Negrut (2020), the finding of these themes within the case study only further ratify their conclusion. Meaning no novel proposition will be built within this direction of found data.

ESG data quality and the aspect of time

As mentioned in the Introduction chapter, sub-chapter Literature background, there are several aspects to ESG data of importance. These aspects illustrate how ESG data quality can be assessed, consisting of reliability, granularity, freshness, comprehensiveness, actionability, and scarcity (In, Rook, & Monk, 2019). This approach to ESG data quality did come forward within the case study through the following theme.

- **'Time focus of data'**, was observed during the indexing current methodology stage. The time focus of data was deemed relevant within the procedure. It was discussed that the time relevance of any analysis using ESG data is important, as the time focus could affect the relevancy of data. E.g., some topics might be of more importance during the summer than during the winter. It further illustrates that data points within ESG data should have a 'timestamp' to be relevant. This is within the case, data were selected within a certain timeframe, observing, for example, biodiversity, over a period and that this theme was mentioned more often in a certain month. Meaning that ESG data is relevant due to the timestamp of the data input (e.g. a news article, measurements made, or a paper, all published on a certain date). This relevance came forward during the discussion with an expert regarding time and the application of ESG data.

When reflecting this on the available literature, one major aspect which came forward was the notion of "freshness". The definition of freshness as given by In, Rook, & Monk (2019) is as follows "Freshness involves the age of a dataset relative to the relevance of phenomena that it reflects. Freshness is not simply equivalent to how old a dataset is; a dataset may have been produced many years ago and still be 'fresh' if it pertains to events of relevance. For example, decades-old records on environmental litigation may be understood as fresh if they relate to the most recent court proceedings against a company for its pollution activities; whereas data on the dividends paid by that company at the same time may no longer be relevant to decisions, and therefore not fresh.".

When comparing the observation as made in the theme 'Time focus of data', I would argue that the notion of freshness should be split up into two aspects. Into time and relevancy. Their definition merges two aspects into one causal relationship. Currently, "freshness" portrays time as a causal factor regarding relevance, time relevant to the relevance of the dataset. A relative simple relationship is illustrated below.

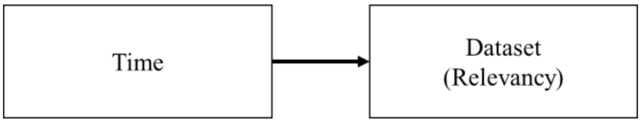

Figure 9 An illustration for clarification of the first statement regarding time

However, this causal relationship neglects the following, time could both be a factor of the dataset, as it could also be a part of the dataset. The first statement as illustrated above furthermore ignores the fact that there could be a continuous measurement of time within a dataset, so-called real-time data. This real-time data illustrates a piling heap of information time-marked where this external factor of time is forced into irrelevancy. The second statement, as illustrated below, illustrates a simple overview for comparison, showing how time is observed.

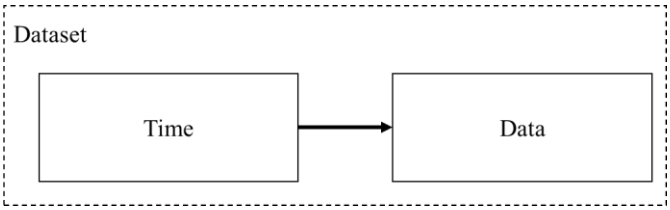

Figure 10 An illustration for clarification of the second statement regarding time

Both statements show the aspect of time, the first one as illustrated by In, Rook, & Monk (2019), the second as observed within the theme of 'Time relevance of data'. This theme regarding the time relevance of ESG data mostly applies to time as a factor within data. For example, the data regarding Biodiversity within the Netherlands, classified under the E of Environment within ESG, could be relevant as follows. An ESG dataset focussing on biodiversity ranging over several years could show several news articles published within one month, thus, showing that biodiversity becomes more prevalent within the public debate. This conclusion could not be deducted only by assessing ESG data according to the first statement.

This reflection of the theme 'Time relevance of data' on literature henceforth illustrates the following; it shows that the factor of time can be interpreted as a significant factor regarding ESG data, as a data point in itself. However, the aforementioned discussion and related theme continue on the notion of ESG data quality by In, Rook, & Monk (2019). A novel proposition could be made regarding time-related to ESG data and its relevancy. However, within this case study, it is not possible. The theme provided a decent base for discussion regarding one paper discussing the quality of ESG data. It did not show any true novel insights, as the aspect of time within an ESG dataset is already known. See by In, Rook, & Monk (2019). It is the mere perspective of how data quality is observed and how time as a factor is related to this, is that could be discussed. Thus, no novel proposition will be created from this theme.

5.1.5. ESG

What has been observed within the results of the case study is that most themes are related to Big Data and AI and that the numerous facets of ESG and ESG data did not fully come forward. This can be seen in the handful of themes categorised under ESG, while categories such as Perception & social, and data, were significantly more prevalent. Perhaps this is due to the way of coding themes, or the approach taken of the case study, as it is about the integration of a toolset based in software, where the kind of data, ESG data, is just a subset used. However, this does not mean that there were no results to discuss within the ESG category. First, ESG data will be discussed, then the future of ESG data, after which the notion of ESG data and that it could be culturally bound is discussed.

The notion of ESG data

One interesting theme found regarding ESG data is the following:

- **'Using ESG data as an umbrella term for all data'**, observed during the defining and introducing phase and the assessing the procedure phase. This was observed during a meeting with another firm that could provide software. It was observed that they used ESG data as an umbrella term to promote their way of using data. The other observation took place during one interview, where it was discussed with an expert that sometimes input data could be interchangeable, and therefore ESG data could be an umbrella term.

This theme illustrates the following; it shows that there is a broad connotation of what ESG data entails within the case, while still being delineated. As stated in the first chapter, ESG data is rooted in alternative data, meaning that it is unstructured, multi-faceted, context-dependent, and ESG related. The data inputs come from different non-conventional data sources. Thus, in essence, it is unstructured. However, what has also been observed within this theme is that ESG data as provided by ESG rating agencies, firms that rate other companies regarding their ESG performance according to ESG data, is rather structured.

This shows a dichotomy to what ESG data entails and according to whom. This thesis adheres to the initial definition of ESG data. The other side of the coin is the structured approach taken by these rating agencies, providing a structure in which ESG data is generated. The cause of such different approaches is that there is still no theoretical foundation for, and a scarcity of high-quality ESG data (In, Rook, & Monk, 2019). This again is shown in the theme as mentioned above.

An interesting difference regarding the comparison of this case study and the literature regarding ESG data is that this case study is focussing on ESG data being a corporate resource, while most studies are focussing on investing based on ESG data. Perhaps this difference is the cause of the approach to how ESG data is being perceived. Within this case and illustrated through this theme, ESG data is perceived as a corporate resource where from unstructured data knowledge is gained regarding, perhaps, managerial decision-making. Or ESG data as a structured approach to rate firms according to their ESG data, as done by ESG rating agencies. The initial statement was that ESG data is context-dependent, where the context is one of how the data looks like. This context can be taken further to what the ESG data is being used for, and it being relevant for the unstructured or structured approach to this kind of data. While there is no initial proposition to model this discussion onto, the following proposition is created to further illustrate these arguments.

- ESG data is context-dependent, illustrating that a structured or unstructured approach to ESG data depends on the application of ESG data.

This proposition aims to further the field of ESG data, as it could be used to create a small part of the currently lacking theoretical foundation regarding ESG data as mentioned by In, Rook, & Monk (2019). The context-dependent aspect of this novel theoretical proposition was already addressed within the literature. The link of this dependency to a structured or unstructured data approach could be considered novel.

As Macpherson, Gasperini, & Bosco (2021) state: "ESG remains an evolving concept and that there are multiple reporting standards and frameworks". Illustrating that the concept of ESG, ESG data, and everything surrounding it is still in a very early stage surrounded by aggregated confusion due to numerous applications and terminology. However, what also has been observed is that (non-mandatory) regulation and the need for better and more standardized ESG data pushes the application of ESG data into a structured data approach. More on the future of ESG data in the next sub-chapter.

A rivalling perspective on the aforementioned proposition could be that it would be fully split up, that the ESG data used here is significantly different when compared to data used as in ESG investing. However, the current hypothesis is on such a high conceptual level, as a theoretical framework is still non-existent, this theoretical framework first needs to be set into place to further discuss what different perspectives might entail.

Future of ESG data

There have been some themes observed regarding the future of ESG data. These, however, were most prevalent during the 'Assessing the procedure' stage. Within the interviews conducted this was a deliberate question regarding ESG data and the future of this combination. Within the following themes, the future of ESG data was observed:

- **'ESG data as a precursor for financial data'**, came forward during the assessing the procedure phase. It was discussed with an expert what ESG data actually portrays. The interviewee stated that there might be a correlation between ESG data and financial performance, as it could be a precursor to financial data.
- **'Future of the combination ESG, Big Data, and AI'**, was apparent in the assessing the procedure phase. This, as it was asked of the interviewees to elaborate on the future of this combination of technology and data. The answers given stated that it could become valuable to financial institutions, combining internal data and AI. However, one interviewee stated that external data and AI as being used by such a firm could be a step too far, meaning that they are not going to use this. Another point made was that it is used to reduce manual labour. Another thing to take into account is that there is upcoming regulation. This could also affect the combination of ESG data & technology.

Within the first theme, it was observed once during an interview that ESG data could be a precursor to financial data. Meaning that if a company scores well in, for example, ESG ratings, ESG data from the market point to a positive assessment of a firm, positive financial results could follow. Thus, a correlation could be made between ESG data as a precursor to financial data. However, as the evidence found regarding this potential proposition is slim, no proposition will be introduced regarding this observation. It is, however, a very interesting correlation as it could future the relevance of the field of ESG data. This, again, can be seen in literature, as stated by In, Rook, & Monk (2019). They asked the question of "what is the missing link between ESG data and financial performance".

In, Rook, & Monk (2019) identified three barriers to ESG data mobilisation and integration, hampering the discovery of this aforementioned link. By which the second one, "the evaluation of ESG data requires more robust tools to ensure their appropriate usage", is of interest within this case. What this case study overall illustrates is that this barrier is mitigated, as a case can be made for novel data technologies as used within this case study to be the required 'robust tools'. Even more, the rise of new fintech start-ups applying these Big Data and AI technologies on ESG data as used within this case could further reduce this barrier.

Perhaps by observing more applications of ESG data through different methodologies hints can be found regarding this missing link. This, as it already comes forward within an observation during this case study. Even more, if a different kind of case study would be conducted, such as an embedded multiple-case design, factors could be distilled to fill in this link. This, as within this type of case study multiple embedded units of analyses are present within multiple cases, providing lots of context regarding the link between ESG data and its potential to be a precursor for financial data.

What has also been observed within the theme of "Future of the combination of ESG, Big Data, and AI" is the upcoming EU (non-mandatory) regulation regarding the reporting on ESG factors through a classification system. This regulation establishes a list of environmentally sustainable economic activities, thus, in essence, structuring the reporting of ESG Data. This development is called the EU Taxonomy (European Commission, 2021). This development goes hand in hand with the reporting on ESG data as seen by ESG rating agencies within the theme "Using ESG data as an umbrella term". As companies are being forced to report on ESG data in a standardized manner, the discussions of what it would mean for the future of such ESG data and how it will be implemented will be opened. There is currently an absence of enough data regarding this topic to propose any novel proposition or to further any discussion regarding this topic. It, in the end, is speculation regarding what could be. However, it would be an interesting development to observe. Perhaps after the future introduction of relevant (non-mandatory) regulation, this subject could again be touched upon. Then, more data could possibly be observed.

ESG data is culturally bound

An interesting theme observed which had the potential to contribute to the theoretical foundation regarding the application of ESG data is 'ESG Data is culturally bound'

- **'ESG Data is culturally bound'**, was observed during all stages of the case study. This theme was initially derived from the discussion regarding the importance of certain ESG topics in different countries. An example given was that an ESG topic such as deforestation was of more importance to the inhabitants of a country with native forests than one without. It was furthermore observed during the designing phase of the procedure, regarding the input of the ESG data. As the operational focus area of the firm is North-western Europe, ESG input should focus on this area. As this focus has a geographic aspect, one should adhere to the culture within this geographic area. This was observed, during the discussion with experts, on how one should assess the output and success criteria of the Big Data and AI software as provided by the related company. It was furthermore addressed during the interviews; This was within the context of input data curated by another party. The observation was made by the interviewee that within a similar process, input data was curated by a company of foreign origin to the firm. Thus, the input data was according to their importance, based on their culture.

The aim of this theme was thereof to illustrate that different cultures could probably value ESG themes differently. Meaning that certain cultures could put more value on Environment, or Social, or Governance. Thus, showing that there is an effect created through cultural context.

This could lead to a novel proposition, one stating that there could be a link. However, this proposition and finding are not novel anymore. Fu, Boehe, & Akhtaruzzaman (2021) recently published an article illustrating these aspects during the writing of the discussion chapter. Applying signalling theory to argue that "the incongruence between ESG scope and stakeholders' cultural values can create signalling noise, which can lead to receivers having differing perceptions of ESG signals.".

While this article reduces the novelty of the found results, significantly when one aims to produce novel propositions for future research, it does solidify the external validity of the results. This, as this theme as observed within this case study, is applicable outside of this case study.

E, S, and G data as combined in ESG data

Within the case study, topics regarding Environment, Social, or Governance have been grouped under the term ESG. Meaning that topics such as 'Biodiversity', classified under Environment, has been similarly addressed by the toolset as 'Human rights', classified under Social, or 'Executive remuneration', classified under Governance. The argument could be made that while the approach to each ESG topic is similar, the outcome might differ. Meaning that the input data could affect a different outcome regarding an Environment, Social, or Governance topic. An example of this would be of information regarding Environment only being reported on negatively, thus skewing the sentiment analysis, while the Governance category could be solely portrayed in a positive light within the data. Thus, showing inherent differences between the ESG categories regarding data. Comparing these results within the same analysis could give a skewed result. However, a counterargument could be made that this would be comparing apples to oranges and the analysis approach would stay the same. The inherent differences to these categories, however, did not directly come to light within the procedure. The discussion regarding the potential difference between the ESG categories, however, could provide additional insights and add to the still non-existent theoretical foundation regarding ESG data. Thus, would make an interesting future research direction.

5.2. Additional findings

This sub-chapter addresses additional remarks regarding the results and the aforementioned discussion.

5.2.1. Non-addressed themes

Within the results, a total of 43 themes were observed. Each theme as found in the results illustrates a concept or abstraction of potential interest. Moreover, as stated within the Methodology chapter, this list of themes offers clues to the emergence of relevant and innovative concepts, albeit novel propositions. These concepts as discussed within the 'Discussion of findings' sub-chapter are steered by the initial propositions, thus, creating also scope for the gathering and discussion of data found. However, the results are gained from a full case study, where a significant amount of relevant and similarly significant amount of non-relevant data is gathered through three successive stages. Meaning that after the gathering and coding of the data, not everything could fit or be used within the discussion, as it would not lead to information relevant to the scope of this research.

5.2.2. Non-addressed proposition

From the initial propositions, as stated within the Methodology chapter, one has not been addressed within the discussion. The proposition not directly addressed is:

- The financial sector is progressive when it comes to the adoption of novel technologies.

This proposition was not addressed due to the following reasons. The main one is that this proposition, in the end, did not fully match the intended research. Meaning that it did not link up with any aspect of AI, Big Data, or ESG and ESG data. It was a more general outlook on the financial industry. This leads to the second reason, that there was, due to a lack of focus on the financial sector itself, no data relevant to address this proposition. If there is no data present, no discussion can be built on this data and therefore no conclusions could be made.

5.3. Limitations

There have been numerous limitations encountered within this study. These limitations were both from a practical and academic nature. First and foremost, it has to be addressed that the COVID-19 pandemic made the conduction of this research challenging. All communication channels were limited to MS Teams, meaning that there was a barrier to access interviewees, people with relevant knowledge within the firm, or to have a meaningful discussion regarding this thesis not delineated by time constraints. It also hampered the accessibility and ease of accessibility of Big Data and AI software, as communication is key.

This was the second limitation of the study, the availability of the software. The case study was in part about the designing and implementation of software. This kind of software package was not available within the firm, thus an external party had to be sought and software acquisition had to be done. This step was a significant time constraint. Without the software, no case study could be conducted. Due to this time constraint, everything was pushed forward. The amount interviews conducted were therefore of a limited number, as most interviewees were on holiday during the period these interviews were conducted. However, I do not expect that more interviews would lead to different answers and novel theoretical propositions. This, as the group of interviewees were selected on their differences while still adhering to the initially stated requirements for interviewees.

The limitations within the academic aspects were the availability of previous research and literature. The research was conducted within a niche of literature, where not a lot of initial literature was available. Thus, while it was niche, it was also significantly broad as no focussed research direction was prevalent. This gave the freedom to explore and to come up with novel propositions and again gave the possibility to assess these according to tangential literature. One benefit was that there were already some papers published regarding the implementation and use of ESG data. However, these papers were published regarding how to apply ESG as investment criteria.

A further limitation of this study is the approach to data collection. Preferably, data would be collected through direct observation, meaning that the participants are observed by creating their own process of Big Data, AI, and ESG data. This could show a pure process uninterrupted by an observer and perhaps provide different theoretical propositions when compared to the approach taken within this study. Perhaps other novel theoretical propositions could be defined within the categories of Current situation & New situation as data and therefore the results were lacking.

5.4. Reflections

Within this sub-chapter, reflections will be discussed regarding the stream of literature, chosen methodology, the societal and managerial perspective, and the academic perspective.

5.4.1. Reflections on the available stream of literature

As illustrated within the Literature background sub-chapter, the availability of literature regarding any of the topics this thesis connects with within the context of the operationalization within a firm is shallow. However, over the last half-year, several studies have been published tangential to this research, even providing similar results or results deemed useful for this study. This came forward in several papers published during the writing of the discussion of this thesis. Simultaneously, it has been observed within the search for a proper software package to be used for this thesis, is that there are a variety of start-ups actually applying ESG data, Big Data and AI as a corporate resource are sprouting.

This could show the following:

- There has not been any significant previous literature regarding the application of ESG data, and Big Data and AI, as a corporate resource within a firm, as the technology was not fully yet matured and adopted.
- AI, for example, "is gaining traction in many sectors" (Radhakrishnan & Chattopadhyay, 2020). Thus, as the technology is matured and adopted, it can be studied through the lens of a corporate resource. Meaning that that is currently on its way to being studied fully and effectively.
- As it will be studied through the lens of a corporate resource, a shift will take place from researching solely the technical aspects within a confined setting, towards it being 'applied within the field'.

Furthermore, during the writing and discussion of this thesis, several discussions were held regarding the application of IT within a firm, and the relevant and available literature. According to these discussions with experts, it is common that literature lags in this regard. This notion with in combination with the aforementioned possible reasons could be why the initially available stream of literature was slim.

An interesting thing noticed, however, is that this lack of literature is currently being filled. The following four papers, which were all published recently, are highlighted to illustrate different parts of this observation.

- Algorithmic bias in data-driven innovation in the age of AI, by Akter, et al (2021), in press since October 2021. The authors identified the sources of algorithmic bias in data-driven innovations. They furthermore provide a future research agenda in this field of bias in AI-based innovations.
- Implications for Artificial Intelligence and ESG Data, by Macpherson, Gasperini, & Bosco (2021), in press since June 10th 2021, stating first that the use of novel data technologies such as illustrated within this thesis is still poorly explained. Furthermore, AI in combination with ESG is not yet fully developed and still creates "aggregate confusion". As stated by one interviewee.
- Culture and Mixed Signals: Does ESG Reduce Risk Everywhere?, BY Fu, Boehe, & Akhtaruzzaman (2021), in press since July 26th 2021. Within this paper, a different lens is used to assess ESG and ESG signals. Confirming the idea, as discussed in the Discussion chapter, that there could be a link between culture and ESG data.
- Addressing bias in big data and AI for health care: A call for open science, by Norori, Quyang, Aellen, Faraci, & Tzovara (2021), in press since October 8th 2021. These authors address bias in such software in another industry, the field of health care. Demanding the need for a more unbiased approach to the introduction of such technologies. Addressing the more "human" aspects of such technologies, as was defined as lacking in the Literature background sub-chapter.

All four papers illustrate some aspect of the research direction taken within this case study and its results.

5.4.2. Reflection on the methodology

The reflection on the methodology has been split up into three parts. First, a more holistic reflection on the methodology, after which the three steps of the case study are reflected upon, and lastly the discussion will be addressed.

Research methodology

The methodology of this thesis was defined through a combination of discussing potential case study approaches with a PhD. student, different papers assessing and conducting case studies, and the book Case Study Research and Applications, Sixth Edition (Yin, 2018). While this proved to create a solid and already proven methodology for such a case study, the initial focus was on relying on an approach for theoretical propositions as proposed by Yin's (2018). This, in the end, might not have been the most ideal approach to such a case study. As the most important part of the current case study, the analysis and discussion of case study evidence, was more built upon Grounded theory. This did not initially come fully forward within the methodology as described within this thesis.

Grounded theory differs from more typical scientific approaches where researchers apply theoretical frameworks and models to the studied phenomenon. Within this Grounded theory, theoretical propositions arise from data, untainted by previously established theory. Within Grounded theory, literature is used for illustrating broad strokes of the knowledge available and provides the research gap to be addressed through Grounded theory (Gibson & Webb, 2012).

By using Grounded theory as a research design framework, it provides a systematic approach to generate theory from gained data using inductive and deductive thinking (Strauss & Corbin, 1997).

Through this method, the theory is 'grounded' in data, it employs "the symbiotic relationship […] between research and theory by reasoning from data to theory and then checking the accuracy of the tentative theory by comparison to more data" (Gibson & Webb, 2012).

If Grounded theory was applied in full, it would provide a more holistic framework for this research, as it can currently be seen as Grounded theory being used through a 'proxy' for the design of this case study, as initially designed through Case Study Research and Applications, Sixth Edition (Yin, 2018). This case study mimics numerous facets of Grounded theory. For example, the aim of grounded theory is that theory can be 'grounded' in data gathered. Meaning that theory arises from the data gained. Within this case study data was being used to create novel propositions, grounding them in the data gained. Thus, this translation of data to the proposition of theory is quite similar.

However, one major difference between Grounded theory and this case study is a deep dive into the available literature. Gibson & Webb (2012) state that little to no prevalent literature study should be conducted before using grounded theory. The literature research conducted however was more thorough than Grounded theory would 'allow'. This is of importance within Grounded theory as the researcher should not be tainted by previously known theories directly related to his or her study.

As the approach of Grounded theory was already interwoven with the initial methodology, the presumption is made that the results would be quite similar if Grounded theory would be leading within the methodology. However, the one difference would be the amount of initial literature research. If this was not done, the scope and direction of this study would be extremely broad and would, perhaps, alter the quality of results and conclusions.

Case study methodology

When reflecting on the first step of the case study, the indexing of the current methodology, the following things can be said. What was mainly noticed is that there is not a lot of technical documentation, as was expected. However, memos of work meetings did provide a significant amount of data. However, this data did not always lead to something useful. Perhaps, if this step would be broadened to also include meetings with different stakeholders such as external parties, novel observations could be made about the interaction between the firm and these stakeholders. Stakeholders in the sense of external companies providing Big Data and AI solutions. This has been incorporated indirectly within the data as it was discussed during the meetings. Perhaps more observations regarding this direction could lead to better data regarding perception and adoption.

When reflecting on the second step of the case study, most has already been mentioned. This has been regarding participant observations. Perhaps direct observations, as previously mentioned within this thesis, could lead to better data and therefore other novel propositions.

The third and last step of the case study was the conducting of expert interviews. One major thing regarding these interviews is the number of people interviewed. It was difficult to find the right interviewees adhering to all the requirements. Perhaps lighter requirements and a more broad method could acquire more and better results. Furthermore, the interviews themselves were based on previously acquired results to assess the procedure, thus to assess what has been designed and used during the case study. This is to steer the interview questions and to find perhaps more information regarding the topics. This worked acceptably well, however, did not fully give the interviewee the possibility to introduce truly novel topics related to the procedure. The semi-structured approach did give some freedom to this, as seen in the theme 'ESG data as a precursor for financial data'. Perhaps a more open approach could have led to different results.

Discussion methodology

When reflecting on the discussion and the creations of the propositions, one interesting thing to be seen is that the results were mostly split over the topics of Big Data, AI, and ESG data. This also gave rise to the more split propositions, meaning that it was hard to combine these topics again under one banner. Perhaps a different methodology regarding discussing the results and the creation of theory could mitigate this. However, this methodology has not yet been identified.

What was also a bit lost regarding the discussion is the whole notion of the inside-out and outside-in perspectives as mentioned in the introduction. This came as the whole study itself was conducted within the scope of the inside-out perspective. It is good again to touch upon this subject, as it provides a basis for this case study. However, no further reflections can be given.

5.4.3. Reflections on societal and managerial relevance

This study illustrates the knowledge gap prevalent in IT, where the industry is often ahead of research. This was observed once during a discussion with a senior IT architect within the firm. The discussion was about the absence of literature describing the social aspects of Big Data and AI and the perception of it. It was concluded that these aspects could become relevant regarding the structuring of implementing software based on Big Data and AI. While this was more from the academic side, perhaps illustrating that a manager would read academic papers to support their management skillset and knowledge regarding the implementation of Big Data, AI, and ESG data, the actual managerial implications also have to be addressed.

One of the more major implications is the perception of Big Data and AI, how it is being used. One of the novel propositions was, for example, on where the focus lies regarding the portraying of information. It was proposed that Big Data and Ai could have a convincing factor regarding such information, that the receiving party of the information would accept the information more easily if Big Data and AI was being used. It would even reduce scepticism of ESG data and information. This implies that managers could use these technologies solely for convincing their own standpoints. This is just one example of several propositions to come forward from this study. Within these propositions, the managerial and societal relevance lies.

5.4.4. Reflection on the academic perspective

The academic perspective within this study is one where I have mixed feelings. The initial literature assessed in the beginning addressed not a lot of knowledge regarding the perception of Big Data and AI or ESG data, or how it's being used within firms. Thus, one of the most logical steps was to conduct an explorative study to assess potential propositions which could be addressed in future studies. This could give an interesting academic perspective on an upcoming combination of technology and data. While an MSc. thesis usually is not a study where extremely novel and revealing information comes forward. It was nevertheless insightful and even a bit demotivating to see that papers have been published during the last two months of my study illustrating what has been done during the study. It did, however, reaffirm that the usage of Big Data and AI in combination with ESG data is poorly explained within the literature, as seen in the paper of Macpherson, Gasperini and Bosco (2021). Even more, getting (partly) similar results as Akter, et al. (2021), as described within the discussion, was something reaffirming. It, in the end, illustrated that the approach taken with the literature, as a vessel through which initially only a knowledge gap is illustrated, is a valid and academic approach to research. This, again, was also reflected upon in the Discussion methodology sub-chapter.

One of the more interesting results is that this study, in a small manner, could contribute to the academic debate regarding the bias of technologies such as AI, as discussed in the 'Learning and adoption' sub-chapter. Here, it was asked what would be inherent to AI, and if XAI would even be AI as the core characteristics of AI would be modified. This debate would turn philosophical real fast due to questions like "what is the meaning of a characteristic" or "what is inherent to AI", as the notion of intelligence within AI is abstract. However, it was mostly out of scope for this thesis. It was interesting however to touch upon.

5.4.5. Reflection on MOT relevance

The relevance regarding the curriculum of the MSc. Study Management of Technology is found throughout the study. It is the link between humans and prevalent technologies within firms, how the interaction is between them, and what the adoption is. This study explores Big Data and AI in combination with ESG data as a corporate resource, and what the 'human' implications are of these kinds of technology. This study furthermore takes both the corporate perspective of how Big Data and AI in combination with ESG data is being implemented within a firm, as the academic MOT perspective of where literature is still lacking regarding the adoption of the aforementioned technology and data.

5.5. Validity of the research

Validity regarding this study was initially addressed within the Methodology chapter. These aspects of validity are reflected upon as follows.

Construct validity is that within this case, the correct operational measures are being used for the studying of concepts. The construct validity within the case was on the addition of ESG data, and a Big Data and AI toolset, and what this addition entails within a firm. The construct validity within this thesis comes from the multiple sources used within the discussion. When a proposition is created, it is based on multiple themes coming from the different kinds of sources used. Furthermore, the interviews were set up in such a way that it would add to the results already gained. Meaning that it would support or debunk previous findings, thus creating substantiated results.

Internal validity came to light during the discussion of the novel propositions. Here, different themes, thus observations made, were grouped together as observed together. Meaning that what was stated within the discussion was also observed. While the internal validity is of lesser relevance with this kind of study, as it is a mere exploration of a topic to create propositions to be tested in future research, it was still taken into account.

External validity was deemed also of lesser importance within the methodology. It was not the initial goal of this study to have a significant external validity, as it cannot be easily achieved through a holistic single-case study. However, what was observed within the discussion is that certain propositions discussed could be substantiated with extremely new literature. Illustrating that the observations made within this single case study could actually be prevalent within the wider world.

However, as stated in the 'Methodology' chapter, what is of importance within this case is the evaluation criterion of applicability regarding external validity. Within qualitative research, the transferability of the research findings show applicability outside of the context of the research setting (Hammarberg & De Lacey, 2016). The applicability of the findings of this case study could be considered of importance, as many propositions reflect real-world decisions.

Reliability within this study is the aspect of replicability. This replicability has two contradicting aspects. As this is a study where a subject is being explored not yet previously known within the literature, it is hard to know when the topic has been fully explored. Thus, there are no tangible research boundaries. It is even harder to substantiate when it is taken into account that the study conducted was a qualitative case study where numerous participant observations were made. It's a study that relied heavily on the researcher. Thus, it is near impossible to know if every facet of Big Data and AI, with the combination of ESG data, has been explored. However, reliability within the case study is defined through the case design (Yin, 2018). Every aspect of the case study and the methodology has been made as explicit as possible. Thus, if another researcher would follow a similar path, with a similar focus within each observation, having a similar case study premise, it could come onto similar results.

In chapter 2, Methodology, several criteria have been devised to judge if the case study was successful or not. These criteria were; that the case and the process itself is conducted in such a way that validity has been taken into account. That, in the end, there are novel theoretical propositions created. Which are substantiated and discussed through results and potentially relevant literature. Concluding, that the main research question and its related sub-questions can be answered by going through the process as defined by the research framework. When reflecting upon these, it can be stated that these criteria have been adhered to. Validity has been taken into account, and the results have been illustrated through novel theoretical propositions substantiated in observations and literature. And it all has been concluded in the next chapter, chapter 6, Conclusion and recommendations.

Chapter 6. Conclusion and recommendations

This chapter answers the research sub-questions and the main research question. First, the sub-questions are addressed after which the main research question is addressed. As the results of this study are several propositions to be researched in the future.

6.1. Answering the sub-questions

A total of three sub-questions were devised to answer the main research question. These are answered within this sub-chapter and are as follows.

Sub-question 1: What observations can be extracted from observing the current state of the application of ESG data, Big Data and AI within a firm?

There were a total of 23 themes observed within the current approach of applying ESG data, Big Data, and AI within the firm. This is further illustrated in Appendix A and Appendix B**.** These were clustered under five categories. These themes are illustrated in the results chapter. Most themes were found regarding the Data and Perception & Social categories.

Sub-question 2: What observations can be extracted from observing the development process of a procedure using ESG data, Big Data and AI within a firm?

There were a total of 34 themes observed within the current approach of applying ESG data, Big Data, and AI within the firm. This is further illustrated in Appendix A and Appendix B**.** These were clustered under six categories. These themes are illustrated in the results chapter. Most themes were found regarding the Data, Learning and adoption, and Perception & Social categories.

Sub-question 3: What observations can be extracted from looking back at the process of creating a procedure using ESG data, Big Data and AI within a firm?

There were a total of 31 themes observed within the current approach of applying ESG data, Big Data, and AI within the firm. This is further illustrated in Appendix A and Appendix B**.** These were clustered under six categories. These themes are illustrated in the results chapter. Most themes were found regarding the Perception & Social category.

6.2. Answering the main research question

The main research question is answered through the discussion of the results gained from the sub-questions. This discussion was done within

Main research question: What observations can be extracted from assessing the introduction of a Big Data and AI toolset applying ESG data within a procedure?

To answer the main research question, the results of the sub-questions have been discussed in Chapter 5, Discussion. From this discussion, novel theoretical propositions have been constructed illustrating the essence of the observations extracted from assessing the introduction of a Big Data and AI toolset using ESG data within a procedure. These propositions are:

- The perception within a firm of using Big Data and AI within a process could affect the learning rate and the learning approach taken by the user. This affects the acceptance of the technology. Thus, the perception could affect the adoption rate of Big Data and AI within a firm.

This proposition illustrates the effect the perception of the combination of these technologies have on the learning rate and learning approach. It, in the end, could affect the acceptance of the combination and thus could affect the adoption of such technologies.

- If Big Data and AI are used within a process, people tend to be convinced by Big Data and AI used within the process, thus Big Data and AI can be used to convince people of the validity of the results of the process.

This proposition illustrates the convincing power of Big Data and AI as observed within the case study. It focuses on the affirmation of the validity of the results of the process, with the focus on convincing people of its validity through the application of these technologies. By creating this proposition and analysing its related observations, a distinct bias became prevalent within these observations. This bias is defined as a 'prophet bias'. The following analogy regarding this prophet bias can be made with respect to AI. The output of AI can be seen as the AI telling prophecies without being able to fully comprehend and address the approach to these "prophecies". Causing people to believe the prophet, as it is outside of one's comprehension.

- If conferred management information is substantiated by an information process using Big Data and AI, then people do not have the tendency to acknowledge the inherent biases in such processes.

This proposition shows the general acceptance regarding the usage of these technologies and their relationship to bias. This proposition is built on the previous one, as it illustrates the convincing power of Big Data and AI. It is interesting to see that this proposition illustrates two things, one of ignoring bias due to the methodology, and the second of accepting and being convinced by the methodology, turning a blind eye to the negative aspects.

- If Big Data and AI are used within a process, data quality and source are perceived as of less importance.

The aforementioned proposition is one of focus. Within the case study, it was observed that within the assessment of a process that uses Big Data and AI, the focus is mainly on the method and the sources are deemed of less importance.

- There could be causation between one's knowledge of Big Data and AI, and the perception of bias when assessing a process that uses Big Data and AI.

Within the case study, it is observed that when Big Data and AI are introduced within a process, one should acknowledge the inherent bias to the process. When these biases are acknowledged, it could positively affect the acceptance of the process, as one perceives more validity regarding the process. Personal knowledge has been observed as an influencing factor regarding this factor of addressing bias, thus, assessing a process using Big Data and AI.

- ESG data is context-dependent, illustrating that a structured or unstructured approach to ESG data depends on the application of ESG data.

This proposition aims to further the theoretical foundation of the field of ESG data. It provides a starting ground for future research to gain insights regarding the currently lacking theoretical foundation of ESG data. The context-dependent aspect of this novel theoretical proposition was already addressed within the literature. This proposition is one where it could provide additional insights into a potential fundamental framework regarding ESG data. The link of this dependency to a structured or unstructured data approach is thus of future interest.

6.3. Recommendations and future research

The following sub-chapter will discuss the recommendations and future research. The recommendations are based on the discussion, as there have been numerous findings that did not lead to current novel theoretical propositions, but have the potential to provide these in the future. Future research will be based on the propositions as stated in the Answering the main research question sub-chapter. Here, it will be discussed how future research could approach these propositions.

6.3.1. Recommendations

The recommendations are discussed in the perspective of where data is lacking, hunches have been found but not materialized, and current technological trends not fully ready to be explored yet.

XAI

One interesting development, as mentioned within Discussion, Learning and Adoption, is the notion of XAI. This recent development is still in its early stages, and it would be recommended to readdress this subject in the near future with a focus on the adoption and bias. If the artificial part of AI is explainable, perhaps novel things can be said about the rate of adoption or bias when assessing and introducing procedures making use of AI.

Company culture

As mentioned within the chapter Discussion, Perception and Social, company culture could play a role within the usage of Big Data, AI, and ESG data within a firm. However, no significant data was found regarding this tendency. One could only hint at a causal relationship between company culture, learning and adoption as the category under which it was classified, and ESG data, Big Data, and AI. As noticed within this case study, ESG and the related notion of 'sustainability' is still a heavily discussed and almost controversial topic. Future studies could perhaps shed more light on the combination of company culture, the notion of sustainability, and the adoption of ESG data, Big Data, and AI.

ESG data

As discussed within the chapter Discussion, ESG, upcoming (non-mandatory) regulation regarding the reporting and classification of ESG factors, could affect the way ESG related information is reported. Meaning that ESG data might become more insightful, plentiful, and available for usage. This is due to companies being forced to report on their ESG performance in a more standardized manner. This would open up the possibility to use this ESG data, perhaps providing a basis for more in-depth research. An example of this is the introduction of the EU Taxonomy, which is in force since July 12th, 2020 (European Commission, 2021). As it is currently being implemented more and more, the stream of relevant data keeps increasing. Within an estimate of three years, this subject should be reassessed. As it gives companies within the EU a decent time frame to process, increase the quality of, and publish ESG data.

E, S, and G data as combined in ESG data

As mentioned within the discussion, sub-chapter E, S, and G data as combined in ESG data, the difference between Environment, Social, and Governance within ESG data has not been addressed. A case could be made for the argument that each category has different data, adhering to a different context, thus giving a different result while being analysed throughout the Big Data and AI toolset. However, throughout the case study, nothing regarding this distinction has been observed. This could be due to the scope and direction of the research, as it did not focus on this distinction and grouped it together. A recommendation for future inquiry could be to assess the differences of each category regarding ESG and what conclusions could be derived from this distinction. This direction of inquiry could add valuable insights to the still non-existent theoretical foundation regarding ESG data.

6.3.2. Future research

The result of the case study was several novel theoretical propositions that could contribute to a future research direction in the field of ESG data, Big Data and AI. Thus, these propositions could be used for future research. Within this sub-chapter, a possible future research direction will be defined for each proposition as stated in the conclusion.

- The perception within a firm of using Big Data and AI within a process could affect the learning rate and the learning approach taken by the user. This affects the acceptance of the technology. Thus, the perception could affect the adoption rate of Big Data and AI within a firm.

For this proposition, future research could go in the direction of an experiment assessing the learning rate, learning approach, and the acceptance of Big Data and AI. Here, the study would aim to assess potential causal relationships and to see if other factors could be relevant regarding this proposition. The relevancy of this research is found in the adoption of novel data technologies, as these could provide competitive value. It could furthermore bolster the adoption rate of such technologies within a firm.

Even more, within this proposition, the starting point is the perception of Big Data and AI within a firm. The author speculates that there are factors leading up to the perception of Big Data and AI within a firm. As the perception of such technologies come from something and are usually not created out of thin air. This, however, has not been observed within the case study. A future research direction could be one of assessing the influencing factors regarding this perception. An explorative case study could be conducted to further explore these influencing factors.

- If Big Data and AI are used within a process, people tend to be convinced by Big Data and AI used within the process, thus Big Data and AI can be used to convince people of the validity of the results of the process.

This proposition could be further researched through experiments, where the 'convincing power' of Big Data and AI is assessed. This assessment could be done comparatively by assessing other forms of conveying management information based on different approaches. Perhaps through other, IT means or conveying management information based on expert assessment. Validity regarding the conveyed information could be an effect of the technology used within a process. The relevance of this study is found where one wants to validify the effect where Big Data and AI can be used to convince people, thus accelerating or even positively enhancing the effect of conveying management information.

Furthermore, regarding this proposition, future research is needed in the direction of the prophet bias accompanying these kinds of technology. As it is the authors' opinion that there are numerous factors involved regarding this prophet bias related to these technologies not yet uncovered by this case study. Perhaps a more in-depth explorative case study could shed light on this combination of bias and technology.

- If conferred management information is substantiated by an information process using Big Data and AI, then people do not have the tendency to acknowledge the inherent biases in such processes.

Future research regarding this proposition could be included in the same research direction as the previous proposition, where the 'convincing power' is assessed. Here, other factors can be included such as acknowledging bias and interaction with the results of such processes. The relevancy is found in better conveying information and the increase of acceptance of certain information, such as ESG related subjects as seen within this case study.

- If Big Data and AI are used within a process, data quality and source are perceived as of less importance.

Future research could be in the perception of the application of Big Data and AI. The question of what is deemed of more importance could help in identifying points of neglect within an IT process. An experiment could be proposed to assess this perception, where perhaps more factors related to such processes involving Big Data and AI can be assessed. The relevance of this research is found

- There could be causation between one's knowledge of Big Data and AI, and the perception of bias when assessing a process that uses Big Data and AI.

Within this proposition, it is assessed how an employee perceives a process using Big Data and AI, where their knowledge is a changing factor. It can be assessed through an experiment, where, perhaps through interviews and surveys, influencing factors such as knowledge and experience, and resulting factors such as bias and perception are assessed. The relevancy regarding such research can be found in how the combination of Big Data and AI can be introduced and applied within companies.

- ESG data is context-dependent, illustrating that a structured or unstructured approach to ESG data depends on the application of ESG data.

This proposition is one where it could provide additional insights into a potential fundamental framework regarding ESG data. As mentioned in the Literature background, this framework is still lacking. An interesting approach to this proposition could be linked with the discussion of ESG data as a precursor to financial data, thus financial performance. As discussed in the Discussion chapter, sub-chapter Future of ESG data, a missing link is still there and barriers are discussed. By observing more and different applications of ESG data through different research methodologies insights regarding this missing link could be found and could shed light on the context-dependency of ESG data. This, as during this case study some relevancy was already found. An interesting approach would be to introduce an embedded multiple-case study to assess relevant factors to fill in this link. This as within this type of case study multiple embedded units of analyses are present within multiple cases, providing lots of context regarding the link between ESG data and its potential to be a precursor for financial performance.

6.4. Contributions of this study

6.4.1. Theoretical contributions of this study

The theoretical contributions of this study are split into two aspects. First, the literary perspective regarding each proposition will be addressed. This, to observe where contributions have been made and how the literary reflection regarding these propositions add value. Second, concluding remarks regarding the theoretical contributions of this study are given. These provide a more holistic view of the contributions of this study.

Perspective through literature

This sub-chapter illustrates the contribution of this case study in perspective to relevant literature. A table of key literature was provided within the Literature background sub-chapter, the contents of this paper will be used to reflect on the results, as it provided the basis for this study. A link will be made regarding each proposition, to the relevant literature, and a summary as illustrated in the Literature background sub-chapter will be given. Thus, the initial table as seen in the Literature background sub-chapter will be split among the propositions. Every proposition will be addressed according to these key papers. Every key paper as mentioned will be addressed in relevance to a proposition.

Proposition	Source	Summary paper
ESG data is context-dependent, illustrating that a structured or unstructured approach to ESG data depends on the application of ESG data.	(Monk, Prins, & Rook, 2019)	One of the more important papers found regarding ESG data. It illustrates that there is still a long way to go to provide substantiated theory regarding ESG data.
	(In, Rook, Monk, & Rajagopal, 2019)	This paper illustrates that ESG data is rooted in alternative data. Furthermore illustrating that the way this kind of data is perceived and how it is applied is just as important as what the actual data entails.
	(Kotsantonis & Serafeim, 2019)	This paper addresses ESG data, illustrating that the sheer variety and inconsistency of the measured data and how companies address relevant ESG data leads to data gaps. Further illustrating that there are no well-defined approaches to integrating ESG data.

Table 7 Literary perspective, ESG data

The first proposition to be reflected upon is the one regarding the context-dependency of ESG data. When reflected upon the paper by Monk, Prins, & Rook (2019) and In, Rook, Monk, & Rajagopal (2019), the added value of this proposition is seen clearly. As this paper illustrates that there is still a long way to go regarding ESG data, the proposition could lay the groundwork for this kind of data. The proposition illustrates a distinction between a structured and unstructured approach, as within literature the main direction taken is one of an unstructured approach.

Furthermore, when reflecting on the paper by Kotsantonis & Serafeim (2019), the following contributions can be derived. Their paper illustrates that there are data gaps within ESG data measured by companies. One question derived could be off, is measured what should be measured, in a correct and relevant approach. This proposition of the context-dependency related to a structured or unstructured approach could illustrate the need to revise these ESG data collection methods. Meaning that, perhaps, the structured or unstructured approach taken could lead to these data gaps or even, when combined, could fill these data gaps. This, however, is still speculation and future research is needed regarding this kind of context-dependency. It does take the notion as described by Kotsantonis & Serafeim (2019) one step further.

Proposition	Source	Summary paper
If Big Data and AI are used within a process, people tend to be convinced by Big Data and AI used within the process, thus Big Data and AI can be used to convince people of the validity of the results of the process.	(Antoncic, 2020)	The integration of Big Data and AI technologies could be used to integrate sustainability data in a firms' business model and decision-making for a possible competitive advantage. Furthermore, the board of a firm needs to be 'sufficiently fluent' in the latest sustainability technology to adopt such a way of operating, giving an arbitrary approach to the adoption of such technologies. This paper states regarding the combination of sustainability, Big Data, and AI that it 'could' be done, giving no further tangible proof or application to his claims.
If conferred management information is substantiated by an information process using Big Data and AI, then people do not have the tendency to acknowledge the inherent biases in such processes.		
If Big Data and AI are used within a process, data quality and source are perceived as of less importance.	(Hasan, Popp, & Oláh, 2020)	This paper illustrates the current landscape and influence of Big Data on the finance sector. Providing future research directions. One significant example from this paper is the statement of "…, there is a need to expand the formal and integrated process of implementing big data strategies in financial institutions.".

Table 8 Literary perspective, implementation strategies of Big Data and AI

The three propositions as illustrated in Table 8 can be linked directly to the following two papers, as described within the Literature background sub-chapter. The common denominator regarding these papers is the Big Data and AI implementation strategies. The initial knowledge gap derived is that there is still information needed regarding the adaptation of Big Data and AI. The current landscape within the financial sector shows that there is a need for knowledge regarding the adoption of these technologies, as described by Hasan, Popp, & Oláh (2020).

The three propositions mentioned contribute to the next steps as illustrated in these papers. These propositions show influencing factors regarding implementation strategies. They lay the groundwork for future research regarding the adoption and implementation of these technologies.

The first and second propositions add to the literature as described by Antoncic (2020). when Big Data and AI are used to gain a competitive advantage, it is being implemented within a firm. It provides tangibility to this integration of these technologies and contributes to the influencing factors surrounding adoption and implementation. Thus, taking a next step in "the arbitrary approach to the adoptions of such technologies", as is mentioned in the description of this paper. The third proposition, the one regarding the perception of data quality and source, further contribute to these influencing factors and provide initial detail regarding such adoption processes. These contributions can be also reflected upon the statement from Hasan, Popp, & Oláh (2020), which state the need for expansion regarding the formal and integrated processes of implementing big data strategies in the financial sector. This, as these influencing factors, thus propositions, could affect this implementation of big data strategies.

Proposition	Source	Summary paper
The perception within a firm of using Big Data and AI within a process could affect the learning rate and the learning approach taken by the user. This affects the acceptance of the technology. Thus, the perception could affect the adoption rate of Big Data and AI within a firm.	(Pencheva, Esteve, & Mikhaylov, 2020)	This paper illustrates the adoption of Big Data and AI within governmental institutions. Stating that "a few barriers at the individual level are noted in the literature, but relatively little attention is paid to these". Even going as far as stating that the mindset of the individual could be of significant importance. But further details are not given.
There could be causation between one's knowledge of Big Data and AI, and the perception of bias when assessing a process that uses Big Data and AI.		

Table 9 Literary perspective, adoption on the individual level

Regarding the individual level, literature illustrated that there are several barriers present. However, little attention is given to these. These barriers, however, have not been directly identified. This case study and its results contribute to the discussion regarding these barriers. Notably regarding the perception within a firm that aims to adopt such technologies. Illustrating that factors such as adoption rate and learning approach could be of importance. Furthermore, showing the notion of one's knowledge in relation to the perception and bios for assessing processes using Big Data and AI.

Concluding remarks regarding the theoretical contributions of this study

The theoretical contributions of this study come in numerous forms. The main ones have been discussed in the previous sub-chapter in relation to key literature. Concluding remarks regarding these contributions are as follows. The following propositions illustrate how this kind of technology is being perceived and how it brings an inherent bias to the process of applying this technology.

- If Big Data and AI are used within a process, people tend to be convinced by Big Data and AI used within the process, thus Big Data and AI can be used to convince people of the validity of the results of the process.
- If Big Data and AI are used within a process, data quality and source are perceived as of less importance.
- There could be causation between one's knowledge of Big Data and AI, and the perception of bias when assessing a process that uses Big Data and AI.
- If conferred management information is substantiated by an information process using Big Data and AI, then people do not have the tendency to acknowledge the inherent biases in such processes.

These propositions illustrate a less well-researched part, the integration and application of the combination of Big Data and AI. Aiming to add to the expansion of "the formal and integrated process of implementing big data strategies in financial institutions" (Hasan, Popp, & Oláh, 2020). This, as there is a need for this expansion, as stated by Hasan, Popp, & Oláh (2020).

Another theoretical contribution of this study is the one of adding insights into the development of a theoretical foundation regarding the application of ESG data. This can be seen in the following proposition:

- ESG data is context-dependent, illustrating that a structured or unstructured approach to ESG data depends on the application of ESG data.

The aforementioned theme opens up the discussion on the application of ESG data in a structured or unstructured data approach. As the case study adhered to an unstructured approach, it was interesting to observe the discussions regarding it being applied as structured data, while still maintaining the "same" definition throughout the discussion.

Thus, this study contributes to numerous aspects of key literature as mentioned in the Summary key literature sub-chapter. The aim was to explore the subject of ESG data, Big Data and AI within a firm.

6.4.2. Practical contributions of this study

The practical contributions of this study are found in the application of these technologies and ESG data. It addresses the bias Big Data and AI bring to the table, as was seen in the proposition of "If conferred management information is substantiated by an information process using Big Data and AI, then people do not have the tendency to acknowledge the inherent biases in such processes.". People applying and receiving information based on processes using these kinds of technology should become aware of the inherent bias of such a process if this proposition holds up.

The awareness of the application of these technologies and where the focus lies is also of importance. This can be seen in the following proposition:

- Big Data and AI are used within a process, data quality and source are perceived as of less importance.

This illustrates the practical contribution to where the initial focus lies when perceiving the outcome of a process using Big Data and AI. If this proposition holds true, this could lead to faulty results as there is being worked with data and sources of insufficient quality. As one interviewee stated, "garbage in, garbage out".

Furthermore, stakeholders within a firm have to be aware of these potential "convincing powers" of such technology. This, if the following proposition holds true:

- If Big Data and AI are used within a process, people tend to be convinced by Big Data and AI used within the process, thus Big Data and AI can be used to convince people of the validity of the results of the process.

Another practical contribution comes in the form of assessing the adoption rate of the combination of Big Data and AI. If the following propositions holds true, the adoption rate within a firm could be positively altered.

- The perception within a firm of using Big Data and AI within a process could affect the learning rate and the learning approach taken by the user. This affects the acceptance of the technology. Thus, the perception could affect the adoption rate of Big Data and AI within a firm.

References

Adadi, A., & Berrada, M. (2018). Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI). *IEEE Access, Volume 6*, 52138 - 52160. doi:10.1109/ACCESS.2018.2870052

Akter, S., McCarthy, G., Sajib, S., Michael, K., Dwivedi, Y., D'Ambra, J., & Shen, K. N. (2021). Algorithmic bias in data-driven innovation in the age of AI. *International Journal of Information Management*. doi:https://doi.org/10.1016/j.ijinfomgt.2021.102387

Aldiabat, K. M., & Le Navenec, C. (2011). Philosophical Roots of Classical Grounded Theory: Its Foundations in Symbolic Interactionism. *The Qualitative Report, 16(4)*, 1063-1080. doi:10.46743/2160-3715/2011.1121

Antoncic, M. (2020). A paradigm shift in the board room: Incorporating sustainability into corporate governance and strategic decision-making using big data and artificial intelligence. *Journal of Risk Management in Financial Institutions 13(4)*, 290-294.

Antoncic, M. (2020). Uncovering hidden signals for sustainable investing using Big Data: Artificial intelligence, machine learning and natural language processing. *Journal of Risk Management in Financial Institutions, 13 (2)*, 106-113.

Asatiani, A., Malo, P., Nagbøl, P. R., Penttinen, E., Rinta-Kahila, T., & Salovaara, A. (2020). Challenges of Explaining the Behavior of Black-Box AI Systems. *MIS Quarterly Executive, 19(4)*. doi: 10.17705/2msqe.00037

Avgar, A., Tambe, P., & Hitt, L. M. (2018). Built to Learn: How Work Practices Affect Employee Learning During Healthcare Information Technology Implementation. *MIS Quarterly, 42(2)*. doi:10.25300/MISQ/2018/13668

Bala, G., Bartel, H., Hawley, J. P., & Lee, Y. (2015). Tracking Companies' Real Time Sustainability Trends: Cognitive Computing's Identification of Short-Term Materiality Indicators. *SSRN*. doi:http://dx.doi.org/10.2139/ssrn.2570616

Barlas, Y., & Carpenter, S. (1990). Philosophical roots of model validation: Two paradigms. *System Dynamics Review, 6 (2)*, 148-166. doi:10.1002/sdr.4260060203

Barocas, S., Rosenblat, A., Boyd, D., & Gangadharan, S. P. (2014). *Data & civil rights: Technological primer.* Data & Civil Rights Conference. doi:10.2139/ssrn.2536579

Barriball, K. L., & While, A. (1994). Collecting data using a semi-structured interview: a discussion. *Journal of Advanced Nursing-Institutional Subscription, 19(2)*, 23-44. doi: https://doi.org/10.1111/j.1365-2648.1994.tb01088.x

Becker, H. S. (1958). Problems of interference and proof in participant observation. *American Social Review, (23)*, 652-660.

Behl, A., Chavan, N., Jain, K., Sharma, I., Pereira, V. E., & Zhang, J. Z. (2021). The role of organizational culture and voluntariness in the adoption of artificial intelligence for disaster relief operations. *International Journal of Manpower, Vol. ahead-of-print No. ahead-of-print.* doi:https://doi.org/10.1108/IJM-03-2021-0178

Belanger, F., & Watson-Manheim, M. B. (2006). Virtual teams and multiple media: Structuring media use to attain strategic goals. *Group Decisions and Negotiation, 15*, 299-321. doi:10.1007/s10726-006-9044-8

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qual. Res. Psych., 3*, 77-101.

Brock, J. K., & Von Wangenheim, F. (2019). Demystifying AI: What Digital Transformation Leaders Can Teach You about Realistic Artificial Intelligence. *California Management Review, 61(4)*, 110-134. doi:https://doi.org/10.1177%2F1536504219865226

Bryant, A., & Charmaz, K. (2007). *The SAGE Handbook of Grounded Theory.* SAGE Publications Ltd.

Chapman, G., Cleese, J., Gilliam, T., Idle, E., Jones, T., & Palin, M. (1970, 09 22). Monty Python's Flying Circus, Series 2 Episode 2. BBC.

Chun Tie, Y., Birks, M., & Francis, K. (2019). Grounded theory research: A design framework for novice researchers. *Sage Open Med, 7*. doi:10.1177/2050312118822927

Colson, E. (2019, July 08). What AI-Driven Decision Making Looks Like. *Harvard Business Review*. Retrieved from https://hbr.org/2019/07/what-ai-driven-decision-making-looks-like

Corea, F. (2019). How AI is transforming financial services. In F. Corea, *Applied Artificial Intelligence: Where AI Can Be Used In Business* (pp. 11-17). SpringerBriefs in Complexity. doi:10.1007/978-3-319-77252-3_3

Denzin, N., & Lincoln, Y. (2018). *Handbook of Qualitative Research.* London: Sage Publications Inc. doi:ISBN 978-1-4833-4980-0

Doody, O., & Noonan, M. (2013). Preparing and conducting interviews to collect data. *Nurse Researcher, 20(5)*, 28-32. doi:10.7748/nr2013.05.20.5.28.e327

Dreyfus, H. (1972). *What computers still can't do: A critique of artificial reason.* Cambrige, MA: MIT Press.

Elish, M. C., & Boyd, D. (2018). Situating methods in the magic of Big Data and AI. *Communication Monographs, 85(1)*, 57-80. doi:10.1080/03637751.2017.1375130

Elish, M., & Hwang, T. (2016). *An AI pattern language.* New York, NY: Data & Society Research Institute.

European Commission. (2021, 10 05). *Europa.eu*. Retrieved from EU Taxonomy sustainable activities: https://ec.europa.eu/info/business-economy-euro/banking-and-finance/sustainable-finance/eu-taxonomy-sustainable-activities_en

Financial Times. (2020, June 30). *Wirecard's real business relied on small number of customers*. Retrieved from Financial Tiems: https://www.ft.com/content/7c466351-02fe-4d66-85a1-53d012de7445

Fu, L., Boehe, D., & Akhtaruzzaman, M. (2021). Culture and Mixed Signals: Does ESG Reduce Risk Everywhere? *Academy of Management Proceedings, 2021 (1)*. doi:https://doi.org/10.5465/AMBPP.2021.10306abstract

Gibson, D. M., & Webb, L. (2012). Grounded Theory Approaches to Research on Virtual Work: A Brief Primer. In *Virtual Work and Human Interaction Research.* doi: 10.4018/978-1-4666-0963-1.ch010

Glaser, B., & Strauss, A. (1967). *The dsicovery of grounded theory.* Chicago, IL.: Aldine Press.

Hacking, I. (1982). Biopower and the avalance of printed numbers. *Humanities in Society, (5)*, 279-295.

Hammarberg, K., & De Lacey, M. (2016). Qualitative research methods: when to use them and how to judge them. *Human Reproduction, 31(3)*, 498-501. doi:https://doi.org/10.1093/humrep/dev334

Hasan, M., Popp, J., & Oláh, J. (2020). Current landscape and influence of big data on finance. *Journal of Big Data, 7 (21)*. doi:https://doi.org/10.1186/s40537-020-00291-z

Hughes, A., Urban, M. A., & Wójcik, D. (2021). Alternative ESG Ratings: How Technological Innovation Is Reshaping Sustainable Investment. *Sustainability, 13*, 3551. doi:https://doi.org/10.3390/su13063551

Hyde, K. F. (2000). Recognising deductive. *Qualitative Market Research*. doi:10.1108/13522750010322089

In, S. Y., Rook, D., & Monk, A. (2019). Integrating Alternative Data (Also Known as ESG Data) in Investment Decision Making. *Global Economics Review, 48(3)*, 237-260. doi:10.1080/1226508X.2019.1643059

In, S. Y., Rook, D., Monk, A., & Rajagopal, R. (2019). Alternative ESG Data for Investment Decision Making. *SSRN Electronic Journal · January 2019*. doi:DOI: 10.2139/ssrn.3380835

Kell, G. (2018, July 11). *The remarkable rise of ESG*. Retrieved from Forbes.com: https://www.forbes.com/sites/georgkell/2018/07/11/the-remarkable-rise-of-esg/?sh=55c1eed61695

Kolbe, R., & Burnett, M. (1991). Content-analysis research: An examination of applications with directives for improving research reliability and objectivity. . *Journal of Consumer Research, 18(2)*, 243-250.

Kotsantonis, S., & Serafeim, G. (2019). Four Things No One Will Tell You About ESG Data. *Applied Corporate Finance, 31(2)*, 50-58. doi:https://doi.org/10.1111/jacf.12346

Laermann, M. (2016). The Significance of ESG Ratings for Socially Responsible Investment Decisions: An Examination from a Market Perspective. *Available at SSRN: https://ssrn.com/abstract=2873126*. doi:http://dx.doi.org/10.2139/ssrn.2873126

Liddy, E. D. (2001). *Natural Language Processing. In Encyclopedia of Library and Information Science, 2nd Ed.* NY: Marcel Decker, Inc.

Lohr, S. (2015, April 28). *Less Noise but more money in data*. Retrieved from New York Times: https://bits.blogs.nytimes.com/2015/04/28/less-noise-but-more-money-in-data-science/?_r=0

Macpherson, M., Gasperini, A., & Bosco, M. (2021). Artificial Intelligence and FinTech Technologies for ESG Data and Analysis. *SSRN*. doi:https://dx.doi.org/10.2139/ssrn.3790774

Macpherson, M., Gasperini, A., & Bosco, M. (2021). Implications for Artificial Intelligence and ESG Data. *Available at SSRN: https://ssrn.com/abstract=3863599*. doi:http://dx.doi.org/10.2139/ssrn.3863599

Madhavan, A., Sobczyk, A., & Ang, A. (2021). Toward ESG Alpha: Analyzing ESG Exposures through a Factor Lens. *Financial Analysts Journal, 77 (1)*, 69-88.

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & & Byers, A. H. (2011). *Big Data: The next frontier for innovation, competition, and productivity.* McKinsey Global institute. Retrieved from http://www.mckinsey.com/business-functions/digital-mckinsey/our-insights/big-data-the-next-frontier-for-innovation

McCorduck, P. (2004). *Machines who think.* Natick, MA.: A.K. Peters.

Monk, A., Prins, M., & Rook, D. (2019). Rethinking Alternative Data in Institutional Investment. *The Journal of Financial Data Science Winter 2019, 1 (1)*, 14-31. doi: https://doi.org/10.3905/jfds.2019.1.1.014

Murale, V., & Preetha, R. (2014). An Integrated Approach towards Legitimization of Single Case Design-working Paper. *Procedia Economics and Finance, 11*, 812-818. doi:https://doi.org/10.1016/S2212-5671(14)00245-7

Norori, N., Quyang, H., Aellen, F. M., Faraci, F. D., & Tzovara, A. (2021). Addressing bias in big data and AI for health care: A call for open science. *Patterns, 2 (10)*. doi:https://doi.org/10.1016/j.patter.2021.100347

Ntoutsi, E., Fafalios, P., Gadiraju, U., Losifidis, V., Nejdl, W., Vidal, M., . . . Kruegel, T. (2020). Bias in data-driven artificial intelligence systems—An introductory survey. *Wires Data mining and knowledge discovery, 10(3)*. doi: https://doi.org/10.1002/widm.1356

Paschek, D., Luminosu, C. T., & Negrut, M. L. (2020). Data—The Important Prerequisite for AI Decision-Making for Business. *Innovation in Sustainable Management and Entrepreneurship*, 539-551. doi:10.1007/978-3-030-44711-3_40

Paschek, D., Luminosu, C. T., & Negrut, M. L. (2020). Data—The Important Prerequisite for AI Decission-Making for Business. *Innovation in Sustainable Management and Entrepreneurship*, 539-551. doi:10.1007/978-3-030-44711-3_40

Pedreschi, D., Giannotti, F., Guidotti, R., Monreale, A., Ruggieri, S., & Turini, F. (2019). Meaningful Explanations of Black Box AI Decision Systems. *Proceedings of the AAAI Conference on Artificial Intelligence, 33(01)*, 9780-9784. doi:https://doi.org/10.1609/aaai.v33i01.33019780

Pencheva, I., Esteve, M., & Mikhaylov, S. J. (2020). Big Data and AI – A transformational shift for government: So, what next for research? *Public Policy and Administration, 35(1)*, 24-44.

Press, G. (2014, September 03). *Forbes.com*. Retrieved from Big Data defenitions, what's yours? [Blog Post]: http://www.forbes.com/sites/gilpress/2014/09/03/12-big-data-definitions-whats-yours/

Radhakrishnan, A., & Chattopadhyay, M. (2020). Determinants and Barriers of Artificial Intelligence Adoption – A Literature Review. *International Working Conference on Transfer and Diffusion of IT*, 89-99. doi:https://doi.org/10.1007/978-3-030-64849-7_9

Rai, A. (2020). Explainable AI: from black box to glass box. *Journal of the Academy of Marketing Science, 48*, 137–141.

Raub, M. (2018). Bots, Bias, and Big Data: Artificial Intelligence, Algorithmic Bias and Disparate Impact Liability in Hiring Practices. *Arkansas Law Review, 71(2)*.

Saldaña, J. (2009). *The coding manual for qualitative researchers.*

Sam, K. M., & Chatwin, C. R. (2018). Understanding Adoption of Big Data Analytics in China: From Organizational Users Perspective. *IEEE International Conference on Industrial Engineering and Engineering Management (IEEM)*, 507-510. doi:0.1109/IEEM.2018.8607652

Sánchez, J., Sánchez, Y., Collado-Ruiz, D., & Cebrián-Tarrasóna, D. (2013). Knowledge Creating and Sharing Corporate Culture Framework. *Procedia - Social and Behavioural Sciecnce, 74(29)*, 388-397. doi:https://doi.org/10.1016/j.sbspro.2013.03.029

SAS. (2021, 04 26). *History of Big Data*. Retrieved from sas.com: https://www.sas.com/en_us/insights/big-data/what-is-big-data.html

Shindelar, S. (2014). Big data and the government agency. *Public Manager, (43)*, 52-56.

Strauss, J., & Corbin, J. M. (1997). *Grounded theory in practice.* San Jose State University, USA: Sage Publishing.

Suchman, L. (2007). *Human-machine reconfigurations: Plans and situated actions (2nd ed).* New York, NY: Cambridge University Press.

Toh, T., Dondelinger, F., & Wang, D. (2019). Looking beyond the hype: Applied AI and machine learning in translational medicine. *EBioMedicine, 47*, 607-615. doi:https://doi.org/10.1016/j.ebiom.2019.08.027

Townsend, B. (2020). From SRI to ESG: The Origins of Socially Responsible and Sustainable Investing. *The Journal of Impact and ESG Investing, 1 (1)*, 10-25. doi:https://doi.org/10.3905/jesg.2020.1.1.010

UNPRI. (2021, 08 10). *What are the principles for responsible investment*. Retrieved from www.unpri.org: https://www.unpri.org/pri/what-are-the-principles-for-responsible-investment

Walsham, G. (1995). Interpretive case studies in IS research: nature and method. *European Journal on Information* , 74-81.

Wilson, E., & Wilson, D. (1988). Degrees of freedom in case research of behavioural theories of group buying . *Advances in Consumer Research, 15*, 587-594.

Yin, R. K. (2018). *Case Study Research and Applications, Sixth Edition.* Los Angeles: Sage Publications Ltd. Retrieved from ISBN 9781506336169

Zeranski, S., & Sancak, I. (2020). Does the 'Wirecard AG' Case Address FinTech Crises? *SSRN*. doi:https://dx.doi.org/10.2139/ssrn.3666939

Zhang, H., & Poole, M. (2010). Virtual team identity construction and boundary maintenance. In Long, s. (Ed). *Communication, Relationships and Practice in Virtual Work*, 100-122. doi:10.4018/978-1-61520-979-8.ch006

Appendix A – Results

Coding - Categories	Coding - Themes	01 - Minutes Expert work meeting	02 - Minutes Expert work meeting	03 - Minutes Expert work meeting	04 - Minutes Expert work meeting	05 - Minutes Expert work meeting	06 - Minutes Expert work meeting	07 - Minutes Expert work meeting	08 - Minutes Expert work meeting	09 - Minutes Expert work meeting	01 - Observations company meeting 1	02 - Observations company meeting 2	03 - Observations company meeting 3	04 - Observations made week 20210524	05 - Observations made week 20210531	06 - Observations made week 20210607	07 - Observations made week 20210614	08 - Observations made week 20210621	09 - Observations made week 20210628	10 - Observations made week 20210705	01 - Company Documents	02 - Company Documents	Interview 1	Interview 2	Interview 3
Current situation	Expert opinion is prevalent/used									x		x		x	x				x				x		x
Current situation	Expert opinion has bias									x		x		x	x				x				x		x
Data	Importance of data sources	x			x	x	x		x		x	x		x	x			x	x					x	x
Data	Importance of data quality	x		x	x	x	x		x		x			x	x									x	x
Data	Importance of how ESG data is formatted	x	x		x		x		x		x	x		x	x	x	x	x							
Data	ESG data origin	x			x			x	x	x	x			x	x	x	x	x						x	
Data	Expert opinion as source of data	x	x		x			x				x											x	x	x
Data	Data pre-processing			x			x																		
Data	Time focus of data								x																
Data	(No) difference between the ESG aspects of datasets and other datasets														x		x						x	x	x
ESG	ESG Data is culturally bound		x				x				x		x											x	

ESG	Using ESG data as an umbrella term for all data												x												x
Learning and adoption	Knowledge of capabilities of AI and Big Data differs												x	x				x	x				x		x
Learning and adoption	Usability of the software tool													x									x	x	x
Learning and adoption	Knowledge of Big Data and AI is lacking														x										
Learning and adoption	Perceived barriers to use the technology																		x						
Learning and adoption	Learning and adoption perspectives														x				x	x			x	x	x
Output	Big Data and AI to reduce manual labour								x		x			x		x								x	x
Output	Big Data and AI to define speed of trends									x	x														
Output	Data driven approach										x														x
Output	Review correct output software(AI)										x	x						x							
Perception & Social	ESG is shown as business risk data	x						x	x	x					x										
Perception & Social	Big Data and AI as nonbiased view		x		x				x		x												x	x	x
Perception & Social	Big Data and AI as a spearpoint to push information				x								x							x			x	x	x
Perception & Social	Big Data and AI just for convincing people				x		x						x												x

Perception & Social	The approach of reporting on the Big Data and AI tool				x											x					x	x		x	
Perception & Social	People looking only at method, not data quality and source					x	x							x						x				x	
Perception & Social	Not invented/purchased here syndrome						x																x		x
Perception & Social	AI is received with scepticism								x								x						x	x	x
Perception & Social	Results of Big Data and AI, not for strategic choices									x															
Perception & Social	Results of Big Data and AI, for tactical choices (timing and handling)									x	x														
Perception & Social	Even in a Big Data and AI process, expert opinion is needed									x													x	x	x
Perception & Social	Big Data and AI are used as buzzwords												x	x			x								x
Perception & Social	Perception of Big Data and AI is that it is near mystical														x										x
Perception & Social	The perception of Big Data & AI is 'state of the art'															x								x	x
Perception & Social	Company culture																		x				x	x	x
Perception & Social	Overwhelmed due to hype																		x						
Perception & Social	Generational aspects of acceptance of Big Data and AI																		x				x	x	x

Perception & Social	Big Data and AI used as a black box												x	x			x		x						
Perception & Social	Combination of business implementation and technology																						x	x	x
Perception & Social	Big Data and AI is perceived as (not) valuable																						x	x	x
Data	ESG data as a precursor for financial data																								x
Perception & Social	Future of the combination ESG, Big Data, and AI																						x	x	x

Appendix B – Clarification results

This Appendix gives a summary of the results gained. Data regarding these results can be found in the confidential appendix.

Results of indexing current methodology

The following themes have been observed during the meetings, talks, and company-provided publicly available documents, as part of the first part of the case study.

Importance of data sources

This theme came forward in five meetings. Within the first meeting sources of ESG data were discussed. This came forward through the discussion regarding the importance of data sources by another process using Big Data and AI. Within the other process, the Big Data set was one of the public news sources, but only in English. As the emphasis within both processes, the one of the case study and the process discussed in the meeting is on the geographic operational area of the firm. Thus, the data sources should also contain news data from other languages.

Importance of data quality

This theme is near tangential to the previous theme. However, the previous discusses sources and the latter discusses the data itself. This difference also lies in the variety of the data itself, meaning that during the meetings the quality was discussed. This theme was observed a total of six times during the meetings. One example of this is the fact that when creating the ESG dataset, one has to 'define the size and relevancy of the [ESG topic]'. Further discussing the quality of the data input.

Importance of how ESG data is formatted

While Big Data is often unstructured and in great quantities, this is one of the prerequisites of being Big Data, the importance of how data is formatted is observed in five different meetings. The focus of this theme is regarding ESG data. This theme is regarding the data as input and data as output. The data output also has to be accessible and usable, which is a prerequisite for the usability of the procedure.

ESG is shown as business risk data

What came forward during four meetings is the theme of ESG data as seen as business risk data. The origin within the firm, as a bank, is to mitigate risk within their financial processes. ESG, as mentioned within the meetings, have their origin in events that could have a material impact on the firm. Meaning that ESG data could affect decisions taken further down the line.

ESG data origin

The theme regarding the origin of ESG data is noticed during five meetings. As the procedure is to assess ESG data, the origin was discussed multiple times. It was discussed that the origin of ESG data could have been initially biased. This observation is split into two themes, 'ESG data origin' and 'Expert opinion as data source'. It shows that the origin of data matters, which was confirmed by a meeting assessing the potential of ESG data and the output, which again is based on the data origin.

Expert opinion as a data source

Within the initial meetings regarding the procedure, the following theme of expert opinion as a source of data came forward within four meetings. It was observed that the effects of ESG data are assessed within the firm through the lens of business risk. This ESG related business risk, as part of a process, is assessed through and based on expert opinion.

Big Data and AI as nonbiased view

The aim of using Big Data and AI within the procedure is to generate a nonbiased view. This theme has been assessed during the meetings. This theme comes in the form of the perception within the firm that such software is able to provide such a view.

ESG Data is culturally bound

This theme is observed a total of two times during the initial meetings. This theme was derived from the discussion regarding the importance of certain ESG topics in different countries. An example given was that an ESG topic such as deforestation was of more importance to the inhabitants of a country with native forests than one without.

Data pre-processing

This theme was discussed two times. Within both meetings where this theme was discussed was in the context of "garbage in, garbage out". Meaning that when you have a bad data input, you have no useful results. Thus, it was discussed that data pre-processing could be required within the procedure.

Big Data and AI as a spearpoint to push information

This theme was addressed during one meeting. An expert discussion was observed that the information regarding ESG subjects is usually disputed. It was discussed that Big Data and AI can help to substantiate ESG information. Illustrating the technology being used to spearpoint "data-driven information" to stakeholders within the firm.

Big Data and AI just for convincing people

This theme was discussed two times. This has roots in the notion of Big Data and AI as a technology other stakeholders within the firm do not know of. The aim of using Big Data and AI to convince people was discussed. It was discussed that the procedure should use Big Data and AI in a sense to analyse ESG topics, thus it is being used to convince stakeholders about the importance of such ESG topics.

The approach of reporting on the Big Data and AI tool

This theme was discussed during one meeting and within 2 documents published publicly. The approach to reporting on such technologies was discussed in the meeting, and it was done within the 2 company, publicly available documents. Within these documents, it was seen that Big Data and AI was mentioned as a method used to analyse data. It showed that it could, perhaps, provide legitimacy to the published documents.

People looking only at method, not data quality and source

This notion was observed during the discussions regarding the procedure. Here, the focus of the discussion was what would be the focus of a process of Big Data and AI. It was mentioned that people would be looking more at the method than the quality during two discussions.

Not invented/purchased here syndrome

This theme was observed once, as it was discussed why the firm would adopt such Big Data and AI technology. Within this discussion, it was mentioned that the not invented here syndrome would apply to the procedure and has to be taken into account.

Big Data and AI to reduce manual labour

This theme came from the observation that it was perceived that Big Data and AI could reduce manual data analysis labour. This was observed once.

Time focus of data

The time focus of data was deemed relevant within the procedure. It was discussed that the time relevance of any analysis using ESG data is important, as the time focus could affect the relevancy of data. E.g. some topics might be of more importance during the summer than during the winter. This theme was discussed once. It further illustrates that certain data points within ESG data should have a 'timestamp' to be relevant. This is within the case, data were selected within a certain timeframe, observing, for example, biodiversity, over a period and that this theme was mentioned more often in a certain month. Meaning that ESG data is relevant due to the timestamp of the data input (e.g. a news article, measurements made, or a paper, all published on a certain date).

AI is received with scepticism

This theme was observed once. It was during the discussion regarding a similar process that also uses NLP. However, it was also discussed that within that other procedure the AI aspect was not that useful, as a lot of human interaction was needed. It was discussed that it could be the future, but it was not perceived as that useful, therefore scepticism was observed within this discussion.

Current situation: expert opinion is prevalent/used

Within the discussion about the creation of the procedure, it was discussed that within the current situation ESG data is assessed through expert opinion. This discussion was observed during one expert work meeting.

Current situation: expert opinion has bias

While assessing the current situation within the firm, the theme was found that expert opinion was used in a number of processes. It was observed once. It was discussed that it, however, has the experts bias in every opinion given.

Big Data and AI to define speed of trends

This point was discussed once. This kind of technology could assess trends, as it analyses societal data. It was discussed that it can be used as an instrument to have a short-term feel for the market. It was also observed that this discussion was a bit of speculation on what this kind of software could achieve.

Results of Big Data and AI, not for strategic choices

This theme was discussed once and specifies one aspect of the previous theme. Results could not be used for strategic choices. That's due to the perception that this procedure, and therefore the technology used within the procedure, has a short time aspect. Thus, it would not be perceived as useful for strategic choices.

Results of Big Data and AI, for tactical choices (timing and handling)

This theme was discussed simultaneously with the previous theme. It was discussed that the procedure and indirectly the technology could be used for tactical choices. One could time their handling for putting a product on the market from assessing market (Big) data through AI.

Even in a Big Data and AI process, expert opinion is needed

This theme was observed during one meeting. It was discussed that when one creates such a procedure, it has to be done by a so-called 'expert'. Thus, someone who has knowledge and brings their own opinions and bias into the process.

Results of defining and introducing procedure

The following themes have been observed during the second part of the case study, where Big Data and AI was introduced.

Expert opinion is prevalent/used

It was observed four times that expert opinion is prevalent. This observation was first made during the discussion with a Big Data and AI solutions provider. Here, the discussion was in regard to how data was provided and by whom. Even more, this was observed during another meeting with the same provider. Showing that within the taxonomy used which the analysis is based on, an expert gives their opinion on what actually goes in the taxonomy of the analysis. Even more, the root source of an entry in the analysis is based on the opinion of the person who does the analysis. This was observed during a discussion within the firm.

Expert opinion has bias

This theme was like 'Expert opinion is prevalent/used', observed four times. The discussions observed in the previous theme caused the discussion regarding the effect of such 'expert opinion'. Within each of the four previous observations, this theme is also observed. As every expert brings their own experience and opinion with them.

Importance of data sources

This theme was observed six times. Twice during meetings with different firms providing Big Data and AI solutions. Here, the discussion was in regarding the source of the data by which the ESG dataset was analysed. The observations overall were from the same calibre. Meaning that there always was a focus on the importance of the source of data. This was shown by questions illustrating the importance of the source, relevance of the source, or sources affecting the potential outcome of the procedure.

Importance of data quality

The importance of data quality as a theme was observed a number of three times. Like the expert work meetings, when the design of the procedure was discussed the notion of 'garbage in, garbage out' was prevalent. This, again, came forward during the discussion in regard to what the success criteria are regarding the software used for the procedure. Here, it was discussed with the Big Data and AI software provider that the Big Data part they provide is unstructured and of 'certain quality'. Showing that the output of the analysis can be checked regarding the data quality and that one of the success criteria is that the firm has the ability to do so.

Importance of ESG how data is formatted

The importance of how ESG data is formatted was prevalent during seven observations. As observed within one discussion regarding the procedure, ESG data within this case shows ESG topics and their relevancy. The output is also ESG data and metrics according to these ESG topics. The amount of observations also lies in the fact that this is a core output of the analysis, thus showing no significant substantive difference regarding the observations. What is furthermore observed and inherent part of this theme is that the way ESG data is formatted is part of the internal validity of the analysis. In the end, it was observed that regarding the procedure, ESG data shows ESG topics and their relevancy.

This is used to scope the Big Data dataset regarding ESG data. Thus, the output of the procedure relies on the input metrics, ESG data. Thus, the way ESG data is formatted is part of the internal validity of the analysis, validating that one analyses what one wants to analyse. Furthermore, it was observed that it was most relevant to the output, as it creates validity regarding the analysis of the procedure.

ESG data origin

This theme was observed six times during the creation and rollout of the procedure. It was observed that the origin of ESG data is of importance, as this was mentioned often during discussions regarding the input and therefore also the output of the procedure. These discussions were about the way the ESG data was defined. Within the firm, an approach has been defined by which this ESG dataset was constructed. Due to this, the discussion regarding this process was prevalent, thus the number of observations regarding this theme was high.

Expert opinion as source of data

This theme has been observed once, during a meeting with a company that provides Big Data and AI solutions. This as a number of aspects their Big Data & AI part of their software was based in the 'expert opinion' of the company. As that company defined the way the AI was trained, what sources such as Twitter, news websites, and company reports, were included in their Big Data solutions, and how their taxonomy by which the analysis is done. As there is no freedom within this solution to define one's own process, one has to rely on the company's expert opinion of these data sources.

(No) difference between the ESG aspects of datasets and other datasets

This theme was observed twice. Once during the analysis of how data is interpreted within an NLP toolset. The discussion was observed that: ESG data, unstructured for example, is not any different from any other data when assessed through NLP, thus the question was raised that there is no significant difference when compared to other data. This was also observed within the Big Data and AI software used within the procedure, as ESG is just like any other dataset that has the potential to be analysed through their system. Within this case study, furthermore, ESG data was essentially thus the subset of the used Big Data dataset.

ESG Data is culturally bound

This was observed twice. The first observation was regarding the input of the ESG data, as the operation area of the firm is Northwest Europe, ESG input should focus on this area. As this focus has a geographic aspect, one should adhere to the culture within this geographic area. This was observed during the discussion on how one should assess the output and success criteria of the Big Data and AI software as provided by the related company. This was also once observed during a meeting with a company that provided a way to assess and plot ESG data on a map. Here, the observation was made that ESG could be used to boast one's own achievements within the field of AI. As this approach to promoting ESG data was based on the culture of where this company comes from, cultural relevance has been included as a theme within this observation.

Using ESG data as an umbrella term for all data

Using ESG data as an umbrella term for all data has been observed once. This was during the meeting with another company that could provide a Big Data and AI software solution. However, it was observed that they used ESG data as an umbrella term to promote their way of using data. Perhaps the way ESG data is used defines the approach taken. That if ESG data is being used for managerial decision-making it is more unstructured as the need for flexibility is there.

Knowledge of capabilities of AI and Big Data differs

This was observed a total of four times. The first time this was observed was during the same observation as with the previous theme. Here, it was observed that the company used AI a Big Data as superficial terms to impress potential customers. They show that their knowledge and capabilities in regard to what is AI and what is Big Data differ from the knowledge of these topics within the firm. Another observation was made during a meeting with another company in regard to defining the potential software to be used for the procedure. Here it was observed that the firm held back on what their software could do, and there was a bit of a knowledge chasm between how the people from the firm in the meeting perceived the capabilities of AI and Big Data and what the company providing this software knows. However, another observation was made where one colleague knew significantly more than the company that was presenting their software. Here, this colleague addressed after the meeting that their use of AI might not even be relevant anymore in a short period.

Usability of the software tool

The usability of the software tool has been observed and classified under one 'observation'. This was done during the initial talks with companies providing these kinds of Big Data and AI software tools. However, this was also observed during the actual use of the software tool within the procedure. This, in itself, was prevalent throughout the procedure, as it is an essential part of the usability of the procedure.

Knowledge of Big Data and AI is lacking

This was observed once. This was during the discussion about the procedure within the firm and how it will be used. This was also done while showing what is needed for the procedure. The observation was made that there was the perception that Big Data and AI was a sort of black box to analyse stuff with. That the procedure analyses ESG data according to a big stream of data and AI, thus it creates a nice overview of information. Where the experts perceived it as showing information in, information out, and not how it's done. And that the actual knowledge regarding the software is lacking.

Learning and adoption aspects

The learning and adoption aspects, regarding how the procedure will be/is used, is observed three times. Once during the design of the procedure where it was offered to teach colleagues how to use the Big Data and AI software. It was also observed that there should be a willingness to learn regarding using the Big Data and AI software, thus the procedure. Furthermore, it was observed that this learning might be correlated to the adoption of the previous theme.

Perceived barriers to use the technology

The perceived barriers to using the technology have been observed once. This is in context with the adoption of such technologies within the firm. It was observed that there is a need for learning as seen in the previous theme. It was observed that this theme was also partly caused by barriers perceived by users of such technologies.

Big Data and AI to reduce manual labour

This was observed three times. The most prevalent is the observation made within the procedure itself. Big Data and AI are used to analyse the ESG dataset, one does not have to do a manual analysis anymore. It's the core of the process, so do speak. It was also observed that manual labour was required the moment the Big Data and AI process was insufficient. Thus, showing within this theme that correctly used Big Data and AI could reduce manual labour.

Big Data and AI to define speed of trends

This theme was observed once. This was during the assessment of the software used in the procedure. Here, it was one criterion of what the software could do, provide for the procedure. Defining the speed of trends, as observed, is one of the capabilities of the software structure behind the software as used within the procedure.

Data driven approach

This theme was observed once. This was observed within the same observation as 'Big Data and AI to define speed of trends'. Within the same observation, it was observed that a data-driven approach is the core perception of the procedure, meaning that the aim of this procedure, by using Big Data and AI, provides a data-driven approach to analyse ESG data.

Review correct output software(AI)

This theme was prevalent during the design and review of the procedure. It was observed a total of three times. These observations were about how the software works within the procedure. Meaning that it had to be assessed if the NLP software was assessing the right kind of words, in the right context. Furthermore, the sentiment of these 'hits' within the NLP software has to be assessed. This, as the software, has to be 'trained' to give the correct output.

ESG is shown as business risk data

This theme was observed once during the introduction and usage of the procedure. This was shown during a discussion regarding the capability of the procedure and its related output. The output of such analysis regarding ESG topics could be used within other departments of the firm, as interest was shown in the outcome due to it also being a business risk. This risk aspect (partially) comes from future regulation.

Big Data and AI as nonbiased view

Big Data and AI as a nonbiased view, meaning that it can be used to show a nonbiased approach to analysing data, is observed once. This was observed during the assessment of the software used within the procedure. Here, it was shown that this was one of the aims of the procedure, thus related to the output and process of the procedure itself.

Big Data and AI as a spearpoint to push information

This theme was observed twice. First, it was observed during a meeting with a company claiming to provide a Big Data and AI toolset regarding ESG data. Here, the salespeople used the terms Big Data and AI as a spearpoint to sell their software. However, the actual big data and AI aspects were perceived as negligible. The second observation was during the illustration of the specific output of the procedure. Here, it was observed that a colleague immediately assessed the potential of the output and the procedure (with the emphasis on Big Data and AI), and stated that it could be used to push information that other colleagues would accept as true.

Big Data and AI just for convincing people

This theme was observed once. While this theme could have overlap with the previous theme in regard to similar observations, this observation has the pure intent to convince people through the usage of terms like 'Big Data and AI'. This was observed during the meeting with a company claiming to provide 'Big Data and AI' software. It was observed that this company used these terms (probably) solely for convincing people their software is useful and should be bought.

The approach of reporting on the Big Data and AI tool

This theme was observed once. This was during the designing phase of the procedure. Here, it was discussed how one should report on such procedures. Similarities were made to some similar procedures where it was stated that a 'state of the art' tool was used, referring to a similar Big Data and AI software used.

People looking only at method, not data quality and source

This theme was observed twice. It was observed that there was interest in the method, bringing a certain 'gravitas' with it. Big Data and AI were observed as something big and illustrious. Furthermore, it was observed during one meeting with a company providing Big Data and AI software. Here, they showed a focus on the methodology, and it was assessed that if the data input was shaky at best, they choose to ignore it.

AI is received with scepticism

This was observed once after a demo given regarding Big Data and AI software, usable for the procedure. "The AI of today will not be classified as what is AI in two years." As one colleague stated. Here, some scepticism was observed regarding how AI is received and perceived within the firm.

Results of Big Data and AI, for tactical choices (timing and handling)

This is observed once, during the meeting discussing the success criteria with the firm that provides the Big Data and AI software. This theme came forward as the usage of the main output of the software the firm is going to use. This theme, in the end, is prevalent throughout the procedure, as the output and the usability of the output rely on such factors. Even before designing such a procedure, one has to ask, what is it going to be used for. Therefore, this theme is observed once, but prevalent throughout the procedure.

Big Data and AI are used as buzzwords

The usage of Big Data and AI as buzzwords is observed three times. Once during the meeting with a company, who provides Big Data and AI as a solution. Once regarding a similar software with the potential to be used, where it was promoted as a state of the art Big Data and AI platform. And once with the combination of such software being used as a 'black box'. Here, it was observed that Big Data and AI was used as a buzzword, without even giving 'a peek under the hood'.

Perception of Big Data and AI is that it is near mystical

This theme, the perception of Big Data and AI, as something near-mystical, is observed once. This observation takes the 'black box' idea a step further. It was observed that the perception of Big Data and AI was of something unknown, and was perceived as something near-mystical to someone who does not know what it entails. This was observed while discussing and showing the procedure including software, as is, with a group of colleagues within the firm.

The perception of Big Data & AI is 'state of the art'

This was observed during the assessment of the potential use of certain Big Data and AI software for the procedure, during the 'designing' steps of the procedure. It was observed in one of the reportings on similar software, where the words 'state of the art' was used. Showing the perception of Big Data and AI as state of the art, novel.

Company culture

The company culture theme was also observed while discussing the procedure with colleagues. This observation came from the idea that the acceptance of software could be assessed through the lens of company culture. It was observed that the tendency of the discussion was that company culture directly affects the adoption of new technology.

Overwhelmed due to hype

This was observed once during the discussion after showing how the procedure works. It was stated that novel technologies, e.g. cloud computing, AI, could sometimes overwhelm people and push them into a dichotomous camp, accepting or rejecting, without even properly assessing the technology.

Generational aspects of acceptance of Big Data and AI

This was observed once during a discussion. It was discussed that each generation of people has a different approach to technology. Meaning that the adoption rate could also be correlated to the age of the people who are the decision-makers within the firm.

Big Data and AI used as a black box

This theme was observed a total of four times. This was observed throughout the interactions of colleagues with the procedure. It was observed that Big Data and AI was perceived as something where data goes in and results come out. Here, the black box is designed by a person, and people put data in and collect results as output, illustrating the perception that it is something where the actual method might be of lesser importance. Otherwise, the perception was not something as observed.

Results of assessing the procedure

Expert opinion is prevalent/used

This theme was discussed twice during the interviews. It was both in the context of how a process was designed and set up. An example was given that with every IT process, there are certain human aspects to it, as one has to initially define the expectations of the output of the IT process.

Expert opinion is used/prevalent

This theme was discussed during two interviews. It relates to the way currently expert input is used regarding the way ESG data is gathered and formatted.

Expert opinion has bias

The bias aspects within expert opinion related to the design and operations of such a procedure have been discussed twice during the interviews. The tendency was to group this together with the previous theme, as expert opinion is prevalent thus bias might creep into one's IT procedure. However, this bias was not deemed as unacceptable, as long as such a system is designed to show how it operates. An example of this was given of a big online retailer, who used AI to assess their own job listing with success criteria required for such jobs and matched it with future candidates. This system had a significant bias towards the male applicants, as the current jobs were mainly filled by men. As there were more men whose job performance were analysed, the results and success criteria were more applicable for men. Thus, if the people who designed such a system made sure that their bias was properly addressed, it is not a big problem. The example of this retailer, for example, illustrated the contrary.

Importance of data sources

The importance of data sources was discussed twice during the interviews. It was usually assessed through the question of what is more important, data or method within such a procedure. The interviewees all stated that this question is a certain predicament, as both aspects are important. However, it also showed that data and methods have different success criteria. The common denominator was that one has to think of data input, thus data sources, when assessing the output.

Importance of data quality

The importance of data quality came forward during two interviews. The data quality was assessed mostly through the lens of ESG data. This lens showed that data quality and comparability between firms still have a long way to go. Meaning that no current regulations are defining how a firm should report on their ESG data. Furthermore, one expert stated that sometimes the data quality may be of lesser importance. Meaning that when assessing data quality in a big data set, the flow of the data is of more importance, mitigating the reduced quality through sheer volume.

ESG data origin

The origin of ESG data was discussed once during the interviews. This origin of ESG data was discussed in the context of ESG rating agencies such as DJSI and GRI, whose requirements do not change very often. The reflection in regard to the procedure was that such input was good to take into account. Furthermore, the origin of ESG data was discussed as the output of the procedure was dependent on this. Meaning that the way the outcome of the procedure could be used in relation to what kind of ESG data is for input.

Expert opinion as source of data

This theme was observed during all interviews. This theme was discussed in the context of normal data input and ESG data input. It was discussed that the source of, for example, ESG data is down the line based on expert opinion, what is taken into account and what is not. A side-track in this discussion was one of documenting this expert opinion, if one chooses a source for data, it should be documented. This is to provide a substantiated description later down the line to verify the process. If the process is not verifiable, as the data sources chosen within the process are not clarified through documentation, the process could become unusable over time.

(No) difference between the ESG aspects of datasets and other datasets

This has been discussed within every interview. The results to this theme differ, however one point of view illustrates that regarding the input of such a procedure, there is nearly no difference between ESG data and other kinds of data. Illustrating that the method, thus AI aspects and related Big Data data sets through which the analysis is done, do not care about the data input. However, another interviewee stated that ESG data is bound to future regulation and reporting directives. Illustrating the different perspectives one could take regarding ESG data. Furthermore, during one interview a point of view was taken of current coding of sustainability data within the firm. There are certain specifics to it. Due to the regulations, a bank has to adhere to, this coding of certain ESG data has to adhere to this (future) regulation imposed by the EU Commission regarding ESG reporting.

ESG Data is culturally bound

The probability that ESG data is culturally bound is observed once during the interviews. This was within the context of input data curated by another party. The observation was made by the interviewee that within a similar process, input data was curated by a company of foreign origin to the firm. Thus, the input data was according to their standards.

Using ESG data as an umbrella term for all data

This was discussed within three interviews. All within the context of how and what ESG data might portray. It was stated during one interview that sometimes input data could be interchangeable, and therefore ESG data could be an umbrella term. Furthermore, during one interview it was stated that there are different approaches to ESG data, such as the approach taken to report ESG data according to the GRI or the SASB framework. Showing that different data reporting methods fall under ESG data. This shows that perhaps there is a structured approach to ESG data, the approach ESG rating agencies for example take.

Knowledge of capabilities of AI and Big Data differs

This theme has been observed during two interviews. This theme illustrates the knowledge gap between people in regard to AI and Big Data. This has been observed while talking to an interviewee who has experience with similar systems. Here he stated that there was "not a perfect fit" between software and firm, rooted in how such IT systems create value for the firm. Furthermore, illustrating that there is a difference between what IT people think such software can do and the potential business case relevant for the adoption. This shows that while one group of people might want to use and play around with the software, other people within the firm might want to solely use it for a business case. This way to approach AI and Big Data shows that the knowledge and capabilities within teams within the firm might differ.

Usability of the software tool

The usability of the software tool was discussed within every interview, thus a total of three times. Here, the discussion was mostly around how the tool could be used, how the results were portrayed, and how the data might be used. The responses of interviewees differ, as one stated that the software tool did not provide any value, as similar products can be created in-house. One interviewee stated that the tool could prove useful. The last interviewee stated that the expectations regarding the tool were high, and expected more of it. Furthermore, as one interviewee stated, there is a difference in what the interviewee wanted to achieve through this tool, what is currently available on the market, and what is possible with the software behind the tool.

Learning and adoption perspectives

The learning and adoption perspectives were discussed within all three interviews. The interviewees agree that there is a link between knowledge sharing, how technology is being presented, and the adoption of such a technology. However, as one interviewee stated, it's also about the size of the program (which uses this technology) which bring different challenges with it in regard to learning and adoption. The "adoption is bigger, it's getting more complex". If more people are involved and the procedures to be changed are bigger, so does the complexity of how such technologies should be taught.

Big Data and AI to reduce manual labour

This theme was discussed within two interviews. The first time it was discussed that such technologies definitely have the capability of making one's life easier. Stating that "the use of big data is the future, we have to go there, we can't do this (the data retrieval process) manually". Even illustrating that certain data retrieval processes could become so big that manual labour is out of the question.

Data driven approach

This theme is related to the portrayal of the output. This has been discussed within the interviews a total of one time. Here, it was discussed that a data-driven approach might be used to convince people, that you, as an expert, are not an expert unless you can provide a solid background of information. Thus, that a data-driven approach is a way to go to present results besides one's expertise.

Big Data and AI as nonbiased view

The discussion regarding Big Data and AI as a nonbiased way to approach an information process has been held within every interview, thus three times. The common denominator within these interviews is that Big Data and AI are software solutions, and part of a process. These software solutions and the process as a whole might include bias through the designing of a process. This perception has also been illustrated in the theme "Expert opinion as data source". Within this theme, however, the reflection was also made from the perspective of the receiver of such information, where one interviewee stated that it has the 'gravitas' to convince people that this is sort of an unbiased way to approach information sharing.

Big Data and AI as a spearpoint to push information

This theme has been observed three times. One of the best analogies given to how such technologies are currently used is given by one interviewee. It is used as "we see it as a stick to hit the dog, we are already planning to hit the dog, but now we got a stick. It is necessary to hit the dog". This is interpreted in two ways. That it can be used as a spearpoint to push information, like Big Data and AI can create a lot of useful information, and it also has a convincing factor, further addressed in the next theme. Furthermore, it has been addressed during one interview that even as one pushes information such as management information, not substantiated in Big Data and AI, this information will be challenged.

Big Data and AI just for convincing people

This theme has been observed once. As stated within the previous theme, it is an approach to shake people up with information. If one uses Big Data and AI, the notion of Big Data and AI alone can convince people that the information provided by such a methodology using Big Data and AI is correct.

The approach of reporting on the Big Data and AI tool

The approach of reporting on the Big Data and AI tool has been mentioned once. This is best encapsulated in the following quote regarding the correlation between reporting on such technologies and the adoption of one: "I've experienced over the years is that it's extremely big, but you are restricted by the people that surround it, you see it also here, you get stuck on definitions, how definitions are formatted, instead of the data definition. The data is not the problem, its available, but how you use it and how you do it, and how to instruct people to use it is difficult". Thus, showing that how one reports is of importance regarding such technologies.

People looking only at method, not data quality and source

This theme was an integral part of the interviews, as the question was asked related to this theme. It, non-surprisingly, was observed in all the interviews. The results were interesting, as it illustrates someones perception of the process. The common denominator was the "Garbage in garbage out" principle. Furthermore, one interviewee stated that if one of the two is garbage, you can not get a correct answer. When reflected on the firm the data quality is of more importance, however, another interviewee stated that the method is more important.

Not invented/purchased here syndrome

What was discussed numerous times within the interviews was that most software within the firm is designed in-house, or adapted to it. This, due to the not invented here syndrome. Meaning that people only like things they had a say in designing. The interviewees said that people within the firm are more likely to reject software and software solutions if it was not invented here.

AI is received with scepticism

This theme was discussed in all three interviews. One interviewee regarded, for example, the software used within the procedure as not that relevant. A comparison is made with Excel in a couple of years. Furthermore, AI is also received with scepticism due to it being a black box, one does not know that is under the hood of the software.

Even in a Big Data and AI process, expert opinion is needed

This theme was discussed within all the interviews. This, as every process in the end is designed by people, putting their bias into the process. Furthermore, results need to be assessed, input has to be assessed, and decisions have to be made regarding data used and data sources.

Big Data and AI are used as buzzwords

This theme was observed once during one interview. It was discussed that these terms were used as buzzwords for convincing people. This could say multiple things in regard to the perception of Big Data and AI, as the receiving end is convinced by it.

Perception of Big Data and AI is that it is near mystical

This theme was observed once during one interview. It was observed during the same interview as the previous theme. There has been overlap in the observation as it has some correlation. If Big Data and AI are perceived as near mystical, it can also be used as buzzwords.

The perception of Big Data & AI is 'state of the art'

This theme has been observed twice during the interview sessions. Both times it was discussed that it could be perceived as something new and impressive, state of the art so do speak. It was perceived as something that could reduce the time worked on some tasks.

Company culture

This was an integral part of the interviews, as it was asked directly. Thus, it has been observed three times. There have been many ways to approach this subject. One interviewee stated that on the front end a lot has to be done to make sure it is accepted within the firm, due to the culture within the firm. Another interviewee stated "Culture is everything!", that with a strong culture the adoption of technologies such as Big Data and AI might be hampered, stating "they only going to do it when they have to".

Generational aspects of acceptance of Big Data and AI

This theme was observed three times during the interview sessions. This theme entails more general aspects of the acceptance of Big Data and AI not illustrated by any specific theme. An example of this is that when a firm uses external sources for such a process of Big Data and AI, it is generally more accepted. Or the conservativeness of the average employee in a financial institution, that this clashes with the introduction of technologies such as Big Data and AI.

Combination of business implementation and technology

This theme was observed within all three interviews, as it aims to illustrate the business implementation of technology and aims to catch relevant information within this theme. An interesting view was that this kind of software can be used for predictions. Furthermore, it could be used as an addition to the in-house data a financial institution already has. Seeing this as an addition of AI as an analysis method.

Another observation regarding this theme is that it could be implemented through current other ESG reporting methods such as GRI and SASB.

Big Data and AI is perceived as (not) valuable

This theme aims to illustrate the perception of Big Data and AI as something (not) valuable. This theme was observed within all interviews. This, as it was asked to the interviewees. The interviewees had a different opinion on the software as used within the procedure. One stated that it was not that valuable, as the software used was overpriced and did not create that much value. Another interviewee stated that he/she thought that it could provide value in the future. The third interviewee stated that it could create value for other firms in society, but perhaps not that much for a big and conservative firm as the one where the case study took place.

ESG data as a precursor for financial data

This theme was observed once during one interview. It was a discussion to what ESG data actually portrays. The interviewee stated that there might be a correlation between ESG data and financial data, as it could be a precursor to financial data.

Future of the combination ESG, Big Data, and AI

This theme was observed within all interviews. This, as it was asked to the interviewees to what the future of this combination of technology and data entails. The answers were that it could become valuable to financial institutions, combining internal data and AI. However, one interviewee stated that external data and AI as being used by such a firm could be a step too far, meaning that they are not going to use this. Another point made was that it is used to reduce manual labour. Another thing to take into account is that there is upcoming regulation. This could also affect the combination of ESG data & technology.

Appendix C – Interview procedure

Goal of the interview

The main goal of the interview is to explore the perception of the interviewee regarding Big Data and AI, the procedure, and the integration, perception, and operability within the firm. It will assess the procedure as designed and introduced to the interviewee.

Start interview

Ask permission for the recording of the interview.

Start with a welcome, give more background regarding the research and interviewer. Then explain here the procedure, show here the procedure and software related to the procedure.

Open question regarding what they have seen

Let them give their opinion on what they've seen, ask through if they state something interesting.

Basic questions

What is your experience with such technological advancements? E.g. big data and AI

Perception and adoption

How do they think such technologies are being perceived as?

How could these technologies add value to a firm?

Do they think the usage of a Big Data and AI process adds credibility, and why?

Will such technology or framework be widely accepted within the bank, and why?

How does (corporate) culture play a role in such processes?

Learning and sharing

How does learning and knowledge sharing fit within this framework, e.g. the acceptance of this kind of software or perhaps the critical stance regarding Big Data and AI?

Ask how learning and adaptation perhaps play a role, look into perception of such technologies?

Data and technology & ESG

Do you think that ESG data differs from any other input?

Is the method more important or the data in such a process?

Do you interpret the results as unbiased? And if so (not), why (not)? (can be asked as ''are the results unbiased?"

How do you think this kind of technology could be used in the future?

Appendix D – Literature research methodology

The aim of this literature research methodology is to address the approach taken to come to the information as provided in the Literature background sub-chapter of chapter 1.

Scale and scope of the literature review

ESG data is currently being adopted by many fintechs, showing that the rise of sustainability-related data is at the forefront of the financial industry. This, in combination with novel data technologies such as Big Data and AI is portrayed as something to increase one's competitive advantage. Firms providing these SaaS products, which effectively all of them are, are for example Datamaran or Refinitiv. As firms are already implementing the combination of ESG data, Big Data and AI, the initial impression and approach to the literature research are to find and provide a holistic overview of the available data. The research itself revolves around mostly the practical and reflective fields of science.

Search description and selection criteria

During the review of relevant literature, the following criteria are adhered to for the inclusion of literature:

- Literature is peer-reviewed.
- Literature is written in the English language.
- Literature is written by an author with expertise and relevant credentials in the field of AI, and/or Big Data, and/or ESG, and/or ESG data.
- Literature can take many forms, theoretical or experimental, preferring experimental due to the research in the application of the research topic.
- Literature mainly focuses on the application of AI, and/or Big Data, and/or ESG, and/or ESG data within a firm
- Literature is published within the last 10 years, to contain the relevancy of this novel field of research.

Due to the novel field of this research, slight deviations regarding the focus or author preference might be taken. Otherwise, literature is excluded from being used within this study of available literature.

Topic relevance of publications

The scope of relevant literature is defined within this sub-chapter. Literature has to be topically relevant, within the many facets of literature itself such as the title, abstract, or keywords. These must relate to ESG data, Big Data or AI.

ESG data

Literature regarding ESG data is in scope with this literature review. Here, the lens is on what it might entail, what the theoretical foundations might be, or even how it is structured. The research aims to seek to answer the following questions:

- How is ESG data used
- On what is ESG data-based?
- What can be said about the perception and integration of ESG data?

Potential keywords: *ESG, ESG data, theoretical foundation AND ESG data*

Novel data technologies

To analyse the topics of ESG data, Big Data and AI, these are grouped together within the notion of novel data technologies. As there has been a significant amount of research already on each individual topic. Thus, the focus of it will be on the implementation and perception of it within a firm. The research aims to answer the following question:

- How can Big Data and AI be used for ESG data?
- What can be said about this integration of Big Data, ESG data, and AI?
- What is the perception of ESG data, Big Data, and AI when it is being applied within a firm?
- What are the strategies for implementing one or more of the facets of ESG data, Big Data and AI within a firm?

Potential keywords: *ESG Data, Big Data and AI, integration AND Big Data and AI, perception*

Big Data and AI operationalized within a firm

As the integration of ESG data is mentioned in the previous sub-chapter, the facets of Big Data and AI as operationalized within a firm should also be addressed. Here, the focus should be more on the operationalization and integration of Big Data and AI. The research aims to answer the following questions:

- What can be said about the operationalization of Big Data and AI in a firm?
- How are Big Data and AI used within a decision-making process?
- How does this integration look like?
- Can something be said about the setting of the firm when integrating Big Data and AI?

Potential keywords: *Big Data and AI, operationalization, integration, decision-making process*

The perception of Big Data and AI

As the previous sub-chapter illustrates the topic relevance of the integration of Big Data and AI, the research direction regarding the perception of Big Data and AI goes one step deeper on a conceptual level. As these technologies are integrated within a firm, this means that employees are learning, using, and applying these technologies. Thus, their perception of these technologies plays a role in these facets. Meaning that it is all based, in the end, on their interactions with it. The research aims to answer the following questions:

- What kinds of social perspectives regarding the adoption of Big Data and AI can be given?
- How are these technologies embraced?
- What can be said about how people within a firm look at these technologies?
- What can be said about the adoption and perception of these technologies?

Potential keywords: *Adoption, perception, embracing Big Data and AI, integration of Big Data and AI*

Search descriptions

The research for relevant literature was done through the following three databases. First Google Scholar, then Scopus, and last Sciencedirect. The decision was made to exclude Web of Science, as previous attempts to use this database were unsuccessful due to problems with accessibility. The search description creation and search process was an organic one, where some descriptions were defined through the previously found results. The table below illustrates the search query, database used, number of results, and the research motivation. The keywords are ordered from first to last used.

Search query	Database	No. of results	Research goal or motivation
ESG data	Google scholar	107000	Broad sweep regarding ESG data
	Scopus	873	
	Sciencedirect	204	
"sustainability metrics" OR "sustainable metrics"	Google scholar	10400	To continue on the previous search query, where a paper was found where a similar connotation was used for ESG data, sustainability metrics
	Scopus	699	
	Sciencedirect	1221	
"ESG Metrics"	Google scholar	958	To find a broad sweep of literature, based on something read in a previous paper from the previous search descriptions.
	Scopus	8	
	Sciencedirect	25	
"ESG data" AND ("AI" OR "Big data")	Google scholar	646	Broad sweep regarding ESG data
	Scopus	3	
	Sciencedirect	10	
"Integration" AND "AI" AND "Big Data"	Google scholar	71500	To find more general literature regarding the integration perhaps applicable
	Scopus	243	
	Sciencedirect	5637	
"Integration" AND ("ESG Data" or "ESG") AND "AI" AND "Big Data"	Google scholar	68	To find relevant literature regarding the integration of this technology in combination with ESG
	Scopus	1	

	Sciencedirect	2	
"operationalization " AND "Big Data" AND "AI"	Google scholar	3080	A look into how this combination of technology could be operationalized
	Scopus	2	
	Sciencedirect	2798	
"operationalization " AND "Big Data" AND "AI" AND "ESG"	Google scholar	33	A continuation of the previous search query.
	Scopus	0	
	Sciencedirect	14	
"sustainability measurement systems" OR "ESG measurement systems"	Google scholar	113	To take a detour, finding perhaps ESG related IT systems
	Scopus	7	
	Sciencedirect	25	
("sustainability" OR "ESG") AND ("Big Data" OR "AI")	Google scholar	1520000	To find literature regarding Big Data and AI in combination with sustainability
	Scopus	1818	
	Sciencedirect	94144	
"ESG" AND ("AI" OR "Big data")	Google scholar	26600	Looking at the combination of Big Data and AI, and ESG data.
	Scopus	13	
	Sciencedirect	748	
"Trend analysis" AND ("AI" OR "Big data")	Google scholar	37800	Looking to see what can be said about predicting trends and the operationalization of Big Data and AI
	Scopus	278	
	Sciencedirect	2208	

("Trend analysis" OR "Market analysis) AND ("AI" OR "Big data") and ("Sustainable" OR "Sustainability")	Google scholar	17000	Looking for how analytical technology can be used to assess value, including sustainability. This was a side-track to see how it could be applied for management information.
	Scopus	17	
	Sciencedirect	936	
"sustainability indices" AND "Big data"	Google scholar	383	Looking for the combination of big data and sustainability
	Scopus	8	
	Sciencedirect	15	
"ESG" AND "AI" AND "Big data" AND "Management"	Google scholar	1110	Looking for a link to match Big Data and AI with management information
	Scopus	1	
	Sciencedirect	21	
"big data" AND "long term" AND "strategic" AND "ESG" AND "Decision" AND "strategic Decision making"	Google scholar	76	Looking for the link between long term decision making and ESG.
	Scopus	0	
	Sciencedirect	0	
"big data" AND "adoption" AND "board room"	Google scholar	245	To see if the adoption of Big Data and AI is researched within a board room of a company
	Scopus	0	
	Sciencedirect	3	
"integrated thinking" and "Big data" and "AI"	Google scholar	90	Based on a company meeting, to see if integrated thinking literature could help and bring novel perspectives
	Scopus	0	
	Sciencedirect	0	

"Alternative data" and "decision making"	Google scholar	15400	To see if alternative data is used for decision making. This as it was previously found that ESG data is based in alternative data.
	Scopus	73	
	Sciencedirect	1196	
"Alternative data" and "management"	Google scholar	34700	Here the goal was to see how alternative data and management fit together
	Scopus	219	
	Sciencedirect	2638	
"Alternative data" AND "management" AND "AI"	Google scholar	5170	Continuation on previous topic including AI.
	Scopus	2	
	Sciencedirect	199	
"implementing" AND "big data" AND "ai" AND "alternative data"	Google scholar	227	Continuation on the previous topic including implementation.
	Scopus	1	
	Sciencedirect	39	
"big data" AND "AI" AND "ESG" AND "case study"	Google scholar	220	The idea of the possibility of a previous case study was researched
	Scopus	0.	
	Sciencedirect	8	
"big data" AND "AI" AND "case study" AND "Financial"	Google scholar	25000	Looking for a Big Data case study in a financial institution. This, due to the scope of the research opportunity within a financial institution.
	Scopus	6	
	Sciencedirect	923	
	Google scholar	3890	

"big data and ai" AND "Social"	Scopus	19	Assessing the social aspects of Big Data and AI.
	Sciencedirect	248	
"Social aspects" AND "big data and ai"	Google scholar	101	Assessing the social aspects of Big Data and AI.
	Scopus	0	
	Sciencedirect	14	
"Human perception" AND "big data and ai"	Google scholar	57	Looking for the perception of Big Data and AI
	Scopus	0	
	Sciencedirect	5	
"Perception" AND "big data and AI"	Google scholar	1330	Broad sweep looking for the perception of Big Data and AI
	Scopus	2	
	Sciencedirect	125	
"employee" AND "big data and ai"	Google scholar	793	Looking for how employees would perceive Big Data and AI.
	Scopus	2	
	Sciencedirect	96	
"ESG" AND " Big data and ai"	Google scholar	36	Broad sweep to check if anything relevant was missed. Backtracking some papers.
	Scopus	0	
	Sciencedirect	2	
"ESG" and "Culture"	Google scholar	34600	Looking for a correlation between ESG and culture, as this came up during a company meeting
	Scopus	67	

	Sciencedirect	1145	
"ESG Data" AND "Big Data" AND "AI"	Google scholar	92	A more general search, realizing that at the end of the literature research this combination was not yet done.
	Scopus	2	
	Sciencedirect	2	

Search Results

A total of 17 papers have been found tangential to the research topic of the application of ESG data, and Big Data and AI. These papers were relevant and within scope. The results have been used within the Literature background sub-chapter to illustrate the knowledge gap addressed within this thesis. These results are a combination of the "hits" within each database and by assessing literature and follow through on what and whom each paper cited.