Program: Industrial Engineering specialization in leadership and work organization
Title: Explainable AI and transparency in AI systems for SMEs
k(PCXFörklarbar AI och transparens i AI system för SMEs
Year of publication: 2024
Authors: Herman Beronius & Hilda Malmfors
Supervisor: : Chandadevi Giri
Examinator: Jonas Waidringer
Keywords: Artificial Intelligence (AI), AI black box, Explainable AI (XAI), AI Trust, AI acceptance, AI transparency, SME

Preface

We would like to express our gratitude to our supervisors for their invaluable guidance and support throughout this research endeavor. We would also like to thank the participating case companies for their cooperation and for providing the information and insights that have enriched this bachelor thesis. Lastly, we want to express our appreciation for the opportunity to participate in this exciting and educational research project on AI implementation in SMEs at the University of Borås, in collaboration with the case companies.

Herman Beronius *Hilda Malmfors*

Abstract

The study examines how explainable AI (XAI), and transparency can increase trust and facilitate the adoption of AI technologies within small and medium-sized enterprises (SMEs). These businesses face significant challenges in integrating AI due to limited technical expertise and resources. The purpose of the study is to explore how XAI could bridge the gap between complex AI models and human understanding, thereby enhancing trust and operational efficiency.

The research methodology includes a case study with a literature review and expert interviews. The literature review provides background and context for the research question, while the expert interviews gather insights from employees in various roles and with different levels of experience within the participating SMEs. This approach offers a comprehensive understanding of the current state of AI adoption and the perceived importance of XAI and transparency.

The results indicate a significant knowledge gap among SME employees regarding AI technologies, with many expressing a lack of familiarity and trust. However, there is strong consensus on the importance of transparency and explainability in AI systems. Participants noted that XAI could significantly improve trust and acceptance of AI technologies by making AI decisions more understandable and transparent. Specific benefits identified include better decision support, increased operational efficiency, and enhanced customer confidence.

The study concludes that XAI and transparency are crucial for building trust and facilitating the adoption of AI technologies in SMEs. By making AI systems more comprehensible, XAI addresses the challenges posed by limited technical expertise and promotes broader acceptance of AI. The research emphasizes the need for continuous education and clear communication strategies to improve AI understanding among stakeholders within SMEs.

To enhance transparency and user trust in AI systems, SMEs should prioritize the integration of XAI frameworks. It is essential to develop user-centered tools that provide clear explanations of AI decisions and to invest in ongoing education and training programs. Additionally, a company culture that values transparency and ethical AI practices would further support the successful adoption of AI technologies. The study contributes to the ongoing discourse on AI adoption in SMEs by providing empirical evidence on the role of XAI in building trust and improving transparency. It offers practical recommendations for SMEs to effectively leverage AI technologies while ensuring ethical and transparent AI practices in line with regulatory requirements and societal expectations.

Sammanfattning

Studien undersöker hur förklarbar AI (XAI) och transparens kan öka förtroendet och underlätta införandet av AI-teknologier inom små och medelstora företag (SME). Dessa företag står inför betydande utmaningar vid integrationen av AI på grund av begränsad teknisk expertis och resurser. Syftet med studien är att undersöka hur XAI kan överbrygga klyftan mellan komplexa AI-modeller och mänsklig förståelse, vilket i sin tur främjar förtroende och operationell effektivitet.

Forskningsmetodiken inkluderar en fallstudie med en litteraturöversikt och expertintervjuer. Litteraturöversikten ger bakgrund och kontext till forskningsfrågan, medan expertintervjuerna samlar insikter från anställda i olika roller och med olika erfarenhetsnivåer i de deltagande SMEs. Detta tillvägagångssätt gav en omfattande förståelse av det nuvarande tillståndet för AI-adoption och den upplevda vikten av XAI och transparens.

Resultaten visar på en betydande kunskapslucka bland SME-anställda när det gäller AI-teknologier, med många som uttrycker en brist på bekantskap och förtroende. Det råder dock stark enighet om vikten av transparens och förklarbarhet i AI-system. Deltagarna angav att XAI avsevärt kunde förbättra förtroendet och acceptansen av AI-teknologier genom att göra AI-beslut mer förståeliga och transparenta. Specifika fördelar som identifierades inkluderar bättre beslutsstöd, ökad operationell effektivitet och ökat kundförtroende.

Studien drar slutsatsen att XAI och transparens är avgörande för att skapa förtroende och underlätta införandet av AI-teknologier i SME. Genom att göra AI-system mer förståeliga adresserar XAI utmaningarna med begränsad teknisk expertis och främjar en bredare acceptans av AI. Forskningen understryker behovet av kontinuerlig utbildning och tydliga kommunikationsstrategier för att förbättra AI-förståelsen bland intressenter inom SME.

För att öka transparensen och användarförtroendet i AI-system bör SME prioritera integrationen av XAI-ramverk. Det är viktigt att utveckla användarcentrerade verktyg som ger tydliga förklaringar av AI-beslut och att investera i kontinuerliga utbildnings- och träningsprogram. Dessutom kommer en organisationskultur som värderar transparens och etiska AI-praktiker ytterligare stödja det framgångsrika införandet av AI-teknologier. Studien bidrar till den pågående diskursen om AI-adoption i SME genom att tillhandahålla empiriska bevis på rollen av XAI i att bygga förtroende och förbättra transparens. Den erbjuder praktiska rekommendationer för SME att effektivt utnyttja AI-teknologier, och säkerställa etiska och transparenta AI-praktiker som är i linje med regulatoriska krav och samhälleliga förväntningar.

Terminology

AI – Artificial Intelligence – involves the development of computer systems capable of performing tasks that typically require human intelligence, such as learning from data, recognizing patterns, and making decisions. AI combines algorithms, data, and computational power to imitate human cognitive functions.

AI black box – refers to AI systems whose internal workings are not transparent or understandable to users or developers. This opacity makes it difficult to discern how the AI arrives at its decisions or predictions, raising concerns about accountability and trustworthiness.

XAI – Explainable AI (Explainable Artificial Intelligence) – is a field of AI that focuses on making the decisions of AI systems understandable to humans. It addresses the black box and aims to provide insights into the reasoning processes of AI models, ensuring transparency and trustworthiness in their decision-making.

AI Trust – refers to the confidence that users, developers, and other stakeholders have in the reliability, fairness, and transparency of AI systems. It encompasses the assurance that AI systems will perform as expected, without bias or unforeseen consequences, and that their decision-making processes are understandable and explainable.

AI acceptance – is the willingness of individuals, organizations, and society to adopt and integrate AI systems into their daily operations and decision-making processes, based on the perceived benefits, usability, and alignment with human values and needs.

AI transparency – refers to the practice of designing AI systems in such a way that their operations and decision-making processes are clear and understandable to users, stakeholders, and regulators. This openness helps ensure that AI actions could be audited and trusted, fostering accountability and ethical use of technology.

SHAP – (Shapley Additive Explanations) – is a method used in XAI to interpret the output of machine learning models. It assigns each feature in a model a contribution value (Shapley Value) based on cooperative game theory, explaining the impact of each feature on the prediction by considering all possible combinations of feature values.

LIME – (Local Interpretable Model-agnostic Explanations) – is an algorithm used to explain the predictions of any machine learning model. It works by approximating the model locally with an interpretable model, like a linear model, around the prediction of interest.

SME – Small and Medium-Sized Enterprises – are businesses defined by their size in terms of number of employees, turnover, or assets, and constitute most businesses in many economies, including Sweden.

CONTENTS

1. INTRODUCTION

This chapter provides a background to the study and the problem area it encompasses. Furthermore, the purpose of the study and the research question are presented, as well as the limitations.

1.1 Background

During the era marked by the fourth industrial revolution, also known as Industry 4.0, we are experiencing a significant transformation of the manufacturing industry through digitalization and the internet. This transformation has led to SMEs facing challenges and opportunities like never before. At the center of this transformation stands AI – a technology that forms a key part for smart manufacturing and the streamlining of industrial processes. AI is defined as a system's ability to correctly interpret external data, to learn from such data, and to use those learnings to achieve specific goals through flexible adaptation (Haenlein & Kaplan 2019). This technology offers numerous opportunities for business operations, including improved decision-making, inventory planning, and optimization of manufacturing processes (Rauch, Dallasega & Unterhofer 2019).

Despite AI's potential, its implementation within SMEs is not without challenges. Lack of technical competence and organizational difficulties are just some of the obstacles SMEs face when it comes to integrating AI into their operations (Issa, Lucke, and Bauernhansl 2017). To overcome these challenges and fully utilize the potential of AI, it is necessary to develop strategies and methods that are customized for the unique needs and conditions of SMEs. As AI technology develops and becomes more integrated across various sectors, the issue of transparency and understanding of these systems has become increasingly important. Here, the importance of XAI emerges. XAI is a field of AI and aims to make the decision-making processes of AI systems more transparent and understandable, which is crucial for building trust and safety in the use of these systems, especially in critical areas where incorrect decisions could have major consequences (Khan et al. 2022). EU legislation, which mandates that AI must be explainable in terms of decisions affecting the citizens of the European Union, has enhanced the relevance of this research. (ibid.). This legislation underscores the need for research and development in the field of XAI, not only to meet legal requirements but also to promote trust and acceptance of AI technology among the public.

Research within XAI has been extensive and multifaceted, including the development of neutrally inspired theoretical frameworks and the exploration of concept-supported interpretations to bridge the gap between complex models and human understanding (Chan 2022; Lecue 2020; Medianovskyi & Pietarinen 2022; Khoozani, Sabri, Seng, Seera & Eg 2024). Recent research has highlighted the challenges of evaluating and selecting effective XAI methods and emphasizes the complexity of XAI and the many considerations required for its successful implementation (Stassin, Corduant, Mahmoudi, Siebert 2024). This thesis builds upon existing research and aims to further investigate the role of XAI within SMEs. By exploring and assessing various XAI methods and their applications, it aims to understand how these methods could improve trust and safety in the use of AI within SMEs. Through this comprehensive analysis, the hope is to contribute to the ongoing conversation around AI and XAI and provide insights that could be beneficial for developers and decision-makers within the AI field, with a particular focus on SMEs.

1.2 Problem Statement

Despite the rapid advancements and potential of AI to revolutionize industrial sectors, SMEs are facing significant challenges regarding the transparency and explainability of AI. These challenges are especially pronounced in SMEs where resources are limited, and technical expertise is often lacking. This situation creates a divide between the theoretical capabilities of AI and its practical applications within these companies. Although AI could offer improvements in efficiency and decision-making, SMEs are unable to fully use these advantages due to a lack of understanding and trust in the operational processes of AI systems (Issa, Lucke, and Bauernhansl 2017).

Where AI is increasingly integrated into core business operations, the need for XAI becomes apparent. XAI aims to make the decision-making processes of AI systems more transparent and comprehensible, which is crucial for building trust and safety in the use of these systems. This need is not only a technical necessity but also an ethical and legal concern, especially in light of regulations such as the European Union's General Data Protection Regulation (GDPR), which demands that AI is explainable, particularly when its decisions impact citizens (Saeed & Omlin 2023).

The study focuses on exploring XAI and transparency within SMEs to improve trust and facilitate the use of AI technologies. By analyzing how explainability and transparency could influence users' attitudes and behaviors towards AI, the study aims to find out the key factors contributing to a successful implementation and integration of AI solutions. Through this research, the hope is to identify effective alternatives that could overcome existing barriers and promote a more sustainable use of AI within the business sector. The goal is to contribute to a broader understanding of the role of XAI in creating more transparent, accountable, and more reliable AI systems, which is crucial for SMEs striving to navigate in the rapidly changing technological landscape.

1.3 Purpose and research questions

As stated in the background and the problem statement, SMEs face considerable challenges concerning the transparency and comprehensibility of AI systems, which restricts their capacity to fully exploit the potential advantages offered by AI technologies. The purpose of the research is therefore to:

Investigate the impact of explainable AI and transparency on enhancing trust levels and streamlining the adoption of AI technologies in SMEs.

This aims to provide a comprehensive picture of how these capabilities interact and influence the use of AI technologies. By understanding the impact of XAI and transparency on trust levels, SMEs could make more informed choices about integrating AI technologies into their operations. The research will be operationalized through three research questions, where the first question aims to gather insights into existing research, while the other two are expected to be answered by the research study. The first question was formulated as:

[1] What is the current state of research on explainable AI and trust in the context of SMEs?

Through a systematic review and synthesis of existing research, the first question endeavors to provide valuable insights into the theoretical bases and practical implications of XAI and trust for SMEs. By doing so, it lays the foundation for the subsequent research questions and helps to advance scientific understanding in this critical area of research. The second question was formulated as:

[2] Does explainable AI and transparency affect SMEs attitudes towards using AI technologies?

This question delves into the crucial aspect of user perception and acceptance of AI technologies in relation to the presence of XAI and transparency. By exploring this question, the study aims to uncover the extent to which these capabilities influence users' attitudes and behaviors regarding the adoption and utilization of AI technologies. The third is described as:

[3] Does improved trust and understanding help to facilitate the adoption of AI technology within SMEs?

This question examines the role of trust and understanding in facilitating the adoption of AI technologies. It aims to understand whether increased trust and comprehension among users lead to a more flexible implementation of AI solutions. By investigating this question, the study aims to highlight the capabilities that contribute to the successful integration of AI technologies.

1.4 Delimitations

This research examines the impact of XAI and transparency to influence levels of trust and facilitate the adoption of AI technologies in SMEs. Given the broad and complex nature of AI systems (Issa, et al. 2017), delimitations are necessary. The delimitations drawn are therefore aimed to reduce complexity and scope, ensuring focused and manageable analysis.

While methodologies and strategies for implementing AI technologies in SMEs will be presented, they will not be verified or tested in this study. The research focuses on providing insights and recommendations based on existing literature and business opinions, aiming to inform decision-making processes regarding the adoption and utilization of AI technologies in SMEs. The study will furthermore focus on the trustworthy factors rather than the structure and functions of AI systems. The study will also discuss numerous aspects and factors relevant to the topic but will then be delimited towards the purpose of the study which focuses on SMEs, in order to effectively answer the study's research questions. The study will not focus on AI as a broad topic, but instead emphasize the area of XAI and transparency in the systems.

2. METHODOLOGY

This chapter outlines the methodology, encompassing a delineation of the selected approach and a rationale behind its suitability for the study. The research adopts a case study approach, employing both a literature review and an expert interview method where the approach will be described.

2.1 Research Design and Strategy

To address the research questions, a case study will be applied. It is a combination of a literature review and an expert interview method, in order to explore the topic. The literature review is intended to address the initial research question and provide a background for the other two. It relies on an extensive review of existing research and literature, mainly academic articles. Through a systematic examination of existing literature, the review aims to identify key themes, theories, methodologies, and findings pertinent to the research inquiries. Additionally, it serves to contextualize the study within the broader academic discourse and to identify gaps or areas requiring further investigation (Thomas 2021). By synthesizing existing knowledge and insights, the literature review lays the foundation for the subsequent phases of the research process, including the implementation of the expert interviews. The expert interviews have a structured and iterative approach that involves gathering input from a panel of experts through two series of questionnaires and rounds of discussion. The combination of the literature review and the expert interviews provides a broad approach to addressing the research questions, drawing upon existing knowledge and expertise to generate new insights and recommendations (Patel & Davidson 2019). The connection between the study´s research questions and the method are illustrated in Figure 1.

Figure 1. The research questions related to the case study.

2.2 Case Study Approach

The case study provides a form of investigation that elevates a view of life in its complexity. It enables the examination and application of a broader research object to a smaller research group, to make it manageable to study (Thomas 2021). The emphasis in a case study is about the particular rather than the general, which means focusing on the specific case and analyze it from all their angles. The case study is thus not a methodological choice but a choice of what is to be studied, the specific case. Then there is a methodological choice to help inquire into the subject (ibid.). In this study, the specific case chosen is the importance of XAI and transparency in AI systems for the implementation of AI technologies. Patel and Davidson (2019) elucidate that it is possible to encompass the examination of multiple cases, such as several organizations.

They also delineate the method as beneficial for examining processes and changes, which also could be applied to this certain study. Patel and Davidson (2019) present an alternative perspective to Thomas (2021) on generalizability and describes that it could vary depending on how the case is selected. This could be exemplified by the involvement of three distinct organizations that contributed insights to this study, enterprise A, B and C. This aspect could be deliberated upon to represent the broader population of SMEs. The interaction is illustrated in Figure 2. The case study examines the selected case but also uses a literature review to provide an overview of the research.

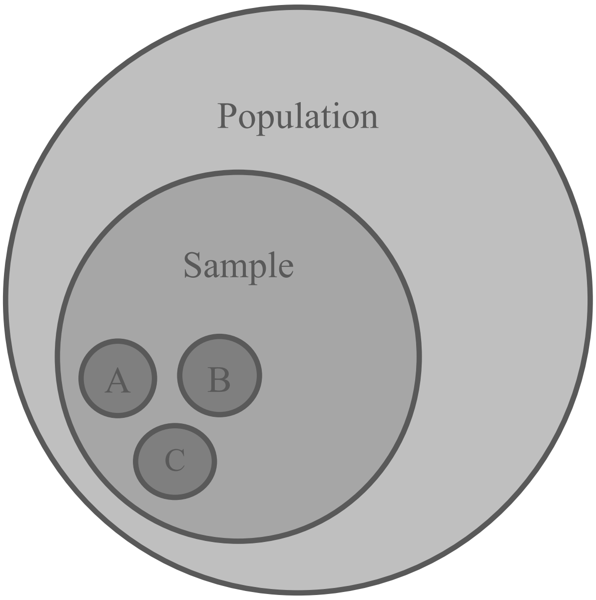

Figure 2. The figure illustrates the connection between the population and the study sample.

2.2.1 Selection of Case Enterprises

In this case study, data from three specifically selected companies, referred to as Enterprise A, Enterprise B, and Enterprise C, was collected and analyzed using an expert interview method. These companies represent various aspects of Swedish industry and vary in their levels of engagement and maturity in the use of AI. Each company has expressed intentions to implement AI in their operations, which guided the selection for this study. The choice of these companies is based on their potential to provide valuable insights into how AI could be implemented to meet unique industrial challenges and opportunities. This is crucial for the research project that focuses on AI within SMEs, specifically targeting XAI and transparency within AI technologies.

The participation of these companies enriches the study by exploring their perceptions of integrating XAI into their processes. By analyzing their feedback and experiences, the study aims to uncover how these companies view the necessity and benefits of XAI. This not only helps to understand the industry's readiness and the barriers to adopting such technologies but also provides a clear view of the expected impacts of XAI on operational transparency and decision-making processes.

2.2.2 Profile of Case Enterprises

The different case enterprises possess distinct needs and objectives for AI implementation, and they could utilize XAI and its potential in various ways. This is illustrated in Table 1. Each enterprise operates in a unique industry, but still within the SME sector, which drives its specific AI implementation considerations and goals. For instance, an online clothing retailer aims to personalize customer experiences, while a manufacturer of advanced laundry systems focuses on optimizing operational efficiency. Meanwhile, a producer of customized metal products looks to enhance its design and manufacturing processes. By employing XAI, these enterprises could achieve greater transparency in their AI applications, fostering trust and confidence among customers and stakeholders. This transparency is crucial for explaining AI-driven decisions, whether it's recommending a product, optimizing a system, or clarifying production methods.

Table 1. Case Study Enterprises and XAI Potential Benefits.

Enterprise	Industry Focus	AI Implementation Consideration	Potential Benefits of XAI
Enterprise A	Online sale of premium men´s clothing	Considering AI to enhance costumer experience by personalizing shopping experiences.	Transparency through XAI could build costumer trust by clarifying AI-driven product recommendations and style advice.
Enterprise B	Design and manufacture of advanced laundry systems	Considering AI to optimize operational efficiency.	Transparency in how AI models make decisions to build trust in the technology.
Enterprise C	Manufacturer of customized metal products	Considering AI to improve design and manufacturing processes.	Transparency in AI decisions could strengthen trust by clarifying processing methods and material usage, enhancing operational efficiency and customer satisfaction.

2.3 Structured Literature Review

The literature chosen is intended to provide an overview of research for this study. Thomas (2021) states that the literature review is a crucial element in any research, even in a case study. This provides a background to the subject and helps to formulate ideas about the object of study. Thulesius (2021) emphasizes the significance of the capacity to impartially scrutinize existing research. This is due to the tendency of researchers to possess ideas and preconceptions about the subject of their research. To objectify the research subject and contextualize it, a thorough and extensive search was conducted on the "Web of Science" database to explore the topic within the context of SMEs.

The topics and factors considered relevant within the SME context were subsequently employed for additional exploration. These were once more queried on the "Web of Science" database as keywords, but this time with a broader focus on the topic in general. Data filters for year of publication were also used as well as articles only. The articles were then selected according to relevance and the data was collected. This iterative data collection method was applied repeatedly for each keyword, resembling the cyclical process depicted in Figure 3. After a systematic process of collecting and refining data, the literature review was carefully assembled to provide a thorough overview of relevant research for this study.

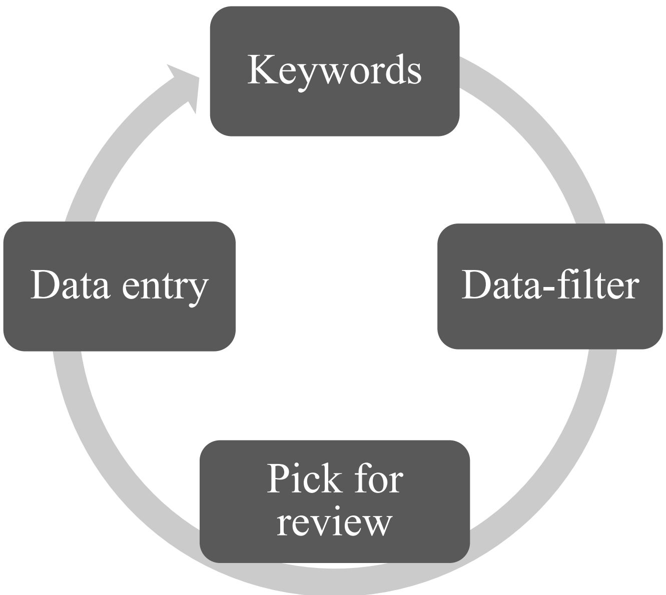

Figure 3. The iterative data collection method.

2.4 The Expert Interview Method

The case study employs an expert interview method. The method relies on the expertise and insights of a group of individuals, chosen for their knowledge and experience in the subject matter under investigation. These participants, often referred to panelists, provide input through a series of structured questionnaires or rounds of discussion. In each round, the responses from participants are aggregated and anonymized before being redistributed to the group. Participants then could review the summarized responses and revise their own opinions or judgments considering the insights shared by others. Further questions may be posed if the initial answers are found to be inadequate. This iterative process continues for several rounds until a consensus or convergence of opinions is reached, or until predefined stopping criteria are met (Loo 2002).

One of the key advantages of the expert interviews is its ability to utilize the collective wisdom of diverse stakeholders while mitigating biases and social pressures that could influence group decision-making. By maintaining anonymity and facilitating structured communication, the expert interviews encourage participants to express their opinions freely and thoughtfully, leading to more robust and reliable outcomes. It allows flexibility in adapting to evolving circumstances or changing information. As new insights emerge or as participants gain a broader understanding of the problem at hand, the iterative nature of the expert interview process enables adjustments to be made iteratively, ensuring that the final conclusions are based on the most up-to-date information available (Loo 2002).

This flexibility and dynamic adjustment are crucial, where variables often change rapidly. The expert interview's ability for continuous adaptation makes it ideal for these sectors because it helps to compile various participants opinions and refines forecasts and decisions progressively as new information becomes available. By anonymizing contributions, the expert interviews ensures that the discussions are not affected by the participants' status or charisma but focus solely on the content of their submissions. Its iterative nature is particularly beneficial for identifying points of consensus and divergence, which refines theories or models with each round (Landeta 2006).

The expert interview method could be tailored to the specific requirements of each study, such as adjusting the number of rounds to achieve consensus based on the complexity or controversy of the issue. The formats of the questionnaires could also vary from open to closed questions, depending on the type of information that the coordinators wish to extract from the participant panel. This contributes to a high degree of response validation and reliability, which exceeds that of one-time surveys or polls. As participants refine their responses over successive rounds, the collected data becomes more robust and reliable, significantly enhancing the quality of the results. Its structured, anonymous communication framework makes it a powerful tool for achieving consensus and making informed decisions (Landeta 2006).

2.4.1 Selection of the Panel Participants

The composition of the participant panel is a reference point for the expert interview method, as the quality of the process and its results is conditioned by the selection of participants. The participants in this study included employees of SMEs with a comprehensive understanding of the business operations. They are represented by the three case companies selected for their commitment and enthusiasm for implementing AI technologies. The number of participants was influenced by their knowledge and capability to effectively address and respond to the questions asked. This eliminated many potential respondents as the quality of the responses was important. The study concluded with the participation of nine persons. A common way to measure panel quality is through self-assessment. The participants will answer a set of questions, expresses the degree of knowledge about the object of research (López-Gómez 2018). The participants were therefore also asked to rate their own knowledge of the topic. This was considered an important point for assessing and analyzing the responses.

2.4.2 Interview Process Overview

Questionnaires were utilized in the study's research methodology to ensure the anonymity of the experts involved. To distribute the questionnaires to the case companies, a contact person was contacted, who then either referred the questionnaires to colleagues or circulated them internally. In connection with this, the purpose of the study was explained, and they were given information on how to complete and return the questionnaire. Thereafter, follow-up reminders were sent regularly to encourage timely responses. In the study, only two iterations were conducted because the enterprises reached consensus by the second round.

2.5 Data Collection, Preparation and Analysis

Two sets of questionnaires for the expert interviews were distributed in two rounds over a three–week period. For the first round, basic questions were asked about the experts' knowledge and familiarity with AI technologies, as well as questions about their perceptions of XAI and transparency. The responses from the first round were then collated and analyzed to develop the second–round questionnaire. Additional questions were included in the second round, as it became clear that some topics needed more context and were better suited for a later stage in the process. Therefore, the questions for the second questionnaire were designed to elicit the best possible answers and to align with the study's objectives. Once the responses were finally collected, they could be compiled and analyzed for the results of the study.

2.6 Quantitative and Qualitative Structures

The study could be regarded as employing both quantitative and qualitative methodologies. Patel and Davidson (2019) highlight the challenge of precisely defining the boundary between quantitative and qualitative research methods. This difficulty arises due to the intrinsic complexity and interdisciplinary character of research methodologies, leading to frequent overlap and evolution of methods. This was a sample survey, in the context of SMEs, which is characterized by a quantitative methodology. Given that the study utilized expert interviews as its chosen approach, makes the outcome distinctive with its iterative character, making replication challenging. Therefore, it could be considered as incorporating qualitative elements into its methodology, despite its predominantly quantitative nature. This hybrid approach allows for a more nuanced and comprehensive understanding of the research topic, capturing both quantitative information and qualitative insights from participant opinions.

Allwood (2021) describes how the research problem evolves with the research process as the empirical research adds more detail to the questions, which is true for both quantitative and qualitative methodology. The questionnaires for the expert interviews were formulated and developed based on previous discussions and the collective and growing knowledge of the researchers and participants involved.

3. THEORETICAL FRAMEWORK

This chapter provides a theoretical framework and explanatory approach. The purpose of the theoretical framework is to provide a structured foundation for exploring how XAI and transparency affect trust and the adoption of AI technologies in SMEs. Therefore, the framework serves to connect the research questions with relevant theories and concepts, guiding the analysis and interpretation of findings.

3.1 Introduction to Theoretical Framework

The purpose of this study was to investigate the impact of XAI and transparency on enhancing trust levels and streamlining the adoption of AI technologies in SMEs. To ensure a reliable outcome, the theoretical framework was constructed through conscious choices with connection to the research questions. The theoretical framework formed the basis for answering and supporting the research questions, as well as meeting the purpose. The main theories and concepts central to this study includes: XAI, the black box, transparency, AI trust, responsible AI, and AI acceptance.

3.2 Foundations of Explainable AI (XAI)

XAI and transparency are crucial for increasing trust and facilitating the adoption of AI technologies in SMEs. XAI aims to make AI systems more understandable by explaining how decisions are made and how they function, which is particularly important in contexts where trust and accountability are central.

Transparency in AI systems involve clearly describing and reproducing the mechanisms through which these systems make decisions and adapt to their environment. This includes disclosing the algorithms, models, and data sources used, as well as the rationale behind the decisions made by AI systems (Tchuente et al. 2024). Transparency not only helps build trust but also empowers users to make informed decisions and assess the risks and benefits of AI technology (Behl et al. 2023). In critical sectors such as healthcare, military, and finance, transparency is particularly important. In these fields, AI decisions could have significant consequences, and users need to understand how and why these decisions are made to trust the systems. By providing insights into AI decision-making processes, XAI helps ensure that AI is used safely, fairly, and ethically (Ali et al. 2023).

A central issue in AI is the "black box" nature of many AI systems. The black box problem means that AI models are often opaque, making their inner workings and decision-making processes difficult for users to understand. This lack of transparency could lead to problems with interpretability, accountability, and fairness, as users cannot see how a specific decision was made (Adadi & Berrada 2018). This opacity could create uncertainty and decrease users' willingness to trust and adopt AI technologies. To address this challenge, researchers, and practitioners in XAI strive to develop methods and techniques that provide insights into AI systems' decision-making processes, making them more transparent and understandable for users (Tchuente et al. 2024). While black box models could indeed provide accurate predictions, they do not offer explanations for how these predictions are generated. This could result in a lack of accountability and potential unfairness in AI system decisions, further undermining trust in the technology. By illuminating the decision-making logic behind AI predictions and actions, XAI aims to reduce this opacity and increase understanding and trust in AI systems (Ali et al. 2023).

For SMEs, trust in AI systems is crucial for their willingness to adopt and integrate these technologies into their operations. XAI could help build this trust by providing users with clear explanations of how and why AI systems make their decisions. By offering insights into AI models' functioning, XAI could help ensure that these systems meet both legal and ethical standards, which is especially important in sectors where decisions could have significant consequences (Ali et al. 2023). SMEs could greatly benefit from AI technologies to improve efficiency, optimize processes, and create new business opportunities. However, without trust and understanding of how these technologies work, AI adoption may be hindered. XAI offers a solution by making AI systems more accessible and comprehensible to users, which could, in turn, lead to broader adoption and integration of AI in SMEs' operations (ibid.).

Jamshidi et al. (2023) emphasizes the need to develop XAI solutions that consider users' varying knowledge levels and needs. Saeed & Omlin (2023) highlight the importance of a systematic meta survey to identify current challenges and future opportunities within XAI. These works contribute to the ongoing dialogue surrounding the development of XAI, an area crucial for enabling responsible and sustainable use of AI technologies. By integrating explainability into AI technologies, researchers and developers could create systems that not only perform efficiently but also provide users with necessary insights into their decision-making processes. This helps make AI systems more accepted and trusted, which in turn facilitates their adoption and use in SMEs (Madsen et al. 2023).

3.3 Components of XAI

XAI is divided into several key components that together contribute to making AI systems more understandable and trustworthy. These components include data explainability, model explainability, post-hoc explainability, and the assessment of explanations. By combining these discussions into one cohesive section, we could gain a broader understanding of how these aspects interact to enhance transparency and trust in AI systems.

Data explainability focuses on the transparency and interpretability of the data used by AI models. It involves ensuring that the input data fed into AI systems is understandable and meaningful, which is crucial for validating the outcomes produced by these models. Techniques for data explainability include pre-processing and understanding data, feature selection, data visualization, and the use of interpretable data representations (Ali et al. 2023). By providing clear insights into the data, stakeholders could critically assess the appropriateness of data for training AI models, identify potential biases, and ensure that the data accurately represents the problem space.

Bozorgpanah et al. (2022) examine the relationship between data privacy and the explainability of AI systems. They specifically investigate how privacy-preserving techniques impact the interpretability of AI decisions through Shapley values. Shapley values are used to explain the importance of each feature or attribute in a dataset by measuring its contribution to the model's prediction. This approach is crucial for understanding how data manipulation for privacy concerns affects the transparency and comprehensibility of AI model decisions. Their research suggests that it is possible to balance data privacy and explainability, ensuring that AI systems could be both secure and transparent, while providing stakeholders with insights into decision-making processes. This supports the development of AI technologies that are both ethically responsible and user trustworthy (Bozorgpanah et al. 2022).

Model explainability involves making the internal mechanisms and processes within AI models transparent to users. This is particularly important for complex models such as neural networks and ensemble methods. The goal is to demystify how inputs are transformed into outputs, thereby providing a clear understanding of the model's decision-making process. Techniques under this component may include simplified models that approximate the behavior of complex ones, feature importance ranking, and visual explanations of model behavior (Ali et al. 2023).

Arora and Chopra (2023) emphasize the importance of bridging data explainability and model explainability. They argue that a deeper insight into the role of data in AI models' decision-making processes could enhance explainability by offering more accurate and meaningful explanations of the model's behavior and decisions. By applying advanced techniques for data visualization and ranking the importance of features, they suggest methods to deconstruct and explain the complex relationships and interactions between input data and the model's outputs (ibid.).

Post-hoc explainability refers to explanations generated after the model has decided. These explanations aim to provide insight into the specific decisions made by AI models, often through local interpretation methods that explain individual predictions. Techniques include Local Interpretable Model-agnostic Explanations (LIME), Shapley Additive Explanations (SHAP), and counterfactual explanations, which help users understand why a model made a particular decision (Ali et al. 2023).

Madsen, Reddy, and Chandar (2023) emphasize the importance of post-hoc explainability in enhancing users' trust and understanding of AI systems. By illuminating in detail how individual inputs contribute to the model's final decision, these explanations enable a deeper insight into the workings of AI models, particularly valuable for SMEs utilizing AI in critical decision-making processes. Understanding and thus trusting the decisions of AI systems is essential for making informed business decisions and improving customer relations through increased transparency. By applying different post-hoc explanation techniques, businesses could tailor the level of explanation to their specific use case and needs, which Madsen et al. (2023) identify as key to implementing effective and reliable AI solutions. This flexibility in choosing the appropriate explanation method not only facilitates internal decision-making processes but also contributes to stronger customer engagement by demonstrating a willingness and ability to account for the basis of AI decisions (ibid.).

Hamm et al. (2023) conducted a study to test the impact of these post-hoc explanations, specifically using SHAP, on user perception. Their experiment involved participants who were tasked with identifying forged signatures, a sensitive and precise activity that benefits significantly from clear and understandable AI decisions. The study revealed that while the introduction of SHAP explanations did slightly but significantly improve the perceived explainability of decisions made by the AI, it had a more pronounced effect on the users' trust and perceived usefulness of the AI system. This highlights the critical role of user interface design in enhancing the effectiveness of post-hoc explanations. An interface that users find intuitive and engaging leads to higher levels of perceived explainability, which in turn enhances trust and satisfaction with the AI system. Hamm et al. (2023) suggest that the aesthetic and interactive elements of the AI application are not just peripheral but central to its success, implying that developers should place as much emphasis on these aspects as on the underlying AI technology.

Hamm et al. (2023) also highlight the importance of hedonic factors, those related to the pleasure and emotional satisfaction derived from using the system. They found that users who enjoyed the interaction with the AI system perceived higher levels of explainability, which bolstered their trust and overall satisfaction. This finding is particularly relevant for applications where AI decisions have significant personal or financial implications for users, such as in healthcare, financial services, and legal industries.

The assessment of explanations involves evaluating the effectiveness and impact of the explanations provided by AI systems. This component focuses on ensuring that explanations meet the needs of various stakeholders, including domain experts, end-users, and regulatory bodies. It includes evaluating the understandability, relevance, and accuracy of explanations, as well as their ability to foster trust and enable informed decision-making (Ali et al. 2023).

Finke et al. (2022) argue that the assessment of explanations is crucial for understanding and improving how AI systems communicate their decisions and reasoning to people. By systematically evaluating the quality of explanations, developers and designers of AI systems could tailor and enhance these systems to better meet the needs of users and society. This process contributes to creating more transparent, reliable, and ethically responsible AI systems that people could trust and understand. Finke et al. (2022) emphasize the importance of integrating a thorough assessment of explanations in the development and implementation of AI systems. This includes developing guidelines and frameworks for how explanations should be designed, presented, and tailored to meet the varied needs and expectations of different users. Such an assessment requires an interdisciplinary approach that combines insights from computer science, psychology, design, and ethics to ensure that AI explanations are not only technically accurate but also meaningful and useful for people in their daily interactions with AI systems.

By combining these components—data explainability, model explainability, post-hoc explainability, and assessment of explanations—a comprehensive framework is created that makes AI systems more understandable and trustworthy. These components are not isolated but interact to enhance users' understanding and trust in AI technologies. This is crucial for building trust and facilitating the adoption of AI technologies in SMEs. Ali et al. (2023), Madsen et al. (2023), and Hamm et al. (2023) collectively illustrate that post-hoc explainability not only needs to be technically proficient in explaining AI decisions but also must be user-centric, addressing both the cognitive and emotional needs of users.

3.4 Challenges and Key Issues in XAI

Implementing XAI and achieving transparency poses significant challenges for SMEs. Understanding and addressing these challenges is crucial for the effective implementation and utilization of AI technologies within these organizations. One of the primary challenges for SMEs is the inherent complexity of XAI methods and tools. Many AI models, especially deep learning models, function as black boxes, making it difficult for non-experts to understand their decision-making processes. This complexity is a significant barrier to the adoption of XAI, as SMEs often lack the specialized knowledge required to interpret and implement these technologies effectively (Adadi & Berrada 2018). Maintaining a balance between high performance of AI models and their transparency is another major challenge. While advanced AI models could deliver highly accurate results, their complexity often reduces their interpretability. SMEs must navigate the trade-off between using sophisticated AI models and ensuring these models are explainable and transparent (Jamshidi et al. 2023). This balance is particularly critical as SMEs may not have the resources to deploy multiple models or conduct extensive testing to find the optimal balance.

Ensuring data privacy while maintaining explainability is a significant issue. Techniques such as Shapley values, used to explain the importance of features in a dataset, could sometimes conflict with privacy-preserving methods. SMEs must find ways to protect sensitive information without compromising the transparency and interpretability of AI models (Bozorgpanah et al. 2022). This balance is crucial for gaining and maintaining the trust of customers and stakeholders. Integrating XAI into existing business processes and systems could be complex and resource intensive. Effective implementation requires a broad understanding of how XAI technologies could complement and enhance current operations without causing disruptions. SMEs often face difficulties in this area due to limited resources and expertise (Jamshidi et al. 2023). Strategic planning and phased integration could help mitigate these challenges, but they require careful consideration and execution.

XAI solutions must be tailored to the varying knowledge levels and needs of different users within an SME. Creating explanations that are accessible and meaningful to both technical and non-technical stakeholders is challenging. Customizing XAI solutions to meet the specific requirements and understanding levels of different users is critical for effective adoption (Jamshidi et al. 2023). This need for customization adds another layer of complexity to the implementation process. Saeed & Omlin (2023) emphasize the importance of formalization in the field, where a clearer definition and quantification of explainability and its quality metrics are crucial. This includes the development of objective and user-centered evaluation methods to measure the effectiveness and usability of XAI solutions. Their meta survey also underscores the need for interdisciplinary research and collaboration to improve the understanding and methodology within XAI.

SMEs must navigate a complex landscape of regulatory and ethical considerations when implementing XAI. Compliance with standards like GDPR for data protection and adherence to ethical guidelines for AI use are critical. This requires SMEs to stay updated with regulatory changes and implement robust governance frameworks (Deshpande & Sharp, 2022; Wulf & Seizov 2022). The cost and effort associated with regulatory compliance could be significant, especially for smaller organizations. Building trust in AI systems is crucial for their adoption. Transparency and explainability are central to creating trust among users and stakeholders. Kartikeya and Arai (2022) point out that the lack of transparency in AI decision-making processes could problematize trust, but XAI could significantly increase trust by providing clear and understandable explanations. Gerlich (2023) further emphasizes the importance of adhering to trustworthy AI principles, such as data security, transparency, accountability, and fairness, to build and maintain trust.

3.5 Responsible AI Practices

Ensuring responsible AI practices in SMEs requires a comprehensive understanding of ethical frameworks and guidelines. These guidelines are crucial for protecting privacy, security, and fairness in AI implementations. By focusing on these aspects, SMEs could build and maintain trust with their customers and stakeholders.

A central element in the development of responsible AI is adhering to established international standards and guidelines. Deshpande & Sharp (2022) emphasize the importance of following standards such as ISO 26000:2010 and ISO/IEC 27001 to ensure ethical, transparent, and fair use of AI. These standards help companies develop and implement AI systems that align with internationally recognized ethical principles. GDPR and CCPA are two significant regulatory frameworks governing data protection and security, with the first protecting EU and the second protecting California. These regulations promote user privacy and transparency, which is critical for building trust in AI systems (Deshpande & Sharp 2022). Wulf & Seizov (2022) underscore the importance of transparency and explainability in AI systems while criticizing the effectiveness of GDPR in ensuring sufficient control for individuals over the automated processing of their personal information. Their research highlights the need to develop rules and guidelines that better address the challenges posed by modern AI technologies.

Data protection is one of the most critical aspects of responsible AI usage. Ensuring that user data is handled securely and transparently is essential to prevent misuse and protect users' privacy. Bozorgpanah et al. (2022) discuss the balance between data protection and explainability, demonstrating how privacy-preserving techniques could impact the interpretability of AI decisions. Their research shows that it is possible to apply data protection methods without compromising the explainability of AI models, which is crucial for creating ethically responsible and reliable AI systems. Gerlich (2023) interprets the European

Commission's principles for trustworthy AI, emphasizing data protection, security, transparency, accountability, and fairness. These principles serve as guidelines to ensure that AI technologies are developed, implemented, and used in a manner that is ethical, responsible, and aligned with societal values.

Fairness is another fundamental principle that must be considered in the development and implementation of AI systems. Inequality and bias in AI systems could lead to unfair and discriminatory decisions. Ali et al. (2023) highlights the importance of developing AI models that are transparent and understandable, which could help identify and address biases and ensure fair decisions. Wulf & Seizov (2022) stress the need for a more nuanced and effective regulatory framework to handle the challenges posed by modern AI technologies. They point out that a holistic approach, including industry-specific guidelines and ethical principles, is necessary to build trust and create fair AI systems.

SMEs face challenges when implementing XAI. These companies often have limited resources and may not have access to the same level of technical expertise as larger organizations. To overcome these obstacles, it is essential for SMEs to have access to guidelines and tools specifically designed for their needs. Norzelan, Mohamed, and Mohamad (2024) emphasize the importance of understanding the factors that influence technology acceptance in SMEs, including performance expectations and attitudes towards AI. They note that clear communication about the benefits of AI, as well as education and skill development, are crucial for increasing acceptance and promoting successful implementation. Hasija and Esper (2022) propose strategies for addressing fears and uncertainties surrounding AI technologies, including involving employees in the implementation process, and providing continuous education. This could help reduce resistance and increase trust in AI systems.

3.6 Relationship between XAI, Transparency and AI Trust

The relationship between XAI, transparency, and trust in AI systems within SMEs is intricate and essential for the responsible development and deployment of AI technologies. XAI plays a crucial role in making AI systems more transparent by explaining how decisions are made and how the systems function, which is particularly important in contexts where trust and accountability are central.

Transparency in AI systems involve clearly describing and reproducing the mechanisms through which these systems make decisions and adapt to their environment. This includes disclosing the algorithms, models, and data sources used, as well as the rationale behind the decisions made by AI systems (Tchuente et al. 2024). Transparency helps build trust and empowers users to make informed decisions and assess the risks and benefits of AI technology (Behl et al. 2023). In critical sectors such as healthcare, military, and finance, transparency is particularly important as AI decisions could have significant consequences. Users need to understand how and why these decisions are made to trust the systems, ensuring AI is used safely, fairly, and ethically (Ali et al. 2023).

A central issue in AI is the black box nature of many AI systems, where models are often opaque, making their inner workings and decision-making processes difficult for users to understand. This lack of transparency could lead to problems with interpretability, accountability, and fairness, as users cannot see how a specific decision was made (Adadi & Berrada 2018). This opacity could create uncertainty and decrease users' willingness to trust and adopt AI technologies. Researchers and practitioners in XAI strive to develop methods and techniques that provide insights into AI systems' decision-making processes, making them more transparent and understandable for users (Tchuente et al. 2024). By illuminating the decision-making logic behind AI predictions and actions, XAI aims to reduce this opacity and increase understanding and trust in AI systems (Ali et al. 2023).

For SMEs, trust in AI systems is crucial for their willingness to adopt and integrate these technologies into their operations. XAI could help build this trust by providing users with clear explanations of how and why AI systems make their decisions. By offering insights into AI models' functioning, XAI could help ensure that these systems meet both legal and ethical standards, which is especially important in sectors where decisions could have significant consequences (Ali et al. 2023). SMEs could greatly benefit from AI technologies to improve efficiency, optimize processes, and create new business opportunities. However, without trust and understanding of how these technologies work, AI adoption may be hindered. XAI offers a solution by making AI systems more accessible and comprehensible to users, which could lead to broader adoption and integration of AI in SMEs' operations (Ali et al. 2023).

Empirical evidence supports the role of XAI in enhancing transparency and trust. Khakurel and Rawat (2022) highlight the transformative potential of XAI in providing maximum transparency by answering questions about how models effectively come up with outputs. Their findings underline the importance of XAI in bolstering trust in AI systems, paving the way for wider acceptance and adoption across various domains. Kartikeya and Arai (2022) delve deeper into the dynamics of trust between humans and AI, noting that the lack of transparency in AI decision-making processes could undermine trust. Their research shows that XAI could significantly increase trust by making the decision-making process more transparent.

Madsen, Reddy, and Chandar (2023) conducted a study on SMEs utilizing AI in critical decision-making processes, showing that XAI significantly improves users' understanding of AI decisions, which enhances trust and acceptance of AI systems. The study highlights the importance of providing explanations tailored to the specific needs of different stakeholders within SMEs. Hamm et al. (2023) tested the impact of post-hoc explanations, specifically using SHAP, on user perception. Participants tasked with identifying forged signatures found that SHAP explanations improved the perceived explainability of AI decisions, significantly enhancing users' trust and perceived usefulness of the AI system. This underscores the importance of user interface design in enhancing the effectiveness of post-hoc explanations. An intuitive and engaging interface leads to higher levels of perceived explainability, enhancing trust and satisfaction with the AI system.

Gerlich (2023) interprets the European Commission's principles for trustworthy AI, emphasizing data protection, security, transparency, accountability, and fairness. These principles serve as guidelines to ensure that AI technologies are developed, implemented, and used ethically and responsibly, aligning with societal values. By adhering to these principles, SMEs could build trust in AI systems among users and stakeholders.

XAI is crucial in building transparency and trust in AI systems within SMEs. By making AI decision-making processes more understandable and transparent, XAI helps SMEs navigate the challenges of AI adoption and fosters trust among users and stakeholders. Empirical evidence highlights the importance of tailored explanations, intuitive interfaces, and adherence to ethical principles in enhancing the transparency and trustworthiness of AI systems, facilitating broader acceptance and integration of AI technologies in SMEs.

3.7 Acceptance and Adoption of AI Technologies

The acceptance and adoption of AI technologies in SMEs are influenced by several key factors, including perceived usefulness, ease of use, organizational readiness, and perceived risk. Understanding these factors is essential for developing strategies to overcome barriers to AI adoption, with a focus on skills development and employee involvement.

One of the primary drivers of AI adoption in SMEs is the perceived usefulness of AI technologies. SMEs are more inclined to adopt AI tools if they believe these technologies could enhance operational efficiency, improve decision-making processes, and provide a competitive advantage. Tangible benefits such as cost savings, increased productivity, and better customer insights play a crucial role in shaping positive attitudes toward AI adoption (Norzelan, Mohamed & Mohamad 2024). For instance, AI could automate repetitive tasks, allowing employees to focus on more strategic activities, and provide data-driven insights that help businesses make more informed decisions. This perceived usefulness directly correlates with the performance expectancy, which is a significant factor influencing technology acceptance.

However, the complexity of AI systems could act as a significant barrier to their adoption. SMEs often lack the technical expertise required to implement and maintain sophisticated AI solutions, making the technology seem daunting. The complexity of AI models, especially those that function as black boxes, further exacerbates this issue as these models could be difficult for non-experts to understand and trust (Adadi & Berrada 2018). Therefore, the ease of use of AI technologies, including user-friendly interfaces and comprehensive documentation, is critical. Simplifying the integration process and providing accessible support could help mitigate these challenges, making AI more approachable for smaller enterprises (Jamshidi et al. 2023). Techniques such as XAI could play a crucial role in this context by making AI systems more understandable and transparent, thereby facilitating their adoption.

Organizational readiness also significantly impacts AI adoption. This readiness is influenced by existing infrastructure, financial resources, and the technical skills of the workforce. Organizations with a robust IT infrastructure and a culture of innovation are better positioned to integrate AI technologies effectively (Jamshidi et al. 2023). Additionally, the availability of financial resources to invest in AI and related training programs is a critical factor. SMEs must consider the initial costs of AI implementation and the ongoing costs of maintenance and training. Without adequate investment, even the most willing organizations may struggle to implement AI solutions effectively. This highlights the importance of strategic planning and phased integration to ensure that AI technologies are adopted in a manner that is sustainable and scalable.

Perceived risk is another important consideration in the acceptance and adoption of AI technologies. Concerns about data security, privacy, and the potential for job displacement could hinder the acceptance of AI. To overcome these barriers, it is essential to address these concerns transparently and proactively. Implementing robust data protection measures and ensuring that AI systems comply with relevant regulations, such as GDPR, could help alleviate fears about data security and privacy (Deshpande & Sharp 2022; Wulf & Seizov 2022). Moreover, involving employees in the adoption process could mitigate fears about job displacement. When employees understand that AI is meant to augment rather than replace their roles, they are more likely to support its implementation (Gerlich 2023). Creating an environment where employees feel secure and valued could significantly reduce resistance to AI technologies.

Skills development and employee involvement are crucial strategies for overcoming barriers to AI adoption. Investing in training programs to enhance the technical skills of the workforce is vital. This includes not only training on how to use AI technologies but also developing a broader understanding of AI's potential and limitations. By improving the technical capacity of employees, SMEs could ensure that their workforce is equipped to leverage AI effectively (Hasija & Esper 2022). Continuous education and skill development could help employees stay updated with the latest advancements in AI, fostering a culture of learning and innovation within the organization.

Employee involvement is equally important in the successful adoption of AI technologies. Engaging employees in the AI adoption process helps build trust and reduces resistance. Involving staff in decision-making processes, seeking their input on AI initiatives, and clearly communicating the benefits of AI could foster a sense of ownership and acceptance. When employees feel involved and valued, they are more likely to embrace new technologies (Vorobeva et al. 2023). This involvement could take various forms, such as forming cross-functional teams to oversee AI projects, organizing workshops to gather feedback, and creating channels for continuous communication about AI developments.

Empirical evidence supports the role of XAI in enhancing transparency and trust, which are crucial for AI adoption. Khakurel and Rawat (2022) highlight the transformative potential of XAI in providing maximum transparency by answering questions about how models effectively come up with outputs. Their findings underline the importance of XAI in bolstering trust in AI systems, paving the way for wider acceptance and adoption across various domains. Additionally, Kartikeya and Arai (2022) delve deeper into the dynamics of trust between humans and AI, noting that the lack of transparency in AI decision-making processes could undermine trust. Their research shows that XAI could significantly increase trust by making the decision-making process more transparent, thereby facilitating broader adoption and integration of AI technologies in SMEs' operations.

The acceptance and adoption of AI technologies in SMEs hinge on addressing the perceived usefulness, ease of use, organizational readiness, and perceived risks associated with AI. By emphasizing skills development and employee involvement, and adhering to ethical principles, SMEs could overcome barriers to AI adoption and harness the potential of AI to drive innovation and growth. As SMEs navigate the complexities of AI adoption, they must prioritize transparency, continuous learning, and employee engagement to build a foundation of trust and acceptance for AI technologies.

4. RESULTS AND ANALYSIS

This chapter provides the findings of the case study, comprising a summary of both the literature review results and the outcomes from the first and second iterations of the expert interviews. For the expert interviews, an overall summary of the collected responses is initially provided, followed by a detailed development and explanation.

4.1 Overview of Literature Review

In the literature review, the significance of XAI and transparency was explored, highlighting their roles in building trust and facilitating the adoption of AI technologies in SMEs. By analyzing core concepts and relevant theories, key insights were identified and integrated into the design of an expert interview and subsequent surveys.

The opacity of AI systems, often referred to as the black box, emerges as a significant barrier to user trust. This characteristic means that systems could deliver accurate predictions without explaining how these conclusions are reached. The issue of the black box is central to the study as it directly affects users' ability to trust and understand the technology. To address this, questions were included in the survey aimed at exploring how AI is used within organizations, with a particular focus on perceptions of AI system transparency (Adadi & Berrada 2018). These questions were intended to reveal how well participants understood the AI systems they interact with and the impact this understanding has on their trust in the technology.

XAI forms another core aspect of the literature review. Research indicates that increasing the explainability of AI could reduce the gap between complex AI operations and users' understanding, which could enhance trust. With this insight, survey questions focused on attitudes towards XAI were formulated to assess its impact on decision-making processes within SMEs (Ali et al. 2023). By exploring familiarity with and attitudes towards XAI, data was gathered on how well participants understood how AI made decisions and the effect this had on their view of AI technology.

In the application of the expert interviews, the integration of these theoretical insights aimed to capture different perspectives on how explainability and transparency are managed within various organizations. The survey reflected the complexities discussed in the theory, including various levels of transparency and the role of explainability in creating trust, as well as how these factors influence strategies for adopting AI (Jamshidi et al. 2023).

By reviewing and integrating theoretical frameworks into an empirical investigation, a link between theory and practice was established. This has strengthened the academic foundation of the research and deepened understanding of practical applications and challenges within the field of XAI. The results from the expert interviews and subsequent analyses of survey data provide a comprehensive picture of how XAI and transparency affect trust and adoption in SMEs, which directly informs the conclusions in the thesis. By bringing together theoretical insights and empirical data, the work contributes to an understanding of the dynamics of AI adoption and the critical factors that shape this process in the business world.

4.2 First Iteration

During the initial iteration, the first questionnaire was distributed to participants within the companies. The findings revealed a variety of perspectives and opinions on AI technology, which were influenced by the knowledge and position of the participants. The survey included a diverse range of participant spanning various roles and domains, such as executives like CEOs and managing directors, operational staff like machine operators, and professionals specializing in business development, product development, mechanical and software development. This diverse representation offered a comprehensive viewpoint on the subject, encompassing a range of expertise levels and insights. The experts also had varying years of experience in their field, as illustrated in Figure 4. In total, nine experts participated: two with 0-5 years of experience, four with 5-10 years, and one each with 10-15, 15-20, and 20-25 years of experience.

Figure 4. Distribution of Expert´s Years of Experience.

4.2.1 Overview of First Iteration

Table 2 presents a summary of the experts' perceptions of the questions in Questionnaire 1. This is presented to provide an overview of the responses and facilitate a deeper understanding of the insights garnered from the questionnaire analysis. Additionally, this overview fosters transparency and ensures that the findings derived from the questionnaire analysis are accessible and comprehensible to all parties involved in the research endeavor. The questions are also presented in Appendix 1.

Table 2. Summary of the Collected Responses of Questionnaire 1.

Question	**Summary**
Current use of AI Technologies	AI technologies are minimally used, with some using AI for translation and purchasing.
Usage of AI in Professional Activities	Several participants have used AI tools like ChatGPT for translation, creating templates, and structuring texts.
Frequency of Interaction with AI Technologies	Most participants rarely interact with AI technologies, with some using AI monthly or daily for small tasks.
Familiarity with AI Technologies (1 to 5 scale)	Most participants have low to medium familiarity with AI technologies, with the majority rating themselves between 1 and 3 on the scale.
Knowledge of Organization's Functions (1 to 5 scale)	Participants generally have good knowledge of their organizations' functions, with the majority rating themselves between 3 and 5 on the scale.
Perception of XAI's Impact on Colleagues' Attitudes	Mixed perceptions, ranging from curiosity and indifference to the belief in the importance of understanding AI. Transparency and clear explanations are seen as crucial for wider AI adoption.
Importance of Transparency in AI Systems for Building Trust	Most participants believe transparency is very important for building trust in AI systems.
Impact of XAI on Decision-Making Processes	Several participants believe XAI would improve decision-making and provide better decision support.
Challenges in Implementing Transparent AI Systems	Challenges include lack of knowledge, fear of job loss, and resource constraints. Proposed solutions involve hiring AI specialists, gradual implementation, and clear communication.
Examples of XAI Integration in Workflows	Suggestions include production planning, more efficient machine programming, and customer order handling.
Strategies for Enhancing Understanding and Trust in AI	Education, clear communication, and showing concrete examples of AI use in other companies are considered suitable strategies.

4.2.2 Experts´ Responses from First Iteration

The initial questions centered on understanding companies' existing familiarity with and utilization of AI technologies. Where the responses were consistent in the limited utilization, in instances where it was employed, it was in the use for translation and through common web-services or in the use of a chat bot. Thus, many of the respondents uses common web-services in their work, even though it is not a fully established method in the organization.

The participants were then asked to rate, on a scale one to five, their own knowledge of AI technologies as well as the knowledge of the operational processes where AI implementation is considered. Where level 1 represents no knowledge at all, and level 5 represents very good knowledge. This was important for understanding the participants´ perception of their own knowledge levels about the subjects. This is illustrated in Table 3. This indicated that the participants have a limited understanding of AI technologies but a better understanding for the operational processes. Therefore, while the initial assessment may reveal areas for improvement, it ultimately portrays participants as valuable assets poised to contribute effectively to the organization's AI endeavors.

Table 3. Participants` Knowledge Levels of AI Technology and Operational Processes.

Knowledge Level	AI Technology Number of Participants	Operational Processes Number of Participants
Level 1	1	0
Level 2	3	0
Level 3	4	2
Level 4	1	5
Level 5	0	2

The subsequent questions in the questionnaire addressed how the experience and perception of XAI and transparency influenced trust in AI technology within the companies. For these questions, there were some disagreements in the thoughts and answers. Certain participants emphasized the importance of XAI and transparency in gaining a deeper understanding of AI systems. Other participants were not convinced that this was important, but also appeared to have misconceptions about the questions or the meaning of the different aspects. Therefore, this was further investigated in iteration 2, to see if it was possible to reach more consensus.

The results also suggest that there is a greater understanding of transparency in AI systems compared to XAI, since the responses were more persuasive in this instance. This means that participants in the study have a better grasp of the concept of transparency in AI systems than they do of the concept of XAI. One of the participants stated that transparency in AI systems is not only important for building trust, but also crucial for handling sensitive information, which appeared to agree with others. Additionally, it was emphasized that addressing the concerns of individuals who experience fear or a sense of insecurity regarding AI technology is imperative. Furthermore, one of the participants described that transparency could provide insights and enable to identify potential areas where AI could enhance the business operation. Currently, AI remains a complex area, with many aspects of its functionality still not fully clarified. However, with staff equipped with knowledge and comprehension of the underlying AI systems, it could be seriously regarded as a viable option for various tasks within the business.

Most of the participants were convinced that XAI is crucial for decision-making processes, and it is pointed out that this understanding that XAI could provide, also could extend to a broader portion of the business, reaching beyond just the individuals most closely involved. Thus, this broader comprehension throughout the organization could facilitate informed decision-making and streamline the integration of AI into various areas of the business.

The participants were asked to share their views on which strategies or approaches they believe are most appropriate to build trust in the technology and facilitate its implementation in their own organization. One of the participants elaborates on the challenge of determining the initial steps to take in the implementation process. Uncertainty is expressed about where to begin, highlighting the complexity of navigating the integration of AI technology within the organization. Another participant emphasizes that the initial step is to ensure that all stakeholders are well-informed about AI and its potential implications for the organization, also to take small steps forward. In addition, the importance of strong leadership to guide the implementation process effectively is emphasized. It is suggested that having clear communication channels and supportive leadership along with education are essential for overcoming the complexities and uncertainties associated with AI integration.

The participants outlined the areas where they foresee the advantages of employing AI technologies, including optimizing production planning, programming machines, translating texts into different languages, improving data system search capabilities, using camera techniques to detect deviations, and generating new quotations and layout designs. These applications demonstrate the potential for XAI to significantly enhance efficiency and decision-making processes across various departments and functions within the organization.

4.3 Second Iteration

During the second iteration of the expert interviews, the four axes of explainability were explored in detail, as well as the knowledge gaps previously identified, focusing on deepening the understanding of specific AI models and their potential applications within the organizations. The aim of this phase was to identify and address any inconsistencies or knowledge gaps that had emerged from the first iteration, and to investigate how XAI could help bridge the gap between AI's technical capabilities and user expectations and needs.

4.3.1 Overview of Second Iteration

Table 4 offers a summary of the experts' views on the questions in Questionnaire 2. This summary is provided to give an overview of the responses and to deepen the understanding of the insights obtained from the questionnaire analysis. Furthermore, this summary promotes transparency and ensures that the findings from the questionnaire analysis are accessible and understandable to all stakeholders involved in the research project. The questions are to be found in Appendix 2.

Tabell 4. Summary of the Collected Responses of Questionnaire 2.

Question	**Summary**
Desired AI Functions and Areas for Integration	Participants suggest AI could help with code refactoring, product text translation, warehouse optimization, production planning, and quality control.
Willingness to Implement AI Without Explanations	The majority are reluctant to implement AI without explanations, with explanations considered necessary to verify AI outputs and build trust.
Acceptance of Transparent AI Systems Without Decision Explanations	Most participants are still reluctant to implement AI without explanations, even if the system is transparent. Explanations are needed for understanding decisions.
Perceived Impact of XAI on Colleagues' Attitudes Towards AI Adoption	Most participants consider XAI and transparency to be crucial for building trust in AI systems.
Main Fears and Concerns About AI Technologies	Main concerns include loss of control, data security, and job displacement.
Significance of Clear and Understandable Input Data for Verifying AI Model Results	Clear and easy-to-understand input data is considered essential for verifying AI results and ensuring accuracy.
Importance of Tracing and Understanding AI Models' Decision-Making Processes	Most participants believe it is important to trace and understand AI models' decisions to detect biases or inaccuracies.

Value of Receiving Detailed Explanations of AI Decisions Afterward for Improving AI Initiatives	The ability to receive detailed explanations afterward is considered important for understanding and improving AI initiatives.
Prioritization of Evaluating the Quality of AI System Explanations to Ensure They Are Understandable and Useful	Evaluating the quality of explanations from AI systems is considered essential to determine when actions can be taken and to improve training data.
Crucial Aspects of Explainability in AI (Data, Model, Post-Hoc, and Assessment) for Enhancing Transparency and Trust in AI Technologies and Their Reasons	Data and model explainability are prioritized for transparency and trust. Post-hoc explainability and assessment of explanations are also considered important.

4.3.2 Experts´ Responses from Second Iteration

The responses from the first survey round showed a varied degree of understanding and acceptance of AI technology among the participants, which pointed to the need for further clarification and education. To better address these areas, more targeted and in-depth questions were introduced in the second round. These questions aimed to explore the participants' specific expectations of AI and how well their perceptions aligned with realistic AI applications.

By reformulating the questions and focusing on specific use cases and scenarios where XAI could play a critical role, a significant improvement in the quality of the responses was observed. The participants expressed an increased understanding of how XAI could contribute to making AI decisions more transparent and reliable, which in turn could improve the acceptance and trust in AI systems within their organizations.

The participants were presented with several options for AI applications, some with explainable and transparent systems for evaluation, and others without. They were then asked to determine whether the application could provide value to the organization and whether it could possibly contribute to a greater acceptance of AI systems. Some participants pointed out that for simple and normal tasks there may be no need for explainable or transparent systems as the results are clear anyway. However, this does not apply to individuals who are unable to perform the task themselves or possess limited knowledge of the subject. But the participants otherwise agree that explainability and transparency in AI applications is necessary for trust in the technology. One of the questions asked whether the participants were willing to trust an AI application that offered transparency but no explanations. Here, several participants hesitated in their answers, indicating that they had developed a greater appreciation for the importance of XAI. Thus, they described that the transparent systems could still be complicated and difficult to understand, which increasing the need for explainable systems.

The participants state that a lack of knowledge and the reluctance to trust AI are causing concern in their organizations. They explain that they require transparent systems to understand where the information is going and to be able to trace the decisions made by AI. They argue that without this transparency, it's challenging to build trust in AI systems within their organizations. XAI could address these concerns by providing insights into how AI systems arrive at their decisions. XAI techniques make AI algorithms more transparent and interpretable, allowing users to understand the reasoning behind AI-generated decisions. By implementing XAI, organizations could gain visibility into the decision-making process of AI systems, thereby increasing trust and confidence among users. With XAI, users could track how data is being processed, which features are influencing decisions, and understand the overall decision-making logic. This transparency not only helps in building trust but also enables organizations to identify and rectify any biases or errors in the AI system. Therefore, XAI serves as a crucial tool in addressing the lack of knowledge and trust in AI within organizations.

This section addressed the importance of the four axes of XAI: data explainability, model explainability, post-hoc explainability, and explanation assessment. Each axis was thoroughly reviewed to emphasize its crucial role in fostering transparency and trust within AI systems, which are increasingly vital for operational success across various industries, particularly SMEs.

For data explainability the importance of clear and understandable data was strongly emphasized. Participants agreed on the necessity of transparent data management practices. Several participants mentioned the challenge of providing the right input to the system, which in turn may require knowledge and adaptation of existing systems and processes. On the other hand, the use of XAI could help to understand the quality and relevance of the input data. By incorporating XAI techniques, organizations could enhance the transparency of their AI systems and improve the interpretability of their decisions, thereby addressing concerns related to data quality and system reliability. Additionally, providing employees with training on how to interpret and utilize AI-generated insights could further support the successful integration of AI technologies within organizations.

For model explainability, the participants agreed on its importance. The ability to trace and understand the decision-making processes of AI models highlights the critical need for transparency. Participants noted that knowing how data is processed and analyzed to produce results allows for greater accountability and reliability in AI systems. A clear and coherent explanation of these processes is crucial for building user confidence and facilitating fair and responsible AI practices. This transparency helps identify and correct errors and fine-tune AI models over time to better serve their intended purposes without compromising ethical standards.

Even for post-hoc explainability, there was consensus on its importance. One of the participants gives their view on how it could be used effectively to fine-tune the processes, but also to compare different results depending on the input data. Additionally, post-hoc explainability enables organizations to identify and address biases and errors in AI decision-making. By scrutinizing the underlying factors influencing AI outcomes, organizations could detect and rectify any biases or errors that may arise due to skewed or incomplete data.

For assessment of explanations, there was strong consensus on the necessity of carefully evaluating how explanations are perceived by end-users. It is important that the explanations are understandable and relevant to all users, regardless of their technical background. This assessment should include user feedback to continuously improve the clarity and usability of the explanations provided by AI systems. This iterative process of feedback and improvement could help tailor AI operations to user expectations and operational needs, increasing the overall effectiveness and acceptance of AI technologies.

4.4 Compilation of The Survey

This section provides an overview of the study's findings and categorizes them into general findings, applicable across various organizations, and those specific to SMEs. This was necessary to delineate the purpose and research questions of the study focused on SMEs. Table 5 visualizes this division to facilitate the comprehension of the results.

Table 5. Comparing general findings for organizations with specific SME findings.

Aspect	**General finding for organizations**	**Specific SME Findings**
XAI and Transparency	Participants showed an increased understanding of XAI and transparency in AI systems. This was deemed important for building trust in AI technologies overall. Participants identified that XAI and transparency increased understanding of how AI decisions are made and helped eliminate uncertainty about the technology's functioning.	SMEs identified a direct connection between XAI, transparency, and trust. Their interest in adopting AI technologies increased when systems were transparent and explainable. An increased understanding of how AI decisions are made, coupled with transparency in the decision-making process, contributed to reducing the risk of errors and improving efficiency within the operations.
Trust and understanding of AI Technologies	Participants believed that increased trust and understanding of AI technologies were necessary to facilitate interest in the adoption of these technologies. Increased trust was built through transparency in the decision-making process and XAI, leading to a deeper understanding of the	SMEs experienced a positive correlation between increased trust and understanding of AI and increased interest in the adoption of AI technologies in their operations. An improved understanding of the benefits and challenges of the technology through XAI and transparency helped SMEs identify

	benefits and limitations of the technology.	specific use cases where AI can add value and enhance competitiveness.
Adoption of AI Technologies within SMEs	The results indicated that XAI, transparency, trust, and understanding were key factors in increasing interest in the adoption of AI technologies, especially within the SME sector. Participants within the SME sector expressed that XAI, and transparency were crucial for reducing risks in adopting new technologies and facilitating exploration of AI possibilities within organizations. By increasing trust in AI technologies and understanding its functions and potential value, SMEs were more inclined to explore and investigate use cases where AI can add value and increase competitiveness.	Interest in the adoption of AI technologies within SMEs increased through XAI and transparency, which enhanced trust and understanding of these technologies. XAI and transparency were considered crucial for reducing uncertainties and risks of errors in exploring AI technologies within the SME sector and facilitated a smoother transition to an exploratory and investigative approach to AI.
Organizational Challenges and Opportunities	Participants identified various organizational challenges and opportunities affecting the adoption of AI technologies. Challenges included resource constraints, cultural factors, and the need for education and skill development. Opportunities included increased productivity, cost savings, and improved customer satisfaction through more efficient services and products.	SMEs encountered similar organizational challenges and opportunities but also experienced specific barriers and opportunities unique to their size and scope of operations. These included the need for flexibility and adaptation, resource constraints, and the benefits of agility and speed in decision-making.
Effects on Business Performance	Participants noted potential improvements in business performance through the adoption of AI technologies. This included increased productivity, cost	SMEs identified opportunities for enhanced business performance through the adoption of AI technologies, including increased competitiveness, greater innovation

	savings, and improved customer satisfaction through more efficient services and products.	capability, and improved customer experience. At the same time, the need to balance potential benefits with challenges such as initial investments, training, and integration of new systems was noted.
Management of Ethical and Legal Issues	Participants recognized the importance of addressing ethical and legal issues related to the use of AI technologies, including issues of privacy, fairness, responsibility, and compliance with applicable laws and regulations.	SMEs experienced similar challenges and opportunities regarding ethical and legal issues but also noted the need to adapt to rapidly changing regulatory environments and the complexity of navigating through various legal and ethical requirements.
Change Management and Organizational Culture	Participants stressed the importance of effective change management and organizational culture in successfully integrating AI technologies. A proactive approach to change management, which includes clear communication, leadership involvement, and fostering a culture conducive to innovation and continuous learning, was deemed essential for addressing resistance to change and facilitating a smooth transition to an AI-driven operation.	SMEs emphasized the significance of tailored change management and organizational culture to suit their specific needs and resource limitations. Flexibility, transparent communication, and involvement from all organizational levels were highlighted as critical factors in establishing an environment conducive to the successful integration of AI technologies and fostering a culture of ongoing learning and innovation.

4.5 Case study analysis

The quantitative analysis of the case study provides valuable insights into how the results relate to the needs and expectations of Enterprises A, B, and C. The analysis highlights the importance of transparency and explainability in AI systems, which is reflected in the participants' responses across the different iterations. These quantitative findings are then linked to the qualitative data collected from expert interviews and questionnaires, providing a comprehensive understanding of the participants' perspectives.

For Enterprise A, the quantitative data shows that participants have varying levels of understanding of AI technologies, with a significant proportion indicating limited familiarity. This aligns with the qualitative feedback from the interviews, where respondents emphasized the need for more educational initiatives to increase AI literacy. In the first iteration of the survey, it was found that many participants only used AI to a limited extent, often through common web services like translation and chatbots. This limited use reflects a general uncertainty and lack of trust in more complex AI applications, something directly influenced by the transparency and explainability of AI systems (Tchuente et al. 2024; Adadi & Berrada 2018). The results indicate that a targeted strategy for implementing XAI could address these concerns. Enterprise A should focus on developing clear communication strategies and providing training programs to improve employees' understanding of AI. By doing so, Enterprise A could increase trust in AI technologies and ensure smoother integration into their business processes (Ali et al. 2023; Behl et al. 2023).

Participants from Enterprise B showed a slightly better understanding of AI compared to their counterparts in Enterprise A, but there was still a noticeable knowledge gap regarding specific AI operational processes. The qualitative interviews revealed an interest in leveraging AI to optimize production and improve efficiency, but there was hesitation due to the perceived complexity and lack of explainability. In the second iteration, it was found that participants valued explainable and transparent AI decisions, especially for more complicated decisions where insight and understanding were critical for building trust (Jamshidi et al. 2023). For Enterprise B, the results indicate the necessity of integrating XAI into their AI strategies. By making AI systems more transparent and explainable, Enterprise B could foster a more positive attitude towards AI technologies. Additionally, the company should focus on leadership initiatives that promote clear communication about the potential benefits and applications of AI, addressing the uncertainties expressed by the participants (Ali et al. 2023).

Participants from Enterprise C demonstrated the highest level of AI understanding among the three enterprises, but there was consensus on the need for improved transparency and explainability in AI systems to facilitate better decision-making processes. The qualitative data from the interviews indicated that participants were particularly concerned about the ethical implications and potential biases in AI systems. Respondents expressed a desire for XAI systems to understand and audit decisions, which would increase their trust in the technology and its applications in sensitive areas such as data security and ethical decision-making (Gerlich 2023; Bozorgpanah et al. 2022). The quantitative data from the surveys also revealed that participants across all three enterprises experienced a general lack of transparency and explainability in AI systems, affecting their willingness to trust and adopt these technologies. This finding is supported by the theory on XAI and transparency, which emphasizes that explainability and insight are crucial for building trust and facilitating the adoption of AI technologies (Adadi & Berrada 2018; Tchuente et al. 2024).

In the first iteration of the survey, participants indicated that their knowledge of AI technologies was limited. The majority chose levels 1 to 3 on a five-point scale, where level 5 represented very good knowledge. This indicates a significant knowledge gap that needs to be addressed through education and clear communication (Ali et al. 2023). This lack of knowledge directly impacts their trust in AI, making the implementation of XAI even more critical to provide clear and understandable explanations of AI system decisions (Behl et al. 2023). The second iteration focused on providing deeper insights into specific AI models and their potential applications. The results showed that participants had an increased understanding of how XAI could contribute to making AI decisions more transparent and reliable. Several participants emphasized that transparency in AI systems is not only important for building trust but also for handling sensitive information and identifying potential areas for improvement within the business (Madsen et al. 2023; Hamm et al. 2023).

The compilation of the findings shows that participants emphasized the necessity of increased trust and understanding of AI technologies to foster interest in their adoption. Trust was enhanced through transparency in the decision-making process and XAI, leading to a deeper understanding of the technology's benefits and limitations. Enterprises A, B, and C observed a positive correlation between increased trust and understanding of AI and heightened interest in adopting these technologies within their operations. Improved understanding of the benefits and challenges through XAI and transparency helped Enterprises A, B, and C identify potentially use cases where AI could add value and increase competitiveness.

Participants also identified various organizational challenges and opportunities affecting the adoption of AI technologies, including resource constraints, cultural factors, and the need for education and skill development. Opportunities included increased productivity, cost savings, and improved customer satisfaction through more efficient services and products. Enterprises A, B, and C faced similar organizational challenges and opportunities but also encountered unique barriers and advantages due to their size and scope of operations, such as the need for flexibility, adaptation, and the benefits of agility and speed in decision-making. Furthermore, participants recognized the potential for improved business performance through AI adoption, including increased productivity, cost savings, and enhanced customer satisfaction. SMEs saw opportunities for increased competitiveness, innovation capability, and improved customer experience, while also acknowledging the need to balance benefits with challenges such as initial investments, training, and system integration.

The participants also emphasized the importance of addressing ethical and legal issues related to AI technologies, including privacy, fairness, responsibility, and compliance. Enterprises A, B, and C experienced similar challenges and opportunities in this regard but highlighted the need to adapt to rapidly changing regulatory environments and navigate complex legal and ethical requirements. Effective change management and organizational culture were deemed crucial for the successful integration of AI technologies. A proactive approach, including clear communication, leadership involvement, and fostering a culture of innovation and continuous learning, was considered essential for addressing resistance to change. Enterprises A, B, and C stressed the importance of tailored change management and organizational culture, emphasizing flexibility, transparent communication, and involvement from all organizational levels as critical factors for creating an environment conducive to the successful integration of AI technologies and fostering ongoing learning and innovation.

Madsen, Reddy, and Chandar (2023) conducted a study on SMEs utilizing AI in critical decision-making processes, showing that XAI significantly improves users' understanding of AI decisions, which enhances trust and acceptance of AI systems. The study highlights the importance of providing explanations tailored to the specific needs of different stakeholders within SMEs. Most respondents in enterprises A, B and C seemed to consider the third axis post-hoc explanation to be the most important XAI model. Hamm et al (2023) tested the effect of post-hoc explanations which demonstrated even more clearly the importance of this. They used SHAP on user perceptions where participants were tasked with identifying forged signatures where they found that SHAP explanations improved the perceived explainability of AI decisions, significantly enhancing users' trust and perceived usefulness of the AI system. Which show that an intuitive and engaging interface leads to higher levels of perceived explainability, which strengthens trust and satisfaction with the AI systems for the SMEs.

Gerlich (2023) interprets the European Commission's principles for trustworthy AI, emphasizing data protection, security, transparency, accountability, and fairness. These principles serve as guidelines to ensure that AI technologies are developed, implemented, and used ethically and responsibly, in line with societal values. By adhering to these principles, enterprises A, B and C could potentially build trust in AI systems among users and stakeholders, where XAI is crucial for building transparency and trust in AI systems within these SMEs. By making AI decision-making processes more understandable and transparent, XAI helps SMEs navigate the challenges of AI adoption and fosters trust among users and stakeholders. Empirical evidence underscores the importance of tailored explanations, intuitive interfaces, and adherence to ethical principles in enhancing the transparency and trustworthiness of AI systems, facilitating broader acceptance and integration of AI technologies in SMEs.

The quantitative and qualitative analysis shows that transparency and explainability are critical for building trust and facilitating the adoption of AI technologies in SMEs. Enterprises A, B, and C should invest in educational programs and develop clear communication strategies to improve employees' understanding of AI. The implementation of XAI is essential for making AI systems more transparent and understandable, which in turn will increase users' trust and acceptance. By addressing the specific needs and concerns of each enterprise, this case study offers actionable insights that could guide future AI strategies and implementations. Empirical evidence from studies by Khakurel and Rawat (2022) and Kartikeya and Arai (2022) supports the importance of XAI in strengthening trust in AI systems, which is crucial for achieving broader acceptance and integration of AI technologies in SMEs.

5. CONCLUSIONS AND RECOMMENDATIONS

This chapter presents the conclusions drawn from the study and offers recommendations based on the findings. The research questions will also be answered to draw these conclusions and connect back to the study's purpose.

5.1 Answering the research questions

This study investigates the impact of XAI and transparency on trust levels and the adoption of AI technologies within SMEs, guided by three specific research questions. The findings from the systematic literature review and expert interviews offer comprehensive insights into these queries.

[1] What is the current state of research on explainable AI and trust in the context of SMEs?

The systematic literature review reveals that research on XAI within SMEs is limited but present, with a predominant focus on operational transparency and user trust. Researchers like Chan (2022) and Khan et al. (2022) highlight the critical gap between theoretical advancements and practical applications. They stress the need for AI frameworks with XAI that are tailored to the unique environments of SMEs, where resources are limited, and technical expertise may be scarce. Moreover, the GDPR mandates the adoption of XAI to enhance trust and compliance, underscoring the necessity for SMEs to integrate transparent AI systems.

Empirical findings from the expert interviews reinforce these points. It has been noted through these interviews that while theoretical knowledge about XAI is growing, practical applications within SMEs lag. This gap is often due to the complex nature of AI algorithms and the limited technical resources available to SMEs. Without clear explanations and transparency about how these algorithms function, SMEs struggle to integrate and fully leverage AI technologies. Future research should therefore prioritize developing user-friendly and easily understandable XAI solutions tailored to SMEs' specific needs and constraints.

[2] Does explainable AI and transparency affect SMEs attitudes towards using AI technologies?

Insights from both the systematic literature review and expert interviews confirm that XAI and transparency significantly influence SMEs' attitudes towards AI technologies. Transparency is essential not only for regulatory compliance but also for building user trust and acceptance. The case study of Enterprises A, B, and C illustrate this vividly. Transparency in AI processes demystify AI operations, enhancing user confidence and willingness to engage with AI technologies. Users from these enterprises reported greater comfort and trust in AI systems where decision processes were clear and justifiable.

For example, in Enterprise A, an online retailer of premium men's clothing, the potential implementation of XAI could be crucial for building customer trust by explaining product recommendations. In Enterprise B, which is involved in designing and manufacturing advanced laundry systems, XAI could play a significant role in optimizing operational efficiency by providing clear explanations of AI-driven decisions, thereby fostering trust among their staff. Similarly, in Enterprise C, a manufacturer of customized metal products, the integration of XAI could enhance design and manufacturing processes by ensuring that AI decisions related to production methods and material usage are transparent and easily understandable. These scenarios illustrate how XAI can be vital in fostering trust and facilitating the adoption of AI technologies within these enterprises.

Supporting literature, such as the study by Nguyen et al. (2023), similarly demonstrates that transparency and explainability in AI systems strongly influence user attitudes and acceptance within SMEs. When users have access to transparent and XAI systems, they feel more comfortable interacting with and using these technologies in their daily operations. This suggests that investments in XAI can increase trust in AI within SMEs and enhance their readiness to integrate and benefit from these technologies, ultimately improving efficiency and competitiveness.

[3] Does improved trust and understanding help to facilitate the adoption of AI technology within SMEs?

Research unequivocally supports the notion that enhanced trust and understanding, facilitated by XAI, are pivotal in adopting AI technologies in SMEs. Expert interviews highlighted that increased transparency and explainability are essential in mitigating AI-related anxieties and resistance among users. A clearer comprehension of AI technologies fosters a broader organizational trust landscape, crucial for adopting these technologies more widely. The empirical data suggest that trust and understanding reduce barriers to technology acceptance, enabling smoother integration and operationalization of AI within SMEs.

In Enterprise A, the potential implementation of XAI could clarify how AI recommendations are made, which would likely increase customer trust and engagement. For Enterprise B, incorporating XAI could improve internal trust and operational efficiency by making AI decisions transparent and easily understandable for their staff. In Enterprise C, integrating XAI could enhance the reliability of their production processes by ensuring that AI-driven decisions are clearly justified and trustworthy. These examples highlight how XAI can play a critical role in fostering trust and facilitating the adoption of AI technologies within these enterprises.

Garcia et al. (2023) provide further evidence, showing that increased understanding and trust in AI technologies within SMEs lead to a greater propensity to adopt these technologies. Their study underscores the importance of providing education and support to SMEs to enhance their understanding of AI and its applications. When companies have sufficient knowledge and confidence in AI technologies, they are more successful in integrating these systems into their operations, thus capitalizing on their potential to improve productivity and innovation. This highlights the critical role of investing in education and support programs to promote AI adoption within the SME sector.

5.2 Conclusions

The research underscores the vital role of XAI in AI systems within SMEs, particularly in making AI decision-making processes transparent. Such transparency is crucial due to the typical resource constraints and limited technical expertise in these enterprises. By elucidating how AI systems function, XAI allows SMEs to effectively harness these technologies, enhancing operational efficiency and fostering an environment where technology could be trusted and ethically applied.

The findings from this study highlight the importance of integrating XAI into the operational frameworks of SMEs. Transparent AI systems do not merely ease the adoption of technology; they also enhance trust among users and stakeholders. This trust is essential as it reinforces the reliability and fairness of AI systems. In an era where ethical considerations are as critical as technological advancements, being able to demonstrate transparent and understandable AI decision-making processes is invaluable. The impact of XAI on user attitudes towards AI technologies within SMEs is significant. Enhanced transparency and clarity provided by XAI help bridge the gap between AI capabilities and user expectations, leading to increased trust. This increased trust facilitates a broader integration of AI solutions into daily business operations. For SMEs, balancing the integration of advanced technologies with ethical considerations and adherence to regulatory requirements is crucial, and XAI plays a pivotal role in this balance.

XAI is indispensable in SMEs not just for demystifying AI technology but also as a cornerstone for establishing a sustainable technological trust framework. This framework enables SMEs to navigate the challenges of digital transformation effectively. As these enterprises strive to remain competitive and compliant in a digitally driven market, XAI not only boosts their operational capabilities but also ensures that their technological strategies align with broader business ethics and regulatory compliance, making XAI a critical enabler of ethical business practices and long-term sustainability in the digital age for SMEs.

5.3 Recommendations

With the conclusions of the study in consideration, the following recommendations for organizations to enhance the adoption and effectiveness of XAI in AI systems for SMEs are following:

- SMEs should include XAI aspects to AI frameworks to increase transparency and understanding of AI systems, increasing user confidence, and facilitating smoother integration.

- Develop and provide ongoing education and training programs on AI and XAI for employees to emphasize the importance of transparency and the role of XAI in demystifying AI processes.

- Collaborate with AI developers to create and use user-centric tools that provide clear, understandable explanations of AI decisions, making AI technologies more accessible and trusted.

- Conduct regular audits of AI systems to ensure they meet ethical standards and transparency requirements, with updates based on the latest XAI research and user feedback.

- Foster an organizational culture that values and promotes transparency in AI implementations, encouraging open discussions about AI decisions and their impacts.

- View XAI as a strategic asset that could offer a competitive advantage by building trust with customers and stakeholders through transparent and ethical AI practices.

- Engage with external XAI experts and academic institutions to stay updated with the latest developments and best practices in XAI.

- Form multi-disciplinary teams to manage AI projects, ensuring diverse perspectives are considered, from ethical implications to technical requirements and business needs, enhancing the robustness and acceptability of AI solutions.

6. DISCUSSION

In the following chapter, the results and methodology are first discussed. The study's implications are then presented, and the chapter is concluded with suggestions for future research areas.

6.1 Discussion of Results

The research presents a case study, which could restrict the generalizability of the findings to a broader context. This is due to the inherent characteristics of case studies, where a smaller research group is used to investigate a broader research object. Despite potential limitations in generalizability, case studies offer detailed exploration of specific cases and phenomena, providing insights that may not be easily obtained through other research methods (Thomas 2021). The participants represented three companies within the SME sector. It is worth noting that these companies, while representative of the SME sector in some respects, may not fully encapsulate the diversity of the entire sector, see Figure 2. But the consistency of participants' perceptions of XAI and transparency suggests that the findings could be applicable to a broader range of SMEs. Another aspect to address is that even though several theories and strategies for XAI and transparency in the implementation of AI in SMEs are discussed, they have not been tested or verified. This limits the results of the study to a theoretical framework rather than empirical validation. The research also identified that while previous research for the topic is linked to SMEs is extensive, there exists a critical gap in practical application analysis compared to theoretical knowledge. This emphasizes the importance of further research on practical applications.

6.2 Discussion of Methodology

The completed preliminary study enabled the definition of the problem, formulation of the purpose, and identification of research questions relevant to the study. The basis of the theoretical framework was established through the research delimitations. This process was guided by an extensive literature review, as Thomas (2021) deems it essential for all research, aiming to identify existing research gaps and relevant theoretical foundations as starting points for the study. The expert interview method was then employed to further examine the research area, involving participants from the case companies. Two rounds of questionnaires and controlled feedback were conducted, indicating a consensus among participants on various aspects of the research problem. The results obtained from the expert interviews were instrumental in refining the research questions, validating the theoretical framework, and providing a broader understanding of the problem at hand.

Landeta (2006) emphasizes that the flexible nature of iterative interview methods means that the approach could be adapted according to the situation, the knowledge of the participants and the number of iterations possible. This adaptability ensured that the approach could be tailored to the specific needs of the study, enhancing its effectiveness, validity, and reliability. Indeed, it turned out that the participants in the study had less knowledge of the field than anticipated and therefore an adaptation and adjustment was required to fulfill the purpose of the study and greater consensus could be inferred for the later round. This knowledge gap also confirms the results and the need for information dissemination and education within SMEs.

6.3 Implications

This thesis has explored the role of XAI in enhancing transparency within AI systems for SMEs, focusing on how this could build trust and facilitate the integration of AI technologies. The findings suggest that XAI could significantly demystify AI operations, making the technology's decision-making processes clear and justifiable to end-users, which is crucial for gaining their trust and acceptance. Specifically for SMEs, where resources and technical expertise may be limited, XAI offers a tailored approach to understand and leverage AI effectively. In contexts similar to those studied, XAI could provide specific strategies and actionable insights that help SMEs overcome barriers to AI adoption, promoting more widespread and confident use of AI technologies. The research emphasizes the importance of transparency and explainability in creating trustworthy AI systems within SMEs.

6.3.1 Theoretical Implications

The examination of XAI and transparency within AI systems for SMEs significantly contributes to the ongoing discourse in AI ethics and usability. While existing theories around AI functionality stress the importance of accuracy and efficiency, the research pivots the focus towards understanding and transparency as pivotal factors in fostering trust and operational integration. Although the study does not introduce new theoretical frameworks, it provides empirical evidence that reinforces the critical role of XAI in enhancing transparency. This is crucial for SMEs where the adoption of technology must be accompanied by a solid understanding due to limited resources and expertise. The findings support the theory that increased transparency not only aids compliance with regulations but also elevates user trust and system acceptance.

The research extends the dialogue on the practical applications of XAI by highlighting how SMEs could leverage transparent AI systems for better decision-making and operational effectiveness. The necessity for AI systems to be understandable to non-expert users is emphasized, which aligns with current trends in human-centered AI development. This study serves as a foundation for further theoretical exploration into how transparency in AI could be effectively measured and implemented across diverse business environments. It invites future research to explore the nuances of XAI in sectors beyond SMEs, potentially leading to broader theoretical advancements in how transparency is integrated and valued in AI systems globally.

6.3.2 Practical Implications

The study on XAI and transparency in AI systems for SMEs provides insights that could guide organizations in their efforts to understand and implement these technologies. Although the research primarily presents measures that could help increase understanding of why such technologies should be implemented, the results could also serve as a basis for developing and refining strategies for more effective use of AI. Practitioners within SMEs could benefit from these insights to address specific needs and challenges associated with AI technologies, particularly to enhance transparency and understanding in their operational processes. The research highlights how organizations could identify critical factors and motivates why it is important to take initiatives to act and prioritize transparency and understanding in their operations.

The study emphasizes the importance of integrating XAI principles to improve user experience and customer satisfaction. By making the decision-making processes within AI systems more understandable and transparent, companies could build stronger trust among users and other stakeholders. This is of great importance for SMEs, where technical expertise may often be limited and where there is a clear need to make the technology accessible and understandable to all users. The study serves as a source of increased awareness and understanding of XAI and transparency. It offers a platform for SMEs to design their AI strategies responsibly, which not only improves their operational efficiency but also strengthens their market position by demonstrating a commitment to responsible use of technology.

6.3.3 Social Implications

Social implications refer to the significance of the study's findings for societal aspects, highlighting how organizations, especially SMEs, could adopt XAI to foster a more transparent and reliable environment. The research emphasizes the societal need for SMEs to adopt greater transparency in AI applications, which is crucial for building public trust. It is stressed that the responsibility to promote transparency should not be isolated within specific departments or roles, instead it should be a collective endeavor across all levels of the organization.

Global initiatives like the EU's regulations on AI transparency underscore the broader societal drive towards ethical AI practices. These regulations reflect a growing consensus on the need for technologies that are not only effective but also understandable and fair, in line with broader societal values and legal standards. The study also highlights how improved transparency in AI could serve as a catalyst for organizations to reduce potential biases and ethical risks in their operations. By doing this, SMEs not only meet legal standards but also contribute positively to social justice and equality. This aligns with societal goals to ensure that technology serves the common good and adheres to ethical standards.

This research illustrates how the adoption of XAI could lead to significant societal benefits, including reduced biases, increased accountability, and enhanced public trust in AI technologies. It suggests that policymakers and regulators should consider these findings when designing guidelines that promote transparency and accountability in AI, to ensure that technological advancements align with societal values and ethical standards. Such measures are crucial for a sustainable integration of AI across various sectors, enhancing technology's role in promoting societal well-being.

6.4 Further Research

Further research should aim to empirically test the proposed theories and strategies to provide more robust insights into the implementation of AI with XAI and transparency in SMEs. By addressing these areas, future research could provide a deeper understanding of the challenges and opportunities associated with XAI in SME settings. One avenue is to evaluate the effectiveness of XAI techniques in the environment and comparative studies could also be conducted to assess the effectiveness of different XAI techniques for different organizations within the SME sector. By testing various methods in real-world settings, researchers could evaluate how well each technique fosters user understanding and trust in AI systems. Additionally, it is crucial to investigate how XAI affects the decision-making process within SMEs, by examining whether XAI empowers SME users to make better decisions with AI or if it leads to information overload.

Another area for future research is the development of XAI tools tailored for specific SME needs. There is a need to explore the development of XAI tools tailored for specific industries or applications relevant to SMEs. This means customizing explanations to the terminology and workflows common in particular sectors could significantly enhance the usability and effectiveness of XAI settings. Additionally, understanding user trust and XAI in SMEs is also a crucial area for future research. Moreover, investigating how XAI could be used to build trust in AI systems over extended periods of use within SMEs is essential. Research should examine whether the initial transparency provided by XAI diminishes over time and how trust in AI systems could be maintained over extended periods of use within SMEs. Additionally, exploring how the level of technical expertise of SME users affects their perception of XAI explanations is critical. Understanding whether users with limited technical backgrounds require fundamentally different XAI approaches compared to more technical users could inform the development of more user-friendly XAI solutions. These areas for future research highlight the need for further investigation into the implementation of XAI and transparency in SMEs. By addressing these unanswered questions, future researchers could contribute to the advancement of this developing field and build upon the findings of this study.

REFERENCES

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

Ali, S., Abuhmed, T., El-Sappagh, S., Muhammad, K., Alonso-Moral, JM., Confalonieri, R., Guidotti, R., Del Ser, J., Díaz-Rodríguez, N. & Herrera, F. (2023). Explainable Artificial Intelligence (XAI): What we know and what is left to attain Trustworthy Artificial Intelligence. *INFORMATION FUSION*. vol. 99. DOI:10.1016/j.inffus.2023.101805

Allwood, CM. (2021). The place of the qualitative approach in the theory of science. In Klingberg, G. & Hallberg, U. (red). *Qualitative methods.* Lund: Studentlitteratur.

Arora, M. & Chopra, A. (2023). Explainability for Clustering Models. *SOFT COMPUTING IN DATA SCIENCE*. vol. 1771, pp. 3-1. DOI:10.1007/978-981-99-0405-1_1

Behl, A., Sampat, B., Pereira, V. & Jabbour, CJC. (2023).
The role played by responsible artificial intelligence (RAI) in improving supply chain perform ance in the MSME sector: an empirical inquiry. ANNALS OF OPERATIONS RESEARCH. DOI:10.1007/s10479-023-05624-8

Bozorgpanah, A., Torra, V. & Aliahmadipour, L. (2022). Privacy and Explainability: The Effects of Data Protection on Shapley Values. *TECHNOLOGIES*, vol.10 (6). DOI:10.3390/technologies10060125

Chan, GKY. (2022). AI employment decision-making: integrating the equal opportunity merit principle and explainable AI. *AI & SOCIETY*. DOI:10.1007/s00146-022-01532-w

Finke, J., Horwath, I., Matzner, T. & Schulz, C. (2022). (De)Coding Social Practice in the Field of XAI: Towards a Co-constructive Framework of Explanations and Understanding Between Lay Users and Algorithmic Systems. *LECTURE NOTES IN ARTIFICIAL INTELLIGENCE*, vol 13336, pp. 149-160. DOI10.1007/978-3-031-05643-7_10

Fleisher, W. (2022). Understanding, Idealization, and Explainable AI. *EPISTEME-A JOURNAL OF INDIVIDUAL AND SOCIAL EPISTEMOLOGY*, 19 (4), pp. 534-560. DOI:10.1017/epi.2022.39

Haenlein, M. & Kaplan, A. (2019). “A Brief History of Artificial Intelligence: On the Past, Present, and Future of Artificial Intelligence”, *CALIFORNIA MANAGEMENT REVIEW,* vol. 61, no. 4, pp. 5–14. DOI:10.1177/0008125619864925

Hamm, P., Klesel, M., Coberger, P. & Wittmann, HF. (2023). Explanation matters: An experimental study on explainable AI. *ELECTRONIC MARKETS*, vol. 33, no. 1. DOI:10.1007/s12525-023-00640-9

Hasija, A. & Esper, TL. (2022). In artificial intelligence (AI) we trust: A qualitative investigation of AI technology acceptance. *JOURNAL OF BUSINESS LOGISTICS,* 43 (3), pp. 388-412. DOI:10.1111/jbl.12301

Issa, A., Lucke, D. & Bauernhansl, T. (2017). "Mobilizing SMEs Towards Industry 4.0-enabled Smart Products", *Procedia CIRP*, vol. 63, pp. 670–674. DOI:10.1016/J.PROCIR.2017.03.346

Jamshidi, S., Nowaczyk, S., Fanaee-T. & Rahat, M. (2023). A Systematic Approach for Tracking the Evolution of XAI as a Field of Research. *MACHINE LEARNING AND PRINCIPLES AND PRACTICE OF KNOWLEDGE DISCOVERY IN DATABASES*, vol. 1753, pp. 461-476. DOI:10.1007/978-3-031-23633-4_31

Kartikeya, A. & Arai, K. (2022). Examining Correlation Between Trust and Transparency with Explainable Artificial Intelligence. INTELLIGENT COMPUTING, VOL 2, vol 507, pp. 353-358. DOI:10.1007/978-3-031-10464-0_23

Khakurel, U. & Rawat, DB. (2022). Evaluating Explainable Artificial Intelligence (XAI): Algorithmic Explanations for Transparency and Trustworthiness of ML Algorithms and AI Systems. *ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING FOR MULTI-DOMAIN OPERATIONS APPLICATIONS IV,* vol 12113. DOI:10.1117/12.2620598

Khan, MS., Nayebpour, M., Li, MH., El-Amine, H., Koizumi, N. & James L. Olds. (2022). Explainable AI: A Neurally-Inspired Decision Stack Framework. *BIOMIMETICS*, 7 (3), Article 127. DOI:10.3390/biomimetics7030127

Khoozani, ZS., Sabri, AQM., Seng, WC., Seera, M. & Eg, KY. (2024). Navigating the landscape of concept supported XAI: Challenges, innovations, and future directions. *MULTIMEDIA TOOLS AND APPLICATIONS*. DOI:10.1007/s11042-023-17666-y

Landeta, J. (2006) Current validity of the Delphi method in social sciences. *TECHNOLOGICAL FORECASTNING AND SOCIAL CHANGE,* 73 (5), pp. 467-482. DOI:10.1016/j.techfore.2005.09.002

Lecue, F. (2020). On the role of knowledge graphs in explainable AI. *SEMANTIC WEB*, 11 (1), p. 41-51. DOI:10.3233/SW-190374

Loo, R. (2002). The Delphi method: a powerful tool for strategic management. *POLICING-AN INTERNATIONAL JOURNAL OF POLICE STRATEGIES & MANAGEMENT*, 25 (4), p. 762-769. DOI: 10.1108/13639510210450677

López-Gómez, E. (2018). The Delphi method in Current Educational Research: A Theoretical and Methodological Review. *EDUCATIÓN XX1*, 21 (1), p. 17-40. DOI: 10.5944/educXX1.15536

Madsen, A., Reddy, S. & Chandar, S. (2023). Post-hoc Interpretability for Neural NLP: A Survey. *ACM COMPUTING SURVEYS*, vol 55 (8). DOI:10.1145/3546577

Medianovskyi, K & Pietarinen, AV. (2022). On Explainable AI and Abductive Inference. *PHILOSOPHIES*, 7 (2), Article 35. DOI:10.3390/philosophies7020035

Patel, R. & Davidson, B. (2019). Forskningsmetodikens grunder – Att planera, genomföra och rapportera en undersökning. Lund: Studentlitteratur.

Rauch E., Dallasega, P., & Unterhofer, M. (2019). "Requirements and Barriers for Introducing Smart Manufacturing in Small and Medium-Sized Enterprises," *IEEE Eng*. Manag. Rev., vol. 47, no. 3, pp. 87–94. DOI: 10.1109/EMR.2019.2931564

Stassin, S., Corduant, V., Mahmoudi, SA. & Siebert, X. (2024). "Explainability and Evaluation of Vision Transformers: An In-Depth Experimental Study," *ELECTRONICS*, 13 (1), Article 175. DOI:10.3390/electronics13010175

Saeed, W. & Omlin. C. (2023). Explainable AI (XAI): A systematic meta-survey of current challenges and future opportunities. *KNOWLEDGE-BASED SYSTEMS*, 263, Article 110273. DOI:10.1016/j.knosys.2023.110273

Tchuente, D., Lonlac, J. & Kamsu-Foguem, B. (2024).
A methodological and theoretical framework for implementing explainable artificial intelligen ce (XAI) in business applications. *COMPUTERS IN INDUSTRY,* vol 155, Article 104044. DOI:10.1016/j.compind.2023.104044

Thomas, G. (2021). How to do a case study. Los Angeles: SAGE.

Thulesius, H. (2021). Grounded Theory. In Klingberg, G. & Hallberg, U. (red). *Qualitative methods.* Lund: Studentlitteratur.

Vorobeva, D., Pinto, DC., Antonio, N. & Mattila, AS. (2023). The augmentation effect of artificial intelligence: can AI framing shape customer acceptance of AI-based services? *CURRENT ISSUES IN TOURISM,* Article 2214353. DOI:10.1080/13683500.2023.2214353

Wulf, AJ & Seizov, O. (2024). "Please understand we cannot provide further information": evaluating content and transparency of GDP-mandated AI disclosures. *AI & SOCIETY,* vol. 39 (1). pp. 235-256. DOI:10.1007/s00146-022-01424-z

APPENDICES

The following questions were answered by participants from the companies A, B and C during the two iterations of the expert interviews. It is important to mention that participants received information and feedback during the process which is not included as it was done on an ongoing basis.

Appendix A: Questionnaire 1

1. How are AI technologies currently used in your organization?

2. Have you used AI technologies in your professional activities? Please describe.

3. How frequently do you interact with AI technologies in your work?

4. On a scale of 1 to 5, how familiar are you with AI technologies?

5. On a scale of 1 to 5, how well do you know your own organization’s functions today?

6. How do you perceive the impact of XAI in influencing your colleagues` attitudes towards adopting AI technologies?

7. How important do you believe transparency in AI systems is for building trust among your colleagues in considering AI adoption?

8. How do you think the introduction of XAI systems would impact the decisions-making processes within your organization?

9. What challenges or concerns do you foresee in implementing transparent AI systems within your organization, and how do you think these could be addressed effectively?

10. Could you provide examples of how XAI systems could be integrated into existing workflows within your organization to improve efficiency and decision-making?

11. In your opinion, what strategies or approaches would be most suitable for enhancing understanding and trust among your colleagues regarding AI technologies in the context of your organization?

Appendix B: Questionnaire 2

1. What specific functions or tasks would you like AI to perform to contribute to your organization? Additionally, could you share more about which processes or areas within your business you are considering integrating AI into?

2. Would your company be willing to implement an AI solution or model that doesn't provide any form of explanation for its decisions or actions? Why or why not?

3. Would your company be willing to implement an AI solution or model that doesn't provide any form of explanation for its decisions or actions, but has a transparent AI system? Why or why not?

4. How do you perceive the impact of XAI in influencing your colleagues' attitudes towards adopting AI technologies?

5. Do you consider XAI and transparency to be important for you and your colleagues' to be able to trust an AI system?

6. What scares you and your colleagues most about AI?

7. What aspects of integrity are most critical for you and your colleagues in AI systems?

8. Data Explainability: How does making the input data clear and easy to understand help check if AI model results are correct?

9. Model Explainability: Do you think it is important to be able to trace and understand how your AI models arrive at their conclusions?

10. Post-hoc Explainability: How do you value the ability to receive detailed explanations of AI decisions afterward to understand and improve your AI initiatives?

11. Assessment of Explanations: How do you prioritize the evaluation of the quality of explanations from your AI system to ensure they are understandable and useful?

12. Considering the four axes of explainability in AI (data, model, post-hoc, and assessment of explanations), which one or ones do you believe is most crucial for enhancing transparency and trust in AI technologies, and why?

HÖGSKOLAN
I BORÅS

Besöksadress: Allégatan 1 · Postadress: 501 90 Borås · Tfn: 033-435 40 00 · E-post: registrator@hb.se · Webb: www.hb.se