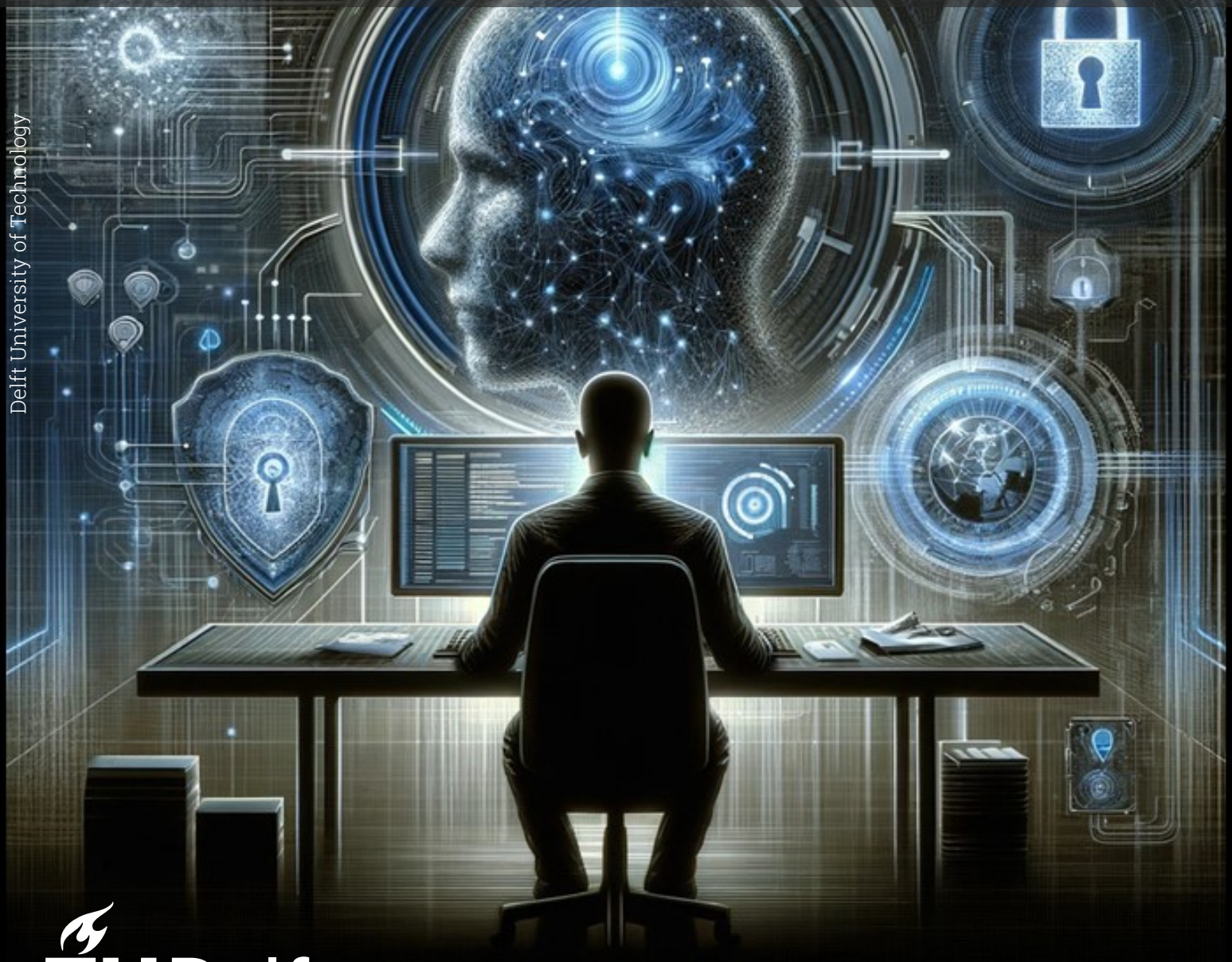
Adoption of AI in Cybersecurity

Bridging the Gap Between Innovation and Application

MSc Graduation Project

Stefani Slavova



Delft University of Technology

TU Delft

Adoption of AI in Cybersecurity

Bridging the Gap Between Innovation and Application

by

Stefani Slavova

to obtain the degree of Master of Science
at the Delft University of Technology,
to be defended publicly on Friday August 30, 2024 at 04:00 PM.

Project duration:	February, 2024 – August, 2024	
Thesis committee:	Y. Zhauniarovich,	TU Delft, Supervisor
	S. Azimi Rashti,	TU Delft, Supervisor
	M. van Eeten,	TU Delft, Supervisor, Chair
	E. Barbaro,	External Supervisor

Cover: Generated by AI through OpenAI's ChatGPT

An electronic version of this thesis is available at http://repository.tudelft.nl/.

TU Delft

Preface

Dear Reader,

Before you lies the culmination of my academic journey at TU Delft—the thesis I wrote on the adoption of artificial intelligence in cybersecurity. The past two years have been a fulfilling and enriching experience, and I believe the work presented in this report is a fitting conclusion to this chapter of my life. Cybersecurity was a relatively new field for me, and I have thoroughly enjoyed the learning experience that this thesis provided.

I would like to express my gratitude to everyone who has been part of this project. My deepest appreciation goes to the members of my committee. First and foremost, I would like to thank my first supervisor, Yury Zhauniarovich, and my external supervisor, Eduardo Barbaro, for their unwavering support, insightful feedback, and encouragement. Their expertise and dedication were invaluable in navigating the challenges that arose during this project. I am also grateful to my second supervisor, Sepinoud Azimi Rashti, and the chair of my committee, Michel van Eeten, for their critical input, which significantly elevated the quality of my work.

My heartfelt thanks also go to the participants in my study, whose willingness to share their experiences and insights made this research possible. Their contributions provided the empirical foundation upon which this thesis is built.

I am especially thankful to my partner, Andrei, for always supporting me and treating me with kindness during challenging times. To my friends, I am truly grateful for the emotional support and for being there to invoke interesting discussions that often sparked new ideas.

I would also like to acknowledge those who may be far away in distance yet remain ever-present in my thoughts. To my mother, Diana, and my father, Ivan, who have always encouraged me to follow my aspirations and supported me unconditionally along the way. Their life lessons have shown me the value of resilience, and their care has allowed me to thrive.

I hope this research advances our understanding of the complexities of AI adoption in cybersecurity. Although my contribution is small, I hope it paves the way for future students and researchers in this area.

Stefani Slavova
Rotterdam, August 2024

Summary

As digitalization continues to reshape industries, cybersecurity departments face an overwhelming number of alerts and potential threats. This leads to decision fatigue among security analysts, making it challenging to maintain robust security measures. In response, cybersecurity departments are turning to Artificial Intelligence (AI) solutions to automate routine tasks, prioritize alerts, and speed up incident responses. While AI holds great promise, industry trends show that companies are adopting AI-driven cybersecurity solutions at a slower pace than threat actors. This discrepancy suggests underlying challenges.

This thesis explores these challenges through a case study of a large European bank's cybersecurity department. By employing a sociotechnical systems (STS) approach, this study examines the interplay between social and technical factors within cybersecurity departments undergoing the transition to AI-enabled cybersecurity. Data were collected through semi-structured interviews with security analysts, data scientists, and leaders. These interviews were analyzed using reflexive thematic analysis, leading to the identification of four key themes.

1. **Mixed perceptions of AI-based solutions:** While AI has the potential to significantly enhance efficiency, there are varied perceptions among stakeholders. Benefits such as improved threat detection and reduced decision fatigue were recognized, but challenges like data quality issues, and the lack of transparency in AI systems were also prevalent.
2. **Organizational influence on AI adoption:** Organizational factors play a crucial role in the success of AI adoption. Senior management support, organizational readiness, and effective change management were identified as critical to overcoming resistance and facilitating smooth integration.
3. **Interdisciplinary development dynamics:** The interviews highlighted the importance of collaboration between data scientists and security analysts in developing AI solutions that are both technically sound and user-friendly. Continuous education and training were emphasized as necessary to empower employees and foster a culture of innovation.
4. **Building trust in AI systems:** Trust emerged as a significant factor in AI adoption. The study found that transparency in AI operations, clear explanations of AI-driven decisions, and active user participation in the development process are essential for building trust and ensuring the successful integration of AI technologies.

In addition to these thematic findings, a root cause analysis was conducted to explore in more depth the underlying issues that hinder the successful adoption of AI in cybersecurity. Based on these findings, the thesis proposes a new conceptual model that integrates both technical and social factors, offering bridging mechanisms to address the challenges of AI adoption in cybersecurity. This conceptual model enhances the understanding of the complex relationships between technology, processes, and people and provides practical recommendations for improving the development and deployment of AI systems. Implementing mentorship programs can address job security concerns and guide career development. Investing in continuous education equips employees with the necessary skills and fosters a culture of innovation. Aligning organizational strategies with AI initiatives and maintaining open communication channels are vital. Establishing effective feedback mechanisms ensures continuous improvement of AI systems, and ensuring transparency in AI operations builds trust and facilitates user acceptance.

Future research should focus on enhancing the generalizability of findings through diverse case studies that compare companies of different sizes and industries, and by incorporating mixed-method approaches. It should also develop human-centered AI frameworks that integrate social and technical factors, explore practical AI trust-building strategies, and conduct longitudinal studies to track AI integration and its impact on trust, cybersecurity roles, and collaboration.

Contents

1

Introduction

Imagine a day in the dynamic atmosphere of a cybersecurity department. Amidst the glow of monitors, analysts are meticulously monitoring and responding to a barrage of alerts. Each alert signals a potential threat, ranging from phishing attempts to sophisticated hacking efforts. The volume and complexity of these alerts make the task daunting, requiring quick, accurate decision-making to protect sensitive customer data. This scenario underscores a critical challenge faced by cybersecurity teams worldwide: managing an overwhelming number of alerts while maintaining robust security measures [1].

Problem statement

In this high-stakes environment, the integration of Artificial Intelligence (AI) offers a promising solution. By providing advanced analytical capabilities, AI has the potential to transform cybersecurity by automating routine tasks, prioritizing alerts, and speeding up incident responses [2]–[4]. However, implementing these solutions presents a set of challenges. Researchers highlight compatibility and configuration issues with legacy systems [4], [5], the need for training and building resilience among security professionals [3], and the difficulties in ensuring transparency and accountability in AI-driven systems [2]. It becomes evident that the success of this transition requires not only technological innovation but also effective change management.

On the market, multiple solutions offer AI capabilities to provide security detection and response services, such as those offered by Darktrace [6] and Cisco [7]. Despite the availability of such vendor solutions, some organizations choose to build and implement their custom AI-enabled solutions too. This allows them to tailor the solutions to the unique requirements and infrastructure of the organization and further strengthen their security operations. However, research shows that these organizations face challenges, such as the absence of well-defined AI strategies or weak technology and data foundations [5].

These challenges contribute to a broader issue observed in the field: although AI has the potential to significantly improve cybersecurity defenses, its adoption by cybersecurity teams has been notably slower than the pace at which threat actors are leveraging AI capabilities [8]. As the implementation of AI in cybersecurity is a relatively recent development [9], the literature offers little detail about the specific challenges and strategies organizations can employ to overcome them. This gap underscores the need for further research to understand the obstacles to AI adoption and how they can be addressed to enhance cybersecurity defenses.

Theoretical framework

We adopt the perspective that this complexity is best examined through the lens of sociotechnical system (STS) theory. STS theory emphasizes the interdependence between social and technical components within an organization and is well-suited for exploring complex, dynamic systems [10], [11] like those found in cybersecurity departments. Some of these interactions include how security professionals use AI tools, the feedback loop between users and developers, and the organizational support

required for successful implementation. By examining these interactions, the study can identify areas where alignment or misalignment occurs, leading to more effective integration of strategies.

In addition to sociotechnical system theory, this study also uses Hameed et al.'s conceptual model of innovation adoption in organizations [12]. While not a theory, Hameed et al.'s model provides a comprehensive framework for understanding how organizations adopt and integrate new technologies. It encompasses various stages of adoption, including initiation, adoption decision, and implementation, each influenced by multiple organizational characteristics. This model helps understand the processes by which organizations recognize the need for innovation, develop strategies for adoption, allocate resources, and overcome resistance to change [12]. By integrating STS theory with Hameed et al.'s conceptual model of innovation adoption in organizations, this study aims to offer a holistic understanding of both the technical and social factors that influence the successful adoption of AI in cybersecurity.

Purpose of this study

One of the central concepts in sociotechnical system theory is the *sociotechnical gap*–the mismatch between what is technically possible and what can be supported by social structures [13]. Every sociotechnical system inherently contains this gap, which must be understood and managed to ensure the effective implementation of new technologies [13]. Hence, understanding this gap is a crucial prerequisite for organizations to effectively develop and implement AI-based systems within their cybersecurity departments. Given that this gap has not been extensively studied in this specific context, this study aims to explore and answer the following main research question:

> *How can organizations bridge the sociotechnical gap in developing and using AI-based tools for cybersecurity?*

To address this research question, the study focuses on the following steps:

- To investigate the specific challenges organizations face in developing and implementing in-house AI cybersecurity solutions, including technical and social barriers.
- To examine the interactions between technical systems and social structures within cybersecurity departments to understand how these influence AI adoption.
- To provide actionable recommendations for organizations to enhance the development, implementation, and integration of AI-based cybersecurity systems, addressing both technical and social factors.

Research design

Figure 1.1 summarizes the research phases and methodologies used in this study. It provides a comprehensive overview of the inputs, processes, and outputs associated with each phase, facilitating a clear understanding of how the study is conducted.

To comprehensively address the main research question, it is essential to break it down into more manageable components that help us understand the sociotechnical dynamics in the adoption of AI in cybersecurity, analyze these dynamics through empirical insights, and propose strategies for improvement. The selected sub-research questions (sRQs) serve this purpose.

sRQ1: *What are the potential interactions between the social and technical components in the development and use of AI-based tools for cybersecurity?*

This question is fundamental because it addresses the sociotechnical perspective: the successful integration of AI in cybersecurity depends on the seamless interaction between social (people, organizational structures) and technical (AI models, infrastructure) components. Understanding these interactions helps identify the sociotechnical dynamics and potential friction points that could impact AI adoption.

sRQ2: *What insights can be gained from empirical results on the development and integration of AI-based tools into cybersecurity practices?*

Empirical insights provide a grounded understanding of how AI models are currently being integrated into cybersecurity practices. This question seeks to gather real-world experiences and challenges faced by security professionals, offering a practical perspective on the theoretical interactions explored

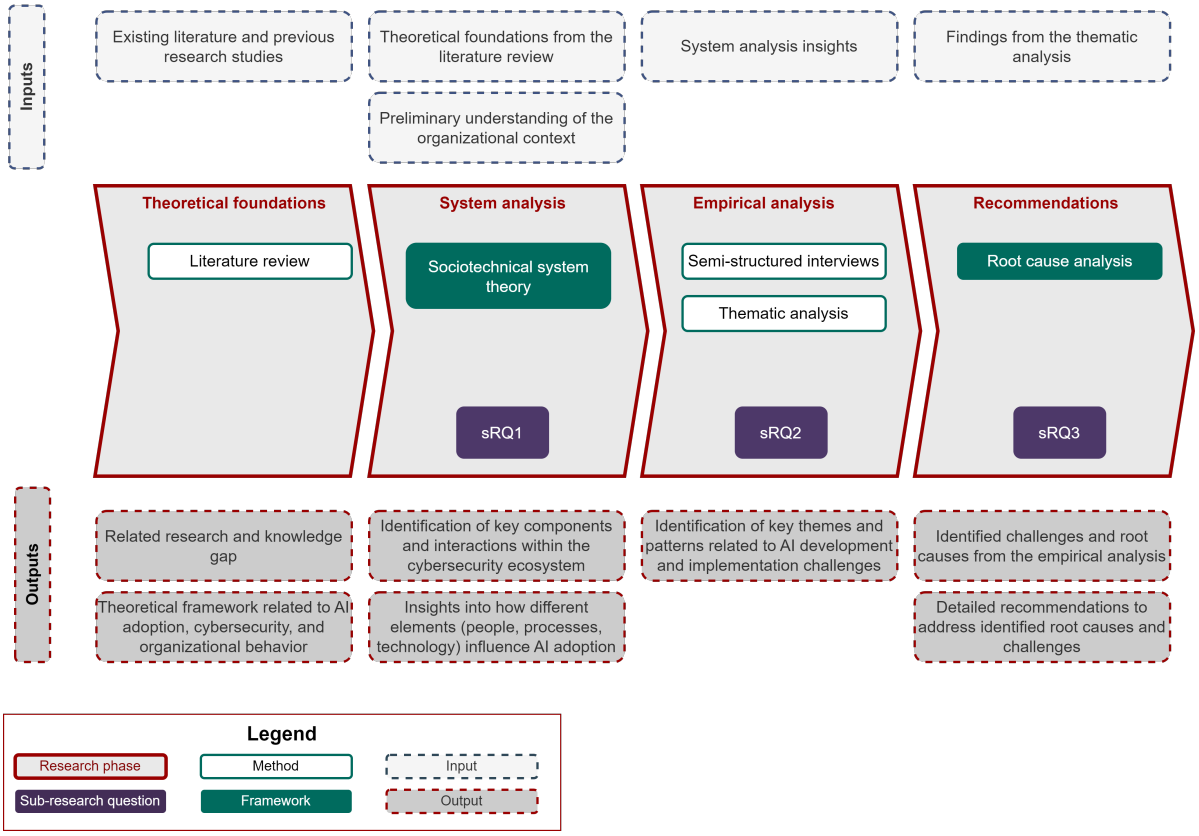


Figure 1.1: Overview of the research design and methodology phases.

in sRQ1. By analyzing empirical data, we can identify common themes, best practices, and barriers to successful AI adoption.

sRQ3:*What mechanisms can enhance the development and use of AI-based tools for cybersecurity in organizations?*

Building on the findings from sRQ1 and sRQ2, this question aims to develop actionable recommendations for improving AI integration. It focuses on providing practical solutions and strategies that organizations can implement to overcome identified challenges and enhance the effectiveness and acceptance of AI models in cybersecurity. The goal is to translate theoretical insights and empirical evidence into concrete steps that can drive positive change.

To achieve its objective, this study adopts a case study approach—a method well-suited for the in-depth exploration of a phenomenon within its real-life context [14]. This study analyzes the case of a large European bank, offering a comprehensive view of the processes, challenges, and outcomes associated with in-house AI development in a highly regulated industry. From a preliminary survey, we discovered a low adoption rate of AI within the bank's cybersecurity department. This finding aligns with broader industry trends [8], making the bank's cybersecurity department an ideal candidate for this case study. Data collection methods include semi-structured interviews with key stakeholders, including security analysts, data scientists, and leaders, ensuring a diverse range of perspectives. The interviews are analyzed through a reflexive thematic analysis, following Braun and Clarke's framework [15]. From the analysis, four main themes emerged that significantly impact the adoption of AI in the studied case. The findings were subsequently subjected to a root cause analysis to identify underlying issues and contributing factors. Based on this comprehensive analysis, several recommendations were developed, intended to provide insights for the studied organization. Lastly, the findings were synthesized, leading to the proposal of a new conceptual model that encapsulates the interactions between technical and social factors in the adoption of AI in cybersecurity, along with suggested bridging mechanisms.

Relation to MSc Complex Systems Engineering and Management

This research is firmly connected to the core principles of the Complex System Engineering and Management (CoSEM) program, which emphasizes the understanding, design, and management of complex, interdisciplinary systems. In the context of cybersecurity, the integration of Artificial Intelligence (AI) represents a system where technological advancements intersect with organizational and social dynamics. This study embodies the CoSEM approach by applying a sociotechnical systems lens to explore the multifaceted challenges and opportunities of AI adoption within a large financial institution. By addressing both the technical aspects of AI implementation and the social aspects that influence its success, the research aligns with the CoSEM program's focus on developing holistic, interdisciplinary solutions to complex problems. Additionally, the study's emphasis on innovation management and organizational behavior underscores the importance of effectively navigating the complexity inherent in the adoption of emerging technologies within organizational contexts, a central theme of the CoSEM curriculum.

Thesis structure

The remainder of this thesis is organized as follows:

- Chapter 2 provides a comprehensive literature review, covering the background of cybersecurity, AI, and the intersection of these fields, and the relevant theories.
- Chapter 3 details the methodology used in this research, including the case study context, data collection and analysis methods, and ethical considerations.
- Chapter 4 presents a sociotechnical system analysis, examining the potential interactions between social and technical components in the adoption of AI in cybersecurity.
- Chapter 5 outlines the results of the interviews, organized into key themes identified through thematic analysis.
- Chapter 6 discusses the implications of these findings and offers practical recommendations for organizations looking to adopt AI in cybersecurity.
- Chapter 7 synthesizes and reflects on the findings, compares them with other studies, highlighting unique insights and common challenges, and suggests areas for future research.
- Chapter 8 concludes the study.

2

Literature review

2.1. Background

2.1.1. Cybersecurity

Cybersecurity is a fuzzy concept defined by researchers in various ways. Fischer [16, p.1] describes it in broader terms as "the act of protecting ICT [Information and Communication Technology] systems and their contents". As technology rapidly advances, cybercrime evolves alongside it. Cyberattacks, driven by motives such as financial profit and operational disruption, intentionally take advantage of weaknesses in computer systems and represent a growing danger [17]. Attackers are constantly developing new techniques to gain unauthorized access to networks, applications, and data, aiming to compromise the confidentiality, integrity, and availability of information [18].

2.1.2. Artificial Intelligence

Artificial Intelligence (AI) is another complex concept with varied definitions. In this paper, we adopt a simplified definition from the European Commission's AI Act [19]. It describes AI systems as machine-based systems designed to work independently and adapt over time. They take in information to make predictions, create content, give recommendations, or make decisions that can affect the physical or digital worlds. Machine learning (ML), a pivotal branch of artificial intelligence, is focused on the development of algorithms that enable systems to learn from and make decisions based on data [20]. At its core, machine learning embodies the convergence of computational power and data-driven insights, leading to a new era of intelligent automation and decision-making. The field has gained significant momentum and is currently used across various sectors.

Within the domain of machine learning, a spectrum of techniques exists (see Figure 2.1). Some are simpler and inherently interpretable, such as decision trees, and offer a clear understanding of how decisions are made, which classifies them as transparent models [21]. Conversely, more sophisticated algorithms, including deep neural networks and ensemble methods, deliver superior performance but often do so at the cost of opacity. This underscores a challenge in the realm of AI: balancing effectiveness with understandability [22].

In response to the growing complexity and opacity of advanced machine learning models, the research community has given rise to the concept of explainable artificial intelligence (XAI). It refers to systems and methods that provide human-understandable explanations or interpretations of the internal mechanisms or outputs of an AI model [21]. The work of Arrieta et al. [21] presents a taxonomy of explainable artificial intelligence methods. Two of the most commonly used approaches in practice are Local Interpretable Model-agnostic Explanations (LIME) [1] [23] and SHapley Additive exPlanations (SHAP) [2] [24].

[1] LIME [23] creates simple models, like decision trees, to approximate and explain a complex model's predictions for a specific instance, highlighting which features were most influential.

[2] SHAP [24] allocates the contribution of each feature to the difference between the actual prediction and the average prediction, using concepts from cooperative game theory to ensure a fair distribution of contributions across features.

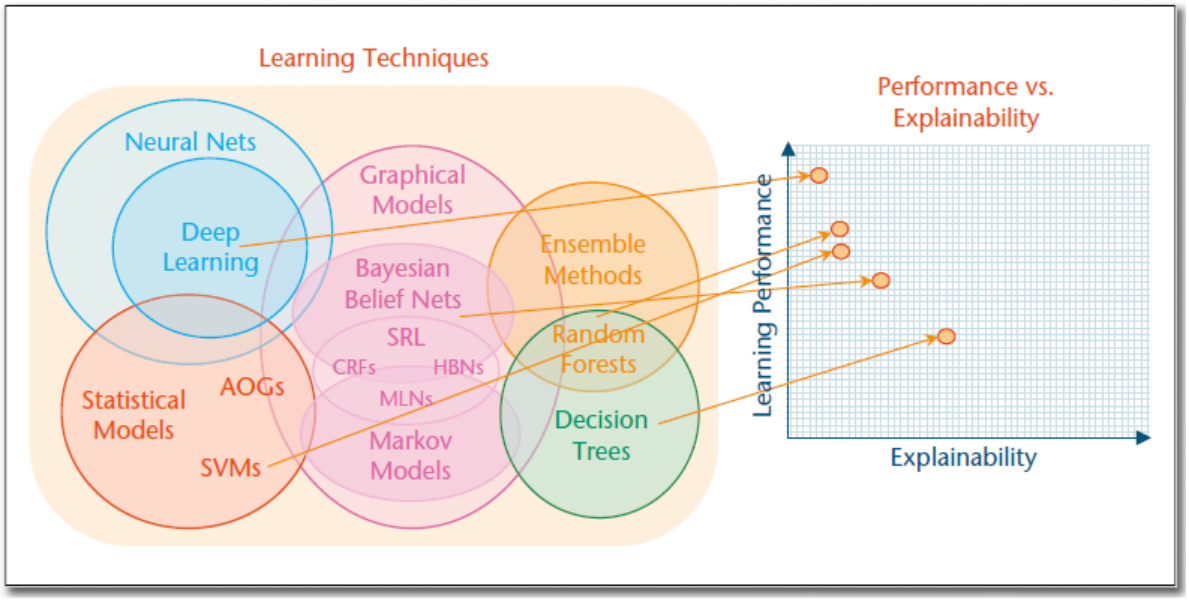


Figure 2.1: The trade-off between learning performance and explainability, from [22].

Explainable AI has become especially important for high-stakes environments. Inaccuracies in predictions are unlikely to lead to serious consequences in less critical settings, like e-commerce product suggestions or social network recommendations. However, in more vital settings, such as the medical, criminal justice, and cybersecurity domains, the absence of explanations presents considerable risks. In the context of cybersecurity, it becomes important to ensure that security professionals can interpret and trust the automated systems they use to make decisions.

2.1.3. Cybersecurity and AI

The field of study focused on applying AI techniques and methods to extract insights from data is known as data science. Cybersecurity Data Science (CSDS) is a rapidly evolving field emerging from the functional integration of its two parent domains: applying data science methods to enhance cybersecurity [25]. CSDS methodologies stem from fundamental data science activities such as ”analytics problem framing, exploratory data analysis, visualization, diagnostics, data preparation, data engineering, statistical analysis, feature engineering, machine learning, optimization, semantic analytics, and ensuring scientific rigor in data-focused inquiry” [25, p.2].

While artificial intelligence has become a cornerstone technology across various industries, its integration into cybersecurity represents a relatively recent development. This gradual adoption may stem from initial skepticism regarding AI’s capability to protect sensitive data and critical systems adequately [9]. Reviewing the literature published in Scopus (see Figure 2.2) reveals a notable paradigm shift in the cybersecurity domain. In recent years AI’s potential for enhancing digital protection mechanisms has been increasingly recognized and explored by researchers.

In parallel, the ongoing digital transformation, characterized by the integration of digital solutions into business processes, has fundamentally changed the way companies operate. While this shift offers numerous benefits, it also poses significant cybersecurity challenges. As companies embrace digital transformation, they must implement robust security measures to ensure that their systems are secure from potential threats [26].

In response to these challenges, various legislative frameworks have been developed by the European Union (EU) to enhance organizational resilience. The Network and Information Security (NIS) Directive [27] provides a broad regulatory framework aimed at improving cybersecurity across multiple sectors, including energy, transport, health, and digital infrastructure. The NIS Directive mandates that operators of essential services and digital service providers adopt appropriate security measures and report significant incidents to national authorities. This directive is complemented by sector-specific regula-

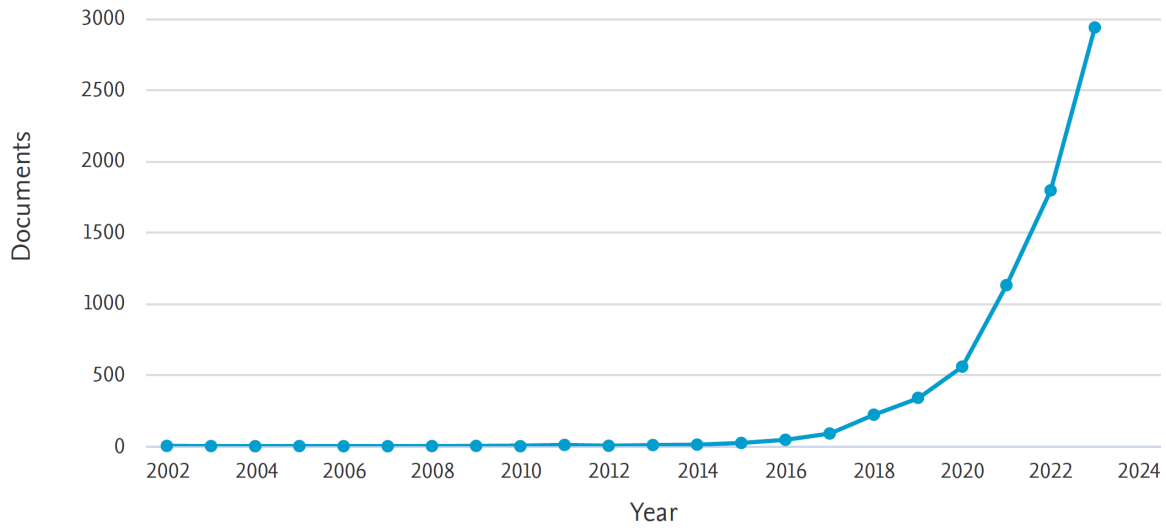


Figure 2.2: Evolution of scholarly publications on machine learning and artificial intelligence in cybersecurity published in Scopus up to and including year 2023.

tions like the Digital Operational Resilience Act (DORA) [28], which focuses on the financial sector. These regulations emphasize the critical need to embed strong cybersecurity measures into the digital transformation journey, ensuring that innovation does not come at the expense of security.

Transitioning from regulatory frameworks to practical cybersecurity measures, it is essential to explore some detection methods employed to safeguard organizations against cyber threats. One traditional method is signature-based detection, also referred to as rule-based detection. It works by comparing incoming data to known patterns or signatures of malicious code to identify threats. This method is effective in spotting known malware and well-defined attack patterns. However, it falls short when it comes to identifying zero-day exploits or new, previously unknown threats [29].

One of the most challenging aspects of cybersecurity is the rapid and continuously evolving nature of cybersecurity breaches, which have increased over time. This necessitates the swift and ongoing development and expansion of cybersecurity methods [30]. According to Ansari et al. [31], this has prompted the wider adoption of AI in this domain, as AI technologies can be designed to effectively identify cyber risks and malicious malware.

This has given rise to machine-learning detection, which utilizes sophisticated algorithms and statistical models to examine large volumes of data and recognize patterns that signal cyber threats. Through training on labeled datasets [3], machine learning algorithms can categorize and prioritize security alerts, detect new and previously unknown threats, and adjust to changing threat landscapes over time [29]. Examples include AI-based intrusion detection models [32], [33], smartphone user authentication methods [34], [35], detection of distributed denial of service attacks [36] and malware detection [37]. An overview of more ML applications in the cybersecurity domain is presented in the works of Li [38], Sarker et al. [33], Chan et al.[9], and Mohamed [39].

Notably, many researchers specifically focus on alleviating the problem of alert fatigue that Security Operation Center (SOC) analysts experience. Baruwal Chhetri et al. [40] propose the A^2C Framework, which advocates for human-AI teaming to alleviate this issue. The framework emphasizes dynamic decision-making modes—automated, augmented, and collaborative—to manage the overwhelming volume of alerts, thereby optimizing SOC operations and reducing cognitive strain on analysts. Further research focuses on using machine learning algorithms to reduce noise and false positives, thereby decreasing the workload on analysts and enhancing the efficiency of SOCs (e.g., [41], [42]). Notably, despite the clear potential of machine learning in SOC operations, its adoption has been limited, and the results have fallen short of expectations [43].

[3] Labeled datasets are collections of data where each example is paired with a specific label or outcome, indicating the correct category or value for that data record.

2.2. Related research

As the integration of AI into cybersecurity continues to evolve, it is crucial to understand both the theoretical and practical implications of this transition. The following section reviews related research to highlight existing findings and identify gaps in the literature.

Sontan and Samuel's [2] paper explores how AI transforms cybersecurity by enhancing threat detection, vulnerability analysis, and incident response. The study reveals AI's effectiveness in automating vulnerability scans, prioritizing threats, and speeding up incident responses, thus reducing human error and strengthening security. Nonetheless, it also points out significant challenges, such as ethical and privacy issues in automated decision-making, potential biases in AI models, and the difficulties in ensuring transparency and accountability in AI-driven systems.

Familoni [3] explores the evolving landscape of cybersecurity influenced by the integration of artificial intelligence (AI), by presenting theoretical foundations and practical implications. The paper highlights AI's role in improving threat detection, authentication, and response while also noting that AI introduces new risks through AI-driven attacks on machine learning algorithms. Familoni emphasizes the importance of human expertise in the ethical use of AI, advocating for robust training and skill development for cybersecurity professionals. He also stresses the need for adherence to ethical standards, awareness of cognitive biases, user-focused design, and building resilience and adaptability within cybersecurity teams.

Gusman [44] conducted a qualitative study on the deployment of AI and ML in cybersecurity, focusing on their impact on intelligent decision-making. Through 10 semi-structured interviews with cybersecurity professionals in the United States, three key themes emerged: the ongoing importance of human decision-makers due to AI limitations, the increasing use of AI-driven decisions in cybersecurity, and the necessary learning curve and training for effective AI integration. Minor themes included a strong interest in learning about AI and ML systems and the significant adjustments needed for their adoption. Despite the rise of AI and ML, IT (Information Technology) professionals expect to remain the primary decision-makers due to current technological limitations. The study's limitations include its small sample size and potential biases inherent in qualitative research.

Al-Dosari, Fetais, and Kucukvar [5] conducted a qualitative study on AI applications and challenges in the cybersecurity of Qatar's banking sector. Through thematic analysis of interviews with nine experts, four main themes emerged: the importance of AI in enhancing cybersecurity, challenges in AI deployment, potential misuse of AI, and vulnerabilities in AI-based tools. The study found that AI is crucial for defending against web-based attacks and fraud but faces obstacles like inefficiencies in in-house development, compatibility issues with legacy systems, and regulatory compliance challenges. Additionally, AI poses risks through adversarial machine learning [4] and AI-powered malware, and inherent vulnerabilities such as data accumulation and privacy risks in chatbots. The study's limitations include its small sample size and potential qualitative research biases.

Gonaygunta's [45] research explores the factors influencing the adoption of ML algorithms for cyber threat detection in the banking industry using the Unified Theory of Acceptance and Use of Technology (UTAUT) model. The study examines how performance expectancy, effort expectancy, social influence, and facilitating conditions affect IT professionals' intentions to use ML in cybersecurity. Key findings show that performance expectancy and facilitating conditions positively influence the intention to adopt ML, while social influence has a negative impact, and effort expectancy has no significant effect. Despite ML's potential to improve cybersecurity through efficient threat detection and response, its adoption is hindered by challenges such as organizational readiness, system compatibility, and regulatory compliance. The study is limited by its focus on a single geographic region and reliance on self-reported data.

Radebe, Tsibolane, and Hart [4] studied the perceptions of cybersecurity experts on AI-enabled tools in large South African enterprises, using the Expectation Confirmation Model (ECM) and semi-structured interviews with 11 professionals, based in South Africa. The findings indicate that experts see substantial benefits in AI tools, such as automation, reduced human intervention, insightful reporting, fewer

[4] *Adversarial machine learning* is a field that focuses on the security vulnerabilities of machine learning models, particularly how they can be deceived or manipulated by carefully crafted inputs.

false positives, decreased risk, and enhanced productivity through quicker data gathering. However, challenges like data privacy concerns, alert fatigue, configuration issues, the risk of AI misuse by cybercriminals, marketing hype, and costs were also identified. Overall, the experts expressed high satisfaction with AI tools and a strong intention to continue using them, underlining AI's crucial role in improving cybersecurity in large organizations. The study's limitations include its small sample size and focus on a specific region, indicating the need for broader research across various countries.

Despite the growing body of research highlighting the potential of AI to transform cybersecurity, a significant gap remains in the literature regarding how organizations transition to developing and using in-house developed AI-based cybersecurity solutions. The papers by Sontan and Samuel [2] and Familoni [3] provide comprehensive reviews of the benefits and challenges associated with AI in cybersecurity, but these studies are largely theoretical and lack real-world insights into the practical implementation and transition processes. The empirical studies conducted by Gusman [44], Al-Dosari et al. [5], Gonaygunta [45], and Radebe et al. [4] deliver valuable findings based on survey and interview data. These studies reveal how AI is perceived in cybersecurity among professionals in the field and identify key challenges and benefits.

An interesting observation by Al-Dosari and colleagues is that banks in Qatar attempting to build in-house security solutions face numerous challenges, including inefficiencies in development and compatibility issues with legacy systems [5]. These challenges include the absence of a well-defined AI plan, weak technology and data foundations, outdated operational strategies, and significant security risks. Given these challenges, studying how a single organization navigates the development and implementation of in-house AI cybersecurity solutions can provide valuable insights.

2.3. Related theories

To understand the complex dynamics of adopting and integrating AI in cybersecurity, it is essential to base our investigation on robust theoretical frameworks.

2.3.1. IT innovation adoption

Innovation is a complex, multi-step process in which organizations convert ideas into new or enhanced products, services, or processes. This transformation allows them to progress, compete, and distinguish themselves effectively in their market [46]. Following this definition, we can classify the development and adoption of AI in cybersecurity as a classic IT innovation adoption problem.

For the past several decades, the adoption of IT innovations has been an intensively researched area [47]. Many models have been developed to determine whether users will adopt an innovation. These include the Technology Adoption Model (TAM) [48], the Diffusion of Innovation (DOI) theory [49], and the Theory of Planned Behaviour (TPB) [50]. However, these models alone could hardly determine whether innovations will be adopted in organizations where adoption is often forced. In response, Hameed et al. [12] proposed a conceptual model to understand the adoption of IT innovations within organizations. It is based on a combination of several established theories: Diffusion of Innovation (DOI) [49], Theory of Reasoned Action (TRA) [51], Technology Acceptance Model (TAM) [48], Theory of Planned Behaviour (TPB) [50], and the Technology-Organization-Environment (TOE) framework [52]. For context, we present three of them below.

Technology Acceptance Model (TAM)

The first developed technology acceptance theory is the Technology Acceptance Model, proposed by Davis [48]. TAM is based on the Theory of Reasoned Action (TRA) [51], a psychological theory that explains the relationship between attitudes, intentions, and behaviors. The TAM specifically applies TRA to the context of technology use, identifying the factors that lead to a user's acceptance or rejection of a technology. The theory posits that the use of a technology is primarily determined by two factors: *perceived usefulness* and *perceived ease of use*. These factors shape the user's attitude towards using a system, defined as the impact of an individual's positive or negative emotions on performing a specific behavior [53].

In the context of Information and Communication Technology (ICT), the Technology Acceptance Model (TAM) plays a significant role in providing theoretical insights into the behaviors related to the use and

acceptance of ICT [54], [55]. Researchers have applied it to explore the adoption of various technologies, establishing it as a key theory in the field [56].

While being acknowledged as an effective, credible, and highly reliable model [57], [58], the TAM has also been openly criticized for being too simple and leaving out important variables [59], such as subjective norms. Researchers also argue that the TAM does not include any potential barriers that would hinder individuals from adopting a specific technology [60].

Technology-Organization-Environment (TOE) Framework

The Technology-Organization-Environment framework, developed by Tornatzky and Fleischer [52], approaches technology adoption from an organizational standpoint, unlike other technology adoption models that focus on the individual perspective, such as TAM.

The TOE framework suggests that the process of adopting innovations is influenced by three distinct contexts within an organization:

- Technology context: This involves the internal and external technologies relevant to the organization, including existing technologies in use and those available in the market [46].
- Organizational context: This includes characteristics and resources of the organization such as connections among employees, communication practices within the firm, the size of the company, and the availability of surplus resources [46].
- Environmental context: This refers to the setting in which the organization operates, including industry characteristics, market structure, the regulatory environment, and the presence of technology suppliers and competitors [46].

Notably, this framework does not specify particular influencing factors for each context. Hence, the exact factors for a given research inquiry should be established by referencing prior research and theoretical insights, as various types of innovations are influenced by distinct factors affecting their adoption [46].

Diffusion of Innovation (DOI) Theory

The Diffusion of Innovation theory, developed by Rogers [49], describes how, why, and at what rate innovations spread through the social system. The author explores five attributes of innovations that affect the rate at which they are adopted. Rogers emphasizes that perceptions are vital in describing human behavior. The characteristics of innovations as perceived by the recipients, not as categorized by experts or change agents, determine how quickly they are adopted. The rate of adoption depends on the following perceived attributes of innovations:

- Relative advantage: This refers to the degree to which an innovation is perceived as better than the idea, practice, or product it replaces.
- Compatibility: This attribute measures how consistent the innovation is with the values, experiences, and needs of potential adopters.
- Complexity: This is the degree to which an innovation is seen as difficult to comprehend and use.
- Trialability: This is the extent to which an innovation can be tested on a limited basis.
- Observability: This refers to how easily others can see the outcomes of an innovation.

The DOI theory can be applied appropriately to individual and organizational contexts [61].

IT innovation adoption in organizations

Hameed et al.'s conceptual model, presented in Figure 2.3, divides the adoption process into three main stages: initiation, adoption decision, and implementation, where each stage is influenced by various characteristics.

We note that this model must be slightly adapted for the context of this study to account for an organization where innovations can be either *acquired* or *developed*.

Initiation

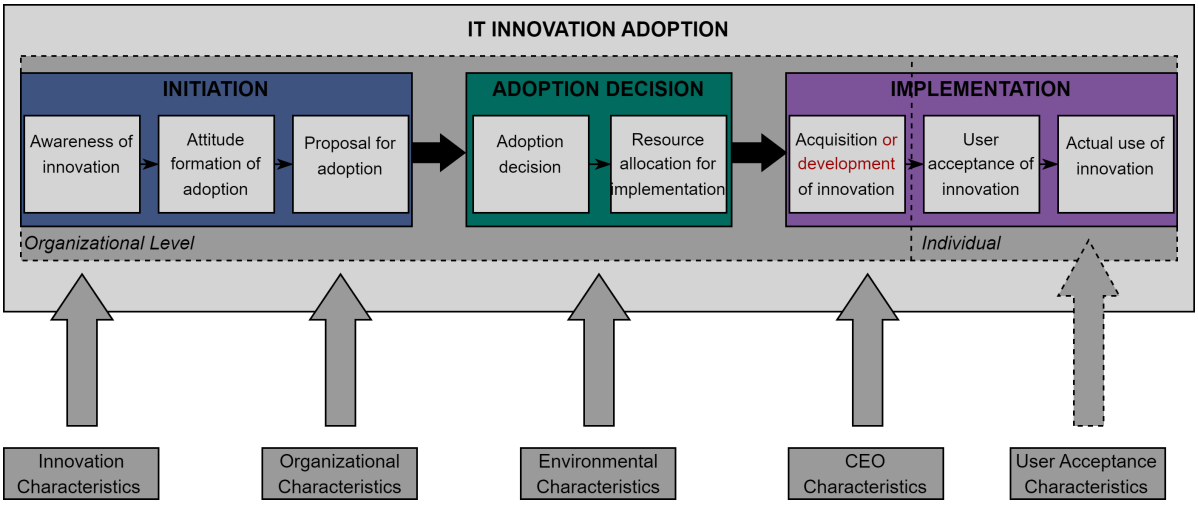


Figure 2.3: Conceptual model for the process of IT innovation adoption, adapted from [12], with modifications highlighted in red.

- Awareness of innovation needs: Recognition within the organization of the need for advanced cybersecurity solutions that can be met through AI.
- Attitude formation of adoption: Initial attitudes towards AI adoption and implementation are formed. These can be influenced by perceived benefits, challenges, and organizational needs.
- Proposal for adoption: A formal proposal of use cases for using AI in cybersecurity is developed.

Adoption decision

- Adoption decision: A decision is made to adopt AI and develop the proposed use cases.
- Resource allocation for implementation: Resources, such as budget, personnel, and time, are allocated to support the development and implementation of AI technologies.

Implementation

- Development of innovation: AI models for cybersecurity are developed, tested, and deployed in-house.
- User acceptance of innovation: Cybersecurity experts are introduced to the AI systems.
- Actual use of innovation: The AI technologies are actively used in the organization's cybersecurity practices.

2.3.2. Sociotechnical systems theory

To study the in-house development of AI tools for cybersecurity and their implementation in the security practices of organizations involves understanding the complex interplay between social and technical components involved in the system. We use the term *system* rather than *process* because it captures the complexity and interconnectivity of developing and implementing AI tools for cybersecurity. Sociotechnical system theory is an appropriate theory to consider in this study. The term sociotechnical captures the interactions between *technical* systems and *social* humans [11]. The theory highlights how important humans in the organization are to solve complex issues instead of relying on technical solutions only [10]. This is how the sociotechnical approach, which focuses on the joint optimization of both subsystems, was introduced [62], [63]. In organizations, individuals work with technological artifacts—tools, devices, and techniques—to convert inputs into outputs, aiming for economic performance and job satisfaction [11], [62]. The social subsystem includes the organization's structure, including authority and reward systems, and involves individuals with their respective knowledge, skills, attitudes, values, and needs [11]. The structure of a sociotechnical system is illustrated in Figure 2.4.

As an open system, the complex environment also influences the sociotechnical system. The interaction between social and technical elements can be non-linear due to unexpected, uncontrolled, and

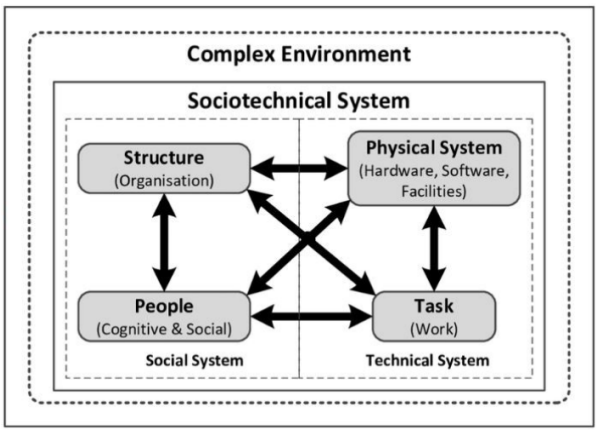


Figure 2.4: Components and interrelationships of sociotechnical systems, sourced from [10].

unpredictable relationships. People have the flexibility and intelligence to reorganize and adapt to environmental challenges and changes [11].

At the heart of sociotechnical theory is the concept that new systems can be optimized and will only function effectively when social and technical elements are combined and considered as interdependent parts of a work system. To guide the design of such systems, including those developing new information technologies, Clegg provides a set of sociotechnical principles [64, p.465]. We display them in Table 2.1 as they provide a holistic view of system development and implementation. They have been used by researchers of human factors, computer and information scientists, engineers, and various other social science disciplines [65].

Table 2.1: Clegg's principles of sociotechnical system design [64, p.465].

Meta-principles	
1.	Design is systemic
2.	Values and mindsets are central to design
3.	Design involves making choices
4.	Design should reflect the needs of the business, its users and their managers
5.	Design is an extended social process
6.	Design is socially shaped
7.	Design is contingent
Content principles	
8.	Core processes should be integrated
9.	Design entails multiple task allocations
10.	System components should be congruent
11.	Systems should be simple in design and make problems visible
12.	Problems should be controlled at source
13.	The means of undertaking tasks should be flexibly specified
Process principles	
14.	Design practice is itself a sociotechnical system
15.	Systems and their design should be owned by managers and users
16.	Evaluation is an essential part of design
17.	Design involves multidisciplinary education
18.	Resources and support are required for design
19.	System design involves political processes

According to Ackerman [13], a central challenge that all sociotechnical systems face is what he describes as the sociotechnical gap. This gap represents "the divide between what we know we must support socially and what we can support technically" [13, p.179]. He argues that exploring and un-

derstanding this gap is a central problem for the field of human-computer interaction that focuses on Computer-Supported Cooperative Work (CSCW) and its technical mechanisms, and emphasizes that neglecting it hinders the development of effective systems [13]. A central aspect of this gap is that it cannot be closed due to the fluidity and nuance of social activities—"user needs are dynamic" [66, p. 2], while technical systems are often "rigid and brittle (...) in their support of the social world" [13, p. 180]. Building on Ackerman's insights, which emphasize the need for a deep understanding of the sociotechnical gap, Ehsan and colleagues develop a framework to describe this gap in the context of (Explainable) Artificial Intelligence. As visualized in Figure 2.5, the technical side comprises Data, Models, and AI-generated Explanations, while the social side consists of Trust in AI, the Actionability of Explanations, and Values [66]. Initially, the infrastructure consists of distinct social and technical components with a broad, undefined sociotechnical gap. As the process of charting the gap begins, each side's building blocks are operationalized, providing detailed insights [66].

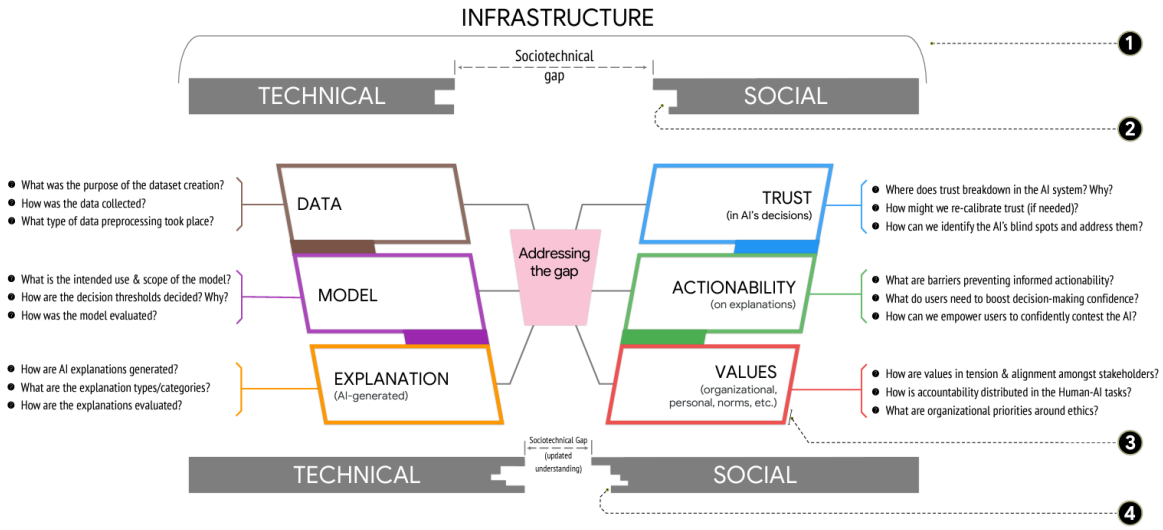


Figure 2.5: A framework for understanding the sociotechnical gap in the context of (Explainable) Artificial Intelligence, sourced from [66].

3

Methodology

For this research on how organizations transition to AI-based cybersecurity solutions through in-house development, a case study approach is particularly suitable. A case study allows for an in-depth exploration of complex phenomena within their real-life context [14]. The challenges and strategies involved in developing AI-based solutions for cybersecurity are multifaceted and influenced by many contextual factors. A case study enables a comprehensive examination of these intricacies that broader survey or experimental methods might miss. Additionally, the case study method is particularly suitable for answering *how* and *why* research questions [14].

3.1. Case study context

This study analyzes the case of a large European bank, specifically its cybersecurity department. To protect the confidentiality and anonymity of the participants and organization involved, the bank's name is not disclosed. Throughout the report, it will be referred to as the Bank. The Bank is a leading financial institution in Europe and offers a broad range of services in both retail and wholesale banking. It operates in over 40 countries and has more than 60,000 employees.

The Bank's cybersecurity department has several hundreds of employees working across various countries. Its operations are divided into three main areas: Security Detection and Response (SDR), Identity and Access Management (IAM), and Attack Surface Management (ASM). For the purposes of this study, the term *threat management* will be used as an umbrella term to collectively refer to these key areas. The department uses both vendor-provided security solutions, some of which incorporate AI-based services, and custom AI-based solutions developed by data scientists in-house.

As any leading financial institution, the Bank faces significant cybersecurity threats and invests substantial resources in advanced AI technologies to enhance its security measures. Despite its resources and commitment, the Bank, like other organizations, faces challenges in fully integrating AI into its cybersecurity practices. These characteristics make it an ideal candidate for exploring the sociotechnical factors that influence this process.

Insights from this case study are particularly relevant to cybersecurity departments within the baking sector. Financial organizations face stricter regulatory requirements due to dealing with sensitive financial data. Security breaches can lead to significant financial losses and reputational damage. These factors necessitate a more cautious AI adoption process to ensure regulatory compliance.

However, the sociotechnical factors this study focuses on are not exclusive to the banking industry. They reflect the dynamics within cybersecurity departments that wish to implement AI systems. Therefore, the study's findings can inform AI adoption strategies in cybersecurity departments across different industries.

3.2. Data collection method

The primary data collection method was semi-structured interviews. Semi-structured interviews rely on a predefined set of questions whose order or phrasing is not set [67]. This interview method was chosen because the research aimed to uncover rich details about the AI implementation process and AI's acceptance in the Bank's cybersecurity department. This includes personal perspectives and details about the organizational dynamics, which might not have been properly captured if e.g., a survey was used. According to Orlikowski and Gash [68], combining individuals' interpretations of the technology and process can generate firm-level knowledge. Semi-structured interviews were preferred over structured or unstructured, as this method allows for exploring specific topics in detail and adapting to the participant's responses. This way, non-anticipated insights can be identified too [69].

The semi-structured interviews were conducted either face-to-face or virtually, based on the participants' location and preference. On average, the interviews lasted 45 minutes. All interviews were recorded with the consent of the participants and transcribed for analysis.

3.2.1. Interview protocols

The interviewees were categorized into three groups: security analysts, data scientists, and leaders. Separate interview guides were developed for each interviewee group and were tailored to their specific roles and perspectives. The interview protocols creation process followed the following steps:

1. **Literature review:** A review of existing literature on technology and innovation adoption provided a foundation for identifying the key factors that contribute to technology adoption (see Chapter 2).
2. **Sociotechnical system analysis:** A sociotechnical system analysis, emphasizing the interactions between technology, individuals, and the organization, provided insights into the multifaceted impacts of AI implementation within the Bank's cybersecurity department. This approach highlighted how technical challenges, organizational dynamics, and human factors collectively can influence the successful integration of AI technology. This analysis is described in Chapter 4.
3. **Development of interview guides:** Based on the information gathered in steps 1 and 2, very comprehensive interview guides were developed for each interviewee group.
4. **Expert review and iteration:** Input from subject matter experts was sought to review the first version of the interview guides and ensure the relevance of the questions. Following this feedback, the interview protocols were adapted, and the number of questions was reduced.
5. **Pilot testing:** The three interview protocols were pilot-tested with three people to refine the clarity of the questions.

Table 3.1 displays a summary of the topics covered by each interview guide. To determine the sequence of the topics, the guidelines given by Annette Lareau [70] were followed. The interviews started with standard questions to make the participants feel at ease, such as asking them to talk about their job responsibilities. More sensitive or speculative questions were asked towards the end of the interview. For more sensitive topics, participants were also asked to reflect on the experiences of their colleagues. This approach is supported by sociological guidelines that emphasize the importance of creating a safe environment for interviewees. Indirect questioning can reduce the perceived personal risk of disclosing sensitive information [71]. This is known as a measure to reduce social desirability bias, one of the most common sources of bias that affect the validity of research findings [72]. Additionally, probing questions were asked, aiming to get more depth from each answer. The comprehensive interview guides are available in Appendix A.

3.2.2. Participants

A total of 15 interviews were conducted, which allowed to reach a sufficient level of saturation—the point where new interviews do not bring new ideas, indicating that additional data collection is no longer needed. Research shows that empirical saturation is usually reached within the range of 9–17 interviews [73].

The interviewees comprised a diverse group, aiming to capture a wider range of perspectives on the

Table 3.1: Interview guide topics per interviewee group.

Interviewee Group	Topics Covered
Security Analysts	• Current use of AI tools for cybersecurity • Integration into daily workflows • User experience • Participation in AI development projects • Trust in AI technologies for cybersecurity • Concerns and barriers for non-users • Perceptions of AI's impact on job roles and decision-making
Data Scientists	• Technical and organizational challenges in developing AI models • Collaboration and communication with other teams • Ensuring user-friendliness and workflow integration • Gathering and incorporating feedback • Factors influencing AI adoption among end users • Future AI trends in cybersecurity • Trust and accountability in AI-assisted decision-making
Leaders	• Benefits and challenges of AI in cybersecurity • Ideation and adoption of new AI innovations • Communication and feedback mechanisms • Factors influencing AI adoption rates • Future impact of AI on job roles and responsibilities • Accountability and liability issues related to AI-assisted decision-making

topic, and included:

- 6 security analysts, directly involved in day-to-day security operations and threat management. Some of them dedicate limited time (e.g., a day per week) to collaborate with data scientists on machine-learning development and implementation projects;
- 6 data scientists responsible for developing and deploying AI solutions in the cybersecurity department;
- 3 leaders overseeing operations in the cybersecurity department.

All participants were selected based on their usage of AI tools, involvement in AI projects and/or cybersecurity expertise. Participants were recruited through purposive sampling, ensuring a mix of roles and experiences relevant to AI integration. The interviewees were contacted via direct invitations or internal referrals.

Notably, two-thirds of the respondents are involved in the Security Detection and Response (SDR) area where AI integration efforts have been ongoing for the longest time within the Bank's cybersecurity department. This focus provided a rich source of data on the long-term challenges and successes of AI adoption.

The study includes the opinions of four security analysts who either use machine learning-based applications for their tasks or have been involved in machine learning projects as domain experts, as well as two analysts who do not use AI tools in their work. The analysts who use machine learning tools provide direct insights into the practical challenges of AI integration, while those who do not use AI shed light on barriers and resistance to adoption. By capturing both user and non-user experiences, we aimed at a thorough analysis of the sociotechnical factors influencing AI adoption.

The leaders selected for the study hold significant influence within the cybersecurity department. Their responsibilities include making critical decisions about AI adoption and implementation, overseeing operational activities, and shaping the strategic direction of cybersecurity initiatives. These leaders offer a mix of strategic and operational perspectives, contributing to a more holistic understanding of the factors influencing AI adoption.

3.3. Data analysis

The interview data were analyzed using the Reflexive Thematic Analysis (RTA) framework developed by Braun and Clarke [15]. The reflexive approach to thematic analysis emphasizes the researcher's active role in creating knowledge [74]. It involves six phases:

Phase 1: Familiarisation with the data

Initial interaction with the data began with transcribing the interviews. This was a lengthy process that allowed to spend considerable time with each interview by listening to the interview recordings and writing the transcripts. Some qualitative researchers argue that the transcription of the interviews is a crucial task that the researcher should undertake themselves because the "analysis begins during transcription" [75, p. 230]. After each transcription, the interview was read through a couple of times and this time was used to note potentially interesting passages and points for analysis.

Phase 2: Generating initial codes

Following the transcription, each interview was coded systematically using the Atlas.ti software. Codes are labels or short descriptions assigned to a segment of qualitative data to capture important aspects relevant to the research questions. When used systematically across the whole sample, they facilitate the organization of the data and the identification of insights during the analysis process.

The coding in thematic analysis can follow an *inductive* (bottom-up), *deductive* (top-down) approach, or a mix of both. This choice depends on the analysis and whether it is primarily guided by the data or by theoretical frameworks [76]. An inductive coding approach was utilized to explore the context-specific factors influencing AI integration at the Bank. Hence, the codebook was developed through engagement with the data rather than before that. Inductive coding offered flexibility and openness to new and unexpected insights that might not have been captured through a deductive approach.

Coding can also be *semantic*, capturing the explicit meanings and using language close to that of the participants, *latent*, focusing on deeper, more implicit, or conceptual meanings, or a combination of both [74]. At first, mostly semantic codes were used, aiming to capture the data as communicated by the respondents. As the process progressed, more latent meanings in participants' words were noticed. As a result, some excerpts were double-coded, using both semantic and latent codes. At the end of this process, the codebook contained roughly 180 codes.

Phase 3: Generating initial themes

Following the coding process, the focus shifts from the interpretation of individual interviews to "interpretations of aggregated meaning and meaningfulness across the dataset" [77, p.1403]. It must be emphasized that themes are not hidden within the data and waiting to be discovered. Instead, the researcher must actively interpret the connections among various codes and explore how these can shape the narrative of a particular theme [76].

First, initial categories were created to organize the codes into groups with similar meanings. Through continuous engagement with the data, these categories were refined. During this process, some codes were merged, while others were disregarded. By the end of this phase, an initial thematic map was developed, encompassing three main themes, each with two, three, and five subthemes, respectively.

Phases 4 and 5: Developing, refining and naming themes

Following the creation of the initial thematic map, the themes were reviewed to ensure coherence and alignment with the overall data set. This involved two levels of review. First, at the coded data level, each theme was examined to determine if the coded data extracts formed a coherent pattern. Secondly, at the data level, the coded excerpts were re-read to confirm that the themes accurately reflect the meanings present within the data. The thematic map was reiterated, one theme was split and additional subthemes were added. This resulted in 4 main themes, with two or three subthemes each. A narrative for each theme was developed, which informed their names.

Phase 6: Writing the report

In the final phase, the focus shifted to producing the report. This involved integrating the themes into a cohesive narrative that addressed the research questions and provided a detailed account of the findings. To illustrate this theme, vivid quotes, and examples were selected from the data. The report, found in Chapter 5, aims to tell the data's story, accounting for its depth and richness.

3.4. Ethical considerations

Ethical approval was obtained from the Human Research Ethics Committee (HREC). All interviewees received detailed information about the study before consenting to participate. A comprehensive data management plan was implemented to ensure the proper handling, storage, and security of the collected data. All interview transcripts were anonymized to protect participants' identities. Due to the specificity of their work, participants' job titles, years of experience, and cybersecurity areas of expertise are not disclosed in the report to prevent their identification among their colleagues. The data was stored on the Bank's secure servers, accessible only to the principal researcher. The data will be retained for two years after the completion of the project, according to ethical guidelines, and will be securely disposed of after this period.

4

Sociotechnical system analysis

As AI technologies become more prevalent in cybersecurity practices, organizations and their workforces must adapt. To understand the potential impact of this change, a sociotechnical system analysis was conducted. This analytical approach allowed us to examine (1) how AI implementation can reshape the social system within cybersecurity and (2) how the social system can affect AI development and implementation. By investigating these interactions between the social and technical systems, we gained a more holistic understanding of the potential determinants and barriers to AI adoption in cybersecurity. These insights were later used to determine the topics for the interviews with employees from the Bank's cybersecurity department.

By examining the adoption and implementation of AI technologies into cybersecurity practices through the sociotechnical system lens, we acknowledge that this transition depends on the interplay between the social and technical subsystems. By definition, the technical subsystem includes all necessary technical elements for operating the system and the tasks performed with it. The social subsystem consists of the organizational structure and the individuals involved. This encompasses their attitudes, competencies, values, and relationships [62].

The social subsystems is made up of the organizational structure and the people within it. In the context of adopting AI in cybersecurity, the people in the social system are the individuals or groups that use, develop, implement, and manage the AI systems. These are the (1) security analysts as end-users, (2) data scientists, who are responsible for the system design, development, and implementation, and (3) leaders, who oversee the systems' integration into daily operations and ensure alignment with strategy. This also includes the cognitive and social aspects, such as how individuals perceive AI tools, their readiness to adopt new technologies, and how they communicate and collaborate. The structure in the social subsystem comprises the organization within which the AI tools are adopted. This includes the hierarchy, roles, organizational support, processes, and procedures within the cybersecurity department.

The technical subsystem comprises all the necessary technical elements required for developing and operating the tools, along with the specific tasks they are designed to perform. In the studied context, these include the developed AI tools and technologies themselves (hardware and software) and the infrastructure necessary to facilitate their operations. The tasks are dependent on the applied cybersecurity area, in this case, threat detection and response, identity and access management, or attack surface management.

Figure 4.1 shows the results of the sociotechnical analysis. Due to a lack of literature sources, which focus on the social implications of implementing AI technologies into cybersecurity, this analysis is largely driven by the general body of knowledge on organizational behaviour and social dynamics.

People $\rightarrow$ Technology

1. Advocacy: Enthusiasm and support from people within the organization, especially those in management, can influence the rate of AI technology adoption. Management resistance is found to

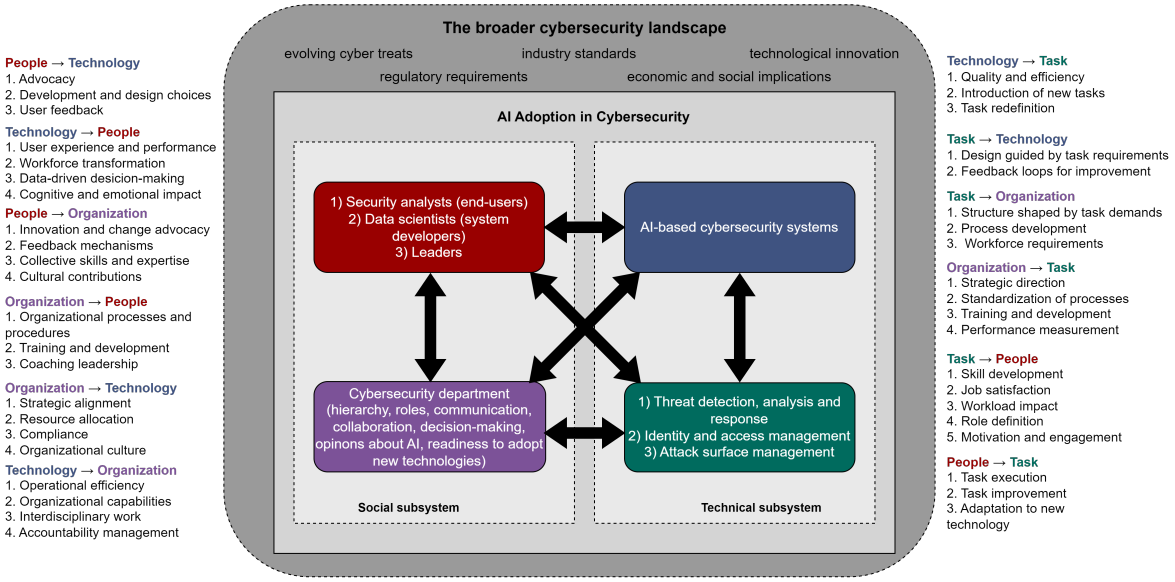


Figure 4.1: Sociotechnical system analysis depicting the adoption of AI in cybersecurity.

be one of the leading barriers to AI implementation in cybersecurity [31].

2. Development and design choices: Data scientists play a crucial role in shaping AI technology. They influence the design, functionality, and overall capabilities of AI tools.
3. User feedback: Security analysts, as end-users and domain experts, offer valuable feedback that drives the iterative improvement of AI systems. Hence, they significantly influence their evolution and refinement.

Technology → People

1. User experience and performance: The design and usability of AI technology can influence user satisfaction, adoption rates, and perceived performance enhancements in cybersecurity tasks.
2. Workforce transformation: The implementation of AI technology can potentially introduce changes in job roles and required skill sets, creating a need for upskilling employees [78]. This can possibly alter the employment landscape within cybersecurity.
3. Data-driven desicion-making: AI tools can enhance human decision-making by providing data-driven insights that surpass the capabilities of manual analysis [79].
4. Cognitive and emotional impact: The introduction of AI technology can reduce the cognitive load for analysts by, for example, reducing the number of false alerts they need to analyze and reducing decision fatigue [2]. However, it can also increase cognitive load in areas such as managing and interpreting AI outputs. It may also affect some emotional aspects like job satisfaction and anxiety regarding job security [80].

People → Organization

1. Innovation and change advocacy: Enthusiastic individuals can act as change agents, pushing for innovative technologies and practices such as the wider adoption of AI in cybersecurity, influencing the organization's strategic direction [49].
2. Feedback mechanisms: User feedback from employees can drive organizational changes in processes, policies, and technology choices, ensuring that these adjustments better align with operational needs.
3. Collective skills and expertise: The collective skills and expertise of the workforce can influence an organization's capacity to integrate and leverage new technologies effectively [81].
4. Cultural contributions: The attitudes, behaviors, and values of individuals contribute to the organizational culture. These can either foster a climate of innovation and learning or resist new

technologies. Thus, this is how people influence the organization's ability to adapt to changes and integrate advanced solutions [82].

Organization → People

1. Organizational processes and procedures: Organizational processes and procedures dictate how people work, including how AI tools should be developed, implemented and used. Whether the organization provides clear guidelines and support can influence the practical transition to AI and users' acceptance. Standardizing, reengineering, and implementing new processes can help integrate AI systems effectively, making AI adoption easier and more successful [81].
2. Training and development: The organization's commitment to training influences how well employees can adapt to new AI technologies and how effectively they can use them [81].
3. Coaching leadership: The organization's coaching leadership practices, emphasizing mentoring and employee development, can significantly impact how employees manage job stress during the transition to AI [83].

Organization → Technology

1. Strategic alignment: Organizational strategy significantly influences the development and implementation of technology, including AI, ensuring that these technological investments align with business objectives [81], [84].
2. Resource allocation: The financial, human, and infrastructural resources that an organization allocates to technology development, deployment, and maintenance shape the capabilities and advancement of those technologies [81].
3. Compliance: Organizational policies, governance structures, and adherence to compliance standards impact how AI technology is designed and used [85], particularly in highly regulated fields such as cybersecurity and finance.
4. Organizational culture: The culture within an organization can foster innovation and risk-taking or, alternatively, resist change. These cultural aspects influence how technology is adopted, integrated, and valued [81].

Technology → Organization

1. Operational efficiency: AI technologies can drastically improve the efficiency of organizational processes by automating tasks and providing decision support, which can reshape business processes [85], [86].
2. Organizational capabilities: The capabilities of AI technologies can enhance the overall capabilities of an organization. By incorporating AI into its cybersecurity practices, the organization can better mitigate and address complex security threats [87].
3. Interdisciplinary work: Successful AI deployment depends on integrating various perspectives, including domain expertise, data insights, and IT [81].
4. Accountability management: Organizations should implement robust governance frameworks and accountability mechanisms to ensure the ethical and responsible use of AI technologies in cybersecurity operations [3].

Technology → Task

1. Quality and efficiency: AI technologies can improve the quality and efficiency of cybersecurity tasks, such as more accurate threat detection and reduction of false alerts [29].
2. Introduction of new tasks: As AI technology evolves, new tasks emerge, such as data curation for AI training, feature engineering, and interpretation of AI-generated insights, which require new skills and workflows [81].
3. Task redefinition: Existing cybersecurity tasks may be redefined as AI technology takes on more of the workload, leading to a shift in the nature and scope of human-involved tasks [88].

Task → Technology

1. Design guided by task requirements: The design of AI technologies is guided by the specific requirements of cybersecurity tasks. This ensures that the systems effectively meet the needs of the tasks they are intended to facilitate.
2. Feedback loops for improvement: The evaluation of AI models on certain tasks creates feedback, which is used to refine and improve the technology [89]. For example, how well AI identifies false positives in threat detection will inform its iterative development.

Task → Organization

1. Structure shaped by task demands: The nature and demands of cybersecurity tasks can shape organizational structure, necessitating certain hierarchies, teams, or communication channels to deal with those tasks effectively.
2. Process development: As tasks evolve with technological changes, such as the introduction of AI, organizations may need to develop or update processes to maintain or increase productivity and performance.
3. Workforce requirements: The complexity and requirements of tasks inform the workforce planning in an organization. This includes the hiring practices, training programs, and team compositions.

Organization → Task

1. Strategic direction: Organizational priorities and strategy dictate which tasks are most important, leading to resource allocation that can enhance those tasks.
2. Standardization of processes: Organizations develop standard operating procedures that can shape the way tasks are performed, including the use of AI in those tasks.
3. Training and development: The organization's investment in training impacts how tasks are performed by equipping employees with the necessary skills and knowledge.
4. Performance measurement: The way an organization measures and rewards task performance can influence how those tasks are executed. By establishing specific KPIs, management can either encourage or discourage the use of AI tools.

Task → People

1. Skill development: The complexity and requirements of cybersecurity tasks demand specific skills. This need drives individuals to acquire new knowledge and competencies. The introduction and integration of AI tools further emphasize the importance of these skills [81].
2. Job satisfaction: The nature of tasks influences job satisfaction. Engaging, meaningful tasks that leverage one's skills can increase satisfaction, whereas monotonous, overly challenging, or mismatched tasks to skill sets can decrease it [90].
3. Workload impact: The number and intensity of tasks impact people's workload. An excessive workload can lead to stress and burnout [91], while a low demanding job can lead to dissatisfaction [92].
4. Role definition: Tasks define roles within an organization, determining what each person is responsible for. As AI changes the nature of certain tasks, it can also redefine roles, requiring adaptation from the workforce.
5. Motivation and engagement: The degree to which tasks are seen as valuable, interesting, or contributing to one's professional growth can significantly influence individual motivation and engagement levels [93].

People → Task

1. Task execution: The way people perform tasks (i.e., their job performance) is shaped by their organizational and technical skills [94].
2. Task improvement: Feedback from individuals performing tasks is essential for identifying improvements. This input can lead to process optimization, better tool usage, and the refinement of AI models in the systems.

3. Adaptation to new technology: People's ability and willingness to adapt to new technologies, such as AI tools in cybersecurity, influence how tasks are performed and how effectively technology is utilized.

5

Empirical results

This chapter presents the key findings from the interviews, organized into four main themes: (1) Mixed perceptions of AI-based solutions in cybersecurity, (2) Organizational influence on AI adoption, (3) Interdisciplinary development dynamics, and (4) Building trust in AI solutions. Each theme is further explored through its respective subthemes. Figure 5.1 displays the thematic map, illustrating the key themes and subthemes identified. The ellipses represent the main themes and the rectangles denote the subthemes associated with each main theme. The dashed lines represent the links between themes, which will be further analyzed in the following chapter. Table 5.1 presents the participants and their identifiers.

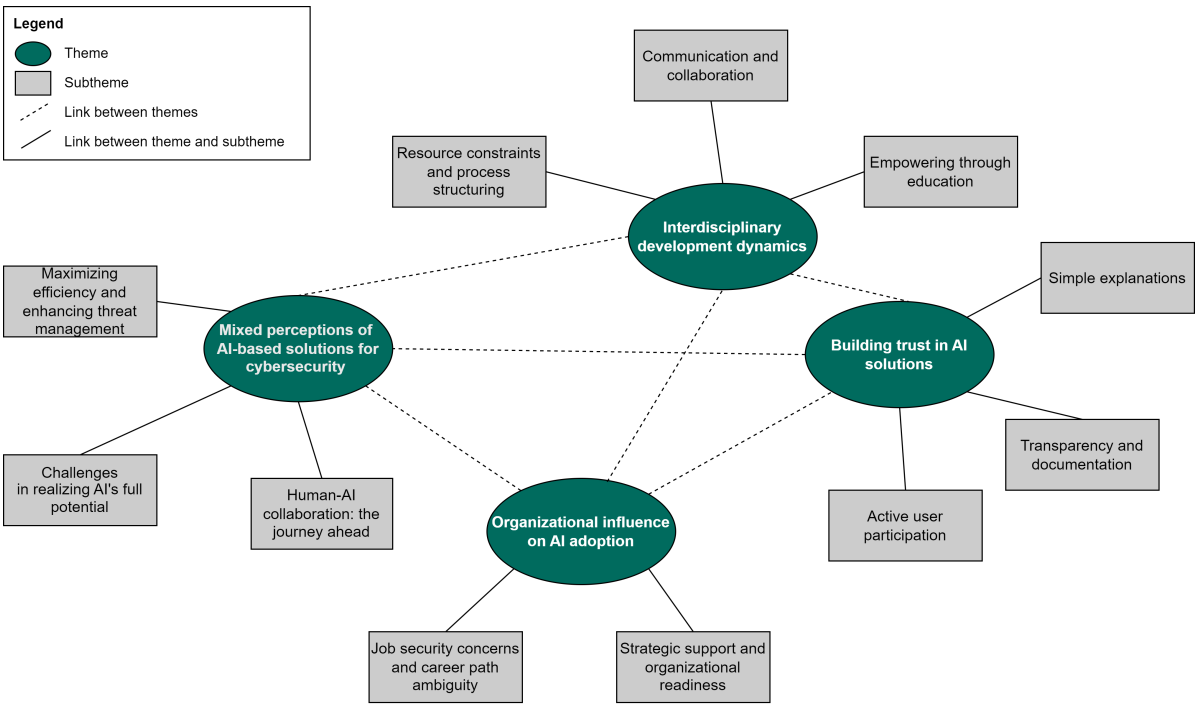


Figure 5.1: Thematic map, based on interview data and reflexive thematic analysis.

5.1. Theme 1: Mixed perceptions of AI-based solutions of cybersecurity

The first theme captures the diverse opinions of cybersecurity experts regarding the implementation and effectiveness of AI in their field. This theme highlights AI's both promising and challenging impact

Table 5.1: Interview participants and their roles.

Participant ID	Role
D1	Data Scientist
D2	Data Scientist
D3	Data Scientist
D4	Data Scientist
D5	Data Scientist
D6	Data Scientist
A1	Security Analyst
A2	Security Analyst
A3	Security Analyst
A4	Security Analyst
A5	Security Analyst
A6	Security Analyst
L1	Leader
L2	Leader
L3	Leader

on modern cybersecurity practices. It reflects the optimism around AI's potential to revolutionize threat management and efficiency, as well as the skepticism and challenges in fully leveraging these technologies. The theme is divided into three subthemes, namely "Maximizing efficiency and enhancing threat management," which highlights the perceived benefits and improvements AI brings to cybersecurity; "Challenges in realizing AI's full potential," which discusses the practical difficulties faced; and "Human-AI collaboration: the journey ahead," which explores the evolving relationship between human professionals and AI tools in cybersecurity contexts.

5.1.1. Subtheme 1a: Maximizing efficiency and enhancing threat management

The use of AI in cybersecurity has the potential to revolutionize the field, bringing considerable efficiency gains and improving threat management. For many security analysts, the daily responsibilities involve repetitive and redundant tasks that can be time-consuming and monotonous. Implementing AI-based solutions can alleviate this burden, automating routine tasks and allowing analysts to redirect their focus toward more critical issues. One analyst shared, "We have so many events that the model could help with this manual work, and we will have more time to do another task. And these are events that you just scroll down." (A2) This highlights how AI can take over mundane tasks, giving analysts more time for higher-level analysis. Another analyst reflected on their early career experiences:

> "When I started [my job], it was really routinal, I was scrolling through the data and (…) I think, we were all overwhelmed by this routine. (…) We are really grateful that the company decided to focus on us and eliminate the manual work." (A3)

This opinion highlights the significant relief AI can provide by removing the repetitive aspects of analysts' jobs.

This shift not only reduces the workload but also helps to combat decision fatigue, a common issue in the cybersecurity environment. An analyst explained, "It [the model] actually reduces sort of your workload. It helps to make it easier and to focus just on the things that require more attention." (A2) Decision fatigue occurs when individuals are required to make too many decisions in a short period, often leading to errors and decreased productivity. By filtering and prioritizing tasks, AI helps mitigate this problem, enhancing overall efficiency.

Beyond automating tasks, AI's most frequently mentioned contribution to cybersecurity is its ability to improve threat detection and management. ML models are used to filter out false alerts and to prioritize potential threats, ensuring that analysts' attention remains focused on the most important events. One data scientist noted, "There are things that they [analysts] cannot do on their own. For example, they have huge amounts of alerts from something and they need to prioritize them, or they need to filter out." (D2) This points to AI's capability to handle large volumes of data and streamline the threat management

process.

AI's advanced computational capabilities enable it to process vast amounts of data quickly and efficiently, a critical requirement in the cybersecurity domain. An AI system's ability to continuously learn from new data and adapt to evolving threats further enhances its effectiveness. This is a major differentiator between using AI systems and traditional rule-based approaches, which prevail in current solutions.

A data scientist highlighted AI's anomaly detection ability by saying: "ML was found to be very useful in finding anomalies."(D2) This is because AI can identify patterns and deviations within large datasets, which are essential for spotting potential threats that might be missed by human analysts. One analyst emphasized the importance of AI in handling numerous events: "And of course, in the case of so many events, there's a possibility that you won't recognize something (...) so there is a chance that the model would choose the events that you missed." (A2) This underscores the role of AI in ensuring comprehensive threat coverage. The need for speed and comprehensive analysis was further highlighted by the analyst: "We really need that because of so many events, we really need the help of AI to find everything, and the second thing is the quicker time." (A2) This quote highlights two key benefits of AI: its ability to handle a high volume of events and its speed in processing and identifying threats.

AI's self-learning capabilities ensure that it constantly evolves, adapting to new threats and improving its performance over time. This dynamic nature of AI is a significant advantage in the ever-changing landscape of cybersecurity. Traditional rule-based systems lack this adaptability, making AI a superior choice for modern threat detection.

5.1.2. Subtheme 1b: Challenges in realizing AI's full potential

Implementing AI-based solutions in cybersecurity at the Bank is constrained by practical challenges. Despite the promising potential of AI outlined in the previous subsection, various factors hinder its effective application, ranging from difficulties in finding suitable use cases to significant data-related issues.

One of the initial hurdles in integrating AI into the Bank's cybersecurity operations is identifying appropriate use cases. The specificity and complexity of the cybersecurity environment often render many AI models unsuitable. As one data scientist noted, "We have unsupervised models out there, a lot of them, but they were not designed for a cybersecurity kind of environment and so you still find that AI suffers in some ways." (D1)

The effectiveness of AI in the bank's cybersecurity operations is heavily dependent on data availability. AI models, especially those relying on supervised learning, require labeled data to learn effectively. Unfortunately, labeled data in cybersecurity is scarce. "One of the biggest challenges for ML in security is the lack of labeled data," (D2) explained a data scientist. Furthermore, cybersecurity datasets often have an imbalanced nature, with very few instances of true attacks compared to normal activities. This imbalance complicates the training process, leading to models that are prone to generating false positives, which can overwhelm security teams with alerts.

Another critical challenge is the quality of data, which directly affects AI model performance. Issues such as missing data, schema imperfections, and bugs can significantly impair model performance. As one data scientist observed:

> "When people create datasets or the data structure in the back-end of the system, for example, they don't really think about the analytics side. So, they just create the field, and they don't care about the quality, just capturing data." (D6)

This quote highlights the lack of strong data governance practices necessary for managing data quality.

The absence of a centralized platform for developing, testing, and deploying AI models adds another layer of difficulty. A data scientist emphasized, "With rule-based models, I believe that you can do it whatever you want because we have those detection systems where you can write something simple, and for machine learning models you need to invest in the platform." (D3) Obtaining data within the bank is another significant challenge due to data being spread across various departments. "Every system that you need to get data from is different or [has] different data owners and is a gigantic challenge

to be able to get, not even to talk about good quality data," (D4) pointed out a data scientist. This fragmentation also complicates data integration and consistency.

Training and tuning AI models for cybersecurity applications at the Bank are time-consuming and resource-intensive tasks. The complexity of these processes often results in models being discarded due to their inefficiency in real-world scenarios. "In some cases, you know, you work on some models, for a long period of time, and then you put in the effort to really making this model do some good things. And this is just trashed to the side, you know because it's generating too many false positives." (D1)

5.1.3. Subtheme 1c: Human-AI collaboration: the journey ahead

AI is viewed by all participants as a complementary tool that enhances the capabilities of human analysts rather than replacing them. By automating routine tasks and analyzing vast amounts of data, AI allows analysts to focus on more complex issues that require human intuition and expertise. Additionally, AI can provide an additional source of information and increase the quality of information available to analysts, enabling them to make more informed decisions. This collaborative approach leverages the strengths of both AI and human analysts, creating a more robust cybersecurity defense. This point was illustrated nicely by an analyst:

> "I think making some models can bring value and more insights to the people that are using them. We have a lot of data, and we show it in reports, like some simple aggregations, but we can bring more [value] from this data (…) for teams that can then do something better, faster, in the time that it should happen and not after damage has been done." (A4)

The adoption of AI by malicious actors (or red teams) underscores the necessity for defensive teams (or blue teams) to also integrate AI into their operations. This ensures that defenders can keep pace with attackers who are continually evolving their tactics using advanced technologies. As one leader pointed out, "Simply to be able to keep up with the attackers, we need to start onboarding and using as many new technologies as we can, so that will definitely change the way we operate, how we build the detections, but also how we do the security analysis." (L1)

However, identifying effective use cases for AI in enterprise-level cybersecurity is not an easy task. While academic research often provides theoretical solutions, these do not always translate seamlessly into practical applications. Models that generate too many false positives, fail to provide new alerts, or do not operate in real time are often rejected by security teams. One leader noted, "Almost every paper that we found on the subject that we were looking for was useless. Because when they apply it to a large organization, it's simply much harder to tune it and in many cases, it turns out that you gotta figure out something on your own from scratch." (L2)

The difficulty in developing effective AI models directly impacts trust and adoption among cybersecurity professionals. Analysts may be hesitant to embrace AI-based solutions, particularly if previous models have not met expectations. Building trust requires demonstrating the reliability and utility of AI in real-world scenarios. One data scientist, reflecting on the challenges faced with an internally developed model, illustrated this point:

> "The model simply wasn't necessary and wasn't well-performing. And I think they lost a lot of confidence in us. (…) So I think it's really important to be very cautious now because if we again send something, it needs to be really good so we can have their confidence back. They can know it's not useless, it's worth their time." (D2)

Many participants indicated that a significant barrier to trust establishment can be the lack of understanding of how machine learning works. Analysts may feel more confident relying on their own judgement rather than on an AI system they do not fully understand. This was nicely described by one of the security analysts:

> "I think it's confusing me. I don't know what to analyze (…) so I'm not really confident to work with them [models] because I don't know how it works exactly. (…) Maybe I feel more confident with my own analytics to verify all data and I know cases from my past that we are looking for something and when I saw all logs in a table on a screen, I saw here is the problem. (…) I'm pretty sure that AI is not going to find it." (A4)

To foster better human-AI collaboration, continuous efforts are needed to build trust and improve AI integration. This includes providing ongoing training opportunities for analysts to understand AI systems and active collaboration between AI developers and cybersecurity professionals. These points will be addressed in Subsection 5.3.3. The journey towards effective human-AI collaboration is closely linked to the broader theme of building trust in AI solutions, which will be addressed in Section 5.4.

5.2. Theme 2: Organizational influence on AI adoption

As described in Theme 1, the adoption of AI in cybersecurity raises mixed feelings. While there is a potential for significant advancements in the field, practical challenges and trust issues arise. AI's technological capabilities play a critical role, but adoption success depends also on organizational factors. The second theme, titled "Organizational influence on AI adoption", investigates how the organizational culture and leadership support at the Bank impact the adoption process. It also highlights the concerns of security analysts regarding their job security and career development.

5.2.1. Subtheme 2a: Job security concerns and career path ambiguity

This subtheme addresses the anxieties that cybersecurity professionals experience about their job security and future career paths in an AI-enhanced environment. It highlights how these concerns can affect their willingness to embrace AI technologies and hinder its successful integration.

Most participants expressed a general fear about the future of the security analyst profession. They highlighted that within the Bank's cybersecurity department, there is a specific concern about AI potentially replacing analysts' jobs. As mentioned in Subsection 5.1.1, AI benefits cybersecurity by reducing false positive alerts and automating routine tasks. While this automation will not eliminate the need for analysts due to the high-stakes nature of the field, there is a fear that the demand for security analysts may decrease in the future. This fear was illustrated by an analyst saying, "The day is going to come when in analytics, some of us will be replaced with AI. That's it, I would say. Because this is the future, unfortunately." (A4)

This statement illustrates that analysts recognize AI's growing capabilities in the field and its potential to outperform humans in certain analytical tasks. The use of the word "unfortunately" suggests that while AI's integration is seen as a progression, it also raises fears about the potential decrease of human roles in the industry. This fear creates resistance and skepticism as people worry about their professional future. An analyst shared:

> "They [analysts] are scared about losing their job position. (…) This is the main reason why they don't want this technology in cybersecurity. (…) When I speak with people, this is one of their arguments." (A6)

A couple of data scientists and leaders acknowledged that there is a need for continuous conversations with analysts on this topic. A leader explained, "We need to take this into consideration while communicating with the teams, so we need to explain the strategy, explain that AI will not replace them, they need to adopt it and learn how to use it." (L1) While data scientists and leaders keep on passing the message that AI is not meant to replace analysts, there seems to be a lack of conversations about how it will affect their career paths. An analyst described:

> "I think that we [the company] have a lot of work to do in this field because right now everyone knows about AI, about the future of AI, but the organization needs to work on education and prepare every single analyst, what is the next step for them in the organization, what is the plan for them." (A1)

This quote underscores the need to focus on education and career planning to help analysts understand their future roles in an AI-integrated cybersecurity environment.

5.2.2. Subtheme 2b: Strategic support and organizational readiness

This subtheme explores the importance of strategic alignment, leadership support, and the overall readiness of the Bank to integrate AI solutions into cybersecurity. A couple of participants indicated that some of the reluctance to adopt AI in cybersecurity comes from the culture within the department. For cybersecurity practices, the Bank has relied on rule-based solutions for a long time. A data scientist

highlighted, "…another barrier might be legacy because usually, we don't do analytics in cybersecurity." (D6) Hence, the integration of AI is considered very innovative. The highly regulated nature of the department further contributes to the reluctance to embrace the new technology.

As cybersecurity professionals are accustomed to traditional methods, they might not want to see the potential benefits of machine learning and AI for the field. One participant noted the department's reliance on dashboards as a primary solution:

> "The other challenge that's specific for analytics is the way that people think about it because the Bank is a world of dashboards. So, for everyone, the solution is always another dashboard and then that's how people think about things." (D4)

This highlights that the Bank's cybersecurity department often relies on familiar tools and methods and this mindset makes it difficult for people to embrace more innovative approaches. Another participant added that this resistance is stronger among more experienced teams with older employees.

A couple of participants also raised concerns about the missing strategic alignment and its effect on the integration of AI into the Bank's cybersecurity operations. While some leaders express verbal support for AI initiatives, this support does not always translate into actionable strategies. A data scientist described this gap:

> "Even though *[Name Redacted]* is super great and supportive of us, analytics is still not a big part of their strategy. So, there is no real push from the top down. They always say, "Yeah, you have my full support, we'll do whatever you need. Just let me know.", but that's not permeated within their teams and their organization." (D4)

Moreover, participants highlighted the disconnect between the enthusiasm for AI at higher management levels and their understanding of what it takes to implement AI solutions in practice. As another data scientist explained:

> "If you go to top-level management or even investors for example, [you hear] "Ohh yeah, AI is amazing," but they don't know what it takes to make a model work. And that's a real problem because that puts pressure on the data scientists when it shouldn't because, if you don't have high-quality data, you cannot do anything. (…) And people think just get some data, press a button and you get something and it's not the case." (D6)

This quote highlights that the adoption of AI into cybersecurity requires educating and preparing all employees, including top managers, about the requirements of AI. This is necessary for the development of clear, actionable strategies to be integrated into the organization's core operations.

5.3. Theme 3: Interdisciplinary development dynamics

The integration of AI solutions in cybersecurity necessitates a collaborative approach that spans several disciplines. This requires coordinated efforts and expertise across the disciplines of cybersecurity, data science, and software development. Theme 3, titled "Interdisciplinary development dynamics", explores the factors that influence successful AI integration in cybersecurity through interdisciplinary collaboration. Three subthemes will be described:

- 3a. Resource constraints and process structuring
- 3b. Communication and collaboration
- 3c. Empowering through education

5.3.1. Subtheme 3a: Resource constraints and process structuring

This subtheme examines the limitations related to resource constraints and the need for structured processes in AI development. Key challenges include time constraints, lack of structured processes, and the need for clear objectives.

Half of the participants shared that the lack of time is among the biggest challenges that they face when collaborating on AI projects and their integration into the Bank's cybersecurity department. Data scientists mentioned that scheduling meetings with analysts is challenging because of the analysts' busy schedules. A data scientist explained, "Everyone is up to their eyeballs in work and priorities

(...) so getting their availability is the hardest." (D4) Additionally, some analysts work on three shifts, including evenings and nights, which complicates scheduling further. Analysts who are interested in machine learning and want to participate in these projects also reflected on this issue. One analyst highlighted the challenge of balancing regular tasks with AI project work:

> "So far, it was mostly the time issue, in the context of actually getting some time for this project along with regular tasks. (…) There wasn't much time available from our side, even though we really wanted to do it." (A5)

One data scientist (D3) pointed out that collaboration improved significantly after establishing a dedicated sub-team of analysts to focus on machine learning projects. Another data scientist (D4), working in a different area, noted that even with limited availability—such as analysts dedicating one day per week to AI projects—collaboration improved significantly. This data scientist shared, "When you start building that network with people that understand what you do, things kind of snowball and get easier." (D4) These insights suggest that having dedicated personnel or even just some limited dedicated time from analysts can greatly enhance collaborative efforts.

A big challenge highlighted by half of the respondents is the lack of structured processes for developing and deploying machine learning models. One data scientist remarked, "We do not have a standard process, you know, we do this and that. It's kind of all hands on deck situation." (D2) This absence of structured processes causes inefficiencies, particularly when it comes to obtaining the data necessary for model development. "I know there are some regulatory obstacles (…) and I do understand that we are a financial institution. But sometimes, I think even the people responsible for it don't really know how to proceed forward for us to be able to do something" (D2), explained the data scientist.

The deployment of models is another bottleneck, highlighted by participants, due to the lack of clear processes and responsible points of contact. "We have a model, the result is amazing, and so on, but then no one knows what to do with it or going to production is very tricky, at least in CISO" (D6), explained a data scientist. This underscores the difficulties in transitioning AI models from development to practical applications, highlighting the need for clearer processes and support systems. About the lack of structure, a leader emphasized the importance of having clearly defined success criteria during the development phase and explained:

> "I think we should already at this [early] phase have a clear understanding of what we are trying to achieve. (…) Of course, there is always part of experimentation there, but we should also say, OK, at this point we stop, if it will not work, then we will not develop it further." (L1)

Knowing when to discontinue the efforts is crucial, especially given the time constraints faced by analysts. An analyst reflected:

> "There was sometime when they [the data scientists] produced a lot of additional work for us. (…) Not all the models worked and brought us good-quality results. So, you want to create something great, but after years, you finally quit it because the results are bad and there is no possibility to use machine learning in such a way." (A4)

This shows how unclear goals can lead to a waste of time on projects that fail to deliver usable results. Such failures could also contribute to trust issues among analysts, reducing their confidence in AI solutions and leading to skepticism, as discussed in Subsection 5.1.3.

5.3.2. Subtheme 3b: Collaboration and communication

This subtheme explores how data scientists and security analysts collaborate on AI projects. It highlights the current challenges and identifies opportunities to foster a more collaborative environment.

Respondents indicated that teamwork between security analysts and data scientists has strengthened in the last few years. Security analysts bring deep cybersecurity knowledge and firsthand experience with operational challenges. This allows them to identify potential use cases where advanced analytics could be useful. Although use cases can originate from various sources, respondents noted that collaboration improves when use cases come from analysts. A data scientist explained:

> "They [analysts] are eager to answer questions. They are eager to collaborate more and help us where necessary and they are eager to see that their opinion of a use case of machine learning was actually useful." (D1)

Analysts help data scientists understand the context of use cases and the underlying datasets. Their input is also essential for assessing model performance. As explained in Subsection 5.1.2, cybersecurity often uses unsupervised machine learning algorithms due to a scarcity of labeled data. Therefore, analysts' expertise is necessary for validating the model outcomes. Respondents agree that tighter collaboration between analysts and data scientists is essential and leads to better solutions.

Cross-team collaboration is generally well-received among the participants. Data scientists appreciate working with security analysts because it allows them to understand their pain points and develop solutions with real impact. Analysts interested in machine learning value the opportunity to work on these projects and learn how models are implemented practically. An analyst shared their experience:

> "Thanks to the collaboration with the [data science] team, we were able to gain more knowledge about how to put something into production, the whole process from the beginning to the end. This was the most valuable." (A5)

Despite positive experiences, several challenges were mentioned. Participants indicated the complexity of combining highly specialized cybersecurity knowledge with data science. A data scientist described this:

> "It [collaboration] is super challenging because it's such a niche and specific domain, which requires a lot of specialized knowledge to be combined. If you are doing data science for a grocery store, you can most of the time understand what's going on. But if you're doing machine learning applied to logs of web servers, you need all these people who deal with web servers and know exactly what is going on. (…) And these are [people with] in-depth technical skills, very narrow, that need to work well together, and that's very difficult." (D5)

Data scientists and analysts explained that the difficulties often stem from ineffective communication. They find it difficult to find the right language to talk with each other, as both disciplines are technical and use different terminology. This requires additional efforts to motivate decisions about the projects. "The communication and the level from where we are starting and the level from where they are starting—that's a big challenge" (D3), explained a data scientist.

Another collaboration issue is the difficulty of obtaining honest feedback from some analysts. "I think they are sometimes afraid of telling us that they are not using all of the models" (D3), said a data scientist. Feedback on model performance is regularly collected during recurring meetings where data scientists, managers, and security analysts are present. However, data scientists have observed that security analysts tend not to speak up during these meetings. Instead, they provide more honest feedback in informal settings, such as small talk. This suggests that the presence of managers in official meetings might create power dynamics that discourage analysts from expressing their true opinions. Additionally, feedback tends to come from the most experienced analysts, which means it might not reflect the broader team's perspectives, leading to a partial understanding of the models' usability.

5.3.3. Subtheme 3c: Empowering through education

The last subtheme in this section focuses on continuous education and training and how these are necessary components to facilitate the effective adoption of AI technologies in cybersecurity. It highlights that ongoing learning opportunities can empower security analysts, enhance their skills, and create a culture of innovation and adaptability.

The interviews revealed that many security professionals lack knowledge about artificial intelligence (AI) and machine learning (ML). This lack of understanding has significant consequences for the adoption of these technologies in the field of cybersecurity. First, if analysts do not understand how machine learning functions, they are less likely to trust the outputs of these tools, as mentioned in Subsection 5.1.3. Second, this lack of understanding can hinder collaboration, as was described in Subsection 5.3.2. A data scientist illustrated the last point:

> "We know how to prepare data for the model, how to prepare features for the model because

> we know what the feature is and why it is important, and I think for the end users, who don't have knowledge around the data science, it's very difficult to start thinking which features we could use." (D3)

Hence, it is evident that continuous education for cybersecurity professionals is essential for improving teamwork between data scientists and security analysts. To address this, in the last couple of years a training course on the foundations of data science has been given to cybersecurity employees in the Bank by data scientists. Respondents note that this has significantly affected the quality of the teamwork on machine learning projects. "It [collaboration] has improved a lot once people were trained" (L2), explained one of the leaders.

Understanding AI also reduces frustration and builds trust. When analysts understand how AI works, they can better interpret unexpected results. They can identify specific reasons for these results instead of assuming the system is faulty. A data scientist explained:

> "… and they [analysts] know how it works, so they're not surprised if they see a peak in the number of results because they can think, more or less, well this is because of this and that. And they are less angry with us when they see more outcomes and more work for them." (D2)

The training is followed by a hackathon, allowing participants to apply their knowledge in practice. This event has several benefits. First, it encourages creative thinking. As described in Subsection 5.2.2, cybersecurity professionals are accustomed to traditional analytical methods. The hackathon pushes them out of their comfort zones, challenging them to create innovative machine learning use cases. Second, it gives participants the chance to collaborate and gain practical experience. Working on hands-on projects during the hackathon helps them see how machine learning is applied in real-world scenarios, as illustrated by one participant:

> "The hackathon was one of the main points of my journey to get into the topic. (…) I had the chance to work with other people and see how it [machine learning] really works." (A2)

While training and hackathons are valuable experiences for analysts, several data scientists argued that for effective learning, hands-on experience should be integrated into analysts' regular job responsibilities. To achieve this, data scientists try to broaden the general skill set of analysts by involving them in the preprocessing of the data, whenever possible. A data scientist highlighted that by saying:

> "I think in order for a person to learn something, they need to do something (…), but people don't have time for that. So, this needs to be something that they do as a part of their job." (D2)

Lastly, leaders indicated that due to the dynamic nature of the cybersecurity domain, it is necessary to create a flexible workforce that can easily adapt to upcoming changes. A leader explained:

> "Maybe not the easiest, but the only way to deal with that [change] is to teach these people to be as flexible as possible. That's the key. You cannot go and spend one year learning one technology because after one year it can turn out to be not useful anymore." (L1)

5.4. Theme 4: Building trust in AI solutions

In Section 5.1.3, we described that the successful adoption of AI technologies depends not only on technical capabilities but also on the trust and confidence of end-users. Theme 4 explores the strategies that can bridge the gap between AI technology and security analysts' trust and is divided into three subthemes:

- Active user engagement
- Transparency and documentation
- Explanations

5.4.1. Subtheme 4a: Active user participation

The first subtheme explains how active engagement of end-users in the AI development process can contribute to building trust and ensuring adoption.

All data scientists agree that involving analysts in the model development process is essential for fostering trust in AI solutions. Analysts gain a sense of ownership over the product when they participate in the process. Their influence on the development increases their satisfaction with the final product and builds trust, as they become partners in the project. A data scientist illustrated this:

> "The more you involve people in developing a product, the more influence they have on what they see, the happier they are because they feel they are part of the process, they feel they have something to say." (D2)

Furthermore, analysts, who are experts in operational security, can perceive data scientists as non-experts in the domain. This leads to reluctance to trust the solutions produced by them. "By showing that we are doing it together, that helps a lot" (D4), reflected a data scientist. Analysts agree that without firsthand experience in developing and implementing AI models, people are more likely to be skeptical about the model's reliability. An analyst, who was actively engaged in a project explained:

> "I think a lot of people don't trust the model because they didn't have the chance to implement anything (…) and they are just afraid that the model would not work, would not give good results. And maybe at the beginning, I also wasn't so sure." (A2)

Another analyst, who participated in a project, acknowledged that without this direct experience, trusting the model would have been much more difficult:

> "It was definitely easier to trust the model after taking part in the development. And if I will get it, without this experience, ooh, [laughs] it could be hard." (A3)

The experience of these analysts shows the significant role of active engagement in building trust. They explained that by participating and learning how models work, they also understand where machine learning excels and where it fails. This helps to demystify the AI process, which many perceive as "magic" otherwise. Removing the sense of mystery and uncertainty surrounding the technology can shift the perception of AI from an unpredictable black box to a well-understood tool.

Moreover, having an opportunity to experiment with a model during the development stage allows analysts to test it with their own examples of incidents. This hands-on experimentation provides concrete evidence of the model's effectiveness, further reinforcing their trust. An analyst described this process:

> "In the development stage, I can easily test it [the model] with my practice and my examples of incidents that are interesting to me. (…) If I see that it's working, (…) it's proof for me that it's OK." (A1)

5.4.2. Subtheme 4b: Transparency and documentation

The second subtheme addresses that transparency is necessary for building trust in AI solutions. It discusses how maintaining transparent processes and providing access to detailed documentation can improve user acceptance among security analysts.

As highlighted in the previous section, active user engagement can foster trust in AI. However, many analysts either do not want to be involved or simply do not have time to participate due to their busy schedules. The interviews suggest that an alternative approach to gaining their trust is by ensuring transparency and keeping analysts informed throughout the development process. This approach keeps analysts updated, allows them to ask questions, and lets data scientists validate their findings along the way. One data scientist explained the benefits:

> "Every now and then we present, this is the progress, this is what we're doing, this is where we at, this is the problems we're having (…) and then people see, OK, there is actual development work being done and people from our side are being involved." (D4)

As discussed in Subsection 5.3.1, due to the lack of structured processes, this practice is not always upheld. An analyst told a story about the development of a model where analysts were neither actively engaged nor properly informed and how this discouraged them from using it:

> "We only got the information that [the model] was prepared and we didn't get any details about this project. I mean, what is the reason to create it? How can it interact with us? I think the [data science] team has some ideas, but they don't verify with us during the

> development. I think that, in some way, this model is a very far project from us. (…) As analysts are treated only as the side of the project, they don't know about it." (A1)

This extract highlights the negative impact of a lack of transparency and communication. When analysts are not actively engaged or adequately informed about a project's purpose and functionality, they feel disconnected and are less likely to trust and use the final product. Proper documentation can mitigate these issues by providing clear, detailed information about the AI model's goals and development process. Another analyst shared their experience:

> "Yeah, I don't know how the model works and what it's supposed to catch. I'm talking about basics. To receive some documentation like that model should look for that, work like that, analyze that data, and it should catch such situations. Yes, now in a lot of cases, this is missing. (…) So I am not confident to work them because I don't know how it works exactly." (A4)

While data scientists have started giving presentations when new models are implemented and then providing access to analysts to these documents, the interviews suggest that this practice must be improved, because comprehensive and accessible documentation can contribute to maintaining transparency and building confidence in AI.

5.4.3. Subtheme 4c: Simple explanations

The final subtheme elaborates on the third identified approach to fostering trust in AI models–by providing analysts with explanations and allowing them to understand how individual AI decisions are made. Subsection 2.1.2 provided a brief introduction to the field of explainable AI (XAI).

The need for providing explanations or some context was discussed by many participants. On the one hand, this is necessary due to regulatory reasons, as in cybersecurity all decisions made by analysts must be traceable. Therefore, the use of explanations is necessary to facilitate the integration of AI into cybersecurity. On the other hand, explanations can build confidence among analysts. If they understand the reasoning behind AI's outputs, they might be more likely to use this technology. A data scientist explained:

> "For that reason, we do not use black box models. We rely heavily on explainability. We cannot tell someone, well, this is suspicious, but not tell them why." (D2)

Analysts agree that having some information on the model's results and the underlying logic can be very helpful for them. They acknowledge that the AI models are not perfect and there is always a chance of the model providing erroneous results. Explanations can help them to make more informed decisions about whether to trust the output or not. In other words, it provides an additional layer of information. An analyst explained:

> "But of course, it's only a model, right? So, it cannot always give good results. But we want to implement something like that: the first thing from the model will be the decision if we should escalate the event or not, and the second will be the proportion of how sure the model is about its reason. This can help to make a decision by ourselves if we should trust or not." (A2)

Effective explanations need to be straightforward and easy to understand. Some data scientists mentioned that they use SHAP plots to explain model decisions but noted that these can be difficult to interpret. A data scientist pointed out, "Explainability, we also try to cover this part, but I believe that they [analysts] don't fully understand it." (D3)

While education might help analysts learn how to read explainability plots, the work of security analysts is often time-sensitive. Hence, for them, it is important to receive context about a model's decision that is easily digestible. Simple and clear explanations are essential for making AI tools practical and trusted among users in high-pressure cybersecurity settings.

6

Implications and recommendations

The insights derived from the interviews reveal a nuanced landscape in the adoption, development, and implementation of AI-based cybersecurity systems in the Bank. Most participants are excited about AI's potential to transform cybersecurity operations, but several challenges and doubts surfaced too. It is necessary to perform a deeper analysis to understand the complex dynamics surrounding this transition.

6.1. Root cause analysis

To gain a deeper understanding of these challenges, we conducted a cause-and-effect analysis, also known as a root cause analysis (RCA). This is a collective term describing a range of systematic approaches applied to identifying the underlying causes of a problem and determining the actions required to resolve it. RCA aims to investigate the problems and find their underlying causes, ensuring that proposed solutions address the fundamental issues rather than merely treating the symptoms of a problem [95]. In the context of this study, this analysis aims to move a level deeper and use the interview findings to identify the logical relationships between the identified themes and subthemes from Chapter 5 and understand the actual causes of the problems mentioned. It should be noted that these relationships are not empirically proven but rather logical connections or potential associations. To establish whether a true causal relationship exists, further statistical studies and empirical validation are necessary.

Figure 6.1 represents these potential links in a cause-and-effect diagram, depicting several levels:

- Undesirable effects: These are the visible problems that emerge as symptoms of deeper issues within the system.
- Contributing factors: These are the intermediate elements that contribute to the undesirable effects, acting as links between the root causes and the observed problems.
- Potential root causes: These are the fundamental issues at the highest level that initiate the entire sequence of cause-and-effect events.
- Potential solutions: These are the proposed actions or interventions aimed at addressing the root causes and mitigating the contributing factors to resolve the undesirable effects. They will be addressed in more detail in Section 6.2.

It should be noted that root cause analysis and systems analysis are complementary methodologies that provide a comprehensive approach to understanding and improving complex processes [96]. The sociotechnical system analysis, described in Chapter 4, offered a holistic view of the cybersecurity ecosystem within the Bank that focuses on the integration of AI technologies. By covering both the social (people and organization) and technical (AI systems and cybersecurity tasks) subsystems, it provided an understanding of how different components and stakeholders interact. For example, as sociotechnical systems analysis places technical issues within the broader organizational and social context, it was easier to identify how technical challenges, such as data quality, can be influenced by social factors, such as the lack of understanding of AI's requirements.

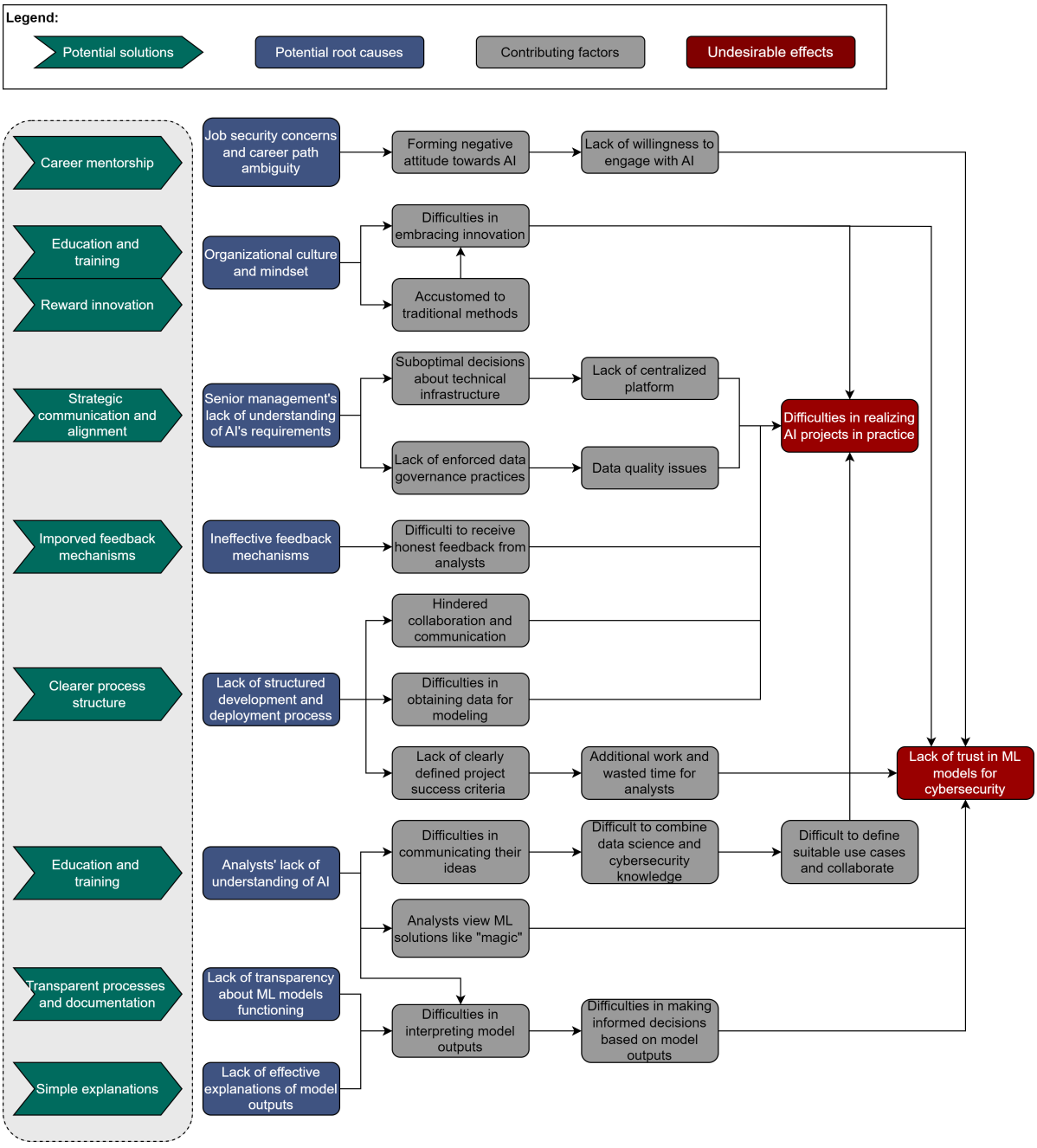


Figure 6.1: Cause-and-effect analysis and potential solutions.

In our analysis, we identified two primary undesirable effects, or symptoms, based on the interview findings: the difficulties in realizing AI projects in practice and the lack of trust in ML models. As illustrated in Figure 6.1, these effects are depicted with arrows pointing toward them and are influenced by various underlying factors. It is also worth noting that these two undesirable effects can be interconnected. For instance, difficulties in realizing AI projects in practice can contribute to the lack of trust in ML models and vice versa, creating a feedback loop that further hinders successful AI development and implementation. This potential bidirectional relationship highlights the complexity of the challenges faced and underscores the importance of addressing both issues simultaneously to achieve meaningful improvements.

Our root cause analysis identified eight potential root causes of these problems. While the term root cause analysis seems to point to a single cause, it is rare, or even impossible, to identify only one reason for a recurring problem [96].

6.1.1. Job security concerns and career path ambiguity

Many employees fear that AI technologies might replace their roles or render their skills obsolete. Interviews revealed that security analysts expressed anxiety about the future of their roles in an AI-driven environment. This aligns with survey results that underscore high anxiety levels about automation [80]. In the context of this study, this concern leads to negative attitudes toward AI, resulting in a lack of willingness to engage with these technologies. This phenomenon can be understood through the lens of cognitive dissonance theory. Cognitive dissonance occurs when individuals experience a conflict between their beliefs and actions [97]. In this case, if analysts believe that AI will take their jobs, it creates a psychological discomfort that makes it difficult for them to trust or cooperate with AI technologies.

6.1.2. Organizational culture and mindset

A traditional organizational culture that resists change can hinder the adoption of innovative technologies like AI. Many professionals are accustomed to established methods and processes and are skeptical about new approaches. This resistance to change makes it difficult for the organization to embrace AI solutions, leading to slow progress in AI project implementation. Participants indicated that even if models are technically successfully developed and implemented, people struggle to understand how exactly to use the insights generated by them, as they are used to their old ways of working. This aligns with Rogers' Diffusion of Innovations (DOI) theory [49] that provides insights into how new ideas and technologies spread within an organization (or a social system). He discusses the effect of organizational culture on the adoption of innovations.

6.1.3. Senior management's lack of understanding of AI's requirements

Participants indicated that senior managers do not have a deep understanding of how AI technologies work and what they require regarding infrastructure, data, and support. This knowledge gap can result in suboptimal decisions about technical infrastructure and resource allocation, which puts more pressure on data scientists and hinders the success of AI projects. It is worth noting that this study did not include interviews with senior managers, and their perspective on this issue is missing. Without their input, it is challenging to fully understand the reasons driving their decisions regarding resources and support for AI initiatives. Future research should aim to include senior management to gain a more comprehensive understanding of their decision-making processes. However, previous studies have shown that top management support and understanding are crucial for the successful implementation of information systems [98].

6.1.4. Ineffective feedback mechanisms

Ineffective feedback processes can lead to communication gaps between data scientists and end-users. The data scientists expressed their concerns about not being able to receive honest feedback from analysts regarding the models' usefulness to them and their performance. Participants noted that the official feedback channel is a biweekly meeting attended by multiple stakeholders, including managers. We assume that mishandled power dynamics may be further influencing these meetings. Without honest feedback, data scientists might not fully understand the practical needs and challenges faced by analysts, leading to the development of AI models that do not address the most critical issues or meet the actual needs of the end-users.

6.1.5. Lack of structured development and deployment process

The lack of structured processes for the development and deployment of machine learning models emerged as a potential cause for the difficulties in realizing AI projects in practice. First, this deficiency hinders collaboration between cross-functional teams because responsible individuals are not always clearly appointed, leading to confusion and project delays. Additionally, data scientists emphasized the need for more structured processes, particularly in obtaining data for modeling, which directly affects project success. Lastly, the absence of clear project success criteria can result in excessive time spent on models that deliver suboptimal performance. Since analysts are actively involved in reviewing model results—a time-consuming task—this can further foster negative attitudes towards AI and increase mistrust.

6.1.6. Analysts' lack of understanding about AI

Many analysts view machine learning as a 'black box' and do not understand how it functions. This lack of understanding affects both the implementation of AI solutions and the analysts' trust in these technologies. Firstly, it leads to difficulties in interpreting model outputs and making informed decisions, thereby reducing trust in AI models. Additionally, their limited knowledge of ML makes it challenging to communicate ideas clearly and integrate their domain expertise with data science. This hampers the definition of suitable use cases and effective collaboration in model development, leading to issues in realizing AI projects. Lastly, not fully understanding how machine learning systems work can cause analysts to see ML solutions as "magic"—a mysterious process. When analysts cannot grasp how ML models arrive at their conclusions, it becomes difficult for them to trust these outcomes.

6.1.7. Lack of transparency about ML models

Another factor that can lead to mistrust is when analysts are unable to interpret model outputs and make informed decisions due to a lack of transparency about the model itself. As domain experts on security data, analysts shared that they want to understand how data scientists process and interpret the data. When there is a lack of transparency, analysts are left in the dark, which makes it challenging for them to trust the results.

6.1.8. Lack of effective explanations of model outputs

The last identified factor that can lead to mistrust is when model outputs are not explained clearly. Clear explanations provide essential information about the model's predictions and assist in improving decision outcomes [99]. AI-assisted decision-making leverages the strengths of both humans and AI to make better decisions. Success in this collaboration depends on calibrating human trust in AI for each situation. Knowing when to trust or question the AI allows experts to apply their knowledge effectively, enhancing decision-making, especially in scenarios where the AI may not perform optimally [100]. The interviews identified that effective explanations are essential to foster confidence and informed decision-making.

6.2. Practical recommendations

6.2.1. Career mentorship

Our findings of the anxiety related to career uncertainties, when AI is introduced into the workforce, are consistent with existing research [80]. To address these concerns, we recommend implementing career mentorship interventions.

In the interviews, leaders from the Bank shared that they are not yet highly concerned about this issue. This is because developing AI models is relatively new in the department, and several models run in production simultaneously. While it may seem premature to discuss future careers with analysts due to the limited integration of AI in cybersecurity practices, research indicates that initial trustworthiness beliefs, or first impressions, can have long-term impacts [101]. Trustworthiness beliefs play an essential role in AI integration as they influence how willing employees are to adopt and utilize AI technologies. If employees trust the technology and the organization's commitment to supporting their career growth alongside AI, they are more likely to embrace and effectively use AI solutions [102].

To address these concerns, it is recommended to establish a career mentorship program that promotes analysts' collaborative behaviors with AI [102]. The program can focus on:

- **Career pathways:** Clearly define career pathways that incorporate AI-related roles. This clarity can help alleviate anxiety and demonstrate the organization's commitment to supporting career growth in the evolving technological landscape [83].
- **Mentorship opportunities:** Pair analysts with experienced mentors (i.e., security analysts experienced with AI) who can guide them through the integration of AI into their work. Mentors can provide insights into how AI can enhance their roles and help them navigate potential career transitions.
- **Early engagement:** Engage employees early in their careers by offering training and development opportunities related to AI. This approach aligns with recommendations from Kong et al. [102], advising employers to hire early-career employees and actively train them to work with AI, thereby cultivating a long-term talent pool.

6.2.2. Education and training

Our findings suggest that education and training can address two key issues. Firstly, they can promote a culture of innovation by changing the organizational mindset and creating a more welcoming environment for AI technologies. Secondly, they can enhance analysts' understanding of AI, effectively bridging the gap between data science and cybersecurity. Notably, the cybersecurity data science (CSDS) practice is a new discipline that is currently developing from practical experience rather than being a purely theoretical or synthetic program. For its success, cross-training and collaborative teaming are essential components [25]. This suggests that training data scientists on cybersecurity topics can be equally important.

Referring back to the sociotechnical system analysis described in Chapter 4, it is evident that individuals influence the organization by contributing to the collective skills and expertise of the workforce. This collective expertise determines how effectively the organization integrates and leverages new technologies [81]. This underscores the importance of addressing this issue.

- **Invest in AI fundamentals education:** From the empirical analysis, it became apparent how important it is to invest resources in educating cybersecurity employees about AI fundamentals. This increased foundational knowledge has already proven to be successful in fostering the collaboration within the Bank's cybersecurity department. The interviews revealed that analysts' increased knowledge of AI has significantly enhanced teamwork dynamics between analysts and data scientists, by improving the communication between them. It has also improved human-AI collaboration, enabling analysts to better interpret model results and discern when to trust or question a model. While participants acknowledge that there is still much progress to be made, they agree that foundational training is an essential building block of this journey.
- **Organize innovative events:** Events such as the organized hackathons were also recognized as positive measures. They allow participants to experience working in a team on a data science project based on a use case they create. This breaks the routine way of working and encourages their creative problem-solving. Hence, we advise to continue organizing such events, as they stimulate innovative thinking. Besides hackathons, this can also be achieved through workshops.
- **Continuous learning opportunities:** While training, hackathons, and workshops can introduce people to the topic, effective learning occurs through continuous practical application. This aligns with the educational theory of constructivism, an action-oriented approach to learning, which states that learners construct knowledge rather than passively absorb information [103]. Among the interviewed analysts, several expressed their interest in learning about the intersection of machine learning and cybersecurity and building a career in this field. It is important to identify such individuals and provide them with opportunities to expand their knowledge and practice. This can be achieved by limited involvement in AI projects that do not interfere with their regular security tasks. This aligns with the recommendation about career mentorship.
- **Serious games**[1]: In order to bridge the gap between cybersecurity and data science, it can be equally important to provide opportunities for data scientists to understand the work of security

[1] Serious games are games that are designed with a primary purpose beyond pure entertainment. While they are still enjoyable and engaging, their primary focus is on achieving a specific educational or practical goal rather than just providing entertainment [104].

analysts better too. A way to achieve this is through serious games. Gamification involves utilizing game-like mechanics, visual elements, and strategic game principles to engage individuals, encourage participation, enhance learning, and address challenges [105]. In this context, serious games can simulate real-world cybersecurity scenarios, allowing data scientists to experience the challenges and thought processes involved in cybersecurity analysis. This experiential learning method can enhance empathy, improve communication, and foster better collaboration between data scientists and security analysts.

We would like to emphasize the importance of providing continuous learning opportunities. These interventions can contribute to enhanced *self-efficacy* among employees, making them more confident in their abilities to understand and work with AI technologies for cybersecurity. According to Albert Bandura, the psychologist who introduced this concept, a strong sense of self-efficacy boosts both personal achievement and well-being. Individuals who are confident in their abilities view challenging tasks as opportunities to overcome, rather than threats to evade [106]. In organizations, higher self-efficacy among employees is associated with greater job satisfaction and performance [107]. The concept of self-efficacy is considered ”one of the most theoretically, heuristically and practically useful concepts formulated in modern psychology” [108, p.47] and can translate into advantages for companies, managers, and employees [109].

6.2.3. Reward innovation

We propose a second solution to stimulating employees to be more adaptive and embrace innovation—by implementing rewards for innovation. According to the Self-Determination Theory (STD) developed by Deci and Ryan [110], there are two types of motivation. Intrinsic motivation is driven by *internal rewards*. These can be feelings of accomplishment, personal worth, or achievement. Extrinsic motivation is driven by *external rewards*, for example, to earn a reward or avoid punishment. Behavioralist researchers distinguish between three types of extrinsic rewards—financial, recognition, and social rewards [111]. Extrinsic motivation has been found to have an interesting effect on intrinsic motivation, as some extrinsic motivators can be internalized [93].

When it comes to work settings, Malek et al. [112] study new product development performance and empirically demonstrate that financial rewards negatively impact intrinsic task motivation. In contrast, they find that recognition and social rewards positively influence intrinsic motivation. The authors argue that recognition and social rewards are seen as forms of managerial acknowledgment that enhance one’s sense of relatedness. In the Self-Determination Theory, relatedness refers to the need to feel connected and significant to others [110].

Kerr and Rifkin, in their book ”Reward Systems: Does Yours Measure Up?” [113], emphasize that the design of reward systems plays a crucial role in shaping employee behavior and organizational outcomes. They argue that effective reward systems align with organizational goals and motivate employees to perform desired behaviors. The authors highlight that there is no one-size-fits-all approach and if reward systems are implemented wrong these can lead to expensive and unwanted outcomes. They outline three essential steps:

1. Defining performance: Establish clear definitions of what performance means by translating the organization’s values, mission statements, and strategies into specific, actionable goals, including ambitious stretch goals. These goals should then be converted into concrete actions.
2. Tracking and assessing performance: Develop comprehensive metrics to monitor activities and evaluate progress toward achieving the defined goals.
3. Aligning rewards with goals: Create reward systems that meet employees’ needs, support the established metrics, and ensure that the company’s objectives are aligned with the efforts and contributions of its employees.

6.2.4. Strategic communication and alignment

Organizational strategy significantly influences how technology, including AI, is developed and implemented [81]. Strategic communication and alignment among senior managers are essential for the successful development and deployment of AI in cybersecurity. At this hierarchical level, it is crucial to demystify AI and clearly outline the requirements for its full-scale implementation. Given that senior

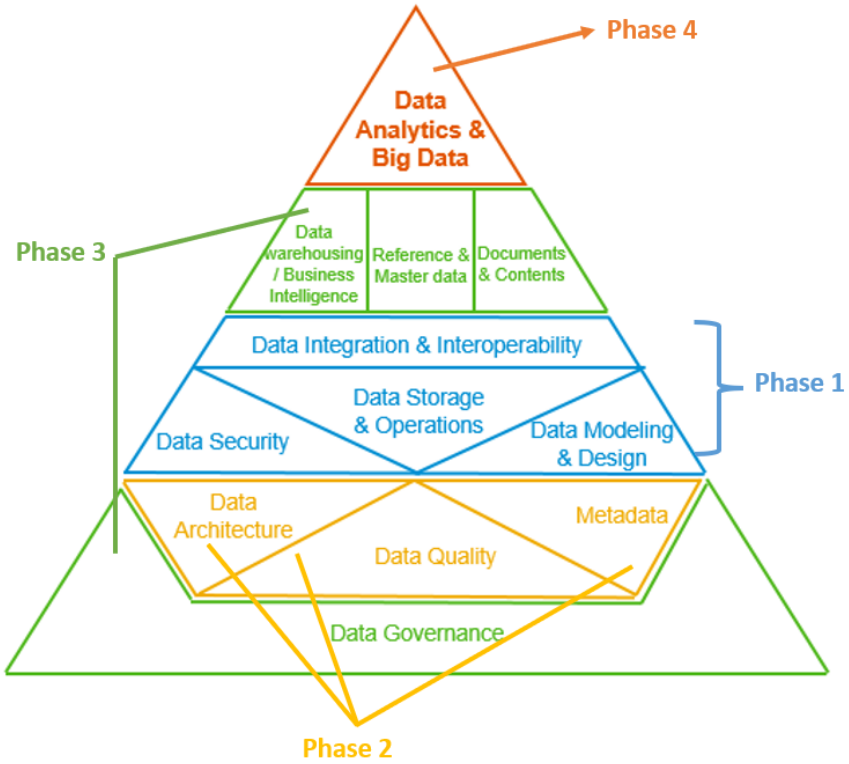


Figure 6.2: Aiken Pyramid of data management [115].

leaders often lack the time to learn about the intricacies of AI, it is imperative to communicate these requirements effectively. This can ensure that senior managers have a comprehensive understanding of what is needed to support and drive AI initiatives, leading to more informed decision-making and successful integration of AI technologies in the Bank's cybersecurity department.

One potential framework that can be used in this regard is the Aiken Pyramid (see Figure 6.2). It draws from the Data Management Body of Knowledge (DMBoK) [114] and aims to help organizations understand and manage their data by organizing the data management pillars. This framework emphasizes the foundational elements for successful AI initiatives, from storage and security to analytics and big data. It breaks down data management into distinct phases, outlining a structured approach. The pyramid emphasizes the importance of data governance and quality, portraying it as the foundation for successful projects.

6.2.5. Improved feedback mechanisms

To address the communication gaps due to ineffective feedback processes identified in our root cause analysis, it is necessary to enhance the feedback mechanisms. Our study highlighted a common challenge in many organizations where hierarchical dynamics can stifle open and honest feedback. This relates to the concept of psychological safety, which is the state of feeling secure enough in a group setting to take interpersonal risks. Essentially, it means that individuals can voice their opinions, seek assistance, or acknowledge errors without fearing criticism or embarrassment. Psychological safety is crucial for increasing team effectiveness and improving performance [116]. In the context of developing and deploying machine learning models for cybersecurity in the Bank, a highly innovative space requiring interdisciplinary efforts, this issue might be further complicated if analysts do not feel competent enough in data science to provide feedback.

To improve the current feedback mechanisms in the Bank, we propose the following options:

- **Feedback culture training:** Conduct workshops focused on building a culture of open feedback. These sessions should showcase instances where feedback has led to significant improvements,

demonstrating the value of feedback and encouraging more open communication.

- **Designated feedback mediator:** Appoint designated individuals who can collect feedback from analysts and present it during meetings. This approach helps ensure that feedback is shared without direct confrontation and facilitate more honest and constructive communication.
- **User feedback modules:** If technology permits it, implement feedback modules within AI tools where users can provide real-time feedback directly within the application.
- **Feedback impact communication:** Communicate back to users how their feedback has been utilized to improve the AI models. This transparency can help build trust and show that their input is valued and acted upon.

Relating this issue to the literature on change management, research has highlighted how important feedback is for fostering change [117]. Hence, working towards an open and transparent collaboration is essential.

6.2.6. Clearer process structure

The lack of a structured development and deployment process has been identified as another barrier to the successful implementation of AI projects in cybersecurity. While the current limited number of ongoing projects and running models does not present a bottleneck, this issue must be addressed proactively. If the Bank intends to implement more AI models in the future, ensuring scalability and efficiency in managing these models will be crucial to avoid potential challenges. To address this, it is essential to establish a clearer process structure that provides a systematic approach to developing and deploying AI models. At this stage, determining the exact structure of the process is challenging. Future work should focus on this aspect.

One potential framework to consider is the methodology outlined in the paper by Foroughi and Luksch [118]. This methodology draws on established data science practices and tailors them to the specific needs of cybersecurity projects. The authors compare several data science methodologies, aiming to identify the most effective approach:

- Knowledge Discovery in Databases (KDD) [119],
- Cross Industry Standard Process for Data Mining (CRISP-DM) [120],
- IBM's Foundational Methodology for Data Science (FMDS) [89],
- Microsoft's Team Data Science Process (TDSP) [121].

Based on the comparison, the authors outline a process with four general steps: (1) problem definition and formulation, (2) data gathering, (3) data analysis, and (4) production (see Figure 6.3). Each general step is broken down into several substeps and this approach is largely based on the Foundational Methodology for Data Science [89]. This blueprint can be used to develop a process tailored to the development and deployment practices at the Bank.

6.2.7. Transparent processes

The interviews indicated that transparency in AI processes is essential for building trust and facilitating informed decision-making. According to Lao et al. [99], providing comprehensive information about AI models can significantly enhance their usability and acceptance. The authors advocate for the use of tools like model cards, fact sheets, and an "About Me" tab to offer detailed insights into various aspects of the models. These tools should include information on model performance, documentation, training data, and other relevant details.

The use of model cards has been advocated as a medium to increase transparency among machine learning developers and users by Mitchell and colleagues, and Figure 6.4 presents an overview of model card sections with recommended prompts for each, as presented in the original paper [122]. The authors emphasize the importance of viewing model cards as one of many transparency tools, such as third-party algorithmic audits (quantitative and qualitative), adversarial testing, and more inclusive user feedback mechanisms.

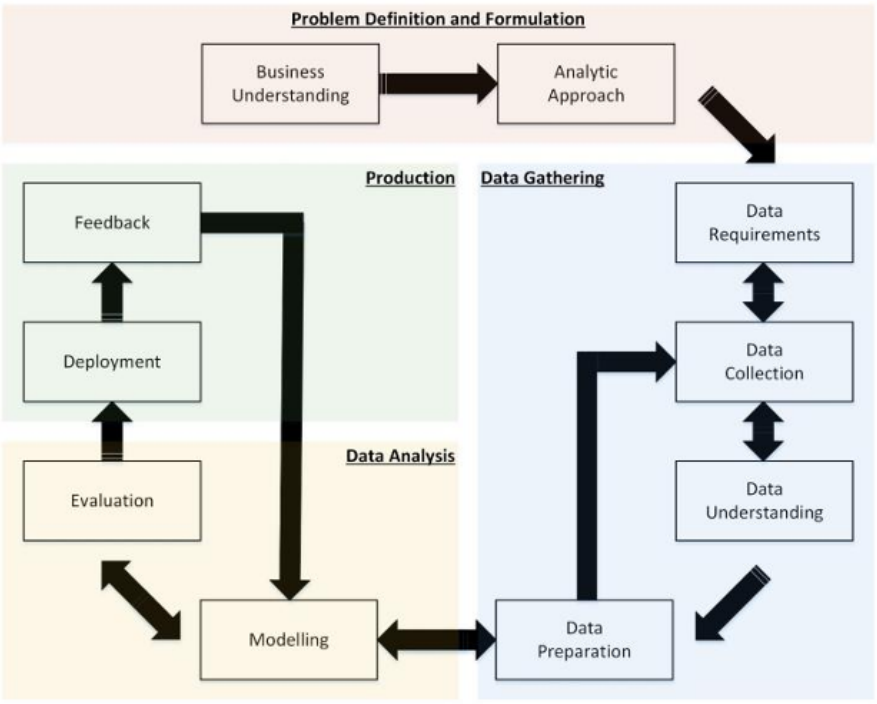


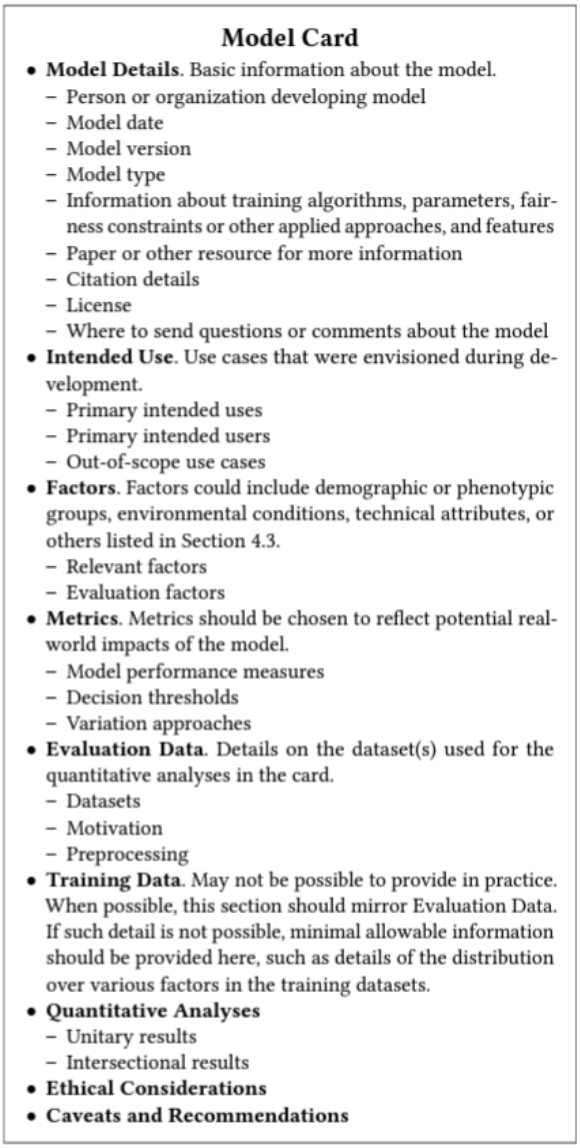
Figure 6.3: The Foundational Methodology for Data Science for Cybersecurity projects, from [118].

6.2.8. Simple explanations

The literature on explainable AI (XAI) is quite limited, particularly when it comes to explanations that are evaluated with real-world users [123]. It remains an open question how to address the actual needs of users for understanding AI [124]. While the technical field of XAI has developed multiple tools to explain AI models (e.g., SHAP, LIME), it is still unclear how to select the best one and translate it into suitable UX designs [125]. The challenge lies in the lack of systematic evaluation of different explanation styles and forms [126].

The community of human-centered explainable AI has emerged because explainability is considered an inherently human property. Therefore, technological choices should be made based on users' explainability needs [99]. For instance, Kim et al. [127] conducted a mixed-methods study with 20 end-users to understand their XAI needs, uses, and perceptions. They found that users prefer practically useful information that resembles human reasoning rather than technical system details. Two human experiments conducted by Zhang et al. [100] reveal that confidence scores can help adjust people's trust in an AI model. However, trust calibration alone is not enough to enhance AI-assisted decision-making. Success also relies on the human's ability to contribute unique knowledge that addresses the AI's errors.

Based on the interview results, it became clear that in the context of cybersecurity, simple explanations are more useful. Participants noted that SHAP plots, for example, can be too complex and time-consuming for analysts to interpret. Instead, a more straightforward approach is needed. Amarasinghe et al. [128] studied the effect of explanations on fraud analysts by using SHAP among other XAI toolkits and provided a simplified interface; see Figure 6.5. Analysts were presented with the top 6 features with the highest absolute importance. The importance values were indicated by colors: green for negative importance (no fraud) and red for positive importance (fraud). A similar interface design could be applied to security analysis, where analysts must decide whether an event is a real threat or not. By providing pairs of feature names and their importance scores in a clear, color-coded manner, we can make the explanations more accessible and actionable for cybersecurity analysts.



Model Card

- **Model Details**. Basic information about the model.
  - Person or organization developing model
  - Model date
  - Model version
  - Model type
  - Information about training algorithms, parameters, fairness constraints or other applied approaches, and features
  - Paper or other resource for more information
  - Citation details
  - License
  - Where to send questions or comments about the model
- **Intended Use**. Use cases that were envisioned during development.
  - Primary intended uses
  - Primary intended users
  - Out-of-scope use cases
- **Factors**. Factors could include demographic or phenotypic groups, environmental conditions, technical attributes, or others listed in Section 4.3.
  - Relevant factors
  - Evaluation factors
- **Metrics**. Metrics should be chosen to reflect potential real-world impacts of the model.
  - Model performance measures
  - Decision thresholds
  - Variation approaches
- **Evaluation Data**. Details on the dataset(s) used for the quantitative analyses in the card.
  - Datasets
  - Motivation
  - Preprocessing
- **Training Data**. May not be possible to provide in practice. When possible, this section should mirror Evaluation Data. If such detail is not possible, minimal allowable information should be provided here, such as details of the distribution over various factors in the training datasets.
- **Quantitative Analyses**
  - Unitary results
  - Intersectional results
- **Ethical Considerations**
- **Caveats and Recommendations**

Figure 6.4: An overview of model card sections with recommended prompts for each, from [122].

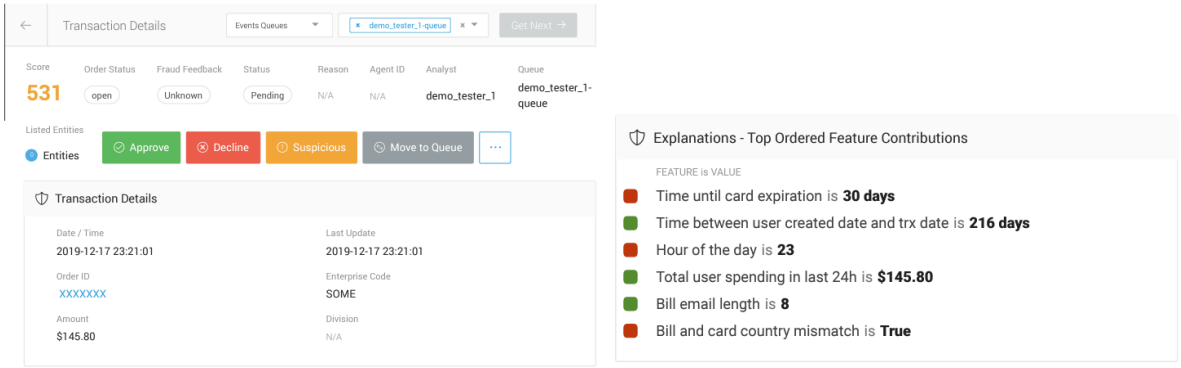


Figure 6.5: Sample interface of an ML-based tool for fraud analysis, including simplified explanations based on feature importance, from [128].

7

Discussion

7.1. Conceptual model

The conceptual model presented in Figure 7.1 encapsulates the interplay between the social and technical factors in the development and use of AI for cybersecurity that were explored in this study. This model illustrates how the social and technical elements interact and what strategies can bridge the gap between them. It is divided into three main components—*Technical Factors*, *Social Factors*, and *Interactions*—supported by *Bridging Mechanisms* that facilitate the development and integration of AI technologies within an organizational context.

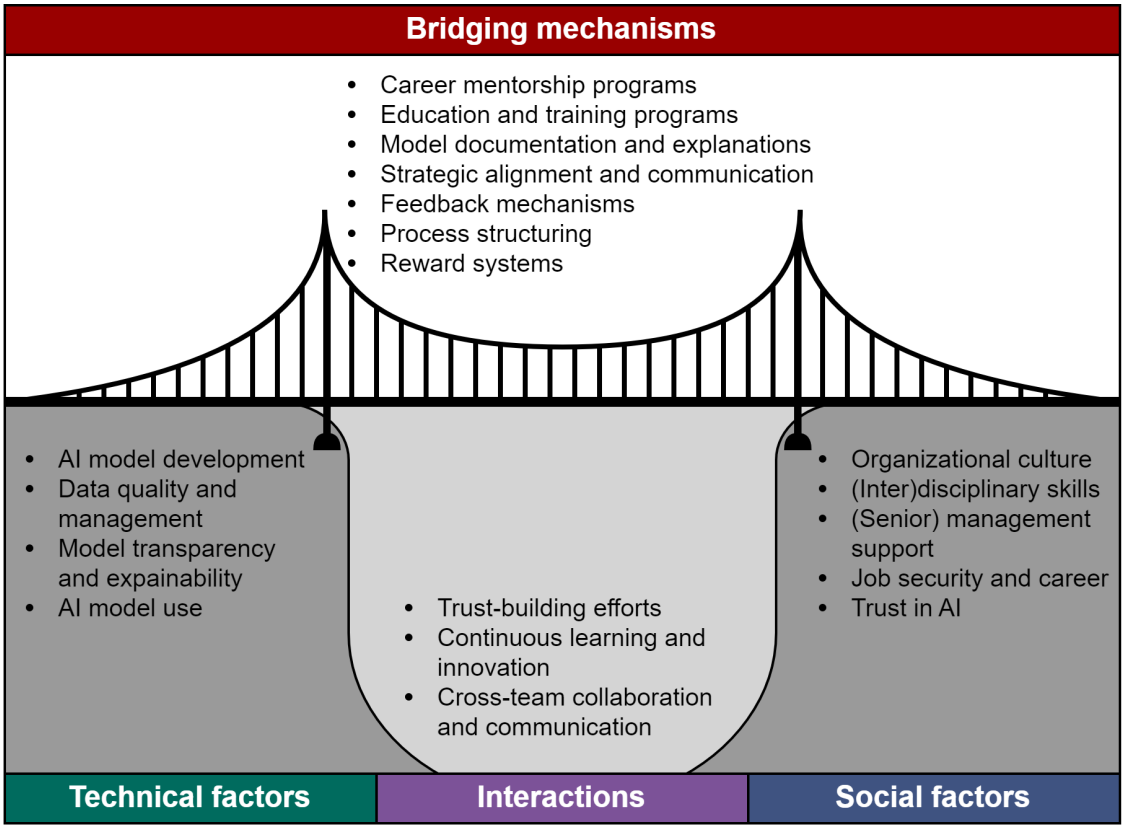


Figure 7.1: Conceptual model for bridging the sociotechnical gap in developing and using AI models for cybersecurity. On both sides, we present the technical (left side) and social (right side) factors. The middle depicts the gap between them. As two shores of a river are connected by the earth beneath the water, we illustrate the interactions between the social and technical systems. Above the gap, the bridge comprises various bridging mechanisms essential for connecting technical and social factors.

Technical factors

AI model development is a critical component that involves creating and refining AI algorithms tailored to detect cybersecurity threats. Developing robust and adaptable AI models is crucial to addressing the dynamic nature of cyber threats. Our empirical findings and analysis show that model development is difficult in practice due to the challenges in identifying suitable use cases and creating models that do not generate too many false alerts (see Subsection 5.1.2). Another essential aspect is **data quality and management**. High-quality data is necessary for training accurate and reliable AI models, and effective data management practices ensure that AI systems receive reliable and relevant data, enhancing their performance and trustworthiness. The difficulties in obtaining high-quality security data were highlighted as a big challenge in the Bank (see Subsection 5.1.2). Moreover, **model transparency and explainability** are vital for fostering trust among users. Understanding how AI models operate and generate results ensures that security analysts can interpret and act on AI-driven insights, which is crucial for practical application in cybersecurity operations. This issue was described in Subsections 5.4.2 and 5.4.3. Finally, the **use of AI models** is a factor that refers to how these models are deployed and operationalized. Effective AI model use involves integrating these models seamlessly into existing systems and workflows, ensuring that the tools are user-friendly and scalable, and implementing continuous monitoring and feedback loops to maintain their effectiveness over time. We discussed the issues related to model deployment and ineffective feedback in Subsections 5.3.1 and 5.3.2, respectively.

Social factors

The adoption of AI in cybersecurity is significantly influenced by **organizational culture**. A culture that fosters innovation and is open to integrating new technologies is necessary for successful AI adoption. Resistance to change within the organization can significantly hinder the implementation of AI solutions. Additionally, management support, particularly **senior management support**, is crucial as it provides the necessary resources and fosters an environment welcoming AI innovations. Leadership commitment has been identified as key to driving strategic initiatives and AI adoption. Both of these social factors were described in Subsection 5.2.2. Moreover, **interdisciplinary skills** are another critical social factor. Creating AI models for security purposes requires a blend of data science and domain-specific knowledge. Developing interdisciplinary skills within teams ensures effective communication and cooperation. We described this point in Subsection 5.3.2. **Job security and career** were also identified as an important social factor due to their impact on organizational resistance (see Subsection 5.2.1). Lastly, **trust in AI** was identified as a fundamental social factor that determines to what extent security professionals feel confident in using AI tools. We found it to be a central issue that needs to be addressed in order to foster the envisioned human-AI collaboration between AI systems and security analysts (see Subsection 5.1.3).

Interaction points

The intersection points in the conceptual model—trust-building efforts, continuous learning and innovation, and cross-team collaboration and communication—are pivotal because they represent the points where the technical and social systems intersect and influence each other. These points are important for bridging the sociotechnical gap in AI adoption for cybersecurity. **Trust-building efforts** involve actions and strategies aimed at creating and maintaining trust between AI systems and their human users. This is an intersection point because trust in AI systems is built through technological aspects like transparency, reliability, and consistent performance, while trust is fundamentally a human experience shaped by perceptions and beliefs. The effectiveness of AI models directly impacts user trust, and the level of trust influences how users interact with AI systems. While trust is essential, the effectiveness of AI models forms the foundation of this trust. Activities aimed at building trust, such as involving end-users and other stakeholders in AI development, providing clear explanations of AI decisions, and demonstrating the reliability of AI models, require a blend of technical transparency and effective communication.

Continuous learning and innovation refer to ongoing efforts to educate employees and foster a culture of innovation through training programs, workshops, and events like hackathons. This is an intersection point because continuous learning involves imparting technical knowledge about AI systems to users. Meanwhile, users utilize this knowledge to generate new ideas for AI use cases in cybersecurity and to provide feedback to data scientists, which is then incorporated to improve AI systems.

Technological advancements necessitate continuous learning to keep users updated, and innovative ideas and practical insights from analysts can lead to technological improvements.

Cross-team collaboration and communication involve interactions between data scientists, cybersecurity analysts, and other stakeholders to develop and implement AI solutions. This is an intersection point because effective AI model development and deployment depend on continuous collaboration between technical teams and operational teams, involving frequent communication and shared problem-solving. The quality of AI solutions is directly influenced by the input and feedback from security analysts, and analysts' ability to use AI tools effectively is enhanced through collaboration. Joint meetings, collaborative development sessions, and feedback loops require both technical and social elements, ensuring that AI solutions are practical and aligned with user needs.

Bridging mechanisms

Career mentorship programs (see Subsection 6.2.1) can play a role in guiding analysts through the integration of AI into their work. By helping them understand how AI enhances their roles, these programs can alleviate anxiety about job displacement and foster a culture of continuous learning and support.

Education and training programs (see Subsection 6.2.2) are essential for bridging the knowledge gap between disciplines. These programs can (1) enhance analysts' understanding of AI and the critical role of data quality, and (2) improve data scientists' understanding of cybersecurity. This dual focus fosters a culture of innovation and collaboration by ensuring that both groups have a mutual understanding of each other's domains.

Model documentation and explanations (see Subsections 6.2.7 and 6.2.8) are key to helping analysts comprehend AI models. By providing clear and accessible explanations, these resources can strengthen analysts' trust in AI and ensure they can effectively incorporate AI outputs into their decision-making processes.

Strategic alignment and communication (see Subsection 6.2.4) are vital for ensuring that AI initiatives are aligned with organizational goals, thereby securing senior management support. Additionally, this alignment ensures that the necessary resources—such as time for employees to collaborate on AI projects and the required infrastructure (e.g., development platforms)—are prioritized to facilitate successful implementation.

Feedback mechanisms (see Subsection 6.2.5) are essential for promoting open communication about AI models and their outputs. By enabling continuous feedback, these mechanisms allow for ongoing improvement and trust-building between analysts and data scientists.

Structured processes (see Subsection 6.2.6) can ensure consistency and efficiency in data management, AI model development, and deployment. These processes are critical for the successful development and integration of AI technologies into organizational practices.

Reward systems (see Subsection 6.2.3) can incentivize innovation in AI technologies, fostering a culture that embraces and supports AI adoption. By recognizing and rewarding efforts in AI development, these systems can encourage continued engagement and innovation.

7.2. Comparison with other studies

Related papers, discussed in the literature review (see Section 2.2), include those by Gusman [44], Al-Dosari and colleagues [5], and Radebe and colleagues [4]. These studies were selected for comparison because they employ a similar qualitative research strategy. Table 7.1 synthesizes relevant information about each study, covering research objectives, methodologies, participant profiles, contextual factors, and key findings.

Our study highlights the potential of AI to enhance threat management and operational efficiency in cybersecurity. Similar findings were observed in the work by Al-Dosari et al. [5] and Radebe et al. [4]. However, it also emphasizes significant challenges such as technical limitations and integration issues, which is consistent with the findings by Radebe et al. [4]. All four studies agree that while there are possibilities for automation and less human intervention in cybersecurity practices, humans

Table 7.1: Comparison of studies on AI in cybersecurity.

Characteristic	This study	Study by Gusman [44]	Study by Al-Dosari et al.[5]	Study by Radebe et al. [4]
Research objectives	To explore the challenges of developing in-house AI systems in cybersecurity and propose ways to mitigate them	To explore how cybersecurity experts make decisions with the help of AI tools	To understand the implications of AI on cybersecurity	To establish how cybersecurity experts perceive the usefulness of AI tools
Methodology	Semi-structured interviews, thematic analysis	Semi-structured interviews, thematic analysis	Semi-structured interviews, thematic analysis	Semi-structured interviews, thematic analysis
Participants	15 cybersecurity experts	10 cybersecurity experts	9 experts	11 cybersecurity experts from large organizations
Contextual factors	A large Bank in Europe	Various industries in the United States	Banking industry in Qatar	Various industries in South Africa
Key findings	- AI can enhance threat management - Technical limitations and integration issues - Human-AI collaboration - Lack of trust - Job security concerns and career path ambiguity - Organizational culture, leadership support, and strategic alignment - Needs resources and well-structured processes - Interdisciplinary collaboration - Need for training and skill development - Need for development with end-users - Need for transparency and clear documentation - Need for model explanations	- Humans will remain the dominant decision-makers; - Decision-making with the help of AI will be more prevalent - Learning curve necessitating training - Cybersecurity employees will need to adjust - Excitement to learn about AI and ML systems	- AI can enhance cybersecurity - AI is more efficient compared to rule-based algorithms - Vulnerabilities of employed models - Accumulation of data - Chatbot privacy - Fictitious data - Redundancy - AI complexity - Potential security risks - Outdated organizational strategies - Lack of compatibility - Need to train employees - Regulatory and compliance requirements - Socioeconomic implications	- Automation - Less human intervention - Insightful reporting - Reduced false positives - Reduced risk - Productivity - Faster and easier data gathering - Data privacy concerns - Alert fatigue - Compatibility and configuration issues - AI used by cybercriminals - Marketing hype - Cost

will remain the dominant decision-makers. This is because of the high risk associated with automated cybersecurity processes, such as financial and reputational damages, as well as due to regulatory and compliance requirements.

A major concern identified in this study is the anxiety among cybersecurity professionals triggered by uncertainty about their future careers. This aligns with the findings of Gusman [44], who notes the need for cybersecurity employees to adjust and learn about AI and ML systems. Al-Dosari et al. [5] also stress the socioeconomic implications of AI integration. Organizational culture, leadership support, and strategic alignment are highlighted in this study as crucial factors for successful AI adoption. This is consistent with findings from Al-Dosari et al. [5], who emphasize how outdated organizational strategies can hinder in-house AI implementation.

The need for continuous education and training is a recurring theme across all studies. Our study highlights the importance of empowering analysts through education, which aligns with Gusman's [44] emphasis on the learning curve and the need for training, as well as the findings by Al-Dosari et al. [5]. As our study focuses predominantly on the in-house development and integration of AI, we emphasize the critical role of educating security analysts. This education is a key facilitator for successful interdisciplinary collaboration, which is essential for achieving successful AI-cybersecurity integration.

Building trust in AI solutions is a critical factor identified in our study, with a focus on how user collaboration, transparency, and simple explanations can bridge this gap. Interestingly, the other studies we compared with did not extensively address the aspects of building trust in AI. This gap highlights a unique contribution of our study, underscoring the necessity of integrating trust-building measures into the AI adoption process in cybersecurity.

Notably, both Al-Dosari et al. [5] and Radebe et al. [4] highlight the vulnerabilities of machine learning models and their potential to become targets for cyberattacks. In our study, this topic was not extensively addressed due to the limited number of participants who discussed it and their varied opinions. One participant expressed concerns about the inherent vulnerabilities of AI models, emphasizing the need for robust security measures to protect against adversarial attacks. Conversely, another participant was more optimistic, suggesting that such attacks would be difficult for threat actors to achieve without first gaining access to the organization's data.

The topic of adversarial learning attacks receives attention from researchers. They stress the possibility of attackers poisoning training data or manipulating the input data of trained models, leading to biased or incorrect predictions [3], [129]. However, while these attacks are technically feasible, researchers also acknowledge that executing them in real-world scenarios presents significant challenges due to factors such as the advanced defenses employed by cybersecurity systems [129]. While this may not be an immediate threat, attackers will likely continue to develop more sophisticated adversarial techniques. Therefore, defenders should proactively test the robustness of AI-enabled cybersecurity systems against these evolving threats [129].

7.3. Reflections

7.3.1. Expected and unexpected findings

It was expected that the interviews would highlight AI's potential in cybersecurity, including its ability to automate tasks and improve security detection. These benefits are consistent with the broader literature and industry trends, which highlight the transformative capabilities of AI in this domain (e.g., see Darktrace's most recent report [130]). Issues such as data quality, the scarcity of labeled data, and the complexity of developing suitable AI models are well documented in the literature. Hence, it was expected that these obstacles would also present significant barriers within the context of the Bank's cybersecurity operations. However, anticipating these challenges does not diminish their significance. Interviews with data scientists highlighted persistent problems related to data quality, noting that many data structures are designed at the back-end of applications solely to capture data, often without ensuring its quality. Implementing effective data management strategies is a foundational requirement for transitioning to an AI-driven cybersecurity department. These strategies should not only include robust data quality monitoring and auditing processes but also emphasize the importance of educating stakeholders about the critical role of data in analytics and their responsibilities within this framework.

The finding that AI will serve as a complementary tool rather than a replacement for human analysts is consistent with the prevailing view in the field. Considering the high-stakes environment of cybersecurity, particularly within a large bank, it was anticipated that participants in this case study would hold similar opinions. Despite AI's significant advancements, it remains susceptible to errors due to technological limitations and the complexity of its operation. Consequently, human analysts play an indispensable role in mitigating these flaws. This highlights the need for future initiatives to focus on optimizing the collaboration between humans and AI within this context, rather than concentrating solely on advancing the technology itself.

Research consistently underscores the significant influence of organizational culture and senior management support on the adoption of AI technologies. In line with this, our case study confirmed the critical role of these factors. For example, resistance to change can be a significant barrier to the adoption of new technologies, especially in established organizations where existing practices and mindsets are deeply ingrained. In this context, strategic alignment from leadership is essential to overcoming these challenges. This finding reinforces the notion that the journey toward AI integration is not solely a technological challenge but also a matter of effectively managing people and organizational dynamics.

While some level of anxiety regarding job security is anticipated with the introduction of AI technologies, the extent to which these concerns and career path ambiguities emerged as significant barriers was surprising. This indicates a deeper level of fear and resistance among security analysts than initially expected, emphasizing the importance of addressing these concerns proactively to mitigate resistance and foster acceptance of AI within the workforce. This means that there is a need for organizations to prioritize people-centric strategies alongside technical solutions. In Subsection 6.2.1, we proposed career mentorship as a key strategy. To develop these career mentorship programs, organizations should begin by conducting needs assessments to identify employees' skills, aspirations, and concerns related to AI integration. These insights will allow mentorship programs to be customized, providing targeted guidance and support that address the identified needs. Additionally, implementing pilot programs can help gather iterative feedback, and refine and adapt the mentorship initiatives to ensure alignment with both organizational objectives and individual goals.

Furthermore, the pronounced lack of trust in AI among security analysts, who often view AI as a "black box", was more significant than anticipated. This distrust implies that the perceived opacity of AI systems is a critical barrier to their adoption in cybersecurity. Interestingly, compared to similar studies (see Table 7.1), our research uniquely identified this issue as a significant challenge in the broader adoption of AI in cybersecurity. As detailed in our root cause analysis (see Section 6.1), the lack of trust in AI among security analysts is the result of several problems, including the analysts' lack of understanding of AI, the lack of transparency about how deployed ML models function, and the lack of effective explanations of model outputs. While it is relatively intuitive that transparency and explanations are essential for building trust, it was less expected that a lack of understanding of AI would also have a significant effect on trust.

One argument for why the lack of understanding of how AI functions significantly affects security analysts' trust in AI lies in the nature of their work and the historical context of cybersecurity practices. Unlike many technologies we use in daily life without fully understanding them, cybersecurity has traditionally relied on rule-based models, which are inherently easier to comprehend. These models operate on clear, predefined rules that analysts can understand, modify, and predict, providing a sense of control and transparency. The shift to AI-driven approaches introduces a level of complexity and opacity that contrasts with the simplicity of rule-based systems.

Furthermore, the high-risk nature of cybersecurity amplifies the need for understanding and trust in the tools used. Security analysts are responsible for protecting sensitive data and critical infrastructure, and their decisions can have significant and far-reaching implications. When analysts rely on AI tools that they do not fully understand, they may feel anxious about the potential consequences of their decisions. This anxiety may stem from the fear of unintended outcomes or missing critical threats due to the "black box" nature of AI, where the decision-making process is not transparent.

Unexpectedly, existing feedback mechanisms were found to be often ineffective, with analysts hesitant to provide honest feedback in formal settings. This suggests that current processes for collecting and integrating user feedback are not adequately fostering open communication, indicating the need

for some adjustments. An important contextual factor from our case study is the Bank's operation in an international environment, with its cybersecurity department distributed across different countries. Cultural differences play a significant role in how feedback is perceived and delivered [131]. This ties into the concept of *power distance*, or the degree to which individuals with less power within organizations accept and expect the unequal distribution of power [132]. In low power distance countries, open communication and feedback are typically more encouraged and expected, whereas in high power distance countries, there may be a greater reluctance to speak openly. To overcome these communication barriers within multicultural teams, one potential approach is to organize workshops focused on cross-cultural communication and feedback.

While the importance of continuous learning and innovation is well-known, the significant positive impact of events like hackathons on collaboration and practical understanding was particularly notable. These events promote creative problem-solving and help demystify AI, making it more approachable and understandable for analysts. Encouraging such innovative activities can enhance practical engagement with AI technologies.

Finally, the preference for straightforward, easily digestible explanations of AI model decisions over more complex technical details was somewhat unexpected. This finding emphasizes that while tools like SHAP are valuable for explainable AI (XAI), their complexity can be a barrier. Analysts benefit more from simplified, clear explanations that meet their practical needs, highlighting the need for user-centered AI explanation development and testing.

Supporting our findings, Suh et al. [133] reported that XAI methods like SHAP and LIME are often perceived as confusing by cybersecurity analysts. A potential solution could be using Large Language Models (LLMs) to translate SHAP outputs into user-friendly narratives. Ali and Kostakos [134] proposed an intrusion detection system using an LLM-based conversational agent, though without user testing to confirm its acceptance. Meanwhile, Zytek et al. [135] demonstrated that LLM-generated natural language explanations consistently outperformed SHAP plots in a user study, suggesting promise in this approach. Notably, LLMs can generate inaccurate explanations ("hallucinations"), emphasizing the importance of thorough evaluation to ensure their reliability.

7.3.2. Necessity of AI in cybersecurity

Given the evident challenges and resistance to adopting AI tools in cybersecurity, as highlighted by the findings of this research and supported by other sources [8], it is essential to consider whether this shift is indispensable for the field.

The necessity of AI in cybersecurity is underscored by several factors. The rapid advancement of digital technologies has changed the nature of cyber threats, creating new challenges for organizations around the world. Traditional cybersecurity measures, once considered sufficient, now struggle to contend with the evolving sophistication and frequency of modern cyberattacks [29]. As Robert Mueller, a former director of the Federal Bureau of Investigation (FBI), once said, "There are only two types of companies: those that have been hacked and those that will be. And even they are converging into one category: companies that have been hacked and will be hacked again" [136].

AI can address these challenges due to its capability to process large amounts of data swiftly and identify anomalies, allowing it to facilitate fast threat detection, a crucial feature for minimizing damage and mitigating risks [137]. This capability is particularly important in detecting advanced threats such as ransomware, zero-day exploits, and insider attacks, which require more proactive and adaptive defense mechanisms [29]. A recent case study by Goswami and colleagues [29], which explored the effects of implementing an AI-driven threat detection system at a bank, reports a significant decrease in false alerts (from 15% to 5%), an improvement in overall detection accuracy (from 68% to 86%), a decrease in the mean time to detect threats and enhanced operational efficiency. A study by Roelofs et al. [41] investigated the use of machine learning to combine weak signals from various independent detection systems within a large organization. The researchers discovered that this approach significantly improved the accuracy of attack detection while also reducing the number of alerts that security analysts needed to review. While our study hypothesized the benefits of AI for cybersecurity based on expert opinions, the literature provides concrete evidence that leveraging AI tools can significantly strengthen organizations' cybersecurity posture.

In contrast to traditional rule-based models, which operate on predefined rules and can be relatively static and inflexible, AI systems are dynamic and adaptable. Rule-based models require continuous manual updates to address new threats, which can be time-consuming and prone to human error. They are often limited in their ability to detect unknown or emerging threats, as their effectiveness is constrained by the rules they are programmed with.

AI can enhance decision-making by providing security analysts with insights and recommendations based on data-driven analysis. This support is crucial in a field where timely and informed decisions can prevent extensive damage. AI systems can prioritize threats, suggest remediation actions, and predict potential outcomes, thereby aiding analysts in making more effective decisions quickly.

7.3.3. In-house AI tools development and vendor solutions for cybersecurity

To ensure the manageability of this project, the research focused on the in-house development and implementation of AI for cybersecurity. The primary focus of the interviews was on the challenges associated with in-house AI development and use. However, it is important to note that these challenges are not unique to in-house development; they also apply to vendor solutions.

When discussing trust development, analysts, data scientists, and leaders indicated that this issue is even more pronounced with vendor solutions. Respondents noted that vendors often do not disclose AI model documentation, which can limit transparency regarding how the models are developed and their inner workings. While trust was identified as a significant issue in our research, addressing it may be more straightforward with in-house tools due to the greater control and transparency that internal development allows.

Vendor solutions, however, play a crucial role in the broader cybersecurity strategy. These tools often provide a robust foundation or baseline, enabling organizations to quickly deploy AI capabilities. By starting with established vendor solutions, in-house teams can then focus on customization and enhancement, adapting these tools to meet the specific needs and challenges of the organization.

In-house development offers several advantages in terms of trust, transparency, and explainability. When analysts and data scientists co-create AI solutions, trust development is facilitated through shared collaboration and involvement. This collaborative process allows stakeholders to gain a deeper understanding of the AI models, fostering a sense of ownership and confidence in the tools.

Additionally, in-house development allows data scientists to implement mechanisms that make AI models more transparent and explainable. By incorporating explainability layers, in-house teams can demystify the decision-making process of AI models. This transparency is particularly crucial in cybersecurity, where understanding the rationale behind AI-driven decisions is essential for analysts to make informed and timely responses.

7.3.4. The need for human-centered AI development and implementation

Many studies on AI in cybersecurity adopt a technocratic view, focusing predominantly on technological capabilities without adequately considering human factors. However, our study employs a sociotechnical framework, emphasizing the dynamics between social and technical components in the integration of AI. This approach emerged from a preliminary survey that revealed a low adoption rate of AI technologies in the Bank's cybersecurity department, a finding consistent with existing literature.

This insight underscores that even highly accurate AI systems may face adoption challenges due to complex human factors. The promise of AI cannot be fully realized if solutions are not designed with human properties and needs in mind. Our case study highlights the pivotal role of security analysts in AI development for cybersecurity. These professionals not only identify use cases based on their daily challenges but also assist data scientists in understanding data, developing model features, and testing the resulting AI models.

Security experts act as both users and domain experts, making their involvement crucial for the successful adoption of AI in cybersecurity. Ignoring their role and the human factors associated with it can impede the effective integration and utilization of AI tools. Therefore, AI solutions for cybersecurity must be engineered with a clear focus on human-centered design, ensuring that the needs and expertise of security analysts are integral to the development process. This approach can enhance the utilization

of AI tools and foster trust and acceptance among the workforce, ultimately leading to more effective cybersecurity practices.

7.4. Limitations

Organizational context
This research is based on a single-case study of a large European bank, providing in-depth insights into AI adoption within a specific organizational and cultural context. While this focus allowed for a detailed exploration of relevant dynamics, the findings may not be fully generalizable to other organizations, industries, or geographic regions. The specific organizational culture, structure, and practices of the studied company can influence the results, meaning that different organizations might face unique challenges and opportunities not captured in this study. However, the alignment of our results with those from comparable studies (see Section 7.2) suggests that the insights gained are likely to be generalizable beyond this immediate context.

Sample size
The study involved 15 participants, which is considered sufficient for qualitative analysis [73] and larger than in some other studies (see Table 7.1). While this sample size allowed for a sufficient examination of key themes, a larger and more diverse sample could provide additional perspectives and capture further nuances.

Participant selection and response bias
A possible limitation of this study is the selection of participants, which may have introduced response bias. Response bias arises when the chosen participants do not accurately represent the broader sample population or the entire industry [138]. In this case, the cybersecurity analysts who were included in the study were predominantly those who were already involved in AI projects or had some level of engagement with AI technologies. This selection potentially skews the findings, as these analysts may hold more favorable views toward AI, given their direct experience and involvement in its implementation. To mitigate this limitation, the interview questions were designed to encourage participants to reflect not only on their own experiences but also on those of their colleagues. This approach aimed to capture a broader range of perspectives, helping to balance the potential bias introduced by the participants' direct involvement with AI.

Qualitative methodology
This research primarily employed qualitative methods, which are well-suited for exploring complex, context-specific phenomena like AI adoption in cybersecurity. While qualitative approaches provide rich, detailed data, they can also be influenced by biases such as interviewer influence and participant subjectivity.

Time constraints
The six-month duration of this study provided a snapshot of AI adoption at a particular moment in time. While this was sufficient to capture key dynamics, it may not fully reflect the long-term impacts and evolving nature of AI integration in cybersecurity. Longitudinal studies could build on this research by offering deeper insights into how AI adoption influences organizations over time.

Researcher bias
Given that the study and subsequent data coding were conducted by a single researcher, there is a possibility of researcher bias. Despite efforts to maintain objectivity through systematic analysis and reflection, the findings may still reflect some degree of personal bias.

7.5. Future research

Generalizability
To increase the generalizability of these findings, future research should consider conducting multiple case studies across a variety of settings and organizational contexts, such as comparing cybersecurity practices in small, medium and large enterprises, or examining different industries like finance, healthcare, and manufacturing. This would help determine whether the findings hold true across different types of organizations and sectors.

Additionally, incorporating mixed-method approaches in future studies could strengthen the evidence

base by complementing qualitative insights with quantitative data. This combination would not only validate the findings through numerical data but also help capture more diverse perspectives by allowing researchers to explore patterns on a larger scale. For example, while qualitative interviews might uncover nuanced experiences and opinions, quantitative surveys could reach a broader audience, revealing trends and variations across different demographics or organizational roles.

Human-centered AI development frameworks
Building on the sociotechnical approach of this study, future research could focus on developing comprehensive frameworks for human-centered AI development in cybersecurity. These frameworks should integrate social factors, such as user trust and organizational culture, with technical considerations to ensure that AI tools are both effective and widely adopted.

Practical exploration of AI trust-building mechanisms
Future research should explore specific strategies that can effectively build trust in AI among security analysts. While this study identified transparency, active user involvement, and clear explanations as some of the key factors, more work is needed to develop and test practical trust-building interventions tailored to different organizational contexts. Future studies should explore which elements of transparency (e.g., visibility into AI decision-making processes, access to model documentation, or clarity in data sources) are most crucial for fostering trust and how these elements vary across different types of AI models or organizational contexts.

In line with our stands of human-centered AI development and use, future work should conduct user studies with security analysts to determine what types of explanations they find most useful. These studies should explore what formats or levels of detail in AI model explanations are most suitable and how these can be tailored to different levels of expertise or different cybersecurity use cases.

Longitudinal studies on AI integration in cybersecurity
Long-term studies that follow the integration of AI into cybersecurity over several years would provide valuable insights into how trust and cross-team collaboration evolve. Insights into how trust in AI develops can be used to design governance frameworks that ensure AI tools are implemented responsibly, with clear guidelines for transparency, accountability, and ethical use. These studies could also track the career development of security analysts as AI tools become more prevalent, providing a clearer picture of the impact on job roles and career trajectories. Understanding the long-term impact of AI on job roles and collaboration can inform the design of targeted training programs that address specific skill gaps and prepare security analysts for the evolving demands of their roles.

Longitudinal insights can provide empirical evidence on how the interaction between social and technical factors evolves. These can contribute to the refinement of the proposed sociotechnical conceptual model and the development of generalized frameworks that could be applied to other sectors facing similar AI integration challenges.

8

Conclusion

In this thesis, we explored the adoption of AI in cybersecurity, specifically within the context of a large European bank's cybersecurity department. By employing a sociotechnical system (STS) approach, we explicitly considered both the technical and social dimensions of AI integration. Our approach was successfully evaluated through a case study that involved semi-structured interviews with key stakeholders, including security analysts, data scientists, and leaders within the department.

Through a thorough sociotechnical system analysis, we identified critical interactions between the technical systems and social structures that influence the adoption of AI in cybersecurity. This analysis revealed that successful AI adoption relies on the cohesive interaction between people, organizational structures, and technological systems. Our qualitative methodology allowed us to explore the intersections between these components, uncovering both the potential and the challenges of AI integration.

Our results confirmed the existing literature, notably the works of Gusman [44], Al-Dosari et al. [5], and Radebe et al. [4], while also expanding on important topics such as the organizational readiness for AI, the influence of job security concerns, and the necessity for effective change management. Additionally, we concluded that while AI offers significant potential for enhancing efficiency in cybersecurity, it also introduces challenges related to user trust and system transparency.

Our methodological framework provided the opportunity to examine in greater detail the interactions between the technical and social elements by conducting a root cause analysis. It particularly highlighted the interdisciplinary aspect of AI-enabled cybersecurity applications within the financial industry and the role of the organization, with its culture, leadership, and processes, in the success of AI adoption. With that in mind, we proposed a new conceptual model that combines both technical and social factors and discusses the interactions among them, providing bringing mechanisms that can help organizations navigate the AI transition. This framework enhances the understanding of AI adoption in cybersecurity and provides practical recommendations for improving the development, deployment, and acceptance of AI systems.

In conclusion, the adoption of AI in cybersecurity represents a significant opportunity to transform threat management and operational efficiency. However, its successful integration requires a holistic approach that addresses both technical and social challenges. By bridging the sociotechnical gap, as outlined in our proposed framework, organizations can better interpret and augment their understanding of the complex relationships among technology, people, and processes, ultimately enhancing their cybersecurity capabilities.

Academic contributions
This research contributes to the academic body of knowledge by providing a detailed examination of the sociotechnical factors influencing AI adoption in cybersecurity. By integrating sociotechnical systems theory and innovation adoption models with root cause analysis techniques, the study offers a comprehensive framework that can be used to understand the complexities of AI implementation in other contexts. Specifically, this research adds to the body of knowledge in the disciplines of orga-

nizational behavior, innovation management, information systems, and cybersecurity. Through this multidisciplinary approach, the study provides valuable insights for researchers and practitioners seeking to navigate the transition to integrating AI technologies within organizational settings.

Practical relevance

For practitioners, this study offers actionable recommendations to enhance the development and implementation of AI-based cybersecurity systems. The insights into the organizational, technical, and social barriers to AI adoption can help cybersecurity departments better prepare for and manage the transition to AI-enhanced operations. The study's findings on user experience, trust in AI systems, and the importance of transparent and structured processes are particularly relevant for improving the practical deployment of these technologies.

Policy and management implications

The study also has implications for policy and management within organizations. It highlights the need for strategic alignment, resource allocation, and the importance of fostering a supportive organizational culture for successful AI adoption. By addressing job security concerns and emphasizing the need for continuous education and training, the research provides managers with essential strategies to ensure a smooth integration of AI technologies, benefiting both the organization and its employees.

References

[1] C. Feng, S. Wu, and N. Liu, "A user-centric machine learning framework for cyber security operations center," in *2017 IEEE International Conference on Intelligence and Security Informatics (ISI)*, IEEE, 2017, pp. 173–175.

[2] A. D. Sontan and S. V. Samuel, "The intersection of artificial intelligence and cybersecurity: Challenges and opportunities," *World Journal of Advanced Research and Reviews*, vol. 21, no. 2, pp. 1720–1736, 2024.

[3] B. T. Familoni, "Cybersecurity challenges in the age of ai: Theoretical approaches and practical solutions," *Computer Science & IT Research Journal*, vol. 5, no. 3, pp. 703–724, 2024.

[4] M. Radebe, P. Tsibolane, and M. Hart, "Perceptions of ai tools for cybersecurity in large enterprises," in *Proceedings of the EWG-DSS 2022 International Conference on Decision Support System Technology (ICDSST 2022)*, European Working Group on Decision Support Systems (EWG-DSS), Thessaloniki, Greece, May 2022. [Online]. Available: https://www.researchgate.net/publication/364899781_Perceptions_of_AI_Tools_for_Cybersecurity_in_Large_Enterprises.

[5] K. Al-Dosari, N. Fetais, and M. Kucukvar, "Artificial intelligence and cyber defense system for banking industry: A qualitative study of ai applications and challenges," *Cybernetics and systems*, vol. 55, no. 2, pp. 302–330, 2024.

[6] Darktrace, *Cyber ai glossary - incident response*, https://darktrace.com/cyber-ai-glossary/incident-response, Accessed: 2024-07-14.

[7] Cisco, *Artificial intelligence in security*, https://www.cisco.com/c/en/us/products/security/artificial-intelligence-ai.html, Accessed: 2024-07-14.

[8] R. Nwaiwu and M. Keeris, *The ai cyber security challenge*, Accessed: 2024-08-11, 2024. [Online]. Available: https://kpmg.com/nl/en/home/insights/2024/06/ai-cyber-security-challenge.html.

[9] L. Chan, I. Morgan, H. Simon, *et al.*, "Survey of ai in cybersecurity for information technology management," in *2019 IEEE technology & engineering management conference (TEMSCON)*, IEEE, 2019, pp. 1–8.

[10] R. Oosthuizen and L. Pretorius, "Modelling of command and control agility," 2014.

[11] G. H. Walker, N. A. Stanton, P. M. Salmon, and D. P. Jenkins, "A review of sociotechnical systems theory: A classic concept for new command and control paradigms," *Theoretical issues in ergonomics science*, vol. 9, no. 6, pp. 479–499, 2008.

[12] M. A. Hameed, S. Counsell, and S. Swift, "A conceptual model for the process of it innovation adoption in organizations," *Journal of Engineering and Technology Management*, vol. 29, no. 3, pp. 358–390, 2012.

[13] M. S. Ackerman, "The intellectual challenge of cscw: The gap between social requirements and technical feasibility," *Human–Computer Interaction*, vol. 15, no. 2-3, pp. 179–203, 2000.

[14] R. K. Yin, *Case study research: Design and methods*. sage, 2009, vol. 5.

[15] V. Clarke and V. Braun, "Thematic analysis," *The journal of positive psychology*, vol. 12, no. 3, pp. 297–298, 2017.

[16] E. A. Fischer, *Cybersecurity issues and challenges: In brief*, 2016.

[17] A. R. D. Rodrigues, F. A. Ferreira, F. J. Teixeira, and C. Zopounidis, "Artificial intelligence, digital transformation and cybersecurity in the banking sector: A multi-stakeholder cognition-driven framework," *Research in International Business and Finance*, vol. 60, p. 101 616, 2022.

[18] A. Bendovschi, “Cyber-attacks–trends, patterns and security countermeasures,” *Procedia Economics and Finance*, vol. 28, pp. 24–31, 2015.

[19] European Parliament and Council of the European Union, *Proposal for a Regulation of the European Parliament and of the Council laying downharmonised rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain Union legislative acts*, en, 2021/0106 (COD), 2021. [Online]. Available: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A52021PC0206 (visited on 05/15/2024).

[20] E. Alpaydin, *Machine learning*. MIT press, 2021.

[21] A. B. Arrieta, N. Díaz-Rodríguez, J. Del Ser, *et al.*, “Explainable artificial intelligence (xai): Concepts, taxonomies, opportunities and challenges toward responsible ai,” *Information fusion*, vol. 58, pp. 82–115, 2020.

[22] D. Gunning and D. Aha, “Darpa’s explainable artificial intelligence (xai) program,” *AI magazine*, vol. 40, no. 2, pp. 44–58, 2019.

[23] M. T. Ribeiro, S. Singh, and C. Guestrin, “” why should i trust you?” explaining the predictions of any classifier,” in *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, 2016, pp. 1135–1144.

[24] S. M. Lundberg and S.-I. Lee, “A unified approach to interpreting model predictions,” *Advances in neural information processing systems*, vol. 30, 2017.

[25] S. A. Mongeau and A. Hajdasinski, *Cybersecurity Data Science*. Springer, 2021.

[26] S. Saeed, S. A. Altamimi, N. A. Alkayyal, E. Alshehri, and D. A. Alabbad, “Digital transformation and cybersecurity challenges for businesses resilience: Issues and recommendations,” *Sensors*, vol. 23, no. 15, p. 6666, 2023.

[27] “Directive (EU) 2016/1148 of the European Parliament and of the Council of 6 July 2016 concerning measures for a high common level of security of network and information systems across the Union,” European Union, 2016, Accessed: 2024-07-09. [Online]. Available: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32016L1148.

[28] “Regulation (EU) 2022/2554 of the European Parliament and of the Council of 14 December 2022 on digital operational resilience for the financial sector and amending Regulations (EC) No 1060/2009, (EU) No 648/2012, (EU) No 600/2014, and (EU) No 909/2014,” European Union, 2022, Accessed: 2024-07-09. [Online]. Available: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32022R2554.

[29] S. Goswami, S. Mondal, R. Halder, J. Nayak, and A. Sil, “Exploring the impact of artificial intelligence integration on cybersecurity: A comprehensive analysis,” *J. Ind Intell*, vol. 2, no. 2, pp. 73–93, 2024.

[30] D. P. Möller, “Cybersecurity in digital transformation,” in *Guide to Cybersecurity in Digital Transformation: Trends, Methods, Technologies, Applications and Best Practices*, Springer, 2023, pp. 1–70.

[31] M. F. Ansari, B. Dash, P. Sharma, and N. Yathiraju, “The impact and limitations of artificial intelligence in cybersecurity: A literature review,” *International Journal of Advanced Research in Computer and Communication Engineering*, 2022.

[32] D. Jin, Y. Lu, J. Qin, Z. Cheng, and Z. Mao, “Swiftids: Real-time intrusion detection system based on lightgbm and parallel intrusion detection mechanism,” *Computers & Security*, vol. 97, p. 101 984, 2020.

[33] I. H. Sarker, Y. B. Abushark, F. Alsolami, and A. I. Khan, “Intrudtree: A machine learning based cyber security intrusion detection model,” *Symmetry*, vol. 12, no. 5, p. 754, 2020.

[34] V. Agrawal, M. Hazratifard, H. Elmiligi, and F. Gebali, “Electrocardiogram (ecg)-based user authentication using deep learning algorithms,” *Diagnostics*, vol. 13, no. 3, p. 439, 2023.

[35] A. Buriro, B. Crispo, and Y. Zhauniarovich, “Please hold on: Unobtrusive user authentication using smartphone’s built-in sensors,” in *2017 IEEE International Conference on Identity, Security and Behavior Analysis (ISBA)*, IEEE, 2017, pp. 1–8.

[36] T. H. Aldhyani and H. Alkahtani, “Cyber security for detecting distributed denial of service attacks in agriculture 4.0: Deep learning model,” *Mathematics*, vol. 11, no. 1, p. 233, 2023.

[37] A. Djenna, A. Bouridane, S. Rubab, and I. M. Marou, “Artificial intelligence-based malware detection, analysis, and mitigation,” *Symmetry*, vol. 15, no. 3, p. 677, 2023.

[38] J.-h. Li, “Cyber security meets artificial intelligence: A survey,” *Frontiers of Information Technology & Electronic Engineering*, vol. 19, no. 12, pp. 1462–1474, 2018.

[39] N. Mohamed, “Current trends in ai and ml for cybersecurity: A state-of-the-art survey,” *Cogent Engineering*, vol. 10, no. 2, p. 2 272 358, 2023.

[40] M. Baruwal Chhetri, S. Tariq, R. Singh, F. Jalalvand, C. Paris, and S. Nepal, “Towards human-ai teaming to mitigate alert fatigue in security operations centres,” *ACM Transactions on Internet Technology*, vol. 24, no. 3, pp. 1–22, 2024.

[41] T.-M. Roelofs, E. Barbaro, S. Pekarskikh, *et al.*, “Finding harmony in the noise: Blending security alerts for attack detection,” in *Proceedings of the 39th ACM/SIGAPP Symposium on Applied Computing*, 2024, pp. 1385–1394.

[42] T. Ban, N. Samuel, T. Takahashi, and D. Inoue, “Combat security alert fatigue with ai-assisted techniques,” in *Proceedings of the 14th Cyber Security Experimentation and Test Workshop*, 2021, pp. 9–16.

[43] A. S. Jacobs, R. Beltiukov, W. Willinger, R. A. Ferreira, A. Gupta, and L. Z. Granville, “Ai/ml for network security: The emperor has no clothes,” in *Proceedings of the 2022 ACM SIGSAC Conference on Computer and Communications Security*, 2022, pp. 1537–1551.

[44] J. Gusman, “A qualitative study on the deployment of artificial intelligence and machine learning within cybersecurity for intelligent decision making,” PhD dissertation, Capella University, Aug. 2023.

[45] H. Gonaygunta, “Factors influencing the adoption of machine learning algorithms to detect cyber threats in the banking industry,” PhD dissertation, University of the Cumberlands, 2023.

[46] J. Baker, “The technology–organization–environment framework,” *Information Systems Theory: Explaining and Predicting Our Digital Society, Vol. 1*, pp. 231–245, 2012.

[47] T. Oliveira and M. F. Martins, “Literature review of information technology adoption models at firm level,” *Electronic journal of information systems evaluation*, vol. 14, no. 1, pp110–121, 2011.

[48] F. D. Davis, “Perceived usefulness, perceived ease of use, and user acceptance of information technology,” *MIS quarterly*, pp. 319–340, 1989.

[49] E. M. Rogers, *Diffusion of Innovations*. Simon and Schuster, 2010.

[50] I. Ajzen, “The theory of planned behavior,” *Organizational behavior and human decision processes*, vol. 50, no. 2, pp. 179–211, 1991.

[51] M. Fishbein and I. Ajzen, “Belief, attitude, intention, and behavior: An introduction to theory and research,” 1977.

[52] L. G. Tornatzky, M. Fleischer, and A. K. Chakrabarti, *The Processes of Technological Innovation* (Issues in Organization and Management Series). Lexington, Mass: Lexington Books, 1990.

[53] R. H. Shroff, C. C. Deneen, and E. M. Ng, “Analysis of the technology acceptance model in examining students’ behavioural intention to use an e-portfolio system,” *Australasian Journal of Educational Technology*, vol. 27, no. 4, 2011.

[54] S.-C. Chen, L. Shing-Han, and L. Chien-Yi, “Recent related research in technology acceptance model: A literature review,” *Australian journal of business and management research*, vol. 1, no. 9, p. 124, 2011.

[55] Y. Malhotra and D. F. Galletta, “Extending the technology acceptance model to account for social influence: Theoretical bases and empirical validation,” in *Proceedings of the 32nd Annual Hawaii International Conference on Systems Sciences. 1999. HICSS-32. Abstracts and CD-ROM of Full Papers*, IEEE, 1999, 14–pp.

[56] A. Alomary and J. Woollard, “How is technology accepted by users? a review of technology acceptance models and theories,” 2015.

[57] P. Legris, J. Ingham, and P. Collerette, “Why do people use information technology? a critical review of the technology acceptance model,” *Information & management*, vol. 40, no. 3, pp. 191–204, 2003.

[58] S. K. Sharma and J. K. Chandel, “Technology acceptance model for the use of learning through websites among students,” *International Arab journal of e-Technology*, vol. 3, no. 1, pp. 44–49, 2013.

[59] R. P. Bagozzi, “The legacy of the technology acceptance model and a proposal for a paradigm shift.,” *Journal of the association for information systems*, vol. 8, no. 4, p. 3, 2007.

[60] S. Taylor and P. A. Todd, “Understanding information technology usage: A test of competing models,” *Information systems research*, vol. 6, no. 2, pp. 144–176, 1995.

[61] M.-B. Owolabi Yusuf and A. Mat Derus, “Measurement model of corporate zakat collection in malaysia: A test of diffusion of innovation theory,” *Humanomics*, vol. 29, no. 1, pp. 61–74, 2013.

[62] R. P. Bostrom and J. S. Heinen, “Mis problems and failures: A socio-technical perspective. part i: The causes,” *MIS quarterly*, pp. 17–32, 1977.

[63] E. L. Trist, *The evolution of socio-technical systems*. Ontario Quality of Working Life Centre Toronto, 1981, vol. 2.

[64] C. W. Clegg, “Sociotechnical principles for system design,” *Applied ergonomics*, vol. 31, no. 5, pp. 463–477, 2000.

[65] P. Waterson and K. Eason, “Revisiting the sociotechnical principles for system design (clegg, 2000),” in *Proceedings of the 20th Congress of the International Ergonomics Association (IEA 2018) Volume VII: Ergonomics in Design, Design for All, Activity Theories for Work Analysis and Design, Affective Design 20*, Springer, 2019, pp. 366–374.

[66] U. Ehsan, K. Saha, M. De Choudhury, and M. O. Riedl, “Charting the sociotechnical gap in explainable ai: A framework to address the gap in xai,” *Proceedings of the ACM on human-computer interaction*, vol. 7, no. CSCW1, pp. 1–32, 2023.

[67] M. Q. Patton, *Qualitative research & evaluation methods: Integrating theory and practice*. Sage publications, 2014.

[68] W. J. Orlikowski and D. C. Gash, “Technological frames: Making sense of information technology in organizations,” *ACM Transactions on Information Systems (TOIS)*, vol. 12, no. 2, pp. 174–207, 1994.

[69] H. Alshenqeeti, “Interviewing as a data collection method: A critical review,” *English linguistics research*, vol. 3, no. 1, pp. 39–45, 2014.

[70] A. Lareau, *Listening to people: A practical guide to interviewing, participant observation, data analysis, and writing it all up*. University of Chicago Press, 2021.

[71] A. Chaudhuri and T. C. Christofides, *Indirect questioning in sample surveys*. Springer Science & Business Media, 2013.

[72] A. J. Nederhof, “Methods of coping with social desirability bias: A review,” *European journal of social psychology*, vol. 15, no. 3, pp. 263–280, 1985.

[73] M. Hennink and B. N. Kaiser, “Sample sizes for saturation in qualitative research: A systematic review of empirical tests,” *Social science & medicine*, vol. 292, p. 114 523, 2022.

[74] V. Braun and V. Clarke, “Reflecting on reflexive thematic analysis,” *Qualitative research in sport, exercise and health*, vol. 11, no. 4, pp. 589–597, 2019.

[75] C. M. Bird, “How i stopped dreading and learned to love transcription,” *Qualitative inquiry*, vol. 11, no. 2, pp. 226–248, 2005.

[76] V. Clarke and V. Braun, “Successful qualitative research: A practical guide for beginners,” 2013.

[77] D. Byrne, “A worked example of braun and clarke’s approach to reflexive thematic analysis,” *Quality & quantity*, vol. 56, no. 3, pp. 1391–1412, 2022.

[78] H. J. Wilson, P. R. Daugherty, and N. Morini-Bianzino, “The jobs that artificial intelligence will create,” 2018.

[79] M. H. Jarrahi, “Artificial intelligence and the future of work: Human-ai symbiosis in organizational decision making,” *Business horizons*, vol. 61, no. 4, pp. 577–586, 2018.

[80] D. Acemoglu and P. Restrepo, “Artificial intelligence, automation, and work,” in *The economics of artificial intelligence: An agenda*, University of Chicago Press, 2018, pp. 197–236.

[81] J. Jöhnk, M. Weißert, and K. Wyrtki, “Ready or not, ai comes—an interview study of organizational ai readiness factors,” *Business & Information Systems Engineering*, vol. 63, no. 1, pp. 5–20, 2021.

[82] K. Szczepańska-Woszczyna, “The importance of organizational culture for innovation in the company,” in *Forum scientiae oeconomia*, vol. 2, 2014, pp. 27–39.

[83] J. Jeong, B.-J. Kim, and J. Lee, “Navigating ai transitions: How coaching leadership buffers against job stress and protects employee physical health,” *Frontiers in public health*, vol. 12, p. 1 343 932, 2024.

[84] N.-A. Perifanis and F. Kitsios, “Investigating the influence of artificial intelligence on business value in the digital era of strategy: A literature review,” *Information*, vol. 14, no. 2, p. 85, 2023.

[85] I. M. Enholm, E. Papagiannidis, P. Mikalef, and J. Krogstie, “Artificial intelligence and business value: A literature review,” *Information Systems Frontiers*, vol. 24, no. 5, pp. 1709–1734, 2022.

[86] S.-L. Wamba-Taguimdje, S. F. Wamba, J. R. K. Kamdjoug, and C. E. T. Wanko, “Influence of artificial intelligence (ai) on firm performance: The business value of ai-based transformation projects,” *Business process management journal*, vol. 26, no. 7, pp. 1893–1924, 2020.

[87] N. G. Camacho, “The role of ai in cybersecurity: Addressing threats in the digital age,” *Journal of Artificial Intelligence General science (JAIGS) ISSN: 3006-4023*, vol. 3, no. 1, pp. 143–154, 2024.

[88] P. Moradi and K. Levy, “The future of work in the age of ai: Displacement or risk-shifting?,” 2020.

[89] J. Rollins, *Foundational methodology for data science*, 2015.

[90] S. E. Humphrey, J. D. Nahrgang, and F. P. Morgeson, “Integrating motivational, social, and contextual work design features: A meta-analytic summary and theoretical extension of the work design literature.,” *Journal of applied psychology*, vol. 92, no. 5, p. 1332, 2007.

[91] C. Maslach, W. B. Schaufeli, and M. P. Leiter, “Job burnout,” *Annual review of psychology*, vol. 52, no. 1, pp. 397–422, 2001.

[92] R. A. Karasek Jr, “Job demands, job decision latitude, and mental strain: Implications for job redesign,” *Administrative science quarterly*, pp. 285–308, 1979.

[93] M. Gagné and E. L. Deci, “Self-determination theory and work motivation,” *Journal of Organizational behavior*, vol. 26, no. 4, pp. 331–362, 2005.

[94] M. R. Wade and M. Parent, “Relationships between job skills and performance: A study of webmasters,” *Journal of Management Information Systems*, vol. 18, no. 3, pp. 71–96, 2002.

[95] B. Andersen and T. Fagerhaug, *Root cause analysis*. Quality Press, 2006.

[96] J. C. Paterson, *Beyond the Five Whys: Root Cause Analysis and Systems Thinking*. United Kingdom: Wiley, 2023.

[97] A. H. Yahya and V. Sukmayadi, “A review of cognitive dissonance theory and its relevance to current social issues,” *MIMBAR: Jurnal Sosial Dan Pembangunan*, vol. 36, no. 2, pp. 480–488, 2020.

[98] J. Y. Thong, C.-S. Yap, and K. Raman, “Top management support, external expertise and information systems implementation in small businesses,” *Information systems research*, vol. 7, no. 2, pp. 248–267, 1996.

[99] V. Lai, C. Chen, Q. V. Liao, A. Smith-Renner, and C. Tan, “Towards a science of human-ai decision making: A survey of empirical studies,” *arXiv preprint arXiv:2112.11471*, 2021.

[100] Y. Zhang, Q. V. Liao, and R. K. Bellamy, “Effect of confidence and explanation on accuracy and trust calibration in ai-assisted decision making,” in *Proceedings of the 2020 conference on fairness, accountability, and transparency*, 2020, pp. 295–305.

[101] R. L. Campagna, A. A. Mislin, K. T. Dirks, and H. A. Elfenbein, “The (mostly) robust influence of initial trustworthiness beliefs on subsequent behaviors and perceptions,” *Human Relations*, vol. 75, no. 7, pp. 1383–1411, 2022.

[102] H. Kong, Z. Yin, Y. Baruch, and Y. Yuan, “The impact of trust in ai on career sustainability: The role of employee–ai collaboration and protean career orientation,” *Journal of Vocational Behavior*, vol. 146, p. 103 928, 2023.

[103] S. O. Bada and S. Olusegun, “Constructivism learning theory: A paradigm for teaching and learning,” *Journal of Research & Method in Education*, vol. 5, no. 6, pp. 66–70, 2015.

[104] D. R. Michael and S. L. Chen, *Serious Games: Games That Educate, Train, and Inform*. Muska & Lipman/Premier-Trade, 2005, ISBN: 1592006221.

[105] K. M. Kapp, *The gamification of learning and instruction: game-based methods and strategies for training and education*. John Wiley & Sons, 2012.

[106] A. Bandura, *Self-efficacy: The exercise of control*. Macmillan, 1997.

[107] T. A. Judge and J. E. Bono, “Relationship of core self-evaluations traits—self-esteem, generalized self-efficacy, locus of control, and emotional stability—with job satisfaction and job performance: A meta-analysis.,” *Journal of applied Psychology*, vol. 86, no. 1, p. 80, 2001.

[108] N. E. Betz, K. L. Klein, and K. M. Taylor, “Evaluation of a short form of the career decision-making self-efficacy scale,” *Journal of career assessment*, vol. 4, no. 1, pp. 47–57, 1996.

[109] J. Cherian and J. Jacob, “Impact of self efficacy on motivation and performance of employees,” *International journal of business and management*, vol. 8, no. 14, p. 80, 2013.

[110] E. L. Deci and R. M. Ryan, “The” what” and” why” of goal pursuits: Human needs and the self-determination of behavior,” *Psychological inquiry*, vol. 11, no. 4, pp. 227–268, 2000.

[111] A. D. Stajkovic and F. Luthans, “Differential effects of incentive motivators on work performance,” *Academy of management journal*, vol. 44, no. 3, pp. 580–590, 2001.

[112] S. L. Malek, S. Sarin, and C. Haon, “Extrinsic rewards, intrinsic motivation, and new product development performance,” *Journal of product innovation management*, vol. 37, no. 6, pp. 528–551, 2020.

[113] S. Kerr and G. Rifkin, *Reward systems: Does yours measure up?* Harvard Business Press, 2008.

[114] D. International, *DAMA-DMBOK: Data management body of knowledge*. Technics Publications, LLC, 2017.

[115] DataCamp, *Pillars of data management*, Accessed: 2024-07-13, n.d. [Online]. Available: `https://campus.datacamp.com/courses/data-management-concepts/pillars-of-data-management?ex=10`.

[116] A. C. Edmondson, *The fearless organization: Creating psychological safety in the workplace for learning, innovation, and growth*. John Wiley & Sons, 2018.

[117] M. Jabri, “Team feedback based on dialogue: Implications for change management,” *Journal of Management Development*, vol. 23, no. 2, pp. 141–151, 2004.

[118] F. Foroughi and P. Luksch, “Data science methodology for cybersecurity projects,” *arXiv preprint arXiv:1803.04219*, 2018.

[119] U. Feyyad, “Data mining and knowledge discovery: Making sense out of data,” *IEEE expert*, vol. 11, no. 5, pp. 20–25, 1996.

[120] A. Azevedo and M. F. Santos, “Kdd, semma and crisp-dm: A parallel overview,” *IADS-DM*, 2008.

[121] Microsoft, *Data science process: Overview*, Accessed: 2024-07-13, n.d. [Online]. Available: `https://learn.microsoft.com/en-us/azure/architecture/data-science-process/overview`.

[122] M. Mitchell, S. Wu, A. Zaldivar, *et al.*, “Model cards for model reporting,” in *Proceedings of the conference on fairness, accountability, and transparency*, 2019, pp. 220–229.

[123] J. van der Waa, E. Nieuwburg, A. Cremers, and M. Neerincx, "Evaluating xai: A comparison of rule-based and example-based explanations," *Artificial intelligence*, vol. 291, p. 103 404, 2021.

[124] Q. V. Liao, D. Gruen, and S. Miller, "Questioning the ai: Informing design practices for explainable ai user experiences," in *Proceedings of the 2020 CHI conference on human factors in computing systems*, 2020, pp. 1–15.

[125] Q. V. Liao, M. Pribić, J. Han, S. Miller, and D. Sow, "Question-driven design process for explainable ai user experiences," *arXiv preprint arXiv:2104.03483*, 2021.

[126] F. Doshi-Velez and B. Kim, "Towards a rigorous science of interpretable machine learning," *arXiv preprint arXiv:1702.08608*, 2017.

[127] S. S. Kim, E. A. Watkins, O. Russakovsky, R. Fong, and A. Monroy-Hernández, "" help me help the ai": Understanding how explainability can support human-ai interaction," in *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, 2023, pp. 1–17.

[128] K. Amarasinghe, K. T. Rodolfa, S. Jesus, *et al.*, "On the importance of application-grounded experimental design for evaluating explainable ml methods," *arXiv e-prints*, 2023.

[129] I. Rosenberg, A. Shabtai, Y. Elovici, and L. Rokach, "Adversarial machine learning attacks and defense methods in the cyber security domain," *ACM Computing Surveys (CSUR)*, vol. 54, no. 5, pp. 1–36, 2021.

[130] "State of ai cyber security 2024," Darktrace, 2024. [Online]. Available: `https://darktrace.com/state-of-ai-cyber-security`.

[131] E. Meyer, *The culture map: Breaking through the invisible boundaries of global business*. Hachette Book Group USA, 2014.

[132] G. Hofstede, "Dimensionalizing cultures: The hofstede model in context," *Online readings in psychology and culture*, vol. 2, no. 1, p. 8, 2011.

[133] A. Suh, H. Li, C. Kenney, K. Alperin, and S. R. Gomez, *More questions than answers? lessons from integrating explainable ai into a cyber-ai tool*, ACM CHI 2024 Workshop on Human-Centered Explainable AI (HCXAI), 2024. arXiv: `2408.04746`. [Online]. Available: `https://arxiv.org/abs/2408.04746`.

[134] T. Ali and P. Kostakos, "Huntgpt: Integrating machine learning-based anomaly detection and explainable ai with large language models (llms)," *arXiv preprint arXiv:2309.16021*, 2023.

[135] A. Zytek, S. Pidò, and K. Veeramachaneni, "Llms for xai: Future directions for explaining explanations," *arXiv preprint arXiv:2405.06064*, 2024.

[136] R. S. Mueller, *Combating threats in the cyber world: Outsmarting terrorists, hackers, and spies*, `https://archives.fbi.gov/archives/news/speeches/combating-threats-in-the-cyber-world-outsmarting-terrorists-hackers-and-spies`, Federal Bureau of Investigation (FBI), Speech at RSA Cyber Security Conference, San Francisco, CA, 2012. [Online]. Available: `https://archives.fbi.gov/archives/news/speeches/combating-threats-in-the-cyber-world-outsmarting-terrorists-hackers-and-spies`.

[137] S. M. Istiaque, M. T. Tahmid, A. I. Khan, Z. A. Hassan, and S. Waheed, "State-of-the-art artificial intelligence based cyber defense model," in *2021 IEEE International Conference on Service Operations and Logistics, and Informatics (SOLI)*, 2021, pp. 1–6. DOI: `10.1109/SOLI54607.2021.9672393`.

[138] J. Park, "Qualitative versus quantitative research methods: Discovery or justification?" *Journal of Marketing Thought*, vol. 3, no. 1, pp. 1–8, 2016.

A

Interview protocols

Standard introduction

- Welcome and thank the interviewees for their participation.
- Provide an overview of the purpose of the interview: to understand the challenges related to in-house development and implementation of AI technologies in cybersecurity.
- Confirm consent for recording.
- Ensure the participant that all data will be anonymized.
- Ensure the participants there are no right or wrong answers and that they can omit any question if they feel uncomfortable.

Standard demographic questions

1. How old are you?
2. What is your highest degree of education?
3. How many years have you been working in your current position?
4. How many years have you been working in cybersecurity?
5. Tell me about your daily responsibilities in your role.

A.1. Interview questions for end users

AI tools properties

1. In your work, do you currently use any AI tools (AI tools = in-house developed models and not commercial models)?
   (a) If yes, go to section: Experience with AI tools for cybersecurity
   (b) If no, go to section: Specific questions for non-users of AI tools

Experience with AI tools for cybersecurity

1. What AI tools do you use or have used for your work tasks? What are the reasons? What do you like/dislike about them?
2. Can you walk me through how you integrate these AI tools into your workflow? How well do the AI tools integrate with other cybersecurity systems or processes?
3. Discuss the overall usefulness of AI technologies in your job. What problem do they solve for you? Have these AI tools helped to reduce the workload or created additional tasks or complexity?
4. Discuss the ease of use of these AI tools. How easy was it for you to become proficient with using them? What support did you receive during this phase?

5. Assess the reliability and trustworthiness of these AI tools. What influences your trust in these tools? What type of explainations or contextual information do you need?

Design and implementation process

1. Were you involved in the design process of any AI tools you use? If not, would you like to be involved? Describe the collaboration. What gaps have you observed?
2. Was there a trial or experimentation phase? How long did it last? How was it carried out? What was your experience with it?
3. How is user feedback collected and incorporated into the development of these AI tools? Do you have a collaboration platform?

Specific questions for non-users of AI tools

1. Which AI tools are you aware of and have access to but do not use? What are your main concerns or reasons for not using the available AI tools in your work? (usefulness, ease of use, integration, reliability, trust)
2. Discuss your knowledge and skills related to AI tools, and how this impacts your confidence in using them?
3. Would you like more exposure or training to AI in general? Would this make you more likely to start using such tools?
4. How important is trust in the tool when you decide to use or not use AI tools? Would greater transparency about how AI tools function and are developed affect your willingness to use them? To what extent do you trust the output of the models?
5. What changes or improvements in AI tools would make you more likely to adopt them?
6. How would being more involved and informed about the development of ML models affect your decision to use them?

Perceptions of AI tool adoption and work transformation

1. What do you perceive as the main factors influencing AI adoption rates among your colleagues? What barriers exist?
2. How do you view the role of AI in decision-making within cybersecurity tasks? How do you feel about AI making some selected decisions on its own? How do you feel about making critical decisions with AI assistance? Who is responsible when AI-related errors occur? How does that affect your confidence in using AI tools?
3. What are your thoughts on how AI could change job roles and job security within the cybersecurity field? How does that affect your decision to use or not use these technologies?

A.2. Interview questions for data scientists

1. How do you view the use of AI for cybersecurity tasks, specifically its benefits and challenges?
2. What are the primary technical and organizational challenges you face when developing and implementing AI models? How do you address these challenges?
3. How are new AI projects initiated? Who is involved in the design and implementation stages? How would you describe your collaboration with other teams during the development process? (challenges, successes, examples) What are some gaps and things that could be improved? What platforms do you use to collaborate? Optional: How do you ensure AI tools are user-friendly for end-users?
4. How do you ensure AI tools integrate with existing cybersecurity systems and workflows, and how do you handle compatibility issues with legacy systems?
5. How do you involve end-users in the development process to ensure the AI tools meet their needs? What feedback mechanisms are in place to gather and incorporate user feedback?

6. From your perspective, what factors influence the adoption rates of AI tools among end-users? What barriers to adoption have you observed, and how are these addressed?
7. Where do you see the future of AI in cybersecurity heading? What changes do you anticipate in job roles/responsibilities and job security due to AI?
8. What is your view on trust in AI? What mechanisms are in place?
9. What is your opinion about AI-assisted decision-making? What mechanisms ensure accountability when AI-related errors occur?

A.3. Interview questions for leaders

1. How do you view the use of AI for cybersecurity tasks, specifically its benefits and challenges?
2. How do new AI innovations typically come to your attention? Who proposes such initiatives?
3. What criteria or factors do you consider when transitioning from a use case proposal to a final adoption decision? Which stakeholders are involved in this process and how?
4. How is information about potential and adopted AI innovations communicated within the organization? What formal mechanisms are in place for collecting and incorporating feedback throughout this process? Do you use any collaborative platforms?
5. What factors influence the adoption rates of AI tools (among your team)? What barriers or forms of resistance have you noticed, and how are these addressed?
6. Where do you see the future of AI in cybersecurity heading in the next few years? How do you anticipate AI will change job responsibilities and job security within the cybersecurity field?
7. What is your opinion about AI-assisted decision-making (i.e., making decisions with AI)? What mechanisms are in place to address accountability and liability when AI-related errors occur?