UPPSALA UNIVERSITET

UPTEC STS 26027

Examensarbete 30 hp

Juni 2026

The Adoption of Autonomous AI Agents in Cybersecurity

A Qualitative Case Study of Early-Stage Adoption in a Security Organisation

Elias Bergman

Civilingenjörsprogrammet i System i Teknik och Samhälle

UPPSALA UNIVERSITET

The Adoption of Autonomous AI Agents in Cybersecurity

Elias Bergman

Abstract

The growing use of artificial intelligence in cybersecurity represents a significant shift in how security work can be organised and performed. As cyber threats become faster, more complex and increasingly difficult to manage manually, autonomous AI agents are emerging as a possible way to support detection, analysis and response. However, adopting such systems raises more than technical questions, because cybersecurity work also considers questions such as reliability, traceability, accountability and control.

This thesis investigates the early-stage adoption of autonomous AI agents within cybersecurity organisations. It aims to understand the most important factors influencing adoption across technological characteristics, organisational conditions and the external environment. Drawing on a combined Diffusion of Innovations and Technology-Organisation-Environment framework, the study takes a socio-technical perspective. A qualitative single case study was conducted at Atea Sweden, based on semi-structured interviews with respondents from the strategic, tactical and operational levels of the security organisation.

The findings show that adoption is shaped by a tension between acceleration and restraint, in which the qualities that make autonomous agents attractive are also those that raise the most caution. The agents are valued for handling growing data volumes, faster threats and complex technical work, especially in operational tasks. The adoption however, is limited by concerns regarding complexity, transparency, data quality, reliability and accountability. The study concludes that the main question is not whether AI should be used, but how much autonomy can be trusted. In particular, data control, transparency and accountability emerge as central conditions across all three contexts.

Teknisk-naturvetenskapliga fakulteten

Uppsala universitet, Utgivningsort Uppsala

Handledare: Elin Carlsson Ämnesgranskare: Göran Lindström

Examinator: Elísabet Andrésdóttir

Populärvetenskaplig sammanfattning

Artificiell intelligens, ofta förkortat AI, har på kort tid blivit en viktig del av diskussionen om hur framtidens arbete kommer att se ut. AI används redan idag för att analysera stora mängder information, ge beslutsstöd och automatisera uppgifter som tidigare krävde mänskligt arbete. Ett delområde inom AI som fått allt större uppmärksamhet är autonoma AI agenter. Enklare AI-verktyg svarar främst på en fråga eller skapar ett resultat när man ber om det och sedan är det människan som avgör hur resultatet ska användas. Autonoma AI agenter fungerar annorlunda: de kan arbeta mot ett mål på egen hand, planera flera steg, hämta information och utföra uppgifter med mindre mänsklig inblandning.

Inom cybersäkerhet är denna utveckling särskilt intressant. Cybersäkerhetsarbete handlar om att skydda system, information och organisationer mot digitala hot. Samtidigt blir hoten allt snabbare, mer avancerade och svårare att hantera manuellt. Säkerhetsorganisationer behöver analysera stora mängder information, varningar, sårbarheter och tekniska signaler. I sådana miljöer kan AI skapa värde genom att hjälpa människor med denna analys, detta genom att upptäcka avvikande beteenden och reagera snabbare mot hot. Autonoma AI-agenter skulle kunna ta detta ett steg längre genom att mer aktivt stödja övervakning, analys och andra säkerhetsuppgifter.

Samtidigt är cybersäkerhet ett område där misstag kan få stora konsekvenser. Det räcker därför inte att en teknik verkar effektiv eller snabb, den måste också vara pålitlig, möjlig att förstå och gå att kontrollera. När en AI-agent får mer självständighet uppstår viktiga frågor: Går det att lita på systemets bedömningar? Går det att förstå hur systemet har kommit fram till ett beslut? Och vem bär ansvaret om något går fel? Dessa frågor gör att införandet av autonoma AI agenter inte bara blir en teknisk fråga, utan också en organisatorisk och juridisk.

Syftet med denna studie är att undersöka vilka faktorer som påverkar den tidiga adoptionen av autonoma AI agenter i cybersäkerhetsorganisationer. Studien har genomförts som en kvalitativ fallstudie hos Atea Sverige. Genom semistrukturerade intervjuer med personer från olika delar av säkerhetsorganisationen har studien undersökt hur autonoma AI agenter uppfattas ur strategiska, taktiska och operativa perspektiv. För att analysera materialet används ett kombinerat ramverk baserat på Diffusion of Innovations och

Technology-Organisation-Environment. Det innebär att studien betraktar tekniken i sig, organisationens interna förutsättningar och den externa miljö som organisationen verkar i.

Resultatet visar att adoptionen präglas av en tydlig spänning mellan möjligheter och försiktighet, där just det som gör autonoma AI agenter attraktiva också är det som väcker mest tveksamhet. Agenterna ses som värdefulla eftersom de kan hjälpa säkerhetsorganisationer att hantera hotaktörer som också använder AI, ökande datamängder och mer komplexa tekniska uppgifter. Detta är särskilt tydligt i det operativa säkerhetsarbetet, där tid, skala och kontinuerlig övervakning är viktigt. Samtidigt begränsas adoptionen av frågor om komplexitet, bristande transparens, datakvalitet, tillförlitlighet och ansvar. Respondenterna är alltså inte negativa till AI i sig, men de är försiktiga med att ge AI-system för mycket självständighet innan frågor om tillit, kontroll och ansvar är tydligt etablerade. För att nå dit behöver organisationen rätt kompetens, tillgång till relevant och kontrollerad data samt gemensamma riktlinjer för hur AI ska användas. Ansvaret för säkerhetsarbetet kan nämligen inte föras över på tekniken även om vissa uppgifter kan automatiseras.

Sammanfattningsvis visar studien att frågan inte främst är om AI ska användas inom cybersäkerhet, utan hur mycket självständighet som organisationer är beredda att ge AI-system i olika säkerhetsuppgifter. Autonoma AI agenter kan skapa stort värde, men bara om organisationer först bygger de tekniska och organisatoriska förutsättningarna som krävs för tillit. Det handlar därför om att gå stegvis från AI-verktyg som stödjer människan mot mer självständiga lösningar, och att se till att kontroll, transparens och ansvar byggs ut i takt med att tekniken får agera mer på egen hand.

Acknowledgements

This master's thesis was carried out during the spring semester of 2026 on behalf of Atea Sweden as a part of the master's programme in Sociotechnical Systems Engineering at Uppsala University.

I would like to express my sincere gratitude to Atea Sweden for the opportunity to carry out this thesis within the organisation and to all the employees who took the time to participate in the interviews. Their openness and willingness to share their insights and experiences made this study possible.

I would also like to thank my supervisor at Atea, Elin Carlsson, for the valuable guidance, support and engagement throughout the project. Her knowledge and feedback have been of great help during the work.

Finally, I would like to thank my subject reviewer at Uppsala University, Göran Lindström, for the constructive feedback and guidance during this thesis.

Uppsala, June 2026
Elias Bergman

Glossary & list of abbreviations

Anomaly-based detection - Detection that established a baseline of normal behavior and flags deviations that may indicate compromise

APT - Advanced Persistent Threat. An attacker that stays hidden in a system over a long period.

Autonomous AI agent - An Artificial Intelligence system that perceives its environment and acts within it to achieve defined goals, with some independence and less human involvement.

Black-box model - A system whose processes are hard to see, making it difficult to know why it gives certain answers.

CTI - Cyber Threat Intelligence. Information about threats and attackers used to support security work.

Digital sovereignty - Keeping control over where data is stored and who handles it.

DOI - Diffusion of Innovations. A theory explaining why and how quickly new technologies are adopted.

Human-in-the-loop - A setup where a human checks, approves or can overrule what an automated system decides.

IAM - Identity and Access Management. Controlling which users and systems have which accounts and permissions.

Legacy system - Older technology that is still in use but hard to replace or connect to newer tools.

Machine learning - Systems that find patterns in data to make predictions, decisions or sort information.

Penetration testing - Attacking systems in a controlled way to find weaknesses.

PKI - Public Key Infrastructure. The systems and certificates used to verify identity and create trust in digital communication.

SaaS - Software-as-a-Service. Software delivered over the internet rather than installed locally.

Signature-based detection - Detection that spots threats by matching them against known attack patterns.

SOC - Security Operations Centre. A function that monitors and responds to security threats.

TOE - Technology-Organisation-Environment. A framework examining adoption through technological, organisational and environmental factors.

Zero-day exploit - An attack on a weakness that is not yet known, before a fix exists.

1 Introduction

Organisations operate in an environment of continuous technological change, where new tools and systems appear with the promise of doing existing work better, faster or more efficiently. Yet, the adoption of a new technology is rarely only a technical matter. Introducing new systems often reshapes the way the organisation functions, affecting already established routines, how decisions are made and governed as well as the competences employees are expected to have (Rogers, 2003). Because of this, the decision to adopt is shaped both by what the technology can do and whether the organisation is ready and willing to take in the changes that come with it (Tornatzky & Fleischer, 1990). Understanding why some technologies are adopted while others are not has therefore become a central question in research on organisational and technological development (Oliveira & Martins, 2011).

Earlier technological shifts, such as cloud computing, illustrate this well. Cloud computing allowed organisations to access computing resources such as storage, processing power and software over the internet, rather than running everything on their own local infrastructure (Mell & Grance, 2011). This offered clear benefits in the form of flexibility, scalability and reduced cost, but the adoption of cloud services was not driven by these advantages alone. Organisations had to consider questions of readiness, security, control over data and responsibility for systems that were no longer fully managed in-house (Low, Chen & Wu, 2011). This shows a pattern that tends to repeat across technological shifts, that adoption is shaped both by the benefits a technology promises and by the organisational concerns it raises.

Artificial intelligence (AI) represents the current major technological shift that organisations are facing. AI is increasingly being used to support analysis, inform decision-making and automate tasks that previously required human effort (Kaplan & Haenlein, 2019). This development has accelerated over the past few years, with AI spreading quickly across research, industry and a wide range of organisational settings (Hajkowicz et al., 2023). A more recent step in this development is the rise of autonomous AI agents, also known as agentic AI. While earlier AI applications mainly produced output in response to specific requests, leaving humans in control of how that output was used, autonomous agents instead

work toward defined goals with some degree of independence. This means that they carry out tasks with reduced human involvement (Yao et al., 2022).

Cybersecurity is one area where these developments are particularly significant. Security organisations must operate in environments characterised by growing data volumes, accelerating threat development and increasing system complexity (Rawat et al., 2019). As cyber threats evolve faster than traditional, manual security practices can respond, the speed at which organisations can detect and act on threats becomes critical. The consequences of falling behind are illustrated by the cyberattack on the system supplier Miljödata, where attackers exploited vulnerabilities to gain access to sensitive information (SVT Nyheter, 2024), showing how gaps can be exploited when threats develop faster than defences.

To keep pace with this development, AI has been positioned in both academic literature and industry discussion as an important enabler for handling the growing demands placed on security work. This by supporting advancements in threat detection, predictive analysis and automated response that traditional methods struggle to achieve (Jada & Mayayise, 2024; Mohamed, 2025). AI-based techniques are commonly presented as essential for handling large volumes of security data, detecting previously unseen attacks and supporting other responses in security operations (Chen et al., 2020; Zhang et al., 2022). As a result, many modern cybersecurity frameworks and strategic guidelines expect organisations to use these techniques to strengthen their security practices (Chen et al., 2020; ENISA, 2023; Zhang et al., 2022). Autonomous AI agents may extend this even further by taking a more active part in monitoring environments, analysing signals or supporting security tasks on a more continuous basis.

At the same time, cybersecurity is a demanding context for adoption. Security work depends heavily on reliability, traceability, accountability and control, making the loss of human oversight sensitive. The more independently a system is expected to act, the more important it becomes to understand how it reaches conclusions and who remains responsible when something goes wrong (Wallace et al., 2020; Kaur et al., 2025). Questions of this kind have already drawn attention in previous research which has examined AI adoption in broader organisational settings (Horani et al., 2025) and the adoption of cybersecurity technologies more generally (Wallace et al., 2020; Kaur et al., 2025). Although AI is seen as a technology

that could reshape security work, less is known about what shapes the early-stage adoption of autonomous AI agents within security organisations.

1.1 Purpose and research questions

The purpose of this study is to identify and analyse the most important factors influencing the early-stage adoption of autonomous AI agents in cybersecurity organisations. With this purpose in mind, the study is structured around three research questions:

1. How do the technological characteristics shape the adoption of autonomous AI agents in cybersecurity work?
2. How do internal organisational factors influence this adoption?
3. How does the external environment affect the adoption?

1.2 Research setting

This study is carried out at Atea Sweden, the Swedish part of one of the largest IT infrastructure and services providers in the Nordics. The company has around 2700 employees and delivers a wide range of IT and security services to organisations both in the private and public sector (Atea, 2025). Within the company, security work is organised across information security, IT security and cyber security. Because of this broad structure, the company makes it possible to capture different perspectives within a single organisational context. Atea is also actively exploring the role of AI in its own security work, which makes it well suited for a study focused on early-stage adoption.

1.3 Delimitations

Two delimitations should be clarified beforehand. First, the study focuses on the early-stage adoption of autonomous AI agents. Post-adoption processes such as long-term implementation and integration are therefore outside its scope. Second, the study does not focus on the adoption of autonomous agents within one specific security function. Instead, it considers the adoption across cybersecurity work more broadly, drawing on perspectives from different parts of the security organisations.

2 Background

The adoption of autonomous AI agents in cybersecurity is hard to understand without first establishing the context in which this adoption takes place. Cybersecurity is a demanding setting, where the work is shaped by the nature of the threats it deals with, the way security organisations are structured and the regulations they have to follow. This background chapter therefore describes the specifics needed to understand the rest of the study. It begins with artificial intelligence and autonomous AI agents as the technology in focus, before turning to how security organisations are structured and how AI is used within cybersecurity. Finally, it covers the regulatory context, since security work in Europe is affected by legal frameworks that govern personal data, artificial intelligence and cybersecurity.

2.1 Artificial intelligence

Artificial intelligence (AI) is a broad concept that has received growing attention in recent years, particularly as AI applications have spread across research, industry and public debate (Hajkowicz et al., 2023). The concept refers to systems capable of performing tasks that historically required human cognition (Russell, 2010). These cognitive functions include learning, reasoning, problem-solving, perception and language understanding. Rather than relying on static, hand-coded instructions to complete a task, modern AI systems can interpret external data, use algorithms to learn from it and lastly apply what they learn to produce the results (Kaplan & Haenlein, 2019).

AI is not one single technology, but a wider field that contains several approaches and subfields. One of the most important is machine learning, where systems identify patterns in data and use them to make classifications, predictions or decisions. A more advanced subfield of machine learning is deep learning, which uses multi-layered neural networks to process complex data. Deep learning has contributed to major advances in areas such as image recognition, speech processing and language-based applications (LeCun, Bengio & Hinton, 2015). Another central area in the field of AI is natural language processing (NLP), which focuses on enabling computers to analyse, interpret and generate human language. NLP forms the basis for technologies such as chatbots, translation systems and large language models (Eisenstein, 2019).

These developments have expanded the role of AI from systems that primarily classify or analyse information to systems that can generate text, support reasoning and interact with users in more flexible ways. In recent years, these developments have also contributed to the growing interest in AI agents, where AI is not only used to produce outputs on request, but to carry out tasks with some level of autonomy (Yao et al., 2022).

2.2 Autonomous artificial intelligence agents

The concept of an AI agent is not completely something new within AI research. In general, an agent can be understood as a system that perceives its environment and acts within that environment in order to achieve certain goals. Wooldridge & Jennings (1995) describe intelligent agents through properties such as autonomy, reactivity, proactiveness and social ability. Autonomy refers to the ability to operate without constant human direction, reactivity to responding to changes in the environment and proactiveness to taking initiative in pursuit of goals. Lastly social ability is being referred to as interacting with other systems or users. These ideas provide an important foundation for understanding more recent discussions of autonomous AI agents and agentic AI.

Recent advances in generative AI, especially large language models, have brought up agent-based systems again. Newer AI agents can go beyond responding to individual prompts by planning steps, using external tools or information sources and adjusting their actions as a task develops (Yao et al., 2022). In this study, autonomous AI agents refer to AI-based systems that can work toward a defined objective with some degree of independence, although the level of human involvement may differ.

2.3 Security organisations

Before turning to how AI is used within cybersecurity, it is useful to understand how security work is typically organised, since the role and impact of these technologies differ depending on where in the organisation they are applied. Security work is broad and reaches well beyond purely technical IT functions, combining technical, legal, organisational and human competencies (Englund & Tullin, 2020).

Security organisations are commonly structured into three organisational levels: Strategic, tactical and operational. This structure helps organisations manage cyber risk effectively by

separating decision-making, planning and execution, while ensuring that information flows between all levels. Within this structure, each level has a different purpose, audience and time horizon, but all three are necessary for a mature and effective security function (ISO/IEC 27001, 2022; ISO/IEC 27002, 2022). In a Swedish context, these same three levels are commonly addressed through the terms information security, IT security and cybersecurity (Englund & Tullin, 2020), which in this study are understood as the strategic, tactical and operational levels respectively.

2.3.1 Strategic level

The strategic level focuses on long-term security direction and business risk management. Its primary goal is to ensure that security activities align with organisational objectives and protect critical business assets. The work is not about specific technical measures, but about setting the organisation's overall security direction and risk appetite over a longer time horizon. Responsibility at this level typically lies with executive management and senior security leadership, such as Chief Information Security Officers (CISOs) and risk committees (ISO/IEC 27001, 2022). In the Swedish context, this strategic level is the one referred to as information security. It is strategic and preventative in nature and includes areas such as risk management, business continuity, information management and the legal and regulatory side of security (*see Figure 1)*.

Information security standards explicitly define security as management-driven and risk-based activity. ISO/IEC 27001 (2022) requires organisations to understand their internal and external context, assess risks and define security objectives aligned with business goals.
The outputs at this level, such as security policies, governance structures and strategic priorities, set the direction that the tactical and operational levels then work from. At the same time, strategic security competence is often described as less developed than technical expertise, with leadership and management of security at this level as an area needing further attention (Englund & Tullin, 2020). This points to the strategic level being a practical condition for effective security work.

2.3.2 Tactical level

The tactical level translates strategic objectives into security designs, controls and defensive plans. The level works over a medium-term horizon and acts as a bridge between strategy and

execution. This level is typically managed by IT security teams, security architects and security managers, with the focus on determining how the organisation should defend itself against relevant threats (NIST, 2020). In the Swedish context, this tactical level is the one referred to as IT security. It focuses on building and maintaining secure technical environments and includes areas such as identity and access management, network security and the security of hardware and operational technology (*see Figure 1).*

The difference between designing security controls, which is within the tactical layer, and operating them is reflected in NIST SP 800-53. It does this by separating the selection of controls from the operational work of monitoring and responding (NIST, 2020). To make these design decisions, the tactical level needs to understand how attackers operate, particularly their tactics, techniques and procedures (TTPs). A common example is the MITRE ATT&CK framework, which documents how attackers behave and supports threat modeling, detection engineering and control validation. This framework is designed for defensive planning and analysis rather than real-time incident response (MITRE, 2023). The tactical level is therefore where the organisation decides which defences to put in place and how the environment should be structured, rather than where those defences are run day to day.

2.3.3 Operational level

The operational level is responsible for the day-to-day security operations and immediate threat response. It works over the shortest time horizon, often in real time. Activities at this level include monitoring security alerts, investigating suspicious activity, responding to incidents and conducting threat hunting and forensic analysis (NIST, 2012). In the Swedish context, this operational level is referred to as cybersecurity. It focuses on defending against active threats and includes areas such as in incident response, security operations, cyber threat intelligence and offensive security (*see Figure 1).*

Work at this level is highly technical and time-sensitive. It relies on concrete information such as indicators of compromise, malicious IP addresses or domains and malware samples to support real-time detection and response (NIST, 2012). Outputs at the operational level include alerts, incident reports, forensic findings and containment actions. These outputs can also later on inform tactical improvements and strategic decisions and reassessments. In this

way, the operational level both relies on the direction set above it and feeds insights back upward.

It is apparent that the strategic, tactical and operational levels are interdependent and form a continuous feedback loop. Strategic priorities guide tactical planning, tactical decisions shape the operational activities and the lessons learned from incidents feed back into both tactical improvements and strategic risk assessment. Such a layered approach to security is also recognised in ISO/IEC 27002 (2022), which highlights the need to collect and analyze strategic, tactical and operational intelligence to support decision-making throughout the organisation.

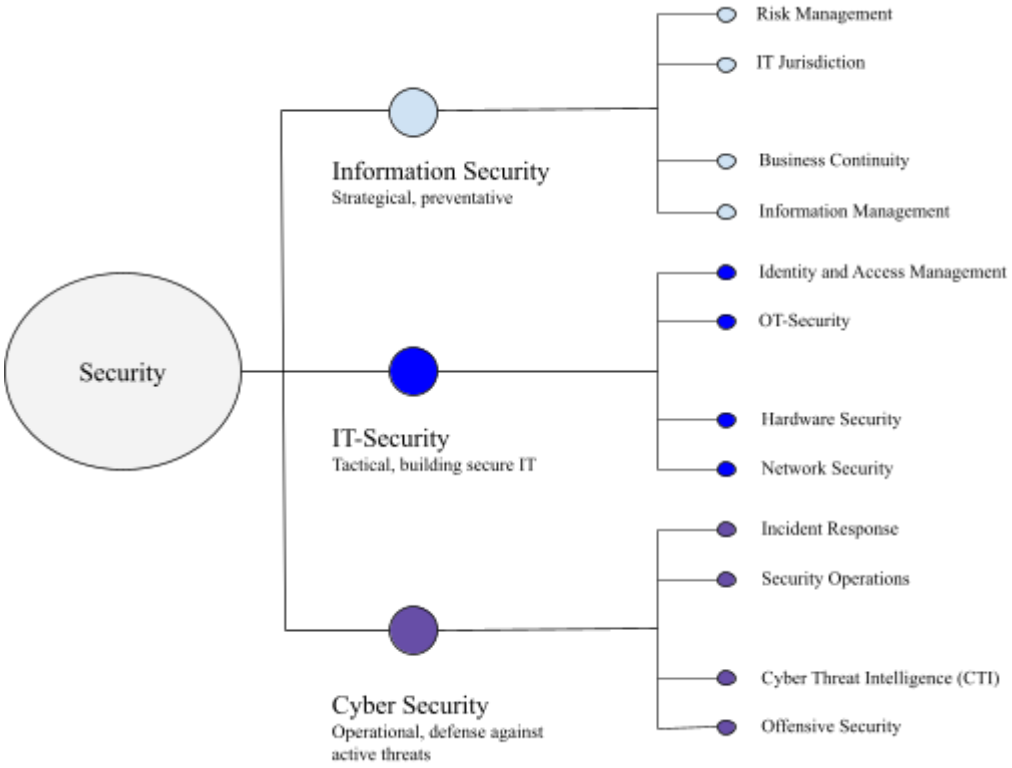

Figure 1: Example structure of a security organisation

2.4 Artificial intelligence in cybersecurity

AI has emerged as a key enabler for addressing the growing scale and complexity of cybersecurity threats. Historically, this development has been concentrated at the operational level, where the volume and speed of monitoring and response work make automation most valuable, although its effects increasingly reach the tactical and strategic levels as well. Cybersecurity work was originally reactive and focused on perimeter-based defenses, in

which security controls were designed to prevent known threats through static rules and predefined signatures (Scarfone & Mell, 2007; Garcia-Teodoro et al., 2009). Early intrusion detection systems and antivirus solutions relied heavily on signature-based detection, which required prior knowledge of an attack's characteristics. While effective against known threats, these approaches were limited in their ability to detect new or rapidly evolving attacks (Sommer & Paxson, 2010).

As cyber threats increase in sophistication, attackers have begun to exploit automation, obfuscation techniques and multi-stage attack chains, resulting in traditional rule-based defenses becoming increasingly insufficient. Zero-day exploits and advanced persistent threats (APTs) challenged conventional security models by operating without being discovered over extended periods and also while avoiding known signatures (Chen et al., 2020). This shift forced cybersecurity work to evolve from a preventative and reactive security work, toward one that emphasizes continuous monitoring, detection and response (Hutchins, Cloppert & Amin, 2011).

The introduction of behavioural and anomaly-based detection marked an important transition in cybersecurity practice. Rather than searching for known malicious patterns, security systems began establishing baselines of normal behaviour and identifying deviations that may indicate compromise (Sommer & Paxson, 2010). This change was felt most directly at the operational level, where analysts had to investigate the deviations these systems produced. However, early anomaly-based systems were often hindered by high false-positive rates and limited scalability, which placed significant strain on security operation teams (Scarfone & Mell, 2007).

AI and machine learning techniques have enabled a further evolution of cybersecurity work by addressing many of these limitations. Machine learning models can process large volumes of diverse security data, correlate events across systems and continuously adapt as new data becomes available (Chen et al., 2020). This has transformed cybersecurity from a largely manual and alert-driven activity into a more automated, data-driven practice. Consequently resulting in human analysts being able to focus on interpretation, decision-making and strategic oversight rather than low-level event prioritization (Zhang et al., 2022).

Despite its potential, the use of AI in cybersecurity also introduces new challenges. AI systems are highly dependent on data quality, explicit and implicit assumptions about system and attacker behaviour, as well as continuous feedback. Poorly trained or poorly operated models may generate false positives, overlook critical threats or lead to an overreliance on automated decision-making (Sommer & Paxson, 2010; Zhang et al., 2022).

2.5 Regulatory context

Using autonomous AI agents in cybersecurity depends heavily on current laws and regulations. Organisations operating in Europe must relate to legal frameworks that govern personal data, artificial intelligence and cybersecurity more broadly. Three regulatory developments are especially relevant in this context: the General Data Protection Regulation, the EU Artificial Intelligence Act and the NIS2 Directive.

2.5.1 General Data Protection Regulation

The General Data Protection Regulation (GDPR) is the central EU framework for the protection of personal data. It applies when organisations process information that can be linked to an identifiable person. The regulation is built around several principles such as transparency, purpose limitation, data minimisation, integrity, confidentiality as well as accountability for how personal data is handled (European Parliament and Council, 2016). This means that organisations must be able to justify why personal data is processed, limit the processing to what is necessary and also to ensure that the information is handled securely.

2.5.2 The EU Artificial Intelligence Act

The EU Artificial Intelligence Act (EU AI Act) is the first actual legal framework specifically aimed at regulating artificial intelligence within the European Union. The regulation is built around a risk-based approach, where AI systems have different requirements depending on the level of risk they may create. Systems considered to pose unacceptable risks are not allowed, while high-risk systems are subject to more strict obligations related to risk management, documentation, transparency, human oversight, accuracy and cybersecurity (European Parliament and Council, 2024).

The AI act places attention on how AI systems are developed, monitored and controlled. In particular, the regulation emphasises that high-risk AI systems should be designed in a way

that allows effective human oversight, so that risks to health and safety can be reduced (European Parliament and Council, 2024). Although not every autonomous agent used in cybersecurity would be classified as high-risk, the AI Act shows that questions of transparency, control and accountability are becoming central in the governance of AI.

2.5.3 NIS2 Directive

The NIS2 Directive is an EU cybersecurity regulation aimed at strengthening the overall level of cyber resilience across the Union. It applies to a broader range of sectors and organisations than the previous NIS Directive and places stronger requirements on cybersecurity risk management, incident reporting and organisational responsibility for security (European Parliament and Council, 2022). Unlike GDPR and the AI Act, NIS2 is not specifically focused on personal data or artificial intelligence. Instead the directive provides a broader regulatory framework for how organisations should manage cybersecurity risks.

The core of the NIS2 Directive focuses on mandatory risk management and fast incident reporting. Covered organisations must implement strong security habits, which include managing supply chain risks, fixing system vulnerabilities and using encryption to protect information (European Parliament and Council, 2022). When a cyberattack occurs, companies must follow a strict timeline, starting with an initial early warning to authorities within 24 hours of noticing the incident, followed by a formal report within 72 hours (European Parliament and Council, 2022). The directive also makes company leaders personally accountable, meaning that top managers can face penalties if they fail to approve and oversee proper cybersecurity measures (European Parliament and Council, 2022). Beyond incident reporting, NIS2 also places stronger demands on how organisations document, monitor and demonstrate their security work. Organisations covered by the directive are expected to keep oversight of their systems and processes and to be able to show that proper cybersecurity measures are in place (European Parliament and Council, 2022). This means that traceability, documentation and ongoing control of security activities become central parts of how organisations are expected to work under the directive.

3 The building of the theoretical framework

This chapter develops the theoretical framework used to analyse the adoption of autonomous AI agents in cybersecurity. It begins by reviewing previous research on technology adoption, focusing on the Diffusion of Innovations (DOI) theory and the Technology-Organisation-Environment (TOE) framework and explains how these two perspectives have been used and combined in earlier studies. From this, a research gap is identified. The chapter then presents the combined DOI-TOE framework used in this study, where the two perspectives are brought together to capture both how the technology is perceived and how organisational and environmental conditions shape adoption.

3.1 Previous research on technology adoption

Technology adoption research commonly draws on theoretical frameworks to explain why organisations adopt and implement new technologies. Two widely used perspectives are Diffusion of Innovations (DOI) theory and the Technology-Organisation-Environment (TOE) framework (Oliveira & Martins, 2011) . While DOI focuses on the perceived characteristics of the innovation itself, the TOE framework provides a broader perspective by examining the organisational and external contexts in which a technology is introduced. Together, these frameworks offer complementary perspectives and are discussed in more detail below.

3.1.1 Diffusion of Innovations

The Diffusion of Innovations theory, established originally by Rogers (1962) is one of the founding pillars of adoption research. The theory has since been revised and expanded across several editions, and this study draws on the updated version presented by Rogers (2003). The theory explains that the adoption rate of an innovation is determined by the potential adopters perceptions of five specific attributes: relative advantage, compatibility, complexity, trialability and observability. These attributes help explain the reasons why individuals choose to accept or reject innovation.

Previous research has shown that DOI can be useful for explaining technology adoption, especially when combined with complementary theories. For example, Al-Rahmi et al. (2019) integrated DOI with Technology Acceptance Model (TAM) to study students' intention to use e-learning systems. The researchers found out that the features compatibility

and relative advantage have significant positive impact on adoption. Their findings suggest that when a tool is seen as better than current methods and fits well with existing work habits, users are much more likely to accept it. On the contrary, they identified complexity as a major barrier that makes users more hesitant to adopt new systems (Al-Rahmi et al., 2019).

Research specifically targeting the adoption of security technologies has further refined how DOI attributes can be understood in high-stakes contexts. Iles et al. (2017) investigated the adoption of national security technology and found that perceptions of the technology's characteristics and the users attitude toward the technology were the strongest predictors of their intention to adopt. Their study included several DOI-related characteristics such as perceived advantage, compatibility, ease of use, demonstrability and trialability. Although these characteristics could not all be analysed separately in their final model, the broader perception of the technology had a strong influence on both attitudes and adoption intention. This supports the relevance of DOI for understanding how security technologies are evaluated before adoption.

However, despite its strengths, researchers have noted that the DOI framework has limitations when applied to organisations. Zhu, Kraemer and Xu (2006) argue that while DOI focuses on the innovation itself, it often overlooks the impact of the external environment. This is also emphasized by Fichman (2004) who argues that IT innovation research becomes limited when it focuses too narrowly on the innovation and its perceived benefits, without considering the wider organisational and contextual conditions that shape adoption.
For this reason, DOI is valuable in this study, but needs to be complemented by a framework that captures these broader contexts.

3.1.2 Combining DOI with the TOE framework

Because DOI focuses mainly on how the innovation itself is perceived, it is often combined with broader organisational adoption frameworks. The Technology-Organisation-Environment (TOE) framework is especially useful in this regard, as it adds attention to the organisational and external conditions that surround adoption. Originally developed by Tornatzky & Fleischer (1990), the framework highlights three contextual dimensions: technological, organisational and environmental factors. Together, the two perspectives provide a more complete view. In this view the DOI helps explain how

adopters evaluate the technology, while TOE provides the necessary context for how the organisation and the external environment influence the decision to adopt (Oliveira & Martins, 2011; Chiu et al., 2017).

This combination has been used in several studies of organisational technology adoption. Martins et al. (2016), for example, combined DOI, TOE and institutional theory to study the diffusion of Software-as-a-Service in firms. They found that the intention to adopt was significantly influenced by relative advantage and complexity. They also discovered that broader organisational and environmental factors such as technological competence, top management support and normative pressure had an impact. Similarly, Horani et al. (2025) combine DOI and TOE to study organisational adoption of AI-based technologies, arguing that innovation characteristics need to be understood together with socio-organisational and environmental factors. Their findings show how AI adoption is positively influenced by relative advantage and compatibility, while complexity had a negative effect. Organisational factors such as top management support, availability of resources and strategic alignment were also deemed significant. Lastly, in the environmental context, competitive pressure as well as vendor support had a great impact.

3.1.3 Research gap

The combination of broader adoption perspectives has also been applied in cybersecurity research, although less extensively. Wallace et al. (2020) use the TOE framework to examine cybersecurity adoption decisions among IT leaders and argue that the traditional framework does not fully capture the specific conditions of this context. Their findings highlight that cybersecurity adoption is strongly shaped by external factors such as cyber threats and cyber-attacks, which they describe as forces that influence decisions to invest in new security measures. They also note that cybersecurity differs from more general forms of technology adoption, since organisations often respond to uncertain and evolving risks that are difficult to measure, rather than to clearly visible efficiency gains. More recently within cybersecurity but with regards to AI, Kaur et al. (2025) use TOE to map barriers to the adoption of AI-enhanced cybersecurity solutions. They identify several technological, organisational and environmental barriers, including integration complexity, lack of transparency and regulatory compliance.

While the studies mentioned above provide a strong foundation for understanding technology adoption, an important gap remains. Existing research has examined the adoption of general AI technologies and cybersecurity measures more broadly. However, less is known about the early-stage adoption of autonomous AI agents within cybersecurity organisations. This is particularly relevant because cybersecurity adoption is shaped by conditions that are different from many other technology contexts, including evolving external threats, high demands on reliability and a strong need for control and accountability (Wallace et al., 2020). At the same time, autonomous AI agents introduce specific concerns, as their complexity and limited transparency can make them harder to understand and supervise. Their autonomy also raises questions about how much independent action should be delegated to them (Hahn et al., 2026). There is therefore a need to examine how these characteristics interact and are shaped by organisational conditions and the wider cybersecurity environment. This study addresses that gap by using a combined DOI and TOE framework to explore the drivers and barriers surrounding the adoption of autonomous AI agents in cybersecurity.

3.2 Theoretical DOI-TOE Framework

By combining the TOE framework with the DOI theory, this study takes on a holistic view of the adoption process. The TOE framework provides the overall structure by examining adoption through three contexts: technological, organisational and environmental. DOI is integrated within the technological contexts, as its five innovation attributes offer a more detailed way of understanding how the technology itself is perceived by potential adopters. Because this study focuses on the early stages of transition and the drivers as well as barriers behind the adoption decisions, it focuses specifically on the perceived characteristics of the innovation. Other parts of DOI, such as post-adoption implementation or long-term confirmation, are therefore not considered, as the research is limited to the pre-adoption phase of the technology lifecycle.

3.2.1 Technological context

The technological context concerns both the characteristics of the innovation being considered and the existing technological conditions within the organisation (Tornatzky & Fleischer, 1990). As previously mentioned, the technological context is developed through Rogers (2003) DOI attributes: relative advantage, compatibility, complexity, trialability and observability.

The first attribute, relative advantage, refers to the extent to which a new technology is perceived as better than the solution it may replace (Rogers, 2003). In adoption research, this is often connected to expected improvements in efficiency, quality, speed or overall performance. Rogers (2003) also emphasizes that this advantage does not have to be purely technical, but may also concern economic benefits, convenience or other outcomes that adopters consider valuable. The greater the perceived relative advantage, the more likely an innovation is to be adopted (Martins et al., 2016; Horani et al., 2025).

Compatibility concerns the extent to which an innovation is perceived as consistent with existing values, needs, technical systems and ways of working (Rogers, 2003). Technologies that fit poorly with current infrastructures or routines are often more difficult to adopt, even when their potential benefits are recognised. Compatibility has also been found to be a significant positive factor for technology acceptance, Al-Rahmi et al. (2019) argued that users were more likely to adopt innovations that align with their existing practices. In the context of organisational technology adoption, compatibility was also seen especially relevant since it had to fit existing systems.

A third important attribute is complexity, which refers to the extent to which a technology is perceived as difficult to understand or use (Rogers, 2003). The more difficult a technology appears to be, the more likely it is to create hesitation or slow adoption. This is especially relevant in relation to AI, since such systems are often experienced as difficult to interpret both technically and practically. Taddeo, McCutcheon & Floridi (2019) argue that AI in cybersecurity is often associated with uncertainty regarding how decisions are produced, which may influence both trust and willingness to rely on the technology. High perceived complexity may therefore contribute to caution, particularly in environments where errors can have serious consequences.

Trialability refers to the extent to which an innovation can be tested on a limited basis before broader adoption takes place (Rogers, 2003). This can be particularly important when a technology is perceived as risky or uncertain, since experimentation allows for evaluation before making larger commitments. Banerjee et al. (2012) argue that trialability may become especially significant in high-risk adoption settings, such as cybersecurity. Lier et al. (2025) similarly argue the importance of trialability, but with specific regards to AI in cybersecurity.

They emphasize the value of iterative evaluation and early risk identification during implementation.

Observability concerns to the extent to which results and benefits of an innovation are visible to potential adopters (Rogers, 2003). When the effects of an innovation are easy to notice or communicate, potential adopters may find it easier to assess its value. Rogers also notes that some innovations are naturally more observable than others, which may affect how quickly their benefits become recognised. This is important for technologies where the outcome is not always visible, for instance in software, which might lead to slower adoption (Rogers, 2003).

3.2.2 Organisational context

The organisational context refers to the internal conditions that may influence whether a firm adopts a new technology. Within the TOE framework, this includes factors such as management structure, available resources, organisational capabilities and readiness to introduce and use new innovations (Tornatzky & Fleischer, 1990).

One commonly emphasised factor is top management support. Management can influence adoption by setting priorities, allocating resources and signalling that a technology is strategically important. In research on organisational AI adoption, Horani et al. (2025) found that top management support had a significant positive influence on the organisation's intention to adopt AI-based technologies. The authors therefore indicate that emerging and complex technologies are more likely to be considered seriously when leadership provides support for further development.

A second organisational factor is organisational readiness. The adoption of complex technologies often depends on whether the organisation has sufficient financial, technical and human resources to evaluate, implement and maintain them. Martins et al. (2016) identify technological competence as an important factor in organisational adoption, while Horani et al. (2025) find that resource availability also positively influences the adoption. Research on organisational AI readiness also stresses the long-term AI adoption not only of technical infrastructure, but also people, processes and data-related capabilities (Uren & Edwards, 2023).

A third factor concerns strategic alignment and governance. New technologies are not adopted in isolation, but are often evaluated in relation to the organisation's goals, priorities and direction. To make sure that this actually happens, there are governance structure guiding specifically how the adoption should take place. In previous AI adoption research, it has been found that strategic alignment had a significant positive effect on intention to adopt (Horani et al., 2025). Similarly, Uren & Edwards (2023) show that this strategic alignment between technical business functions benefit the adoption of AI.

3.2.3 Environmental context

The environmental context refers to the external conditions that may influence an organisation's adoption of new technologies. Within the TOE framework, this includes factors such as industry characteristics, competitive pressure, government regulation, suppliers and other external actors surrounding the organisation (Tornatzky & Fleischer, 1990). In previous adoption research, environmental factors have been shown to matter because organisations do not evaluate new technologies on their own, but also in relation to the broader setting they operate in (Oliveira & Martins, 2011; Chiu et al., 2017).

One environmental factor that has received attention in adoption research is external pressure from the surrounding field or market. Organisations may become more decisive to adopt a technology when it gains legitimacy within their sector, or when other actors such as customers create expectations that the technology should be adopted. Martins et al. (2016) found out in their study of Software-as-a-Service diffusion that normative pressure influenced adoption intention.

A second important factor is regulation and external standards. Regulatory requirements may shape adoption by setting both expectations as well as force how organisations should manage risk, information and accountability. In cybersecurity, Wallace et al. (2020) show that regulation, disclosure requirements and IT standards influence adoption decisions. The authors connect this to organisations having to ensure that new security measures align with external legal and professional expectations. Kaur et al. (2025) make a similar point in relation to AI-enhanced cybersecurity. They identified that regulatory compliance is one of the barriers that organisations must consider when adopting AI-based security solutions.

Suppliers and external technology providers are also something that is included in the environmental context. This is due to the fact that organisations may depend on vendors for access to new technologies, implementation and ongoing service, which in turn makes supplier relationships relevant for adoption. Wallace et al. (2020) show that cybersecurity decisions are influenced by supplier and third-party actors, especially where sensitive information or network access is involved. Kaur et al. (2025) similarly highlight external dependency and vendor-related concerns as important when organisations are discussing the adoption of AI-enhanced cybersecurity solutions.

Finally, in cybersecurity, the environmental context also includes the external threat landscape. Wallace et al. (2020) argue that one of the differences between cybersecurity and many other technology adoption areas is the need for cybersecurity organisations to respond to evolving and uncertain threats that are difficult to measure. Their findings show that cyber threats and cyber attacks act as important external forces shaping adoption decisions.

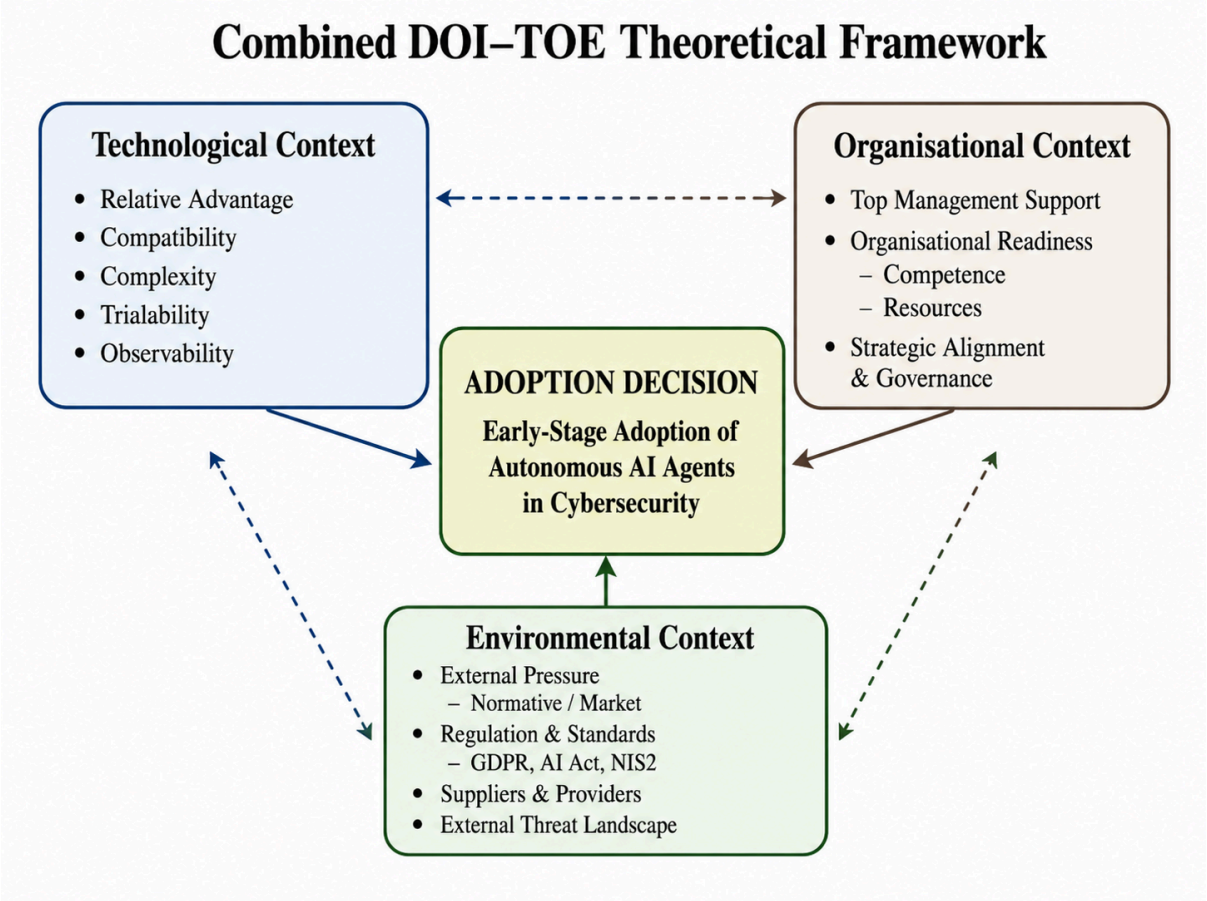

Figure 2: The synthesized theoretical framework.

4 Methodology

4.1 Research strategy

This section presents the overall research strategy of the study. It outlines the methodological choices that guided the research and helps clarify the basis of the methodological approach used throughout the study.

4.1.1 Philosophical foundations of the research

In social research, it is important to clarify the philosophical assumptions that guide a study. These assumptions influence how researchers understand social reality and how knowledge about such reality can be developed. Bryman (2016) highlights two central dimensions in this discussion: ontology and epistemology. Ontology concerns assumptions about the nature of social reality, while epistemology relates to how knowledge about this reality can be generated and interpreted. By acknowledging these perspectives, the study provides transparency regarding the assumptions that guide the research approach.

From an ontological perspective, this study assumes that organisational realities are shaped by the interpretations and actions of individuals within organisations. This view follows a constructionist perspective, where social phenomena, such as technological adoption, are understood as the outcomes of interactions between actors, technologies and organisational contexts (Saunders, 2009; Bryman, 2016). Epistemologically, the study takes an interpretivist approach, which focuses on understanding how individuals interpret and make sense of social processes. Since this research aims to explore how actors within cybersecurity organisations perceive and approach the adoption of AI and automation, an interpretivist perspective is suitable for capturing these experiences (Saunders, 2009; Bryman, 2016).

4.1.2 Qualitative research strategy

This study adopts a qualitative research strategy in order to examine the factors influencing the early-stage adoption of AI and automation within cybersecurity organisations. The research question focuses on understanding the conditions and drivers shaping this process, which requires insights into how individuals within organisations perceive and interpret technological change. A qualitative approach is therefore appropriate, as it allows for

exploring experiences, interpretations and organisational processes in greater depth (Bryman, 2016).

The research is conducted as a single case study focusing on the Atea organisation. Case study research is commonly used when the aim is to gain an in-depth understanding of complex phenomena within their real-life setting, especially in new and relatively under-researched areas (Njie & Asimiran, 2014). This was considered appropriate for the study, since the adoption of emerging technologies is shaped by the interaction between technological systems, organisational structures and the individuals working within them. Although a single case study limits the possibility of statistical generalization, it was still a suitable choice because the aim was not broad generalization, but a deeper understanding of how this transition is perceived and approached within a specific organisational context (Bryman, 2016; Yin, 2014).

4.1.3 Research approach

In social research, the relationship between theory and empirical observations is often described through deductive and inductive approaches (Bryman, 2016). A deductive approach typically starts from existing theories, from which hypotheses are derived and then tested through empirical observations. An inductive approach instead begins with empirical observations, from which theoretical insights are gradually developed (Bryman, 2016). While these approaches represent two ways of linking theory and empirical data, they do not always fully capture the iterative nature of many qualitative studies.

When studying complex and emerging phenomena, such as the adoption of new technology, both existing theory and empirical observations can play important roles during the research process. For this reason, this study follows an abductive research approach. Clark et al. (2021) argues that abduction allows the researcher to move back and forth between theory and empirical findings during the study. Rather than strictly testing theory or developing theory only from observations, the researcher continuously compares empirical insights with existing theoretical perspectives. This process makes it possible to refine and further develop theoretical understanding as new insights emerge from the empirical material (Clark et al., 2021; Dubois & Gadde, 2002).

In line with the abductive research approach, the study began with an exploratory phase in which several interviews were held with security professionals within the case company. This phase aimed to develop an initial understanding of the organisational context and the current use of AI and automation within cybersecurity practices. Insights from this phase also helped develop the theoretical framework and the interview guide used in the latter data collection phase. Consequently, the data collection was conducted through qualitative interviews with individuals working within the security organisation. These phases allowed the study to iteratively connect the theoretical perspectives with the insights gathered from the empirical material (Bryman, 2016).

4.2 Research process

Based on the philosophical foundations and research strategy established in the previous sections, it is necessary to operationalize these choices into a practical execution plan. This operationalization is defined through the research process, which shows how the theory and methods were put into action. To achieve this, the study was structured into two distinct stages: an exploratory phase and a data collection phase. This iterative two-phase approach was designed to explore and refine the understanding of the complex transition towards AI and automation within security organisations. Consequently to ensure a continuous dialogue between empirical observations and theoretical development.

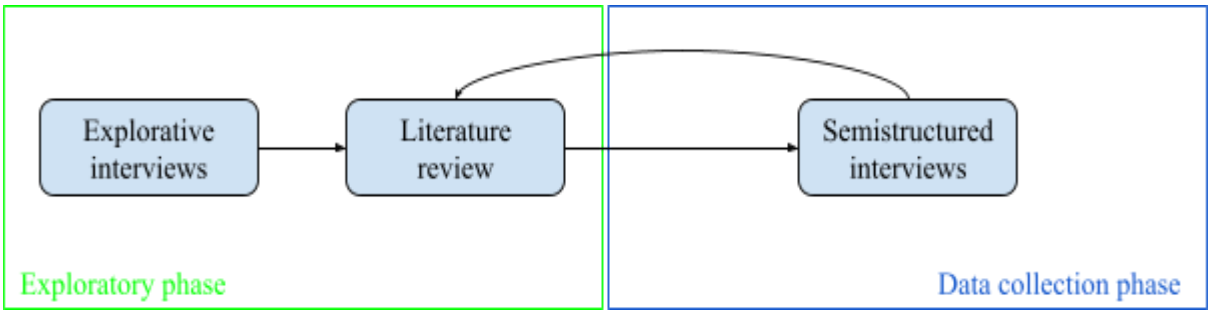

Figure 3: The research process, divided into the exploratory phase and the data collection phase.

4.2.1 Exploratory phase

The exploratory phase was designed as the initial stage of the research process, with the primary objective of gaining an understanding of the research problem as well as the organisational context at the case company. This stage followed an open ended approach to

map out the field and identify key themes related towards AI. To obtain the initial insights, unstructured and open interviews were conducted internally with stakeholders within the security organisation of the case company. These participants were selected because they offered a broad perspective on the organisation's security landscape as well as agentic AI . The reason for unstructured interviews is due to them being highly effective in early stage research as they provide the flexibility to explore viewpoints and experiences without the constraints of a fixed questionnaire (Bryman, 2016). This open format allowed for the identification of organisational challenges and helped refine the focus of the research by utilizing professional insights from various security roles.

Table 1: Compilation of exploratory interviews within the case company.

Respondent	Description	Department	Time
1	Concept Manager Cyber Security	Cyber Security	45 Min
2	Concept Manager Agentic AI and Intelligent Automation	Agentic AI and Automation	35 Min
3	Concept Manager Cyber Security	Cyber Security	45 Min

At the same time as the internal exploration, a systematic literature review was conducted to build a theoretical base for the study. This process was done in combination with the insights gathered from the explorative interviews, where key concepts and terms identified by the participants were used to guide the academic search. The search was done within the Uppsala University Library databases which include hubs like Web of Science, Scopus, DiVA and ScienceDirect. Focusing on scientific articles from these databases was important to ensure the use of high quality, peer reviewed research. Not only giving the research scientific weight but also allowing other researchers to trace the search process and verify the theoretical base (Webster & Watson, 2002). This approach ensured that the study remained connected both to current, real-world insights and existing research about technology adoption and cybersecurity. To make the search more complete, backward searching was also used. This

specific method involves checking the reference lists of relevant articles to find further important works and theories (Webster & Watson, 2002).

4.2.2 Data collection phase

Continuing after the exploratory phase, the main data collection was carried out through semistructured qualitative interviews with employees working in different parts of the security organisation within the case company. Semistructured interviews were considered suitable because they provide a balance between structure and flexibility. They make it possible to cover key topics related to the research question, while still allowing respondents to elaborate on issues they consider important and enabling follow up questions when relevant themes emerge during the conversation (Bryman, 2016; Kvale and Brinkmann, 2015). To support this process, an interview guide (*see Appendix A)* was developed and structured into thematic section based on the purpose of the study and the theoretical framework. This structure helped create consistency across interviews, while still allowing room for variation depending on the respondents role and experience. The interviews typically began with broader questions about the respondent's role and experience, before moving into more focus discussions related to AI, automation and adoption. Prior to each interview, participants were sent brief information about the purpose of the study, the interview process and the main themes that would be discussed. This was done to make the participants feel more comfortable and better prepared, which may support a more open and reflective conversation (Patel & Davidsson, 2019; Kvale & Brinkmann, 2015).

The interviews were held digitally through Microsoft Teams and were recorded for later transcription and analysis, with the consent of the participants. Since recording may affect how openly participants respond, the option to decline recording was also offered, in line with good qualitative interview practice (Jacobsen, 1993). Conducting the interviews online offered several practical advantages. It made it easier to schedule meetings with respondents working in different parts of the organisation and located in different cities across Sweden, while also reducing the time and practical constraints associated with in-person interviews. Digital interviews can therefore increase accessibility and make qualitative data collection more manageable, especially in organisational settings, such as within cybersecurity, where participants often have limited time (Archibald et al., 2019). At the same time, online interviews also have its limitations. Compared to conducting face-to-face interviews, they

make it harder to read non-verbal communication, establish the same level of personal presence and manage small technical disruptions during the conversation (Khan & MacEachen, 2022). Despite this, the efficiency and accessibility of the online format were deemed to outweigh these drawbacks.

4.3 Selection of interviewees

The selection of interviewees was carried out through a combination of purposive and snowball sampling. Purposive sampling was used in order to identify respondents who were considered relevant to the purpose of the study and who could contribute with knowledge about AI, automation and security work within the case company. This form of sampling is common in qualitative research when the aim is not statistical representation or to provide a fully representative picture of the organisation as a whole, but rather to gather rich and relevant perspectives (Bryman, 2016).

The intention was to gain a broad view of how the transition towards AI and automation is understood across different parts of the security organisation. For this reason, interviewees were selected from several functions within the organisation, including information security, IT security and cyber security. These areas were considered important because they represent different perspectives on security work but also since it made possible for stronger connections to different parts of the theoretical framework. Below are the different departments of the case company's security organisation, in *Table 2* a more specific description can be seen of each of the roles to the coherent department:

- Information security: Individuals with a strategic and preventative perspective.
- IT security: Individuals with a more technical and tactical perspective related to building secure systems.
- Cyber security: Experts on active threats and defensive work, offering more of an operational perspective.

In addition to the purposive selection, snowball sampling was also used during the interview process. At the end of each interview, participants were asked whether they knew other individuals within the organisation who could be relevant to interview. Snowball sampling is often used in qualitative research to identify additional participants through the networks of

initial respondents, particularly when the researchers seek access to knowledgeable individuals within a specific setting (Bryman, 2016). This was especially important in this study due to the lack of previous experience and limited network within the company. Furthermore this helped identify perspectives that may otherwise have been difficult to access.

Table 2: Overview of conducted interviews within the security organisation of Atea.

Respondent	Description	Department	Duration
1	IT Legal Counsel	Information Security	30 Min
2	Information Security Consultant	Information Security	40 Min
3	Information Security Consultant	Information Security	44 Min
4	Identity and Access Management (IAM) Solution Architect	IT Security	60 Min
5	Identity Management Solution Architect	IT Security	50 Min
6	Public Key Infrastructure (PKI) Consultant	IT Security	53 Min
7	Incident Responder	Cybersecurity	34 Min
8	Penetration Tester	Cybersecurity	44 Min
9	Senior Penetration Tester	Cybersecurity	43 Min

4.4 Qualitative data analysis using thematic analysis

The interviews were first transcribed in order to prepare the empirical material for analysis. Transcription was supported by digital transcription tools, after which the transcripts were

carefully reviewed and manually corrected to ensure that the content accurately reflected the recordings. Kvale & Brinkmann (2015) emphasizes the importance of having a reliable transcription of the interview material within qualitative research. The authors further acknowledge that the accuracy of the transcripts directly affects the quality and credibility of the following data analysis.

The transcribed material was then used as a basis for the thematic analysis. This method is widely used for identifying, analysing and interpreting patterns within qualitative data (Braun & Clarke, 2006). The approach provides a structured, yet flexible way of organising large amounts of qualitative material into themes that capture important aspects of the data in relation to the research question. Because thematic analysis is not tied to a specific theoretical framework, it can be applied within different research perspectives. This flexibility makes it suitable for studies such as this one, in which empirical findings are interpreted in relation to several theoretical perspectives (Braun & Clarke, 2006; Terry et al., 2017). These findings then help identify patterns in how actors understand adoption of new emerging technologies.

By following the six different phases outlined by Braun & Clarke (2006), the analysis was conducted. The first step involved becoming familiar with the empirical material through repeated readings of the interview transcripts in order to develop an overall understanding of the data. Once this familiarity had been established, initial codes were generated by identifying parts of the material relevant to the research questions. These codes helped structure the data and made it possible to identify recurring patterns across the interviews. Building on this, the third phase focused on examining the codes and grouping them into preliminary themes that reflected broader patterns of meaning within the empirical material.

From this point, attention shifted from coding individual sections of the transcriptions, to reviewing the themes and patterns in relation to the whole material (Terry et al., 2017). This fourth phase made it possible to assess whether the emerging themes were clear, relevant and supported by the data. Once the themes had been reviewed, they were further defined and named in order to make their meaning clearer. The final phase consisted of interpreting and presenting these themes in relation to the research questions and the theoretical framework of the study.

Throughout the analysis, movement took place between the interview material, the emerging codes and the developing themes. This iterative process allowed the analysis to move from descriptive patterns in the data towards more refined thematic understandings. In line with the abductive research approach, this process also involved continuous movement between the empirical observations and the theoretical perspectives guiding the study (Clark et al., 2021; Dubois & Gadde, 2002). This made it possible not only to develop the themes from the empirical material, but also to refine the theoretical framework as the analysis progressed.

4.5 Trustworthiness of the study

In qualitative research, the quality of a study is often discussed a bit differently than in quantitative research. Concepts such as reliability, replication and validity remain relevant, but they do not always fit interpretivist studies in a straightforward way. This is because such studies aim to understand meanings, interpretations and context-dependent experiences rather than to measure variables or produce statistically generalizable results (Bryman, 2016). For this reason, this study draws on the concept of trustworthiness, as developed by Lincoln & Guba (1988), which is often used to assess the quality of qualitative research through the criteria of credibility, transferability, dependability and confirmability.

Credibility concerns whether the study provides a credible and accurate understanding of the phenomenon being studied (Lincoln & Guba, 1988; Bryman, 2016). In this study, credibility was supported by conducting semi-structured interviews with respondents from different parts of the security organisation. This made it possible to capture variation in perspectives across strategic, tactical and operational roles. Transferability refers to the extent to which the findings may be relevant to other settings compared to the specific one in which the study was conducted (Lincoln & Guba, 1988). Since this thesis is based on a qualitative single case study, the aim is not statistical generalization. Instead, the ambition is to provide contextual insights into how AI and automation are perceived within a security organisation in the early stages of adoption. By describing the research setting, the selection of interviewees and the methodological choices as transparently as possible, the study makes it easier to understand the context in which the findings were produced. This gives readers the possibility to assess whether the findings may also be relevant in other organisational contexts with similar characteristics (Bryman, 2016).

Dependability concerns the consistency and transparency of the research process, while confirmability relates to whether the findings are grounded in the empirical material rather than shaped by the judgement of the researcher (Lincoln & Guba, 1988). In this study, both were supported by clearly documenting the research process, including both the exploratory and the data collecting-phase, the development of the theoretical framework, the interview guide and lastly also the steps of thematic analysis. Furthermore, the abductive approach also contributed to this transparency by making clear how theory and empirical findings were related throughout the study. Since qualitative analysis always involves interpretation, complete neutrality is difficult to achieve, making transparency in methodological and analytical choices important (Bryman, 2016).

4.6 Ethical considerations

Due to the nature of qualitative research, ethical considerations were kept in mind during the process. The four main principles outlined by Bryman (2016) were followed: Information, consent, utility, and confidentiality. Prior to each interview, every participant was fully informed that their participation was voluntary and that they could withdraw at any given point. Furthermore they were also informed about the study's methods and purpose. The respondents were assured that the collected data would be anonymized with a non-identifiable general description of the respondents role, and used strictly for academic purposes only (Bryman, 2016). This to ensure that participation would not result in any negative consequences for the respondents.

Beyond these principles, additional ethical considerations relevant to qualitative research were also taken into account. Arifin (2018) highlights the importance of responsible data handling and protection throughout the process. In this study, interview recordings and transcripts were stored securely and were only accessible to the supervisor within the organisation and the researcher. Furthermore, care was taken during the analysis and presentation of the findings to ensure that no sensitive organisational information or identifiable details about the participants were disclosed. These considerations helped ensure that ethical standards were maintained throughout both the data collection and reporting stages of the study (Arifin, 2018).

5 Empirical findings

This chapter presents the empirical findings from the interviews. The findings are then structured into three parts. The first describes how security work is carried out across the organisation and how AI has entered that work so far. The second presents the opportunities respondents associate with autonomous agents, covering the technical work, organisational value and external pressure. The third part presents the concerns that may limit further adoption, relating to areas such as reliability, accountability and legal-, regulatory- and supplier-related uncertainty.

5.1 Security work and the introduction of AI

The interviews show that security work within the case company is carried out differently depending on where in the organisation the respondent is positioned. At the same time, a shared pattern across the interviews is that AI has already entered the work in different forms, although mostly in supportive ways rather than through fully autonomous systems.

5.1.1 Strategic security work

The strategic part of the security organisation is represented by respondents working with information security, legal interpretation and broader governance-related issues. The IT legal counsel works in an area that places them between legal, regulatory and information security issues. Their work mainly concerns how rules, legal requirements and questions of responsibility relate to new technologies and organisational practice. The two information security consultants both focus on information protection, risk awareness and helping organisations work in a more structured, secure way. Their security work is tied to policies, routines and organisational support rather than hands-on technical defence.

In the strategic interviews, AI is often spoken about in broader organisational terms, although there is also some limited use of vendor tools such as Microsoft Copilot, that is generative AI, in support oriented tasks. Rather than being deeply built into everyday workflows, AI appears more often through advisory work, governance and broader questions about how new services can be used safely. One of the information security consultants describes how AI-related questions increasingly enter the work through clients asking whether new services can be used and what must be done before that happens: “*much AI-related questions are*

coming in now, can we use this service? what do we need to do before we do that?". At the same time, the consultant is clear that their work is not mainly centered around new technologies, explaining that information security work is much more organisational, physical and personnel-related.

A similar pattern appears across the rest of the strategic level. The other information consultant explains that they work *"very operationally with information security, but not close to the technical side"*, and instead focuses on standards, classification, risk analysis and regulatory matters. The IT legal counsel further emphasizes this by stating that *"AI tools are not tools I relate to directly; rather, I come into contact with technology through the requirements"*. Taken together, this suggests that AI has so far entered the strategic part of the organisation mainly through advisory work, legal and regulatory interpretation as well as broader organisational preparation.

5.1.2 Tactical security work

The tactical part of the security organisation is represented by respondents working with IT security, system design, identity, access, and secure technical environments. One of the solution architects works within identity and access management (IAM), meaning that the role involves designing solutions that ensure users and systems have the right accounts and permissions. In practice, this means working with how access is structured and controlled across technical environments. The other solution architect works in Microsoft cloud environments, with a strong focus on customer security workshops, technical environments and security design.

The Public Key Infrastructure (PKI) consultant is similarly involved in identity and access management, but more focused on PKI itself. PKI can be described as the systems and certificates used to verify identity and create trust in digital communication.

In this part of the organisation, AI has entered the work more directly. The IAM solution architect is described as someone who has always worked closely with new technologies, both in regards to previous emerging technological solutions but also to AI. They describe using AI mainly in development-related work, where it helps carry out technical tasks that would otherwise take more time or have to be delegated. Noting that: *"I can send away a task and then validate the result. I use it a lot to replace boring technical work that I do not have*

time to do myself", showing how AI has become part of the technical workflow itself. The other solution architect is also described in a similar way, as an early adopter of new technology, although the current use of AI is lighter. Rather than using it directly in technical development, the respondent mainly describes using AI in meetings, where it helps with transcription, summaries and note-taking.

The PKI consultant describes a highly technical day-to-day role, saying that: *"PKI is machine-, as well as human-identities… It's my area of expertise and I'm in there actively working on things in 60 to 80 percent of my everyday work"*. At the same time, AI does not appear to be as established in the PKI consultant's own workflow as for the other tactical respondents. The interview instead suggests that the role is close to the environments where these questions matter, even if AI has not yet entered daily work in the same immediate way. Altogether, the tactical interviews show that AI has entered this part of the organisation through practical technical support, especially in repetitive, bounded and support-oriented tasks, while the degree of actual use still differs between respondents.

5.1.3 Operational security work

The operational part of the security organisation consists of respondents working directly with active threats, testing and day-to-day cybersecurity tasks. The incident responder helps organisations handle cyberattacks when they have already happened. This includes stopping the incident, understanding how the attacker got in and lastly making sure that they are no longer still present in the environment. The penetration tester, on the contrary, works in offensive security, meaning that the role involves attacking systems in a controlled way in order to find weaknesses before real attackers do. Lastly, the senior penetration tester works as a team lead and consultant in offensive security, where the work includes penetration testing, technical offensive security and practical evaluation of systems, vulnerabilities and security controls.

In this group, AI is described as having entered everyday work already and there is a consensus that there has been a steady acceleration just the last couple of years. A common pattern across the interviews is that AI has also begun to function as a new kind of search engine in daily work. Rather than replacing their tasks, respondents describe using it to find solutions faster, check technical issues and support problem-solving in the middle of ongoing

work. This is also reflected in the senior penetration tester's description, where the respondent mentions that: *"the whole development process has been affected a lot by AI. It is so easy to generate code nowadays, it is almost as if you don't have to write anything by yourself anymore"*.

A direct and practical use of AI appears in the penetration tester's interview, especially through tools adapted for penetration testing. The penetration tester explains that recent advancements have made these tools more relevant to everyday work, stating that: *"We have seen automated solutions before that have been quite talentless in the sense that they have not been thinking by themselves. If they find vulnerability X, they do not automatically continue to find vulnerability Y. But now with better trained models, the solutions can definitely make a difference"*. Suggesting that AI, and more automated solutions has begun to enter the operational work.

5.2 Perceived opportunities with autonomous agents

While AI has so far entered security work mainly in supportive forms, the interviews also point to several reasons why respondents see value in moving toward more autonomous agents. These reasons appear at different levels, from the practical technical work to broader organisational value and pressure from the surrounding environment. The following sections present these as the opportunities that may push adoption forward.

5.2.1 Reasons pushing adoption forward in the technical work

One reason pushing the adoption of autonomous agents and more advanced AI solutions forward is that several respondents describe security work as increasingly difficult to handle manually, especially when large amounts of data, logs and technical signals need to be reviewed. This is particularly clear in the more tactical and operational interviews, where AI is described as useful in situations where the amount of material is simply too large for a person to work through efficiently. One of the IAM solution architects explains this clearly when talking about defensive work and log analysis, stating that: *"it is not reasonable for a human to sit and analyse log files anymore. That is really a task for AI"*. The point is further emphasized, in which another solution architect describes how autonomous agents in identity and access management would act as a force multiplier: *"if you are to guarantee what a user account has been used for… you suck in logs from all systems and look for abnormalities,*

and then an alarm goes off". Indicating the technology's ability to analyze huge amounts of data points and look for abnormalities is something not humanly possible.

Another reason mentioned in the interviews is that AI is seen as improving the ability to carry out more technically demanding work within the same role. This is not only described as a matter of doing existing tasks faster, but also as a way of complementing the person's own strengths and weaknesses. The incident responder, for example, explains that AI helps in areas where their own technical knowledge is more limited: *"I am not good at programming. It helps me extremely much to develop solutions to my problems"*. In this way, AI is described as helping the respondent move further in tasks that would otherwise have been harder to solve independently, especially when the work requires knowledge outside their main area of expertise.

Ultimately, the motivation for moving toward more autonomous AI systems appears to be rooted in the visible efficiency benefits that respondents are already experiencing in their daily work. These early successes serve as a proof of concept, demonstrating that the technology can handle the tough task of for instance security administration and technical analysis. One of the information security consultants highlights this by describing how current AI tools have already fundamentally changed the time needed for documentation work, stating that: *"Instead of it taking me a week and a half, full-time… I do it maybe in half a day instead, and that is quite a large efficiency gain"*.

5.2.2 Organisational value pushing adoption

The efficiency gains described previously are also connected to financial and organisational value. In the interviews, AI and autonomous agents are not only discussed as tools that can make the security work faster, but also as something that may reduce costs and make the use of resources more effective. The penetration tester makes this clear when discussing their own work, arguing that: *"the question of money is absolutely central… If an agent can perform a penetration test in one day instead of four weeks, you save massive amounts of money"*.

Several respondents also describe an internal willingness to try new tools when they are seen as useful in practice. The same penetration tester explains that the organisation is: *"quite*

receptive as long as it generates value and simplifies the work", and connects this directly to whether a tool can show practical or financial benefit. There is also a great positivity towards tools that make work easier, pointing to an internal pull from respondents who are curious about new technology and willing to test it when it appears relevant to their work.

A similar pattern appears across the operational level, where respondents describe ongoing experimentation with AI tools and agents rather than waiting only for formal implementation from above. This is not described as a lack of organisational support, but rather as a sign that adoption is partly driven from within the specialist groups themselves. The incident responder describes this as a form of internal curiosity, where: *"there are many groups starting around how we should create agents to do smart things"*. In the same way, the penetration tester explains that it is often the people working closest to the tools who need to evaluate what is useful, since those higher up may not always have detailed knowledge of the specific technologies. This gives the impression that adoption is pushed forward not only by formal decisions, but also by employees who test, discuss and adapt new tools in relation to their own work.

5.2.3 External pressure and the need to keep up

Beyond internal willingness as well as efficiency benefits, a major reason for automating parts of the security work with the help of AI is due to the external pressure from a rapidly shifting environment. This is described by several respondents as a: *"cyber arms race"*, in which the organisation must adopt AI simply to keep pace with the AI already utilized by threat actors. There also seems to be a shared view that both the quality and quantity of attacks have increased. This with regards to malware, vulnerabilities and phishing, with a respondent claiming that: *"there have been more threats, but also more qualitative threats"*. Another respondent summarizes this broader development by stating: *"Because the threat actors are doing it, i.e using AI, we have to match it… AI will be a force multiplier that helps us work faster"*.

This increased pace is also visible in how respondents describe vulnerability exploitation and zero-day attacks. The incident responder explains that a proof of concept for a vulnerability could previously take months to appear, while today it may be available within hours because of AI. An IAM solution architect also talks about this, but with regards to zero-day exploits,

where attackers manage to exploit a vulnerability almost immediately. They argue that organisations previously had more time to patch systems before this happened, but today, this window has become much shorter. Because of this, security work is described as needing to move beyond known signatures and fixed patterns, and instead focus more on behaviours that fall outside of what is normal. They state: *"Is this a behaviour that is outside the norm? Then it can be a threat that no one knows about. AI will help in the analytical phase since we can analyze so much more data and find these unknown threats"*. Across the interviews, this is also where autonomous AI agents are described as especially relevant: they could work continuously in the background, searching for unusual behaviours and weaknesses before they develop into larger security incidents.

A broader market development is also visible in the interviews, although it is described less directly than the pressure from threat actors. Respondents do not generally describe the reason for AI adoption as being driven mainly by fear of competitors. Instead, AI is described as something that is increasingly becoming part of the tools and services surrounding security work as well as any technological solutions nowadays. A penetration tester describes this by saying that AI is now being added to almost everything and compares it with previous emerging technologies: *"they put AI-driven on almost anything. Just like everything had to have blockchain a couple of years ago, it is AI now. Some tools, like vulnerability scanners are being marketed with them being driven by AI in a way that they can identify specific vulnerabilities better, but it still feels like a honeypot that just exists"*. In their opinion, AI in the market is therefore described with some caution, as something that is often used as a selling point even when it is not yet clear whether it makes the tool meaningfully better in practice. At the same time, the fact that AI is becoming part of the wider security market creates a need for the organisation to follow the development.

The market development is also connected to customer expectations, although this is described more as a gradual shift than as a direct competitive pressure. Respondents describe AI as something customers increasingly ask about, and as something security providers need to understand in order to remain relevant. AI can therefore be seen as part of how the organisation can create value towards its customers.

5.3 Concerns limiting further AI adoption within security

Although the interviews show several reasons for adoption of AI and autonomous agents in security work, respondents also describe a number of concerns that may slow down or limit further adoption. These concerns are especially clear when AI is discussed as something more autonomous, rather than as a support tool controlled by a human. Across the interviews, respondents return to questions such as trust, control, responsibility and technical readiness. The following sections present these concerns as factors that may make organisations more cautious when moving from limited AI support toward more autonomous security solutions.

5.3.1 Reliability, data and technical readiness

When discussing the implementation of autonomous agents, one common concern across the interviews is that the technology is still moving very quickly, while the ability to understand and control it does not always develop at the same pace. Several respondents describe a hesitation toward letting AI systems make decisions on their own, especially when the reasoning behind the output is unclear. One of the information security consultants describes this uncertainty by saying that: *"there is a certain caution when it comes to introducing AI, because it is still unknown ground. You do not really know what it may involve"*. A similar point is also made by the other information security consultant, who argues that: *"the technology often runs ahead of the human, and it goes so incredibly fast that we do not have time to reflect"*. In these interviews, the issue is not only that AI is new, but that its development creates uncertainty around when it is safe to rely on it.

The hesitation is closely connected to questions of reliability and transparency. Respondents describe autonomous AI agents as difficult to trust when it is unclear how a conclusion has been reached. The IT legal counsel explains that organisations: *"need to have insight into how the agent reaches conclusions and what it is actually concluding"*. The incident responder describes the same issue but more directly, stating that: *"we do not know what it does. Therefore we do not trust it"*. This becomes especially important when AI is expected to act with less human involvement, since the user must still be able to understand whether the system has made a reasonable decision. The senior penetration tester adds that transparency alone may not solve this problem, since: *"even if the model is transparent, it is such a complex system that someone also has to understand it. That is required in order to be able to take responsibility for it"*. In the interviews the respondents are not only concerned

with whether AI gives useful answers, but also with whether they can follow how those answers have been produced and lastly understand them.

Several respondents also connect this problem to the so-called black-box character of AI systems. The penetration tester describes the technology as difficult to fully understand, stating that: *"fundamentally, it is a black hole, and even researchers find it difficult to know why certain models behave the way they do"*. This lack of insight creates hesitation, especially in security work where wrong decisions can have serious consequences. One of the IAM solution architects therefore stresses the need for caution, whilst arguing that: *"we must maintain a healthy skepticism and always have control mechanisms in place. AI is a fantastic tool, but it is no replacement for human judgement and ethics"*. Rather than rejecting AI, the respondents describe a need for checks, validation and human control before it can be trusted in more autonomous forms.

Another recurring concern is the quality and control of data. Several respondents describe data as a basic condition for using AI safely. One of the information security consultants explains this clearly by saying: *"You need to have control over your data. What data do we have? Where is it? Is it classified? If you do not have that base ready, then you can't really adopt these new tools in a secure way"*. The IT legal counsel continues down a similar path, arguing that: *"if we can't manage information and data in our organisation as things stand today, then AI will not be a support tool that can be applied in a good way"*. They further warn that organisations may expose themselves to enormous risks if they implement AI without knowing: *"what is sensitive, what is worth protecting and what is important for the organisation"*. Across the interviews the concern is therefore not only whether AI tools work, but also whether the organisation has enough control over the information it will rely on.

Beyond the question of data, respondents also describe the surrounding technical environment as important. Some of them see existing systems and previous investments as a possible limitation. One of the IAM solution architects describes this as a legacy problem, stating that *"I often see organisations with a technological heritage, a 'legacy', that is difficult to get out of. It makes it hard to apply modern technology fully"*. Furthermore they add that: *"technical debt is one of the biggest roadblocks on the way toward autonomous systems"*. At the same time, not all respondents describe the technical environment as a major obstacle. Some point out that implementation may be easier when AI solutions are connected

to already established platforms and cloud-based services. This gives a divided picture of technical readiness, meaning that in some environments, existing platforms may make adoption easier, while in others, legacy systems, poor data structure and unclear information flows may slow it down.

5.3.2 Responsibility, control and organisational readiness

Another concern that is recurring in the interviews is that AI has the potential to automate parts of the security work, but that responsibility can't be moved over to the technology in the same way. This becomes especially important when respondents discuss autonomous agents and automated decisions. The senior penetration tester explains that: *"if you have an automated decision process, there must somewhere be a human responsibility at the foundation, someone who is responsible for the algorithm or the process being correct, and who also carries responsibility for the outcome"*. A similar view appears among almost all interviews, where respondents place the responsibility on the person using or interpreting the system. The incident responder summarizes this by stating: *"It is the analyst who makes a wrong assessment who becomes responsible"*, even if the analyst relies on a security product that is automated by AI. In this sense there is a consensus that AI may support the work, but with responsibility still described as something that remains within the security organisation.

This also creates caution around removing human control completely. Respondents do not reject automation, but several of them describe full autonomy as something that requires strong safeguards. The IT legal counsel explains that: *"if we are not going to have a human-in-the-loop, then we really need to be confident in the data it is going to work with, or have a very extensive understanding of the algorithms"*. Together, these views give the impression that fully autonomous systems are not completely rejected. Instead they do not see them as realistic in the present situation unless questions of data, responsibility and technical understanding are first handled.

Furthermore, the interviews reveal a significant gap in strategic vision and steering. The IT legal counsel argues that without a clear framework, AI adoption risks becoming uncoordinated and dangerous. They warn: *"I think it is important to have a steering, a rulebook or frameworks that describe what you want to achieve with AI… If we have not defined that, we risk creating a company or an organisation where it becomes the wild west"*.

Without this top-down guidance, they suggest that the technology will never reach its full potential, remaining limited to simple administrative tasks: *"Without steering or goals, the technology risks just becoming an 'amplified googling' or a chatbot that replaces the intranet. We do not make the big leap that the technology actually enables if we do not know how we are going to use it"*. This concern is also connected to how AI initiatives may become spread across separate parts of the organisation. The senior penetration tester describes silos as a possible barrier to further adoption, since they can make it harder to coordinate knowledge, tools and ways of working across different groups. One of the information security consultants makes a similar point when explaining that AI adoption is rarely owned by one person alone, and that different parts of the organisation need to come together in a joint effort.

A final organisational concern highlighted in the interviews is that the broader implementation of autonomous agents requires a dual understanding of both security work and the underlying AI systems. Several respondents express that it is not enough to simply judge whether an output looks correct. The user also needs to understand what information the AI has processed, how the results were produced and when the output should be questioned. Without this level of knowledge, respondents describe a greater hesitation toward letting agents act independently. Alongside this knowledge gap, established routines and ways of working are also identified as obstacles. One of the IAM solution architects connects this to the challenge of moving away from older systems, stating that: *"It is hard to get people to realize that they should have gotten rid of a system 15 years ago when they think it still works well"*. They also acknowledge that because existing routines are so deeply connected to these older systems, the transition to autonomous agents might be slowed by a preference for maintaining methods that have long been considered reliable.

5.3.3 Legal, regulatory and supplier-related uncertainty

The concern about responsibility and accountability also appears in the interviews as a legal and regulatory question. If autonomous agents make decisions or carry out actions in security work, several respondents describe uncertainty around who would be held accountable if something goes wrong. While the respondents generally place responsibility on the person or organisation using the system, this also creates hesitation toward allowing autonomous agents to act independently. Even if AI supports the decision, the legal responsibility may still

remain with the user, which makes the consequences of relying on an autonomous system more serious. The senior penetration tester describes this as a significant hurdle, noting that it is difficult for organisations to clearly define responsibility for processes that to a large extent lie outside your own walls, which naturally leads to increased caution.

Regulation is also discussed as something that affects how autonomous agents can be introduced, especially when personal data is involved. The IT legal counsel explains that legal requirements become clearer when AI systems process personal data: *"It is always important that we have a basic understanding in order to say that we have control over the technology, but it is put to the test when we know that the AI technology processes personal data"*. The IT legal counsel connects this to existing legislation and supervision, where organisations need to understand how the technology works and what information it handles. GDPR-related issues are therefore described as important because autonomous agents may need access to sensitive information, logs or internal environment in order to be useful.

At the same time, the interviews do not describe regulation as the main obstacle in every case. The IT legal counsel further explains that many of the current legal requirements are not impossible to handle, but that they require a more systematic approach before implementation. They state that the AI ACT mainly creates demands on organisations to become more competent in relation to AI, arguing that the biggest challenges often are connected to: *"information management, architecture and how to work with AI strategically"*. NIS2, which is aimed at elevating cybersecurity standards across critical sectors to meet growing digital threats, is also discussed among the respondents. In general, it is also talked about in a similar way to the AI Act. During the discussions it is made clear that this directive is not mainly about AI, unlike the AI Act , but more about security and the controls organisations need to have in place. This might still affect the implementation of AI indirectly, since autonomous agents would need to fit into existing requirements for control, documentation and security management. However, regulation is not generally described as something that stops adoption. It is instead described as something organisations need to relate to while AI develops at a faster pace than laws and formal guidance.

A more concrete uncertainty concerns suppliers and where data is handled, especially when autonomous agents are not built in-house. Since many AI tools are provided by large international vendors, respondents describe questions of data location, control and trust as

important. The IT legal counsel describes this through the issue of digital sovereignty, stating that organisations have become more sceptical toward American technology and are increasingly looking for European alternatives. The same respondent concludes it by saying: *"we have never talked as much about digital sovereignty as we do today"*, and connects this to a declining trust in how information is handled by major international providers. One of the penetration testers also describes a similar concern, explaining that much of the AI use in the west is based on hosted models, while local models may allow greater control over what information is sent away.

6 Analysis

This chapter analyses the factors that influence the early-stage adoption of autonomous AI agents within cybersecurity organisations. The analysis is structured through the combined DOI-TOE framework presented in Chapter 3. The technological context is examined through the five DOI attributes: relative advantage, compatibility, complexity, trialability and observability. The organisational context focuses on internal conditions such as readiness, governance, competence and responsibility. Lastly, the environmental context addresses external forces such as threat developments, market expectations, regulation and supplier-related uncertainty.

6.1 Technological conditions shaping adoption

The technological context is central for understanding why autonomous AI agents are considered attractive, but also why respondents remain cautious. Through the DOI perspective, the findings show that the technology is evaluated according to its perceived usefulness, fit, complexity, ability to be tested and the visibility of its outcomes. It is also apparent that these attributes do not work independently, but instead interact closely.

6.1.1 Relative advantage

The most apparent technological driver is the perceived relative advantage of autonomous AI agents. In line with Rogers (2003), the findings suggest that adoption becomes more likely when the technology is seen as offering clear improvements over existing ways of working. Across the interviews, these improvements are mainly connected to speed, analytical capacity and the ability to handle tasks that are increasingly difficult to manage manually. Respondents describe security work as involving large volumes of logs, signals, alerts and technical information, where AI agents can help identify patterns and abnormalities at a scale that would be unrealistic for humans alone.

This perceived advantage appears especially clear among respondents closer to the tactical and operational levels of security work. These respondents encounter the practical limits of manual analysis more directly, which may explain why they speak more concretely about AI as a force multiplier in log analysis, identity monitoring and penetration testing. The technology is not only viewed as making existing tasks faster, but also as extending what one

professional can realistically accomplish within a role. For example, respondents describe AI as useful for solving programming-related problems, supporting technical development and allowing more complex tasks to be handled without relying entirely on additional specialists.

This helps explain why autonomous agents become attractive in the first place. Their perceived advantage is not based only on AI being seen as a promising technology, but on its ability to address a growing gap between the demands of cybersecurity work and what human attention and analytical capacity can handle. Where security work is experienced as increasingly data-intensive and time-sensitive, autonomous agents appear valuable because they promise continuity, speed and analytical capacity. This supports earlier research that identifies relative advantage as an important factor in adoption decisions (Martins et al., 2016; Horani et al., 2025), but the findings add that the perceived advantage is closely tied to the inability of manual work to keep pace with the growing demands of cybersecurity.

6.1.2 Compatibility

A second important technological factor is compatibility. Rogers (2003) describes compatibility as the extent to which an innovation fits existing values, needs and practices. In the findings, compatibility appears in three closely connected forms: fit with the organisation’s data environment, fit with existing technical systems, and fit with the logic of security work itself.

The issue of data is particularly important. Respondents describe AI and autonomous agents as dependent on access to information that is relevant, structured and possible to trust. If an organisation does not have control over what data exists, where it is stored or how sensitive it is, the conditions for letting an autonomous system process or act on that data become weaker. This means that the adoption of autonomous agents depends on how mature the organisation's information handling is. In this sense, poor data control both reduces technical performance as well as it affects whether the technology is perceived as adoptable at all.

Compatibility also concerns integrations with existing systems and the findings here suggest a mixed picture. Some respondents describe legacy systems and technical debt as substantial barriers and argue that older systems can make it difficult to make full use of modern and more autonomous solutions. Others are less convinced that this technical legacy will be a

decisive obstacle in every case, particularly where AI functions can be connected through existing platforms or newer infrastructures. This difference suggests that compatibility is likely to vary depending on the specific part of the security environment that is being discussed. A respondent working close to systems architecture may see integration problems more clearly compared to someone further away from the actual technology, for instance in the strategic part of the organisation.

Beyond the data and systems, compatibility also concerns whether autonomous agents fit with established expectations in cybersecurity work. Security environments are often organised around traceability, validation and control. Technologies that introduce greater autonomy may therefore be perceived as difficult to align with current ways of working, especially if they reduce the user's ability to follow how a decision has been made.

6.1.3 Complexity

The strongest technological barrier in the findings is complexity. Rogers (2003) defines complexity as the extent to which an innovation is perceived as difficult to understand or use. The empirical material supports the relevance of this attribute, but it also shows that the concern is not only about whether users can operate an interface. Instead, complexity is tied to whether the system's reasoning can be understood, assessed and trusted.

Respondents repeatedly describe hesitation toward allowing AI systems to make decisions when it is unclear how they reach conclusions. This concern is reinforced by the black-box character of many systems. Some respondents describe the technology as a black hole, while others stress that organisations need insight into what the system concludes and how it reaches those conclusions. These concerns directly reflect the kind of uncertainty that has also been discussed by Taddeo, McCutcheon and Floridi (2019), who argue that AI in cybersecurity raises questions about opacity and trust.

What makes this especially important for autonomous agents is that the consequences of complexity increase as the degree of autonomy increases. When AI is used as a support tool, a human can review, reject or correct the output before acting. When an agent is expected to work more independently, the human may only intervene after the system has already prioritised, recommended or even carried out a task. The findings suggest that this shift

creates a different level of hesitation. Respondents are questioning whether they can rely on a system whose internal reasoning is not fully transparent in a domain such as cybersecurity where errors can be costly.

A further aspect of complexity is the speed of technological development itself. Respondents describe AI as moving very quickly, at a pace that makes it difficult for organisations and professionals to reflect fully before new capabilities come to earth. This creates uncertainty about what the technology can already do, what it may soon be able to do and when it becomes safe enough to rely on more heavily.

There are also some differences between organisational levels. Respondents closer to strategic and governance-oriented work appear to focus more on transparency, accountability and the ability to justify decisions. Tactical and operational respondents also raise trust concerns, but often in relation to whether the technology actually behaves reliably in practical security work. Together, these perspectives show why complexity becomes such a central barrier. For autonomous agents to be trusted, they need to function reliably and be understandable enough for users to rely on their decisions.

6.1.4 Trialability

The findings also highlight the relevance of trialability, meaning the extent to which a technology can be tested before broader adoption (Rogers, 2003). The interviews show that AI adoption is already developing through a degree of experimentation. Respondents describe groups testing new tools, exploring possible agentic functions and assessing whether they generate practical value. This is particularly visible among operational respondents, who appear close to early experimentation with AI-supported technical work.

Testing matters because autonomous agents are still surrounded by uncertainty. Trialability gives organisations a way to evaluate potential benefits and limitations before making larger commitments. This aligns with Banerjee et al. (2012), who argue that trialability becomes especially important when technologies are perceived as risky or uncertain. In this study, experimentation can help respondents move from general curiosity about agents to a more concrete understanding of where they may be useful and where caution is still needed.

6.1.5 Observability

Observability concerns how visible the results of an innovation are to potential adopters (Rogers, 2003). In the findings, some benefits of AI are easy to observe, especially when respondents describe clear time savings, faster technical support or reduced manual work. These visible improvements appear to strengthen the belief that more advanced AI solutions may also provide value.

However, the value of autonomous agents in cybersecurity may often be harder to demonstrate. Their purpose may be to identify signals, monitor unusual behaviour or prevent incidents before they occur. Since these benefits are more indirect, they may be less visible than simple efficiency gains. This may make it harder to judge whether autonomous agents create enough value to justify their adoption. Observability is therefore closely connected to relative advantage. The technology becomes more attractive when its benefits are clear, but harder to trust when those benefits remain uncertain or are mainly promised by vendors.

6.2 Organisational conditions shaping adoption

The organisational context shows that while adoption is shaped by the technology, there are also other factors coming into play. The organisation's ability to support, govern and take responsibility for its use also has an impact on adoption.

6.2.1 Top management support

Top management support has been identified in previous adoption research as an important factor because the leadership can set priorities, provide resources and signal that a technology is strategically important (Martins et al., 2016; Horani et al., 2025). In the findings, this appears more of a question of what kind of support is currently present, than whether management is positive or negative toward AI.

Several respondents describe AI-related development as being pushed forward by employees and specialist groups who test tools, discuss possible use cases and explore what AI and consequently agents can do in practice. However, this grassroots activity does not take place in isolation from the organisation. The fact that employees are allowed to experiment with AI tools and explore possible applications can itself be understood as a form of organisational support. Rather than being blocked by restrictive leadership, early experimentation appears to

be given some room to develop. This may help explain why interest in AI and agents is able to grow from within the specialist groups.

The findings also suggest that there may be a gap between those closest to the technology and those responsible for broader organisational decisions. Some respondents, especially within the operational level of the organisation, indicate that specialists are often better positioned to judge which tools are actually useful, while those higher up may not always have the same detailed understanding. This does not reduce the importance of management support, but it shows that such support likely needs to be built through dialogue between leadership and technical expertise rather than through top-down decisions alone.

6.2.2 Organisational readiness

A second important organisational condition is readiness. In the theoretical framework, organisational readiness concerns whether the organisation has the competence, resources and internal capabilities needed to evaluate and implement new technologies. Previous research has connected such readiness to technical competence, resource availability and long-term AI preparedness (Martins et al., 2016; Horani et al., 2025; Uren & Edwards, 2023).

The findings show that readiness becomes especially important for autonomous agents because these systems require knowledge from more than one area. Respondents describe a need to understand both the security work itself and how the AI system functions. It is not enough that a person can judge whether an output appears reasonable. They also need to understand what information the system has processed, how the results have been produced and when the output should be questioned. Without this knowledge, respondents describe a greater hesitation toward giving agents a more independent role. This is also connected to the consequences of poor decisions in security work, where errors may affect the confidentiality, availability or integrity of systems and information. The level of competence required is therefore higher than for many simpler uses of AI.

What the findings also reveal is that this readiness gap does not look the same across the organisation. Respondents at the operational level are the ones most actively experimenting with AI tools and engaging with their practical limitations. Their readiness gap is primarily about deepening technical understanding of how autonomous systems function and where

they can be trusted. At the tactical level, respondents are close to system architecture and integration questions, which means that their readiness concerns are around how agents connect to existing environments and whether those environments are mature enough to support them. At the strategic level, respondents engage with AI more through governance, legal interpretation and organisational preparation than through direct use. Their readiness gap is less about technical understanding and more about whether the organisation has the frameworks and competence to make responsible decisions about adoption.

This variation matters because it means that organisational readiness cannot be treated as a condition that is either present or absent. Different parts of the organisation face different gaps and closing them likely requires different measures. For autonomous agents in particular, this creates a coordination challenge. Decisions about where and how to deploy them responsibly cannot rest with one single part of the organisation alone. Instead, they require inputs from operational specialists that understand the technology in practice, tactical architects who see the integration constraints and lastly also strategic decision makers who are responsible for governance and risk.

6.2.3 Strategic alignment and governance

The findings also show that strategic alignment and governance have some importance for the broader adoption of autonomous agents. In previous research, strategic alignment refers to whether a new technology is connected to the organisation's goals, priorities and direction (Horani et al., 2025; Uren & Edwards, 2023). In this study, this becomes relevant because the adoption is not only about testing useful tools, but also about deciding what role AI should have in security work.

As described earlier, much of the current AI use appears to be driven by employees and specialist groups who test tools in relation to their own work. This can support early adoption, since those closest to the work understand where AI may create practical value. However, the findings also suggest that this type of experimentation may become limited if it is not connected to a clearer organisational direction. Respondents describe a need for steering, shared rules and a clearer understanding of what the organisation wants to achieve with AI. Without this, AI risks remaining as separate tools or local experiments rather than becoming part of a more coordinated development.

This also helps explain why silos are described as a possible barrier. If different groups explore AI independently, knowledge and ways of working may become fragmented. For supportive AI tools, this may not be a major problem. However, autonomous agents are more complex because they may depend on shared data, common rules and coordination across different parts of the organisation. Governance therefore becomes important not only for reducing risk, but also for making sure that AI adoption can develop in a consistent and useful way.

6.2.4 Accountability and control

Responsibility and control stand out as central organisational conditions in the adoption of autonomous agents. Across the interviews, respondents repeatedly see the potential of tasks being automated. The problem here however, is that the responsibility can not be transferred to the technology in the same way. Even when an AI system supports or automates a decision, the responsibility is still placed on the human professional or the organisation using it.

This concern becomes stronger when AI is discussed as something autonomous rather than supportive. As it stands today, AI is a tool that helps an analyst search for information or summarise material, leaving the final decision clearly with the human. An autonomous agent may instead monitor, prioritise or act with less direct involvement in each step. This helps explain why respondents describe full autonomy as unrealistic unless strong safeguards are in place. Their hesitation does not appear to be based on a general rejection of automating tasks. Instead, it is connected to the need for confidence in data, understanding the underlying algorithms and lastly also clarity about who remains responsible when something goes wrong.

This shows that accountability becomes important both at the organisational and individual level. Organisations need to define how responsibility should be handled when autonomous agents are introduced, while the professionals using these systems must still be able to stand behind the assessments or actions they rely on. This further helps explain why respondents are cautious toward greater autonomy and that there is a two-way perspective on accountability and responsibility.

6.3 Environmental conditions shaping adoption

The environmental context shows how factors outside the organisation influence the adoption of autonomous AI agents. In the findings, the external threat landscape appears to be the strongest driver, while market expectations, regulation and supplier-related uncertainty mainly shape the conditions under which the adoption can take place.

6.3.1 External threat landscape

The external threat landscape appears to be the most important environmental driver in the findings. Respondents describe cybersecurity as a field where the pace and quality of attacks are increasing, and where defenders need to keep up with threat actors who are also beginning to use AI. This creates a sense that AI adoption is partly about improving internal efficiency, partly about maintaining the ability to respond to a changing threat environment. This also plays into the nature of cybersecurity, in which it is described as better to prevent an attack before it happens, rather than to address the problem afterwards.

This helps explain why autonomous agents are seen as relevant in cybersecurity specifically. The pressure does not only come from the amount of work inside the organisations, but from the fact that attacks are described as becoming faster and harder to manage through traditional methods. Both operational and tactical respondents mention shorter time windows for vulnerability exploitation and zero-day attacks, where organisations may have less time to detect, understand and respond before damage is done. In such a context autonomous agents become attractive because they are imagined as systems that can work continuously, identify unusual behaviour and support faster analysis. All these things to stay one step ahead of the attackers.

6.3.2 External pressure: normative and market

Compared to the external threat landscape, market and normative pressure appear less direct in the findings. Respondents do not describe AI adoption as mainly being driven by fear of competitors. Instead, AI is described as something that is increasingly becoming part of the wider security market, customer discussions and vendor offerings. This creates a softer form of external pressure, where organisations are expected to understand AI and be able to explain how it may create value, even if adoption is not forced by competition alone.

AI therefore becomes both attractive and questioned at the same time. On one hand, customer interest and market developments may encourage organisations to explore AI-based solutions, since security providers need to stay relevant in a field where AI is becoming more common. On the other hand, respondents also express scepticism toward AI being used mainly as a marketing label. When tools are described as "AI-driven" without clear evidence of improvement, this may create caution rather than stronger willingness to adopt.

6.3.3 Regulations and standards

Regulation and external standards have a more complex position in the findings than either the threat landscape or market developments. Rather than being described as strong drivers or blockers to adoption, GDPR, the AI Act and NIS2 are described as conditions that must be managed as adoption proceeds. This supports earlier research by Wallace et al. (2020) and Kaur et al. (2025), who argue that regulation mainly shapes how security technologies are introduced, rather than stopping adoption from happening. The findings add to this by showing that GDPR, the AI Act and NIS2 each affect autonomous agents in different ways.

GDPR becomes most relevant when autonomous agents need access to personal data in order to function. Since security work often involves logs, identity information and internal activity that can be linked to individuals, the conditions for letting an autonomous system process such data become important. Respondents describe a need to understand what data is being processed, where it is stored and whether the processing can be justified and controlled. This connects closely to the earlier discussion of data readiness, since the same lack of control that limits technological compatibility also makes it harder to meet GDPR requirements. In this way, GDPR points to the same need for good data control that autonomous agents already require in order to work.

The AI Act is described in a slightly different way. Its emphasis on transparency, human oversight and risk management aligns closely with the concerns that respondents already express independently of regulatory awareness. Rather than being experienced as an external requirement that limits adoption, the AI Act appears to reinforce values that are already present in how respondents think about autonomous agents. At the same time, the findings suggest that the AI Act mainly creates demands on the organisation to become more competent in relation to AI. The biggest challenges are therefore not described as the legal

text itself, but as the underlying organisational work connected to information management, architecture and the strategic use of AI. In this sense, the AI Act is shaping the conditions for adoption indirectly, by raising expectations on what organisations need to have in place before autonomous agents can be used responsibly.

NIS2 has a somewhat different role in the findings, since it is not focused on AI but on broader cybersecurity requirements. Respondents describe it in a similar way to the AI Act, in the way that it is not seen as something that stops adoption, but as something organisations need to relate to. For autonomous agents, this means fitting into existing requirements for control, documentation and security management. NIS2 therefore shapes adoption indirectly, by reinforcing the need for structure and oversight when new technologies are introduced. This becomes especially relevant when agents act more independently, since the traceability and accountability NIS2 emphasises match the concerns respondents already raise about reduced human involvement.

With these regulations taken all together, the regulatory landscape is described less as a barrier and more as part of the conditions under which adoption takes place. The three regulations all point in the same direction by emphasising control, oversight and responsibility, which are also the areas where respondents already express the strongest concerns. It is also clear that regulation may even develop at a slower pace than the technology itself, meaning that organisations both need to consider existing rules while also trying to prepare for how autonomous agents may be regulated going forward.

6.3.4 Suppliers and technology providers

Suppliers and external technology providers introduce a further environmental dimension that is particularly important for autonomous agents, which are rarely built in-house. This creates questions about where data is processed, who controls the infrastructure and how much trust can be placed in the vendor.

The findings show that this concern is mainly connected to control over sensitive information. If autonomous agents need access to logs, internal systems or security-related data in order to be useful, the choice of supplier becomes more important. Respondents therefore raise concerns about hosted models, foreign technology providers and the lack of

digitally sovereign AI infrastructure in Europe. There is a described growing scepticism toward large international providers, with organisations increasingly looking for alternatives to keep data processing within a more controlled or geographically closer environment.

The distinction between hosted and locally deployed models is therefore practically relevant. A hosted model may offer more capability, but it also means that sensitive data leaves the organisation's own infrastructure. A locally deployed model offers greater control, but may come with limitations in capability or accessibility. For autonomous agents operating in a security environment, where the data being processed is often sensitive by default, this trade-off shapes how solutions are evaluated and which to consider, before other questions are raised.

6.4 Concluding analysis

This study aimed to understand the most important factors influencing the early-stage adoption of autonomous AI agents in cybersecurity organisations. Using a combined DOI-TOE framework, it examined adoption through the technological, organisational and environmental contexts. This through interviews across the strategic, tactical and operational levels of a security organisation. The overall finding is that adoption is shaped by a tension between acceleration and restraint, where the qualities that make autonomous agents attractive are also the qualities that raise the most caution.

Regarding the first research question, on how the technological characteristics shape adoption, the study finds that they pull in two directions at the same time. Autonomous agents are seen as offering a clear relative advantage in handling work that is becoming too large and too fast for manual effort. However, complexity and limited transparency set the limit on how much independence respondents are willing to accept. The advantage makes the technology attractive, but what ultimately decides how far adoption can go is whether security workers can understand and trust how the system reasons.

Turning to the second research question, on how internal organisational factors influence adoption, the study finds that they shape how much independence the organisation can allow. Interest is already growing among specialist groups who test tools in their own work. But

going further depends on readiness and governance, while accountability sets the clearest limit.

Regarding the third research question, on how the external environment affects adoption, the study finds that the environment drives adoption mainly through the threat landscape, while the other external factors shape the conditions under which it can take place. Faster and more capable attacks, and threat actors who are themselves using AI, make autonomous agents appear necessary rather than only useful. Regulation, the market and suppliers do not force adoption to happen, but shape how it proceeds and may give indication of why.

The study contributes to research on technology adoption by applying a combined DOI-TOE framework to a technology and a setting that have received little attention so far, namely autonomous AI agents within cybersecurity. Earlier research has examined the adoption of general AI technologies and cybersecurity measures more broadly. This study instead shows how the specific characteristics of autonomous agents, particularly their autonomy and limited transparency, interact with the conditions of cybersecurity work. In doing so, it adds to the understanding of how adoption happens in a context shaped by evolving threats and high demands on reliability, control and accountability.

7 Discussion

While the analysis examined the technological, organisational and environmental contexts separately, this chapter looks at the findings as a whole. This by focusing on how the three contexts connect rather than treating them as separate explanations. From these connections, the chapter also draws out a number of practical implications for security organisations considering autonomous agents followed by limitations and future research.

7.1 The same conditions surface across all three contexts

One of the clearest patterns in the findings is that the three contexts are not separate explanations, but different views of a few shared underlying conditions. Data control is the most obvious example. In the technological context, poor data control reduces compatibility, since autonomous agents depend on data that is relevant, structured and possible to trust. In the organisational context, the same issue appears as a question of readiness, since the organisation needs to know what data it has and where it is. In the environmental context, it returns again through regulation, since the same lack of control that limits technological compatibility also makes it harder to meet GDPR requirements. Data control is therefore not a technological, organisational or regulatory issue on its own, but one condition that shows up in all three.

Transparency and understanding follow the same pattern. In the technological context, this appears as complexity and the black-box character of the systems. In the organisational context, it returns as accountability, since responsibility for an automated decision can only be taken by someone who understands how it was produced. In the environmental context, it appears again in several of the regulations, especially in the AI Act, which emphasises transparency and human oversight. This is why regulation is described as reinforcing values that respondents already hold, rather than forcing new ones on them. This connection helps explain why the barriers to autonomous AI agents feel so consistent across the interviews, since complexity, accountability and regulation to a large extent are the same problem seen from different perspectives.

7.2 Perception is shaped by position and current use

A further pattern is that the findings cannot be fully explained by the contexts alone, but also by who the respondents are and how they already work. The level a respondent works seemingly shapes which context in the theoretical framework they find affecting the adoption the most. Respondents closer to the tactical and operational levels meet the limits of manual work directly, which is why they describe the relative advantage most clearly. This in areas such as log analysis, identity monitoring and penetration testing. Respondents closer to the strategic level relate to AI more through requirements, governance and legal interpretation than through the direct use of it. The same technology is therefore evaluated differently depending on where in the organisation the respondent sits, which connects back to the strategic, tactical and operational structure described in the background. It also reflects the background's description of AI in cybersecurity, where the use of AI has so far been concentrated at the operational level and to some extent the tactical level. The respondents who describe the relative advantage most clearly are the same ones working where AI has historically been applied, which helps explain why the perceived value is strongest there.

How respondents already use AI also helps explain both the optimism and the caution. AI has so far entered security work mainly as a support tool, used to search for information, generate code, help with log analysis or support documentation. These uses are bounded and easy to observe, and the efficiency gains they produce act as a kind of proof of concept that makes respondents willing to consider going further. At the same time, this current use is also what might explain the hesitation toward autonomy. When AI is used as a support tool, the human reviews the output before acting on it. An autonomous agent removes that step, which means respondents are reasoning from a use of AI they already trust toward a use that works in a different way. The caution is therefore not a rejection of AI, but rather a reflection of the gap between the support-oriented use they know today and the more independent use they are being asked to imagine. This is also why respondents want to keep a human in the loop before moving toward automation.

7.3 Different conditions change at different speeds

A further way of reading the findings is to consider not only what conditions shape adoption, but how quickly each of them changes. The DOI-TOE framework is useful for identifying the

technological, organisational and environmental conditions that matter, but it largely treats these conditions as a picture of the current time. It captures what the conditions are at the point of decision, rather than the speed at which each of them is moving. However, when viewed over time, the contexts in the framework do not move at the same pace. This difference in pace may itself be a reason for the hesitancy this study identifies.

The fast-moving contexts are mainly technological and in some part environmental. The capabilities of autonomous AI agents, and other areas of AI, are developing extremely quickly. This also means that the external threat landscape is developing alongside them, as threat actors adopt the same tools. The market reinforces this acceleration, since AI is increasingly attached to the tools and services surrounding security work, sometimes as an actual capability but also sometimes mainly as a label. From an organisation's perspective, the outside world is therefore moving rapidly. Where the technology it might adopt, the threats it must answer to and the offerings it is presented with, all change faster than it can fully evaluate them.

The slow-moving conditions, on the contrary, are mainly organisational and connected to the regulatory part of the environmental context. Data control, competence, governance structures and accountability are not things that can be acquired quickly. Instead they are built gradually and depend on coordination across the organisation. Regulation moves slowly too, often more slowly than the technology it is meant to govern. These are precisely the conditions this study identifies as necessary for trusting agents with greater independence. They are also the conditions that take the longest to put in place.

Seen this way, the tension between acceleration and restraint that runs through the findings is not only a weighing of benefits against risks, but also a mismatch of speeds. Hesitancy then becomes easier to understand. It is not simply caution about the technology, but a reasonable response to being asked to adopt it faster than the conditions for trusting it can be built. This also explains why organisations can be positive toward AI as a support tool while remaining cautious about autonomy. Support-oriented uses fit within conditions that already exist, whereas greater autonomy depends on the slow-moving conditions that have not caught up yet. For the DOI-TOE framework, this points to a wider implication: in fast-moving domains such as cybersecurity, how quickly each context changes may matter as much for adoption as

the state of the conditions at any moment. If adoption is limited by how slowly the organisational conditions can be built, then the practical task is to start building them early.

7.4 Practical implications

Because the conditions for trusting autonomous AI agents are the slow-moving ones, the practical task for organisations is to start building them early, before the pressure to adopt outpaces them. Several practical implications follow from this for security organisations considering these agents.

Because data control surfaces in all three contexts, it should be treated as a precondition rather than a later concern. Organisations need to know what data they have, where it is and how sensitive it is before relying on autonomous systems. This kind of data control matters on three fronts at once, since it supports technological performance, strengthens organisational readiness and makes regulatory compliance easier. Readiness should also be understood as a combination of security and AI competence. Organisations should expect these gaps to differ across the strategic, tactical and operational levels, rather than assuming readiness is the same everywhere.

Governance and steering matter early. Without a shared direction, AI risks remaining as scattered experiments spread across silos. This is a particular problem for autonomous agents, since they depend on shared data and coordination across the organisation. Organisations should also define how accountability and human oversight will work before granting agents more independence, since responsibility cannot be delegated to the technology. Finally, supplier choice should be treated as a security decision of its own, where the trade-off between hosted and locally deployed models is weighed against the sensitivity of the data involved. Taken together, these implications suggest that adoption is most likely to progress responsibly when organisations build the conditions for trust first and extend autonomy afterwards, since these conditions take time that a fast-moving environment does not always allow.

7.5 Limitations and future research

This study has several limitations that should be considered when interpreting the findings. Firstly, the research is based on a single case study within one cybersecurity organisation.

This made it possible to gain a detailed understanding of how autonomous AI agents are perceived in a specific organisational setting, but it also limits the extent to which the findings can be generalised to all cybersecurity organisations. Other organisations may have different levels of AI maturity, different security structures, different regulatory exposure or different experiences with automation. The findings should therefore be understood as context-specific insights rather than as universal conclusions.

Secondly, the study focuses on the early-stage adoption of autonomous AI agents. This means that the analysis is based on perceptions, expectations and current experimentation rather than on full-scale implementation. Since autonomous AI agents are still an emerging technology, many respondents discussed possible future uses rather than established everyday practices. This is valuable for understanding early adoption, but it also means that the study cannot fully explain how these systems will function once they are implemented more broadly over time.

A third limitation concerns the fast development of AI technology. As discussed in a previous section, this pace is itself a part of the findings, but it also has a methodological consequence. The capabilities of generative AI and autonomous agents are changing quickly, which means that some of the opportunities and concerns identified in this study may develop further in the near future. For example, improvements in transparency, reliability, local deployment or integration with existing security tools could change how organisations evaluate the technology. At the same time, new risks may also appear as agents become more capable and more deeply connected to organisational systems.

Lastly the study is also limited by the selection of respondents. Although the interviews included strategic, tactical and operational perspectives, the number of respondents within each organisational level was relatively small. This means that the findings may not capture the full variation of views within each level. For example, the operational level included perspectives from incident response and penetration testing, but not from roles such as cyber threat intelligence or security operations centre analysts. Similarly, other roles within the strategic and tactical level may have seen opportunities or concerns regarding the adoption differently. The findings should therefore be understood as reflecting selected perspectives within each organisational level and not as representing all possible roles in a cybersecurity organisation.

Future research could therefore examine the adoption of autonomous AI agents across several cybersecurity organisations in order to compare how different organisational contexts shape adoption. Such studies could investigate whether the same tensions between efficiency, control, accountability and readiness appear in organisations with different levels of AI maturity or different types of security work.

Another important direction for future research is to study actual implementations of autonomous AI agents in cybersecurity. While this study focuses on early-stage adoption, future studies could follow organisations over time as the agents are tested, introduced and integrated into security workflows. This would make it possible to examine how trust, responsibility and governance for instance, develops in practice, rather than only how they are perceived before broader adoption.

Since the findings suggest that adoption conditions may differ depending on the task and organisational level, more focused studies could provide a deeper understanding of where autonomous agents are most likely to create value and where they are more difficult to introduce. The potential use areas differ substantially across the security organisations. Autonomous AI agents may create clearer value in operational work, where large amounts of logs, alerts and technical signals need to be handled quickly, while their value may be less direct in strategic work. This is due to the nature of strategic work, where tasks are more connected to judgement, governance and organisational responsibility. Future research could therefore examine specific use cases within different parts of the security organisation instead of treating cybersecurity adoption as one single process.

Finally, further research is needed on governance and accountability for autonomous AI agents in cybersecurity. This study shows that responsibility and control are central concerns, but more work is needed to understand how organisations can design oversight structures, define human involvement and manage responsibility when AI systems act with increasing independence. This will become increasingly important as autonomous agents move from experimental tools toward more active roles in security work.

References

Al-Rahmi, W. M., Yahaya, N., Aldraiweesh, A. A., Alamri, M. M., Aljarboa, N. A., Alturki, U., & Aljeraiwi, A. A. (2019). Integrating technology acceptance model with innovation diffusion theory: An empirical investigation on students' intention to use E-learning systems. Ieee Access, 7, 26797-26809.

Archibald, M. M., Ambagtsheer, R. C., Casey, M. G., & Lawless, M. (2019). Using zoom videoconferencing for qualitative data collection: perceptions and experiences of researchers and participants. International journal of qualitative methods, 18, 1609406919874596.

Arifin, S. R. M. (2018). Ethical considerations in qualitative study. International journal of care scholars, 1(2), 30-33.

Atea. (2025). *Om Atea* [About Atea]. Atea Sverige. Retrieved June 2, 2026, from https://www.atea.se/om-atea/

Banerjee, P., Wei, K. K., & Ma, L. (2012). Role of trialability in B2B e-business adoption: theoretical insights from two case studies. Behaviour & Information Technology, 31(9), 815-827.

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

Bryman, A. (2016). Social research methods. Oxford university press.

Chen, C. Y., Quan, W., Cheng, N., Yu, S., Lee, J. H., Perez, G. M., ... & Shieh, S. (2020). IEEE access special section editorial: Artificial intelligence in cybersecurity. *IEEE Access*, *8*, 163329-163333.

Chiu, C. Y., Chen, S., & Chen, C. L. (2017). An integrated perspective of TOE framework and innovation diffusion in broadband mobile applications adoption by enterprises. International Journal of Management, Economics and Social Sciences (IJMESS), 6(1), 14-39.

Clark, T., Foster, L., Sloan, L., & Bryman, A. (2021). Bryman's social research methods (Sixth edition). Oxford University Press.

Dubois, A., & Gadde, L. E. (2002). Systematic combining: an abductive approach to case research. Journal of business research, 55(7), 553-560.

Eisenstein, J. (2019). Introduction to natural language processing. MIT press.

Englund, E., & Tullin, L. (2020). *Cybersäkerhet: En kartläggning av Sveriges nuläge 2020 och framtidsutsikter för branschen*. Unitalent AB; Security Link; Linköping Science Park; Tillväxtverket.

ENISA (2023) European Union Agency for Cybersecurity. (2023). *ENISA threat landscape 2023*. Publications Office of the European Union.

European Parliament and Council of the European Union. (2016). Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). Official Journal of the European Union, L 119, 1–88.

European Parliament and Council of the European Union. (2024). Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending certain Union legislative acts (Artificial Intelligence Act). Official Journal of the European Union, L 2024/1689.

European Parliament and Council of the European Union. (2022). Directive (EU) 2022/2555 of the European Parliament and of the Council of 14 December 2022 on measures for a high common level of cybersecurity across the Union, amending Regulation (EU) No 910/2014 and Directive (EU) 2018/1972, and repealing Directive (EU) 2016/1148 (NIS2 Directive). Official Journal of the European Union, L 333, 80–152.

Fichman, R. G. (2004). Going beyond the dominant paradigm for information technology innovation research: Emerging concepts and methods. *Journal of the association for information systems*, *5*(8), 1.

Garcia-Teodoro, P., Diaz-Verdejo, J., Maciá-Fernández, G., & Vázquez, E. (2009). Anomaly-based network intrusion detection: Techniques, systems and challenges. *computers & security*, *28*(1-2), 18-28.

Hahn, M., Tretter, M., & Dabrock, P. (2026). Ethical perspectives on AI Agents and Agentic AI. *AI and Ethics*, *6*(2), 218.

Hajkowicz, S., Sanderson, C., Karimi, S., Bratanova, A., & Naughtin, C. (2023). Artificial intelligence adoption in the physical sciences, natural sciences, life sciences, social sciences and the arts and humanities: A bibliometric analysis of research publications from 1960-2021. Technology in Society, 74, 102260.

Horani, O. M., Al-Adwan, A. S., Yaseen, H., Hmoud, H., Al-Rahmi, W. M., & Alkhalifah, A. (2025). The critical determinants impacting artificial intelligence adoption at the organizational level. Information Development, 41(3), 1055-1079.

Hutchins, E. M., Cloppert, M. J., & Amin, R. M. (2011). Intelligence-driven computer network defense informed by analysis of adversary campaigns and intrusion kill chains. *Leading Issues in Information Warfare & Security Research*, *1*(1), 80.

Iles, I. A., Egnoto, M. J., Fisher Liu, B., Ackerman, G., Roberts, H., & Smith, D. (2017). Understanding the adoption process of national security technology: An integration of diffusion of innovations and volitional behavior theories. *Risk Analysis*, *37*(11), 2246-2259.

ISO/IEC 27001 (2022) International Organization for Standardization. (2022). *Information security, cybersecurity and privacy protection — Information security management systems — Requirements* (ISO/IEC 27001:2022).

ISO/IEC 27002 (2022) International Organization for Standardization. (2022). *Information security, cybersecurity and privacy protection — Information security controls* (ISO/IEC 27002:2022).

Jacobsen, J. K. (1993). Intervju, konsten att lyssna och fråga. Studentlitteratur.

Jada, I., & Mayayise, T. O. (2024). The impact of artificial intelligence on organisational cyber security: An outcome of a systematic literature review. *Data and Information Management*, *8*(2), 100063.

Kaplan, A., & Haenlein, M. (2019). Siri, Siri, in my hand: Who's the fairest in the land? On the interpretations, illustrations, and implications of artificial intelligence. Business horizons, 62(1), 15-25.

Kaur, R., Klobučar, T., & Gabrijelčič, D. (2025). Adoption of artificial intelligence in cybersecurity for organizations: Barriers and roadmap. *IETE technical review*, *42*(2), 246-258.

Khan, T. H., & MacEachen, E. (2022). An alternative method of interviewing: Critical reflections on videoconference interviews for qualitative data collection. International journal of qualitative methods, 21, 16094069221090063.

Kvale, S., & Brinkmann, S. (2015). *InterViews: Learning the craft of qualitative research interviewing* (3rd ed.). SAGE Publications.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. nature, 521(7553), 436-444.

Lier, S. K., Eppers, T. M., Gerlach, J., Müller, P., & Breitner, M. H. (2025). An iterative five-phase process model to successfully implement AI for cybersecurity in a corporate environment. Electronic Markets, 35(1), 56.

Lincoln, Y. S., & Guba, E. G. (1988). Criteria for Assessing Naturalistic Inquiries as Reports.

Low, C., Chen, Y., & Wu, M. (2011). Understanding the determinants of cloud computing adoption. *Industrial management & data systems*, *111*(7), 1006-1023.

Martins, R., Oliveira, T., & Thomas, M. A. (2016). An empirical analysis to assess the determinants of SaaS diffusion in firms. Computers in Human Behavior, 62, 19-33.

Mell, P., & Grance, T. (2011). The NIST definition of cloud computing.

MITRE Corporation. (2023). *MITRE ATT&CK* [Knowledge base]. Retrieved April 4, 2026, from https://attack.mitre.org/

Mohamed, N. (2025). Artificial intelligence and machine learning in cybersecurity: a deep dive into state-of-the-art techniques and future paradigms. *Knowledge and Information Systems*, 1-87.

NIST SP 800-53 (2020) National Institute of Standards and Technology. (2020). *Security and privacy controls for information systems and organizations* (NIST Special Publication 800-53, Revision 5). U.S. Department of Commerce.

NIST SP 800-86 (2012) National Institute of Standards and Technology. (2012). *Guide to integrating forensic techniques into incident response* (NIST Special Publication 800-86). U.S. Department of Commerce.

Njie, B., & Asimiran, S. (2014). Case study as a choice in qualitative methodology. Journal of research & method in Education, 4(3), 35-40.

Oliveira, T., & Martins, M. D. R. F. D. O. (2011). Literature review of information technology adoption models at firm level.

Patel, R., & Davidson, B. (2019). *Forskningsmetodikens grunder: Att planera, genomföra och rapportera en undersökning* (5th ed.). Studentlitteratur.

Rawat, D. B., Doku, R., & Garuba, M. (2019). Cybersecurity in big data era: From securing big data to data-driven security. *IEEE Transactions on Services Computing*, *14*(6), 2055-2072.

Robert K. Yin. (2014). Case Study Research Design and Methods (5th ed.). Canadian journal of program evaluation. [Online] 30 (1), 108–110.

Russell, S. J. (2010). Artificial intelligence a modern approach. Pearson Education, Inc..

Saunders, M. (2009). Research methods for business students. Person Education Limited.

Scarfone, K., & Mell, P. (2007). Guide to intrusion detection and prevention systems (idps). *NIST special publication*, *800*(2007), 94.

Sommer, R., & Paxson, V. (2010, May). Outside the closed world: On using machine learning for network intrusion detection. In *2010 IEEE symposium on security and privacy* (pp. 305-316). IEEE.

SVT Nyheter. (2024). *Cyberattack mot datasystem – känsliga uppgifter kan ha läckt*. SVT Nyheter.https://www.svt.se/nyheter/inrikes/cyberattack-i-datasystem-kansliga-uppgifter-kan-ha-lackt

Taddeo, M., McCutcheon, T., & Floridi, L. (2019). Trusting artificial intelligence in cybersecurity is a double-edged sword. *Nature Machine Intelligence*, *1*(12), 557-560.

Terry, G., Hayfield, N., Clarke, V., & Braun, V. (2017). Thematic analysis. The SAGE handbook of qualitative research in psychology, 2(17-37), 25.

Tornatzky, L.G. and Fleischer, M. (1990) The processes of technological innovation. Lexington, MA: Lexington Books.

Uren, V., & Edwards, J. S. (2023). Technology readiness and the organizational journey towards AI adoption: An empirical study. International Journal of Information Management, 68, 102588.

Wallace, S., Green, K. Y., Johnson, C., Cooper, J., & Gilstrap, C. (2020). An extended TOE framework for cybersecurity-adoption decisions. Communications of the Association for Information Systems, 47(1), 51.

Webster, J., & Watson, R. T. (2002). Analyzing the past to prepare for the future: Writing a literature review. MIS quarterly, 26(2), xiii-xxiii.

Wooldridge, M., & Jennings, N. R. (1995). Intelligent agents: Theory and practice. *The Knowledge Engineering Review, 10*(2), 115–152

Yao, S., Zhao, J., Yu, D., Du, N., Shafran, I., Narasimhan, K., & Cao, Y. (2022). React: Synergizing reasoning and acting in language models. arXiv preprint arXiv:2210.03629.

Zhang, Z., Al Hamadi, H., Damiani, E., Yeun, C. Y., & Taher, F. (2022). Explainable artificial intelligence applications in cyber security: State-of-the-art in research. *IEEe Access*, *10*, 93104-93139.

Zhu, K., Kraemer, K. L., & Xu, S. (2006). The process of innovation assimilation by firms in different countries: a technology diffusion perspective on e-business. Management science, 52(10), 1557-1576.

Appendix

Appendix A: Intervjuguide

Bakgrund

- Kan du kortfattat beskriva din roll inom säkerhetsorganisationen och hur du arbetar med säkerhetsfrågor i vardagen?
- Hur ser din kontakt ut med nya teknologiska lösningar eller verktyg i säkerhetsarbetet?
- Hur upplever du att säkerhetsarbetet har förändrats de senaste åren, särskilt i relation till automation eller AI?

Nuvarande användning av AI och autonoma agenter

- På vilket sätt används AI eller automatiserade verktyg i ditt eget arbete idag?
- Hur ser du på skillnaden mellan AI som stödjer dig i arbetet och AI som agerar mer på egen hand?

Innovation och osäkerhet i tidiga skeden

- När nya lösningar som AI eller autonoma agenter diskuteras, vilka typer av osäkerheter brukar uppstå?
- Utifrån det du har insyn i, hur går det vanligtvis till när nya säkerhetsteknologier diskuteras eller utvärderas?
- Hur bedömer ni värdet eller nyttan av en ny säkerhetsteknologi i ett tidigt skede?
- Finns det exempel där organisationen varit försiktig eller avvaktande kring en ny teknologi? Vad berodde det på?

Adoption - Teknologisk kontext

- På vilket sätt ser du att autonoma AI-agenter skulle kunna göra säkerhetsarbetet bättre, snabbare eller mer effektivt jämfört med hur arbetet sker idag?
- Hur väl skulle sådana lösningar passa in i organisationens nuvarande tekniska miljö och arbetssätt?

- Hur lätt eller svårt är det att förstå hur en autonom AI-agent fungerar och kommer fram till sina slutsatser?
- Finns det några teknologiska faktorer som särskilt underlättar eller försvårar införandet av sådana lösningar?

Adoption - Organisatorisk kontext

- Hur skulle du beskriva de etablerade arbetssätten inom din avdelning av säkerhetsorganisationen idag? På vilket sätt kan införandet av AI och agenter förändra dessa arbetssätt?
- Hur påverkar organisationens struktur, kompetens och resurser möjligheten att arbeta med AI säkerhetslösningar?
- Hur ser stödet och styrningen ut från organisationen för att utforska och införa sådana lösningar?

Adoption - Miljö-kontext

- Hur påverkar externa faktorer som hotlandskapet och branschutvecklingen synen på autonoma AI-agenter?
- Hur påverkar juridiska och regulatoriska krav (t.ex. GDPR, AI Act, NIS2) möjligheten att införa sådana lösningar?

Ansvar, kontroll och tillit

- Hur påverkar frågor om ansvar, kontroll eller transparens hur organisationen ser på användningen av AI i säkerhetsarbete?
- Vilka frågor kring ansvar eller transparens uppstår när säkerhetsbeslut i större utsträckning fattas av automatiserade system?
- Finns det situationer där du skulle vara tveksam till att låta ett automatiserat system fatta säkerhetsbeslut?

Rollspecifika frågor

Informationssäkerhet (strategisk nivå):

- Hur ser du att styrmodeller eller policyer behöver anpassas när autonoma AI agenter börjar användas i säkerhetsfunktioner?

IT-säkerhet (taktisk nivå):

- Hur ser du på hur autonoma AI agenter skulle kunna kopplas in i och integreras med de tekniska miljöer och system du arbetar med att utforma?

Cybersäkerhet (operativ nivå):

- Hur skulle autonoma AI agenter kunna förändra hur incidenter upptäcks och hanteras i det dagliga arbetet?

Avslutande fråga

- Vad tror du är de viktigaste faktorerna som kommer att avgöra om AI och agenter får en större roll i säkerhetsarbetet framöver?