Kaunas University of Technology

School of Economics and Business

Governing Shadow Artificial Intelligence in Organizations

Master's Final Degree Project

Tadas Semionovas

Project author

Assoc. Prof. dr. Inga Stankevičė

Supervisor

Kaunas, 2026

Kaunas University of Technology
School of Economics and Business

Governing Shadow Artificial Intelligence in Organizations

Master's Final Degree Project
Innovation Management and Entrepreneurship (6211LX031)

Tadas Semionovas
Project author

Assoc. Prof. dr.
Inga Stankevičė
Supervisor

Assoc. Prof. dr.
Vytautė Dlugoborskytė
Reviewer

Kaunas, 2026

Kaunas University of Technology

School of Economics and Business
Tadas Semionovas

Governing Shadow Artificial Intelligence in Organizations

Declaration of Academic Integrity

I confirm the following:

1. I have prepared the final degree project independently and honestly without any violations of the copyrights or other rights of others, following the provisions of the Law on Copyrights and Related Rights of the Republic of Lithuania, the Regulations on the Management and Transfer of Intellectual Property of Kaunas University of Technology (hereinafter – University) and the ethical requirements stipulated by the Code of Academic Ethics of the University;

2. All the data and research results provided in the final degree project are correct and obtained legally; none of the parts of this project are plagiarized from any printed or electronic sources; all the quotations and references provided in the text of the final degree project are indicated in the list of references;

3. I have not paid anyone any monetary funds for the final degree project or the parts thereof unless required by the law;

4. I understand that in the case of any discovery of the fact of dishonesty or violation of any rights of others, the academic penalties will be imposed on me under the procedure applied at the University; I will be expelled from the University and my final degree project can be submitted to the Office of the Ombudsperson for Academic Ethics and Procedures in the examination of a possible violation of academic ethics.

Tadas Semionovas
Confirmed electronically

Semionovas Tadas. Governing Shadow Artificial Intelligence in Organizations. Master's Final Degree Project, supervisor Assoc. Prof. dr. Inga Stankevičė; School of Economics and Business, Kaunas University of Technology.

Study field and area (study field group): Management, Business and Public Management.

Keywords: Shadow AI; Shadow IT; AI governance; cybersecurity risk management; organizational governance; innovation enablement.

Kaunas, 2026. 68 p.

Summary

This thesis examines the governance of Shadow Artificial Intelligence in organizations. Shadow AI is mostly understood as the usage of AI tools, models, or AI-enabled features by employees or business units without explicit approval or outside established organizational governance controls. The topic is relevant because AI tools have become easy to access and use in everyday work. At the same time, organizational rules, approval processes, technical controls, and risk management practices often remain unclear or slow to adapt.

The study treats Shadow AI as an organizational governance problem with cybersecurity consequences. Its roots are similar to those of Shadow IT, but its risk profile is very different. AI tools can generate content, process sensitive information, support decisions, automate actions, and produce outputs that appear reliable even when they are completely inaccurate. While employees often use these tools for practical reasons: to write faster, summarize information, translate text, support coding, prepare analysis, or reduce repetitive work. Because of this, a governance approach based only on restrictions is unlikely to work.

This thesis aims to develop a conceptual framework for governing Shadow AI in a way that supports responsible AI use while managing cybersecurity and information governance risks. The research follows a qualitative two phase design. In the first phase, academic literature, industry sources, and governance standards are analyzed to build the theoretical foundation. The reviewed areas include Shadow IT, Shadow AI, AI governance, IT governance, risk management, compliance, and the balance between innovation and control.

Based on this analysis, the thesis develops the Shadow AI Governance Framework (SAIGF). The framework is structured into three layers: governance foundation, core governance mechanisms, and innovation enablement with sustainability. These layers define the main areas organizations need to address, including responsibility, regulatory alignment, risk appetite, visibility, policy, authorization, risk assessment, technical controls, approved AI options, faster approval paths, awareness, culture, and ongoing improvement.

The second phase validates and refines the framework through ten semi structured interviews with participants from six organizations. The sample includes managerial respondents involved in governance or AI-related decisions, as well as employees who use AI tools in their work. The findings generally support the framework, but they also show that several areas require adjustment. The empirical results expand risk appetite to include AI cost governance, investment failure risk, and

operational dependency risk. They also introduce agentic AI permission governance, embedded AI reassessment, and human oversight of AI supported decisions.

One of the key findings is the gap between managerial and employee perspectives. Managers often describe policies, rules, or approval processes as already present. Employees, however, frequently experience the same governance mechanisms as unclear, difficult to find, or detached from everyday work. This means that formal governance may exist, but still fail in practice if employees cannot understand or use it.

The thesis concludes that Shadow AI cannot be managed effectively through blocking, generic awareness messages, or policy documents alone - i.e. simple black or white controls. Organizations need to have clearer ownership, proportionate controls, usable approved tools, practical guidance, and governance processes that match how employees actually work. The proposed SAIGF offers a structured way to manage Shadow AI risks while still preserving the productivity and innovation value that makes AI adoption attractive in the first place.

Semionovas Tadas. Šešėlinio dirbtinio intelekto valdymas organizacijose. Magistro baigiamasis projektas vadovė dr. Inga Stankevičė; Kauno technologijos universitetas, Ekonomikos ir verslo fakultetas.

Studijų kryptis ir sritis (studijų krypčių grupė): Vadyba, Verslas ir viešoji vadyba.

Reikšminiai žodžiai: šešėlinis DI; šešėlinės IT; DI valdysena; kibernetinio saugumo rizikos valdymas; organizacinė valdysena; inovacijų įgalinimas.

Kaunas, 2026. 68 p.

Santrauka

Šiame tyrime nagrinėjamas šešėlinio dirbtinio intelekto valdymas organizacijose. Šešėlinis DI dažniausiai suprantamas kaip darbuotojų ar verslo padalinių vykdomas DI įrankių, modelių ar DI funkcijų naudojimas be aiškaus organizacijos patvirtinimo arba už nustatytų organizacinio valdymo kontrolės mechanizmų ribų. Tema yra aktuali, nes DI įrankiai tapo lengvai prieinami ir naudojami kasdieniame darbe. Tuo pačiu metu organizacijų taisyklės, patvirtinimo procesai, techninės kontrolės priemonės ir rizikos valdymo praktika neretai išlieka neaiškūs arba lėtai prisitaiko prie pokyčių.

Darbe šešėlinis DI nagrinėjamas kaip organizacinio valdymo problema, turinti kibernetinio saugumo pasekmių. Jo ištakos panašios į šešėlinės IT ištakas, tačiau rizikų pobūdis gerokai skiriasi. DI įrankiai gali generuoti turinį, apdoroti jautrią informaciją, padėti priimti sprendimus, automatizuoti veiksmus ir pateikti rezultatus, kurie atrodo patikimi net tuomet, kai yra netikslūs. Darbuotojai šiuos įrankius dažnai naudoja praktiniais tikslais: norėdami greičiau rašyti, apibendrinti informaciją, versti tekstus, padėti programuoti, rengti analizę ar sumažinti pasikartojančio darbo apimtį. Dėl to vien tik ribojimu paremtas valdymo požiūris vargu ar būtų veiksmingas.

Darbo tikslas – sukurti konceptualų šešėlinio DI valdymo modelį, kuris padėtų organizacijoms atsakingai naudoti DI ir kartu valdyti kibernetinio saugumo bei informacijos valdymo rizikas. Tyrimas vykdomas taikant kokybinį dviejų etapų dizainą. Pirmajame etape analizuojama akademinė literatūra, praktiniai šaltiniai ir valdymo standartai, siekiant suformuoti teorinį pagrindą. Apžvelgiamos sritys apima šešėlinę IT, šešėlinį DI, DI valdyseną, IT valdyseną, rizikų valdymą, atitiktį bei inovacijų ir kontrolės pusiausvyrą.

Remiantis šia analize, darbe sukurtas šešėlinio DI valdymo modelis, angl. Shadow AI Governance Framework, SAIGF. Modelis sudarytas iš trijų sluoksnių: valdymo pagrindo, pagrindinių valdymo mechanizmų bei inovacijų įgalinimo ir tvarumo. Šie sluoksniai apibrėžia pagrindines sritis, į kurias organizacijos turėtų atkreipti dėmesį: atsakomybę, reguliacinį suderinamumą, rizikos toleranciją, matomumą, politiką, autorizaciją, rizikos vertinimą, technines kontrolės priemones, patvirtintų DI įrankių pasirinkimą, pagreitintus patvirtinimo procesus, sąmoningumą, kultūrą ir nuolatinį tobulinimą.

Antrajame etape modelis patikrinamas ir patikslinamas atliekant dešimt pusiau struktūruotų interviu su šešių organizacijų atstovais. Tyrimo imtį sudaro tiek vadovaujamo lygmens respondentai, atsakingi už valdymo ar su DI susijusius sprendimus, tiek darbuotojai, naudojantys DI įrankius savo darbe. Empiriniai rezultatai iš esmės patvirtina modelio aktualumą, tačiau taip pat parodo, kad kai

kurias sritis būtina patikslinti. Empiriniai duomenys išplečia rizikos tolerancijos sampratą, įtraukiant DI kaštų valdymą, nesėkmingų investicijų riziką ir veiklos priklausomybės nuo DI įrankių riziką. Taip pat įtraukiami agentinio DI prieigos teisių valdymas, į patvirtintus įrankius integruoto DI pakartotinis vertinimas ir žmogaus priežiūra priimant DI palaikomus sprendimus.

Vienas svarbiausių rezultatų yra atotrūkis tarp vadovų ir darbuotojų požiūrių. Vadovai dažnai nurodo, kad politikos, taisyklės ar patvirtinimo procesai jau egzistuoja. Darbuotojai tuos pačius valdymo mechanizmus dažnai vertina kaip neaiškius, sunkiai surandamus ar atitrūkusius nuo kasdienio darbo. Tai rodo, kad formalus valdymas gali egzistuoti, tačiau praktikoje neveikti, jeigu darbuotojai jo nesupranta arba negali juo pasinaudoti.

Darbe daroma išvada, kad šešėlinio DI negalima veiksmingai valdyti vien blokavimu, bendro pobūdžio sąmoningumo žinutėmis ar politikos dokumentais, tai yra paprastu juodai baltu kontrolės principu. Organizacijoms reikia aiškiau paskirstytos atsakomybės, proporcingų kontrolės priemonių, naudotojams tinkamų patvirtintų įrankių, praktiškų gairių ir valdymo procesų, atitinkančių realų darbuotojų darbo pobūdį. Siūlomas SAIGF modelis suteikia struktūruotą būdą valdyti šešėlinio DI rizikas, išsaugant produktyvumo ir inovacijų vertę, dėl kurios DI naudojimas organizacijose tampa patrauklus.

Table of contents

List of tables

List of figures

List of abbreviations

Abbreviations:

IT – Information technology
NIST – National Institute of Standards and Technology
CIS – Center for Internet Security
AI – Artificial Intelligence
ML – Machine Learning
LLM – Large Language Models
SAI – Shadow Artificial Intelligence
GPT - Generative pre-trained transformer
DLP - Data Loss Prevention
BYOD – Bring your own device
EU – European Union
GDPR – General Data Protection Regulation
CASB - Cloud Access Security Broker
SIEM - Security Information and Event Management
HIPAA - The Health Insurance Portability and Accountability Act
PCI-DSS - Payment Card Industry Data Security Standard
DPIA - Data Protection Impact Assessment
SaaS - Software as a Service

Introduction

Research relevance:

With the rapid expansion of large language models (LLMs) based on artificial intelligence (AI), their use has been increasing not only in everyday life but also across professional environments. This trend is reflected in the *Artificial Intelligence Index Report (2025), as* it is increasingly being used more and more not only in people's daily lives, but also within the professional environment as well. This creates new governance gaps as employees adopt AI tools faster than organizational controls can adapt. As AI becomes increasingly tightly integrated within professional environments, this brings information technology (IT) and cybersecurity teams a new challenge: how to manage this rapid growth of technology. This gives rise to a relatively new phenomenon dubbed shadow AI (SAI) (Chin et al., 2025a). This is usually defined as the deployment and/or usage of SAI tools, models systems by employees of an organization without explicit endorsement, supervision, or management by IT and cybersecurity departments (Puthal et al., 2025a). But this also creates another dilemma for organization leadership – how to manage these AI tools while still maintaining and even leveraging these tools for organizational and employee benefit, but without stifling innovation (Silic et al., 2025a). However, practical governance approaches for SAI that balance risk control with innovation enablement are still underdeveloped in the literature, motivating this research to create a management framework for organizations.

Research problem analysis:

The aim of the research is to develop a conceptual model (framework) for organizational governance of SAI that balances innovation enablement with cybersecurity risk management in the context of rapid AI adoption.

The object of the research is the organizational governance of SAI usage, specifically, the mechanisms used to manage risks while enabling innovation.

Research Objectives:

1. To conduct a systematic available literature review on SAI, shadow IT, AI governance, and innovation governance to identify relevant concepts and existing approaches
2. To identify and classify the key organizational drivers, use cases, and risks associated with SAI adoption and usage
3. To develop a conceptual SAI management framework that balances innovation enablement with cybersecurity and information governance requirements, enabling innovations, and specifying core governance mechanisms
4. To validate and refine the proposed conceptual SAI management framework using empirical input, and to provide managerial recommendations for the model's implementation and suggestions for future research directions.

Thesis methodology and structure:

This thesis employs a qualitative, two-phase approach to develop and validate an organizational governance framework for SAI management. In the first phase, academic literature, industry sources, and relevant governance standards are examined to establish the theoretical foundations of Shadow

AI governance. The analysis covers Shadow IT, Shadow AI, AI-related risk areas, IT governance principles, regulatory and compliance requirements, and the balance between innovation and control. Based on this examination, a conceptual Shadow AI Governance Framework is created.

In the second phase, the conceptual framework is validated through empirical evidence gathered from semi-structured interviews. These interviews include both managerial-level respondents involved in governance, cybersecurity, IT, or AI-related decision-making, and non-managerial employees who use AI tools in their professional roles. This approach enables the research to compare governance intentions with actual employee experiences and to identify practical gaps between formal policies and real-world AI use.

Chapter 1 provides a thorough and systematic problem analysis. Chapter 2 then expands on this by offering an organized review of academic, industry-standard, and practical research, aiming to gather key data points on approaches relevant to SAI management within organizations. Using this foundational knowledge, Chapter 2 discusses the theoretical background and ultimately develops the theoretical SAI management framework.

Chapter 3 outlines the methodology and sampling approach. To validate the theoretical model, semi-structured interviews are conducted with individuals in both managerial and non-managerial roles.

Chapter 4 presents the empirical findings, validates and refines the conceptual framework, and provides discussion, managerial recommendations, technical implications, limitations, and directions for future research.

1. Problem analysis Shadow Artificial Intelligence management within organizations

1.1. Rapid AI adoption and emergence of Shadow AI

With the implementation and general availability of AI models, plug-and-play AI services, and tools, not so long ago considered as inaccessible for the layman, AI is now embedded in almost all everyday life's and workflows of employees across the world, no matter the sector or organizational level (Ross et al., 2025). This rapid explosion of AI and AI-based solutions has left various organizations and their IT departments scrambling – employees of various levels started using these tools without IT having any governance structures in place. "69% of IT leaders feel pressured to adopt AI faster than they can secure it."(PAUBOX, 2025)

This situation is especially relevant because AI adoption is no longer limited to centrally planned digital transformation initiatives. In the past, adopting more advanced technologies often required budget approval, procurement involvement, IT deployment, system integration, administrator privileges, and at least some form of technical support. In contrast, many AI tools can now be accessed directly through a browser, a personal account, a mobile application, a browser extension, or an embedded feature in an existing business application. This creates a situation in which technology adoption can occur at the employee level before the organization has even decided how such tools should be assessed, approved, controlled, or monitored.

The situation becomes more intricate as AI is no longer just a standalone tool. It is increasingly embedded into productivity suites, communication apps, coding environments, document processing systems, data analytics platforms, and other software used by organizations. Employees might perceive these AI features as just regular updates or new functionalities. Yet, from a governance perspective, such updates can create new data flows, processing activities, output capabilities, and dependencies on third parties that weren't considered during initial approvals. Consequently, organizations must now oversee not only the use of public tools like ChatGPT but also AI features integrated into existing, possibly pre-approved, tools that might not have been individually reviewed.

1.2. What Makes Shadow AI Distinct from Shadow IT

As Silic, Silic, and Kind-Trüller, (2025b) observed, "As artificial intelligence (AI) matures, the Shadow IT landscape is evolving into what can be termed "Shadow AI". This phenomenon involves the use of AI tools and applications without formal organizational approval, mirroring the core dynamics of traditional Shadow IT but introducing far more complex risks". SAI has its origins in SIT, and the behavioral dynamics that produce it are recognizably similar. However, the technology's properties make SAI a qualitatively different governance challenge.

The threat is somewhat understood. However, it raises many topic-specific issues that almost make it a separate subject. "While Shadow AI shares roots with Shadow IT, its generative, opaque, and autonomous nature introduces novel risks related to data privacy, algorithmic bias, hallucination, and governance drift." (Silic et al., 2025).

This distinction holds significant importance. Traditional Shadow IT typically pertains to employees or departments utilizing hardware, software, cloud services, or digital solutions without formal approval from the Information Technology department. The principal concern is that such technology operates within the organizational environment without official oversight and control. SAI follows a

similar pattern; however, the potential consequences are often more challenging to predict. AI tools are not limited to traditional methods of storing, transferring, or processing information. These AI tools can generate new content, summarize documents, write code, provide recommendations, support decision-making, automate actions, and produce outputs that may subsequently be regarded as reliable business intelligence. In addition, SAI extends the conventional Shadow IT dilemma from the unauthorized use of technology to the unauthorized generation of knowledge and decision support (Ross et al., 2025).

While traditional unauthorized SaaS tools pose risks related to data storage, access control, licensing, or supplier management, unauthorized AI tools introduce similar concerns plus additional issues like output reliability, hallucinations, bias, accountability, and intellectual property. For instance, when an employee uploads internal data to a public AI service, the organization loses control over what data is shared, how it is processed, and whether it is retained by the provider. If this employee later uses the AI-generated output in a report, customer communication, code update, legal interpretation, or management decision, operational and accountability risks arise. The core issue is not just data leaving the organization, but also that AI-generated outputs may re-enter organizational processes without clear identification, validation, or governance.

1.3. SAI usage drivers

A major reason why SAI becomes an organizational problem is that it tends to emerge within environments where speed and productivity are rewarded. Employees are often driven by efficiency needs, reduced cognitive load, and the desire to produce outputs faster. For example: "AI 'reduces cognitive load, allowing employees to focus on higher-value tasks instead of spending hours on manual work.'"(Silic et al., 2025b) . AI tools provide immediate benefits with low adoption friction (often requiring only a browser). At the same time, organizational approval processes for new tools are frequently slow, compliance-heavy, or unclear, which incentivizes bypass behavior even when employees do not intend harm. “People tend to do it without thinking, just wanting to speed up their work […] you just uploaded a bunch of company data […] and your security team does not know about this” (PAUBOX, 2025)

This shows that SAI should not automatically be seen as a sign of malicious employee behavior. Often, it is driven by practical work pressures. Employees may try to boost their efficiency by writing faster, synthesizing information, translating texts, preparing presentations, generating ideas, fixing technical problems, improving client communication, or filling skill gaps. In these situations, AI is attractive because of its immediacy, ease of use, and often faster response compared to organizational support channels. Employees might see the tool as a productivity helper rather than a new technology that needs management review. This explains why purely punitive or restrictive measures may be ineffective: they may punish the symptom without solving the underlying organizational need.

At the individual level, SAI is therefore driven by convenience, speed, productivity, and perceived usefulness. Employees utilize AI as it facilitates faster and less effortful completion of tasks. It can support writing, summarizing, coding, research, translation, communication, analysis, and decision-making preparations. Additionally, AI can assist employees in areas where they may feel less confident. For example, a technically proficient employee might employ AI to enhance writing quality, while a non-technical employee may utilize AI to comprehend technical documentation or

automate repetitive tasks. This versatility renders AI applicable across various departments, roles, and levels of seniority (PAUBOX, 2025; Ross et al., 2025; Silic et al., 2025b).

At the organizational level, SAI is often influenced by governance immaturity, ambiguous ownership, and a misalignment between business requirements and formal approval procedures. Many entities have yet to explicitly define the ownership of AI governance, the approved tools, the permissible data inputs into AI systems, the review requirements for specific use cases, and the process for employees to seek approval for new tools. Even when policies are established, they may be stored in inaccessible locations, articulated in language that is challenging for non-technical personnel, or disseminated solely through generic awareness messages. Practically speaking, a policy that exists but is neither comprehensible nor readily accessible may effectively function as if no policy were in place (PAUBOX, 2025; Ross et al., 2025; Silic & Back, 2014).

From a technological perspective, the barriers to adopting artificial intelligence are considerably lower. Public AI tools are readily accessible without the necessity for installation, administrator privileges, or occasionally registration. Browser extensions and plug-ins further facilitate ease of adoption. Embedded AI functionalities add additional complexity, as organizations might approve the primary application while AI features are implemented subsequently, bypassing the initial review process. Locally operated models and open-source AI tools introduce an extra layer of complexity, especially when technically skilled employees are capable of operating or modifying models beyond the supervision of centralized IT departments (PAUBOX, 2025; Silic et al., 2025b).

At the business and market levels, organizations are also under significant pressure to adopt AI promptly. Leadership frequently anticipates that AI will lead to productivity enhancements, cost reductions, accelerated innovation, and improved competitiveness. Nevertheless, such expectations may precede the development of suitable governance capabilities within the organization. Consequently, employees may receive conflicting signals: on one hand, they are encouraged to be innovative and to utilize AI; on the other hand, there may be no clear, practical, and sanctioned pathway for doing so. This environment is amplified by the risk of shadow usage becoming not only feasible but widespread (Chin et al., 2025b; Ross et al., 2025; Xu, 2025).

1.4. SAI governance failures

Shadow AI can be viewed as a combination of four observable governance failures.

- First, lack of visibility: organizations often do not have reliable inventories of which AI tools are being used, by whom, for what purpose, and with what data.
- Second, lack of accountability and decision rights: it is frequently unclear who is responsible for approving AI use cases, who owns the risk, and who is accountable when AI outputs affect business decisions.
- Third, inconsistent rules and enforcement: even when policies exist, employees may not know them or may interpret them differently across teams, and enforcement can be uneven across departments and tool types.
- Fourth, lack of evidence and auditability: governance requires the ability to demonstrate what controls exist and whether they work; Shadow AI often lacks logging, documentation, and traceability, which becomes critical in compliance, incident response, and reputational risk contexts.

These four governance failures demonstrate that SAI is not solely a technological issue but also a comprehensive organizational governance challenge. Without proper visibility, the organization cannot accurately assess the extent of AI utilization. A lack of accountability hampers the clear allocation of responsibility regarding approvals, risks, or incidents. Inconsistent or inaccessible rules undermine employee adherence. The absence of evidence and auditability prevents the organization from demonstrating compliance or deriving lessons from incidents. Consequently, managing SAI necessitates more than merely blocking websites or issuing broad policies; it requires a structured governance framework that integrates decision rights, risk appetite, approval processes, technical controls, employee awareness, and innovation support.

1.5. Risk Landscape created by SAI

The main concern with SAI is data exposure. Staff may accidentally upload internal documents, customer details, employee records, financial information, source code, contracts, meeting notes, or other sensitive materials into AI tools without fully understanding the risks. While many organizations implement measures like Data Loss Prevention (DLP), web filters, endpoint security, and access controls, these might not completely cover AI-specific data flows. AI tools can process data externally, save prompts, or use information to improve services, often relying on complex supply chains that are not transparent. These issues pose threats to cybersecurity, data privacy, confidentiality agreements, and compliance with regulations.

Another issue is legal and compliance risk. Laws and industry rules require organizations to clearly specify what data they process, why, where, who handles it, and what safeguards are in place. Employees using AI tools improperly might lead to gaps in data processing agreements, risk assessments, legal justifications, retention policies, or audit records. This is especially critical when dealing with personal information, customer data, regulated datasets, financial records, healthcare information, legal documents, or intellectual property. Often, organizations are unaware of their own data processing activities, which compounds the risk (Balogun et al., 2025).

A third challenge involves the reliability of AI outputs. Results generated by AI can seem confident and polished even if they are inaccurate. This can be problematic if employees rely on AI for decision-making, customer communication, legal review, troubleshooting, coding, or summarizing complex documents. The danger isn't just hallucinations but also over-reliance, where errors go unnoticed because the AI's output appears authoritative. Without proper review or quality checks, mistakes can creep into workflows. This risk is heightened in SAI environments, where formal oversight is less common (Ross et al., 2025; Silic et al., 2025).

Another concern focuses on accountability. When AI influences decisions or outcomes, it can be unclear who is responsible. If an AI recommendation leads to a poor decision, questions arise about whether the employee, manager, provider, or organization should be held accountable. Well-structured setups with clear roles and reviews can help reduce this issue. However, in cases of Shadow AI, accountability often slips through the cracks, creating a governance gap and making it difficult to assign responsibility.

Finally, operational dependence and tool proliferation pose risks. When departments adopt AI independently, usage can become fragmented, with duplicate subscriptions, inconsistent practices, and informal reliance on various tools without proper evaluation. Teams might depend on AI for

routine tasks, reporting, customer interactions, or analysis without oversight. Changes in tools, outages, higher costs, or errors could disrupt operations. Overall, SAI introduces risks that go beyond information security, threatening overall business continuity and process stability.

1.6. Innovation dilemma and proportionate governance

At the same time, it is important not to ignore the positive side of AI usage. AI tools can increase productivity, support creativity, improve employee experience, reduce repetitive work, and help organizations innovate. This creates the central dilemma for organizational leadership: how to manage these AI tools while still maintaining and even leveraging them for organizational and employee benefit, without stifling innovation (Silic, Silic, & Kind-Trüller, 2025a). If organizations respond only with prohibition, employees may continue using AI through personal devices, personal accounts, or less visible channels. This would reduce organizational visibility even further and may increase rather than decrease risk.

Therefore, the issue is not whether AI should be permitted or banned. The real concern is how AI can be governed to align with actual employee behavior and organizational needs. An effective governance approach must acknowledge that employees use AI because it addresses practical problems. If approved tools are unavailable, slow, hard to use, or much worse than public options, employees are more likely to turn to shadow alternatives. If approval processes are unclear or too slow, employees are less inclined to follow them. If policies are written solely from a compliance perspective without explaining how to use them practically, employees may not understand how to apply them. Successful SAI governance, therefore, requires both control and support.

This highlights the need for proportionate governance. Low-risk AI applications, such as enhancing the wording of non-sensitive internal texts, should not necessarily undergo the same review process as high-risk applications involving personal data, customer information, source code, regulated decisions, or automated actions. If all AI use is regarded as equally risky, governance can become too slow, leading employees to bypass it. Conversely, if all AI use is considered low risk, organizations might fail to address serious exposures. Therefore, an effective SAI governance framework must differentiate between various tools, use cases, data types, output impacts, and levels of autonomy.

1.7. Research gap and problem

Despite the urgency and scale of the SAI governance challenge, the academic literature has not yet produced a comprehensive response to it. As expressed by (Silic et al., 2025b): "Despite growing interest in Shadow IT and ethical AI governance, existing research has not fully addressed the intersection of these domains" or SAI adjacent fields. Existing Shadow IT research provides a solid foundation for understanding unauthorized technology adoption, but it does not fully address the generative, opaque, and decision-influencing nature of AI. Existing AI governance research offers valuable principles for responsible AI, risk management, and compliance, but it often assumes that AI systems are known, approved, and formally managed by the organization. Shadow AI falls between these areas: it is AI use that occurs before or outside formal governance.

This research seeks to address existing deficiencies in SAI management by contributing to the collective body of knowledge through the development of a conceptual framework for the organizational governance of SAI. This framework aspires to harmonize the promotion of innovation with the management of cybersecurity risks within the context of rapid AI integration. Its practical

relevance resides in assisting organizations to move beyond reactive and fragmented approaches. Rather than exclusively relying on blocking strategies, awareness communications, or conventional software approval procedures, organizations necessitate a structured methodology to identify SAI, assess associated risks, allocate accountability, provide sanctioned alternatives, establish expedited approval processes, implement technical controls, and continuously improve governance by monitoring employee behavior and technological advancements.

2. Theoretical Presumptions

This section discusses the theoretical foundations of SAI and based with this knowledge proposes a theoretical framework on how to effectively manage this phenomenon.

2.1. Foundations of Shadow IT

The following section discusses the origin of SAI – Shadow IT. It has been and still is an issue that organizations have to face and tackle, but this area is usually somewhat well understood, has lover complexity and has already established controls. This Section will review it to gain necessary knowledge before proceeding to SAI.

2.1.1. Definitions and evolution of Shadow IT

To start will be establishing the baseline definition of SIT based on the existing literature.

Table 1. Definitions of SIT

Author, year	**SIT Definition**
(Ross et al., n.d.)	*Shadow AI—the unsanctioned, informal use of artificial intelligence technologies within organisations.*
(Silic & Back, 2014)	*Shadow IT is a currently misunderstood and relatively unexplored phenomena. It represents all hardware, software, or any other solutions used by employees inside of the organisational ecosystem which have not received any formal IT department approval.*
Huber et al. (2018)	*Business departments and users autonomously implement IT solutions outside of the organizational IT service management. This phenomenon is called Shadow IT.*
(Huber et al., 2018)	*Shadow IT describes IT systems that business units implement individually in their business processes, whereby they are not involved in an organizational IT management*
(Raković et al., 2020)	*The following terms are used in the literature for systems developed by end users or used without the knowledge of IT departments: shadow system(s), shadow IT, feral system(s), grey IT, hidden IT, rogue IT, workaround systems, workaround IT, unofficial IT, bolt-on*

Within **Table 1** we can see various definitions over many years that have been used within different research. Even though there are a lot more research articles that cover this topic, we can already see that most of these definitions focus on voluntary or involuntary misuse of organizational resources without having a formal approval from the company's IT department.

Therefore, building on existing literature (Silic & Back, 2014; Huber et al., 2018; Raković et al., 2020), this thesis defines SIT as the autonomous adoption and use of technology solutions - including hardware, software, cloud services, and applications - by employees or business units outside the

formal IT governance structure, without explicit approval or oversight from the organizational IT department.

This definition captures several important characteristics identified within the reviewed literature:

- Lack of authorization – Solutions lack explicit approval from the IT department.
- User-driven adoption – Employees or business units initiate this usage.
- Governance bypass – This usage occurs outside established IT controls or structures.

It is also important to note that SIT can also be referred to as follows, as defined by Raković et al. (2020):

- Feral systems
- IT Workaround
- Grey IT
- Hidden IT
- Rogue IT
- Workaround systems
- Unofficial IT

This thesis will use the already established unifying term for all the above – Shadow IT.

Having discussed and established the unifying definition of SIT, next take a look at what drives employees to adopt such technologies and tools.

2.1.2. Drivers of SIT adoption

Based on the already reviewed literature, we can split out four key areas that are driving the adoption of SIT, and review each area individually.

Individual/Personal drivers

For individual drivers there is a multitude of personal and other outside factors that cause emergence or usage of SIT. For example “Employees resort to Shadow IT not merely due to convenience but also because they perceive official tools as misaligned with their job requirements." (Walters, 2021). Also employees change, they change companies, change positions and some may just continue usage of certain tools just “Due to inertia on an individual as well as an organizational level, business users continue using shadow IT." (Huber et al., 2018).

Organizational drivers

For organizational drivers usage of SIT becomes a bit different, it’s no longer just a simple inertia, but can be something just “the lack of a good alignment between business and IT." (Silic & Back, 2014). That’s just plane misalignment. Additional factors may include power aspects (Spierings et al. 2012; Kerr et al. 2007) or individual behavior (Ortbach et al. 2013)." (Zimmermann et al., 2014), of certain key stakeholders that point entire organizational unit SIT way. Also business organizations quite often feel that a long response time (fulfilment of user requests) by IT department and low initial (perceived) costs of shadow IT.

Technological drivers

With emergence of various SaaS based services – adoption of SIT tools has never been easier. Users just need to sign up to a website, using click through agreements makes it easy, quick and painless for the end user. This is based on authors personal professional experience.

Business/Market drivers

Business side always has to be highly reactive, in just intime, technological economy – has a extremely high demand of speed and reactivity, hence sometimes perception develops that "without involving a central IT department to create flexible and innovative solutions. Self-reinforcing effects lead to an intertwinement of SIT with the organization." (Conceptualizing Shadow IT Integration Drawbacks). And this often results from several options on how business can fulfill IT need themselves: Implement solution autonomously or initiate proper change involving IT Department (Zimmermann et al., 2014)

A summary table of drivers is provided in

Table 2.

Table 2 Key drivers of SIT

Driver category	**Key factors**	**Primary sources**
Individual/Personal	Convenience, misalignment with job requirements, inertia	Zimmermann et al. (2014)
Organizational	IT-business misalignment, power aspects, individual behavior of key stakeholders, long IT response times, low perceived costs of SIT	Spierings et al. (2012), Kerr et al. (2007), Ortbach et al. (2013), Zimmermann et al. (2014)
Technological	SaaS services, easy sign-up, click-through agreements	Author's own observation
Business/Market	Speed and reactivity demands, self-reinforcing effects, autonomous solution implementation vs. IT involvement	Zimmermann et al. (2014)

2.1.3. Risks and organizational impacts of SIT

From a risk perspective, SIT commonly results in: data leakage and unauthorized sharing (e.g., through personal cloud storage or unmanaged SaaS), increased likelihood of malware or account compromise, non-compliance exposure to legal requirements, and exposure to operational fragility. Organizationally, SIT drives tool sprawl, duplicate spending, integration gaps, and inconsistent data - creating "shadow processes" that are hard to audit and hard to scale. It can also strain business–IT relationships: business units perceive IT as a blocker, while IT perceives uncontrolled risk - setting the stage for the later transition toward Shadow AI, where the same dynamics persist but with higher-speed, higher-impact decision automation and data exposure.

2.1.4. Transition from SIT to SAI

With emergence of AI starting at around 2023 with the release of back in 2023, that allowed for SIT to evolve and a new area of SAI has came into picture, especially as this new and still very young technology matures.”This phenomenon involves the use of AI tools and applications without formal organizational approval, mirroring the core dynamics of traditional Shadow IT but introducing far more complex risks." (Silic et al., 2025)”. While traditional SIT is somewhat well understood and managed phenomena by now, SAI is opaque and almost always operates as “Black box” principle – meaning we know what goes in, what goes out, but what happens in the middle – is complete unknown.

While SAI is part of SI, it brings different and quite unique challenges, therefore this thesis will keep them as separate topics, while still having in mind that SAI is just part of SIT.

Key differentiators of SAI from SIT include:

- Autonomous and generative nature – differs from other technologies as it can independently process and generate new content or make decisions based on known data (Silic et al., 2025)
- Opacity and unpredictability – unlike traditional SIT, that mainly consists of traditional tools that are just not explicitly authorized, SAI usually has no or little transparency on how the output has been achieved, nor what training data was used in training particular AI model. (Silic et al., 2025)
- Seamless integration – traditional SIT tools are composed of simple applications installed on user devices or usage of cloud services. SAI differs from this as it can be integrated with any of these tools (Silic et al., 2025), sometimes without the knowledge of the end user.
- Attribution and accountability - "Unlike traditional Shadow IT, Shadow AI tools can generate outputs, process sensitive data, and even automate decision-making—all without organizational visibility." (Puthal et al., 2025 – cited in Silic et al.)

2.2. Shadow AI Theoretical Foundations

This Chapter will review existing theoretical foundations of SAI, reviewing its definitions, origins within Shadow IT, explore the various factors of SAI occurrence, and what innovation factors are driving its existence.

2.2.1. Shadow AI Definition and characteristics

In many cases SAI is not yet a common or well understood term, especially as it is quite new term, mainly as evolution or subcategory from Shadow IT (Chin et al., 2025b). In order to advance the research further, a clear and unified understanding of this term and its boundaries is required.

Table 3 Definitions of SAI

Author, year	SAI Definition
(Chin, Li, Mirone, & Papa, 2025b)	“shadow AI usage could be viewed as a new type of unethical pro-organizational behavior (UPB) performed by employees who do bad things for good reasons “
(Puthal et al., 2025b)	“deployment and usage of AI tools, models, or systems by employees of an organization without explicit endorsement, supervision, or management by IT and cyber security departments is defined as Shadow AI”
(Xu, 2025)	“More recent work highlights the rise of “shadow AI,” or the

Author, year	SAI Definition
	unsanctioned use of third-party tools, which creates new risks related to security, liability, and governance “
	“AI adoption is defined here as the organizational decision and implementation process of integrating AI technologies into business operations, decision-making, and strategic activities”
(*The Ethical and Legal Implications of Shadow AI in Sensitive Industries: A Focus on Healthcare, Finance, and Education*, n.d.)	“Shadow AI refers to AI-driven tools and models implemented without regulatory oversight, often adopted by employees, third-party vendors, or decentralized teams without adhering to security protocols and compliance frameworks.”
(Silic et al., 2025b)	"Shadow AI refers to the informal or unauthorized use of AI tools within organizations, often beyond formal oversight."
	"The proliferation of artificial intelligence (AI) tools in organizations has given rise to 'Shadow AI'—the unsanctioned use of AI systems outside approved governance frameworks."

Within **Table 3** we can see some select quates in reference to definition of SAI phenomena. It is best summarized by (Silic et al., 2025b), being as simply any instance of informal or unauthorized usage.

But looking into this a bit deeper and understanding some of the additional components that make this Shadow AI construct:

- Deployment and usage of AI and(or) associated tools (Puthal et al., 2025b)
- Not having explicit allow from IT team (Puthal et al., 2025b), (Chin et al., 2025b)

Therefore, although other works focus on explicit “bad” or “good” factors of such usage, they generally agree that it’s just not explicitly authorized use. Of course, all other implications cannot be ignored either, such as Information Security, compliance, and general IT hygiene.

Therefore, this work will use the fallowing operational definition:

Shadow Artificial Intelligence (SAI) is the employee - or business-unit - initiated use or deployment of AI tools, AI models, or AI-enabled features to perform organizational tasks, where such use occurs without explicit authorization and/or outside established organizational governance controls (e.g., risk assessment, data handling rules, monitoring, and accountability mechanisms).

2.2.2. Drivers of Shadow AI emergence

In the section 2.2.1 we already established the definition of SAI, and in the section 2.1.2 discussed the drivers of SIT adoption. Here, we would take a much deeper look at the exact drivers that are driving SAI's emergence and continued use. Will be analyzing within the same segments for the sake of consistency and to be able to compare SIT and SAI ("Shadow AI Governance, Risk, and Organizational Resilience").

Individual/Personal Drivers

Factors for individuals, are quite similar to SIT, yet differ. AI allows for employees to solve problems in new ways, using new approaches (Executive Interview I3, Silic et al., 2025). Also due to ease of use of AI, as a lot of AI chat interfaces do not even require an account to use, employees start to use it, without even thinking it (Limor Kessem, cited in Paubox). In addition, AI quite usually allow

employees to fill in gaps where employees are not as strong in certain areas, allowing for much more efficient and faster work execution.

Organizational Drivers

To start, with AI and SAI being such a new phenomenon – there is severe lack of governance around it (The Ethical and Legal Implications of Shadow AI) and as with SIT – speed of IT teams and their approval processes have not increased significantly (Paubox, 2025)

Another striking issue – organizational awareness about the risks imposed by such tools. (Silic et al., 2025). According to (Silic et al., 2025) and their conducted survey, only an average rating of 4.4 was achieved when asking employees weather they are aware about existing policies and procedures in regards to AI usage

Technological Drivers

With advent of Generative pre-trained transformer (GPT) AI – it became massively more accessible. While ML and other technologies were available, these GPT basically levelled the playing field. "This democratization of AI has accelerated innovation, boosted productivity, and reshaped how knowledge work is performed. (Shadow AI Governance paper). In addition, AI tools have became widely available to the masses (Conflicting impacts paper).

Another issue also arises from already existing tools – those also now are getting embedded AI (Paubox, 2025) out of the box. Also with the COVID-19 shift to remote work another issue also arises – "Employees increasingly rely on personal devices and unauthorized software to maintain productivity." (IBM, 2024 – cited in Silic et al.)

Business/Market Drivers

Baseline AI technology is freely available to anyone, therefore any copetitors can also use it without restriction. So not using it – becomes a detriment to the organization (Xu, 2025), especially to use it in order to increase work efficiency. In addition, executive leadership is quite ecstatic about AI usage, and see it as massive market advantage (Shadow AI is outpacing our email security: 4 - 4 ( ). According to market research "69% of IT leaders feel pressured to adopt AI faster than they can secure it." (Paubox survey, 2025)

A summary table of drivers is provided in **Table 4**.

Table 4 Key drivers of SAI adoption

Driver Category	Key Factors	Primary Sources
Individual/Personal	Productivity seeking, cognitive load reduction, creativity, convenience, skill gap compensation, speed	Silic et al. (2025), Paubox (2025), Shadow AI Governance paper
Organizational	Absence of AI governance, slow approvals, no sanctioned alternatives, training gaps, awareness gaps, innovation-rewarding culture, governance drift	Silic et al. (2025), Paubox (2025), Shadow AI Governance paper
Technological	AI democratization, generative AI accessibility, low-code/no-code, embedded AI, cloud services, remote work, personal devices	Shadow AI Governance paper, IBM (2024), Conflicting impacts paper

Driver Category	Key Factors	Primary Sources
Business/Market	Competitive pressure, executive enthusiasm, fear of falling behind, 69% pressured to adopt faster than secure	Paubox (2025), AI adoption literature, Conflicting impacts paper

2.2.3. Typology of Shadow AI tools and applications

SAI typology can be evaluated in several ways: by delivery (how users interact with it), by function (its purpose), and by data interaction (the level of risk involved). Since this research primarily analyzes user behavior related to SIT/SAI, which stems from altered user behavior, it will focus on a delivery perspective.

- Public AI SaaS is the most accessible option of SAI, as many vendors and services offer AI for free, often without registration. Examples include Google (Gemini), OpenAI (ChatGPT), and others.
- Embedded AI in approved SaaS occurs when the AI service is integrated via an update by the provider. This makes the application an approved application, but since the AI service is not part of the original review process, it creates SAI usage.
- Browser extensions and add-ins are also popular, with various AI providers offering browser extensions for easier AI access. A good example of this is Microsoft's Edge browser, which has this feature built in (Microsoft's Copilot AI).
- Locally executed models — tech-savvy users can even download various AI models onto their devices or use software that leverages AI, thus creating SAI usage.

In summary, using a delivery-based typology provides a practical lens for this thesis, because it reflects how employees actually adopt and embed AI into daily work. This also allows a glance to potential enforcement mechanisms depending on whether AI is accessed as public SaaS, embedded inside approved platforms, via extensions, or executed locally.

Therefore, the remainder of this thesis will mostly focus on this - delivery-focused typology as the baseline structure for analyzing Shadow AI usage and its implications.

2.3. Risk and Threat Landscape of Shadow AI

SAI presents a distinct risk profile and landscape of risks. Unlike AI managed centrally or by IT, SAI often involves unclear, hidden data flows with limited visibility. This section will discuss the risks associated with SAI from five different perspectives.ngles.

2.3.1. Cybersecurity risks

When discussing SIT and SAI, one of the first questions IT professionals ask is about the information security risks these tools expose organizations to.

One of the primary and easiest to see risks is Data exposure (leakage) risk. Organisations may have implemented sophisticated DLP tools, but even these can not protect everything. Especially as this is new vector of information loss that has not yet been fully understood. In addition employees quite oftenly tend to just upload sensitive information even without thinking about it (Limor Kessem, cited in Paubox). And as AI processes such data, and it usually process such data vithout restriction, geneate

content, that directly influences business decisions while working in completely unregulated environment (Silic et al., 2025), easily amplifies any inherent and AI made risks exponentialy.

Secondly, SAI is difficult to detect, usually gets integrated with other products, thus generating third-party and supply chain risks. For example "The healthcare sector faces unique challenges with shadow AI usage. Clinicians use unauthorized AI diagnostic tools to improve patient outcomes, but these tools often lack HIPAA compliance and expose sensitive patient data to third-party cloud providers." (Silic et al., 2025 – coding example)

Remote work and BYOD possess another risk, especially as that is also one of the motivators for SAI usage, which creates another layer of separation and detection difficulty.

Lastly AI brings its own new threats like:

- Prompt injection attacks - "prompt injection attacks, where attackers try to manipulate an AI system into leaking sensitive data or taking harmful actions through crafted [...] content" (PAUBOX, 2025)
- Model manipulation - "an AI can be taught using data that has been strategically corrupted in what is called a model poisoning attack [...] hackers may have gained access to the model and altered it so that the predictions are not accurate or may alter the functionality of underlying mechanisms" (Puthal et al., 2025)
- AI system misuse/repurposing - "OpenAI's transcription tool, Whisper, despite advisories against its use in high-risk domains, has been widely adopted in clinical settings, raising concerns about patient safety" (Balogun et al., 2025)
- Recursive error amplification – “Small errors in complex systems can amplify over time, leading to large, negative consequences, especially in sustainability” (*Recursive Error Amplification → Term*, n.d.) as referenced by (Silic et al., 2025)
- Autonomous decision-making without oversight - "AI systems can autonomously process large datasets and make critical decisions with minimal human oversight." (Silic et al., 2025)

2.3.2. Data privacy and protection risks

One of the most immediate and high-impact risk categories of SAI is data privacy and protection and potential for its exposure. Because many AI tools operate as external services (or introduce new “invisible” data flows through embedded features), employees may unintentionally expose organizational information by pasting text into prompts, uploading files, or connecting internal systems to AI-powered automation. In contrast to SIT, SAI can process and regenerate sensitive content at high speed, increasing both the volume and reach of potential data leakage (Puthal et al., 2025).

2.3.3. Compliance and legal risks

SAI creates compliance and legal risk because it introduces unapproved processing of data, unvetted third parties, and undocumented decision support into business workflows. Even when the intent is productivity, the absence of formal review means AI usage can easily bypass required controls such as vendor due diligence, data processing agreements, retention rules, and regulatory accountability. As a result, organizations may be unable to demonstrate who processed what data, on what legal

basis, using which provider, and under what safeguards - creating both compliance gaps and legal exposure (Balogun et al., 2025).

2.3.4. Operational and reputational risks

As it is quite usual, SAI creates quite unique risks in regards to Operational and reputational risks. Previously discussed data loss risk has implications on reputational risks. But even higher risks is if the data is not being disclosed, but retained by the 3rd party AI provider. These risks become more severe in sectors with strict confidentiality and regulatory obligations, for example:

- Healthcare – patient data that is closely protected by various regulations like HIPAA (Health Insurance Portability and Accountability Act) or GDPR (General Data Protection Regulation).
- Financial services – protection of financial records, regulated by standards like PCI-DSS.
- Legal – information covered under attorney-client privilege or in highly sensitive cases (protected by law).
- HR data – personal, private information.

Another risk is inconsistency and lack of transparency. AI companies are unwilling to be open about what occurs within their models and have also been shown to apply their controls unevenly. Silic et al., 2025 has mentioned “Furthermore, transparency emerged as a critical issue, as the average rating for transparency in AI use was only 4.0 [author note: in a scale of 1 to 10], suggesting that many employees feel their organizations do not effectively communicate how AI is being used internally.”

Lastly data integrity also becomes a risk as well. Despite massive efforts and attempts to reduce it, AI hallucinations is still a massive issue, as AI models still attempt to satisfy requests at any costs, even generating fake information to satisfy that need (Silic et al., 2025). "While AI can help speed up certain processes, employees need to remain critical of its outputs, as mistakes can be subtle but highly consequential." (Executive Interview I1, Silic et al., 2025). Also, Recursive usage of AI is an issue as well, that causes self-re-enforcing behavior and complicates fishability of auditability. (Silic et al., 2025)

2.3.5. Ethical, bias, and intellectual property risks

Proliferation of AI creates new and novel Ethical and Intellectual property risks As one of the biggest issues with AI is – Data. The need for it is massive and various AI companies have already resorted to stealing copyrighted works without any attribution or remorse, completely disregarding any ethical obligations, just to fulfil data needs. Companies as big as Meta, Nvidia, OpenAI and others have already been implicated in various public announcements of data theft or even been named as defendants in various civil cases.

Another issue is AI generate content ownership and its attribution. While legal implications have already been discussed within section 2.3.3 as Silic et al., 2025 has expressed “This raises problems of attribution and control, as responsibility for biased, harmful, or incorrect outputs becomes blurred.” This also erodes at human creativity and Critical thinking, by just doing the hard work for the person (Silic et al., 2025)

2.4. IT Governance Theoretical Foundations

This section will review established IT governance and Information security governance frameworks and any additional theoretical foundations and take it as foundational knowledge suggestions on SAI/SIT management. The emphasis is placed on the IT governance perspective as a baseline for later constructing the conceptual SAI governance framework.

2.4.1. IT governance concepts and principles

IT governance as defined by (Weill & Ross, 2004), is “The framework of decision rights and accountability to encourage desirable behavior in the use of IT “. This basically means that a framework, that includes leadership, organizational structures and processes that ensure that IT is an extension of organization strategy and allows/helps in execution of it. Therefore, it is vital for organizations to have a strong IT governance framework, as it can allow or deny fulfillment of organizational strategy.

"The lack of a good alignment between business and IT creates an ideal context among human factors for the creation of shadow systems." (Shadow IT – A view from behind the curtain). As an example ISO/IEC 38500 (Information technology — Governance of IT for the organization) defines several fundamental principles of IT governance.

Effective IT governance is therefore not merely a technical concern but an organizational design problem. It requires the coordination of leadership structures, decision-making processes, and oversight mechanisms to ensure that technology use supports, rather than undermines, organizational strategy. This understanding is essential for SAI governance: the emergence of Shadow AI is not primarily a technology failure but a governance failure - a breakdown in the mechanisms through which organizations define, communicate, and enforce acceptable technology use.

ISO/IEC 38500 (Governance of IT for the organization) provides one of the most widely referenced codifications of IT governance principles. It defines six principles - responsibility, strategy, acquisition, performance, conformance, and human behavior - each of which has direct implications for SAI management. **Table 5** summarizes these principles and their relevance to SAI.

Collectively, these six principles reveal that effective IT governance operates as an integrated system: it assigns decision rights (Responsibility), aligns those decisions with organizational purpose (Strategy), controls how technology enters the organization (Acquisition), monitors its effectiveness (Performance), ensures legal and regulatory compliance (Conformance), and accounts for the human factors that determine whether governance is followed or circumvented in practice (Human Behavior). SAI, by definition, bypasses all six. This is precisely why it represents a governance failure rather than merely a technology management challenge.

Table 5 ISO/IEC 3850 principles for IT governance

ISO/IEC 38500 principle	**Description**	**Relevance to Shadow AI handling (SAI)**
Responsibility	Responsibilities for IT-related decisions and actions are clearly assigned and understood	Stops “everyone uses it, no one owns it.” Mentality. Suggests that SAI tools need to have defined owners for: AI tool approval/denial

ISO/IEC 38500 principle	Description	Relevance to Shadow AI handling (SAI)
		Use-case ownership, data handling rules, exception approvals, monitoring, and incident response for AI misuse/data leakage must be established with clear lines of responsibility and accountability.
Strategy	IT use is aligned with the organization's strategy and supports current and future business needs.	SIT and SAI is a governance gap and not just a IT issue. Establishing organizations stance on AI, and align controls and risk appetite enables organization to reach innovation or other strategic goals
Acquisition	IT investments/acquisitions are made for valid reasons, based on appropriate analysis, and with clear expected value and risk consideration.	Directly targets "shadow procurement" (method of how payed SAI tools are acquired) Build an approved AI tools catalog; require vendor due diligence, this limits the risks of users just going out and acquiring any tool they want
Performance	IT performs well and delivers the services and outcomes required to meet business needs	Demands SAI outputs to require validation for high-impact outputs, defining quality/accuracy thresholds, tracking hallucination/defect rates, and monitor business process impact (rework, errors, downtime, customer impact).
Conformance	IT conforms with applicable laws, regulations, and internal policies/standards.	Prevents accidental regulatory exposure via SAI. Enforce data classification rules for prompts/uploads, ensure legal, regulatory and contractual requirements are met.
Human behavior	IT governance respects human factors - how people work, incentives, capabilities, and the social/ethical impact of IT use.	Explains why bans fail and how SAI emerges. Focus on usable approved tools, training, clear guidance, and transparency while reducing workarounds and addressing overreliance, bias risks, and accountability for AI-assisted decisions.

2.4.2. AI-specific governance standards

ISO/IEC 42001 (Information technology — Artificial intelligence — Management System) is one of the latest and newest industry standards, that directly governs requirements and provides guidance for establishing, implementing, maintaining, and continually improving an Artificial Intelligence Management System

In practice, ISO/IEC 42001 directly governs and guides organizational structure around AI through elements such as:

- Policies, objectives, and accountability for AI use and oversight
- Risk management and impact assessment across AI use cases and lifecycle
- AI system lifecycle management (from design/deployment to monitoring and improvement)
- Third-party / supplier governance for AI services and dependencies

(Reto P. Grubenmann, n.d.)

While ISO/IEC 42001 represents a significant step forward in formalizing AI governance, it has an important structural limitation: it is designed to govern sanctioned AI deployments—systems that the organization has knowingly adopted and integrated into its management system. SAI, by definition,

exists outside this scope. ISO/IEC 42001 provides no mechanisms for discovering, assessing, or managing AI tools that employees have adopted without organizational awareness. This gap is not a flaw in the standard itself, which effectively serves its intended purpose, but it highlights the need for a governance framework that specifically addresses the shadow side of AI adoption.

2.5. Regulatory and Compliance Environment

SAI governance does not operate in a regulatory vacuum. Organizations are subject to an evolving set of legal and regulatory requirements that directly constrain how AI tools may be used, what data may be processed, and what obligations of transparency and accountability apply. This section examines the three most relevant regulatory instruments for the purposes of this research: the EU Artificial Intelligence Act, the General Data Protection Regulation, and sector-specific compliance requirements.

2.5.1. EU AI Act and its implications

The EU AI Act, which entered into force in August 2024 with phased implementation through 2026, represents the first comprehensive legislative framework for AI governance globally. Its central mechanism is a risk-based classification system that categorizes AI systems into four tiers: unacceptable risk (prohibited), high risk (subject to strict requirements), limited risk (transparency obligations), and minimal risk (no specific regulation). High-risk AI systems are subject to requirements for data governance, technical documentation, transparency, human oversight, accuracy, robustness, and cybersecurity.

The implications for SAI are direct and significant. When employees deploy AI tools outside formal governance structures, the organization has no mechanism to determine whether a given tool or use case falls within the high-risk category and therefore triggers regulatory obligations. The organization cannot demonstrate compliance with documentation, oversight, or transparency requirements for AI systems it does not know are in use. The EU AI Act thus transforms SAI from a risk management concern into a potential regulatory violation - organizations may be non-compliant without knowing it, simply because AI usage is occurring outside their visibility.

2.5.2. Data protection regulations (GDPR)

The General Data Protection Regulation (GDPR), in force since 2018, remains directly relevant to SAI governance because many AI tools process personal data. For example, when an employee pastes customer information, employee records, or other personal data into an external AI service, several GDPR requirements are implicated: the need for a lawful basis for processing, the requirement for a data processing agreement with the AI service provider, the obligation to inform data subjects about how their data is processed, and the requirement to ensure appropriate technical and organizational safeguards.

Under SAI conditions, none of these requirements can be fulfilled. The organization lacks a data processing agreement with the AI vendor, has not conducted a data protection impact assessment, cannot inform data subjects about the processing, and cannot demonstrate the technical safeguards mandated by GDPR. Similar to the EU AI Act, the main issue is that organizations cannot prove

regulatory compliance for data processing they do not have visibility into. Therefore, SAI represents a fundamental compliance gap, not just an operational risk. risk.

2.5.3. Industry-specific requirements

Beyond mentioned regulations, specific sectors impose additional data protection and AI governance obligations that SAI may violate.

In healthcare, regulations such as HIPAA (Health Insurance Portability and Accountability Act) in the United States impose strict controls on the processing, storage, and transmission of protected health information.

In financial services, requirements such as PCI-DSS (Payment Card Industry Data Security Standard) govern the handling of financial data.

In legal practice, information covered under attorney-client privilege is subject to confidentiality protection that unsanctioned AI processing may breach.

The common thread across all of these is that they require organizations to maintain documented, auditable control over how sensitive data is processed and by whom. SAI fundamentally undermines this requirement by introducing data processing pathways that are undocumented, unaudited, and invisible.

2.5.4. Regulatory implications for SAI governance

The regulatory landscape sets a clear baseline: any governance framework for SAI must enable the organization to identify where AI-related data processing occurs, assess its regulatory implications, and demonstrate compliance to regulators. Governance that reduces risk but cannot prove compliance remains legally inadequate.

2.6. Innovation and Control Balance

Organizations face a fundamental "Yin-Yang paradox" (Chin et al., 2025b) where Shadow AI simultaneously drives productivity and innovation while introducing security vulnerabilities, data leakage, and compliance risks that traditional governance cannot address. Restrictive policies often backfire - employees circumvent bans using personal devices, creating a "governance drift zone" where formal policies exist but lack real-world traction. Effective governance therefore requires shifting from prohibition to controlled enablement: offering sanctioned AI alternatives, AI tool registries, and role-specific training that channels employee innovation needs productively rather than suppressing them.

Reviewed existing literature, primarily identifies three practical mechanisms that reduce SAI not through prohibition but by reducing the conditions that cause it:

- sanctioned AI alternatives - providing approved tools that are genuinely comparable to what employees find on their own
- fast-track and tiered approval processes - eliminating the speed mismatch between AI adoption in minutes and IT approval in weeks

– experimentation sandbox environments - allowing controlled exploration without exposing production data.

Therefore, effective SAI governance must serve a dual purpose - be strict enough to manage risk and meet regulatory requirements, while also being flexible enough to support legitimate innovation needs that promote shadow adoption. A framework that achieves only one of these goals will either be bypassed or prove inadequate. This dual necessity directly influences the design of the conceptual framework in the next chapter - governance and enablement are not competing goals but are interconnected.

2.7. Conceptual Framework/Model Creation

This chapter introduces the conceptual SAI Governance Framework (SAIGF), created as a synthesis of the theoretical foundations, risk analysis, governance principles, regulatory requirements, and innovation-control dynamics examined in Chapter 2. The framework is designed to provide organizations with a structured approach to governing SAI that balances risk management and regulatory compliance with innovation enablement.

2.7.1. Framework development approach

The existing literature review conducted in Chapter 2 established several findings that collectively justify the development of a new governance framework rather than the direct application of existing ones.

1. SAI represents a qualitatively different governance challenge from traditional Shadow IT. Its generative, opaque, and autonomous nature introduces risks - including data privacy exposure, algorithmic bias, hallucination, and governance drift - that are not adequately addressed by conventional IT governance controls (Silic et al., 2025b).
2. Existing governance standards such as ISO/IEC 38500 and ISO/IEC 42001 provide valuable structural principles but are designed for sanctioned technology and sanctioned AI respectively; neither addresses the shadow dimension in which AI usage precedes and bypasses governance.
3. The regulatory environment - particularly the EU AI Act and GDPR - create compliance obligations that organizations cannot meet for AI usage they cannot see, transforming SAI from an operational risk into a legal exposure.
4. Prohibitive governance approaches have been shown to be counterproductive: they drive AI usage underground rather than eliminating it, reducing visibility without reducing risk (Chin et al., 2025b)

These findings define the design requirements for the framework:

– It must address the full lifecycle of SAI - from discovery through management and enablement.
– It must bridge the gap between IT governance principles and the practical realities of employee AI adoption.
– It must satisfy regulatory compliance obligations.
– It must balance control mechanisms with innovation enablement to ensure practical sustainability.

2.7.2. Framework architecture

The SAIGF is structured as three interdependent layers, each necessary but insufficient on its own. The three layers correspond to the three foundational issues of SAI governance: what must be in place before governance can function (Foundation), how governance operates in practice (Core Mechanisms), and what makes governance sustainable and acceptable over time (Enablement).

Table 6 Overview of the SAIGF three-layer architecture

Layer	Components	Purpose
Layer 1 Governance Foundation	1. Governance roles and accountability 2. Regulatory and compliance alignment 3. Risk appetite definition	Establishes the structural prerequisites without which operational governance cannot function: who decides, what rules apply, and how much risk is acceptable.
Layer 2 Core Governance Mechanisms	1. Visibility and discovery 2. Policy and authorization framework 3. Risk assessment and management 4. Technical controls	Provides the operational backbone: the mechanisms that identify SAI, define acceptable use, evaluate risk, and enforce compliance through technical and procedural safeguards.
Layer 3 Innovation Enablement and Sustainability	1. Sanctioned AI provision 2. Fast-track and tiered approval 3. Culture, awareness, and continuous improvement	Ensures governance is sustainable by addressing the innovation needs that drive SAI, maintaining employee awareness, and creating feedback loops for ongoing adaptation.

The layered structure reflects a deliberate design logic. Layer 1 must be established before Layer 2 can function: technical controls and policies cannot be designed without knowing who owns governance decisions, what regulatory obligations apply, and what risk appetite governs control calibration. Layer 2 must be operational before Layer 3 can be effective: sanctioned alternatives and awareness campaigns have limited value if the organization has no visibility into what AI is being used and no policy framework against which compliance can be assessed. Layer 3, in turn, feeds back into Layers 1 and 2 by generating user feedback, evolving the risk landscape, and driving continuous improvement.

2.7.3. Layer 1: Governance Foundation

The foundation layer establishes the structural prerequisites for all subsequent governance activity. Without these elements, operational governance mechanisms lack the authority, direction, and calibration needed to function effectively. More detailed requirements are explored within Table 7 SAIGF Layer 1 explanation table.

Table 7 SAIGF Layer 1 explanation table

Component	Explanation/Justification
Governance roles and accountability	Explicit assignment of governance roles, decision rights, and accountability for AI-related activities across the organization. Taking from ISO/IEC 38500 Responsibility principle and the governance structure literature, this component requires organizations to define: – who has authority to approve or deny AI tools for organizational use – who owns the risk when an AI tool is deployed (whether sanctioned or discovered as shadow) – who is responsible for monitoring compliance – who leads incident response when AI-related incidents occur.
Regulatory and compliance alignment	requires the organization to map its AI usage obligations against applicable regulatory requirements - like the EU AI Act, GDPR, etc. and any sector-specific regulations - and to establish a minimum compliance baseline that governance must satisfy. This mapping defines the non-negotiable requirements that all other governance components must serve. It transforms regulatory compliance from an abstract obligation into a concrete set of governance design constraints: what documentation must be maintained, what assessments must be conducted, what transparency obligations must be met, and what data handling rules must be enforced.
Risk appetite definition	requires the organization to clearly state its position on AI-related risk: how much uncertainty is acceptable, which risk categories are tolerable, and where zero-tolerance applies. Risk appetite serves as the calibration tool for Layer 2—it determines how strict technical controls need to be, how thorough the approval process is, and what threshold differentiates low-risk AI use (requiring minimal governance) from high-risk AI use (needing full assessment). Without a defined risk appetite, governance controls tend to be either overly strict (causing shadow adoption by creating unnecessary friction) or too permissive (failing to address real risks).

2.7.4. Layer 2: Core Governance Mechanisms

The second layer contains the operational mechanisms through which governance is enacted. These are the components that directly identify, assess, regulate, and enforce AI usage within the organization. More detailed requirements are explored within **Table 8** SAIGF Layer 2 explanation table

Table 8 SAIGF Layer 2 explanation table

Component	Explanation/Justification
Visibility and discovery	Visibility is the prerequisite for all governance activities - an organization cannot govern AI usage it does not know exists. This component encompasses both technical and procedural mechanisms for identifying what AI tools are in use, by whom, for what purposes, and with what data. Technical approaches include: – network traffic analysis – Cloud Access Security Broker (CASB) deployment – Data Loss Prevention (DLP) monitoring – AI-specific discovery tools.

Component	Explanation/Justification
	Procedural approaches include: – self-declaration mechanisms – team-level AI usage inventories – integration of AI tool questions into existing audit and review processes.
Policy and authorization framework	This component provides the formal ruleset governing AI usage: what is permitted, what is prohibited, what requires approval, and what conditions apply. It encompasses acceptable use policies, a tool approval process with defined criteria and timelines, and an organizational AI tool registry that classifies tools by approval status and risk category. Policies must be written in accessible language, communicated through multiple channels, and designed to be findable and understandable by employees outside of IT and governance functions.
Risk assessment and management	This component provides a structured process for evaluating the risk profile of AI tools and use cases, aligned to the risk appetite established in Layer 1. The assessment should cover at a minimum: – data sensitivity - what data is shared with the AI tool and how it is processed and retained – security posture - authentication, encryption, and access controls of the AI service – compliance status - regulatory implications of the specific use case – operational dependency - what business processes depend on the AI tool's output – output reliability - the consequences of AI errors, hallucinations, or bias in the specific context of use Risk assessment should be proportionate and tiered. Low-risk use cases - such as using a general-purpose AI assistant for non-sensitive text editing - should be subject to a lightweight, rapid assessment. High-risk use cases - such as processing personal data, generating customer-facing content, or informing business decisions - should trigger full assessment with documented outcomes. This tiered approach is essential for practical sustainability. Applying the same rigor to all use cases creates the bureaucratic friction that drives shadow adoption.
Technical controls	Technical controls provide an enforcement layer that ensures policy compliance through technological means rather than relying solely on employee behavior. This component encompasses – Data Loss Prevention (DLP) tools configured to detect and prevent sensitive data from being shared with unauthorized AI services – network security controls, including CASB platforms that can identify and manage AI tool access – access control and authentication mechanisms for sanctioned AI tools – monitoring and logging technologies that create audit trails for AI-related activities – and AI-specific safeguards such as prompt filtering, output validation, and sandbox environments that contain AI interactions within controlled boundaries. Technical controls are necessary but not sufficient. Existing literature consistently emphasizes that purely technical approaches to SAI management are limited by the speed of AI tool evolution, the diversity of access pathways, and the practical impossibility of blocking all unsanctioned AI usage without simultaneously blocking legitimate productivity tools. Technical controls are most effective when combined with the cultural and enablement mechanisms in Layer 3.

2.7.5. Layer 3: Innovation, Enablement and Sustainability

The third layer addresses the conditions that determine whether governance is sustainable over time. Without these components, governance mechanisms from Layers 1 and 2 degrade: employees find workarounds, policies become outdated as AI tools evolve, and the gap between governance intent and organizational reality widens. More detailed requirements are explored within Table 9 SAIGF Layer 3 explanation table

Table 9 SAIGF Layer 3 explanation table

Component	**Explanation/Justification**
Sanctioned AI provision	This component requires the organization to maintain an actively curated catalogue of approved, pre-vetted AI tools that are available for employee use. The catalogue should be organized by use case (text generation, data analysis, coding assistance, image generation, etc.) and should include clear guidance on what each tool may and may not be used for. Sanctioned tools must be genuinely competitive with the unsanctioned alternatives employees would otherwise find - if the approved tool is markedly inferior in capability, usability, or speed, employees will continue to seek shadow alternatives regardless of policy.
Fast-track and tiered approval	This component addresses the speed mismatch between AI tool adoption and traditional IT approval processes. It requires the creation of a tiered approval mechanism that differentiates between risk levels and provides appropriately rapid decisions for each. A low-risk AI tool - for example, a general-purpose writing assistant that does not process sensitive data - should be approved within days, not weeks. A high-risk tool that processes personal data or generates outputs used in regulatory contexts should be subject to full assessment but with a defined timeline and clear communication with the requestor. While similar control on layer 2 focuses on the review itself, this control focuses primarily on the end user facing side.
Culture, awareness, and continuous improvement	Awareness and education - regular, practical, role-relevant communication about AI risks, data handling rules, and the availability of sanctioned tools. Training should be short, specific, and framed around practical scenarios rather than abstract policy recitation. Trust and transparency - governance mechanisms should be communicated openly, including the rationale behind restrictions and the criteria used for tool approval. Employees who understand why controls exist are more likely to comply than those who perceive governance as arbitrary or uninformed. This component also requires feedback mechanisms through which employees can report governance friction, suggest improvements, and request new tools. Continuous improvement - the AI tool landscape evolves rapidly. Governance mechanisms, approved tool catalogues, risk assessments, and policies must be reviewed and updated on a regular cycle - not as a one-time exercise but as an ongoing operational process. This component includes post-incident review mechanisms, periodic governance effectiveness assessments, and integration of employee feedback into governance evolution.

The SAIGF is proposed as a theoretical model to be validated and refined through empirical research. As a synthesis of Shadow IT foundations, AI-specific risk analysis, established IT governance principles, regulatory requirements, and innovation-control dynamics, the framework represents a structured response to the governance gap identified in Chapter 1. It does not claim to be exhaustive - the SAI landscape is evolving, and any framework must be treated as a living instrument rather than a fixed prescription. What it does provide is a principled starting point: a layered architecture that connects governance prerequisites, operational mechanisms, and sustainability conditions in a way that neither purely technical nor purely policy-based approaches have achieved in isolation.

Chapter 3 will focus on the empirical phase of the research, explaining the methodology used to test the framework in real-world settings and refine it based on practitioner evidence.

3. Shadow Artificial Intelligence Empirical Research design

To validate the conceptual model and identify potential gaps, the following research methodology was developed. This methodology is expected to validate and improve upon the created conceptual model, and especially to make sure that it is grounded in practice and takes into account real-world scenarios.

3.1. Research Goal and Objectives

The goal of empirical research is to validate the conceptual model for managing SAI and ensure that the model is not only grounded in practice but also considers real-world scenarios. Additionally, it aims to provide potential directions for further research.

Objectives of the empirical research are as follows:

1. Prepare the methodology for two types of semi-structured interviews (management and non-management levels) that will facilitate a deeper practical understanding.
2. Validate and ground conceptual model in the practical implications of managing SAI
3. Disclose any and all associated challenges, risks, and opportunities seen in practice.
4. Provide management-level recommendations on designing SAI management programs that are accepted by all levels of employees within a company. In addition, highlight any future academic research directions, or highlight areas of under-research.

These objectives will assist in verifying the SAI management model and grounding it in real-world scenarios and related restrictions.

3.2. Research Design

This study follows an interpretivist epistemological approach. In this view, social reality is understood as something shaped through the meanings people give to their own experiences. Because of that, to understand these meanings, the research focuses on participants' subjective perspectives rather than on measuring reality as a fully objective phenomenon (Creswell, 2014). This is appropriate for SAI research, because SAI phenomenon is still an emerging phenomenon and it is highly context dependent, different organizations' responses to it vary considerably across sectors and structures, and the existing literature explicitly acknowledges that governance failures are as much a product of culture, perception, and behavior as they are of policy design (Silic et al., 2025b).

Keeping with that, a qualitative research approach was selected. Qualitative methods are particularly well-suited to exploration and framework-validating research in areas where theoretical development is still at an early stage (Clark et al., 2021). This approach allows for the collection of information about a highly nuanced and contextual topic, where a quantitative research model would fail to capture the necessary depth or nuance of the information presented.

Semi-structured interviews were selected as the primary data collection method. Semi-structured interviews offer a balance between thematic consistency across participants and the flexibility to pursue emergent topics and unanticipated lines of inquiry (King, 2014). This flexibility is particularly valuable given the novelty of SAI as a governance domain: participants may have encountered governance challenges or workaround dynamics that existing frameworks do not yet capture. A fully

structured instrument would risk foreclosing precisely the insights that the empirical phase is intended to generate.

The overall research design is two-phased. The first phase, documented in Chapter 2, involved a systematic analysis of academic and practitioner literature, resulting in the construction of a theoretical SAI governance framework. The second phase subjects the created theoretical framework to real empirical scrutiny. This sequence reflects the logic of framework development research, in which theoretical construction precedes empirical validation and refinement (Yin, 2018).

Quite a distinct feature of this research (from research that has been reviewed in this field) is the deliberate inclusion of how different groups – Managers (the decision and governance makers) and employees who are just the consumers of AI tools in professional contexts. The literature has predominantly captured the managerial and governance design perspective; the employee perspective, their motivations, their awareness of controls, and their experience of friction remain substantially underdeveloped. Current research primarily stems from the background established by research into SIT, well before the introduction of AI and its associated phenomena. In addition, as this research is also presenting a SAI management phenomenon, including both sides of it is beneficial in crafting beneficial frameworks that are grounded in real-world restrictions.

3.3. Research Sampling

Participants were selected using purposive sampling, a non-probability strategy in which individuals are deliberately chosen based on their relevance to the research questions (Michael Quinn Patton, 2002). Purposive sampling is standard practice in qualitative interview-based research. For this study, it is directly relevant to review firsthand experiences of attempting to use and manage SAI and its tools, to work around company restrictions, and to examine how users (employees) are coping with these restrictions.

Manager level respondent selection: These employees have direct responsibility for AI governance decisions. C-level personnel IT (or related field) personnel and CEOs, IT directors and team leads, Cybersecurity (Information Security) team leads, Senior managers involved in AI tool adoption decisions and their usage, data governance, or information security policy personnel. Cross-sector representation is desirable, and a target of six to eight participants is set for this group.

Non-manager level (employees) respondent selection: This group consists of personnel who do not fit the "Manager level respondent" selection criteria, also includes non-IT, non-governance employees who use AI tools within a professional context, regardless of internal company policies and the specific tool approval status. Key inclusion criteria are that these personnel are using or have used AI tools in a professional context, and the usage of the tool has been deliberate and willful. Sector (employment and employer) diversity within this group is considered particularly valuable, as patterns of employee workaround behavior tend to be more universal than sector-specific governance architectures. A target of six to eight participants is set for this group.

Note on sector diversity – as this study can be highly influenced by sector of the company and the sector that employee works in and the experiences in both can be radically different, diversity in both is highly valued. Both of mentioned factors will be noted and taken into account.

The target combined sample size of twelve to sixteen interviews is appropriate this level qualitative research and consistent with established guidance for semi-structured interview studies in organizational contexts(Guest et al., 2006). However, the final sample size is governed by the principle of theoretical saturation - data collection within each group continues until new interviews cease to produce meaningfully new themes or categories Strauss & Corbin (1998). This approach allows for flexibility without arbitrariness in sample size determination.

Participant access for the managerial group is facilitated by the researcher's professional background within the cybersecurity space, which provides established networks within IT and security communities. For the employee group, participants are recruited through professional and personal networks, with care taken to ensure recruitment is independent of any organizational gatekeepers. This is essential to the integrity of the employee sample, as participants must feel free to speak candidly about workaround behavior without concern about organizational consequences.

All participants receive a written information sheet before the interview, explaining the purpose of the research, their right to withdraw at any time, and the anonymization procedures used for all data. No personal, organizational, or sector identifiers are included in the research outputs. Informed consent is obtained in writing prior to each interview.

3.4. Data Collection and Analysis

In order to collect the necessary data from the interviewees, a single interview guide has been developed. The interview guide contains both managerial and non-managerial level questions. In addition to allowing information to be collected from various levels of personnel, an interview guide has been prepared with questions in both Lithuanian and English. Interviewees were given both language options to choose from. Interview guide has been based guidelines of (Galletta, 2020), interviews have been split out to three segments – opening, middle, and closing.

The opening segment served as the introductory phase of each interview, with the primary purpose of establishing rapport between the interviewer and the respondent, introducing the research topic and its scope, confirming informed consent, and clarifying the voluntary and confidential nature of participation. This segment also included background questions designed to contextualize the respondent's professional role and organizational setting, ensuring that subsequent questions could be interpreted appropriately.

The middle segment constituted the core data collection phase and accounted for the majority of the interview duration. Questions in this segment were structured around the principal thematic domains of the research - including SAI prevalence, governance mechanisms, risk management, the innovation-control tension, and monitoring and enforcement - and were directly aligned with the components of the theoretical governance framework developed in Chapter 2.7. This segment was intentionally semi-structured, allowing the interviewer to explore emergent topics and pursue relevant threads beyond the prepared questions, which added analytical value.

The closing segment provided an opportunity for respondents to offer forward-looking reflections, raise topics not addressed during the middle segment, and contribute any additional insights they considered relevant. It also served as the professional close of the interview, during which next steps - including the anonymization of data and the potential sharing of research findings - were confirmed with each respondent.

All interviews have been conducted either in person or with the use of communication equipment/software, like Microsoft Teams. Both types of interviews have been recorded and transcribed with respondents' explicit agreement. All interviews were completed from April 1st, 2026, to April 30th, 2026. Each interview had a one-hour time slot allocated.

Afterward, all automated interview transcripts were reviewed alongside the audio recordings to correct any errors or inconsistencies. All transcripts were then imported into MAXQDA for structured qualitative analysis. Coding was performed deductively using a predefined code system based on the components of the theoretical SAIGF framework. In addition, each coded segment was simultaneously assigned three attributes:

- mode, capturing the type of statement made - whether descriptive of current state, evaluative of how well something functions, normative in prescribing what should exist, or speculative about future or counterfactual conditions
- Grounding, capturing the evidential basis of the statement - whether first-hand experiential, observational of colleagues, secondhand from others, or reflexive in acknowledging the participant's own knowledge limits
- Valence, applied conditionally to evaluative, normative, and judgment-bearing speculative segments, distinguishing between positive, negative, mixed, and neutral assessments.

Context codes were additionally applied to anchor segments to participant role, sector, AI usage patterns, and organizational governance stance. Where segments carried analytical content that could not be put under any SAIGF thematic code without distortion, provisional inductive codes were applied and flagged for review as candidates for framework refinement. During the second pass interview, those were either removed, combined, or established as concrete inductive points. Following individual transcript coding, interviews were grouped into managerial and non-managerial level sets and analyzed separately to enable systematic cross group comparison.

A multi-dimensional coding system was chosen to capture not only what each statement was about, but also the context in which it was made - whether the participant was describing, evaluating, or speculating, and whether they spoke from direct experience or observation. Drawing on King, (2014)template analysis approach, this allowed segments to be analyzed across several axes rather than being reduced to a single flat code.

The expected outcome of these interviews is to validate the theoretical model for SAI management, while also aiming to gather additional insights and help ground the model in real-world experiences.

4. Findings of the Empirical Research

In the following, chapter SAI usage patterns and different approaches by companies and personnel, along with information from interviews and their documentation, will be analyzed and reviewed. Based on these results, the conceptual SAIGF will be adjusted to better reflect real-world companies and ensure that recommendations remain relevant and practical.

4.1. Overview of empirical research results

Ten participants from six organizations were recruited through purposive sampling and interviewed between 1st and 30th April 2026. The sample achieved is below the lower end of the planned range (twelve to sixteen) but consistent with the principle of theoretical saturation: the final two interviews produced no new SAIGF-level themes, and only one provisional inductive theme that did not recur across cases.

Six participants belong to the managerial-level group (decision-makers with direct or indirect responsibility for AI tool adoption, governance design, or information security policy) and four to the non-managerial group (employees who use AI tools in a professional context but do not own governance design). The split reflects access realities rather than a deliberate weighting: the researcher's professional networks in cybersecurity provided stronger access to managerial respondents.

It should also be noted that recruitment posed significant challenges that influenced the final sample composition. At the managerial level, many potential participants declined to take part, citing concerns about organizational confidentiality and governance arrangements, security policies, and AI tool decisions, which were often viewed as sensitive internal matters that respondents were unwilling or unauthorized to discuss with an external researcher. This primarily limited the size of the managerial group beyond what professional networks could otherwise provide.

At the non-managerial level, recruitment challenges were different. Many potential participants hesitated to openly discuss AI tool use that might go against their organization's official policies, even when assured of full anonymity - the fear of admitting to workaround behavior, even if common, led more cautious employees to opt out. Another issue was finding non-managerial participants who used AI tools enough to provide meaningful insights: some potential respondents either rarely used AI in real life or didn’t realize that the tools in their daily routines counted as AI use relevant to this research.

Table 10. Participant sample characteristics

ID	Group	Role	Organization	Organization Sector	Organization AI stance described	Interview duration (recorded segment) in minutes	Interview Language
R01	Manager	Security operations lead	C1	Electronics manufacturing	Enabling, policy-immature	40	EN
R02	Non-manager	Cybersecurity analyst	C1	Manufacturing	Cautious, restrictive	50	EN

ID	Group	Role	Organization	Organization Sector	Organization AI stance described	Interview duration (recorded segment) in minutes	Interview Language
R03	Non-manager	Information Security engineer	C2	IT services	Strongly enabling, risk-aware	40	EN
R04	Non-manager	Information security engineer	C1	Manufacturing	Unclear, enabling but uncertain	30	EN
R05	Manager	Digital innovation manager	C1	Electronics manufacturing	Mixed, in transition	40	EN
R06	Manager	Transport / logistics manager	C3	Logistics	Strategically enabling, cautious	40	LT
R07	Manager	Co-Owner	C4	Creative comms / foodservice	Enabling but informal	30	LT
R08	Manager	Team lead / senior data engineer	C5	IT outsourcing	Enabling, client-mediated	40	LT
R09	Manager	Executive (Baltics)	C6	Security and defence	Controlled, treats AI as tool	45	EN
R10	Non-manager	Sales support engineer	C1	Electronics manufacturing	Enabling but uncertain ("grey zone")	30	LT

4.1.1. Organizational contexts

With the sample composition established, subsequent subsections briefly overview each participating organization, as organizational context is needed to interpret the findings.

4.1.1.1. C1 - Large multinational electronics manufacturer

C1 is the most heavily represented organization in the sample, contributing five participants (R01, R02, R04, R05, R10). It is a globally distributed manufacturer of electronic components employing approximately sixteen to seventeen thousand staff across more than seventy international sites. Its AI governance posture is enabling but in active development at the time of interview: Microsoft Copilot is positioned as the recommended sanctioned tool, AI capabilities are being embedded into multiple internal products, and a dedicated cross-functional AI team is developing organization-wide policy alongside the existing security and compliance functions. Technical visibility infrastructure (firewall logging, endpoint management, web filtering, DLP, browser extension monitoring) is mature, but AI-specific policy artefacts are not yet uniformly published or accessible to non-IT employees, and the sanctioned tool catalogue is not centrally maintained in a form that employees outside IT can consult. The five C1 participants span supply-side governance roles (security operations, security analysis, security engineering, digital innovation management) and a demand-side user role (sales support engineering), providing both perspectives within a single organizational context.

4.1.1.2. C2 - Mid-sized European IT service provider

C2 is represented by R03, an information security engineer. The organization has approximately 450 employees with offices in several European locations. Its AI posture is strongly enabling AI use, which is described as encouraged, widely adopted, and in some cases expected as part of daily work, with training, an approved tools catalog, and accessible internal guidance referenced. The participant nonetheless identified a governance tension between the existence of approved tools and continuing uncertainty about the formal review process and monitoring mechanisms. Client-data leakage through unapproved tools and excessive employee reliance on AI outputs were identified as the most prominent residual risks.

4.1.1.3. C3 - Lithuanian logistics operator

C3 is represented by R06, a transport and logistics manager with approximately twenty years of sector experience. The organisation has around eighty-five administrative employees, with the driver function delivered through subcontracting arrangements. Digitalization is described by the participant as mid-level: standard fleet management, accounting and transport-exchange systems are in place, with active preparation for NIS2 compliance through an EU-funded cybersecurity project. AI usage is currently limited to general-purpose tools (ChatGPT-class assistants used at management level for communication, information search, and legal-question scaffolding), following a pricing-automation pilot that was being wound down at the time of interview. Governance is informal: no AI-specific policy, no approved-tools catalogue, and no formal AI risk-assessment process exists. The participant frames the highest risks not as data leakage but as investment-failure risk on poorly chosen AI tools and operational-continuity risk if AI-dependent processes were to fail.

4.1.1.4. *C4 — Small Lithuanian creative-and-services business*

C4 is represented by R07, an owner-manager. The organisation operates two business lines - creative communications services and foodservice - with approximately fifteen employees in total and around eleven years of operating history. AI usage is concentrated at management and ownership level (ChatGPT, Gemini, image-generation and voice-generation tools), funded through company subscriptions. The participant explicitly characterized the governance posture as "wild" - without formal policies, controls, or risk-assessment procedures. A distinctive theme from this case is the value tension between AI-enabled efficiency and human creative authenticity, particularly in communications work, which the participant identified as a deliberately preserved boundary. Future expectations are that AI use will expand into recurring processes, scheduling and accounting, contingent on a clearer organisational understanding of tool capabilities. C4 is the smallest organisation in the sample and exemplifies the small-organisation pattern in which formal governance components are largely absent and replaced by personal trust and ad-hoc decision-making.

4.1.1.5. C5 - Lithuanian arm of an international IT outsourcing group

C5 is represented by R08, a developers team-lead and senior data engineer. The organisation is part of an international group of approximately 1,500 employees, with around 150 in the Lithuanian entity. The AI governance posture is enabling and pragmatic, particularly within software development and client-facing delivery. A distinctive feature of this organisational context is the split between internal and client-side AI governance: when employees work inside client environments, the client typically provides funds and governs AI tooling, with the outsourcing provider integrating into the client's

existing controls. Internal governance is partially informal - training, recommendations, centralized licensing and tool-usage dashboards exist, but no organization-wide AI policy document is in place. This case also produced the most concrete agentic-AI incident in the dataset: an AI development agent updated a dependency to a compromised version of a software package, resulting in a repository compromise.

4.1.1.6. C6 - European security-and-defence sector group

C6 is represented by R09, a Baltics-region executive. The organisation comprises a small local Baltics entity (fewer than ten employees) operating under a European group exceeding one thousand employees. The governance posture treats AI as "just another tool" within existing information-security, data-handling and IT approval processes rather than as a separate governance category, with regulatory and policy decisions centralized at European headquarters. Enterprise licensing of approved AI tools is emphasized as the primary control mechanism, although the participant acknowledged that not all employees currently hold enterprise licenses - a residual risk that the participant identified as the central concern. The sensitive nature of the sector drives a more conservative posture in practice than the headline framing suggests: the participant explicitly rejected giving AI decision rights over system interconnections or making AI a core organizational dependency. C6 is also the only case in the sample where regulatory engagement (with the EU AI Act in particular) is described as being handled centrally outside the local entity, rather than being a matter for local operational governance.

4.1.2. Coding distribution and analytical weight

Fig 1 shows the initial coding results of each respondent's interview coding according to the chosen deductive code system. In total, 1963 codes were identified. It is worth noting that, due to the chosen coding system, multiple codes may be applied to the same segments, greatly expanding the total count of coded segments.

Code System	R10 - ...	R09 - ...	R08 - ...	R07 - ...	R06 - ...	R05 - ...	R04 - ...	R03 - ...	R02 - ...	R01 - ...	SUM
SAIGF											0
L1_Foundation		7	11	4	13	22	3	9	5	9	83
L2_Mechanisms	21	29	23	6	20	48	25	21	41	27	261
L3_Enablement	17	16	22	22	21	44	16	19	20	19	216
MODE	44	49	41	32	50	84	37	40	47	40	464
GROUNDING	46	49	41	32	53	85	40	46	56	44	492
VALENCE	27	26	17	18	29	57	19	21	32	17	263
CONTEXT	19	16	15	12	17	35	16	16	14	9	169
IND			3			3	2	1	6		15
SUM	174	192	173	126	203	378	158	173	221	165	1963

Fig 1 Code distribution per Respondent.

Immediately, one respondent stands out - R05. This is not surprising given their role: as the organization's Digital Innovation Manager and de facto AI lead within C1, R05 held a unique dual position in the sample - simultaneously a governance designer working on policy development and a practitioner who had personally tried to get AI tools approved and found the process frustrating. This combination meant they had meaningful insights across all three SAIGF layers, providing the broadest thematic coverage of any single participant.

Beyond R05, the distribution does not reveal any particularly broader patterns. Managerial respondents (R01, R05, R06, R07, R08, R09) generally produced denser transcripts across Layer 1 and Layer 3 components, where governance design perspective and sanctioned tool decisions gave them more to articulate. Non-managerial respondents (R02, R03, R04, R10) produced comparatively leaner overall counts but contributed disproportionately to Layer 2 components-particularly risk assessment and policy - where personal experience of governance friction, tool inadequacy, and workaround decisions generated concrete, grounded content that managerial accounts often could not replicate.

4.2. Findings by SAIGF layer

The following subsections present participant evidence organized by SAIGF layer. For each layer, a table with representative quotes is provided, drawing directly from the coded corpus, followed by analytical commentary on what the evidence collectively reveals. All quotes are from interview transcripts; Lithuanian-language quotes are included with English translations.

4.2.1. Layer 1 - Governance Foundation findings

Table 11 shows the selected interview quotes, that provides deeper overview of Layer 1 analysis for SAIGF.

Table 11 Selected interview quotes for Layer 1: Governance foundation of SAIGF

Component	Participant Quotes
Governance roles and accountability	*"The compliance member from the security team and the AI lead are jointly drafting policy, but I'm not aware of how procurement-stage decisions are actually made."* (R01)
	"The conscious decisions come when executive leadership wants something - then things move very fast. Otherwise it is relatively slow. Push from the bottom up is very little, unless it comes from IT or R&D." (R05)
	"When sanctioned tools fail to gain adoption, IT left it off — because the underlying problem is an HR thing, a management and leadership thing, not necessarily IT." (R05)
Regulatory and compliance alignment	*"We reviewed the Act with our legal department and concluded that as deployers of AI tools the regulatory impact is very, very minimal."* (R05)
	"All the regulations are coming centralized from our headquarters in Europe - it is not something we handle at the local level." (R09)
Risk appetite definition	*"Data leaks and stuff honestly don't concern me that much. Sure, it's a risk. But for me personally it sits a little bit lower on the risk registry. For me, the biggest risk first of all is spend."* (R05)
	"Don't make AI the core of the organization. It is very dangerous when you stop using it as a tool and start giving it decision rights over how to make interconnections." (R09)
	"The highest risk for us is investing in a tool that does not meet expectations and burning internal resources on an unsuccessful pilot." (R06)

The findings from the foundation layer reveal a consistent pattern across the sample: governance ownership exists only in name and lacks clear boundaries in practice. Responsibility for AI governance is usually shared between a security or compliance function and a newly created AI lead

role. However, the handoff between technical and behavioral ownership, such as who is responsible when employees do not adopt sanctioned tools or when AI is used informally outside any review process remains undefined in most cases. R01's account of joint policy drafting without awareness of procurement decisions and R05's observation that sanctioned tool adoption failure is primarily an HR issue rather than an IT one both highlight the same structural gap from different perspectives. Governance has been designed to address the tool selection problem but not the organizational behavior issue that determines whether those selections are followed.

The most noticeable finding in this layer is the regulatory alignment. Only three participants reported any active engagement with the EU AI Act. In each case, the engagement was mediated - R05 analyzed it through a legal department assessment, concluding that exposure as a deployer was "very, very minimal," while R09 positioned regulatory responsibility as entirely centralized at European headquarters rather than a local operational concern. This indicates that even when organizations have established a regulatory translation mechanism at the managerial level, that translation is not reaching the practitioners who make daily decisions about AI tool usage, data handling, and approval. Regulatory alignment, as the framework requires, presumes not only that leadership understands the obligations but that this understanding is put into practice via governance artifacts - policies, assessment criteria, approval conditions - that are visible and accessible to those who need to act on them. The data suggests this second step is mostly absent.

The findings on risk appetite revealed the most unexpected analytical result within this layer. The academic literature reviewed in Chapter 2 considers data leakage to be the one of the primary concern driving governance of strategic artificial intelligence. However, the evidence from participants reverses this hierarchy. R05 explicitly ranked spending risk above data leakage, noting that it "sits a little bit lower on the risk registry." R06 identified the risk of investment failure the danger of committing organizational resources to an AI tool that does not deliver as the main concern. R09 introduced a third perspective not present in the literature: dependency risk. This refers to the organizational danger that arises when AI becomes so integral to decision-making and system interconnection that its unavailability or failure results in an operational crisis. Together, these three perspectives—cost governance, investment failure, and operational dependency expand the concept of risk appetite far beyond its original focus on cybersecurity. They also have direct implications for how this component should be specified in any practical implementation of the framework.

4.2.2. Layer 2 - Core Governance Mechanisms findings

Table 12 presents representative quotes from the coded corpus for each Layer 2 component, selected to illustrate the primary patterns identified across the ten interviews. The commentary that follows draws directly on these additional coded segments to develop the analytical argument for each component.

Table 12 Selected interview quotes for Layer 2: Core governance mechanisms of SAIGF

Component	Participant Quotes
Visibility and discovery	*"It's hard to control. There is no official data. You never know - because everybody also has a personal phone, and you do not know what data they use or how."* (R09)
	"There was a significant initiative to increase monitoring for various browser extensions and to increase visibility into what people tend to use or misuse - but that project is still ongoing, because a lot of other things took priority. It is definitely one of the blind spots." (R02)

Component	Participant Quotes
	"Of approximately one thousand expected users of the new sanctioned agent, only around one hundred people actually opened it - and IT had no visibility into why." (R05)
Policy and authorization framework	*"There is some guidance — you should not input anything sensitive into public tools. But it's very soft. It's not enforced."* (R05)
	"The guidance is not easy to find. Not easy. Employees need to ask." (R05)
	"There is a list of approved tools. I don't think it is available to all employees." (R05)
	"If everybody uses ChatGPT and loads data from it into Copilot just to get results faster and formally appear compliant - then you have a problem. That means organizational culture has already moved past your recommendations." (R09)
Risk assessment and management	*"Since agentic AIs are being rolled out everywhere, they really need to be checked and reviewed so that malware doesn't get launched by a rogue AI on the endpoints."* (R04)
	"An AI agent did not know that with a certain package version a code injection was possible. It updated to that version and essentially the GitHub repository was compromised." (R08)
	"You can go through tokens very, very quickly. If there are no limits applied, AI spend can go sky high." (R05)
Technical controls	*"Our organization chose - I'm not sure if consciously or unconsciously - not to block these tools."* (R05)
	"We have not signed onto any agent gateway that would ensure how the agent responds and what it should not respond to. We use completely what AI tools themselves provide." (R08)
	"I'm still uncertain whether our tools would prevent it or not. I have no idea. Maybe yes, maybe no - and that is also a concern." (R02)

The core mechanisms layer produced the highest absolute volume of coded segments across the entire dataset and the most negatively valenced findings within any single layer. Taken together, the four components reveal that operational governance exists in outline in larger organisations but consistently fails at the boundary between governance design and the employee experience of that governance.

Visibility findings divide sharply along organisation size. Larger organisations described some combination of firewall logging, endpoint management, web filtering, DLP, and browser extension monitoring — but participants were consistent that this infrastructure was not designed for AI-specific signals and leaves meaningful gaps. R02's account of a browser extension monitoring initiative still ongoing due to competing priorities illustrates a pattern that recurred across C1: the gap is acknowledged internally, remediation is planned, but it is consistently deprioritised against other security demands. The most structurally intractable blind spot, identified independently by R09, is the personal device: organisational monitoring perimeters end at the corporate device boundary, while AI usage does not. A further visibility problem, articulated by R05, operates on the supply side rather than the demand side — the organisation knew what it had deployed, but had no visibility into whether employees were actually using it, or why they were not. These are two distinct governance failures: one about detecting unsanctioned behaviour, the other about understanding sanctioned behaviour. Both require attention.

Policy findings were the most heavily coded component in Layer 2 and the most negatively valenced in absolute terms. Three specific failure modes emerge clearly from the evidence. The first is that policies exist but carry no enforcement weight — R05's characterisation of guidance as "very soft, not enforced" describes a governance artefact that is present in form but absent in practice. The second is that policies are inaccessible to the employees they are meant to govern — employees cannot find them without asking IT directly, which effectively makes the compliance pathway dependent on

knowing to ask rather than on the policy being discoverable. The third and most consequential is that the approved tool list, where it exists, is not published in a channel that non-IT employees consult. R09's formulation distils the practical consequence of all three failures together: when employees find a workaround that produces better results and carries no visible consequence, organisational culture migrates to the workaround and the formal policy becomes decorative. A policy that exists but cannot be found, enforced, or practically acted upon is functionally equivalent to no policy at all.

Risk assessment findings produced the most concrete evidence of materialised risk in the dataset. R08's account of an AI development agent updating a software dependency to a compromised version — resulting in a GitHub repository compromise — is the clearest documented case of agentic permissions risk in the corpus and should be read as a sentinel event rather than an isolated incident. It demonstrates that the risk is not theoretical: an AI agent operating with broad write permissions, without a human-in-the-loop checkpoint, can introduce a vulnerability that a human developer would have been expected to catch. R04's concern about agentic AI being "rolled out everywhere" without corresponding review processes reflects the same governance gap from the anticipatory rather than the retrospective direction. The cost dimension of risk assessment, raised by R05, introduces a concern absent from the academic literature reviewed in Chapter 2: token consumption without configured limits creates an uncontrolled financial exposure that sits alongside data leakage as a material organisational risk, yet receives no treatment in current AI governance standards.

Technical controls followed the visibility pattern: present in larger organisations but rarely AI-specific in design or intent. The dominant posture was permissive by default — R05's observation that the organisation chose "not to block these tools," whether consciously or otherwise, describes a posture that is widespread in the sample. Where controls exist, they are inherited from general IT security infrastructure rather than designed for AI. R08's disclosure that no agent gateway is in place, leaving the organisation reliant entirely on vendor-provided guardrails, represents the technical controls gap in its sharpest form: the organisation has ceded the control plane for agentic AI behaviour to the AI vendor, retaining no independent enforcement capability. R02's reflexive uncertainty about whether existing tools would even detect AI-specific attacks - "I have no idea, maybe yes, maybe no" - is coded with reflexive grounding precisely because that uncertainty is itself analytically meaningful: a cybersecurity analyst who cannot assess their own organisation's AI-specific defensive coverage is experiencing a governance visibility gap, not a personal knowledge gap.

4.2.3. Layer 3 - Innovation Enablement and Sustainability findings

Table 13 presents representative quotes from the coded corpus for each Layer 3 component. The commentary that follows draws directly on these additional coded segments to develop the analytical argument for each component.

Table 13 Selected interview quotes for Layer 3: Innovation Enablement and Sustainability of SAIGF

Component	Participant Quotes
Sanctioned AI provision	*"I'm not aware of a list of which tools are approved or not. I have information that our organisation uses a specific one and should be using a specific one - I would call it a recommendation rather than a list."* (R01)
	"Most people think Copilot is not very good. You cannot forbid AI tools because employees will simply use them on personal devices." (R01)

Component	Participant Quotes
	"One strong driver for not going through the approval process is the output quality of the approved model." (R04)
	"If everybody uses ChatGPT and loads data from it into Copilot just to get results faster and formally appear compliant - then you have a problem. That means organisational culture has already moved past your recommendations. You need to align policies and take those tools into your tool library." (R09)
Fast-track and tiered approval	*"Fast track? No. The approval process is general for everything."* (R04)
	"If your organisation introduced a fast-track process yielding a decision within two or three working days, I think a lot more tools would be tried out - in a potentially safe, non-data-leaking way." (R04)
	"If it comes from management, yes, it gets done on priority. I don't think we have a formal priority-review process for other requests - but I would need to check." (R01)
	"If I wanted to onboard a new tool, I wouldn't know exactly how to start. I would ask my team who to contact - the applications team, the security team? I'm not exactly sure." (R02)
Culture, awareness, and continuous improvement	*"Global communications are read by around five percent of people. So there might be a gap."* (R01)
	"There are yearly trainings and AI is mentioned - but I'm not aware that we have a specifically AI-focused training. I think it will be implemented in the near future." (R01)
	"I don't think so" — when asked whether any formal training or guidance on AI data security had been received. (R04)
	"Colleagues are trusting that the LLM knows the answers and stop thinking themselves." (R02)
	"People are not taught, educated, or encouraged to use AI intelligently - it is often mystified." (R06)
	"You need to educate employees - raise general common sense and understanding of the tools. Not make people afraid of AI, and not make them too enthusiastic either. Both extremes are critical, just in different directions." (R09)
	"How are they using the output? That is the biggest risk - not that they used AI, but what they did with what it produced." (R05)

The enablement layer findings complicate the framework's central assumption. The framework as theorized in Chapter 2 presupposes that adequate sanctioned provision reduces shadow adoption by removing the employee incentive to seek alternatives. The empirical data confirms the logic but challenges the premise: in most participating organizations, the primary sanctioned tool - most commonly Microsoft Copilot - was rated mixed to inadequate by both managerial and non-managerial participants alike.

The findings related to sanctioned provisions are the most critically analyzed in this layer. Inadequacy is described in practical terms rather than abstract ones. R01 clearly states that because Copilot is seen as not very effective, employees bypass the policy by using AI on personal devices - turning the sanctioned tool into an ineffective governance mechanism that actually fosters the very shadow behavior it aims to prevent. R04 further emphasizes that output quality is the main reason employees skip the approval process entirely - they do not choose the shadow option because of governance friction, but because the sanctioned option fails to meet their needs. R09 captures the systemic impact: when employees route data through an unsanctioned tool to get better results, then transfer outputs into the sanctioned tool to appear compliant, the organization experiences the worst of both worlds - shadow use risks without any oversight. Therefore, the sanctioned provision component needs to be

reframed, from a list of acceptable tools to a measure of competitiveness. A list of approved tools that go unused is not a governance control; it's merely a governance artifact.

Fast-track approval findings were fewer in number than the sanctioned provision component but offered the most operationally precise evidence in the layer. No participating organization had a tiered or AI-specific expedited approval process at the time of the interview. When it existed, the process applied uniform procedures regardless of risk level or urgency. The practical impact, identified by R01, is a structural imbalance: management-supported initiatives get accelerated treatment because executive backing creates an informal priority channel, while bottom-up requests from employees with genuine productivity needs lack an equivalent route. This imbalance matters because it means the path to compliance is accessible to those with authority and blocked for those most affected by the governance gap, which often drives shadow behavior. R02's account of not knowing whether to contact the applications team or the security team to begin a tool onboarding - and ultimately not trying - is the experiential sign of this failure: the compliance path exists in theory but is not practical, making shadow use the easiest option by default, not by design.

The culture, awareness, and continuous improvement findings showed the strongest and most consistent signals across the entire dataset. Out of 116 coded segments with a strongly negative valence balance, L3_CULTURE was the most heavily evidenced component in the corpus - concentrated in two distinct clusters. The first is the awareness deficit: governance artifacts exist in most larger organizations but do not reach the employees they are meant to serve. R01's statistic that global communications are read by about five percent of recipients is not a criticism of employees but an empirical observation of how organizational communication functions at scale, with a direct implication: policies distributed through global communication channels are effectively invisible to ninety-five percent of the workforce. R04's straightforward statement that no AI-specific training or guidance had been received, despite working in an information security role at a large organization, highlights the same gap at the individual level.

The second cluster pertains to decision quality -a governance issue that was entirely missing from the framework in Chapter 2 and emerged solely from empirical data. R02's observation that colleagues trust LLM outputs and stop thinking independently, R06's concern that AI is misunderstood rather than understood, and R05's identification of how employees rely on AI outputs as the main risk—all describe a failure mode that no technical control, policy, or approval process can address: employees acting on AI-generated content without the critical evaluation skills needed to identify inaccuracies. R03's forward-looking concern that organizations should avoid a path where everything is done by AI highlights the same issue at the governance level instead of focusing on individual behavior. The decision-quality aspect is not a secondary concern related to awareness; it is a governance focus in its own right, requiring not only training but also organizational design features such as human oversight checkpoints for AI-assisted work in high-stakes contexts.

4.2.4. Cross-cutting findings

Three patterns emerged from the analysis that do not correspond to a single SAIGF component but influence the operation of multiple components at the same time.

The first issue is the gap in perception between managers and employees, which led to the dual-group sampling design. The two groups did not give conflicting descriptions of the same governance setup,

instead, they highlighted different aspects. Managers consistently described policies, sanctioned tool lists, and approval processes as existing and fairly developed. In contrast, employees from the same or similar organizations consistently described these artifacts as hidden, unavailable, or operationally unclear. A clear example is the difference between R05's managerial statement - "there is a list, I don't think it is available for all employees" - and R10's non-managerial perspective from the same organization, which expressed the need for a centralized, accessible place to check approved tools and noted that they are unaware of any such resource (R10, translated from Lithuanian). The same artifact, but two entirely different accounts of its practical availability. This reveals a governance failure mode that is not the absence of artifacts but their accessibility: organizations create governance artifacts that are visible to managers but not accessible to employees. This finding also has a methodological implication: studies that involve only managerial respondents will likely overestimate the operational maturity of governance because managers have better visibility into artifacts their employees cannot find. R04 also confirms this within the same organization, stating that policies "are not easily accessible, especially for non-IT employees" - a judgment based on direct experience rather than uncertainty.

The second cross-cutting pattern is the organizational size effect, which is qualitative rather than purely quantitative. In large organizations - primarily represented by C1 across five participants - all SAIGF components were at least partially in place, but each faced the same type of failure: communication and accessibility issues that disconnected governance design from the employee experience. The infrastructure, policies, tools, and roles existed; what was missing was the channel that made them understandable and accessible to employees outside of governance. In medium organizations - C2, with its strong enabling stance but unclear formal review processes; C5, with its client-mediated governance split; and C6, with European headquarters-centralized regulatory and policy decisions - governance was somewhat formalized but minimal at the operational local level, relying more on external parties than on internal design. In small organizations - C3, where no AI-specific policy, approved-tool catalog, or formal risk assessment process existed at the time of interview; and C4, where R07 explicitly described governance as "laukinis" (wild, ungoverned) without contradiction - formal governance components were mostly absent, replaced by owner-level trust and ad-hoc decision-making. This last setup isn’t a failure in the same way as the large-organization accessibility gap; rather, it reflects a sensible governance approach for organizations of that size, where the cost of formal governance outweighs its practical benefit. The practical implication is that the SAIGF, as designed, is most applicable to medium-to-large organizations. Proportional or tiered implementation guidance would be necessary to make it usable for small organizations with informal governance models.

The third cross-cutting pattern involves the promotion of two inductive themes as components within a formal framework (this is indicated within Fig 2). Agentic permissions - labeled as IND_AGENTIC_PERMISSIONS across ten segments in four participant transcripts, with each segment showing negative sentiment - highlight the concern that AI agents are granted broad system permissions by default. This creates governance risks because traditional access control and software approval processes are not designed to manage such permissions. The theme appeared independently in R02, R04, R05, and R08. R08 provided the only concrete documented incident in the dataset: an AI development agent upgraded software dependency to a compromised version without human review, leading to a GitHub repository breach. R04 described the concern normatively - with the idea that agentic AI being "rolled out everywhere" requires oversight to prevent "malware from being

launched by rogue AI on the endpoints." R02 approached it from a cybersecurity analyst perspective, highlighting permissions granted to agentic tools and automated email actions as primary issues. R05 added the conceptual distinction that underpins the code, differentiating between embedded Copilot and standalone Copilot Studio, which have different permission scopes within what appears to be the same authorized tool family.

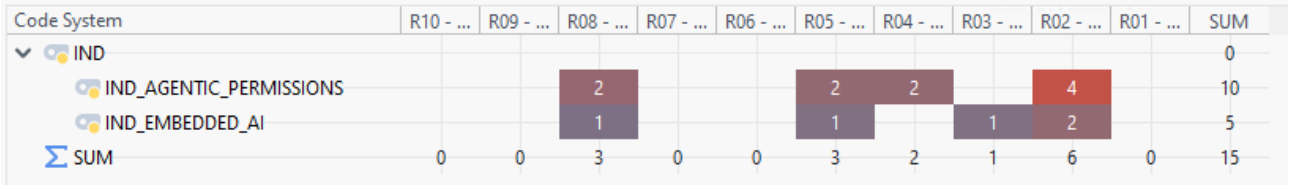

Fig 2 Inductive codes from interview analysis

Embedded AI - coded as IND_EMBEDDED_AI across five segments in three transcripts - describes AI features arriving inside already-approved tools as routine product updates, inheriting sanctioned status without triggering any re-assessment. The code was first observed in R05's interview, where the participant explicitly distinguished embedded Copilot from Copilot Studio as carrying different governance implications despite both operating under a Microsoft licence. R02 provided the most operationally concrete case: an AI agent embedded in a previously trusted security product that had degraded rather than improved investigative utility while retaining full approved status, with no re-assessment event triggered by the feature introduction. R03 confirmed the phenomenon existed in their environment as well, describing it as "a mix of both" sanctioned and shadow in practice. Both inductive themes identify governance objects - agentic capability scope and AI-feature-bearing approved tools - that the original framework, substantially derived from Shadow IT antecedents, did not address.

4.3. Updated Shadow AI governance framework

The empirical findings presented in Section 4.2 are translated here into a set of targeted revisions to the theoretical SAIGF. The update follows a conservative principle: no component is removed, no layer boundary is redrawn, and no component is relocated. Refinements are evolutionary - deepening or extending existing components rather than replacing them. Two new sub-components are added through promotion of inductive codes that met the three-case recurrence threshold, and one cross-cutting thread is introduced that operates horizontally across all three layers. **Table 14** Updated SAIGF provides a consolidated overview of all changes, refinements, and findings of empirical review.

Table 14 Updated SAIGF

Layer	Component	Status	Empirical refinement
L1 Foundation	Governance roles and accountability	Validated	Chapter 2 involves defining who approves tools, owns risk, monitors compliance, and leads incident response - drawing on ISO/IEC 38500 Responsibility principle to prevent the situation where "everyone uses it, no one owns it." Empirically, this is necessary but not enough: while technical ownership exists in most organizations, the boundary between IT, HR, and business

Layer	Component	Status	Empirical refinement
			leadership ownership of behavioral governance - particularly regarding what happens when employees do not adopt sanctioned tools - remains unclear; the AI lead role may exist informally but lacks formal decision-making authority.
	Regulatory and compliance alignment	Validated	The initial SAIGF needed to map the EU AI Act, GDPR, and sector obligations into specific governance design constraints. It was empirically confirmed that this mapping happens - but only at the managerial level through legal departments or headquarters; it does not reach operational decision-makers. The framework must explicitly address this institutional translation gap and not assume it happens automatically.
	Risk appetite definition	Refined	Initial SAIGF framed this as the calibration mechanism for Layer 2, separating low-risk from high-risk AI usage. The original framing assumed data leakage and security risk as the primary factors. Empirically, practitioners prioritized AI cost and token spend, investment-failure risk on unsuccessful pilots, and operational dependency risk above data leakage in their effective risk hierarchies - factors missing from the original component and all reviewed standards; they must be added explicitly.
L2 Mechanisms	Visibility and discovery	Refined	Initial SAIGF specified technical mechanisms (CASB, DLP, network traffic analysis) and procedural ones (self-declaration, team inventories). Empirically confirmed as partially implemented in larger organizations but with a structural blind spot not addressed initially: personal devices and non-enrolled endpoints fall entirely outside the technical monitoring perimeter. This gap cannot be closed by technical controls and must be addressed through policy scope and behavioral reinforcement.
	Policy and authorization framework	Refined	Initial SAIGF required policies to be written in clear language, communicated through multiple channels, and easily found by employees outside IT functions. Empirically, this requirement is stated but consistently unmet: policies exist, but employees must ask IT to find them, and approved tool lists are not published in channels employees consult. The difference between policy existence and policy accessibility must be made an explicit, measurable design requirement - not just an aspiration.
	Risk assessment and management	Refined	The initial SAIGF assessment scope included data sensitivity, security posture, compliance, operational dependency, and output reliability - applied during tool procurement. However, this scope must be expanded: AI features added to previously approved tools do not trigger re-assessment and automatically inherit the approved status; agentic permission scope is not part of any defined

Layer	Component	Status	Empirical refinement
			assessment process; and AI cost exposure is a key concern for practitioners that was missing from the original component.
	Embedded AI re-assessment	New	Not addressed in the initial SAIGF or any reviewed standard. AI features arriving in approved tools via product updates create a governance gap, allowing new data flows and permission scopes to enter the organization without review. Promoted from IND_EMBEDDED_AI code from three participants' interviews, in five segments. One concrete case of degraded utility with inherited approved status
	Technical controls	Refined	Initial SAIGF covered DLP, CASB, access control, monitoring and logging, prompt filtering, output validation, and sandboxing - confirmed as the correct approach based on empirical evidence; an agentic permissions component is added below as the main governance gap not addressed by the original list.
	Agentic permissions and tool-use governance	New	This was not addressed in the initial SAIGF or any reviewed standard. AI agents are granted broad write permissions by default; no human-in-the-loop checkpoint is required in any participating organization; and one incident of repository compromise has been documented within conducted intervies.
L3 Enablement	Sanctioned AI provision and adequacy	Refined	Initial SAIGF explicitly requires that sanctioned tools must be genuinely competitive with unsanctioned alternatives - if they are markedly inferior, employees will seek shadow alternatives regardless of policy. Empirically, this requirement is universally unmet: the primary sanctioned tool is rated inadequate by both managers and employees across multiple organizations; catalog existence without adequacy assessment does not reduce shadow adoption; adequacy and competitiveness assessment must become active, periodic governance activities.
	Fast-track and tiered approval	Validated	The initial SAIGF required low-risk tools to be approved within days and high-risk tools within specified timelines, with transparency directed at end users. Empirically confirmed as completely absent: no participating organization has a tiered or AI-specific pathway; expedited review exists only for management-sponsored initiatives, creating a structural access inequality; a time-bound trial pathway must be formalized as a standard requirement.
	Culture, awareness and continuous improvement	Refined	Initially SAIGF addressed awareness, training, transparency of governance rationale, and continuous improvement. Empirically, training covers safe AI use - what data to input - but not critical evaluation of AI outputs. Decision-quality risk - employees acting on incorrect AI-generated outputs without validation - was identified as the primary unmitigated risk by practitioners across

Layer	Component	Status	Empirical refinement
			multiple organizations and is absent from all reviewed governance standards; must be added as an explicit sub-component
Cross-cutting	Human oversight thread	New	Not included in the initial SAIGF as a specific requirement, but implicitly present across components. Empirical evidence shows decision-quality and human-in-the-loop concerns arise across all three layers. Organizations need to define which decision categories require human approval of AI outputs; this must be specified at Layer 1, enforced at Layer 2, and trained at Layer 3. Currently, this is missing from all participating organizations and reviewed standards.

The framework maintains its three-layer architecture and order-of-precedence logic. Of the original ten components, seven are validated or refined, while three additions - including two new sub-components and the cross-cutting thread - address governance objects that Shadow IT-derived frameworks do not capture: agentic capability scope, AI features embedded inside already-approved tools, and the decision-quality risk that arises when employees act on AI outputs without sufficient critical oversight. These three additions collectively represent the most significant theoretical contribution of the empirical phase.

4.4. Discussion and Recommendations

This section analyzes the empirical findings within a broader theoretical and practical framework, exploring implications for the academic understanding of Shadow AI governance, offering actionable recommendations for organizational leadership and technical teams, and highlighting the study's limitations along with suggestions for future research.

4.4.1. Theoretical Implications

The empirical findings have four implications for the academic understanding of Shadow AI governance.

First, the data support the theoretical position from Chapter 2 that SAI extends rather than replaces Shadow IT theory. The drivers of unsanctioned AI adoption identified by participants – inadequate sanctioned tools, approval process friction, productivity pressure, and peripheral organizational visibility - closely match the Shadow IT drivers described in the literature reviewed in Chapter 2. However, the data also supports the idea that SAI introduces three governance objects not addressed by the Shadow IT literature: agentic permissions, embedded AI within already-sanctioned tools, and decision-making risks from relying on AI outputs. This clarifies the theory: SAI extends Shadow IT in its dynamics but diverges in its governance objects. Frameworks based solely on Shadow IT, without these three additions, will systematically underestimate the SAI risk landscape.

Second, and arguably the most important single theoretical contribution from the empirical phase, the data reveal the visibility and awareness gap - not the technical control gap - as the main failure mode at the operational level. The most heavily coded SAIGF component (L3_CULTURE, 116 segments)

and the most negatively valenced (45 negative segments) were behavioral and informational, not technical. Along with the manager employee accessibility gap. This suggests that the focus of SAI governance research should move away from the technical control issues that have mostly dominated early literature toward the communication, accessibility, and decision-making quality issues that the field has largely overlooked. This aligns with the Chin et al. (2025) finding that prohibitive technical approaches push AI use underground but goes further: even when prohibition isn't the main approach, communication failures alone can cause shadow adoption.

Third, the asymmetry in perceptions between managers and employees was empirically confirmed as an important research focus. The two groups did not provide conflicting accounts but instead highlighted different aspects of the same governance setup: managers are aware that policies, lists, and approval pathways exist, while employees view them as invisible or inaccessible. This finding both methodologically supports the dual-group sampling approach and conceptually indicates that studies with only one group - especially those that only include managerial respondents, which is common in the SAI literature - tend to overestimate the maturity of governance practices. Future research on SAI governance should consider dual-group sampling as a standard methodological practice rather than an optional extension.

Fourth, the theoretical literature has often treated the sanctioned/shadow distinction as binary: a tool is either approved or it is not. Data show two phenomena that challenge this binary. Embedded AI exists within already-approved tools and inherits their approved status by default, even if the AI features themselves have not been assessed. Agentic AI is embedded within approved tools or extends them with permission scopes that surpass any review the original tool received. Both phenomena suggest that future theoretical work should reconceptualize the sanctioned-versus-shadow distinction not as a property of tools but as a property of capabilities within tools - a more detailed unit of analysis that reflects the reality that a single tool can host both approved and unapproved capabilities at the same time. This is a significant theoretical shift; it indicates that approval during procurement, the main control point in current governance practice, is fundamentally inadequate for governing AI.

4.4.2. Management Implications

Six recommendations for organizational leadership arise from the empirical findings.

First, prioritize policy accessibility as a key goal in policy design. Collected data shows that the biggest cause of perceived governance failure is the gap between having a policy and making it accessible. A policy that exists but cannot be found is essentially the same as having no policy at all. Organizations should publish AI policies through the same channels employees use for tool selection, such as the IT service catalog or self-service portal. They should also include a simple, non-technical summary alongside any technical version, and maintain a single, clear answer to the question "which AI tools may I use?" that can be accessed within two clicks from the employee intranet. R10's response - "I would have to ask IT security; where on the internet I could check, I cannot say right now" - exemplifies this failure mode and should serve as the standard for evaluating the recommendation.

Second, clearly define risk appetite, including the cost aspect. Most participants in the dataset described their organizations' risk stance as implied or implicit; it only became evident when specific cases required a decision. Leadership should make an explicit statement that outlines

- the types of AI use the organization will never tolerate, such as confidential customer data in unauthorized tools, automated actions on production systems without human approval, or AI-driven decisions in regulated fields without proper audit trails
- the categories that need formal review
- the uses allowed under ongoing approval
- the budget and risk tolerance for AI pilots. The data showed that cost appetite is not just a back-office matter: R05 ranked it above data leakage in the organization's overall risk hierarchy, and R06 prioritized investment-failure risk over all other risks.

Third, view the provision of sanctioned tools as a competitiveness issue rather than a catalog problem. The most consistent normative theme across the dataset was that adequate sanctioned alternatives decrease shadow adoption, while inadequate alternatives encourage it. Leadership should conduct regular reviews of the adequacy of sanctioned AI tools compared to the alternatives employees would otherwise pursue, treat user-reported inadequacy as a signal for action (such as replacement, expansion, or documented acceptance of the gap), and invest in enterprise licensing for the tools that pass the adequacy review. R09's identification of insufficient enterprise licensing as a key risk and R01's remark that "most of the people thinks that Copilot is not very good" are both signs that a catalog without adequate alternatives is empirically unstable.

Fourth, establish a time-limited fast-track approval process. R04's specific proposal - "some sort of limited-time fast-track process just so they can get approval to check out the tool and if they don't like it, not proceed with the full actual review" - is endorsed. The process should deliver a simple yes-or-no preliminary decision (approved-for-trial / not-approved-for-trial) within a set service-level window (data indicate two to three business days is acceptable to users), any employee should be able to trigger it, and it should clearly define the trial as time-limited and data-restricted (no confidential data, automatic expiry without moving to full approval). This addresses both the speed concern and the imbalance where only management-priority initiatives currently receive expedited processing.

Fifth, avoid placing AI at the center of decision-making. R09 expressed this clearly: AI should be seen as "another tool in the organizational tool library" rather than becoming "the core of the organization's decision-making or system interconnection logic." This recommendation emphasizes the human oversight aspect of the updated framework: leadership should explicitly identify decision categories where AI outputs are only advisory and require human approval, should oppose organizational structures where AI is at the heart of decision processes rather than at the inputs, and should establish clear fallback procedures if AI services become unavailable. R06 illustrated this concretely as a business continuity issue: an organization that depends operationally on AI is vulnerable to its outage.

Sixth, address discovered shadow AI politically rather than punitively. R04 explicitly noted that the unresolved question in their organization is "how to bring shadow AI usage to the light, politically." The empirical evidence shows that shadow AI use is rarely malicious; it is mostly driven by productivity needs and often lacks transparency. A punitive response tends to push the behavior further into the shadows rather than stopping it. Leadership should create an amnesty pathway allowing employees to disclose existing shadow AI use without facing sanctions, in exchange for the organization either officially adopting the tool (if it passes review) or documenting the gap and offering a sanctioned alternative.

4.4.3. Technical implications

Six technical recommendations stem from the empirical findings, mainly aimed at IT, information security, and AI engineering functions.

First, treat DLP and CASB as necessary but not enough. The collected data shows that the current security tools - firewall logs, web filtering, DLP, CASB, SIEM signals - offer partial visibility into AI use at the network and endpoint levels but are not meant to detect AI-specific signals like prompt content, agent activity, or embedded AI features. Organizations should add AI-specific discovery methods: AI-aware DLP that can recognize prompt content as a separate category, browser extension monitoring with an AI-tool taxonomy, identity and licensing telemetry analyzed for AI-tool usage patterns, and, where relevant, purpose-built shadow-AI discovery tools. The goal is to narrow the gap between supply-side visibility (what the organization has deployed) and demand-side visibility (what employees are using).

Second, address the personal-device blind spot through behavioral controls rather than technical ones. Data shows that controlling personal-device use of AI tools cannot rely solely on endpoint management because the device is not enrolled. The technical approach should include: explicitly prohibiting confidential-data entry into AI tools accessed from personal devices as outlined in the acceptable-use policy; implementing mobile device management policies for access to corporate data on non-enrolled devices; and, where possible, applying conditional-access controls that restrict corporate resource access based on the device-management status. Behavioral reinforcement (training, awareness, periodic attestation) should complement these measures, not replace them.

Third, design agentic-permission architectures based on a least privilege principle. The R08 incident - where an AI agent updated a dependency to a compromised version, leading to repository compromise - is the clearest documented case of agentic-permissions risk in the dataset and should be regarded as a sentinel event. Recommended architectural patterns include: read-only access by default for AI agents interacting with corporate systems; write operations that require human approval for any production environment; full agentic mode (autonomous action without per-step approval) limited to designated sandbox environments with no production data and no external network access; an agent inventory that logs every agent's connected systems, permission scope, and approval mode; and, where justified by deployment scope, an agent gateway serving as a control plane for permission enforcement and auditing.

Fourth, re-evaluate already-approved tools when AI features are added. The data shows that AI features integrated into trusted tools as updates do not currently require a procurement-stage review. The technical recommendation is to keep a tool inventory that records AI-feature launches as separate events, to apply the same data handling and permission-scope review at the time of AI-feature implementation as was applied during the original procurement, and to treat "AI feature added" as a reclassification trigger that could raise a previously low-risk tool to a higher risk level. R02's

observation that an AI agent embedded in a security product had reduced perceived usefulness while inheriting the product's approved status highlights this gap.

Fifth, instrument cost telemetry at a granularity that supports control. The data shows that AI spend is a key risk factor. The technical solution is to make AI consumption measurable: dashboards for team and tool spend; token usage limits at the application or tenant level; clear cost ownership assigned to specific teams (using team budgets rather than a central pool); and alerts for consumption anomalies. R05's observation that "if there's no limits applied, you can go rocket and the sky AI will spend on token usage" is technically feasible; the absence of limits is a configuration choice, not a technical constraint.

Sixth, implement AI-specific defensive controls where possible. Data shows that prompt-injection defenses, output filtering, AI-aware DLP, and similar AI-native controls are lacking in the participating organizations. Although the maturity of available tools varies, several features are now feasible: input filtering for known prompt-injection patterns at the AI gateway; output filtering to prevent confidential data leaks in generated content; redaction of personally identifiable information before sharing with external AI services; and anomaly detection in agent-execution patterns. These are recommended not as complete security measures but as essential controls, as without them, the AI-specific risk surface remains unaddressed by existing security tools.

4.4.4. Limitations and Future research

Four categories of limitations should be considered when interpreting the findings.

First, the sample size is small and falls below the upper limit of the planned range (12–16). It is also geographically and somewhat sectoral concentrated. All participants are based in Lithuania or work for organizations with significant Lithuanian operations; the largest sector cluster is electronics manufacturing (four of ten participants), and the security and IT functional cluster is overrepresented (six of ten participants work in security, IT, or digital innovation roles). The representativeness of these findings beyond this context cannot be assumed. The findings on regulatory engagement, especially the relatively distant operational involvement with the EU AI Act, may reflect sector and seniority effects: highly regulated sectors such as healthcare, financial services, and critical infrastructure, along with senior in-house counsel within these sectors, would likely have a different perspective. This was also somewhat hinted at by R09 in the Defense sector.

Second, the design is cross-sectional. All interviews were conducted within a three-week period in April 2026, during which several participants explicitly described their organizations' AI governance arrangements as actively developing. The findings therefore reflect a snapshot at a particularly transitional moment in the field; a longitudinal study would be needed to determine which arrangements stabilize and which are still intermediate. Several participants (R01, R05, R08) described AI governance as a work in progress at the time of the interview; the findings on policy immaturity, in particular, should be understood in this context.

Third, the data are self-reported. Participants described their organizations' governance arrangements as they understood them; the findings on policy invisibility, sanctioned-tool inadequacy, and decision-quality risk all rely on participant accounts that were not cross-checked against

organizational artifacts (policy documents, tool catalogs, audit logs). The reflexive grounding code explicitly captures the segments where participants indicated their own uncertainty about their organization's governance status, but the broader self-report limitation applies to all segments. Future research should cross-verify self-reported accounts with organizational artifacts where access allows.

Fourth, the framework refinement presented in Section 4.2 is theoretical and has not been empirically validated. The updated framework was developed by analyzing the empirical evidence against the Chapter 2 framework; it has not been tested against new empirical evidence. Validating the updated framework - whether through case-study implementation, quantitative testing in a larger sample, or an iterative second qualitative round - is a logical next step that this study does not undertake.

Several directions for future research are indicated by the findings.

- The most immediate is the empirical validation of the updated framework, ideally through a multi-organization case-study design in which the framework's components are mapped onto the actual governance arrangements of participating organizations and the gaps measured.
- A second direction is sectoral comparison: the finding that regulatory engagement is mediated through legal departments and central headquarters rather than reaching operational governance directly should be tested in highly-regulated sectors where the AI Act's provisions create more direct operational exposure (healthcare, financial services, public administration).
- A third direction is the longitudinal study indicated by the transitional state of governance at the time of interview: participants in this study were, in effect, describing their organizations' first-generation AI governance arrangements. Second-generation arrangements, after 12–24 months of operation, would likely produce different findings on which components stabilized, which were abandoned, and which evolved.
- Fourth direction concerns the inductive themes that did not pass the three-case promotion threshold but recurred at lower frequencies in the dataset AI cost governance as a discrete component, expertise bypass, client-mediated AI governance in outsourcing contexts, creative-authenticity tensions in communications work, and AI investment / operational-dependency risk in smaller organizations.

Each of these merits requires a dedicated investigation in samples chosen to surface the relevant phenomenon, and any of the four could productively become the focus of further research.

Conclusions

1. **The systematic literature review on shadow AI, shadow IT, AI governance, and innovation governance identified relevant theoretical foundations, existing approaches, and gaps that justified developing a new governance framework.**
   The analysis covered shadow IT definitions, drivers, and established controls; shadow AI as a distinct phenomenon characterized by generative, opaque, and autonomous properties; IT governance principles codified in ISO/IEC 38500 and ISO/IEC 42001; the regulatory environment shaped by the EU AI Act and GDPR; and the innovation control balance that influences whether governance is sustainable or bypassed. The review found that existing shadow IT research provides a solid basis for understanding unauthorized technology adoption but does not address AI's decision influencing and content generating nature. Existing AI governance standards, in turn, assume AI systems are known, approved, and formally managed by the organization. Shadow AI falls between these areas it is AI use that occurs before or outside formal governance. Two specific gaps were identified: the lack of an integrated framework that manages unauthorized AI adoption as a separate governance object, and the predominance of managerial-only perspectives in current research, which risks overestimating the operational maturity of governance arrangements. These gaps directly informed the design requirements for the conceptual framework developed in this thesis.
2. **The research identified the main reasons why SAI appears in organizations, the ways it is used, and the risks it creates.**
   Shadow AI adoption drivers were classified into four categories: individual and personal factors (productivity seeking, cognitive load reduction, skill gap compensation, convenience), organizational factors (governance immaturity, slow approval processes, inaccessible policies, absence of sanctioned alternatives, awareness gaps), technological factors (AI democratization through generative pre-trained transformer models, embedded AI in approved applications, browser extensions, locally executed models, personal device accessibility), and business and market factors (competitive pressure, executive enthusiasm, the mismatch between the speed of employee AI adoption and the pace of organizational governance response). A delivery-based typology of shadow AI tools was established, distinguishing public AI SaaS, embedded AI in approved SaaS, browser extensions and add-ins, and locally executed models - reflecting how employees actually adopt AI in practice and providing a lens for the design of enforcement mechanisms. The risk landscape was classified into six domains: data exposure and leakage; compliance and legal risk; AI output reliability and hallucination; accountability and decision-making rights gaps; intellectual property and confidentiality concerns; and operational dependencies and tool proliferation. Four observable governance failures were identified as the structural conditions that allow shadow AI to persist: lack of visibility, lack of accountability and decision rights, inconsistent rules and enforcement, and lack of evidence and auditability.
3. **A conceptual Shadow AI Governance Framework was developed to support the structured governance of Shadow AI in organizations..**
   The framework was built as a combination of SIT foundations, AI-specific risk analysis, established IT governance principles, regulatory requirements, and innovation–control dynamics. Layer 1 (Governance Foundation) sets the basic requirements for all governance

activities by defining governance roles and responsibilities, ensuring regulatory and compliance alignment, and establishing risk appetite.
Layer 2 (Core Governance Mechanisms) offers the operational core: visibility and discovery, policy and authorization systems, risk assessment and management, and technical controls.
Layer 3 (Innovation Enablement and Sustainability) focuses on the conditions that influence whether governance is followed or bypassed over time: approved AI programs, fast-track and tiered approval processes, as well as culture, awareness, and ongoing improvement. The layered setup follows a deliberate order - Layer 1 must be in place before Layer 2 can function, and Layer 2 must be operational before Layer 3 can have an impact - while Layer 3 provides feedback to Layers 1 and 2 through user insights, changing risks, and continuous enhancements. Including innovation enablement as its own layer directly tackles the core research issue: governance that only controls or only enables will either be bypassed by employees or be insufficient for managing risks.

4. **The empirical validation confirmed that the SAIGF is applicable in practice but also showed where the initial framework needed to be refined.**
Ten semi-structured interviews with participants from six organizations, including both managerial-level governance designers and non-managerial AI tool users, broadly confirmed the framework's three-layer structure and component logic. However, the empirical phase led to several important refinements. The risk appetite component was expanded to include AI cost governance, investment-failure risk, and operational dependency risk - factors that practitioners consistently prioritized alongside, and sometimes above, data leakage. Sanctioned AI provision was reframed from a catalog requirement to a competitiveness measure, as inadequately approved tools were shown to directly encourage shadow adoption. Two new sub-components emerged from inductive analysis: agentic AI permission governance, supported by a documented incident of repository compromise through an unsupervised AI agent, and embedded AI re-assessment, which addresses AI features that arrive in approved tools via product updates without triggering governance review. A cross-cutting human oversight element was introduced across all three layers to address the decision-quality risk, which emerged as the most heavily emphasized and negatively assessed finding in the dataset. The most significant cross-cutting finding was the manager–employee perception gap: governance artifacts are consistently present from the managerial perspective but remain inaccessible to employees, indicating that policy accessibility - not policy existence - is the key constraint on governance effectiveness. Based on these findings, six managerial recommendations were developed, focusing on policy accessibility, risk appetite definition including cost considerations, adequacy of sanctioned tools, fast-track approval pathways, human oversight of AI-assisted decisions, and the political rather than punitive handling of shadow AI. Six technical recommendations were also offered, covering AI-specific discovery and DLP capabilities, the personal device blind spot, least-privilege agentic permission architectures, embedded AI re-assessment triggers, cost telemetry, and AI-native defensive controls. The study's limitations include a small, geographically concentrated sample, a cross-sectional design, and reliance on self-reported data. Future research should validate the updated SAIGF through multi-organization case studies, compare findings across highly regulated sectors, conduct longitudinal studies to track governance evolution, and explore inductive themes that did not meet the promotion

threshold - including AI cost governance as a separate component, client-mediated governance in outsourcing, and creative-authenticity tensions in AI-assisted work.

List of references

1. Agarwal, R., & Karahanna, E. (2000). Time Flies When You're Having Fun: Cognitive Absorption and Beliefs About Information Technology Usage1. Management Information Systems Quarterly, 24(4), 665–694. https://doi.org/10.2307/3250951
2. Ameen, N., Tarhini, A., Shah, M. H., Madichie, N., Paul, J., & Choudrie, J. (2021). Keeping customers' data secure: A cross-cultural study of cybersecurity compliance among the Gen-Mobile workforce. Computers in Human Behavior, 114, 106531. https://doi.org/10.1016/J.CHB.2020.106531
3. Artificial Intelligence Index Report 2025. (2025).
4. Balogun, A. Y., Metibemu, O. C., Olutimehin, A. T., Ajayi, A. J., Babarinde, D. C., & Olaniyi, O. O. (2025). The Ethical and Legal Implications of Shadow AI in Sensitive Industries: A Focus on Healthcare, Finance and Education. Journal of Engineering Research and Reports, 27(3), 1–22. https://doi.org/10.9734/JERR/2025/V27I31414
5. Balzano, M., & Marzi, G. (2025). At the Cybersecurity Frontier: Key Strategies and Persistent Challenges for Business Leaders. Strategic Change, 34(2), 181–192. https://doi.org/10.1002/JSC.2622;WEBSITE:WEBSITE:PERICLES;JOURNAL:JOURNAL:10991697;ISSUE:ISSUE:DOI
6. Behrens, S. (2009). Shadow systems: The good, the bad and the ugly. Communications of the ACM, 52(2), 124–129. https://doi.org/10.1145/1461928.1461960;TAXONOMY:TAXONOMY:ACM-PUBTYPE;PAGEGROUP:STRING:PUBLICATION
7. Bhattacherjee, A., Davis, C. J., Connolly, A. J., & Hikmet, N. (2018). User response to mandatory IT use: a coping theory perspective. European Journal of Information Systems, 27(4), 395–414. https://doi.org/10.1057/S41303-017-0047-0
8. Bucher, T. (2019). The algorithmic imaginary: exploring the ordinary affects of Facebook algorithms. The Social Power of Algorithms, 30–44. https://doi.org/10.4324/9781351200677-3
9. Chin, T., Li, Q., Mirone, F., & Papa, A. (2025a). Conflicting impacts of shadow AI usage on knowledge leakage in metaverse-based business models: A Yin-Yang paradox framing. Technology in Society, 81. https://doi.org/10.1016/j.techsoc.2024.102793
10. Chin, T., Li, Z., Huang, L., & Li, X. (2025). How artificial intelligence promotes new quality productive forces of firms: A dynamic capability view. Technological Forecasting and Social Change, 216, 124128. https://doi.org/10.1016/J.TECHFORE.2025.124128
11. Clark, Tom., Foster, Liam., Sloan, Luke., & Bryman, Alan. (2021). Bryman's Social Research Methods. 670.
12. Creswell, J. W. (2014). Research and Design Qualitative, Quantitative, and Mixed Methods Approaches. SAGE Publication Inc.
13. Dalal, R. S., Howard, D. J., Bennett, R. J., Posey, C., Zaccaro, S. J., & Brummel, B. J. (2021). Organizational science and cybersecurity: abundant opportunities for research at the interface. Journal of Business and Psychology 2021 37:1, 37(1), 1–29. https://doi.org/10.1007/S10869-021-09732-9

14. Dang-Pham, D., & Pittayachawan, S. (2015). Comparing intention to avoid malware across contexts in a BYOD-enabled Australian university: A Protection Motivation Theory approach. Computers & Security, 48, 281–297. https://doi.org/10.1016/J.COSE.2014.11.002
15. Folorunso, A., Adewumi, T., Adewa, A., Okonkwo, R., Olawumi, T. N., Folorunso, A., Adewumi, T., Adewa, A., Okonkwo, R., & Olawumi, T. N. (2024). Impact of AI on cybersecurity and security compliance. Global Journal of Engineering and Technology Advances, 21(1), 167–184. https://doi.org/10.30574/GJETA.2024.21.1.0193
16. Furneaux, B., & Wade, M. (2010). The End of the Information System Life: A Model of IS Discontinuance. Data Base for Advances in Information Systems, 41(2), 45–69. https://doi.org/10.1145/1795377.1795381;REQUESTEDJOURNAL:JOURNAL:SIGMIS;JOURNAL:JOURNAL:SIGMIS;TOPIC:TOPIC:ACMSIG
17. Fürstenau, D., Rothe, H., & Sandner, M. (2020). Leaving the Shadow: A Configurational Approach to Explain Post-identification Outcomes of Shadow IT Systems. Business & Information Systems Engineering 2020 63:2, 63(2), 97–111. https://doi.org/10.1007/S12599-020-00635-2
18. Galletta, A. (2020). Mastering the Semi-Structured Interview and Beyond. Mastering the Semi-Structured Interview and Beyond. https://doi.org/10.18574/nyu/9780814732939.001.0001
19. Gregory, R. W., Kaganer, E., Henfridsson, O., & Ruch, T. J. (2018). It Consumerization and the Transformation of IT Governance1. Management Information Systems Quarterly, 42(4), 1225–1253. https://doi.org/10.25300/MISQ/2018/13703
20. Guest, G., Bunce, A., & Johnson, L. (2006). How Many Interviews Are Enough?: An Experiment with Data Saturation and Variability. Field Methods, 18(1), 59–82. https://doi.org/10.1177/1525822X05279903
21. Hu, Y., & Jia, X. (2025). Empowering the Intelligent Transformation of the Manufacturing Sector Through New Quality Productive Forces: Value Implications, Theoretical Analysis, and Empirical Examination. Sustainability 2025, Vol. 17, Page 7006, 17(15), 7006. https://doi.org/10.3390/SU17157006
22. Huber, M., Zimmermann, S., Rentrop, C., & Felden, C. (2018). Conceptualizing Shadow IT Integration Drawbacks from a Systemic Viewpoint. Mdpi.Com. https://doi.org/10.3390/systems6040042
23. Zimmermann, S., Rentrop, C., & Felden, C. (2014). Managing shadow IT instances – A method to control autonomous IT solutions in the business departments. In Proceedings of the 20th Americas Conference on Information Systems (AMCIS 2014). Savannah, Georgia, USA. https://aisel.aisnet.org/amcis2014/StrategicUse/GeneralPresentations/12
24. Kerr, D. V., Houghton, L., & Burgess, K. (2007). Power relationships that lead to the development of feral systems. Australasian Journal of Information Systems, 14(2), 141–152. https://doi.org/10.3127/ajis.v14i2.473
25. King, N. (2014). Using Templates in the Thematic Analysis of Text. Essential Guide to Qualitative Methods in Organizational Research, 256–270. https://doi.org/10.4135/9781446280119.n21
26. Kopper, A., Westner, M., & Strahringer, S. (2020). From Shadow IT to Business-managed IT: a qualitative comparative analysis to determine configurations for successful management of IT by business entities. Information Systems and E-Business Management, 18(2), 209–257. https://doi.org/10.1007/S10257-020-00472-6

27. Kshetri, N. (2025). Transforming cybersecurity with agentic AI to combat emerging cyber threats. Telecommunications Policy, 49(6), 102976. https://doi.org/10.1016/J.TELPOL.2025.102976
28. Machlev, R., Heistrene, L., Perl, M., Levy, K. Y., Belikov, J., Mannor, S., & Levron, Y. (2022). Explainable Artificial Intelligence (XAI) techniques for energy and power systems: Review, challenges and opportunities. Energy and AI, 9, 100169. https://doi.org/10.1016/J.EGYAI.2022.100169
29. Michael Quinn Patton. (2002). Qualitative Research Evaluation Methods (3rd ed.). https://aulasvirtuales.wordpress.com/wp-content/uploads/2014/02/qualitative-research-evaluation-methods-by-michael-patton.pdf
30. Ortbach, Kevin & Bode, Martin & Niehaves, Björn. (2013). What Influences Technological Individualization? – An Analysis of Antecedents to IT Consumerization Behavior. 19th Americas Conference on Information Systems, AMCIS 2013 - Hyperconnected World: Anything, Anywhere, Anytime. 5. https://www.researchgate.net/publication/258963087_What_Influences_Technological_Individualization_-_An_Analysis_of_Antecedents_to_IT_Consumerization_Behavior
31. PAUBOX. (2025). RPT.202510.ShadowAI. https://www.paubox.com/hubfs/Report%20Assets/RPT.202510.ShadowAI/RPT.202510.ShadowAI.pdf
32. Puthal, D., Mishra, A. K., Mohanty, S. P., Longo, A., & Yeun, C. Y. (2025a). Shadow AI: Cyber Security Implications, Opportunities and Challenges in the Unseen Frontier. SN Computer Science 2025 6:5, 6(5), 405-. https://doi.org/10.1007/S42979-025-03962-X
33. Puthal, D., Mohanty, S. P., Bhavake, S. A., Morgan, G., & Ranjan, R. (2019). Fog Computing Security Challenges and Future Directions [Energy and Security]. IEEE Consumer Electronics Magazine, 8(3), 92–96. https://doi.org/10.1109/MCE.2019.2893674
34. Raković, L., Sakal, M., Matković, P., & Marić, M. (2020). Shadow IT – Systematic Literature Review. Information Technology and Control , 49(1), 144–160. https://doi.org/10.5755/J01.ITC.49.1.23801
35. Recursive Error Amplification → Term. (n.d.). Retrieved January 22, 2026, from https://fashion.sustainability-directory.com/term/recursive-error-amplification/
36. Reto P. Grubenmann. (n.d.). ISO/IEC 42001: a new standard for AI governance. Https://Kpmg.Com/Ch/En/Insights/Artificial-Intelligence/Iso-Iec-42001.Html .Com. Retrieved January 23, 2026, from https://kpmg.com/ch/en/insights/artificial-intelligence/iso-iec-42001.html
37. Ross, J. A. J., Hibbert, L., & Moss, E. J. (2025). Shadow AI: Governance, Risk, and Organisational Resilience. International Conference on Artificial Intelligence, Computer, Data Sciences, and Applications, ACDSA 2025. https://doi.org/10.1109/ACDSA65407.2025.11166415
38. Shan, B., Liu, K., Lu, X., & Liu, X. (2025). Artificial intelligence, knowledge coupling, and dynamic capabilities in China's GEM listed enterprises: the role of human–AI collaboration. Journal of Knowledge Management. https://doi.org/10.1108/JKM-04-2025-0588/1276614
39. Silic, M., & Back, A. (2014). Shadow IT – A view from behind the curtain. Computers & Security, 45, 274–283. https://doi.org/10.1016/J.COSE.2014.06.007

40. Silic, M., & Lowry, P. B. (2020). Using Design-Science Based Gamification to Improve Organizational Security Training and Compliance. Journal of Management Information Systems, 37(1), 129–161. https://doi.org/10.1080/07421222.2019.1705512
41. Silic, M., Silic, D., & Kind-Trüller, K. (2025). From Shadow It to Shadow AI–Threats, Risks and Opportunities for Organizations. Strategic Change. https://doi.org/10.1002/JSC.2682;WGROUP:STRING:PUBLICATION
42. Spierings, A., Kerr, D., & Houghton, L. (2012). What drives the end user to build a feral information system? In Proceedings of the 23rd Australasian Conference on Information Systems (ACIS 2012). Geelong, Victoria, Australia. https://aisel.aisnet.org/acis2012/6
43. Strauss, A., & Corbin, J. (1998). Basics of Qualitative Research: Grounded Theory Procedures and Techniques, Sage, Newbury Park, CA. https://www.sciepub.com/reference/69708
44. Walters, Richard. (2013). Bringing IT out of the shadows. Network Security. 2013. 5–11. 10.1016/S1353-4858(13)70049-7. https://www.researchgate.net/publication/257523475_Bringing_IT_out_of_the_shadows
45. Wang, Z., Feng, T., & Zhang, Y. (2025). Decoding the AI carbon reduction code: Improving corporate carbon performance from the perspective of industry chain spillovers. Journal of Environmental Management, 391, 126596. https://doi.org/10.1016/J.JENVMAN.2025.126596
46. Weill, P., & Ross, J. W. (2004). IT governance: How top performers manage IT decision rights for superior results. Harvard Business School Press. https://www.researchgate.net/publication/236973378_IT_Governance_How_Top_Performers_Manage_IT_Decision_Rights_for_Superior_Results
47. Xu, W. (2025). Adoption of Artificial Intelligence in Business Management: Opportunities and Challenges. Journal of Digitainability, Realism & Mastery (DREAM), 4(09), 10–24. https://doi.org/10.56982/DREAM.V4I09.317
48. Yin, R. K. (2018). Case Study Research and Applications. Sixth Edition.

Appendices

1 Appendix. Interview guidelines for semi-structured interviews.

This appendix presents the consolidated semi-structured interview guide used across both Management-level and Employee (non-management) participant groups. The guide is organized by thematic domain, with a single shared question set used for both groups wherever possible to maximize cross-group comparability. Each question carries two annotations:

- [Must] or [Optional] — priority indicator to support time management within the 60-minute interview slot. [Must] questions are required to validate SAIGF components or support cross-group analysis; [Optional] questions add depth and may be skipped if time is constrained or already covered through respondent elaboration on a related must-ask question.
- [Management] or [Non-management] - group restriction. Untagged questions are posed to both groups.

Atvėrimas - Dalyvio kontekstas	
Opening - Participant Background	
SAIGF ryšys / SAIGF link:	
Tikslas	***Objective***
Nustatyti dalyvio profilį ir organizacinį kontekstą; suderinti kalbos registrą bei pradinį DI brandos lygį, kad būtų galima tinkamai pateikti tolesnius klausimus.	*Establish participant profile and organizational context; calibrate language register and baseline AI maturity to frame subsequent questions.*
Klausimas (lietuvių k.)	***Question (English)***
[Privalomas] Ar galėtumėte trumpai apibūdinti savo dabartines pareigas ir sektorių, kuriame veikia jūsų organizacija? Darbdavio įvardyti nereikia.	**[Must]** *Could you briefly describe your current role and the sector your organization operates in? You do not need to name your employer.*
[Privalomas] Apytiksliai kiek darbuotojų dirba jūsų organizacijoje, ir ar jūsų sektorius susiduria su konkrečiais reguliaciniais ar sektoriniais apribojimais, kurie veikia DI įrankių valdymą (pvz., finansinės paslaugos, sveikatos priežiūra, viešasis sektorius, gynyba)?	**[Must]** *Roughly how large is your organization in headcount, and does your sector face any specific regulatory or sectoral constraints that influence how AI tools are governed (e.g., financial services, healthcare, public sector, defence)?*
[Pasirenkamas] Kaip apibūdintumėte savo bendrą komforto lygį naudojant DI ir skaitmenines priemones kasdieniame darbe?	**[Optional]** *How would you describe your general comfort level with AI and digital tools in your day-to-day work?*
[Privalomas] Kaip apibūdintumėte bendrą DI diegimo tempą jūsų organizacijoje šiuo metu?	**[Must]** *How would you describe the general pace of AI adoption in your organization currently?*
[Pasirenkamas] [Vadovas] Kiek laiko užimate pareigas, kuriose esate atsakingas (-a) už IT, kibernetinį saugumą ar DI valdymą?	**[Optional] [Management]** *How long have you been in a role with responsibility for IT, cybersecurity, or AI governance?*
[Pasirenkamas] [Darbuotojas] Savais žodžiais - ką jums reiškia „DI“ jūsų darbo kontekste?	**[Optional] [Non-management]** *In your own words, what does 'AI' mean to you in the context of your work?*
1 tema - SAI paplitimas ir įrankių naudojimas	
Theme 1 - SAI Prevalence & Tool Usage	
SAIGF ryšys / SAIGF link: *L2 - Matomumas ir aptikimas / Visibility & discovery*	
Tikslas	***Objective***
Nustatyti empirinį DI naudojimo ir SAI paplitimo pagrindą tiek iš valdymo (sąmoningumo) pusės, tiek iš darbuotojų faktinės praktikos pusės.	*Establish the empirical baseline for AI usage and SAI prevalence from both the governance-side awareness and the employee-side actual practice.*
Klausimas (lietuvių k.)	***Question (English)***

[Privalomas] Kokius DI įrankius jūs (ar jūsų darbuotojai) šiuo metu naudojate ar buvote naudoję praeityje savo darbe?	**[Must]** *What AI tools do you (or your employees) currently use, or have used in the past, as part of your work?*
[Privalomas] Ar DI naudojate daugiausia per atskirus įrankius (pvz., ChatGPT, Claude, Gemini), per esamoje programinėje įrangoje integruotas DI funkcijas (pvz., Microsoft 365 Copilot, Google Workspace AI, Notion AI), ar abu būdus? Kur taikoma - ar kuris nors iš jų pasiekiamas per asmenines paskyras ar prenumeratas, o ne organizacijos?	**[Must]** *Are you using AI primarily through standalone tools (e.g., ChatGPT, Claude, Gemini), AI features embedded in existing software (e.g., Microsoft 365 Copilot, Google Workspace AI, Notion AI), or both? Where applicable, are any of these accessed through personal accounts or subscriptions rather than organizational ones?*
[Pasirenkamas] Kiek svarbūs šie DI įrankiai tapo kasdieniam darbo atlikimui?	**[Optional]** *How important have these AI tools become to how the work gets done day to day?*
[Privalomas] Kaip apibūdintumėte dabartinį DI įrankių naudojimo paplitimą ir mastą jūsų organizacijoje?	**[Must]** *How would you describe the current prevalence and scale of AI tool usage in your organization?*
[Pasirenkamas] [Vadovas] Kada jūsų organizacija pirmą kartą sužinojo, kad darbuotojai naudoja DI įrankius, kurie nebuvo oficialiai patvirtinti per valdymo procesus?	**[Optional] [Management]** *When did your organization first become aware of employees using AI tools that had not been formally approved through governance channels?*
[Pasirenkamas] [Vadovas] Ar tam tikri padaliniai ar funkcijos - pavyzdžiui, rinkodaros, inžinerijos, klientų aptarnavimo, finansų - yra labiau linkę diegti DI įrankius nei kiti? Kur matote didžiausią nesankcionuoto DI naudojimo koncentraciją?	**[Optional] [Management]** *Are particular departments or functions - for example marketing, engineering, customer service, finance - more prone to AI tool adoption than others? Where do you see the highest concentration of unsanctioned AI usage?*
2 tema - Politika ir įgaliojimai	
Theme 2 - Policy & Authorization	
SAIGF ryšys / SAIGF link: *L1 - Vaidmenys ir atskaitomybė / Roles & accountability; L2 - Politikos ir įgaliojimų sistema / Policy & authorization framework*	
Tikslas	***Objective***
Įvertinti valdymo struktūros sandarą, atsakomybę ir prieinamumą tiek iš politikos rengėjų, tiek iš jos adresatų perspektyvos; atskleisti komunikacijos spragas.	*Evaluate governance design, ownership, and accessibility from both the policy-author and policy-subject perspectives; surface communication gaps.*
Klausimas (lietuvių k.)	***Question (English)***
[Privalomas] Ar žinote kokias nors taisykles, gaires, politikas ar valdymo mechanizmus, kurie jūsų organizacijoje reglamentuoja DI įrankių naudojimą? Ar galėtumėte juos apibūdinti?	**[Must]** *Are you aware of rules, guidelines, policies, or governance mechanisms in your organization that govern AI tool usage? Could you describe them?*
[Privalomas] Ar žinote, ar darbe naudojami DI įrankiai yra oficialiai patvirtinti jūsų organizacijos?	**[Must]** *Do you know whether the AI tools used in your work are officially approved by your organization?*
[Privalomas] Kiek šios politikos yra prieinamos ir suprantamos darbuotojams, kurie nedirba valdymo (governance) pareigose?	**[Must]** *How accessible and understandable are these policies to employees outside governance roles?*
[Privalomas] [Vadovas] Kaip jūsų organizacijoje priimami sprendimai dėl DI valdymo ir kas turi formalius įgaliojimus juos priimti?	**[Must] [Management]** *How are AI governance decisions made in your organization, and who has formal authority over them?*
3 tema — Rizika, atitiktis ir duomenų tvarkymas	
Theme 3 — Risk, Compliance & Data Handling	
SAIGF ryšys / SAIGF link: *L1 - Reguliacinis suderinamumas / Regulatory alignment; L1 — Rizikos toleravimas / Risk appetite; L2 — Rizikos vertinimas / Risk assessment*	
Tikslas	***Objective***
Užfiksuoti rizikos suvokimą, atitikties laikyseną ir rizikos toleravimo kalibravimą iš abiejų perspektyvų; įvertinti realų informuotumą apie duomenų ir privatumo pasekmes.	*Capture risk perception, compliance posture, and risk-appetite calibration from both perspectives; assess felt awareness of data and privacy implications.*
Klausimas (lietuvių k.)	***Question (English)***

[Privalomas] Kai DI įrankiai naudojami darbe, kiek yra galvojama, kokia informacija ar duomenys jiems pateikiami?	**[Must]** *When AI tools are used for work, how much thought is given to what information or data is being shared with them?*
[Privalomas] Kurios su DI įrankių naudojimu susijusios rizikos jums kelia didžiausią susirūpinimą?	**[Must]** *Which risks associated with AI tool usage concern you most?*
[Pasirenkamas] Ar gavote iš savo organizacijos kokį nors mokymą, gaires ar pranešimus apie duomenų saugumą ir privatumą DI įrankių kontekste?	**[Optional]** *Have you received any training, guidance, or communications from your organization about data security or privacy in relation to AI tools?*
[Privalomas] [Vadovas] Ar jūsų organizacijoje yra formalus DI rizikų vertinimo procesas ar struktūra? Kaip apibrėžiamas priimtinos rizikos lygis ir kaip ši riba pritaikoma sprendimuose dėl įrankių patvirtinimo?	**[Must] [Management]** *Does your organization have a formal framework or process for assessing AI-related risks? How does it define what level of risk is acceptable, and how is that threshold operationalized in approval decisions?*
[Privalomas] [Vadovas] Kaip jūsų organizacija orientuojasi DI reguliacinėje aplinkoje, ypač atsižvelgiant į ES DI aktą ir duomenų apsaugos reikalavimus?	**[Must] [Management]** *How does your organization navigate the regulatory environment around AI, particularly the EU AI Act and data protection requirements?*
4 tema - Veiksniai ir motyvai	
Theme 4 - Drivers & Motivations	
SAIGF ryšys / SAIGF link: *Tarpgrupinė spragų analizė (informuoja visus sluoksnius) / Cross-group gap analysis (informs all layers)*	
Tikslas	***Objective***
Atskleisti motyvus ir sprendimo veiksnius tiek iš darbuotojų patirties, tiek iš vadovų suvokimo; sudaryti pagrindą tarpgrupinei valdymo suvokimo spragų analizei.	*Surface motivations and decision drivers from both lived employee experience and managerial perception; provide the basis for cross-group governance perception gap analysis.*
Klausimas (lietuvių k.)	***Question (English)***
[Privalomas] Kokios buvo pagrindinės priežastys, dėl kurių jūs (ar jūsų organizacijos darbuotojai) pradėjote naudoti DI įrankius darbe?	**[Must]** *What were the main reasons you (or employees in your organization) started using AI tools at work?*
[Pasirenkamas] Kaip žmonės jūsų organizacijoje paprastai sužino apie naujus DI įrankius, kuriuos verta išbandyti darbe — per kolegas, socialinius tinklus, tiekėjus, profesines bendruomenes ar kitus kanalus?	**[Optional]** *How do people in your organization typically discover or hear about new AI tools to try at work — through colleagues, social media, vendors, professional communities, or other channels?*
[Privalomas] Kai DI įrankiai darbe buvo naudojami neformalia tvarka — be oficialaus patvirtinimo — kokie veiksniai paprastai lemia tokį sprendimą?	**[Must]** *When AI tools have been used at work without going through official approval, what factors typically influence that decision?*
[Pasirenkamas] Ar kada nors norėjote naudoti DI įrankį darbe, bet nusprendėte to nedaryti dėl jūsų organizacijos taisyklių? Kas tada nutiko?	**[Optional]** *Have you ever wanted to use an AI tool at work but decided not to because of your organization's rules? What happened?*
[Privalomas] [Vadovas] Kiek esate įsitikinęs (-usi), kad darbuotojai supranta, kas yra ir kas nėra leidžiama DI naudojimo srityje?	**[Must] [Management]** *How confident are you that employees understand what is and is not permitted regarding AI usage?*
5 tema - Inovacijų ir kontrolės pusiausvyra bei trintis	
Theme 5 - Innovation–Control Balance & Friction	
SAIGF ryšys / SAIGF link: *L3 — Inovacijų įgalinimas / Innovation enablement*	
Tikslas	***Objective***
Tirti inovacijų ir kontrolės įtampą tiek iš valdymo struktūros, tiek iš realios patirties pusės; identifikuoti, kur valdymo trintis skatina apeinamąjį elgesį.	*Examine the innovation–control tension from both the governance-design and lived-experience sides; identify where governance friction drives workaround behavior.*
Klausimas (lietuvių k.)	***Question (English)***
[Privalomas] Kaip apibūdintumėte bendrą jūsų organizacijos požiūrį į naujas technologijas ir DI įrankius — įgalinantis, ribojantis, neaiškus ar dar kitoks?	**[Must]** *How would you describe your organization's general approach to new technology and AI tools — enabling, restrictive, unclear, or something else?*
[Privalomas] Ar kada nors bandėte oficialiai patvirtinti DI įrankį per savo organizaciją? Kokia buvo ši patirtis?	**[Must]** *Have you ever tried to get an AI tool officially approved through your organization? What was that experience like?*

[Pasirenkamas] Ar DI įrankiai paprastai finansuojami per organizacijos viešųjų pirkimų sistemą, padalinių biudžetus, ar per asmenines prenumeratas? Ar kaina, atsiskaitymo ar pirkimų trintis kada nors paveikė tai, kurie DI įrankiai realiai naudojami?	**[Optional]** *Are AI tools typically funded through organizational procurement, departmental budgets, or personal subscriptions? Has cost, billing, or procurement friction ever influenced which AI tools are actually used in practice?*
[Pasirenkamas] Kaip vertinate pusiausvyrą, kurią jūsų organizacija išlaiko tarp produktyvumo / inovacijų ir DI naudojimo valdymo poreikio?	**[Optional]** *How do you feel about the balance your organization strikes between productivity / innovation and the need to govern AI usage?*
[Pasirenkamas] [Vadovas] Ar esate susidūręs (-usi) su darbuotojų ar verslo padalinių pasipriešinimu DI valdymo priemonėms? Kaip jis buvo valdomas?	**[Optional] [Management]** *Have you experienced resistance from employees or business units regarding AI governance measures? How was this managed?*
6 tema - Stebėsena, techninės priemonės ir taikymas	
Theme 6 - Monitoring, Technical Controls & Enforcement	
SAIGF ryšys / SAIGF link: *L2 - Techninės kontrolės priemonės / Technical controls*	
Tikslas	***Objective***
Įvertinti stebėsenos, aptikimo ir taikymo brandą bei veiksmingumą; gilintis į konkrečias techninės kontrolės kategorijas, įvardytas SAIGF 2 sluoksnyje, kad būtų galima validuoti komponentų lygmenį.	*Assess maturity and effectiveness of monitoring, detection, and enforcement; probe specific technical-control categories named in SAIGF Layer 2 to enable component-level validation.*
Klausimas (lietuvių k.)	***Question (English)***
[Privalomas] Ar žinote kokius nors mechanizmus, kuriuos jūsų organizacija naudoja DI naudojimui aptikti ar stebėti? Ar esate su jais susidūręs (-usi)?	**[Must]** *Are you aware of any mechanisms your organization uses to detect or monitor AI usage? Have you encountered any?*
[Pasirenkamas] Ar jūsų organizacija patyrė kokį nors didesnį ar mažesnį incidentą, susijusį su netinkamu DI įrankių naudojimu, duomenų atskleidimu ar atsako klaidomis (pvz., haliucinacijomis ar šališka išvestimi, patekusia į realų darbo procesą)? Kaip jis buvo sprendžiamas ir kas dėl to pasikeitė?	**[Optional]** *Has your organization experienced any incident — large or small — related to AI tool misuse, data exposure, or output errors (such as hallucinations or biased output entering a real workflow)? How was it handled, and what changed as a result?*
[Privalomas] Kaip, jūsų patirtimi, jūsų organizacija reaguoja, kai aptinkamas nesankcionuotas DI naudojimas?	**[Must]** *How does your organization respond when unsanctioned AI usage is discovered, in your experience?*
[Privalomas] [Vadovas] Kokios konkrečios techninės kontrolės priemonės šiuo metu įdiegtos jūsų organizacijoje — duomenų nuotėkio prevencija (DLP), debesijos prieigos saugumo brokeriai (CASB), užklausų filtravimas, izoliuotos (sandbox) DI aplinkos ar kitos? Kurios pasirodė veiksmingiausios, o kurios — mažiausiai?	**[Must] [Management]** *Which specific technical controls are currently deployed in your organization — Data Loss Prevention (DLP), Cloud Access Security Broker (CASB), prompt filtering, sandboxed AI environments, or others? Which have proven most and least effective?*
7 tema — Įgalinimo mechanizmai	
Theme 7 — Enablement Mechanisms	
SAIGF ryšys / SAIGF link: *L3 — Sankcionuotas DI tiekimas / Sanctioned provision; L3 — Pagreitinto patvirtinimo procedūra / Fast-track approval; L3 — Kultūra ir sąmoningumas / Culture & awareness*	
Tikslas	***Objective***
Simetriškai patikrinti SAIGF 3 sluoksnio komponentus: ar organizacijos iš tikrųjų teikia įgalinimo mechanizmus ir ar darbuotojai juos suvokia kaip pakankamus arba reaguotų į patobulintas jų versijas.	*Test SAIGF Layer 3 components symmetrically: whether organizations actually provide enablement mechanisms, and whether employees perceive these mechanisms as sufficient or would respond to enhanced versions.*
Klausimas (lietuvių k.)	***Question (English)***
[Privalomas] Ar jūsų organizacija šiuo metu siūlo ar palaiko prižiūrimą iš anksto patvirtintų DI įrankių katalogą? Kiek gerai jis atitinka realius naudotojų poreikius?	**[Must]** *Does your organization currently offer or maintain a curated catalogue of pre-approved AI tools? How well do they meet actual user needs?*
[Privalomas] Ar jūsų organizacija taiko kokią nors pakopinę ar pagreitintą DI įrankių patvirtinimo procedūrą? Kokia jūsų patirtis su ja?	**[Must]** *Does your organization operate any tiered or fast-track approval pathway for AI tools? What is your experience of it?*

[Privalomas] Kokie sąmoningumo, mokymų ar grįžtamojo ryšio mechanizmai egzistuoja DI naudojimo srityje, ir kiek jie veiksmingi?	**[Must]** *What awareness, training, or feedback mechanisms exist around AI use, and how effective are they?*
[Pasirenkamas] [Darbuotojas] Jeigu jūsų organizacija įdiegtų greitą, aiškią patvirtinimo procedūrą (pvz., sprendimas per 2–3 darbo dienas) ar iš anksto patikrintų DI įrankių rinkinį, pritaikytą jūsų darbui — kiek tai pakeistų jūsų požiūrį į naujų įrankių naudojimą?	**[Optional] [Non-management]** *If your organization were to introduce a fast, clear approval process (e.g., a decision within 2–3 working days), or a set of pre-vetted AI tools tailored to your work, how likely would these mechanisms change how you approach using new tools?*
[Pasirenkamas] [Vadovas] Kaip užtikrinate, kad sankcionuotos alternatyvos išliktų realiai konkurencingos, palyginti su nesankcionuotais įrankiais, kuriuos darbuotojai kitu atveju pasirinktų?	**[Optional] [Management]** *How do you ensure sanctioned alternatives remain genuinely competitive with the unsanctioned tools employees might otherwise adopt?*
8 tema — Modelio vertinimas	
Theme 8 — Framework Evaluation	
SAIGF ryšys / SAIGF link: *Visi sluoksniai / All layers*	
Tikslas	***Objective***
Validuoti ir patikslinti teorinį SAIGF; iš valdymo praktikų ir galutinių naudotojų rinkti įrodymais grįstas rekomendacijas, kurie komponentai svarbiausi.	*Validate and refine the theoretical SAIGF; gather evidence-based recommendations from both governance practitioners and end users on which components matter most.*
Klausimas (lietuvių k.)	***Question (English)***
[Privalomas] Kuriuos DI valdymo modelio elementus, remdamiesi savo patirtimi, laikote svarbiausiais?	**[Must]** *Based on your experience, which elements of an AI governance framework do you consider most critical?*
[Privalomas] Ar yra valdymo komponentų, kurie, jūsų manymu, jūsų organizacijoje arba dabartinėse modeliuose apskritai yra trūkstami ar nepakankamai išplėtoti?	**[Must]** *Are there governance components that you feel are missing or underdeveloped in your organization, or in current frameworks more broadly?*
[Pasirenkamas] [Vadovas] Kaip šiuo metu vertinate, ar jūsų taikomas DI valdymo požiūris yra veiksmingas?	**[Optional] [Management]** *How do you currently measure whether your AI governance approach is effective?*
Uždarymas — Apibendrinimas ir ateities perspektyva	
Closing — Reflection & Forward View	
SAIGF ryšys / SAIGF link: —	
Tikslas	***Objective***
Užfiksuoti į ateitį orientuotas įžvalgas ir iškelti likusias temas, neaptartas ankstesniuose segmentuose; profesionaliai užbaigti interviu.	*Capture forward-looking insight and surface any residual themes not addressed in earlier segments; provide a professional close to the interview.*
Klausimas (lietuvių k.)	***Question (English)***
[Pasirenkamas] Žvelgiant į ateinančius dvejus–trejus metus, kokius didžiausius iššūkius matote organizacijoms valdant DI?	**[Optional]** *Looking ahead two to three years, what do you see as the most significant challenges organizations will face in governing AI?*
[Privalomas] Jeigu galėtumėte pakeisti vieną dalyką, kaip jūsų organizacija šiandien valdo šešėlinį DI ar DI įrankių naudojimą — kas tai būtų?	**[Must]** *If you could change one thing about how your organization currently approaches Shadow AI governance or AI tool usage, what would it be?*
[Privalomas] Kas, jūsų manymu, palengvintų ar padarytų natūralesnį darbuotojų lūkestį laikytis oficialių procedūrų diegiant naujus DI įrankius?	**[Must]** *What do you think would make it easier or more natural for employees to follow official processes when adopting new AI tools?*
[Pasirenkamas] Žvelgiant į priekį, ar tikitės, kad DI įrankių naudojimas jūsų darbe per artimiausius vienus–dvejus metus didės, mažės ar liks maždaug toks pat?	**[Optional]** *Looking ahead, do you expect AI tool usage in your work to increase, decrease, or stay roughly the same over the next year or two?*
[Privalomas] Ar yra dar kas nors apie DI įrankius ar šešėlinio DI valdymą jūsų darbe ar organizacijoje, kas, jūsų nuomone, yra svarbu, bet dar nebuvo aptarta?	**[Must]** *Is there anything about AI tools or Shadow AI governance in your work or organization that you feel is important and that we have not yet discussed?*