Hellenic Open University

Master in Business Administration (MBA)

Postgraduate Dissertation

A.I. for Security and Security for A.I.:

A Strategic Evaluation in the Financial Services Sector

Spyridon Dafalias

Supervisor: Athanasios Rentizelas

Athens, Greece, June 2026

Theses / Dissertations remain the intellectual property of students (“authors/creators”), but in the context of open access policy they grant to the HOU a non-exclusive license to use the right of reproduction, customisation, public lending, presentation to an audience and digital dissemination thereof internationally, in electronic form and by any means for teaching and research purposes, for no fee and throughout the duration of intellectual property rights. Free access to the full text for studying and reading does not in any way mean that the author/creator shall allocate his/her intellectual property rights, nor shall he/she allow the reproduction, republication, copy, storage, sale, commercial use, transmission, distribution, publication, execution, downloading, uploading, translating, modifying in any way, of any part or summary of the dissertation, without the explicit prior written consent of the author/creator. Creators retain all their moral and property rights.

A.I. for Security and Security for A.I.:

A Strategic Evaluation in the Financial Services Sector

Spyridon Dafalias

Supervising Committee

Supervisor:
Athanasios Rentizelas

Co-Supervisor:
Dimitrios Fouskakis

Athens, Greece, June 2026

"To my parents, wife and son, who support me in every way all these years,
I love you.

Challenges promote strength."

Abstract

This dissertation studies the critical relationship between Artificial Intelligence and Security within the Financial Services Sector (FSS). The research is structured around two main topics: “AI for Security” and “Security for AI”.

“AI for Security” represents the strategic value and opportunity of using AI in the FSS as a powerful defensive tool, for vulnerability management, threat analytics, and compliance monitoring. The aim is to identify which are the key benefits of adopting AI tools in the FSS and address how leaders take (or should take) advantage of them to be more competitive and resilient.

On the other hand, “Security for AI”, represents the strategic risk management of using AI and addresses the critical challenge of securing the same AI systems from data poisoning and adversarial attacks that target algorithms. The aim is to identify the key security risks from adopting AI tools in the FSS and address how leaders mitigate (or should mitigate) them to prevent disastrous financial and reputational harm.

Financial institutions are in a strategic race to deploy AI and gain competitive advantage. However, this quick adoption creates a new, high-risk attack surface. A compromised AI model could trigger financial losses, compliance regulatory penalties, and reputational losses. FSS senior leaders do not only have a technical problem to solve but also an important strategic dilemma to make on how to balance the clear ROI of AI driven security with the complex, unquantified risk of insecure AI. This research has a main objective to assess how financial organizations use AI to strengthen cybersecurity systems and to evaluate the security threats and vulnerabilities inherent to AI models.

Methodologically a mixed methods approach is applied on the research design. Primary data is collected through a semi structured questionnaire answered from senior experts in Greek FSS. Collected data is analyzed through descriptive statistics supported by qualitative answers, five reliability tests and twelve inferential hypothesis tests. Findings from statistical analysis are interpreted along with readings from literature review.

Results show that “AI for Security” and “Security for AI” represent a single organizational capability rather than two independent processes. Financial institutions have mostly moved beyond the pilot stages of experimental AI deployment to the integration of AI in their security operations. This is linked to actual results that strengthen the strategic value of AI and the realization of ROI. However, value is acknowledged through the level of threat/fraud detection automation, rather than their actual AI maturity stage.

Results also reveal a clear gap between AI deployment and AI governance. Governance maturity is not necessarily associated with deployment maturity stage, ROI or cross-team alignment. Instead, it appears to be mainly driven by external regulatory pressure and ERM integration maturity. Furthermore, risk perception appears to be mainly sector dependent rather than AI deployment maturity driven, with technology vendors having higher awareness of specific AI threats.

Overall, this study concludes that institutions treating “AI for Security” and “Security for AI” as a single organizational capability, are best positioned to gain competitive advantages and operational efficiency.

Keywords

Artificial Intelligence, Cyber Security, Financial Services, Governance Framework, Machine Learning, Regulatory Compliance

“Τεχνητή Νοημοσύνη για την Ασφάλεια και Ασφάλεια για την Τεχνητή Νοημοσύνη:

Μια Στρατηγική Αξιολόγηση στον Τομέα των Χρηματοπιστωτικών Υπηρεσιών”

“Σπυρίδων Δαφαλιάς”

Περίληψη

Αυτή η διατριβή μελετά την κρίσιμη σχέση μεταξύ Τεχνητής Νοημοσύνης και Ασφάλειας στον Τομέα Χρηματοοικονομικών Υπηρεσιών (FSS). Η έρευνα διαρθρώνεται γύρω από δύο κύρια θέματα: «Τεχνητή Νοημοσύνη για την Ασφάλεια» και «Ασφάλεια για την Τεχνητή Νοημοσύνη».

Ο όρος «Τεχνητή Νοημοσύνη για την Ασφάλεια» εκφράζει τη στρατηγική αξία και την ευκαιρία χρήσης της Τεχνητής Νοημοσύνης στον Τομέα Χρηματοοικονομικών Υπηρεσιών ως ένα ισχυρό αμυντικό εργαλείο, για τη διαχείριση τρωτών σημείων, την ανάλυση απειλών και την παρακολούθηση της κανονιστικής συμμόρφωσης. Στόχος είναι να προσδιοριστούν ποια είναι τα βασικά οφέλη από την υιοθέτηση εργαλείων Τεχνητής Νοημοσύνης στον Τομέα Χρηματοοικονομικών Υπηρεσιών και να εξεταστεί πώς οι ηγέτες τα εκμεταλλεύονται (ή θα έπρεπε να τα εκμεταλλευτούν) για να είναι πιο ανταγωνιστικοί και ανθεκτικοί οι οργανισμοί τους.

Από την άλλη πλευρά, ο όρος «Ασφάλεια για την Τεχνητή Νοημοσύνη» εκφράζει τη στρατηγική διαχείριση κινδύνου από τη χρήση της Τεχνητής Νοημοσύνης στον Τομέα Χρηματοοικονομικών Υπηρεσιών και αντιμετωπίζει την κρίσιμη πρόκληση της ασφάλειας των ίδιων συστημάτων Τεχνητής Νοημοσύνης από «δηλητηρίαση» δεδομένων και εχθρικές επιθέσεις που στοχεύουν τους αλγόριθμους. Στόχος είναι να προσδιοριστούν οι βασικοί κίνδυνοι ασφαλείας από την υιοθέτηση εργαλείων Τεχνητής Νοημοσύνης στον Τομέα Χρηματοοικονομικών Υπηρεσιών και να εξεταστεί πώς οι ηγέτες τους αντιμετωπίζουν (ή

θα έπρεπε να τους αντιμετωπίζουν) για να αποτρέψουν καταστροφικές οικονομικές απώλειες και απώλειες στη φήμη των οργανισμών τους.

Τα χρηματοπιστωτικά ιδρύματα βρίσκονται σε έναν στρατηγικό αγώνα για την ανάπτυξη της Τεχνητής Νοημοσύνης και την απόκτηση ανταγωνιστικού πλεονεκτήματος. Ωστόσο, αυτή η γρήγορη υιοθέτηση δημιουργεί μια νέα επιφάνεια επίθεσης υψηλού κινδύνου. Ένα παραβιασμένο μοντέλο Τεχνητής Νοημοσύνης θα μπορούσε να προκαλέσει οικονομικές απώλειες, κυρώσεις συμμόρφωσης με τους κανονισμούς και απώλειες φήμης. Τα ανώτερα στελέχη του Τομέα Χρηματοοικονομικών Υπηρεσιών δεν έχουν μόνο ένα τεχνικό πρόβλημα να λύσουν, αλλά και ένα σημαντικό στρατηγικό δίλημμα σχετικά με το πώς να εξισορροπήσουν την σαφή απόδοση επένδυσης (ROI) της ασφάλειας που βασίζεται στην Τεχνητή Νοημοσύνη με τον πολύπλοκο, μη ποσοτικοποιημένο κίνδυνο της ανασφαλούς Τεχνητής Νοημοσύνης. Αυτή η έρευνα έχει ως κύριο στόχο να αξιολογήσει τον τρόπο με τον οποίο οι χρηματοπιστωτικοί οργανισμοί χρησιμοποιούν την Τεχνητή Νοημοσύνη για την ενίσχυση των συστημάτων κυβερνοασφάλειας και να αξιολογήσει τις απειλές και τα τρωτά σημεία ασφαλείας που είναι εγγενή στα μοντέλα Τεχνητής Νοημοσύνης.

Μεθοδολογικά, στον σχεδιασμό της έρευνας εφαρμόζεται μια προσέγγιση μικτών μεθόδων. Τα πρωτογενή δεδομένα συλλέγονται μέσω ενός ημιδομημένου ερωτηματολογίου που απαντάται από έμπειρους ειδικούς στον Ελληνικό Τομέα Χρηματοοικονομικών Υπηρεσιών. Τα συλλεγόμενα δεδομένα αναλύονται μέσω περιγραφικής στατιστικής υποστηριζόμενης από ποιοτικές απαντήσεις, πέντε δοκιμών αξιοπιστίας και δώδεκα δοκιμών επαγωγικών υποθέσεων. Τα ευρήματα από τη στατιστική ανάλυση ερμηνεύονται μαζί με τις αναγνώσεις από την ανασκόπηση της βιβλιογραφίας.

Τα αποτελέσματα δείχνουν ότι η «Τεχνητή Νοημοσύνη για την Ασφάλεια» και η «Ασφάλεια για την Τεχνητή Νοημοσύνη» αντιπροσωπεύουν μια ενιαία οργανωτική ικανότητα και όχι δύο ανεξάρτητες διαδικασίες. Τα χρηματοπιστωτικά ιδρύματα έχουν ως επί το πλείστων προχωρήσει πέρα από τα πιλοτικά στάδια της πειραματικής ανάπτυξης της Τεχνητής Νοημοσύνης στην ενσωμάτωσή της στις λειτουργίες ασφαλείας τους. Αυτό συνδέεται με πραγματικά αποτελέσματα που ενισχύουν τη στρατηγική αξία της Τεχνητής Νοημοσύνης και την επίτευξη απόδοσης επένδυσης (ROI). Ωστόσο, η αξία αναγνωρίζεται

μέσω του βαθμού αυτοματοποίησης ανίχνευσης απειλών/απάτης, παρά μέσω του πραγματικού σταδίου ωριμότητας της Τεχνητής Νοημοσύνης.

Τα αποτελέσματα αποκαλύπτουν επίσης ένα σαφές χάσμα μεταξύ της ανάπτυξης της τεχνητής νοημοσύνης και της διακυβέρνησης της τεχνητής νοημοσύνης. Η ωριμότητα της διακυβέρνησης δεν σχετίζεται απαραίτητα με την ωριμότητα της ανάπτυξης, την απόδοση επένδυσης (ROI) ή την ευθυγράμμιση μεταξύ ομάδων. Αντίθετα, φαίνεται ότι καθοδηγείται κυρίως από την εξωτερική ρυθμιστική πίεση και την ωριμότητα ενσωμάτωσης στην διαχείριση επιχειρηματικού κινδύνου (ERM). Επιπλέον, η αντίληψη του κινδύνου φαίνεται να εξαρτάται κυρίως από τον τομέα και όχι από το στάδιο ωριμότητας ανάπτυξης της τεχνητής νοημοσύνης, με τους προμηθευτές τεχνολογίας να έχουν υψηλότερη επίγνωση συγκεκριμένων απειλών της τεχνητής νοημοσύνης.

Συνολικά, αυτή η μελέτη καταλήγει στο συμπέρασμα ότι τα ιδρύματα που αντιμετωπίζουν την «Τεχνητή Νοημοσύνη για την Ασφάλεια» και την «Ασφάλεια για την Τεχνητή Νοημοσύνη» ως μια ενιαία οργανωτική ικανότητα, βρίσκονται σε καλύτερη θέση για να αποκτήσουν ανταγωνιστικά πλεονεκτήματα και λειτουργική αποτελεσματικότητα.

Λέξεις – Κλειδιά

Τεχνητή Νοημοσύνη, Κυβερνοασφάλεια, Χρηματοοικονομικές Υπηρεσίες, Πλαίσιο Διακυβέρνησης, Μηχανική Μάθηση, Κανονιστική Συμμόρφωση

Table of Contents

List of Figures

List of Tables

List of Abbreviations & Acronyms

AI	Artificial Intelligence
AIBOM	AI Bill of Materials
AIRS	AI Risk Scanning
AML	Adversarial Machine Learning
ATLAS	Adversarial Threat Landscape for Artificial Intelligence Systems
CCF	Credit Card Fraud
CCFD	Credit Card Fraud Detection
DevSecOps	Developer Security Operations
DL	Deep Learning
DLP	Data Loss Prevention
DNN	Deep Neural Networks
DoS	Denial of Service
ERM	Enterprise Risk Management
FSS	Financial Services Sector
GDPR	General Data Protection Regulation
IDS	Intrusion Detection Systems
IOA	Indicators of Attack
IQR	Interquartile Range
ISACA	Information Systems Audit and Control Association
KQL	Kusto Query Language
LLM	Large Language Models
MCP	Model Context Protocol
MHO	Meta-Heuristic Optimization
ML	Machine Learning
MLOps	Machine Learning Operations
MLSecOps	Machine Learning Security Operations
MRM	Model Risk Management
MTTD	Mean Time to Detect
MTTR	Mean Time to Respond
MVMO	Maximum Violated Multi-Objective
NIST	National Institute of Standards and Technology

OECD	Organization for Economic Co-operation and Development
OWASP	Open Worldwide Application Security Project
PGD	Projected Gradient Descent
PII	Personal Identifiable Information
RACI	Responsible, Accountable, Consulted, Informed
ROI	Return of Investment
SBOM	Software Bill of Materials
SID	Sensitive Information Disclosure
SIEM	Security Information and Event Management
SIFIs	Systemically Important Financial Institutions
SIRTs	Security Incident Response Teams
SOAR	Security Orchestration, Automation, and Response
SOC	Security Operations Centers
XAI	Explainable Artificial Intelligence

1. Introduction

1.1 Problem Statement

Financial institutions are in a strategic race to deploy AI to gain competitive advantage, operational efficiency and a powerful defensive tool, for vulnerability management, threat and fraud detection. Mastercard (2025) states that over 80% of respondents in the banking industry reported seeing returns from using AI against fraud, while speeding up their fraud prevention processes and gaining customer satisfaction. At the same time, this quick adoption introduces a new high-risk attack surface, with AI systems prone to data poisoning and adversarial attacks that target algorithms. Financial Stability Board (2024) warns about these specific AI related vulnerabilities in the global financial system that pose risks to financial institutions due to the fast paced AI deployment. A compromised AI model can trigger financial and reputational losses, as well as compliance regulatory penalties. Senior leaders in the FSS do not only have a technical problem to solve but a strategic dilemma on how to balance the clear ROI of AI driven security with the complex, unquantified risk of insecure AI.

1.2 Purpose of the Dissertation

This dissertation has a main objective to present a strategic evaluation of the relationship between AI and Security in the Financial Services Sector. It is structured around two main topics: “AI for Security” and “Security for AI”. The first topic assesses how financial institutions can leverage AI as a defensive tool for fraud detection, threat analytics, SOC automation and compliance monitoring and how they can translate into business value. The second topic assesses the emerging risks of using AI, including model opacity, adversarial threats, supply-chain vulnerabilities, and unsanctioned use. The purpose is to provide FSS leaders with a balanced view of opportunities and risks, as well as to highlight the governance and technical controls required to exercise them properly.

1.3 General Methodology and Approach

The research contains a structured literature review with information from current industry data in tech and finance, both from a technological and a strategic point of view. The literature review is divided in two main parts which contain six pillars: Part A reviews the strategic literature on management, business value, regulation, and corporate governance (Pillars A1, A2, A3) and Part B reviews the technical literature on algorithms, exploits, technical value and engineering controls (Pillars B1, B2, B3). Evidence is drawn from peer-reviewed academic publications, regulatory frameworks, standards, industry reports and journals.

The literature review is complemented by an original primary data collection tool: a custom semi structured questionnaire. Questions came from the six pillars of the literature review and data were collected via Google Forms and one-to-one interviews. Their aim is to obtain quantitative data (5-point Likert Scales and Percentages) as well as qualitative data that support some of the quantitative ones. Processing the results is conducted through statistical analysis. It includes (i) descriptive statistics, (ii) Cronbach's Alpha reliability tests of composite values, (ii) hypothesis tests on the relationships between AI maturity, perceived ROI, governance maturity, perceived AI-attack impact, etc. and (iv) comparative tests between groups of respondents based on their role or sector.

1.4 Contribution and Significance

The most fundamental conclusion of this work is that “AI for Security” and “Security for AI” cannot be treated as separate topics. The combination of those brings value and maturity in the Financial Services Sector. It is safe to say that the value of the first, “AI for Security” is conditional on the maturity of the second, “Security for AI” and vice versa.

This dissertation contributes to the academic community by combining two different literatures that are rarely seen together. The strategic management research and the technical research for two separate topics that, combined together, bring value and maturity in FSS organizations: “AI for Security” and “Security for AI”.

Another main contribution of this dissertation is the empirical evaluation of senior practitioners’ perception, regarding the two topics. This evaluation is based on processing

original data, that were obtained through survey questionnaires and interviews, rather than relying only on literature review sources. By providing evidence from people who actually deploy and defend AI, this dissertation further contributes to academic discussion and reflects a snapshot of the current status in the Greek FSS.

1.5 Outline of the Dissertation

The following chapters are:

Chapter 2. Literature Review: This chapter contains the review of the existing literature about the two topics of "AI for Security" and "Security for AI". It establishes the theoretical foundation for the dissertation on how financial institutions use AI to strengthen security and to evaluate the security threats and vulnerabilities of the AI models themselves.

Chapter 3. Research Methodology: This chapter contains the research methodology approach. It describes the research design and methodological approach, the data source collection method, data weaknesses and limitations and how data was manipulated to reach the final results.

Chapter 4. Data Review and Results: This chapter contains the findings from the primary data collection research. It contains the descriptive statistics and inferential statistics tests supplemented with the relevant analysis, figures and tables.

Chapter 5. Discussion: This chapter contains an assessment of the findings in association with the existing literature and how do they respond to the research questions.

Chapter 6. Conclusion and Recommendations: This chapter contains a summary of the main conclusions derived from the assessment, with a summary of limitations and recommendations for future use.

2. Literature Review

2.1 Introduction: The Strategic Landscape of AI in Finance

The integration of Artificial Intelligence (AI) into the Financial Services Sector (FSS) represents a fundamental change in the economics of banking sector, as it enables them to move from simple digitization to advanced prediction. In this transformation journey, AI provides a significant economic advantage, as well as a high perceived value in prediction and analysis for key operations, like credit scoring, algorithmic trading, fraud detection and risk modeling, that were previously considered too complicated and the costs were high.

However, the research into the overall effect of using AI in the banking sector shows that there are risks incorporated, along with gaining advantages. While AI can effectively process large, unstructured data and identify hazards that human analysts could overlook, there is also a high risk of manipulating AI models by adding a new attack surface which makes them fragile.

This review is divided into two parts, evaluating the strategic improvements and dangers, as well as the technical advancements and risks when using AI.

- Part A: Strategic Literature (Management, Business Value, Regulation and Corporate Governance).
- Part B: Technical Literature (Algorithms, Exploits, Technical Value and Engineering Controls).

2.2 Pillar A1: AI for Security (The Strategic Opportunity)

Pillar A1 evaluates AI as a strategic opportunity in financial services, in three areas: The transition from simple automation to advanced prediction, its efficiency in fraud detection over traditional rule-based models and the empirical evidence on adoption and ROI. The literature indicates that AI delivers competitive advantage only when treated as an organizational capability rather than a software acquisition.

2.2.1 From Automation to Strategic Prediction

Artificial intelligence in banking has advanced quickly, moving from process automation to sophisticated forecasting capabilities. While, in former years process automation was the most efficient procedure, AI now gives a competitive advantage with cognitive insight. Tobias Adrian, Financial Counsellor and Director of the Monetary and Capital Markets Department, supports this view, pointing out that human analysis is becoming obsolete when it comes to vast amounts of unstructured data, in modern finance (IMF, 2024). For the past ten years, tech-savvy investment firms have been using machine learning and neural networks to analyze data and now with the use of large language models, they are able to enhance their powerful analytic tools.

McKinsey & Company (2025) highlights that this change is bringing value as AI enables banking risk professionals to concentrate on high-value decision-making rather than gathering data from long regulatory texts and threat intelligence reports. Especially, with generative AI models, firms can automatically generate new loan term sheets by analyzing similar loans that have already been conducted, reducing manual work, speeding up the closing process and ultimately enhancing the borrower's overall experience.

At the same time, banking sector is transforming its business model with the use of the same AI tools, thus moving its strategy from detecting events to predicting future events. Ahirrao et al. (2025) states that AI predictive analytics has become a revolutionary approach to modern financial planning strategy and risk management, as it allows organizations to harness the power of large data volumes to make even more accurate predictions.

Ajuwon et al. (2025) reinforces the statement that AI's impact on the financial services transformation journey goes beyond simple tasks automation and changes how organizations manage portfolios, assess risks and forecast market trends, making data analysis, modeling and decision-making more effective and efficient.

2.2.2 Efficiency in Fraud Detection

AI offers a remarkable strategic value in fraud detection with its real-time data analysis, behavior pattern analysis and continuous improving accuracy. In contrast, legacy systems have certain limits, because they rely on predefined parameters, such as using rule-based mechanisms (ex. if transaction > 50.000€, then flag). Paul and Ogburie (2025) provide a

critical comparison, arguing that traditional systems are inflexible and reactive compared to AI fraud detection systems. While early studies on fraud detection relied on statistical and rule-based models, they were limited in their adaptability to new and evolving fraudulent transactions. On the other hand, AI fraud detection systems can significantly reduce false positives and improve the accuracy of identifying new fraud transaction patterns.

This is further backed up by many researchers who witness a transformation in financial firms' approach to enhance their fraud detection systems. Traditional systems are not sufficient anymore due to a high rate of false positives, whereas AI can leverage contextual data to make sophisticated decisions. Faizal et al. (2025) points out that due to rapid digitalization of transactions, financial fraud has progressed making old rule-based systems insufficient. AI offers powerful analytics through processing large amounts of financial information, thus it greatly enhances fraud detection and risk management.

In addition, the rise of online sales in the twenty-first century has made credit card fraud (CCF) a severe threat. Current, rule-based fraud detection systems cannot keep up with the latest technologies used for criminal activities, leading to major financial losses. This leads to an increasing demand for using deep learning (DL), machine learning (ML) and meta-heuristic optimization (MHO) algorithms to successfully identify fraud (Hafez et al., 2025).

2.2.3 Adoption, Investment, and ROI

Boards typically require evidence that investments in AI deliver measurable returns. The integration of generative AI in financial risk management has reportedly produced a "*25% increase in detection accuracy and a 30% improvement in operational efficiency*" (Joshi, 2025 as cited in Poonum et al., 2025).

So, one would expect that all financial firms are on board, but the truth is that adoption is uneven. ACFE's (2024) anti-fraud technology benchmarking report highlights a gap between intent and execution: "*Although 83% of organizations plan to use AI to eliminate fraud, the actual implementation lags because of data quality barriers and budget limitations*" (ACFE, 2024).

This pattern indicates that AI implementation is more of a managerial challenge than a technological one. AI delivers competitive advantage only when treated as an organizational capability rather than a software acquisition. Organizations need to develop a distinct set of

resources and must be able to coordinate them appropriately, to successfully translate their investments into business value. Consistent with this view, Mikalef and Gupta (2021) reported results showing that "AI capability has a positive effect on organizational creativity and performance".

This leads to the paradox of rising investment and elusive returns (Horton et al., 2025), where boards discuss openly about the AI race, but the reality is more complicated. While organizations are investing more, the ROI becomes harder to measure and slower to materialize. Deloitte's (2025) survey showed that "*85% of organizations increased their investment in AI and 91% plan to increase it again this year, yet 67% of them are expecting a return of investment in 2 to 4 years, when a typical technological ROI period is about 7 to12 months*". Lead executives mentioned that actual ROI is difficult to measure because many of the benefits of AI initiatives are intangible.

2.3 Pillar A2: Security for AI (The Emerging Strategic Risk)

Pillar A2 addresses the emerging risks of using AI models, in three areas: Model opacity, adversarial threats and shadow AI. It establishes that AI models should be treated as a corporate asset that must be defended and explores their vulnerabilities.

2.3.1 The "Black Box" and Systemic Risk

Risk management processes are greatly affected by the opacity of AI models, because they are unable to confirm if a model is acting on past biases or valid financial indicators, thus it is difficult to understand it's logic behind decisions. Lumenova AI (2025), points out that AI models excel at using data to make informed decisions, but if the historical data they are trained on contain biases, then those biases are likely to be reinforced, leading to unfair decisions.

In the financial sector, biased results were observed in algorithms for determining credit card limits, where higher limits were allocated to men than to women (Firth, 2021, as cited in Solaimani & Long, 2025). Also, there were serious concerns about privacy, security and bias in chatbots that are gradually used in the financial services sector (Srivastava et al., 2024, as cited in Solaimani & Long, 2025).

To further support the "Black Box" argument, the Bank of England (2023), states that this concern could lead to "herd behavior" and "market manipulation", which are going to disrupt the financial system if more than a few banks employ similar biased models that respond to market stress in the same incorrect way.

2.3.2 Adversarial Threats as Business Risks

Until now, attackers were trying to hack infrastructures, but nowadays, in the era of AI, attackers are trying to hack logic instead. Global Risk Institute (2024), states that adversarial machine learning (AML) is a deliberate attack against financial stability, because attackers can now "poison" credit models or "evade" fraud filters with deceptive inputs inside a large ML model.

AML has a nuanced role in malware and network intrusion detection, because on one hand it can uncover vulnerabilities in ML algorithms to enhance defense, but on the other hand it can act as an attack vector to manipulate detector behavior (Venturi & Zanasi, 2021, as cited in Ijiga et al., 2024).

As of writing this literature review, on April 2026 Claude Mythos was released by Anthropic in a limited preview form, which according to Carlini et al. (2026) demonstrates significant adversarial capabilities, including the autonomous identification and exploitation of software vulnerabilities. This caused a huge disturbance in the banking industry as financial institutions are rushing to close security gaps, by implementing system patching more often, to avoid zero-day vulnerabilities, or else they are risking exposure and consequently business disruption.

Organizations now face bigger risks, as AI can leverage attackers with the ability to create a new threat landscape, by expanding existing threats and introduce new ones, where attacks are automated, scalable and elusive, while lowering the entry barrier. Attacks are expected to be more effective, more finely targeted, more difficult to attribute and now by far more likely to exploit vulnerabilities (Brundage, et al., 2018).

2.3.3 Shadow AI

A new organizational risk named Shadow AI has emerged as a result of the rapid democratization of generative AI. Shadow AI extends the concept of Shadow IT and

essentially describes the unsanctioned and unmonitored use of external AI applications by employees who need to increase their productivity outside of approved governance frameworks (Silic et al., 2025), like an official Enterprise Risk Management (ERM) channel.

The phenomenon is already widespread: Gartner (2025) reports that "*69% of organizations either suspect or have proof that employees are using banned public generative AI tools at work. By 2030 more than 40% of organizations worldwide will have to address at least one security or compliance incident tied to unauthorized use of AI*".

For financial institutions, the strategic and regulatory implications are especially serious. When employees paste a client's personal identifiable information (PII), sensitive business data or proprietary code into public LLMs, organizations lose control over their data in terms of residency, intellectual property and audit traceability. The OWASP Top 10 for LLM Applications (2024) classifies this as a critical vector for Sensitive Information Disclosure (SID), noting that without centralized governance, Shadow AI bypasses existing Data Loss Prevention (DLP) controls and exposes institutions to regulatory penalties driven by frameworks such as the EU AI Act and GDPR.

According to IBM's (2025) Cost of a Data Breach Report, "*20% of studied organizations had already experienced a breach linked to unsanctioned AI use. Furthermore, the shadow-AI-related incidents resulted in an average of $670,000 total breach costs for exposed customer personally identifiable information (PII) and intellectual property*".

2.4 Pillar A3: Strategic Governance and Regulation

Pillar A3 addresses how leadership should align the two conflicting pillars of strategic opportunity and emerging risks, to pursue compliance and support their governance frameworks against the speed and complexity of AI.

2.4.1 The Regulatory Compliance

The EU AI Act (2024), created by the European Union Council, is the strictest regulatory compliance framework, shifting financial firms' strategies from applying voluntary principles to comply with law requirements. The regulation follows a structured risk-based approach, that differentiates uses of AI among creating (i) an unacceptable risk, (ii) a high risk, (iii) limited risk and (iv) minimal risk. Especially for AI systems used in finance, such

as credit scoring, the EU AI Act marks them as high-risk AI systems, requiring them to be transparent, robust and strictly controlled.

For the financial services sector, it means that institutions are bound by AI specific legal requirements, especially those whose systems fall into the categories of credit scores, risk assessments, insurance, etc. and are considered as high-risk AI systems. This means that financial institutions need to develop additional capabilities, such as technical and legal expertise in AI to overcome any implications due to strict regulations (Almada, 2025).

One the other hand, strict regulations and agile technology, often create a gap, that should be addressed. Akinrele (2025), states that current legislation is ineffective, because technology advances quickly. To address that gap, he proposes a "Principle-based" regulation system that prioritizes regulating results over technical implementations to bridge AI Governance and Financial Regulation.

Furthermore, the Organization for Economic Co-operation and Development, states in their OECD AI Principles (2023) that there should be an accountability matrix that places AI actors directly accountable for any AI results, thus preventing the shifting of accountability only to AI software providers, for risks related to harmful bias, human rights including safety, security, privacy, as well as labor and intellectual property rights.

2.4.2 Governance Frameworks

The Information Systems Audit and Control Association (ISACA) released a framework to comply with these regulations in 2025. According to them, financial firms should enforce the COBIT framework, meaning that AI governance should be incorporated in the enterprise risk management (ERM) process. The framework's structured approach provides a holistic, life-cycle based model that guides organizations on how to align AI initiatives with strategic business objectives, to mitigate AI specific risks.

McKinsey & Company (2025) claims that current governance frameworks are fragmented, because existing Model Risk Management (MRM) committees might not have the technical know-how to assess the risks of AI. To address risks associated with AI, financial firms should support a collaborative effort among legal, risk and tech teams.

Furthermore, Gartner issued a report in 2024, claiming that there is a gap between adoption and governance, meaning that internal auditors lack the confidence and expertise to

successfully oversee AI control failures and unreliable outputs, even though these risks are growing, due to the technological speed of AI.

In conclusion, as Mosch (2025) states, governance is a strategic asset rather than a cost of compliance, thus financial firms with a stronger governance model may be able to implement AI more quickly and securely than others, which might lead in a competitive advantage. That said, financial institutions can reposition as top players when they have "strong leadership commitment, advanced digital infrastructure, effective data governance and an innovation-orientated culture".

2.5 Pillar B1: Technical Advantages in Operations (AI for Security)

Pillar B1 evaluates the technological mechanisms that enable the strategic opportunities reviewed in part A. The literature contains information about automation and telemetry in security operations centers, an analysis of detection systems and the use of AI in security orchestration, automation and response platforms.

2.5.1 SOC Automation & Telemetry

One of AI's key advantages is that it can help the Security Operations Center (SOC) of financial firms to be more efficient, reducing time for detection and respond. AI systems can consume raw telemetry, filter it out, take actions for trivial tasks and send high-fidelity warnings to humans. As Khalid, et al. (2024) mention in their research, the remediation of threats is mainly with the introduction of AI-driven playbooks that reduce the mean time to detect (MTTD) and respond (MTTR). Also, AI goes a step further, improving insider threat detection by analyzing behaviors and reporting deviations from normal user behavior. This way financial institutions can protect themselves from exfiltrating valuable information from someone inside their company.

Furthermore, with "Deep Learning" AI systems can read high-velocity logs and detect complicated, non-linear attack patterns, in contrast to traditional Security Information and Event Management (SIEM) systems, which rely mainly on rules. Khayat et al. (2025), point out that newly developed AI-powered SIEM systems can use complex algorithms to enhance threat detection, by searching for patterns that may have been missed during manual analysis. Nowadays it's easier to recognize false-positives and generate meaningful alerts.

Anashkin and Zhukova (2022) mentioned that the combination of behavioral indicators and attack indicators into Behavioral Indicators of Attack (BIOA), is the next step in cybersecurity. SOC teams are able to predict and identify patterns that suggest malicious intent, for future incidents before they happen. With the help of predictive analytics, AI can identify possible cybersecurity incidents before they actually happen, giving organizations the ability to improve their defense.

2.5.2 Intrusion Detection and Malware Detection Systems

Hussein (2024), published a technical overview of machine learning use in intrusion prevention systems, comparing current signature-based models and behavior-based models. In this technical overview, he claims that current signature-based systems can detect known threats, whereas behavior-based systems can also detect zero-day vulnerabilities, with the probability of alerting false alarm detections.

Sarker (2021) adds that transformer-based networks that employ “attention mechanisms” can recognize false alerts by employing the context of network packets more effectively than current systems, with the use of Deep Neural Networks (DNN). These networks (DNN) can analyze complicated patterns and detect malicious packets even in encrypted network traffic, making intrusion detection and malware analysis more efficient.

According to Hasanah (2025), techniques used by machine learning models, such as Gradient Boosting or XGBoost are better at detecting malicious code in malware analysis, because these models do not examine the actual code, but they concentrate at the execution path, to combat obfuscation. He continues claiming that several models are exceeding 90% on benchmark datasets such as “Virus Share”.

2.5.3 Generative AI in Security Orchestration, Automation and Response (SOAR)

Cyber Security Companies can now leverage Generative AI’s power to provide more intelligent products. Horschig (2023) from Trellix, explains that the implementation of Large Language Models (LLMs) in SOAR platforms, enables security teams to make better decisions, based on dynamic response generated information, analyzed by AI. Until now security teams were depending on pre-coded playbooks, which is the foundation of SOAR

platforms. Nowadays, AI can examine information from different data sources and provide useful recommendations for remediation.

Additionally, Jakkal (2023) states that Microsoft expanded it's Security platform with "security co-pilot", which helps security teams convert natural human language into detailed Kusto language queries (KQL). So, now, threat hunting becomes more accessible to junior security analysts that can execute advanced searches without having to know the actual syntax of queries.

SOC teams are reshaping and becoming more sophisticated as a result of using Generative AI. As Patel (2025) examines through his research, with the use of cloud computing and it's vast power, Generative AI can now solve sophisticated security issues as demonstrated in industry giants, such as Netflix and JPMorgan Chase, who have used AI to minimize risk factors while increasing operational efficiency.

Many Financial Services Institutions went through the same cloud transformation journey, as well, establishing operational transformation. Their SOC teams can now have AI-powered cybersecurity solutions that enable quicker reaction times and less noise in alerts through automation, intelligent alert handling, threat analysis, and predictive analytics. Bijoy (2025) states, that Financial Services have gained modern "*defense mechanisms through real-time threat detection, behavioral anomaly analysis, and predictive security analytics that identify potential cyber-attacks before they compromise system integrity or customer data*".

2.6 Pillar B2: Technical Vulnerabilities in AI (Security for AI)

Pillar B2 evaluates the technical vulnerabilities and attack methods used against AI models, that need to be addressed: Adversarial taxonomies, DoS, model poisoning and prompt injection. AI can be manipulated, just like everything else.

2.6.1 Adversarial Taxonomies, Attacks and Denial of Service

Adversarial Taxonomies are structured classification methods, used to categorize and understand attacks against Artificial Intelligence (AI) and Machine Learning (ML) systems.

National Institute of Standards and Technology (NIST) (2025) has established a technical standard documentation for categorizing various attack methods. The attacks are classified

according to the stage of AI lifecycle, such as inference (evasion) and training (poisoning). As AI becomes an integral part of applications and infrastructure in financial firms, there is a need to document all possible attack methods and establish a guide for mitigation.

Ibitoye et al. (2025) in their survey, refer to this as an "Arms Race" where they are describing possible methods, such as projected Gradient Descent (PGD) in adversarial attacks. In these attacks, the model's learning gradients are used by the attackers to produce undetectable perturbations that cause misclassification. That way the machine learning model, which was attacked, produces a specific wrong prediction, affecting it's integrity.

Another different kind of attack called "Sponge Examples" is documented from Shumailov et al. (2021), where the attacker focuses on availability, rather than integrity, creating a Denial of Service (DoS) attack in the model, In this case the attacker is creating inputs in order to maximize energy consumption and increase latency with ultimate goal to affect it's availability and gain precious time to complete his invasion plan.

2.6.2 Financial Model Poisoning

An interesting technical demonstration of altering financial status to deceive AI systems, especially the ones used in banks and financial firms, is presented by Raff et al. (2025). In this case, a malicious actor such as a fraudulent company sends Maximum Violated Multi-Objective (MVMO) attacks that aim to modify their financial information or reporting data, in their favor, and gain a "clean" or "safe" profile to achieve, for example, getting a big loan. The results are interesting, "*in ≈ 50% of cases, a company could inflate their earnings by 100-200%, while simultaneously reducing their fraud scores by 15%*".

Furthermore, Xia et al. (2025) showcases that an attacker can manipulate a local LLM's parameters that are used for example in inter-bank cooperation, manipulating processes in risk management, payments, financing etc. affecting stability of the global system. Their observations show that poisoning attacks are becoming more effective, stealthy, and powerful, commenting that there are many issues to be addressed by security.

2.6.3 Prompt Injection & Privacy

Open Worldwide Application Security Project (OWASP) foundation published a top 10 hazard list for generative AI in 2025. In this whitepaper by OWASP (2024), various

malicious methods were being identified, such as prompt injection and data model poisoning, which an attacker can use to inject dangerous instructions within the data that the AI model is processing. For example, this could be a fake loan application document in favor of the attacker, leading the AI to disregard any security measures.

Another example of prompt injection is given by researcher Greshake et al. (2023) in which an attacker can embed harmful content in a website, which is then parsed by an AI model, leading to corruption of the application that integrates the LLM model and in turn provides false information.

In terms of privacy, Shokri et al. (2017) demonstrated a technique in which LLM outputs can theoretically leak information about their training inputs, by using membership inference attack methods. They state that models created using machine-learning-as-a service platforms can leak a lot of information.

As a result to privacy concerns, Stack Overflow's research shows mistrust by developers in using AI for coding, due to technical risks. But the research shows another factor and that is poor processes in reviewing code thoroughly before publishing. That is a major concern, especially when using AI to provide code for a large percent of your applications. Hamberg (2025), points out that "*Stack Overflow research shows that 66.2% of developers don't trust AI code" and continues that "The technical risks are visible in your code, but process risks hide in your team's daily habits*".

2.7 Pillar B3: Technical Governance & Controls

Pillar B3 evaluates the technical processes that need to be enforced, to close the gap between static defend policies and rapid AI adoption rate. Those include the AI Bill of Materials, IT Operations and frameworks establishment.

2.7.1 The AI Bill of Materials (AIBOM)

One of the processes that need to be secured is the supply chain of AI models. For software applications there is already a Software Bill Of Materials (SBOM) in place for large organizations, so they would benefit from applying an AI Bill Of Materials for all AI models, AI agents, or MCP servers, that they develop in-house. Vandendriessche et al. (2026) state that having an AIBOM means that organizations should have a repository that

lists all components of an AI model, such as the architecture, the model's lineage, the source of training data, the hardware that supports it, the library dependencies etc. While AIBOM offers transparency, it's data integrity remains a challenge. Minchae et al (2026), propose a blockchain-based governance framework for AIBOM to guarantee integrity and automated regulatory compliance.

2.7.2 Machine Learning Security Operations (MLSecOps)

MLSecOPs extends MLOps and DevSecOps principles to address specific AI or ML vulnerabilities like data poisoning, adversarial attacks and model theft, by integrating security practices throughout the entire ML lifecycle. Nathanson et al. (2025) proposes to apply the Assurance for Artificial Intelligence (AIRS) framework to the entire process in order to ensure that the model hasn't been altered before moving to production. The key ingredient of this process is that an automated procedure automatically checks the pipeline, instead of a human developer, based on the standards that the framework proposes. This framework integrates in ML Sec Operations and by aligning with MITRE ATLAS threat categories can help reveal code contamination, tampering, or unsafe packaging.

This derives from the certainty that now-a-days the pipeline itself is identified as the new attack surface. Threat actors are more prone to attack and compromise MLOps platforms, to extract sensitive data, or poison training data, than attacking the model itself. Hawkins and Thompson (2024) from IBM, share this view and comment that when there is an absence of strict access controls and regulations, the pipeline turns into a poisoning route.

2.7.3 Auditing and Framework Implementation

AI governance has become a topic of concern globally and governments, companies and institutions are attempting to develop rules and principles in order to establish a framework for ethical and responsible use of AI.

But to be able to establish a framework, one must know how to address the Black Box problem, where inputs and outputs are well defined, but the process incorporated until reaching the results is obscure by humans. With that in mind, there is a need for AI systems to be able to provide explanations for their decisions or predictions, so that everybody understands the model's influences and conclude if an AI model succeeds or fails to respond.

To address these challenges, and provide explanations for how models reach decisions, Mersha et al (2025) conclude that the use of Explainable Artificial Intelligence (XAI) is critical in ensuring transparency, accountability and fairness.

In a more technical aspect of controlling AI, companies can use the standard (ISO/IEC 42001:2023) for embedding AI practices into AI management systems, which covers standards for explainability, fairness, bias prevention and human oversight. Standards are designed to give companies the confidence they need to implement technologies in a reliable way. Mateo-Casali et al (2026), provide a reference architecture for the design and implementation of AI systems to conform with the standard. In their article, they point out that integrating AI technologies poses significant challenges. Companies must first address barriers such as data quality, legacy systems, organizational resistance to change and compliance with regulatory standards in order to ensure successful implementation.

According to PwC's 2025 report, the biggest difficulty for companies is to translate principles for responsible AI, into scaled and operational processes. On the other hand, AI agents are expected to be used everywhere within the next year, suggesting an endless pursuit, where "*87% of leaders expect AI agents to reshape governance within the next year*". Radanliev et al. (2024) stress the significance of a complete approach that integrates technological innovation with ethical and regulatory methods to harness the power of AI.

Meanwhile, MITRE corporation (2024) has already published it's MITRE ATLAS (Adversarial Threat Landscape for Artificial-Intelligence Systems) which is a large knowledge database of adversary attacks against AI systems, helping auditors categorize, catalogue and assess AI threats.

2.8 Conclusion: Synthesis and Open Research Questions

The literature review addresses the main objective of this research, that was mentioned in the final proposal: To assess how financial institutions use AI to strengthen security and to evaluate the security threats and vulnerabilities of the AI models themselves. It is divided into two parts where part A examines the opportunities, systemic risks and regulatory expectations under a strategic background and part B examines the advantages, vulnerabilities and engineering controls under a technological background.

Evidence shows that financial institutions leverage AI as a powerful defensive tool, to combat fraud detection, anti-money-laundering, identity verification, false-positives, system vulnerabilities and cybersecurity threats. On the other hand, it shows that financial institutions recognize AI as new attack surface and need to protect their models, agents and pipelines from adversarial inputs, model theft, evasion and poisoning.

Together, these parts address the thesis title "AI for Security and Security for AI".

In "AI for Security" pillars A1 and B1 show that AI delivers measurable operational gains when it comes to fraud detection, shorter times to detect and respond for SOC teams, behavioral and predictive analytics for advancing threats and generative AI that lowers the expertise barrier. Furthermore, adoption is reported by gains of "*25% in detection accuracy and 30% in operational efficiency*" (Joshi, 2025, as cited in Poonum et al., 2025). However, value does not materialize quickly, as "*83% of organizations intend to adopt AI against fraud but execution lags"* (ACFE, 2024) and "*85% have raised AI investment while 67% expect a 2 to 4 year payback period well beyond typical technological ROI horizons*" (Deloitte, 2025). On the other hand, companies that treat AI as an organizational capability rather than a software purchase (Mikalef and Gupta, 2021) see a competitive advantage against others.

In "Security for AI" pillars A2 and B2 show that attackers nowadays have moved from compromising infrastructure to compromising logic with techniques such as adversarial inputs, model evasion and data poisoning (Global Risk Institute, 2024; Ibitoye et al., 2025). Also, with techniques like financial-model poisoning "*in ≈ 50% of cases, a company could inflate their earnings by 100-200%, while simultaneously reducing their fraud scores by 15%*" (Raff et al., 2025). Furthermore, biased models impose systemic risks of opacity which lead to "herd behavior" and "market manipulation" when organizations rely on them (Bank of England, 2023); Lumenova AI, 2025). Finally, the risk of Shadow AI already connected to "*20% of studied breaches has an average cost of $670,000 per incident*" (IBM, 2025; Gartner, 2025). All the above create regulatory exposure under the EU AI Act.

In governance, pillars A3 and B3 show that there are many initiatives like the EU AI Act, the OECD AI Principles, the AIRS and COBIT frameworks, and the AI standard ISO/IEC 42001:2023 that suggest proper AI governance models. The reality is that governance frameworks are most likely fragmented and Model Risk Management (MRM) committees

might not have the technical know-how to assess the risks of AI (McKinsey, 2025) and internal auditors lack confidence (Gartner, 2024). Furthermore, "*87% of leaders identify the "translation of responsible-AI principles into scaled operational processes" as an ongoing challenge*" (PwC, 2025) and while there several tools to help there is limited evidence on how they are deployed within financial institutions.

This dissertation will evaluate how financial institutions perceive the strategic value from deploying AI for security and if there is a corresponding maturity in securing AI. Additionally, what is the operational reality inside those institutions concerning the already in place governance frameworks. Governance maturity will be evaluated regarding deployment scale, cross-team alignment between Data Science and Cybersecurity teams, integration of AI risk into traditional Enterprise Risk Management and external regulatory pressure under instruments such as the EU AI Act. Furthermore, it will discover what the disaggregated implementation barriers are as well as the perceived business impact from AI threats. Finally, it will find out on what level their future preparedness is, concerning new and upcoming threats, with the most recent example of Anthropic's (2026) Claude Mythos that has shifted during the writing of this dissertation (Carlini et al., 2026).

3. Research Methodology

3.1 Research Design and Methodological Rationale

To capture the current state of financial institutions, on the topics the literature review elaborated, a mixed methods approach was applied on the research design. The main data collection tool is a custom semi structured questionnaire (detailed in Appendix: "Primary Data Collection Method) that was designed to be used as a targeted survey tool, as well as a mini-interview tool in some cases, where the respondent's time allowed. This questionnaire aims to obtain quantitative data (5-point Likert Scales and Percentages) as well as qualitative data that support some of the quantitative ones. The questionnaire contains 12 core questions, 5 optional qualitative questions and 3 budget questions, organized into 4 sections:

- Section 1: Introduction, Control Variables & Demographics
- Section 2: Pillar 1 - "AI for Security" (Adoption & Strategic Benefits)
- Section 3: Pillar 2 - "Security for AI" (Strategic Risks & Vulnerability)
- Section 4: Strategic Alignment & Resource Allocation

3.2 Data Collection and Sources

This research relies on primary data collected from senior experts in the Financial Sector. They are divided mainly into two large categories: (a) Technology Vendors that have the expertise and help institutions on designing and implementing AI technologies in large transformation projects and (b) Senior roles in Banking, Fintech and Financial Services. The sample contains 47 respondents with high-profile roles like Chief Technology Officers, AI Directors, Security Directors etc. and low-profile roles like AI Architects, Security Architects, etc., which are still relevant due to the visibility of large projects in AI. The sample is small due to the highly expert roles and very specific industry in Greece.

Data were collected anonymously (due to GDPR), via Google Forms or during an one-to-one video call through Microsoft Teams using the same questionnaire, with a duration of about 15 minutes. The core data of this questionnaire are quantitative and consist of

categorical demographics, operational metrics in the form of percentages and 5-point Likert Scales. These are complimented with qualitative data in most cases, to help better understand the answers.

Data were gathered during a defined 3-week time interval between March 30th, 2026 and April 19th, 2026. The selected time interval was scheduled in the vast majority beforehand, for the interviews, whereas for the survey it was quite adequate given the small number of respondents. Also, this time window reflects the current situation in this industry where large AI transformation projects are in progress and at the same time EU AI Act has been already formalized, and Anthropic's Claude Mythos has appeared in the technological and financial landscape.

The questionnaire is based on theoretical structures derived from literature review and the empirical data analyzed are original to this dissertation, nor have been used in previous research. All data are used exclusively for statistical processing and analysis as part of this dissertation study.

3.2.1 Data Weaknesses and Limitations

The statistical power of inferential tests is constrained due to the small sample size ($n = 47$), which is logical due to the high level of expertise in this specific domain. Also, findings are limited to a small financial services population since the majority of organizations are based in Greece, with the exception of some Technology Vendors that are global, but their answers are restricted to the Greek landscape. These factors mentioned limit findings from showcasing a wider European or global landscape in financial services. Findings would benefit from a larger sample.

Optional budget questions were limited to respondents in the financial services, since respondents from Technology Vendors cannot answer or estimate how budget is allocated inside financial institutions.

Also, data that relies on self-reporting by executives may introduce potential social desirability bias, although their answers are anonymous. Especially, high-profile roles may overstate AI deployment maturity or governance maturity. Another thing to consider is that AI technology is evolving rapidly, so this data may represent a "snapshot in time".

Finally, some composite values did not meet the 0,70 reliability threshold on Cronbach's Alpha test, so they are treated as separate values and analyzed individually. Specifically, the "Disaggregated Implementation Barriers" reported a threshold of $\alpha = 0{,}627$ and the "Alignment & Future Actions" reported a threshold of $\alpha = 0{,}658$. The remaining composite values' thresholds were acceptable and can be used for the statistical hypothesis tests (Strategic Value & Competitive Advantage $\alpha = 0{,}746$, Disaggregated AI Risk Exposure $\alpha = 0{,}795$, Governance, Opacity, and Regulation $\alpha = 0{,}713$).

3.3 Data Manipulation and Software Tools

The main software used for data manipulation and statistical analysis is Microsoft Excel, with the Analysis Toolpak add-in installed. It was selected, due to the small sample size, and since the results came directly from Google Forms and were stored in a Google Sheets file, it was easy to keep the raw responses and proceed with manipulation in a single file.

Data manipulation was separated into two workflows: quantitative and qualitative. Workflows for quantitative data include data capture, cleaning, encoding, descriptive statistics and inferential testing. Workflows for qualitative data don't benefit from a specific analysis software since the sample size is small and the answers were summarized and included in support of the quantitative ones in descriptive statistics.

The final processing includes:

- Descriptive statistics with (i) bar charts and pie charts for the demographic and maturity variables and (ii) tables with descriptive data such as means, medians, standard deviations and skewness for each Likert Scale.
- Cronbach's Alpha test for the five composite values of "Strategic Value & Competitive Advantage", "Disaggregated Implementation Barriers", "Disaggregated AI Risk Exposure" "Governance, Opacity, and Regulation" and "Alignment & Future Actions", with a threshold of $\alpha \geq 0.7$.
- Hypothesis tests on the relationships between AI maturity, perceived ROI, Governance maturity, Perceived AI-attack impact, etc.
- Comparative tests between groups (Technology Vendors vs Financial Services respondents, Technical vs Security/Risk/Compliance roles, and High-profile vs Low-profile roles).

3.3.1 Methodological Justification for the Selected Tests

For the hypothesis tests 1-9 on the relationships between variables, Spearman's rank-order correlation was selected for the statistical analysis. This is a non-parametric statistic test that measures the strength and direction of a monotonic relationship between two variables. Hypotheses 1-9 were stated as monotonic rather than strictly linear relationships, meaning that variables move together but not necessarily at a constant rate. The variables of the data set are ordinal, originating from five-point Likert scale rankings. Spearman's rank-order correlation looks at the ranks of the data points rather than their raw values (unlike the standard Pearson Correlation test). Also, being a non-parametric test, it does not require a normal distribution and works with skewed distributions from Likert scales, making Spearman's rank-order correlation appropriate to use.

For the comparative tests 10-12 between groups, Welch's independent-samples t-test was selected for the statistical analysis. This is a statistical test that compares the means of two independent groups that have unequal sizes and does not assume them to have equal variances (unlike the classic Student's t-test). The selected test estimates each group's variance separately and adjusts the degrees of freedom based on the variance, making it Type I error robust. Also, since all variables in tests 10-12 can be treated as continuous numeric measures whose values can be meaningfully averaged, Welch's independent-samples t-test is appropriate to use.

4. Data Review and Results

4.1 Descriptive Statistics Analysis

This section presents the descriptive statistics analysis for the categorical and ordinal values captured. Relevant questions from the survey questionnaire will be embedded under titles for better visibility. Likert Scales have been converted to ordinal values (1-5). The sample size is $n = 47$.

4.1.1 Demographics

Respondent Roles

There are 27 unique roles among the 47 total responses. The most frequent role is IT Architect with 7 entries (14,9%), followed by Technology Officer with 6 entries (12,8%), Security Architect with 5 entries (10,6%), Compliance Officer with 4 entries (8,5%) and Security Specialist along with AI Director with 2 entries each (4,3%). The remaining 21 roles are distinct with 1 entry each (2,1%). Results showcase a long-tail distribution of unique roles.

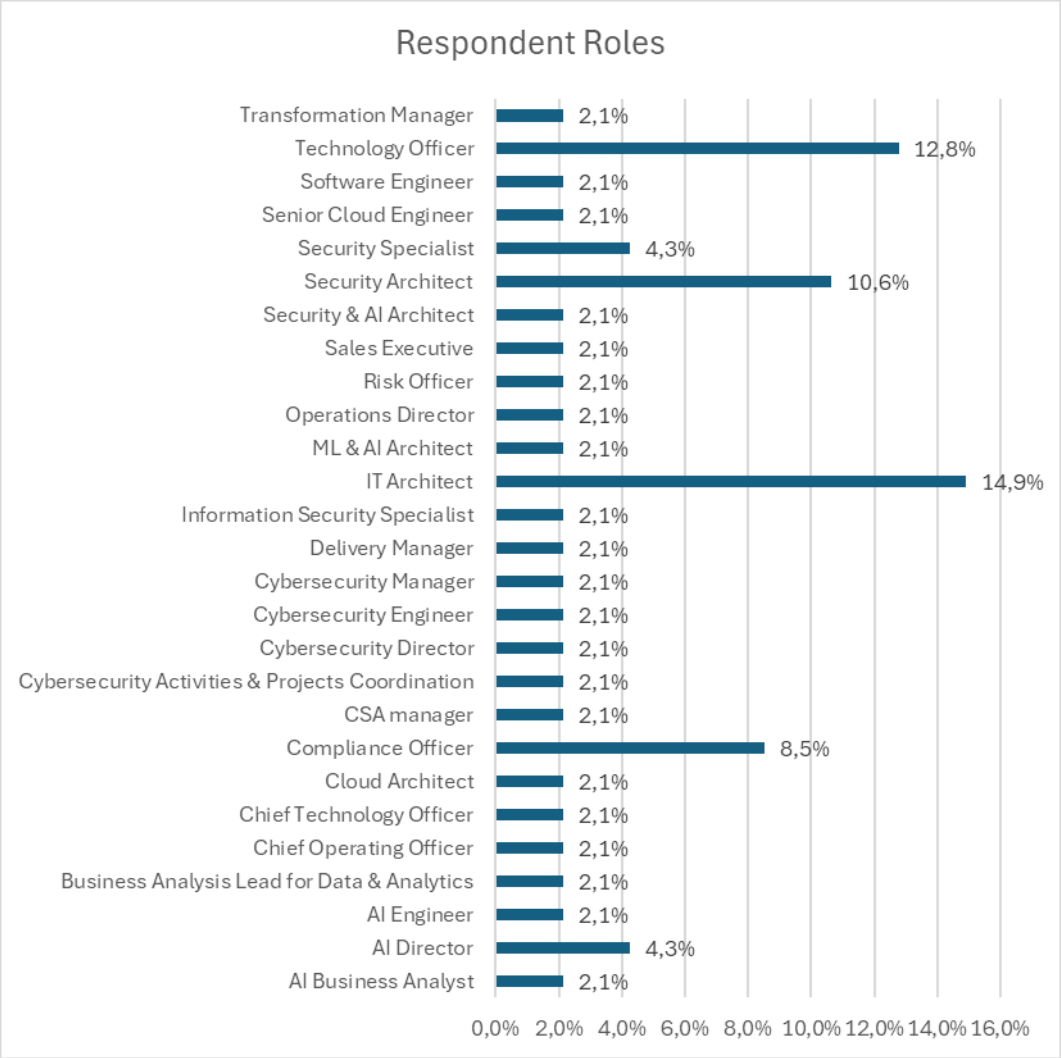


Figure 1 - Respondent Role Distribution

Role Group Analysis

To become easier to interpret, roles are consolidated in two groups, (i) based on their role profile (High profile vs Low profile) and (ii) based on their functional profile (Security/Compliance/Risk vs Technology).

(i) Role Profile Groups

High profile roles are considered the ones that are in a high-level management and can take decisions and responsibilities (AI Director, Chief Operating Officer, Chief Technology Officer, Cybersecurity Director, Operations Director, Technology Officer, Cybersecurity

Manager, CSA manager, Delivery Manager, Transformation Manager, Business Analysis Lead for Data & Analytics). They account for 17 entries (36,2%).

Low profile roles are considered the ones that are highly expertised but are in a low to mid-level management (AI Business Analyst, AI Engineer, Cloud Architect, Cybersecurity Engineer, Information Security Specialist, IT Architect, ML & AI Architect, Sales Executive, Security & AI Architect, Security Architect, Security Specialist, Senior Cloud Engineer, Software Engineer, Cybersecurity Activities & Projects Coordination, Risk Officer, Compliance Officer). They account for 30 entries (63,8%).

From the analysis, the sample is weighted towards low profile roles that account for nearly double the high profile roles (63,8%, 36,2%).

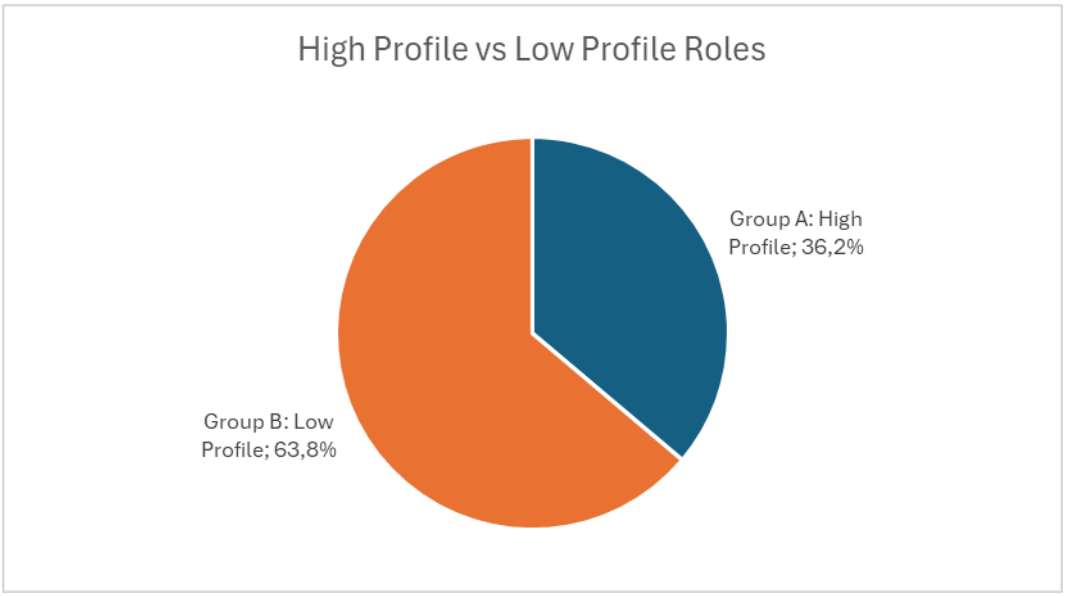


Figure 2 - High Profile vs Low Profile Roles

(ii) Functional Profile Groups

Security/Compliance/Risk roles are grouped together because they are complimentary roles and are responsible for protecting an organization (Security Architect, Security & AI Architect, Security Specialist, Cybersecurity Director, Cybersecurity Manager, Cybersecurity Engineer, Cybersecurity Activities & Projects Coordination, Compliance Officer, Risk Officer, Information Security Specialist, CSA manager). They account for 19 entries (40,4%).

Technology roles are the ones responsible for the deployment of AI (Operations Director, Senior Cloud Engineer, Sales Executive, Transformation Manager, IT Architect, Technology Officer, ML & AI Architect, AI Director, Chief Operating Officer, Chief Technology Officer, Software Engineer, AI Business Analyst, Business Analysis Lead for Data & Analytics, AI Engineer, Cloud Architect, Delivery Manager). They account for 28 entries (59,6%).

From the analysis, the sample is weighted towards technology roles (59,6%) with security/compliance/risk roles holding a substantial percentage (40,4%).

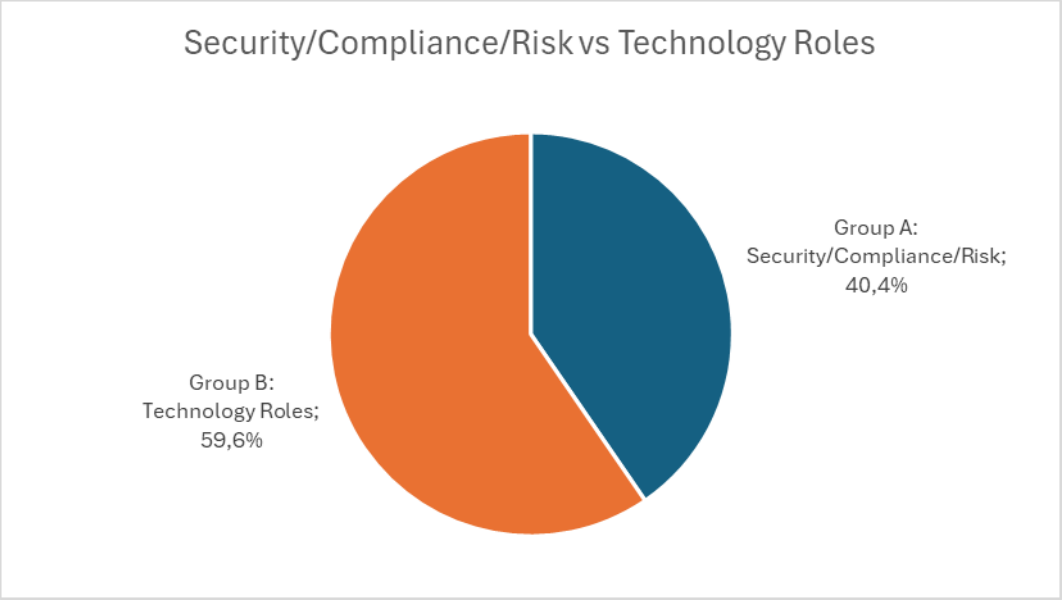
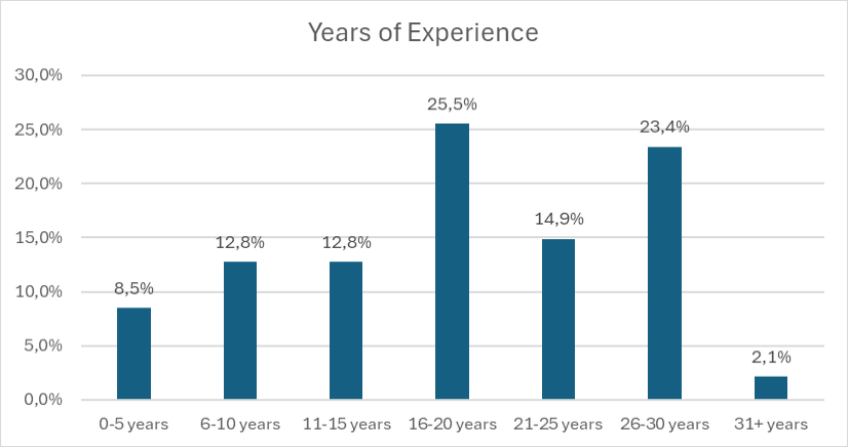


Figure 3 - Security/Compliance/Risk vs Technology Roles

Years of Experience

Years of experience were grouped into six ranges for a sample of 47 entries. The two ranges with the most entries count, are 16-20 years (12 entries, 25,5%) and 26-30 years (11 entries, 23,4%). Also, the combined ranges 11-20 (18 entries, 38,3%) and 21-30 (18 entries, 38,3%) are the most dominant. On average, the respondents have over 18 years of experience (Mean 18,51) and the most common value is 20 years, reported 6 times (Median 20, Mode 20). There are only 4 entries (8,5%) with less than 5 years of experience. The distribution is left-skewed (-0,39) meaning that there are more experienced professionals than junior ones. These results indicate a highly experienced sample.



Statistic	Value
Mean	18,51
Median	20,00
Mode	20
Standard Deviation	8,19
Variance	67,08
Minimum	1
Maximum	32
Range	31
Count	47
Skewness	-0,39

Figure 4 - Years of Experience Distribution

Company/Institution Type

There are 7 unique company types among the 47 total responses in the FSS. The most common company type is Technology Vendor with 20 entries (42,6%). Second most common type is Banking Services with 19 respondents (40,4%). The remaining 8 entries are distributed across FinTech with 3 entries (6,4%), Financial Services with 2 entries (4,3%) and Insurance Services, Payments Institutions and Consulting Services with 1 entry each (2,1% each).

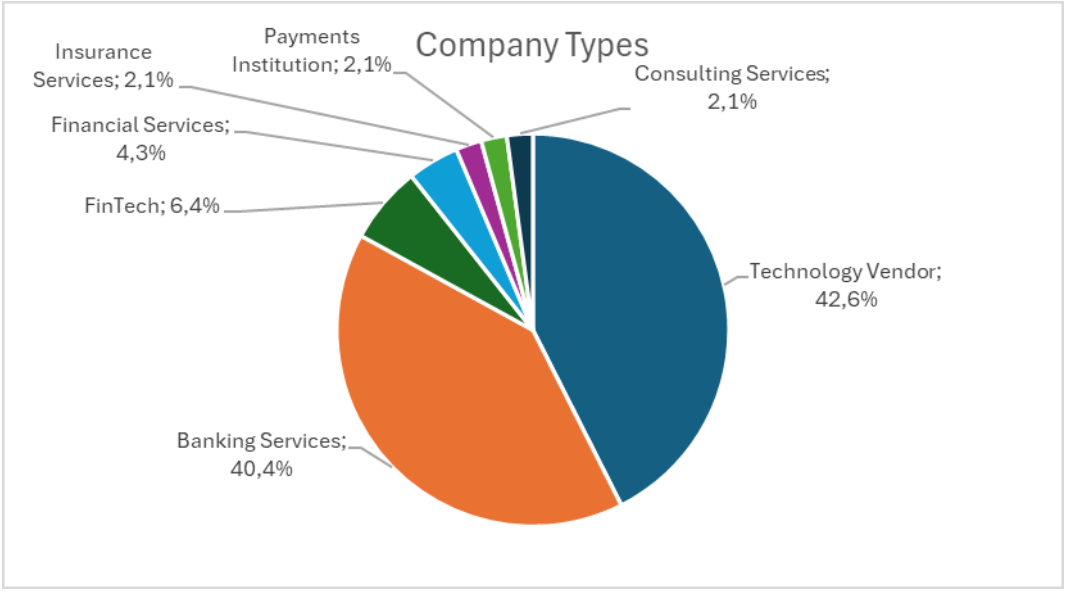


Figure 5 - Company Types

Functional Company Groups

To become easier to interpret, company types are consolidated in two groups based on their functional type. Financial Services Institutions (Banking Services, Fintech, Financial Services, Insurance Services, Payment Institution, Consulting Services) account for a total of 27 entries (57,4%). Sample is fairly balanced distributed between Technology Vendors (42,6%) and Financial Services Institutions (57,4%). This allows further processing for comparative testing.

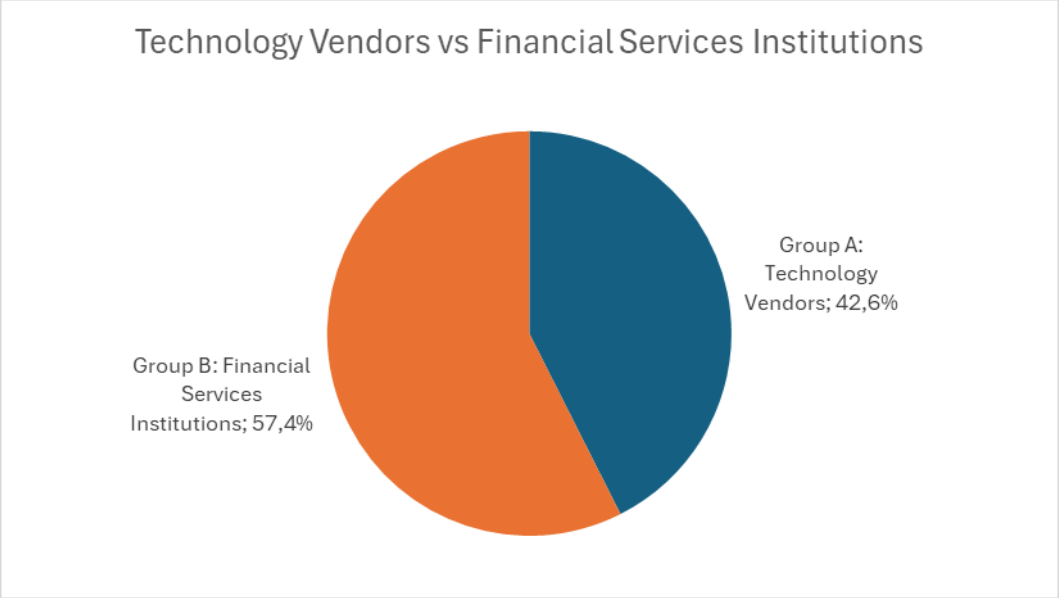


Figure 6 - Technology Vendors vs Financial Services Institutions

Primary Region of Operation

From a total of 47 responses, 3 regions of operation were reported. The sample is strongly weighted towards Greece, with 35 entries (74,5%). The remaining entries mean that companies may primarily operate in Europe (6 entries, 12,8%) or globally (6 entries, 12,8%), but they also have an autonomous branch in Greece.

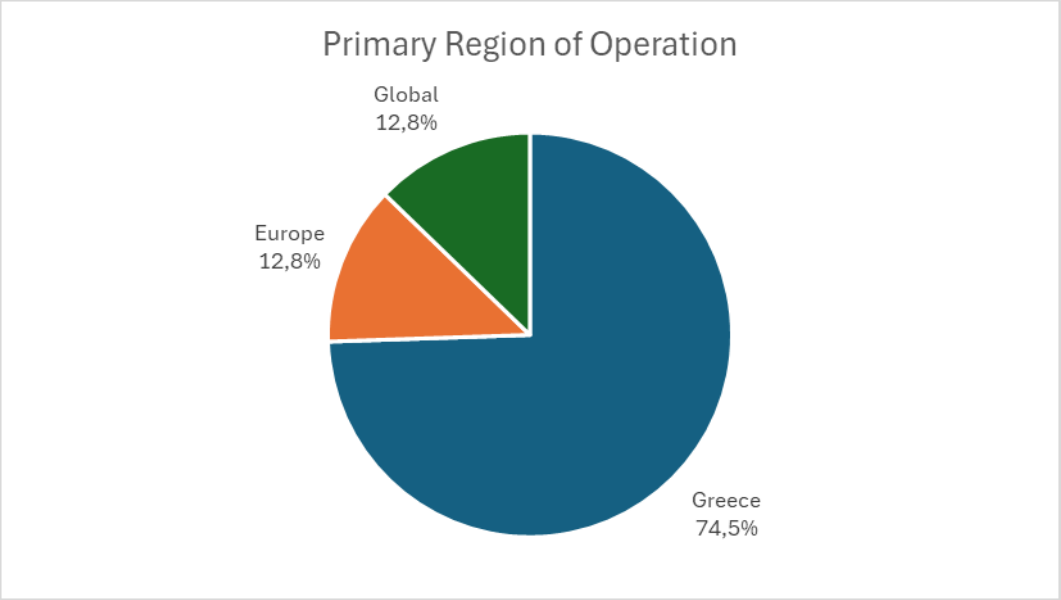


Figure 7 - Primary Region of Operation

4.1.2 AI Deployment Maturity Stage

"How would you classify your organization's current AI deployment?"

There are four AI Deployment Maturity stages in the questionnaire. The most common stage is Limited Production with 20 entries (42,6%). Next are Enterprise-wide and Strategic Transformation, with 12 entries each (25,5% each), that together account for over half of the sample's size 51% (24 entries). Last, is Pilot with only 3 entries (6,4%). The sample is weighted towards institutions that moved beyond pilot stage, into production, with more than half of respondents reporting that their institutions have already deployed AI at full scale production.

If scale is treated as ordinal, the distribution has mode 2, mean 2,70 and median 3, confirming that Limited production is the most common stage, institutions are on average between Limited and Enterprise-wide and half of them are above Enterprise-wide maturity stage.

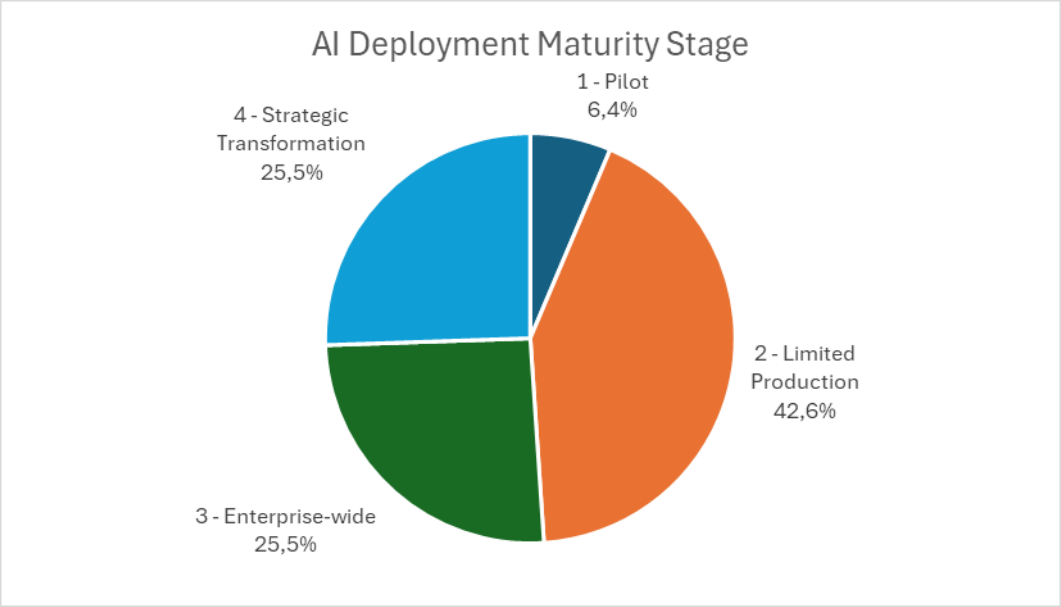
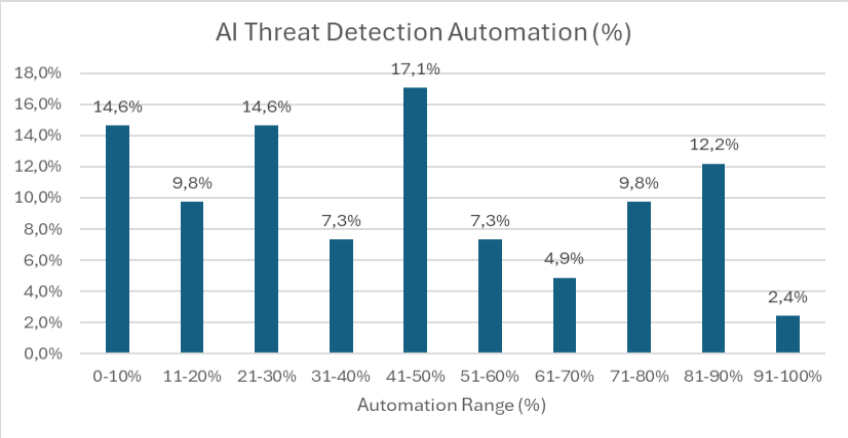


Figure 8 - AI Deployment Maturity Stage

4.1.3 AI Threat Detection Automation

"Approximately what percentage (%) of your threat detection or fraud analysis is currently automated by AI/ML models?"

On average, institutions have automated nearly half of their threat or fraud analysis (mean 47,2%). The most common percentage reported is 50% (mode 50%). Half of the respondents reported an automation level of 50% or lower, while the other half reported 50% or higher (median 50%). These results indicate that institutions are at half of their automation journey. The most common range is 41%-50% (17% of institutions). The ones that have reported less automation levels (0%-30%) are 39% of the sample and the ones that reported higher automation levels (71%-100%) are 24,4% of the sample size. Also, values range from 0% to 100% covering the whole range as the sample has high variability (SD 28,51, Variance 812,56).



Statistic	Value
Mean	47,20
Median	50,00
Mode	50
Standard Deviation	28,51
Variance	812,56
Minimum	0
Maximum	100
Range	100
Count	41
Skewness	0,13

Figure 9 - AI Threat Detection Automation (%)

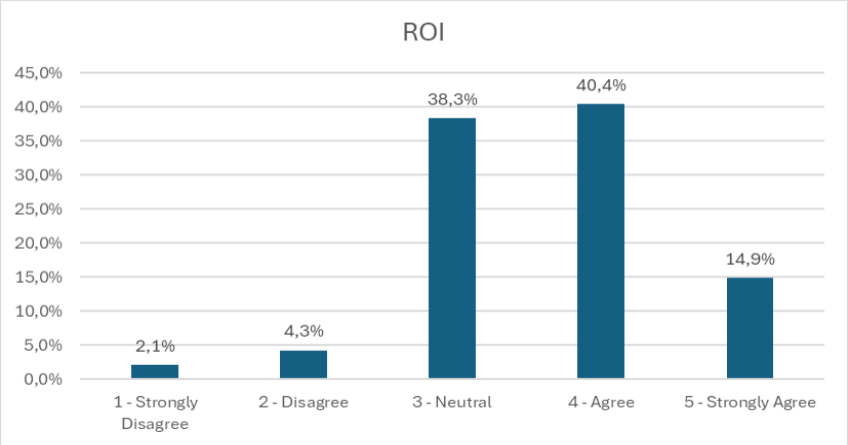
4.1.4 Strategic Value and Competitive Advantage

"Rate your agreement with the following statements."

ROI

"Our deployment of AI-powered security tools has delivered a highly measurable and mission-critical Return on Investment (ROI)."

The distribution of ROI is left-skewed (-0,38), with a mean 3,62, median 4 and mode 4. This means that responses are towards the positive side, with 55,3% of responses being Agree and Strongly Agree. Negative responses are only 6,4%. Institutions that deploy AI-powered security tools have perceived measurable results that add up to AI's strategic value. It is worth mentioning that 38,3% are neutral about the perceived value of AI deployment. This It is possible that they belong to the "Limited Production" maturity stage (42,6%). Further analysis will be conducted on the inferential statistics section of the dissertation (Hypothesis Test 1).



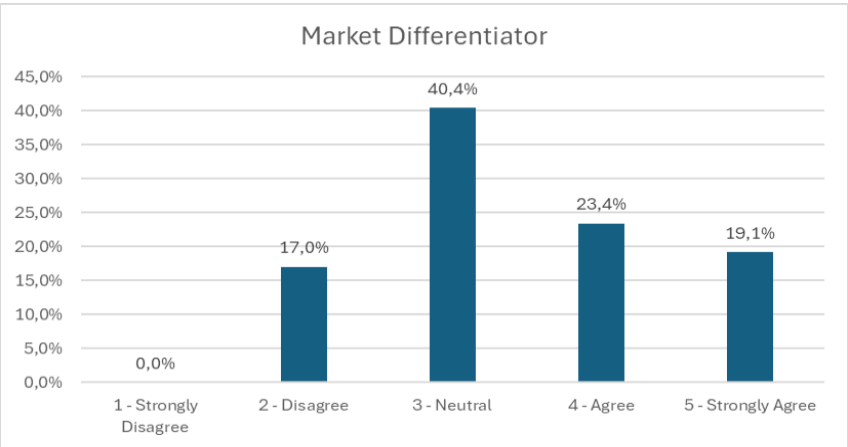
Statistic	Value
Mean	3,62
Median	4,00
Mode	4
Standard Deviation	0,87
Variance	0,76
Minimum	1
Maximum	5
Range	4
Count	47
Skewness	-0,38

Figure 10 - ROI

Market Differentiator

"Our use of AI for security serves as a significant market differentiator against our competitors."

The distribution of Market Differentiator is centered-based (skewness 0,22), with a mean 3,45, median 3 and mode 3. This means that responses are towards the neutral side, with 42,6% of responses being Agree and Strongly Agree, whereas the Neutral responses are 40,4%. Negative responses are 17%. Institutions that deploy AI-powered security tools don't always see these tools as a market differentiator.



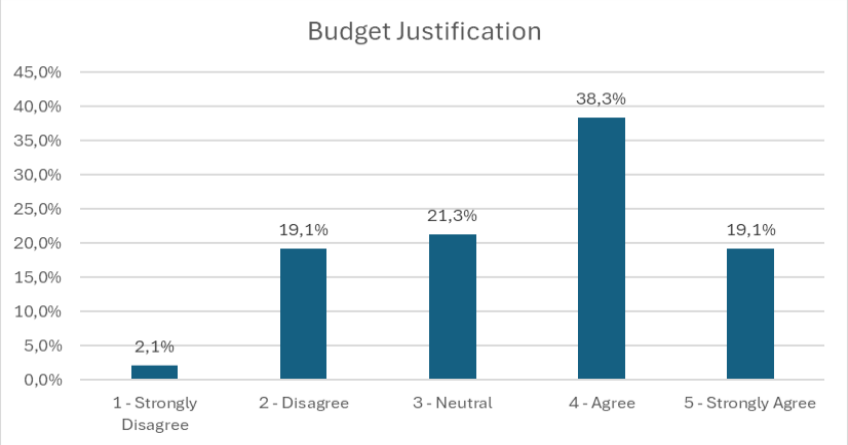
Statistic	Value
Mean	3,45
Median	3,00
Mode	3
Standard Deviation	1,00
Variance	0,99
Minimum	2
Maximum	5
Range	3
Count	47
Skewness	0,22

Figure 11 - Market Differentiator

Budget Justification

"It is currently straightforward to justify increased budget requests for AI security tools to the Board/CFO."

The distribution of Budget Justification is left-skewed (-0,36), with a mean 3,53, median 4 and mode 4. Distribution has a bigger variance than the others with 1,17. Responses are more spread, with 57,4% of responses being positive. Negative responses are 21,2% and neutral are 21,3%. Institutions that deploy AI-powered security tools are somewhat divided. The majority 57,4% can justify spending to the Board, but the 21,2% struggles to justify it, while 21,3% don't believe that is always easy to justify. The majority of 57,4% that can justify spending is also close to the majority of 55,3% that perceive measurable results from AI-powered deployment tools.



Statistic	Value
Mean	3,53
Median	4,00
Mode	4
Standard Deviation	1,08
Variance	1,17
Minimum	1
Maximum	5
Range	4
Count	47
Skewness	-0,36

Figure 12 - Budget Justification

Overall

Respondents reported that they perceive ROI better from deploying AI-powered security tools (mean 3,62). They are not very sure if their institutions have a competitive advantage against others (mean 3,45). Furthermore, budget justification to Board varies among them, while others find it easy, others struggle (mean 3,53). Budget justification and market differentiation perception can benefit from understanding ROI, through better metrics.

Furthermore, all boxplots show a narrow Interquartile Range (IQR) of 1 point, meaning that the middle 50% of respondents largely agree with each other, with most responses ranging

from neutral to agree. Generally, opinions are consistent for all categories, with ROI and Budget Justification sharing the strongest agreement, with a median of 4.

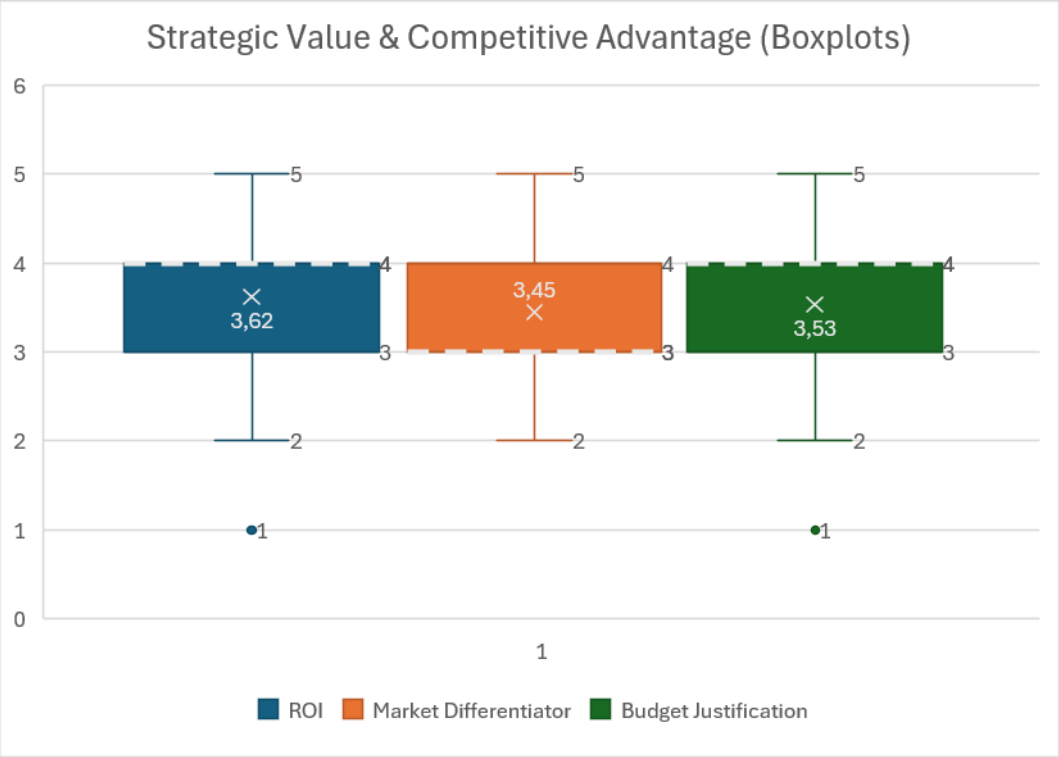


Figure 13 - Strategic Value and Competitive Advantage (Boxplots)

Qualitative Analysis

"Looking at these scores, how specifically does this AI advantage manifest? Are you reducing operational costs, or increasing customer trust through fewer false positives?"

Responses are categorized in themes, based on the frequency of certain words or phrases. Cost reduction and customer trust are the most frequent.

- Cost Reduction: Most respondents reported that AI automation leads to cost reduction. Automating repetitive tasks, processing information faster than humans and cutting down manual work are few of the key points of cost reduction.
- Increased Customer Trust: Also, many respondents reported that AI's decision accuracy leads to fewer false positives, thus increasing customer trust.
- Not there yet: A minority of respondents reported that AI deployment is still at a primitive state in their institution and use it only as a second opinion.

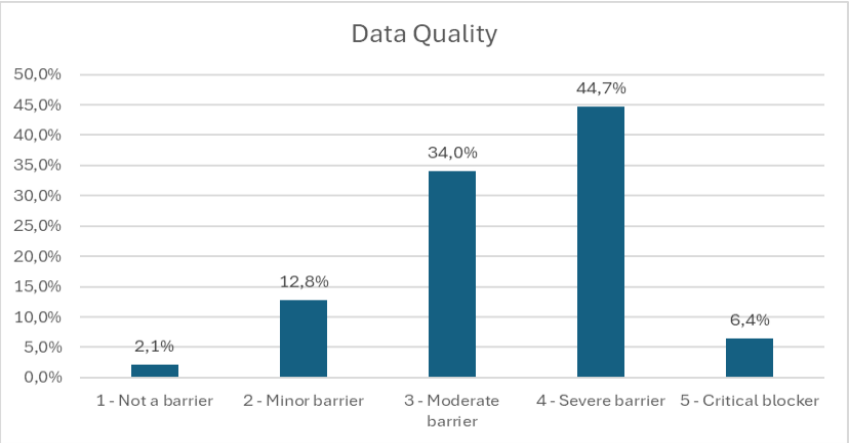
4.1.5 Disaggregated Implementation Barriers Severity Rating

"Rate the severity of the following internal barriers preventing optimal AI security leverage."

Data Quality

The distribution of Data Quality is left-skewed (-0,50), with a mean 3,4, median 4 and mode 4. Responses are towards the most severe barrier side, with 51,1% of responses being Severe Barrier and Critical Blocker. Only 14,9% reported data quality as not a barrier or just minor, while 34% reported Moderate.

Overall data quality is rated as a severe barrier (44,7%), with 6,4% rating it as critical, because an AI model's accuracy depends on data quality.



Statistic	Value
Mean	3,40
Median	4,00
Mode	4
Standard Deviation	0,88
Variance	0,77
Minimum	1
Maximum	5
Range	4
Count	47
Skewness	-0,50

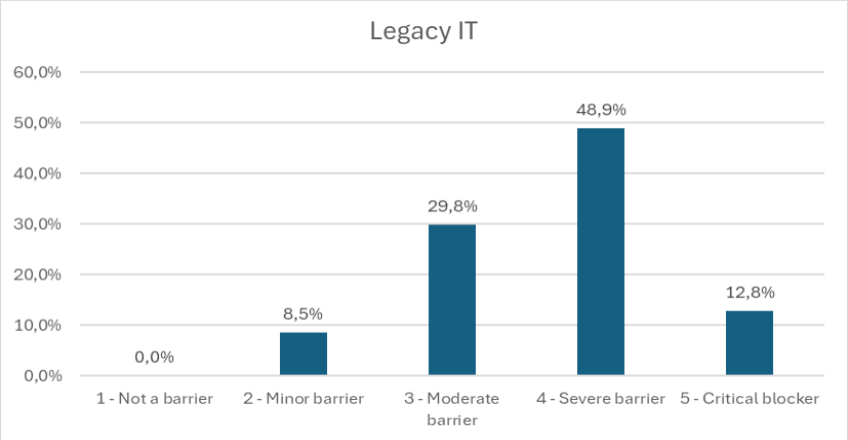
Figure 14 - Data Quality

Legacy IT

The distribution of Legacy IT is left-skewed (-0,29), with a mean 3,66, median 4 and mode 4. Responses are again towards the most severe barrier side, with 61,7% of responses being Severe Barrier and Critical Blocker. Only 8,5% reported legacy IT as a minor barrier, while 29,8% reported Moderate. Also, it is worth mentioning that no respondent stated that Legacy IT is not a barrier.

Overall legacy IT systems are rated as a severe barrier (48,9%), with 12,8% rating it as critical, because an old infrastructure cannot support modern AI tools. On the other hand,

most of the modern AI tools are being deployed in cloud infrastructures. That said, legacy IT systems like mainframes that run on monolithic architectures cannot support modern business operations that are built on microservices or modern APIs.



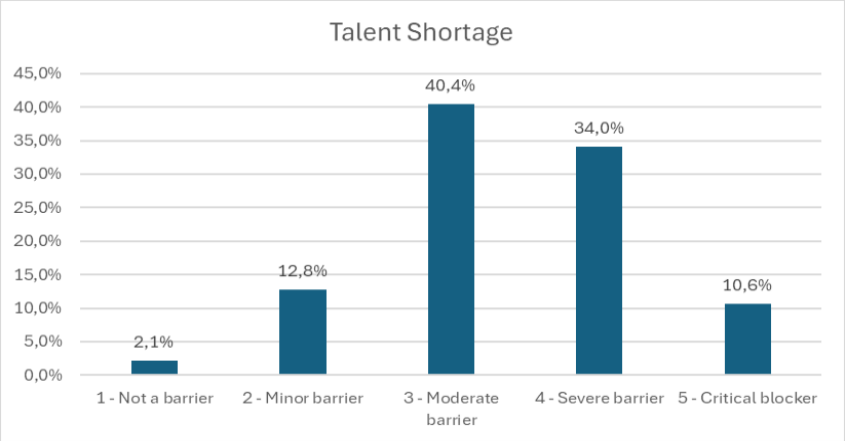
Statistic	Value
Mean	3,66
Median	4,00
Mode	4
Standard Deviation	0,81
Variance	0,66
Minimum	2
Maximum	5
Range	3
Count	47
Skewness	-0,29

Figure 15 - Legacy IT

Talent Shortage

The distribution of Talent Shortage is center-based (skewness -0,16), with a mean 3,38, median 3 and mode 3. Responses are towards the most severe barrier side, with 44,6% of responses being Severe Barrier and Critical Blocker. Some respondents (14,9%) reported talent shortage as not a barrier or just minor, while 40,4% reported Moderate.

Overall talent shortage is rated as a moderate to severe barrier (74,4%), with 10,6% rating it as critical. In general, the skills required for both AI and security are hard to find and even more the combination of AI and security expertise.



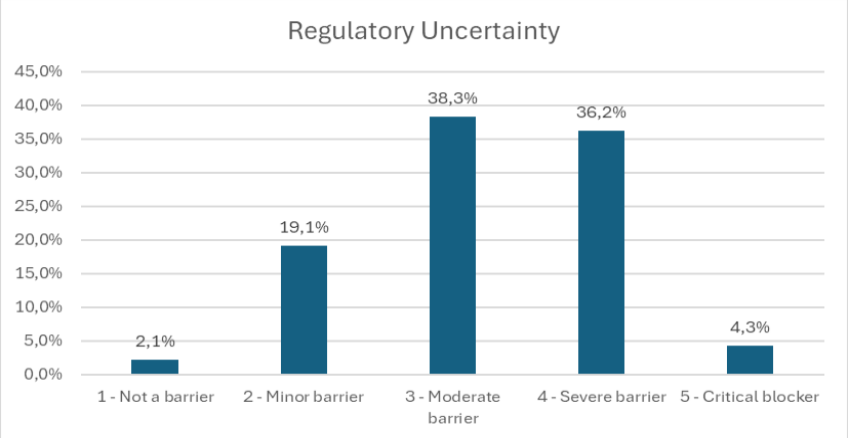
Statistic	Value
Mean	3,38
Median	3,00
Mode	3
Standard Deviation	0,92
Variance	0,85
Minimum	1
Maximum	5
Range	4
Count	47
Skewness	-0,16

Figure 16 - Talent Shortage

Regulatory Uncertainty

The distribution of Regulatory Uncertainty is center-based (skewness -0,24), with a mean 3,21, median 3 and mode 3. Responses are slightly towards the most severe barrier side, with 40,5% of responses being Severe Barrier and Critical Blocker. Some respondents (21,2%) reported regulatory uncertainty as not a barrier or just minor, while 38,3% reported Moderate.

Overall regulatory uncertainty or compliance fears are rated as a moderate to severe barrier (74,5%), with only 4.3% rating it as critical. In general, compliance fears or regulatory uncertainty under laws such as the EU AI Act are preventing organizations from scaling AI.



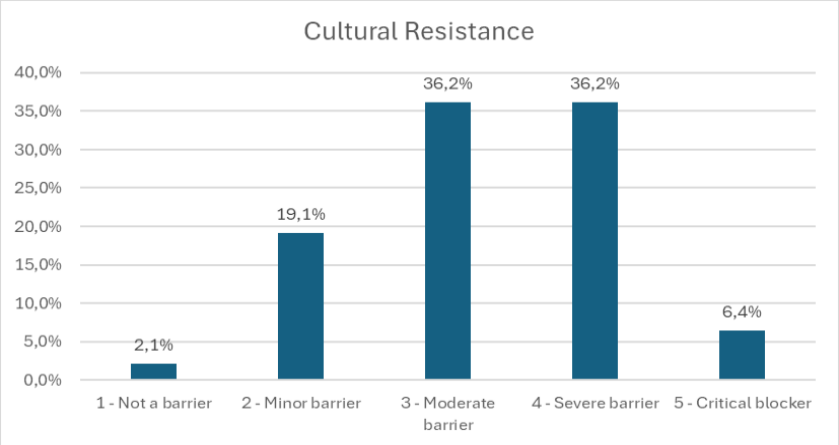
Statistic	Value
Mean	3,21
Median	3,00
Mode	3
Standard Deviation	0,88
Variance	0,78
Minimum	1
Maximum	5
Range	4
Count	47
Skewness	-0,24

Figure 17 - Regulatory Uncertainty

Cultural Resistance

The distribution of Culture Resistance is center-based (skewness -0,19), with a mean 3,26, median 3 and mode 4. Responses are slightly towards the most severe barrier side, with 42,6% of responses being Severe Barrier and Critical Blocker. Some respondents (21,2%) reported cultural resistance as not a barrier or just minor, while 36,2% reported Moderate.

Overall cultural resistance is rated as a moderate to severe barrier (72,4%), with only 6,4% rating it as critical. In general, cultural resistance is all about employees' fear of job replacement and concern about opaque algorithmic decisions.



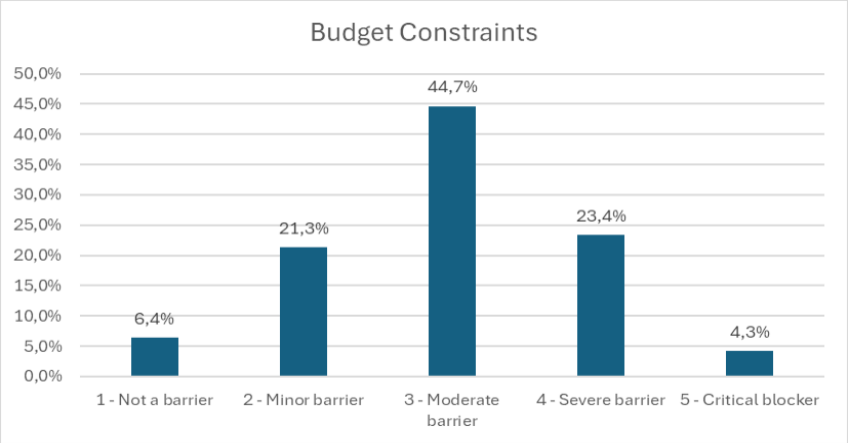
Statistic	Value
Mean	3,26
Median	3,00
Mode	4
Standard Deviation	0,92
Variance	0,85
Minimum	1
Maximum	5
Range	4
Count	47
Skewness	-0,19

Figure 18 - Cultural Resistance

Budget Constraints

The distribution of Budget Constraints is center-based (skewness -0,12), with a mean 2,98, median 3 and mode 3. Responses are slightly towards the moderate barrier side, with 44,7% reporting Moderate. Some respondents (27,7%) reported budget constraints as not a barrier or just minor, while 27,7% of responses being Severe Barrier and Critical Blocker.

Overall budget constraints are rated as a moderate barrier (44,7%), with only 6,4% rating it as critical. In general, budget doesn't seem like a big barrier, as institutions are willing to invest.



Statistic	Value
Mean	2,98
Median	3,00
Mode	3
Standard Deviation	0,94
Variance	0,89
Minimum	1
Maximum	5
Range	4
Count	47
Skewness	-0,12

Figure 19 - Budget Constraints

Overall

All Barrier metrics have means around 3, suggesting that institutions face moderate to severe challenges in implementing AI. The most critical barrier is legacy IT (mean 3,66) with data quality (mean 3,40) and talent shortage (mean 3,38) following closely. Cultural resistance (mean 3,26) and regulatory uncertainty (mean 3,21) are considered relatively moderate barriers. The lowest barrier is budget constraints (mean 2,98).

Overall, there is not a clear result. Institutions should first take into consideration the decommission of their legacy systems. Data cleaning should be a top priority and new expertised personnel in AI and Security should be acquired, though training is an option, too. Regulatory uncertainty and cultural resistance exist and require a methodological approach to overcome. Finally, budget doesn't matter as much as finding the right personnel, though it is a barrier if we combine this with the infrastructure upgrade and staff payrolls.

Furthermore, all boxplots except Budget Constraints show a narrow IQR of 1 point, meaning that the middle 50% of respondents largely agree with each other, with most responses ranging from moderate to severe barrier. Data Quality and Legacy IT share the strongest agreement, with a median of 4, indicating that these are severe barriers. On the other hand, Budget Constraints boxplot shows a wide IQR of 2 points, meaning that the middle 50% of respondents have polarized opinions, with responses ranging from minor to severe barrier. Opinions are more divided on whether budget is an implementation barrier, but more consistent for all other categories.

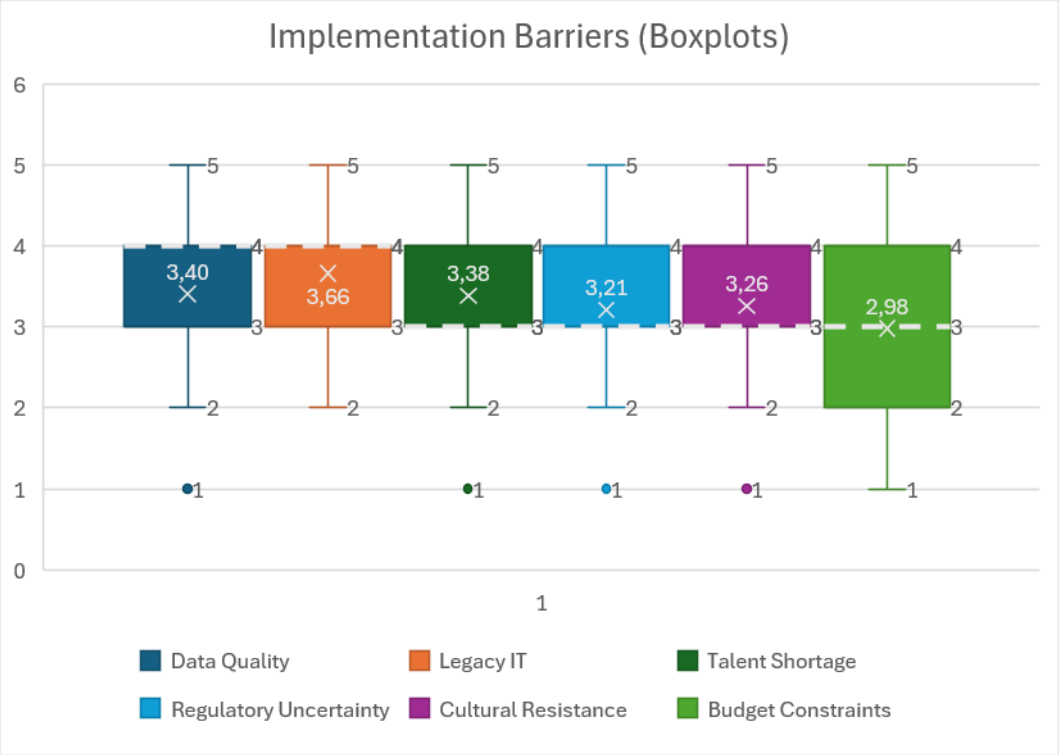


Figure 20 - Implementation Barriers (Boxplots)

Qualitative Analysis

"If you rated one of the above as your critical blocker, can you elaborate on why this specific barrier is so difficult to overcome?"

Responses are categorized in themes, based on the frequency of certain words or phrases. Responses are quite polarized, following the pattern of the mean comparison.

- Data Quality: Some respondents reported that everything in AI starts with data. Data quality indicates the level of AI's analyzing capability.
- Legacy IT: Some respondents mentioned AI's incompatibility with older mainframes. Migration to newer platform is cost intensive.
- Cultural Resistance: Some respondents stated that AI's adoption within the institutions starts from C-level executives. A top-down adoption approach is required.

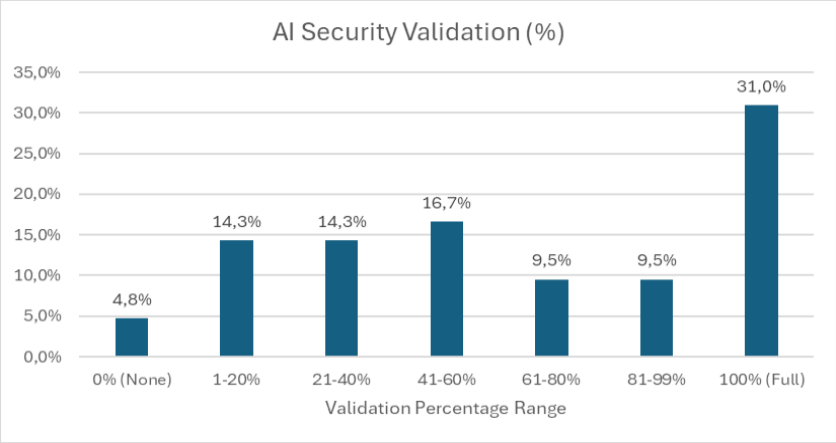
- Regulatory Uncertainty: External regulation authorities (EU AI Act, GDPR, etc.) make AI deployment difficult to progress even within institutions that both capital and talent exist.
- Talent Shortage: A minority of respondents stated that the combination of AI and security skills are rare. Some respondents propose a dedicated AI security team.
- Budget: A minority of respondents stated that with no visible returns from AI in security, budget is difficult to justify.

4.1.6 AI Security Validation

"Approximately what percentage (%) of your production AI models undergo a formal, documented security validation process (e.g., adversarial stress testing) prior to deployment?"

On average, institutions validate more than half of their AI models before deployment (mean 63,19%). Half of the respondents reported that their organization's validation process is 70% or lower, while the other half reported 70% or higher (median 70%). The most common response is 100% validation (mode 100%) which corresponds to 13 respondents (31%). 40,5% of institutions have a high validation rate (81%-100%), 26,2% have a moderate validation rate (41%-80%) and 33,3% have a low validation rate (0%-40%). Distribution is weighted towards higher validation rates (Skewness -0,38) and has a high variability as answers cover the whole range (SD 35,14, Variance 1234,60).

Overall, low validation rates indicate risk exposure and a significant governance gap. Validation maturity is high but uneven. Full validation is the most common state, yet half of organizations are still at 70% or below. There is a critical area (4,8%) that there is no validation and needs to be addressed. 33,3% that have low validation rates could benefit from best practices and governance frameworks that the 31% with full validation rate already achieved.



Statistic	Value
Mean	63,19
Median	70,00
Mode	100
Standard Deviation	35,14
Variance	1234,60
Minimum	0
Maximum	100
Range	100
Count	42
Skewness	-0,38

Figure 21 - AI Security Validation (%)

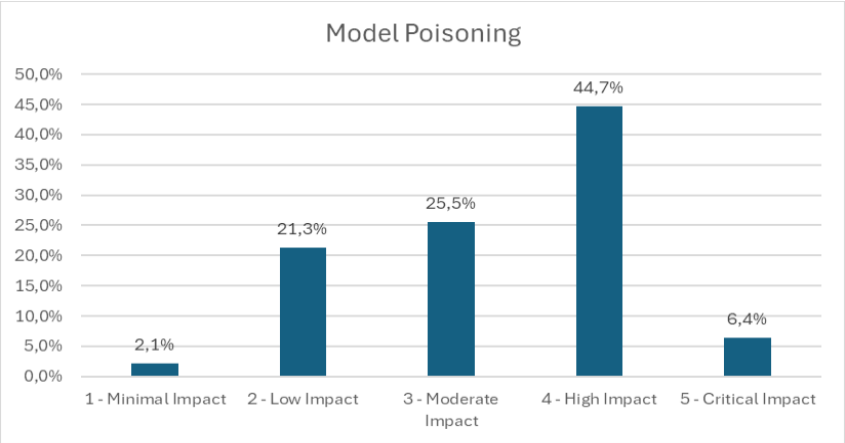
4.1.7 Disaggregated AI Risk Exposure

"Rate the perceived potential business impact of the following AI-specific attack vectors on your organization."

Model Poisoning

The distribution of Model Poisoning is left-skewed (-0,39), with a mean 3,32, median 4 and mode 4. Over half of the responses are towards the highest impact, with 51,1% being High or Critical Impact. 23,4% reported Low or Minimal Impact, while 25,5% reported Moderate.

Overall, model poisoning is rated as high impact (44,7%), with only 6,4% rating it as critical, as this technique corrupts the model's core logic, triggering false predictions or leak sensitive information.



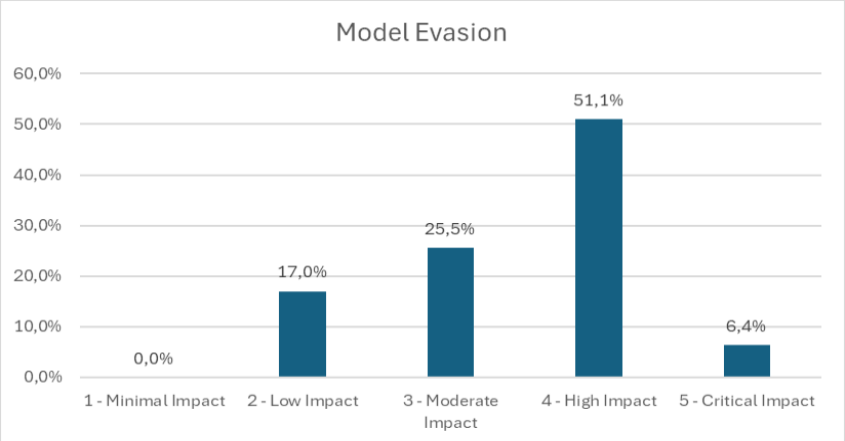
Statistic	Value
Mean	3,32
Median	4,00
Mode	4
Standard Deviation	0,96
Variance	0,92
Minimum	1
Maximum	5
Range	4
Count	47
Skewness	-0,39

Figure 22 - Model Poisoning

Model Evasion

The distribution of Model Evasion is left-skewed (-0,44), with a mean 3,47, median 4 and mode 4. Over half of the responses are towards the highest impact, with 57,5% being High or Critical Impact. 17% reported Low Impact, while 25,5% reported Moderate. None of the respondents consider model evasion as a minimal impact on risk.

Overall, model evasion is rated as high impact (51,1%), with only 6,4% rating it as critical, as this technique is part of adversarial attacks where an input is manipulated to trick models into making a false prediction.



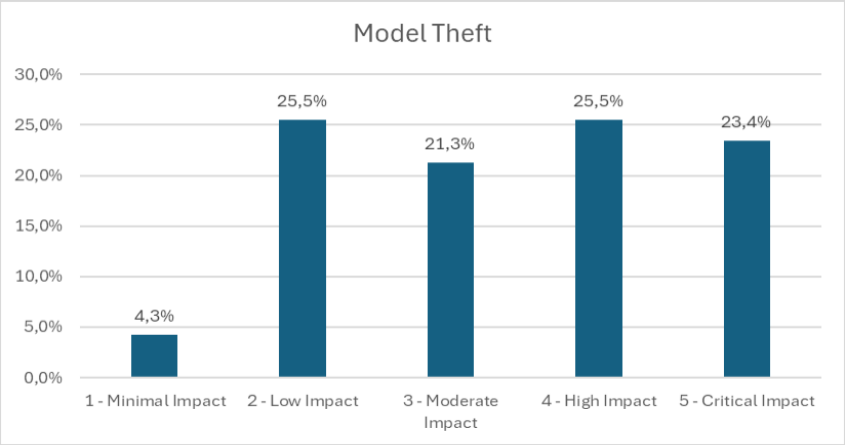
Statistic	Value
Mean	3,47
Median	4,00
Mode	4
Standard Deviation	0,86
Variance	0,73
Minimum	2
Maximum	5
Range	3
Count	47
Skewness	-0,44

Figure 23 - Model Evasion

Model Theft

The distribution of Model Theft is fairly balanced (-0,12), with a mean 3,38, median 3 and mode 4. Responses are towards the highest impact, with 48,9% being High or Critical Impact. 29,8% reported Low Impact, while 21,3% reported Moderate. Responses are polarized as they have high variability (SD 1,23, Variance 1,50) and in all impacts except Minimal, the percentage is fairly the same around 21%-26%

Overall, model theft has the highest rating in critical impact (25,5%) among all AI risk exposure techniques, as this technique enables attackers to extract a model's parameters or architecture and lead to intellectual property loss.



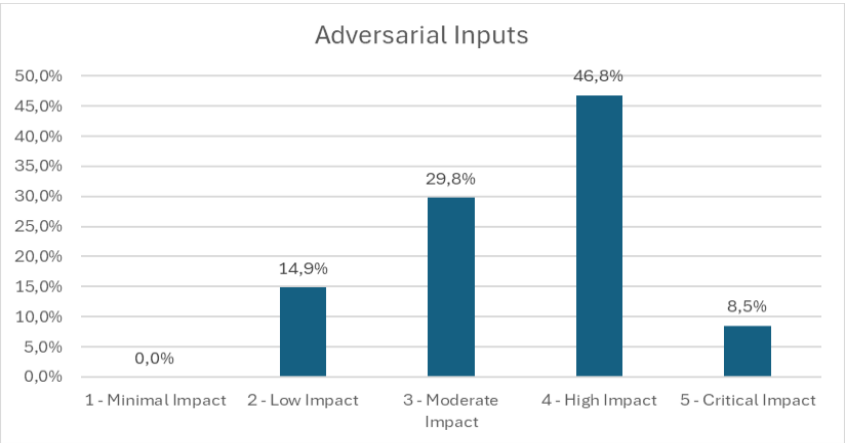
Statistic	Value
Mean	3,38
Median	3,00
Mode	4
Standard Deviation	1,23
Variance	1,50
Minimum	1
Maximum	5
Range	4
Count	47
Skewness	-0,12

Figure 24 - Model Theft

Adversarial Inputs

The distribution of Adversarial Inputs looks a lot like the Model Evasion's distribution. It is left-skewed (-0,29), with a mean 3,49, median 4 and mode 4. Over half of the responses are towards the highest impact, with 55,3% being High or Critical Impact. 14,9% reported Low Impact, while 29,8% reported Moderate. None of the respondents consider adversarial inputs as a minimal impact on risk.

Overall, adversarial inputs are rated as high impact (46,8%), with 8,5% rating it as critical, as this technique exploits vulnerabilities in how AI models process data, that leads to security holes.



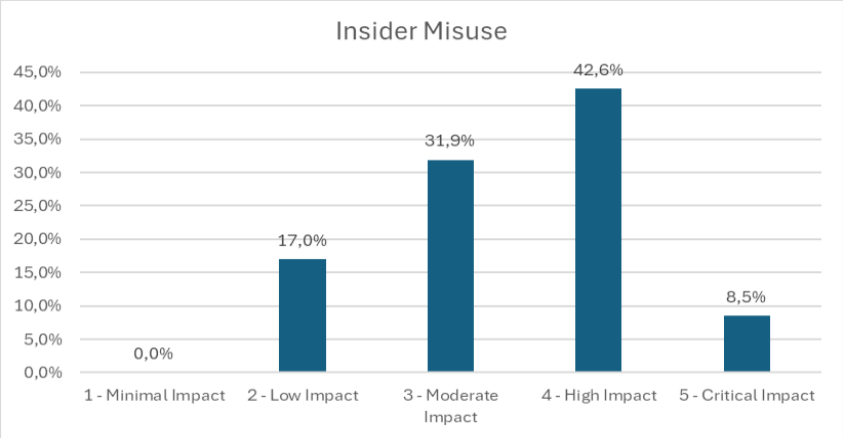
Statistic	Value
Mean	3,49
Median	4,00
Mode	4
Standard Deviation	0,86
Variance	0,73
Minimum	2
Maximum	5
Range	3
Count	47
Skewness	-0,29

Figure 25 - Adversarial Inputs

Insider Misuse

Again, the distribution of Insider Misuse looks a lot like the previous distribution. It is left-skewed (-0,17), with a mean 3,43, median 4 and mode 4. Over half of the responses are towards the highest impact, with 51,1% being High or Critical Impact. 17% reported Low Impact, while 31,9% reported Moderate. None of the respondents consider insider misuse as a minimal impact on risk.

Overall, insider misuse is rated as high impact (42,6%), with 8,5% rating it as critical, as this often refers to autonomous AI agents or employees that leak information and disregard compliance regulations.



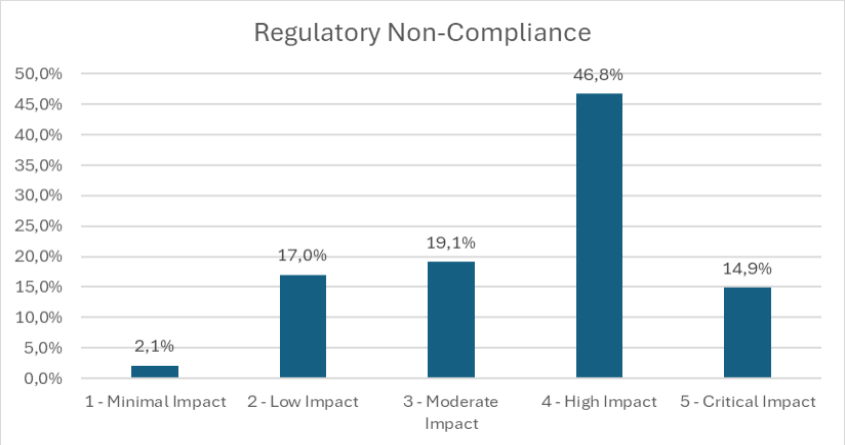
Statistic	Value
Mean	3,43
Median	4,00
Mode	4
Standard Deviation	0,88
Variance	0,77
Minimum	2
Maximum	5
Range	3
Count	47
Skewness	-0,17

Figure 26 - Insider Misuse

Regulatory Non-Compliance

The distribution of Regulatory Non-Compliance is left-skewed (-0,54), with a mean 3,55, median 4 and mode 4. Over half of the responses are towards the highest impact, with 61,7% being High or Critical Impact. 19,1% reported Low or Minimal Impact, as in Moderate (19,1%).

Overall, regulatory non-compliance is rated as high impact (46,8%), with 14,9% rating it as critical, as this is due to AI errors that might generate biased or hallucinated outputs or disregard the GDPR. Furthermore, institutions are accountable for these risks, as regulations like EU AI Act or GDPR state.



Statistic	Value
Mean	3,55
Median	4,00
Mode	4
Standard Deviation	1,02
Variance	1,04
Minimum	1
Maximum	5
Range	4
Count	47
Skewness	-0,54

Figure 27 - Regulatory Non-Compliance

Overall

All Risk Exposure metrics have means between 3 and 4, suggesting that respondents rate risk exposure from attack techniques from moderate to high levels. Regulatory non-compliance has the highest mean (3,55) with adversarial inputs (3,49) and model evasion (3,47) and insider misuse (3,43) following closely. Model poisoning has the lowest mean (3,32).

Overall, there is a high risk awareness among respondents. The most frequent value in all datasets is 4 and all distributions are weighted towards the higher impact. Results show that apart from the respondents' expertise in this field, strict compliance requirements in the FSS and the external pressure from regulatory authorities and the EU AI Act are also creating awareness. Further analysis will be conducted on the inferential statistics section of the dissertation (Hypothesis Test 8).

Furthermore, all boxplots except Model Theft show a narrow IQR of 1 point, meaning that the middle 50% of respondents largely agree with each other, with most responses ranging from moderate to high impact. They also share the strongest agreement, with a median of 4, indicating that they have a high impact. On the other hand, Model Theft boxplot shows a wide IQR of 2 points, meaning that the middle 50% of respondents have polarized opinions, with responses ranging from low to high impact. Opinions are more divided on whether model theft is major risk, but more consistent for all other categories.

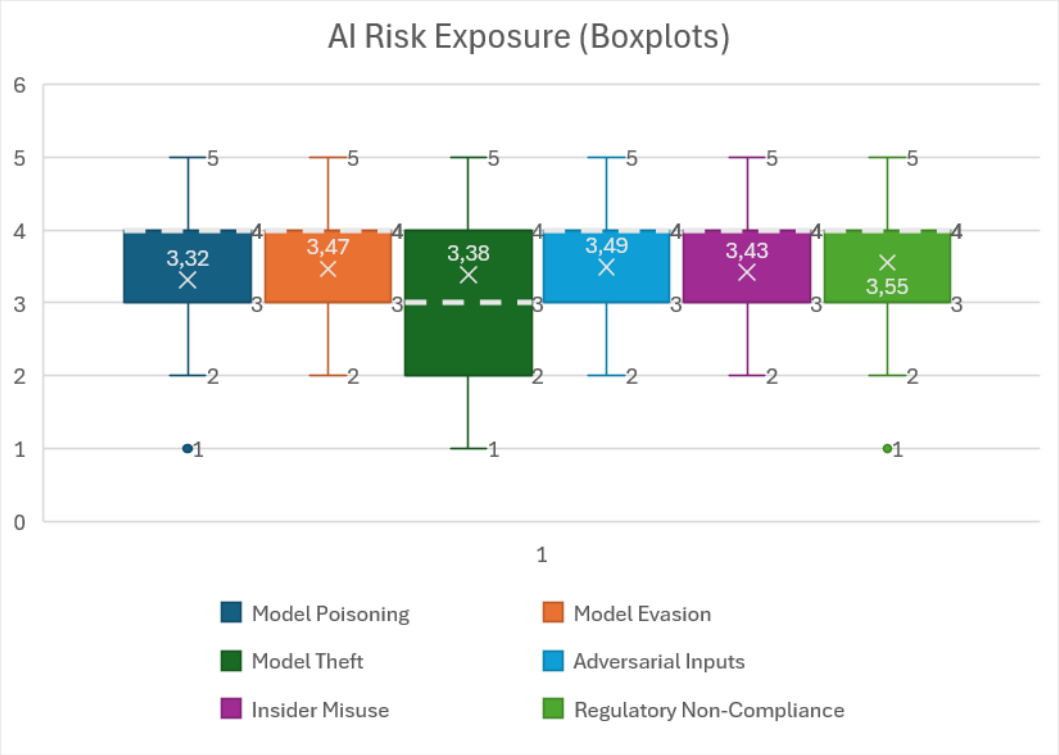


Figure 28 - AI Risk Exposure (Boxplots)

Qualitative Analysis

"Given these ratings, what specific vulnerability scenario keeps you up at night?"

Responses are categorized in themes, based on the frequency of certain words or phrases. Responses are quite polarized, following the pattern of the mean comparison.

- Model Poisoning: Many respondents stated that model poisoning is hard to detect and leads to regulatory catastrophe.
- Adversarial Inputs: Many respondents stated that this technique is quite dangerous as it can exfiltrate sensitive data without breaching infrastructure.
- AI Explainability: Some stated that trusting AI's decisions without the ability to explain or audit them accordingly, is a dangerous approach.

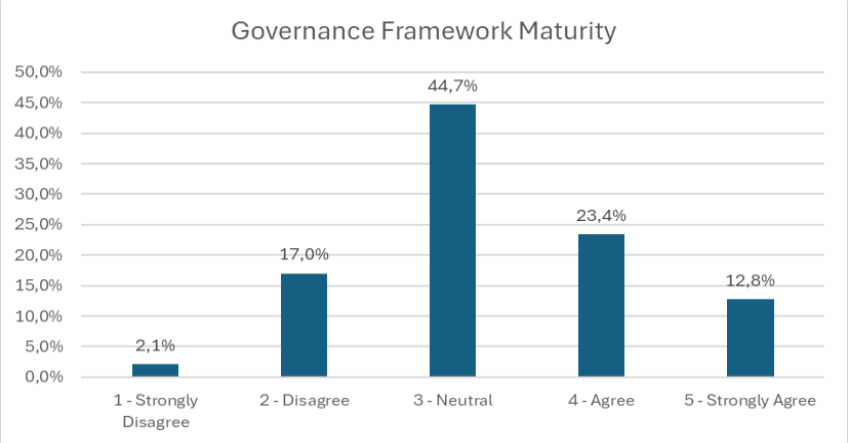
4.1.8 Governance, Opacity, and Regulation

"Rate your agreement with the following statements."

Governance Framework Maturity

"Our organization possesses an audit-ready, mature governance framework specifically tailored for AI risks."

The distribution of Governance Framework Maturity is centered-based (skewness 0,15), with a mean 3,28, median 3 and mode 3. Responses are towards the neutral side, with 36,2% of responses being Agree and Strongly Agree, whereas the Neutral responses are 44,7%. Negative responses are 19,1%. Results show that institutions may lack AI risk specific frameworks, as 63,8% of respondents are either not sure or disagree.



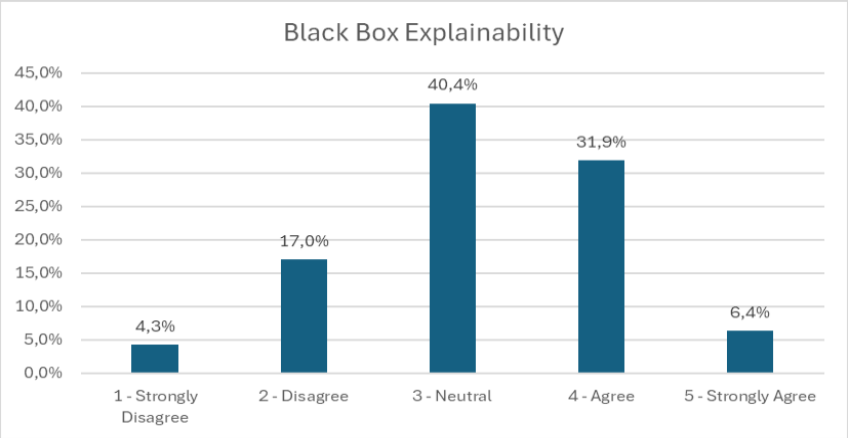
Statistic	Value
Mean	3,28
Median	3,00
Mode	3,00
Standard Deviation	0,97
Variance	0,94
Minimum	1
Maximum	5
Range	4
Count	47
Skewness	0,15

Figure 29 - Governance Framework Maturity

Black Box Explainability

"I am fully confident in our ability to explain the logic of our 'Black Box' AI decisions to external regulators."

The distribution of Black Box Explainability is slightly center-balanced (skewness -0,24), with a mean 3,19, median 3 and mode 3. Responses are towards the neutral side, with 38,3% of responses being Agree and Strongly Agree, whereas the Neutral responses are 40,4%. Negative responses are 21,3%. Results show that 38,3% of respondents are confident in explaining the Black Box to auditors and regulatory authorities, while 61,7% are not. This indicates the lack of XAI capabilities within the institutions.



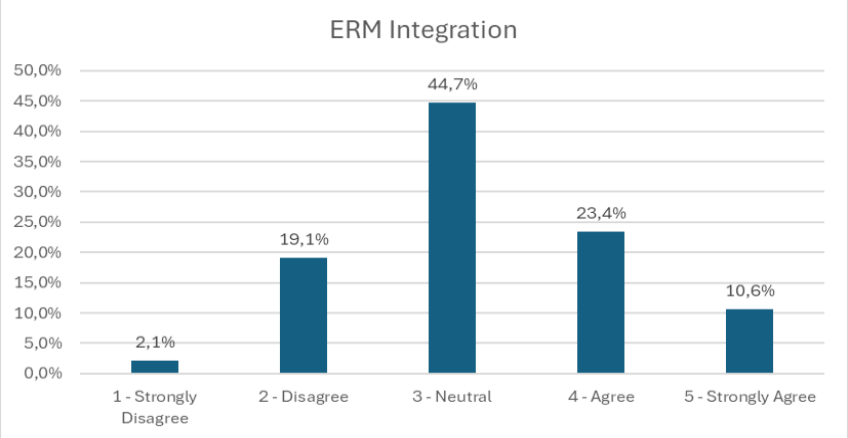
Statistic	Value
Mean	3,19
Median	3,00
Mode	3,00
Standard Deviation	0,95
Variance	0,90
Minimum	1
Maximum	5
Range	4
Count	47
Skewness	-0,24

Figure 30 - Black Box Explainability

ERM Integration

"AI-specific vulnerabilities are seamlessly integrated into our traditional Enterprise Risk Management (ERM) system."

The distribution of ERM Integration is centered-based (skewness 0,18), with a mean 3,21, median 3 and mode 3. Responses are towards the neutral side, with 34% of responses being Agree and Strongly Agree, whereas the Neutral responses are 44,7%. Negative responses are 21,2%. Results show only 34% have integrated AI specific vulnerabilities in their ERM system, while 66% are either not sure or they haven't. This indicates that AI security is managed in isolation, separately from ERM.



Statistic	Value
Mean	3,21
Median	3,00
Mode	3,00
Standard Deviation	0,95
Variance	0,91
Minimum	1
Maximum	5
Range	4
Count	47
Skewness	0,18

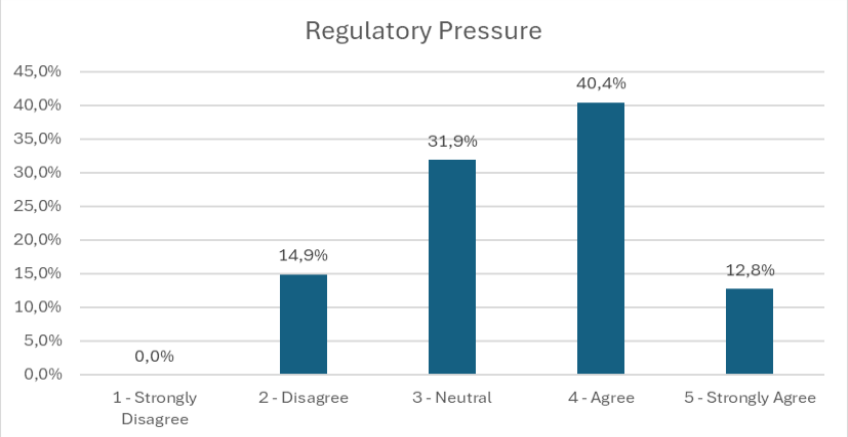
Figure 31 - ERM Integration

Regulatory Pressure

"Impending regulatory pressure (e.g., EU AI Act) is the primary driver for our current investments in AI governance."

The distribution of Regulatory Pressure is left-skewed (-0,12), with a mean 3,51, median 4 and mode 4. Responses are weighted towards the positive side, with 53,2% of responses being Agree and Strongly Agree, whereas the Neutral responses are 31,9%. Negative responses are 14,9%. Results show that 53,2% of the respondents believe that external regulatory pressure is driving governance investments, while 46,8% are either not sure or disagree.

In the previous AI Risk Exposure analysis, regulatory non-compliance is rated as 61,7% being High or Critical Impact, yet only 53.2% say regulatory pressure is what drives investments in AI governance. This indicates either that governance maturity does not depend on external regulatory pressure, or that there is a misalignment between governance and risk perception.



Statistic	Value
Mean	3,51
Median	4,00
Mode	4,00
Standard Deviation	0,91
Variance	0,82
Minimum	2
Maximum	5
Range	3
Count	47
Skewness	-0,12

Figure 32 - Regulatory Pressure

Overall

The results show that ERM integration (mean 3,19) and explainability (mean 3,21) fall behind, indicating that investing in XAI capabilities and integrating AI risks in ERM systems is crucial. On the other hand, there seems to be a gap between the governance

maturity (mean 3,28) and regulatory pressure (mean 3,51), that needs to be addressed by using external pressure as a driver to develop AI ready governance documentation.

Furthermore, all boxplots show a narrow IQR of 1 point, meaning that the middle 50% of respondents largely agree with each other, with most responses ranging from neutral to agree. Generally, opinions are consistent for all categories with Regulatory Pressure sharing the strongest agreement, with a median of 4, indicating that is indeed a primary driver in AI governance investments.

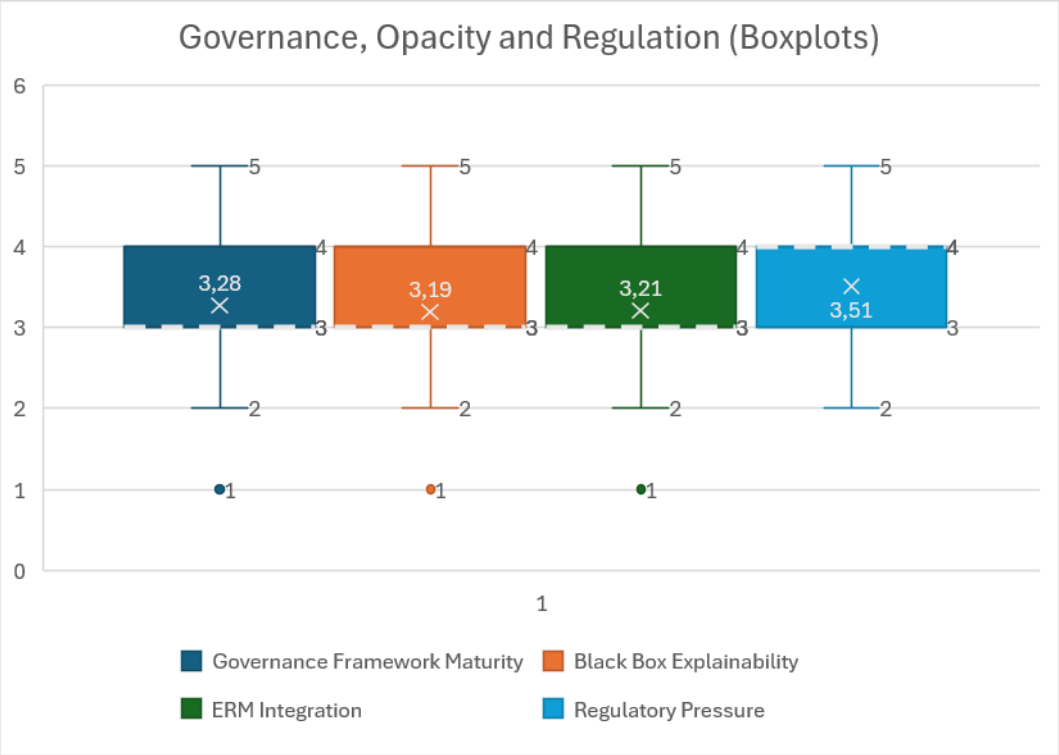


Figure 33 - Governance, Opacity and Regulation (Boxplots)

Qualitative Analysis

"Who ultimately owns the risk of an insecure or biased AI model in your firm, and how is that accountability enforced?"

Responses are categorized in themes, based on the frequency of certain words or phrases. In this case, responses are quite fragmented, of who ultimately owns the risk, which indicates a variety of RACI models inside institutions.

- Diverse Answers: Answers include the following roles ordered by frequency appearance: Business owner, CISO, C-level Directors, Dedicated AI Committee, Shared Accountability, IT, Unclear.

4.1.9 Budget Allocation

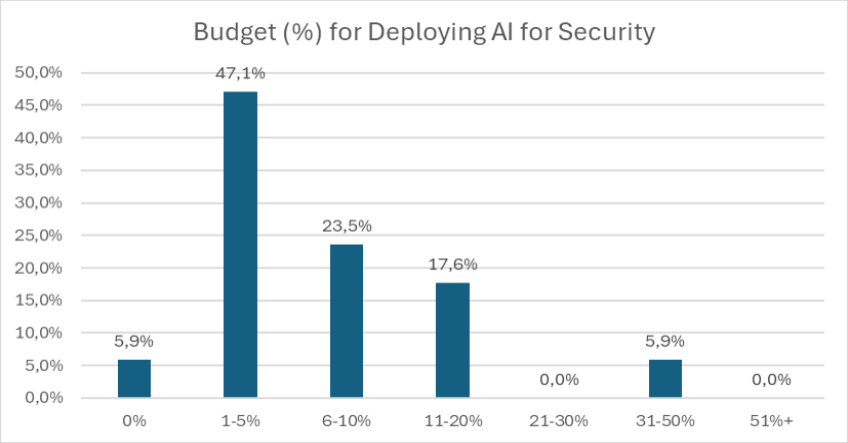
Important Notice: Questions about budget allocation were optional and the final sample size is $n = 17$*, which is quite small.*

AI for Security – Budget Percentage

"Of your total budget dedicated to AI & Security, roughly what percentage (%) is allocated to 'Deploying AI for defense'?"

On average, institutions allocate 11,18% (mean) in AI deployment. Half of the respondents estimate that budget allocation for AI is 5% or lower, while the other half reported 5% or higher (median 5%). The most common response is 5% (mode). The distribution is right-skewed (2,56) and is weighted towards lower percentages. 47,1% of respondents reported low budget allocations (1%-5%). 23,5% of them reported 6%-10% and 17,6% of them reported 11%-20% budget allocation. Also, one respondent reported 0% (min value) and another one reported 50% (max value) budget allocation.

Overall, most of the budget allocation percentages for deploying AI for defense are between 1%-20%. Most institutions allocate very little budget (5%) to AI deployment, but a minority is investing significantly more, increasing the average up to 11,18%.



Statistic	Value
Mean	11,18
Median	5,00
Mode	5,00
Standard Deviation	11,66
Variance	136,03
Minimum	0
Maximum	50
Range	50
Count	17
Skewness	2,56

Figure 34 - Budget (%) for Deploying AI for Security

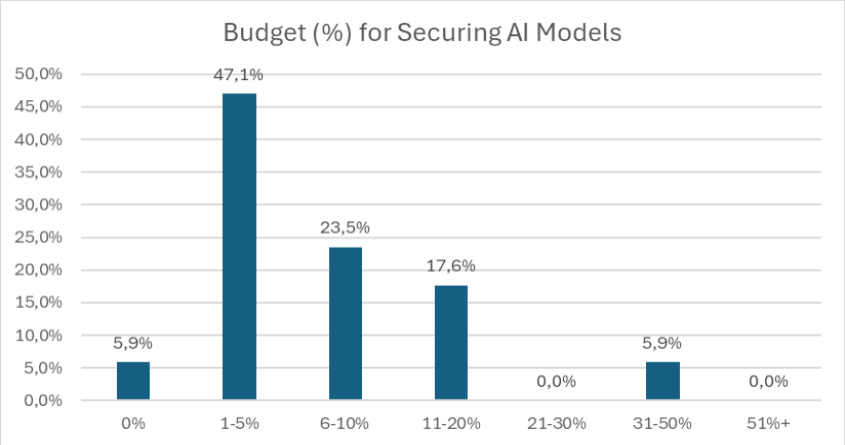
Security for AI – Budget Percentage

"Of your total budget dedicated to AI & Security, roughly what percentage (%) is allocated to 'Securing the AI models themselves'?"

On average, institutions allocate 11,06% (mean) in securing AI. Half of the respondents estimate that budget allocation for securing AI is 5% or lower, while the other half reported 5% or higher (median 5%). The most common response is 5% (mode). The distribution is again right-skewed (2,53) and is weighted towards lower percentages. 47,1% of respondents reported low budget allocations (1%-5%). 23,5% of them reported 6%-10% and 17,6% of them reported 11-20% budget allocation. Also, one respondent reported 0% (min value) and another one reported 50% (max value) budget allocation.

Overall, most of the budget allocation percentages for securing AI are between 1-20%. Most institutions allocate very little budget (5%) to securing AI, but a minority is investing significantly more, increasing the average up to 11,06 %.

Interestingly, results are the same as above, meaning that institutions allocate the same budget percentage for deploying AI and securing AI.



Statistic	Value
Mean	11,06
Median	5,00
Mode	5,00
Standard Deviation	11,74
Variance	137,81
Minimum	0
Maximum	50
Range	50
Count	17
Skewness	2,53

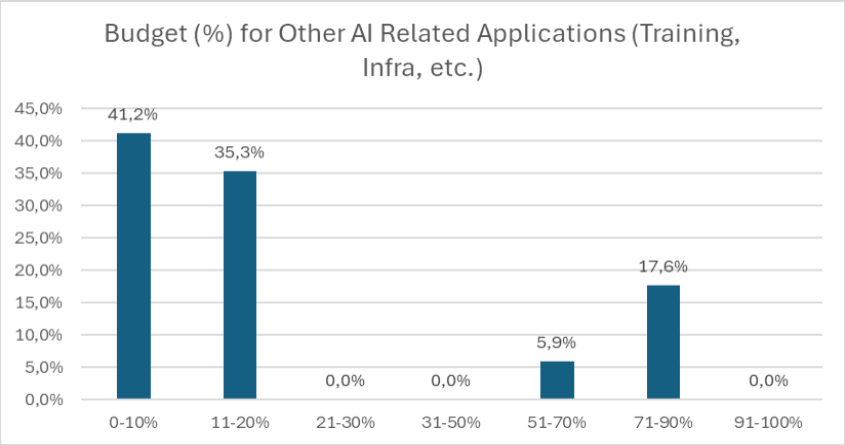
Figure 35 - Budget (%) for Securing AI Models

Other AI-related Applications – Budget Percentage

"Of your total budget dedicated to AI & Security, roughly what percentage (%) is allocated to other things (ex. AI-related staff training, Infrastructure, etc.)?"

On average, institutions allocate 29,53% (mean) in other AI-related applications. Half of the respondents estimate that budget allocation for other AI-related applications is 20% or lower, while the other half reported 20% or higher (median 20%). The most common response is 20% (mode). The distribution is u-shaped (skewness 1,39) and is weighted towards lower and upper percentages. 41,2% of respondents reported low budget allocations (1%-10%) and 35,3% of them reported 11%-20%. 23,5% of them reported 51%-90% budget allocation. Minimum value was 2% and maximum value was 90%.

Overall, most of the budget allocation percentages are between 2-20%, which indicates that institutions also consider investing in other AI-related applications, like training, infrastructure, etc. Most institutions allocate 20% of their budget, but a minority is investing significantly more, increasing the average up to 29,53 %.



Statistic	Value
Mean	29,53
Median	20,00
Mode	20,00
Standard Deviation	29,98
Variance	898,76
Minimum	2
Maximum	90
Range	88
Count	17
Skewness	1,39

Figure 36 - Budget (%) for Other AI Related Applications (Training, Infra, etc.)

Overall

The results are limited due to the small sample size $n - 17$. Also, percentages don't add up to 100% in each response. It might be an indication that respondents reported the percentage of the whole budget in IT, and this is a portion of it. In any case, these results need to be taken into consideration with precaution. The main indicators are that institutions spent on average 11,18% for deploying AI and 11,06% for securing AI, suggesting a balanced budget allocation between them. On the other hand, in other AI-related applications institutions allocate on average 2,6 times more budget (29,53%) than in AI deployment or AI security.

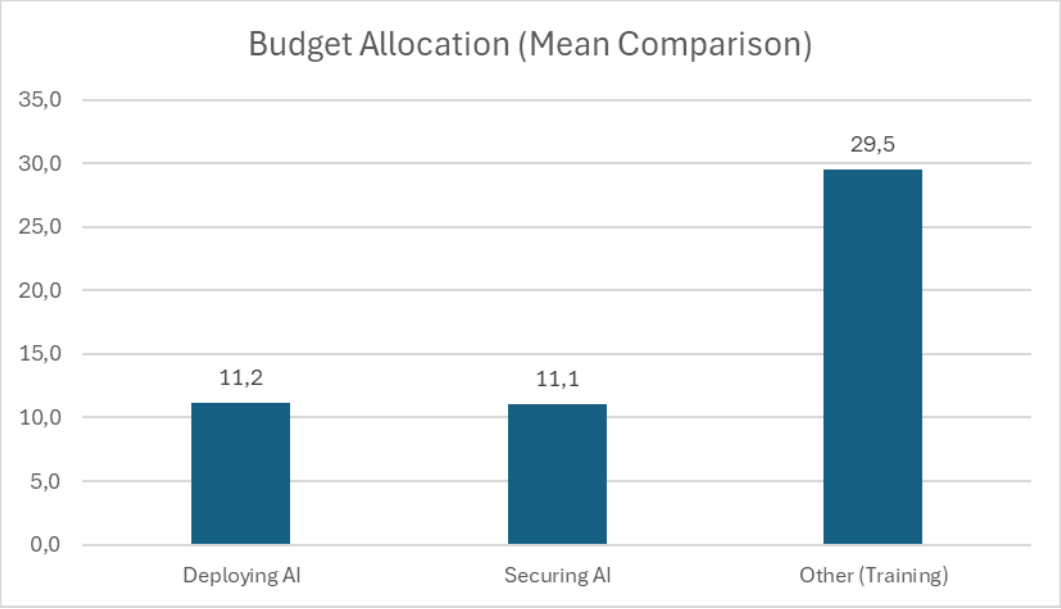


Figure 37 - Budget Allocation (Mean Comparison)

4.1.10 Cross-Team Alignment and Future Actions

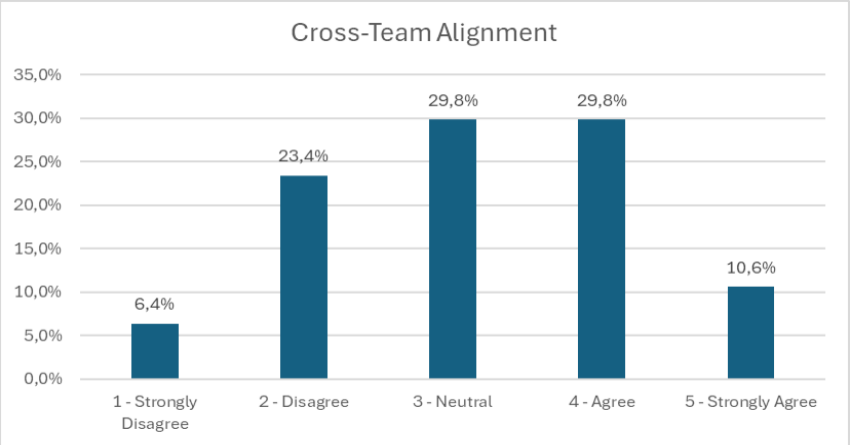
"Rate your agreement with the following statements."

Cross-Team Alignment

"Our AI Data Science teams and Cybersecurity teams operate in full partnership from the initial design phase."

The distribution of Cross-Team Alignment is slightly center-balanced (skewness -0,10), with a mean 3,15, median 3 and mode 4. Responses have a high variability (SD 1,10, Variance 1,22) with 29,8% of responses being Neutral and 29,8% of responses being Agree. Negative responses are 29,8%, whereas positive responses are 40,4%.

Overall responses in general are quite polarized. Results show that 29,8% of respondents report poor cross team alignment between the Security and Data Science teams, 40,4% report good collaboration, whereas 29,8% remain neutral about this relationship.



Statistic	Value
Mean	3,15
Median	3,00
Mode	4
Standard Deviation	1,10
Variance	1,22
Minimum	1
Maximum	5
Range	4
Count	47
Skewness	-0,10

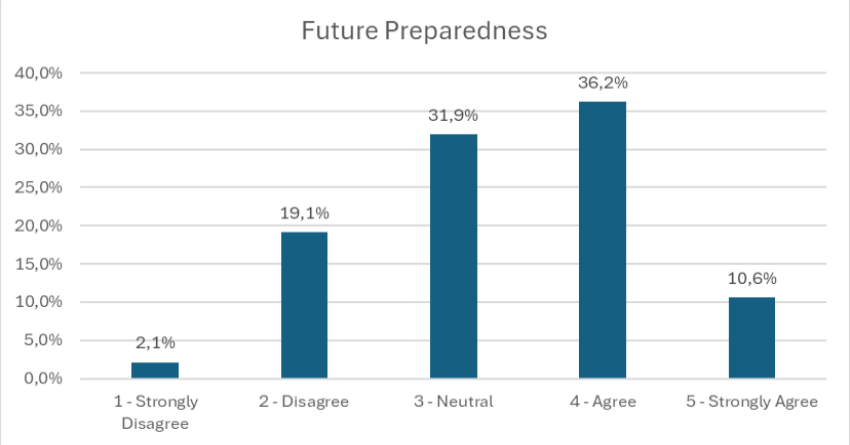
Figure 38 - Cross-Team Alignment

Future Preparedness

"Our organization is proactively prepared to defend against the next wave of AI-generated threats (e.g., deepfakes, autonomous malware)"

The distribution of Future Preparedness is left-skewed (-0,18), with a mean 3,34, median 3 and mode 4. Responses are towards the positive side, with 36,2% of responses being Agree and 10,6% being Strongly Agree. Only 2,1%% reported Strongly Disagree, whereas 19,1% reported Disagree. 31,9% remained Neutral.

Overall, 46,8% of respondents feel that their institutions are quite prepared for future AI threats, whereas 21,2% feel unprepared.



Statistic	Value
Mean	3,34
Median	3,00
Mode	4
Standard Deviation	0,98
Variance	0,97
Minimum	1
Maximum	5
Range	4
Count	47
Skewness	-0,18

Figure 39 - Future Preparedness

Overall

The results show that Cross-Team Alignment has a lower mean (3,15) from Future Preparedness (3,34). Both distributions have negative skewness which indicates that are weighted towards the negative answers, but the most frequent value is 4, which indicates that most of the institutions have good collaboration among teams and are future proof against AI threats.

Overall, 59,6% (including Neutral answers) of institutions need to address the poor collaboration through cross-functional team training. Also, 53,1% (including Neutral answers) of institutions need to prepare better for the future implementing a more proactive strategy.

Furthermore, Cross-Team Alignment boxplot shows a wide IQR of 2 points, meaning that the middle 50% of respondents have polarized opinions, with responses ranging from disagree to agree. On the other hand, Future Preparedness boxplot shows a narrow IQR of 1 point, meaning that the middle 50% of respondents largely agree with each other, with

most responses ranging from neutral to agree. Opinions are more divided on whether cross-teams are aligned, but more consistent for future preparedness.

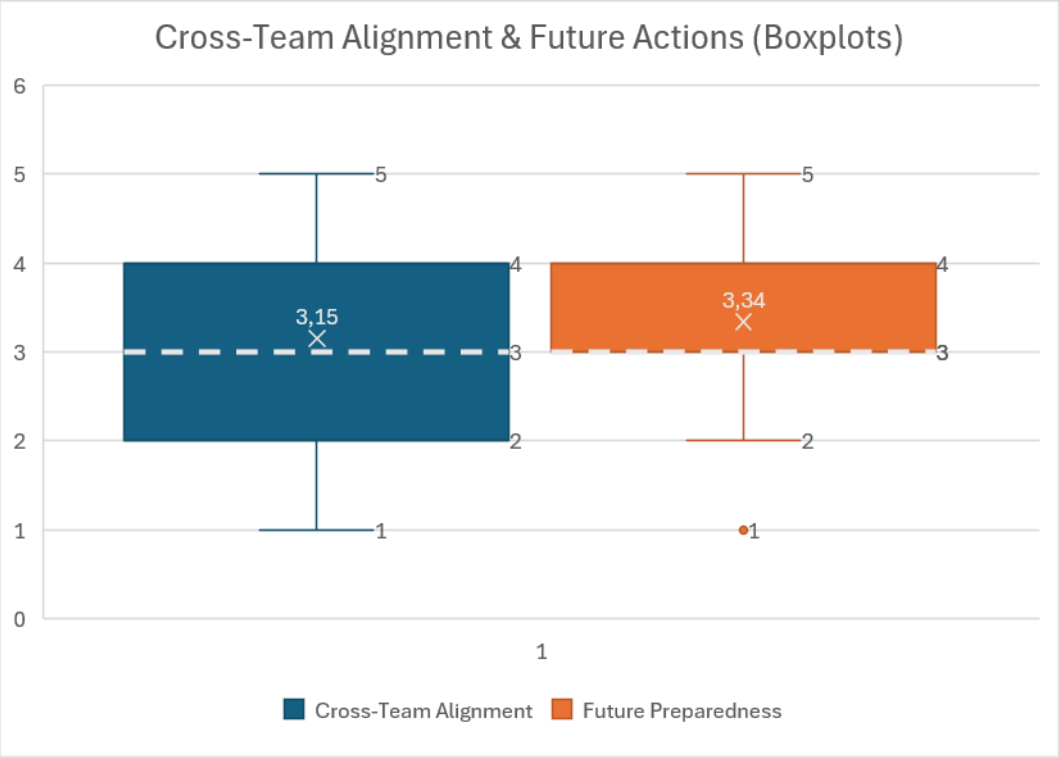


Figure 40 - Cross-Team Alignment and Future Actions (Boxplots)

Qualitative Analysis

"Looking at your budget allocation and team alignment, what major strategic shift is required in your organization over the next 3 years to close the gap between AI adoption and AI security?"

Responses are categorized in themes, based on the frequency of certain words or phrases. Governance frameworks, budget reallocations, team alignment and culture change are the most frequent answers.

- Governance Frameworks: Many respondents stated that robust AI governance frameworks need to be implemented across AI models' lifecycles.

- Budget: Many respondents stated that budgets need to be reallocated towards securing AI models and AI governance frameworks.
- Team Alignment: Many stated that AI, Security, Risk and Data Science teams should be in full alignment. These statements could again be an indication that a robust AI governance framework that all teams adhere, is a necessity.
- Culture change: Staff training and a top-down adoption approach from C-level executives.
- Regulation: A minority stated that AI regulation is required and proposed a "regulation-led roadmap" to that direction referencing the EU AI Act.
- Threat Testing Scenarios: A minority proposed end to end threat testing scenarios driven by AI agents.

4.2 Cronbach's Alpha Reliability Tests

The internal consistency of scales within the same structure was measured with Cronbach's alpha reliability test. Proving the strength of their reliability enables structures to be used in inferential statistics analysis. If a structure fails to pass the reliability test, only individual scales will be used in inferential statistics analysis. Acceptable internal consistencies are the ones that exceed the threshold $\alpha \geq 0{,}70$.

4.2.1 Cronbach's Alpha - Strategic Value & Competitive Advantage

Running the reliability test, proves that "Strategic Value & Competitive Advantage" construct is reliable, as the Cronbach's alpha exceeds the threshold 0,70 ($\alpha = 0{,}746$). All three items measure reliably the same construct.

Cronbach's α - Strategic Value & Competitive Advantage			
Items	**ROI**	**Market Differentiator**	**Budget Justification**
Variance (sample, VAR.S)	0,763	0,992	1,167
n (sample)	47	47	47
Cronbach's α Calculation			
k (number of items)	3		
Sum of item variances $\Sigma\sigma^2 i$	2,922		
Total variance $\sigma^2 t$	5,811		
Cronbach's α	**0,746**		

Table 1 - Cronbach's Alpha - Strategic Value & Competitive Advantage

4.2.2 Cronbach's Alpha - Disaggregated Implementation Barriers

Running the reliability test, proves that "Disaggregated Implementation Barriers" construct is not reliable, as the Cronbach's alpha is below the threshold 0,70 ($\alpha = 0{,}627$). All six items don't measure reliably the same construct.

Cronbach's α - Disaggregated Implementation Barriers						
Items	**Data Quality**	**Legacy IT**	**Talent Shortage**	**Regulatory Uncertainty**	**Cultural Resistance**	**Budget Constraints**
Variance (sample, VAR.S)	0,768	0,664	0,850	0,780	0,846	0,891
n (sample)	47	47	47	47	47	47
Cronbach's α Calculation						
k (number of items)	6					
Sum of item variances $\Sigma\sigma^2 i$	4,799					
Total variance $\sigma^2 t$	10,054					
Cronbach's α	**0,627**					

Table 2 - Cronbach's Alpha - Disaggregated Implementation Barriers

Running a diagnostic test, by removing one item from the structure and re-running the reliability test to see if alpha passes the threshold, unfortunately does not raise the alpha over the 0,70 threshold. These items may represent distinct dimensions and not a single construct, which is quite reasonable as implementation barriers are usually different for each institution and not all barriers appear in the same degree.

Diagnostic - Disaggregated Implementation Barriers			
Item Deleted	**$\Sigma\sigma^2 i$ (remaining)**	**$\sigma^2 t$ (remaining)**	**α if deleted**
Data Quality	4,031	8,429	**0,652**
Legacy IT	4,135	7,444	**0,556**
Talent Shortage	3,949	7,647	**0,604**
Regulatory Uncertainty	4,019	7,309	**0,563**
Cultural Resistance	3,953	6,671	**0,509**
Budget Constraints	3,908	7,514	**0,600**

Table 3 - Cronbach's Alpha - Diagnostic - Disaggregated Implementation Barriers

4.2.3 Cronbach's Alpha - Disaggregated AI Risk Exposure

Running the reliability test, proves that "Disaggregated AI Risk Exposure" construct is reliable, as the Cronbach's alpha exceeds the threshold 0,70 ($\alpha = 0{,}795$). All six items measure reliably the same construct.

Cronbach's α - Disaggregated AI Risk Exposure						
Items	**Model Poisoning**	**Model Evasion**	**Model Theft**	**Adversarial Inputs**	**Insider Misuse**	**Regulatory Non-Compliance**
Variance (sample, VAR.S)	0,918	0,733	1,502	0,734	0,772	1,035
n (sample)	47	47	47	47	47	47
Cronbach's α Calculation						
k (number of items)	6					
Sum of item variances $\Sigma\sigma^2 i$	5,693					
Total variance $\sigma^2 t$	16,888					
Cronbach's α	**0,795**					

Table 4 - Cronbach's Alpha - Disaggregated AI Risk Exposure

4.2.4 Cronbach's Alpha - Governance Opacity Regulation

Running the reliability test, proves that "Governance, Opacity and Regulation" construct is reliable, as the Cronbach's alpha exceeds the threshold 0,70 ($\alpha = 0{,}713$). All four items measure reliably the same construct.

Cronbach's α - Governance, Opacity, and Regulation				
Items	**Governance**	**Opacity**	**Integration**	**Regulatory Pressure**
Variance (sample, VAR.S)	0,944	0,897	0,910	0,821
n (sample)	47	47	47	47
Cronbach's α Calculation				
k (number of items)	4			
Sum of item variances $\Sigma\sigma^2 i$	3,572			
Total variance $\sigma^2 t$	7,680			
Cronbach's α	**0,713**			

Table 5 - Cronbach's Alpha - Governance Opacity Regulation

4.2.5 Cronbach's Alpha - Cross-Team Alignment and Future Actions

Running the reliability test, proves that "Cross-Team Alignment and Future Actions" construct is not reliable, as the Cronbach's alpha is below the threshold 0,70 ($\alpha = 0{,}627$). The two items don't measure reliably the same construct.

Cronbach's α - Cross-Team Alignment and Future Actions		
Items	**Cross-Team Alignment**	**Future Preparedness**
Variance (sample, VAR.S)	1,216	0,969
n (sample)	47	47
Cronbach's α Calculation		
k (number of items)	2	
Sum of item variances $\Sigma\sigma^2 i$	2,185	
Total variance $\sigma^2 t$	3,255	
Cronbach's α	**0,658**	

Table 6 - Cronbach's Alpha - Cross-Team Alignment and Future Actions

Cronbach's alpha often underestimates reliability because it assumes more than two items. Using the Spearman-Brown coefficient, treats the two items as two halves of a test. Unfortunately, the structure has questionable reliability, as the Spearman-Brown coefficient is below the threshold 0,70 ($r_{SB} = 0{,}660$). The two items may represent distinct dimensions and not a single construct, which is quite reasonable as better collaboration between teams surely helps but doesn't also mean better preparedness for the future.

Diagnostic - Cross-Team Alignment and Future Actions	
Inter-item correlation (r)	0,493
Spearman-Brown coefficient	**0,660**

Table 7 - Cronbach's Alpha – Diagnostic - Cross-Team Alignment and Future Actions

4.3 Inferential Statistics Analysis

This section presents the inferential statistics analysis on the ordinal values captured by the questionnaire from respondents in the FSS. There are 12 hypothesis tests based on the two parts "AI for Security" and "Security for AI" including 3 comparative tests across subgroups. The methodology justification for the selected tests was elaborated in section 3.3.1.

4.3.1 Hypothesis Test 1 - AI Maturity & Strategic Value

Hypothesis statement

FSS institutions' AI deployment maturity stage is associated with the perceived strategic value from AI for Security.

The null hypothesis (H0: $\rho = 0$) is that there is no monotonic relationship between AI deployment maturity stage and perceived strategic value from AI for security.

The alternative hypothesis (H1: $\rho \neq 0$) is that AI deployment maturity stage is monotonically associated with perceived strategic value from AI for security.

Variables, measurement and test selection

The variables used in this hypothesis test are:

- AI Deployment Maturity Stage: Values 1-4 that contain Pilot, Limited Production, Enterprise-Wide and Strategic Transformation.
- Strategic Value (composite): Mean of ROI, Market Differentiator and Budget Justification. The composite structure is acceptable through Cronbach's alpha ($\alpha = 0{,}746$).
- Sample size: $n = 47$ respondents.

The hypothesis will be tested through Spearman's rank-order correlation (two-tailed) because it tests the monotonic association between two variables, which matches the hypothesis statement.

Results and Conclusion

A Spearman rank-order correlation was calculated to assess the relationship between AI deployment maturity stage and perceived strategic value from AI for Security (composite of ROI, Market Differentiator, Budget Justification, (Cronbach's $\alpha = 0{,}746$)) among FSS respondents ($n = 47$). There is a statistically significant positive monotonic correlation, with $\rho = 0{,}388$, $t(45) = 2{,}822$, $p = 0{,}007$, at 95% CI $[0{,}113, 0{,}607]$.

Thus, the null hypothesis H0 is rejected at $\alpha = 0{,}05$. FSS institutions at higher AI deployment maturity stages perceive greater strategic value from AI for security.

Hypothesis Test 1 - AI Maturity / Strategic Value	
Statistic	**Value**
n	47
Spearman coefficient ρ	**0,388**
t-statistic	2,822
degrees of freedom	45
p-value (two-tailed)	**0,007**
95% Confidence Interval via Fisher's z	
z' = atanh(ρ)	0,409
Standard Error (z')	0,151
ρ_low (95%)	0,113
ρ_high (95%)	0,607

Table 8 - Hypothesis Test 1 - AI Maturity & Strategic Value

4.3.2 Hypothesis Test 2 - Automation and ROI

Hypothesis statement

The percentage of threat detection or fraud analysis automated by AI in FSS security operations is associated with perceived ROI realization from AI-powered security tools.

The null hypothesis (H0: $\rho = 0$) is that there is no monotonic relationship between the percentage of threat detection or fraud analysis automated by AI and perceived ROI realization from AI-powered security tools.

The alternative hypothesis (H1: $\rho \neq 0$) is that the percentage of threat detection or fraud analysis automated by AI is monotonically associated with perceived ROI realization from AI-powered security tools.

Variables, measurement and test selection

The variables used in this hypothesis test are:

- Percentage of cybersecurity tasks automated by AI: Values 0-100 of percentages.
- Perceived ROI: Values 1-5 in Likert Scale.
- Sample size: $n = 41$ respondents (6 missing responses in AI Automation).

The hypothesis will be tested through Spearman's rank-order correlation (two-tailed) because it tests the monotonic association between two variables, which matches the hypothesis statement.

Results and Conclusion

A Spearman rank-order correlation was calculated to assess the percentage of threat detection or fraud analysis automated by AI and perceived ROI realization from AI-powered security tools among FSS respondents ($n = 41$). There is a statistically significant positive monotonic correlation, with $\rho = 0{,}591$, $t(39) = 4{,}578$, $p = 0{,}000$, at 95% CI $[0{,}347, 0{,}761]$.

Thus, the null hypothesis H0 is rejected at $\alpha = 0{,}05$. FSS institutions that report higher percentage of their threat detection or fraud analysis automated by AI, perceive higher ROI realization from AI-powered security tools than those who do not.

Hypothesis Test 2 - Automation / ROI	
Statistic	**Value**
n	41
Spearman coefficient ρ	**0,591**
t-statistic	4,578
degrees of freedom	39
p-value (two-tailed)	**0,000**
95% Confidence Interval via Fisher's z	
z' = atanh(ρ)	0,680
Standard Error (z')	0,162
ρ_low (95%)	0,347
ρ_high (95%)	0,761

Table 9 - Hypothesis Test 2 - Automation and ROI

4.3.3 Hypothesis Test 3 - Governance Maturity and ROI

Hypothesis statement

FSS institutions' AI governance maturity is associated with perceived ROI realization from AI-powered security tools.

The null hypothesis (H0: $\rho = 0$) is that there is no monotonic relationship between AI governance maturity and perceived ROI realization from AI-powered security tools.

The alternative hypothesis (H1: $\rho \neq 0$) is that AI governance maturity is monotonically associated with perceived ROI realization from AI-powered security tools.

Variables, measurement and test selection

The variables used in this hypothesis test are:

- AI Governance Maturity: Values 1-5 in Likert Scale.
- Perceived ROI: Values 1-5 in Likert Scale.
- Sample size: $n = 47$ respondents.

The hypothesis will be tested through Spearman's rank-order correlation (two-tailed) because it tests the monotonic association between two variables, which matches the hypothesis statement.

Results and Conclusion

A Spearman rank-order correlation was calculated to assess the relationship between AI governance maturity and perceived higher ROI realization from AI-powered security tools among FSS respondents ($n = 47$). The correlation is not statistically significant, with $\rho = 0{,}199$, $t(45) = 1{,}365$, $p = 0{,}179$, at 95% CI $[-0{,}093, 0{,}460]$.

Thus, the null hypothesis H0 failed to reject at $\alpha = 0{,}05$. There is not sufficient evidence that AI governance maturity is associated with perceived ROI realization. FSS institutions with higher AI governance maturity do not necessarily perceive higher ROI realization from AI-powered security tools than those with lower governance maturity.

Hypothesis Test 3 - Governance Maturity / ROI	
Statistic	**Value**
n	47
Spearman coefficient ρ	**0,199**
t-statistic	1,365
degrees of freedom	45
p-value (two-tailed)	**0,179**
95% Confidence Interval via Fisher's z	
z' = atanh(ρ)	0,202
Standard Error (z')	0,151
ρ_low (95%)	-0,093
ρ_high (95%)	0,460

Table 10 - Hypothesis Test 3 - Governance Maturity and ROI

4.3.4 Hypothesis Test 4 - Cross-Team Alignment and Strategic Value

Hypothesis statement

Cross-team alignment between Data Science and Cybersecurity teams in FSS institutions is associated with perceived strategic value from AI for Security.

The null hypothesis (H0: $\rho = 0$) is that there is no monotonic relationship between cross-team alignment and perceived strategic value from AI for Security.

The alternative hypothesis (H1: $\rho \neq 0$) is that cross-team alignment is monotonically associated with perceived strategic value from AI for Security.

Variables, measurement and test selection

The variables used in this hypothesis test are:

- Cross-Team Alignment: Values 1-5 in Likert Scale.
- Strategic Value (composite): Mean of ROI, Market Differentiator and Budget Justification. The composite structure is acceptable through Cronbach's alpha ($\alpha = 0{,}746$).
- Sample size: $n = 47$ respondents.

The hypothesis will be tested through Spearman's rank-order correlation (two-tailed) because it tests the monotonic association between two variables, which matches the hypothesis statement.

Results and Conclusion

A Spearman rank-order correlation was calculated to assess the relationship between cross-team alignment between Data Science and Cybersecurity teams and perceived strategic value from AI for Security (composite of ROI, Market Differentiator, Budget Justification, (Cronbach's $\alpha = 0{,}746$)) among FSS respondents ($n = 47$). The correlation is not statistically significant, with $\rho = 0{,}285$, $t(45) = 1{,}994$ $p = 0{,}052$ at 95% CI $[-0{,}002, 0{,}529]$.

Thus, the null hypothesis H0 failed to reject at $\alpha = 0{,}05$. There is not sufficient evidence that Cross-team Alignment between Data Science and Cybersecurity teams is associated with perceived Strategic Value from AI for Security. The result narrowly fails to reach significance and would benefit from a larger sample.

Hypothesis Test 4 - Cross-Team Alignment / Strategic Value	
Statistic	**Value**
n	47
Spearman coefficient ρ	**0,285**
t-statistic	1,994
degrees of freedom	45
p-value (two-tailed)	**0,052**
95% Confidence Interval via Fisher's z	
z' = atanh(ρ)	0,293
Standard Error (z')	0,151
ρ_low (95%)	-0,002
ρ_high (95%)	0,529

Table 11 - Hypothesis Test 4 - Cross-Team Alignment and Strategic Value

4.3.5 Hypothesis Test 5 - AI Maturity and Risk Exposure

Hypothesis statement

FSS institutions' AI deployment maturity stage is associated with perceived exposure to AI-specific security risks.

The null hypothesis (H0: $\rho = 0$) is that there is no monotonic relationship between AI deployment maturity and perceived exposure to AI-specific security risks.

The alternative hypothesis (H1: $\rho \neq 0$) is that AI deployment maturity is monotonically associated with perceived exposure to AI-specific security risks.

Variables, measurement and test selection

The variables used in this hypothesis test are:

- AI Deployment Maturity: Values 1-4 that contain Pilot, Limited Production, Enterprise-Wide and Strategic Transformation.
- AI Risk Exposure (composite): Mean of Model Poisoning, Model Evasion, Model Theft, Adversarial Inputs, Insider Misuse and Regulatory Non-Compliance. The composite structure is acceptable through Cronbach's alpha ($\alpha = 0{,}795$).
- Sample size: $n = 47$ respondents.

The hypothesis will be tested through Spearman's rank-order correlation (two-tailed) because it tests the monotonic association between two variables, which matches the hypothesis statement.

Results and Conclusion

A Spearman rank-order correlation was calculated to assess the relationship between AI deployment maturity stages and AI Risk Exposure (composite of Model Poisoning, Model Evasion, Model Theft, Adversarial Inputs, Insider Misuse and Regulatory Non-Compliance, (Cronbach's $\alpha = 0{,}795$)) among FSS respondents ($n = 47$). The correlation is not statistically significant, with $\rho = 0{,}021$, $t(45) = 0{,}140$, $p = 0{,}889$ at 95% CI $[-0{,}268, 0{,}306]$.

Thus, the null hypothesis H0 failed to reject at $\alpha = 0{,}05$. There is not sufficient evidence that AI deployment maturity stages are associated with perceived AI Risk Exposure. FSS institutions at higher AI maturity stages do not necessarily perceive higher exposure to AI specific threats than institutions at lower maturity stages. AI risk exposure perception is independent of maturity stage, or institutions with higher AI maturity do not perceive higher risk exposure.

Hypothesis Test 5 - AI Maturity / Risk Exposure	
Statistic	**Value**
n	47
Spearman coefficient ρ	**0,021**
t-statistic	0,140
degrees of freedom	45
p-value (two-tailed)	**0,889**
95% Confidence Interval via Fisher's z	
z' = atanh(ρ)	0,021
Standard Error (z')	0,151
ρ_low (95%)	-0,268
ρ_high (95%)	0,306

Table 12 - Hypothesis Test 5 - AI Maturity and Risk Exposure

4.3.6 Hypothesis Test 6 - AI Maturity and Governance Maturity

Hypothesis statement

FSS institutions' AI deployment maturity stage is associated with mature AI governance practices.

The null hypothesis (H0: $\rho = 0$) is that there is no monotonic relationship between AI deployment maturity and AI governance maturity.

The alternative hypothesis (H1: $\rho \neq 0$) is that AI deployment maturity is monotonically associated with AI governance maturity.

Variables, measurement and test selection

The variables used in this hypothesis test are:

- AI Deployment Maturity Stage: Values 1-4 that contain Pilot, Limited Production, Enterprise-Wide and Strategic Transformation.
- AI Governance Maturity: Values 1-5 in Likert Scale.
- Sample size: $n = 47$ respondents.

The hypothesis will be tested through Spearman's rank-order correlation (two-tailed) because it tests the monotonic association between two variables, which matches the hypothesis statement.

Results and Conclusion

A Spearman rank-order correlation was calculated to assess the relationship between AI deployment maturity and AI governance maturity among FSS respondents ($n = 47$). The correlation is not statistically significant, with $\rho = 0{,}104$, $t(45) = 0{,}704$, $p = 0{,}485$ at 95% CI $[-0{,}188, 0{,}380]$.

Thus, the null hypothesis H0 failed to reject at $\alpha = 0{,}05$. There is not sufficient evidence that AI deployment maturity stages are associated with AI Governance maturity. FSS institutions at higher AI deployment maturity stages do not necessarily have more mature AI governance practices. This is an indication of a possible existing maturity gap in the FSS.

Hypothesis Test 6 - AI Maturity / Governance Maturity	
Statistic	**Value**
n	47
Spearman coefficient ρ	**0,104**
t-statistic	0,704
degrees of freedom	45
p-value (two-tailed)	**0,485**
95% Confidence Interval via Fisher's z	
z' = atanh(ρ)	0,105
Standard Error (z')	0,151
ρ_low (95%)	-0,188
ρ_high (95%)	0,380

Table 13 - Hypothesis Test 6 - AI Maturity and Governance Maturity

4.3.7 Hypothesis Test 7 - Cross-Team Alignment and Governance Maturity

Hypothesis statement

Cross-team alignment between Data Science and Cybersecurity teams in FSS institutions is associated with mature AI governance practices.

The null hypothesis (H0: $\rho = 0$) is that there is no monotonic relationship between cross-team alignment and AI governance maturity.

The alternative hypothesis (H1: $\rho \neq 0$) is that cross-team alignment is monotonically associated with AI governance maturity.

Variables, measurement and test selection

The variables used in this hypothesis test are:

- Cross-Team Alignment: Values 1-5 in Likert Scale.
- AI Governance Maturity: Values 1-5 in Likert Scale.
- Sample size: $n = 47$ respondents.

The hypothesis will be tested through Spearman's rank-order correlation (two-tailed) because it tests the monotonic association between two variables, which matches the hypothesis statement.

Results and Conclusion

A Spearman rank-order correlation was calculated to assess the relationship between cross-team alignment and AI governance maturity among FSS respondents ($n = 47$). The correlation is not statistically significant, with $\rho = 0{,}196, t(45) = 1{,}339, p = 0{,}187$ at 95% CI $[-0{,}097, 0{,}457]$.

Thus, the null hypothesis H0 failed to reject at $\alpha = 0{,}05$. There is not sufficient evidence that Cross-team alignment is associated with AI governance maturity. Cross-team alignment is not the driver for AI governance maturity.

Hypothesis Test 7 - Cross-Team Alignment / Governance Maturity	
Statistic	**Value**
n	47
Spearman coefficient ρ	**0,196**
t-statistic	1,339
degrees of freedom	45
p-value (two-tailed)	**0,187**
95% Confidence Interval via Fisher's z	
z' = atanh(ρ)	0,198
Standard Error (z')	0,151
ρ_low (95%)	-0,097
ρ_high (95%)	0,457

Table 14 - Hypothesis Test 7 - Cross-Team Alignment and Governance Maturity

4.3.8 Hypothesis Test 8 - Regulatory Pressure and Governance Maturity

Hypothesis statement

FSS institutions receiving regulatory pressure for AI, have more mature AI governance practices.

The null hypothesis (H0: $\rho \leq 0$) is that there is no positive monotonic relationship between regulatory pressure and AI governance maturity.

The alternative hypothesis (H1: $\rho > 0$) is that regulatory pressure is positively associated with AI governance maturity.

Theoretical explanation on hypothesis statement.

The hypothesis statement's direction is based on literature review and not based on results.

As mentioned in chapter 2.4.1 "*The EU AI Act, created by the European Union Council in 2024 is the strictest regulatory compliance framework, shifting financial firms' strategies from applying voluntary principles to comply with law requirements*".

Furthermore, chapter 2.4.2 mentions "*The Information Systems Audit and Control Association (ISACA) released a framework to comply with these regulations in 2025. According to them, financial firms should enforce the COBIT framework, meaning that AI governance should be incorporated in the enterprise risk management (ERM) process*".

Finally, from the previous hypothesis tests 5,6 and 7 the conclusions were:

(i) "*AI risk exposure perception is independent of maturity stage, or institutions with higher AI maturity do not perceive higher risk exposure*",
(ii) "*FSS institutions at higher AI deployment maturity stages do not have more mature AI governance practices*" and
(iii) "*Cross-team alignment is not the driver for AI governance maturity*".

Variables, measurement and test selection

The variables used in this hypothesis test are:

- Regulatory Pressure: Values 1-5 in Likert Scale.
- AI Governance Maturity: Values 1-5 in Likert Scale.
- Sample size: $n = 47$ respondents.

The hypothesis will be tested through Spearman's rank-order correlation (one-tailed). As mentioned in the hypothesis statement, a one-tailed test is selected based on theoretical

explanation existing in the literature review. The 95% confidence interval is reported for transparency reasons.

Results and Conclusion

A Spearman rank-order correlation was calculated to assess the relationship between regulatory pressure and AI governance maturity among FSS respondents ($n = 47$). The correlation is statistically significant in the positive direction, with $\rho = 0{,}250$, $t(45) = 1{,}729$, $p = 0{,}045$.

Thus, the null hypothesis H0 is rejected at $\alpha = 0{,}05$. The 95% confidence interval is reported for transparency reasons. Regulatory pressure is positively associated with AI governance maturity. AI governance in FSS institutions is mostly affected by regulatory pressure and not by internal factors.

Hypothesis Test 8 - Regulatory Pressure / Governance Maturity	
Statistic	**Value**
n	47
Spearman coefficient ρ	**0,250**
t-statistic	1,729
degrees of freedom	45
p-value (one-tailed)	**0,045**
95% Confidence Interval via Fisher's z (Calculated for transparency reasons)	
z' = atanh(ρ)	0,255
Standard Error (z')	0,151
ρ_low (95%)	-0,041
ρ_high (95%)	0,501

Table 15 - Hypothesis Test 8 - Regulatory Pressure and Governance Maturity

4.3.9 Hypothesis Test 9 - ERM Integration and Governance Maturity

Hypothesis statement

ERM integration in FSS institutions is associated with mature AI governance practices.

The null hypothesis (H0: $\rho = 0$) is that there is no monotonic relationship between ERM integration and AI governance maturity.

The alternative hypothesis (H1: $\rho \neq 0$) is that ERM integration is monotonically associated with AI governance maturity.

Variables, measurement and test selection

The variables used in this hypothesis test are:

- ERM Integration: Values 1-5 in Likert Scale.
- AI Governance Maturity: Values 1-5 in Likert Scale.
- Sample size: $n = 47$ respondents.

The hypothesis will be tested through Spearman's rank-order correlation (two-tailed) because it tests the monotonic association between two variables, which matches the hypothesis statement.

Results and Conclusion

A Spearman rank-order correlation was calculated to assess the relationship between ERM integration and AI governance maturity among FSS respondents ($n = 47$). There is a statistically significant positive monotonic correlation, with $\rho = 0{,}646$, $t(45) = 5{,}673$, $p = 0{,}000$ at 95% CI $[0{,}440, 0{,}787]$.

Thus, the null hypothesis H0 is rejected at $\alpha = 0{,}05$. ERM integration is significantly associated with AI governance maturity. FSS institutions with more mature ERM integration have more mature AI governance practices. Together with the previous hypothesis results (6,7 and 8), this hypothesis reaches the conclusion that FSS institutions without a mature ERM platform (often driven by regulatory pressure) do not have a mature AI governance framework, regardless of their AI deployment maturity stage or team alignment level.

Hypothesis Test 9 - ERM Integration / Governance Maturity	
Statistic	**Value**
n	47
Spearman coefficient ρ	**0,646**
t-statistic	5,673
degrees of freedom	45
p-value (two-tailed)	**0,000**
95% Confidence Interval via Fisher's z (Calculated for transparency reasons)	
z' = atanh(ρ)	0,768
Standard Error (z')	0,151
ρ_low (95%)	0,440
ρ_high (95%)	0,787

Table 16 - Hypothesis Test 9 - ERM Integration and Governance Maturity

4.3.10 Hypothesis Test 10 - Comparative - Technology Vendors vs All Others on perceived Risk Exposure

Hypothesis Statement

FSS respondents in Technology Vendors perceive on average different AI specific risk exposure than respondents in other institution types in the FSS.

The null hypothesis (H0: $\mu(Tech\ Vendors\ Risk) - \mu(Others\ Risk) = 0$) is that there is no significant difference on perceived risk exposure between technology vendors and other institution types.

The alternative hypothesis (H1: $\mu(Tech\ Vendors\ Risk) - \mu(Others\ Risk) \neq 0$) is that there is a significant difference on perceived risk exposure between technology vendors and other institution types.

Variables, measurement and test selection

The variables used in this hypothesis test are:

- Group 1: Technology Vendors, $n = 20$.

- Group 2: Other Institution Types (Banking Services, Consulting Services, Financial Services, FinTech, Insurance Services, Payments Institution), $n = 27$.
- AI Risk Exposure: (Model Poisoning, Model Evasion, Model Theft, Adversarial Inputs, Insider Misuse and Regulatory Non-Compliance). The composite structure is acceptable through Cronbach's alpha ($\alpha = 0{,}795$).

The hypothesis will be tested through Welch's independent samples t-test (two-tailed) because it compares the means of two independent groups without assuming equal variances.

Results and Conclusion

A Welch's independent samples t-test was calculated to compare the perceived AI risk exposure between Technology Vendors ($n = 20$) and Other Institution types (Banking Services, Consulting Services, Financial Services, FinTech, Insurance Services, Payments Institution) ($n = 27$). Tech Vendors reported significantly higher AI risk exposure, with $mean\ difference = 0{,}439$, $t(44{,}6) = 2{,}401$, $p = 0{,}021$ at 95% CI $[0{,}070, 0{,}807]$.

Thus, the null hypothesis H0 is rejected at $\alpha = 0{,}05$. Tech Vendors perceive higher AI risk exposure than other institutions in the FSS. Furthermore, the disaggregation of AI risks resulted in Adversarial Inputs and Model Theft as the ones that made the difference. The other risks did not result in significant differences between the two groups. Finally, Hypothesis Test 5, resulted in "*AI risk exposure perception is independent of maturity stage, or institutions with higher AI maturity do not perceive higher risk exposure*". Compared to this result it seems that the AI risk exposure perception is sector dependent, given that respondents in tech vendors may have deeper knowledge and more information at their disposal.

Hypothesis Test 10 - Tech Vendors vs Others on perceived Risk Exposure					
Group	**n**	**Mean**	**Std Deviation**	**Std Error**	
Group 1 - Tech Vendors	20	3,692	0,502	0,112	
Group 2 - Other Institution Types	27	3,253	0,749	0,144	
Mean difference (Tech Vendors - Others)		**0,439**			
Statistic	**Value**				
Standard Error of difference (Welch's)	0,183				
t-statistic	2,401				
degrees of freedom	44,6				
p-value (two-tailed)	**0,021**				
95% CI lower	0,070				
95% CI upper	0,807				
Disaggregated Risk Exposure					
Item	**Mean Tech**	**Mean Others**	**Mean difference**	**t (Welch)**	**p (2-tail)**
Model Theft	**3,65**	**3,07**	**0,58**	**2,18**	**0,035**
Model Poisoning	3,65	3,33	0,32	1,28	0,208
Adversarial Inputs	**4,00**	***2,93***	**1,07**	**3,42**	**0,001**
Model Evasion	3,55	3,44	0,11	0,41	0,683
Insider Misuse	3,55	3,33	0,22	0,82	0,418
Regulatory Non-Compliance	3,75	3,41	0,34	1,14	0,259

Table 17 - Hypothesis Test 10 - Comparative - Tech Vendors vs Others on perceived Risk Exposure

4.3.11 Hypothesis Test 11 – Comparative - Security, Risk, Compliance vs Technology Roles on AI Validation Percentage

Hypothesis Statement

FSS respondents with Security, Risk, Compliance Roles report on average different percentages of AI Validation than respondents with Technology Roles.

The null hypothesis (H0: $\mu(Sec, Risk, Comp\ Validation) - \mu(Tech\ Validation) = 0$) is that there is no significant difference in AI validation process percentage between Security, Risk, Compliance Roles and Technology Roles.

The alternative hypothesis (H1: $\mu(Sec, Risk, Comp\ Validation) - \mu(Tech\ Validation) \neq 0$) is that there is a significant difference in AI validation process percentage between Security, Risk, Compliance Roles and Technology Roles.

Variables, measurement and test selection

The variables used in this hypothesis test are:

- Group 1: Security, Risk, Compliance Roles, $n = 17$.
- Group 2: Technology Roles, $n = 25$.
- 5 missing responses.
- AI Validation Percentage.

The hypothesis will be tested through Welch's independent samples t-test (two-tailed) because it compares the means of two independent groups without assuming equal variances.

Results and Conclusion

A Welch's independent samples t-test was calculated to compare the AI validation percentages between Security, Compliance, Risk roles ($n = 17$) and Technology Roles ($n = 25$). There is not a significant difference between the two groups, with $mean\ difference = 1{,}56$, $t(30{,}3) = 0{,}135$, $p = 0{,}894$ at 95% CI $[-22{,}06, 25{,}18]$.

Thus, the null hypothesis H0 failed to reject at $\alpha = 0{,}05$. There is not sufficient evidence that AI validation process in the FSS differs between roles. AI validation process seems to be a shared responsibility.

Hypothesis Test 11 - Sec/Risk/Comp vs Tech Roles on AI Validation (%)				
Group	**n**	**Mean**	**Std Deviation**	**Std Error**
Group 1 - Sec/Risk/Comp roles	17	64,12	39,22	9,51
Group 2 - Tech roles	25	62,56	32,90	6,58
Mean difference (SecRisk – Tech)		**1,56**		
Statistic	**Value**			
Standard Error of difference (Welch's)	11,57			
t-statistic	0,135			
degrees of freedom	30,3			
p-value (two-tailed)	**0,894**			
95% CI lower	-22,06			
95% CI upper	25,18			

Table 18 - Hypothesis Test 11 - Comparative - Security, Risk, Compliance vs Technology Roles on AI Validation Percentage

4.3.12 Hypothesis Test 12 - Comparative – High Profile vs Low Profile Roles in perceived Strategic Value

Hypothesis Statement

FSS respondents with High Profile Roles perceive on average different Strategic Value from deploying AI than respondents with Low Profile Roles.

The null hypothesis (H0: $\mu(High\ Profile\ Value) - \mu(Low\ Profile\ Value) = 0$) is that there is no significant difference in perceived strategic value from deploying AI between high profile and low profile roles.

The alternative hypothesis (H1: $\mu(High\ Profile\ Value) - \mu(Low\ Profile\ Value) \neq 0$) is that there is a significant difference in perceived strategic value from deploying AI between high profile and low profile roles.

Variables, measurement and test selection

The variables used in this hypothesis test are:

- Group 1: High Profile Roles, $n = 17$.
- Group 2: Low Profile Roles, $n = 30$.
- Strategic Value (composite): Mean of ROI, Market Differentiator and Budget Justification. The composite structure is acceptable through Cronbach's alpha (α= 0,746).

The hypothesis will be tested through Welch's independent samples t-test (two-tailed) because it compares the means of two independent groups without assuming equal variances.

Results and Conclusion

A Welch's independent samples t-test was calculated to compare the AI validation percentages between High Profile Roles ($n = 17$) and Low Profile Roles ($n = 30$). There is not a significant difference between the two groups, with $mean\ difference = 0{,}250$, $t(30{,}1) = 0{,}990, p = 0{,}330$ at 95% CI $[-0{,}265, 0{,}765]$.

Thus, the null hypothesis H0 failed to reject at $\alpha = 0{,}05$. There is not sufficient evidence that perceived strategic value from deploying AI in the FSS differs between high profile and low profile roles.

Hypothesis Test 12 - High vs Low Profile Roles in Strategic Value				
Group	**n**	**Mean**	**Std Deviation**	**Std Error**
Group 1 - Senior leadership	17	3,373	0,865	0,210
Group 2 - Operational	30	3,622	0,767	0,140
Mean difference (Senior – Operational)		**0,250**		
Statistic	**Value**			
Standard Error of difference (Welch's)	0,252			
t-statistic	0,990			
degrees of freedom	30,1			
p-value (two-tailed)	**0,330**			
95% CI lower	-0,265			
95% CI upper	0,765			

Table 19 - Hypothesis Test 12 - Comparative - High Profile vs Low Profile Roles in perceived Strategic Value

5. Discussion

5.1 Interpretation of Findings

This chapter contains an assessment of the observed results (Chapter 4) in relation to the existing literature review (Chapter 2) and how do they respond to the research questions. The dissertation is structured between two main topics: “AI for Security” and “Security for AI”. These topics encompass both the strategic and technical aspects of AI and Security to help in better understanding the connection between the two main topics. The strategic part focuses on Management, Business Value, Regulation and Corporate Governance. On the other hand, the technical part focuses on Algorithms, Exploits, Technical Value and Engineering Controls.

Chapter 4 contains the descriptive statistics analysis from the survey’s questionnaire along with qualitative responses from the optional open questions to further strengthen the descriptive statistics. Additionally, inferential statistics analysis contains 12 hypothesis tests, including 3 comparative tests that address the research questions. The discussion will be arranged in the following 5 topics that emerge from the data analysis, starting with general sample observations and following with specific findings.

5.2 Descriptive Profile

The sample contains 47 experts from the financial services sector, with different profiles and roles. From the 47 respondents, there are 27 distinct roles, covering a wide spectrum of expertise, with a mean of 18,5 years of experience, median and mode 20. The distribution is left-skewed (-0,39) indicating a highly experienced sample. They come from 2 main institution types, technology vendors (42,6%) and banking services (40,4%) with the rest, fintech, financial services, insurance services, payment institutions and consulting services accounting for 17%. Geographically, the main country of operation is Greece (74,5%) with the rest entries in Europe and global accounting for 25,5%. Role distribution by profile is split between 36,2% for high profile roles and 63,8% for low profile roles, while by function is split between 59,6% for technology roles and 40,4% for security, compliance and risk roles. These results reveal a fairly balanced sample.

The AI deployment maturity stage distribution reveals that the majority of institutions have passed the pilot stage, with only 6,4% still in this stage. 42,6% of respondents stated that their institution is at the Limited Production stage and the remaining 51% is split in half between the Enterprise-wide and the Strategic Transformation stages. This result reveals the high level of expertise in deploying AI among institutions and in combination with the respondents' high expertise and experience, helps this research to output useful findings.

However, AI threat detection automation presents a high variability in coverage percentage. While the mean is 47,20%, which indicates that institutions on average have automated nearly half of their threat/fraud detection systems with AI, the high variability (SD 28,51) shows that there is a wide range of automation across institutions, regardless of their AI deployment maturity stage. The AI security validation percentage also presents a high variability in coverage percentage. On average institutions validate 63,19% of their AI models. The mean is high due to the most common answer which indicates the mode 100%. But answers vary from 0% to 100% meaning that there is also a wide range of validation across institutions regardless of their maturity stage.

This wide distribution in both automation and validation indicates that these processes reflect organization specific operation practices and not an organizational capability that represents how far institutions have advanced in their AI deployment journey. In literature review (chapter 2.2.3), ACFE (2024) frames AI adoption as a managerial challenge rather than a technological one and is consistent to Mikalef and Gupta's (2021) opinion, that AI delivers value only when treated as an organizational capability and not as a software acquisition. Therefore, internal practices in both automation and validation could not predict institutions' position in the maturity cycle.

5.3 The Strategic Value of AI

The first finding involves the perceived strategic value of deploying AI for Security in relation to institutions' AI maturity stage and AI threat/fraud detection automation. **Hypothesis test 1** resulted in a positive correlation between AI deployment maturity stages and strategic value and competitive advantage from AI for Security ($\rho = 0{,}388, p = 0{,}007$). Also, **hypothesis test 2** resulted in a positive correlation between the percentage of threat detection or fraud analysis automated and perceived ROI realization from AI-powered

security tools. ($\rho = 0{,}591, p = 0{,}000$). Looking at those two together, indicates that strategic value is not created only by simply deploying AI, but from incorporating AI in operations and measuring its actual output. A mean of 3,62 on ROI with 55,3% of agreement and a mean of 3,53 on budget justification with 57,4% positive answers reinforces the fact that the actual value is realized.

In literature review (chapter 2.2.3), Joshi (2025) reports "25% in detection accuracy and 30% in operational efficiency" from automating security with AI. This evidence suggests that AI adoption is reported by gains. In the sample, institutions on average have automated nearly half of their threat or fraud analysis (mean 47,2%). Half of them reported that their automation level is at 50% or lower, while the other half reported 50% or higher (median 50%) and the most common automation level response is 50% (mode 50%). These results indicate that FSS institutions are at half of their automation journey and are ahead of what the literature suggests. However, with the fast pace of AI adoption this is typical. Also, Deloitte (2025) finds that 85% of financial institutions have increased AI investment with 67% expecting payback within two-to-four years. In the sample the AI maturity stage suggests that FSS institutions with a mean of 2,70 (corresponding to 67,5%) are past the Limited Production stage and moving forward to next more mature stages. This means that they have already realized value and thus they continue investing in AI as they have measurable returns. Mikalef and Gupta (2021) state that institutions that treat AI as an organizational capability rather than a software purchase see a competitive advantage against others and this seems to be the case here. In **hypothesis test 2** results suggest that ROI is realized by seeing the actual results in operations. Qualitative answers from respondents also support this claim by reporting that they see results in the form of cost reduction, decreased false-positives and increased customer trust.

5.4 The Maturity Gap

The second finding involves AI governance maturity in FSS institutions in relation to ROI, AI deployment maturity stage, cross-team alignment, regulatory pressure and ERM integration. In **hypothesis test 3** results showed that AI governance maturity does not necessarily correlate with ROI realization from deploying AI powered security tools ($\rho = 0{,}199, p = 0{,}179$). Also, in **hypothesis test 6** results showed that AI governance maturity

does not necessarily correlate with AI maturity stages ($\rho = 0{,}104, p = 0{,}485$). Again, in **hypothesis test 7** results showed that AI governance maturity does not necessarily correlate with cross-team alignment between Data Science and Cybersecurity teams ($\rho = 0{,}196, p = 0{,}187$). So, what drives AI governance maturity remains to be seen in hypothesis tests 8 and 9. While in **hypothesis test 8** results showed that AI governance maturity is associated with regulatory pressure, it has to be acknowledged that this assumption came from a one-tailed test, based on literature review in chapters 2.4.1 and 2.4.2. These chapters mention the EU AI Act as an external regulatory pressure point for institutions, as well as other organizations that release regulatory framework guides. While this is true, the **hypothesis test 8** ($\rho = 0{,}250, p = 0{,}045$, one-tailed) looks only in the positive direction to assess if this is a valid indication. However, what really drives AI governance maturity is ERM integration, as seen in **hypothesis test 9** ($\rho = 0{,}646, p = 0{,}000$). Also, in descriptive analysis, only 30% of respondents stated that AI is integrated into their existing ERM infrastructure, while 65,9% are either not sure or they haven't. Results in governance maturity aren't great either, with 63,8% of respondents being either not sure or disagree that their institutions have mature governance frameworks. Furthermore, in **hypothesis test 6** results showed that AI deployment maturity stages are not necessarily associated with AI governance maturity. Together, these results reveal a gap between AI deployment and AI governance, where institutions may rely on external pressure and ERM integration capabilities to close that gap, rather than internal alignment or strategic decisions.

In literature review (chapters 2.4.1 and 2.4.2), McKinsey (2025) states that current governance frameworks are fragmented, because existing MRM teams might not have the technical know-how to assess the risks of AI. Furthermore, Gartner (2024) claims that there is a gap between adoption and governance, meaning that internal auditors lack the confidence and expertise to successfully oversee AI control failures. In addition, ISACA (2025) states that enforcing the COBIT framework and incorporating it in the ERM process provides a holistic model that aligns AI initiatives with strategic business objectives. All the above strengthen the **hypothesis test 9** results, that FSS institutions with more mature ERM infrastructure, have higher governance maturity. On the other hand, Almada (2025) and Akinrele (2025) strengthen the **hypothesis test 8** (one-tailed) that external pressure (like the EU AI Act), forces strict regulations, as AI advances in a fast pace. Finally, Mosch (2025) states governance should be considered as a strategic asset rather than a cost of compliance.

Maybe this is the reason that in **hypothesis test 3**, AI governance maturity does not necessarily correlate with higher ROI realization.

5.5 Risk Perception

The third finding involves the distribution of perceived risk across institutions. In **hypothesis test 5** results showed that AI maturity stages do not necessarily correlate with the perceived AI risk exposure ($\rho = 0{,}021, p = 0{,}889$). Therefore, AI risk exposure perception is independent of the level of AI maturity stage. The real difference, though, is between sectors where technology vendors perceive higher AI risk exposure than other institutions in the FSS. This significant result comes from **hypothesis test 10** ($\rho = 0{,}439, p = 0{,}021$). Furthermore, the disaggregation of AI risks resulted in Adversarial Inputs (mean 4,00 vs 2,93) and Model Theft (mean 3,65 vs 3,07) as the ones that made the biggest difference in perception between sectors. The other risks did not result in significant differences. Overall, it seems that the AI risk exposure perception is sector dependent and not AI maturity driven. Respondents in tech vendors may have deeper knowledge and more information at their disposal, due to their advisory and implementation role across financial institutions.

In the descriptive analysis, aggregated risks have means ranging from 3,32 to 3,55. The most frequent value in all datasets is 4 and all distributions are weighted towards the higher impact level. Interestingly, Regulatory Non-Compliance has the highest mean (3,55), with 61,7% of respondents rating it high or critical, but Model Theft, where opinions are more divided (mean 3,38 and 48,9% high/critical rating) and Adversarial Inputs (mean 3,49 and 55,3% high/critical rating) are left behind. Being at the top of the perceived AI risk exposure is also evidence of the maturity gap, analyzed in chapter 5.4.

In literature review (chapter 2.4.1), EU AI Act (2024) classifies all AI systems used in finance for credit scoring and risk assessment, as high-risk systems that must be transparent, robust and strictly controlled. In addition, Almada (2025) states that financial institutions need to develop additional capabilities, such as technical and legal expertise in AI to overcome any implications due to strict regulations. Therefore, respondents may perceive AI risk exposure more through this compliance regulatory lens, rather than a technical threat lens.

The most significant result is that AI deployment maturity stages do not necessarily correlate with perceived AI risk exposure. This is surprising, because FSS institutions at higher AI deployment maturity stages should be able to perceive higher exposure to AI specific threats, since wider deployment expands the attack surface. In literature review (chapters 2.3.2, 2.6.1), there seems to be a disagreement, as Global Risk Institute (2024) describes adversarial attacks as threats that affect financial stability, while Ibitoye et al. (2025) make an argument for an escalating "Arms Race" against AI threats. However, since AI risk exposure perception is independent of the level of AI maturity stage, this leads to a risk perception gap in financial institutions, driven mainly but external regulations. This argument is reinforced by the higher risk perception of technology vendors, given their advisory and implementation role. On the other hand, recent developments like the Anthropic's "Claude Mythos" (2026), that demonstrates significant adversarial capabilities, has caused a huge disturbance in the banking industry. By demanding more mature AI governance practices, this event can be the driving force that will close any risk perception gap in financial institutions, regardless of their AI maturity stage level.

5.6 Operational Context

The fourth section involves several smaller findings from operational activities, such as implementation barriers, budget, team alignment and future preparedness. In the descriptive analysis, aggregated barriers have means ranging from 2,98 to 3,66. Legacy IT has the highest mean (3,66) with 61,7% of respondents rating it as severe or critical, followed by Data Quality (3,40), Talent Shortage (3,38), Cultural Resistance (3,26), Regulatory Uncertainty (3,21) and Budget Constraints (2,98). Qualitative answers from respondents don't differ from the criticality order. Interestingly, results show that budget does not limit the deployment of AI and is not a critical barrier as the others. Obviously, institutions that are past the budget constraints, have bigger barriers to overcome with Legacy IT and Data Quality being the hardest.

Results are consistent with the literature review (chapters 2.7.3, 2.2.3). Mateo-Casali et al. (2026) identify data quality, legacy systems and organizational resistance as the main barriers to successful AI implementation. Also, ACFE (2024) points out that AI implementation lags due to data quality barriers and budget limitations. One thing to notice

is that budget constraints, while having the lowest mean score, had also the most divided opinions. Horton's et al. (2025), state that as institutions are investing more, the ROI becomes harder to measure and slower to materialize. Therefore, budget is a real constraint when institutions already invested in AI, have to decide if reinvesting is a viable option.

Budget spending, which had the lowest participation with 17 respondents, showed consistent percentages among them, with AI for Security and Security for AI having almost the same mean percentages (11,2 and 11,1). Also, half of the respondents reported 5% for each which was also the most common percentage. Although there are limitations, such as the number of respondents ($n = 17$) and that percentages don't add up to 100% in each response, there is valuable information that will be dealt with precaution. There is a balanced budget allocation between AI for Security and Security for AI.

In Cross-Team Alignment responses are polarized, since they have a high variability (SD 1,10, Variance 1,22) with 29,8% of responses being Neutral, 29,8% negative 40,4% positive. Also, Cross-team Alignment is marginally not significantly associated with perceived Strategic Value from AI for Security. The **hypothesis test 4** ($\rho = 0{,}285, p = 0{,}052$) test would benefit from a larger sample size, as it rejected for 0,002.

Furthermore, Cross-team Alignment is not significantly associated with AI governance maturity, too. Thus, it is not the driver for AI governance maturity as seen in hypothesis test 7. However, in literature review (chapter 2.4.2), McKinsey & Company (2025) claims that current governance frameworks can benefit from supporting a collaborative effort among legal, risk and tech teams. This points to a gap between best practice and current reality. Governance maturity in financial institutions appears to be driven externally, by regulatory pressure and ERM integration (hypothesis tests 8 and 9), rather than from cross-team collaboration as McKinsey (2025) recommends.

In Future Preparedness 46,8% of respondents consider their institution prepared to defend against the next wave of AI threats, whereas 21,2% are not prepared. This might be a sign that 46,8% of institutions are more confident in their AI systems deployed for defense, indicating that they have shifted their defense from reactive to proactive. In literature review (chapter 2.5.1), Anashkin and Zhukova (2021) mentioned that with the help of predictive analytics, AI can identify possible cybersecurity incidents before they happen, giving organizations the ability to improve their defense.

Finally, the statistical analysis has two more comparative tests that strengthen the internal consistency of the sample. **Hypothesis tests 11** ($p = 0{,}894$) and **12** ($p = 0{,}330$) results show that:

(i) The validation process in the FSS does not differ among function roles and it seems to be a shared responsibility. So, validation process may not differ between function roles, but it still reflects organization specific operation practices and does not relate with institutions' AI deployment maturity stage, as seen in chapter 5.2.

(ii) The perceived strategic value from deploying AI does not differ between high profile and low profile roles. As seen in hypothesis test 1, FSS institutions at higher AI deployment maturity stages perceive greater strategic value from AI for security and is independent of role profiles.

6. Conclusion and Recommendations

6.1 Conclusion

The aim of this research is to evaluate the relationship between AI and Security in the Financial Services Sector and provide a balanced view of opportunities and risks, as well as to highlight the governance and technical controls required to exercise them properly. Through primary evidence from senior experts in the Financial Services Sector, in Greece, it presented results concerning the strategic value of using AI in security operations and the required governance maturity on securing the same AI systems.

The literature review was structured around six pillars, examining the two main topics of "AI for Security" and "Security for AI", both from a technological and a strategic point of view. The research was conducted through twelve hypothesis tests, including three comparative tests, and addressed the following areas: AI deployment maturity stages, threat detection automation, strategic value, AI risk exposure, governance, regulation and ERM integration. Findings revealed that "AI for Security" and "Security for AI" represent a single organizational capability rather than two independent processes.

Evidence suggests that financial institutions have moved from the pilot stages of experimental AI deployment to the integration of AI in their security operations. This integration in security operations is linked to actual results that strengthen the strategic value of AI and the realization of ROI. Specifically, value is acknowledged through the level of threat/fraud detection automation, rather than the actual AI maturity stage of institutions. This evidence suggests that AI adoption is reported by gains and institutions that treat AI as an organizational capability rather than a software purchase see a competitive advantage against others. So, value comes through existing operational effectiveness and not by adoption alone.

However, governance maturity is not balanced with the deployment maturity in terms of adoption. Evidence suggests that governance maturity does not necessarily correlate with ROI realization, AI maturity stages or team alignment. On the other hand, regulatory pressure and ERM integration were significantly associated with AI governance maturity. These results reveal a gap between AI deployment and AI governance, where institutions

may rely on external pressure and ERM integration capabilities to close that gap, rather than internal alignment or strategic decisions. So, current ERM systems exist and governance is built upon them, but being driven by regulatory pressure does not make them part of a single organizational capability.

Another discovery from the results is that AI risk perception is not associated that much by AI deployment maturity stages, but it is dependent on sector. Technology vendors have increased awareness on AI specific threats in relation to other FSS institutions, regardless of their AI deployment maturity stage. While these institutions rate regulatory non-compliance as their top concern, technology vendors are more concerned about model theft and adversarial inputs. It seems that institutions are more influenced by regulatory frameworks whereas technology vendors are more influenced by their hands-on experience from their advisory and implementation role across financial institutions. So, AI risk exposure perception is sector dependent and not AI maturity driven.

The above results are supported by the operational findings. Budget spending is the lowest implementation barrier, whereas legacy IT and data quality are the most fragmented. Results show that there is a balanced budget allocation between AI for Security and Security for AI. On the other hand, cross-team alignment between Data Science and Cybersecurity teams is quite polarized. Team alignment might not be associated with strategic value, but spending seems to be associated with it, regarding the realized ROI from existing operational effectiveness.

Overall, this dissertation connects the strategic and technological aspects of AI and Security in the literature and through empirical evidence provides useful information on (i) how AI deployment translates in strategic value, (ii) how the deployment race creates a governance maturity gap and (iii) how AI risk perception is formed. “AI for Security” and “Security for AI” are not two distinct processes, but two dimensions of a single organizational capability. FSS institutions that want to lead the strategic race to deploy AI and gain competitive advantages and operational efficiency, will be the ones that treat both dimensions as one organizational capability.

6.2 Main Limitations

There are some limitations both theoretical and technical.

Technology and AI industry are moving fast. Nearly every month new models arise, new specialized AI software is created, new threats, vulnerabilities and attack techniques are discovered and new regulatory requirements surface. So, this research only captures a "snapshot in time".

Empirical evidence on AI ROI, breach impact etc. comes from self-reporting industry surveys (ex. Deloitte, IBM, Gartner, ACFE) and may contain selection or social-desirability bias. Also, self-reporting by senior executives may introduce social-desirability bias, particularly in the form of overstatement of AI deployment or governance maturity.

Financial Institutions hardly ever disclose detailed information on security incidents, governance failures, or internal finance reports of AI investments, which is a limiting factor in the study analysis.

This dissertation focuses on the use of AI in the Financial Services Sector, which is a small portion of a broader sector of regulated industries that also use AI to achieve the same results (ex. Communications, Power & Energy, etc.).

The sample size is small ($n = 47$), due to the high level of expertise in the Financial Services Sector, in Greece and limits the statistical power of inferential tests. So, results are not representative of a wider European or global landscape.

Optional budget questions were limited to respondents in the financial services, since respondents from Technology Vendors cannot answer or estimate how budget is allocated inside financial institutions.

6.3 Future Research

This dissertation can be used for further research in the future and for several directions, as AI and Security apply in nearly every sector. The research combines two different literatures that are rarely seen together. The strategic management research and the technical research for two separate topics: "AI for Security" and "Security for AI".

However, the sample has a small size ($n = 47$) which can be a limiting factor and constraints the power of statistical analysis as it is geographically located in the Greek Financial Services Sector. A wider replication of the questionnaire to the European Financial Services Sector could potentially upgrade the power of statistical analysis. This

could enhance and allow re-testing of hypothesis test 4 (Cross-Team Alignment and Strategic Value), where the p-value (0,052) was marginally over the 0,05 threshold. In addition, the hypothesis test 8 (Regulatory Pressure and Governance Maturity, one-tailed) could be run under a two-tailed test, without the need for theoretical based observations. Finally, further data collection such as AI incident registries, or anonymized data from Security Incident Response Teams (SIRTs), could further enhance the statistical analysis by comparing sector-based results.

While this research is focused on the Financial Services Sector, it could be expanded to other regulated sectors, such as Telecommunications, Energy and Power, Healthcare and Public Administration. These sectors are already in the AI transformation journey as well and are expected to encounter similar situations as in the Financial Services Sector. This research could help them acknowledge the strategic value of AI, evaluate the maturity gap between AI deployment and governance, strengthen the risk perception awareness and assess their current governance frameworks in relation to existing ERM systems and external regulatory pressures. Furthermore, this research could be conducted again in three years and track the same institutions to assess the AI governance maturity gap after the EU AI Act regulations implementation. Finally, future research could benefit from assessing both the strategic and technical aspects of AI and Security through the lens of a governance framework.

References

Ahirrao, Y. S., Ansari, I., Azim, K. S., Bhujel, K., & Panchal, S. S. (2025). AI-powered financial strategy: Transforming business decision-making through predictive analytics. *Emerging Frontiers Library for The American Journal of Engineering and Technology*. https://emergingsociety.org/index.php/efltajet/article/view/289

Ajuwon, A., Oladuji, T. J., Akintobi, A. O., & Onifade, O. (2025). AI-powered transformations in financial services: Automation and innovation in investment and risk models. *Engineering and Technology Journal*. https://www.researchgate.net/profile/Tolulope-Joyce-Oladuji/publication/392907503_AI-Powered_Transformations_in_Financial_Services_Automation_and_Innovation_in_Investment_and_Risk_Models/links/6884e2754eccfb3f29c5707b/AI-Powered-Transformations-in-Financial-Services-Automation-and-Innovation-in-Investment-and-Risk-Models.pdf

Akinrele, O. (2025). Bridging AI governance and financial regulation. *WJARR*. https://wjarr.com/sites/default/files/fulltext_pdf/WJARR-2025-3710.pdf

Ali, H.A., Charfeddine, M., Ammar, B., Hamed, B.B., Albalwy, F., Alqarafi, A., & Hussain, A. (2024). Unveiling machine learning strategies and considerations in intrusion detection systems: A comprehensive survey. Frontiers. https://www.frontiersin.org/journals/computer-science/articles/10.3389/fcomp.2024.1387354/full

Almada, M. (2025). The EU AI Act: Logic, content, and implications for finance. *SSRN*. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5125735

Anashkin, Y. & Zhukova, M. (2021). Implementation of behavioral indicators in threat detection and user behavior analysis. *AISMA*. https://ceur-ws.org/Vol-3094/paper_1.pdf

Association of Certified Fraud Examiners (ACFE). (2024). *Anti-fraud technology benchmarking report*. https://www.acfe.com/fraud-resources/anti-fraud-technology-benchmarking-report

Bank of England. (2023). Artificial intelligence and machine learning: Discussion paper. https://www.bankofengland.co.uk/prudential-regulation/publication/2023/october/artificial-intelligence-and-machine-learning

Bijoy, T. (2025). Modernizing financial services with cloud technologies and generative AI. *IJCNIS*. https://www.researchgate.net/profile/Bijoy-Thomas-2/publication/395462695_Modernizing_Financial_Services_with_Cloud_Technologies_and_Generative_AI/links/68c576ca4eef4b024b8b5cb7/Modernizing-Financial-Services-with-Cloud-Technologies-and-Generative-AI.pdf

Brundage, M., Avin, S., Clark, J., Toner, H., Eckersley, P., Garfinkel, B., Dafoe, A., Scharre, P., Zeitzoff, T., Filar, B., Anderson, H., Roff, H., Allen, G. C., Steinhardt, J., Flynn, C., Ó hÉigeartaigh, S., Beard, S., Belfield, H., Farquhar, S., ... Amodei, D. (2018). The malicious use of artificial intelligence: Forecasting, prevention, and mitigation. https://arxiv.org/abs/1802.07228v1

Carlini, N., Cheng, N., Lucas, K., Moore, M., Nasr, M., Prabhushankar, V., Xiao, W., Angulu, H., Ben Asher, E., Bow, J., Bradwell, K., Buchanan, B., Forsythe, D., Freeman, D.,

Gaynor, A., Ge, X., Graham, L., Guru, K., Lakhani, H., . . . Troy, K. (2026, April 7). Assessing Claude Mythos Preview's cybersecurity capabilities. Anthropic. https://red.anthropic.com/2026/mythos-preview/

Council of the European Union. (2024). The AI Act. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52021PC0206

Faizal, N., Vadla, A., Devi, R., Madhuri, U., & Radha, K. (2025). AI in financial services: Fraud detection and risk management. *Research Gate*. https://www.researchgate.net/publication/390522361_AI_in_Financial_Services_Fraud_Detection_and_Risk_Management

Financial Stability Board. (2024). The Financial Stability Implications of Artificial Intelligence. https://www.fsb.org/2024/11/the-financial-stability-implications-of-artificial-intelligence/

Gartner. (2024). Gartner survey shows AI-related risks see greatest audit coverage increases in 2024 [Press release]. https://www.gartner.com/en/newsroom/press-releases/2024-03-21-gartner-survey-shws-ai-related-risks-see-greatest-audit-coverage-increases-in-2024

Gartner. (2025, November 19). Gartner identifies critical GenAI blind spots that CIOs must urgently address [Press release]. https://www.gartner.com/en/newsroom/press-releases/2025-11-19-gartner-identifies-critical-genai-blind-spots-that-cios-must-urgently-address0

Global Risk Institute. (2024). Adversarial machine learning risks. https://globalriskinstitute.org/publication/adversarial-machine-learning-risks-and-opportunities-for-financial-institutions/

Greshake, K., Abdelnabi, S., Mishra, S., Endres, C., Holz, T., & Fritz, M. (2023). Not what you've signed up for: Compromising real-world LLM-integrated applications with indirect prompt injection. *Cornell University*. https://arxiv.org/pdf/2302.12173

Hafez, I. Y., Hafez, A. Y., Saleh, A., Abd El-Mageed, A. A., & Abohany, A. A. (2025). A systematic review of AI-enhanced techniques in credit card fraud detection. *Journal of Big Data*. https://link.springer.com/article/10.1186/s40537-024-01048-8

Hamberg, L. (2025). The risks of using AI in software development. *Jellyfish*. https://jellyfish.co/library/ai-in-software-development/risks-of-using-generative-ai/

Hasanah, I. N., Insany, P. G., Kharisma, L. I., & Rahayu, D. N. (2025). Recent advancements in machine learning models for malware detection: A systematic literature review. *MDPI*. https://www.mdpi.com/2673-4591/107/1/78

Hawkins, B., & Thomson, C. (2025). Abusing MLOps platforms to compromise ML models and enterprise data lakes. *IBM*. https://www.ibm.com/think/x-force/abusing-mlops-platforms-to-compromise-ml-models-enterprise-data-lakes

Horschig, T. (2023). Harnessing the power of AI in SOAR: Enhancing incident response and automation. *Trellix*. https://www.trellix.com/blogs/platform/harnessing-the-power-of-ai-in-soar/

Horton, R., Michalski, J., Winters, S., Gunn, D., & Holland, J. (2025, October 22). AI ROI: The paradox of rising investment and elusive returns. *Deloitte*. https://www.deloitte.com/uk/en/issues/generative-ai/ai-roi-the-paradox-of-rising-investment-and-elusive-returns.html

Ibitoye, O., Abou-Khamis, R., Elshahaby, M., Matrawy, A., & Shafiq, O. M. (2025). The threat of adversarial attacks against machine learning in network security: A survey. *Research Gate*. https://www.researchgate.net/publication/387843025_The_Threat_of_Adversarial_Attacks_against_Machine_Learning_in_Network_Security_A_Survey

IBM. (2025, November 12). What data leaders need to know from the Cost of a Data Breach Report 2025. https://www.ibm.com/think/insights/data-matters/cost-of-a-data-breach

Ijiga, O. M., Idoko, I. P., Ebiega, G. I., Olajide, F. I., Olatunde, T. I., & Ukaegbu, C. (2024). Harnessing adversarial machine learning for advanced threat detection: AI-driven strategies in cybersecurity risk assessment and fraud prevention. *Open Access Research Journal of Science and Technology*. https://www.researchgate.net/profile/Godslove-I-Ebiega/publication/380459159_Harnessing_adversarial_machine_learning_for_advanced_threat_detection_AI-driven_strategies_in_cybersecurity_risk_assessment_and_fraud_prevention/links/663cf58c06ea3d0b7446a401/Harnessing-adversarial-machine-learning-for-advanced-threat-detection-AI-driven-strategies-in-cybersecurity-risk-assessment-and-fraud-prevention.pdf

International Monetary Fund (IMF). (2024). Artificial intelligence and its impact on financial markets and financial stability. https://www.imf.org/en/news/articles/2024/09/06/sp090624-artificial-intelligence-and-its-impact-on-financial-markets-and-financial-stability

ISACA. (2025). Leveraging COBIT for effective AI system governance. https://www.isaca.org/resources/white-papers/2025/leveraging-cobit-for-effective-ai-system-governance

Jakkal, V. (2023). Introducing Microsoft Security Copilot: Empowering defenders at the speed of AI. *Microsoft*. https://blogs.microsoft.com/blog/2023/03/28/introducing-microsoft-security-copilot-empowering-defenders-at-the-speed-of-ai/

Khalid, I., & Purdie, S. M. (2024). AI-powered SOC operations: Revolutionizing cyber security incident response and management. ResearchGate. https://www.researchgate.net/publication/389350761_AI-Powered_SOC_Operations_Revolutionizing_Cyber_Security_Incident_Response_and_Management

Khayat, M., Barka, E., Serhani, A. M., Sallabi, F., Shuaib, K., & Khater, H. M. (2025). Empowering security operation centers with artificial intelligence and machine learning—A systematic literature review. *IEEE*. https://ieeexplore.ieee.org/document/10850912

Lumenova AI. (2025). 3 hidden risks of AI for banks. https://www.lumenova.ai/blog/risks-of-ai-banks-insurance-companies/

Mastercard. (2025). AI is helping banks save millions by transforming payment fraud prevention. https://www.mastercard.com/global/en/news-and-trends/Insights/2026/ai-is-helping-banks-save-millions-by-transforming-payment-fraud-prevention.html

Mateo-Casali, M. A., Maaben, M., Heymann, H., Boza, A., Fraile, F., Leyendecker, L., Grunert, D., & Schmitt, R. H. (2026). Reference architecture for the design and implementation of AI systems in manufacturing in conformity to ISO/IEC 42001. *Taylor & Francis*. https://www.tandfonline.com/doi/full/10.1080/0951192X.2026.2664187

McKinsey & Company. (2025). How financial institutions can improve their governance of gen AI. https://www.mckinsey.com/capabilities/risk-and-resilience/our-insights/how-financial-institutions-can-improve-their-governance-of-gen-ai

Mersha, M., Lam, K., Wood, J., AlShami, A., & Kalita, J. (2025). Explainable artificial intelligence: A survey of needs, techniques, applications, and future direction. *Cornell University*. https://arxiv.org/pdf/2409.00265

Mikalef, P., & Gupta, M. (2021). Artificial intelligence capability: Conceptualization, measurement calibration, and empirical study on its impact on organizational creativity and firm performance. *Research Gate*. https://www.researchgate.net/publication/348738492_Artificial_Intelligence_Capability_Conceptualization_measurement_calibration_and_empirical_study_on_its_impact_on_organizational_creativity_and_firm_performance

Minchae, K., Heemin, C., & Heeyoul, K. (2026). A blockchain-based governance framework for AIBOM integrity and automated regulatory compliance. *IEEE*. https://ieeexplore.ieee.org/abstract/document/11431396

MITRE Corporation. (2024). MITRE ATLAS (Adversarial Threat Landscape for Artificial-Intelligence Systems). https://atlas.mitre.org/

Mosch, M. M. M. (2025). Artificial intelligence as a strategic resource for competitive repositioning. https://arno.uvt.nl/show.cgi?fid=186219

Nathanson, S., Lee, A., Kieffer, C. C., Junkin, J., Ye, J., Saeed, A., Lockhart, M., Fink, R., Peterson, E., & Watkins, E. (2025). AI bill of materials and beyond: Systematizing security assurance through the AI risk scanning (AIRS) framework. *Cornell University*. https://arxiv.org/pdf/2511.12668

Organisation for Economic Co-operation and Development (OECD). (2023). OECD AI Principles. https://oecd.ai/en/dashboards/ai-principles/P9

OWASP GenAI Security Project. (2024). OWASP Top 10 for LLM applications 2025. https://genai.owasp.org/resource/owasp-top-10-for-llm-applications-2025/

Patel, A., Pandey, P., Ragothaman, H., Molleti, R., & Peddinti, R. D. (2025). Generative AI for automated security operations in cloud computing. *IEEE*. https://www.researchgate.net/profile/Advait-Patel/publication/388722404_Generative_AI_for_Automated_Security_Operations_in_Cloud_Computing/links/68a3f0031bee4d42a240b0e2/Generative-AI-for-Automated-Security-Operations-in-Cloud-Computing.pdf

Paul, A. A., & Ogburie, C. (2025). The role of AI in preventing financial fraud. *GSC*. https://gsconlinepress.com/journals/gscarr/sites/default/files/GSCARR-2025-0086.pdf

Poonum, S. R., & Shwetha, S. (2025). Artificial intelligence and financial fraud detection. *IJARSCT*. https://www.ijarsct.co.in/Paper26801.pdf

PwC. (2025). Responsible AI Survey: From policy to practice. https://www.pwc.com/us/en/tech-effect/ai-analytics/responsible-ai-survey.html

Radanliev, P., Santos, O., Brandon-Jones, A., & Joinson, A. (2024). Ethics and responsible AI deployment. Frontiers. https://www.frontiersin.org/journals/artificial-intelligence/articles/10.3389/frai.2024.1377011/full

Raff, E., Benaroch, M., Kukla, K., & Comprix, J. (2025). Adversarial machine learning attacks on financial reporting via maximum violated multi-objective attack. *Syracuse University*. https://arxiv.org/pdf/2507.05441

Sarker, H.I. (2021). Deep cybersecurity: A comprehensive overview from neural network and deep learning perspective. *Research Gate*. https://www.researchgate.net/publication/350216987_Deep_Cybersecurity_A_Comprehensive_Overview_from_Neural_Network_and_Deep_Learning_Perspective

Shokri, R., Stronati, M., Song, C., & Shmatikov, V. (2017). Membership inference attacks against machine learning models. *Cornell University*. https://arxiv.org/pdf/1610.05820

Shumailov, I., Zhao, Y., Bates, D., Papernot, N., Mullins, R., & Anderson, R. (2021). Sponge examples: Energy-latency attacks on neural networks. *Cornell University*. https://arxiv.org/pdf/2006.03463

Silic, M., Silic, D., & Kind-Trüller, K. (2025). From shadow IT to shadow AI: Threats, risks and opportunities for organizations. *Strategic Change*. https://onlinelibrary.wiley.com/doi/abs/10.1002/jsc.2682

Solaimani, S., & Long, P. (2025). Beyond the black box: Operationalising explicability in artificial intelligence for financial institutions. *International Journal of Business Information Systems*. https://www.inderscienceonline.com/doi/epdf/10.1504/IJBIS.2025.146837

Vandendriessche, W., Thijsman, J., D'hooge, L., Volckaert, B., & Sebrechts, M. (2026). AIBoMGen: Generating an AI bill of materials for secure, transparent, and compliant model training. *Cornell University*. https://arxiv.org/pdf/2601.05703

Vassilev, A., Oprea, A., Fordyce, A., Anderson, H., Davies, X., & Hamim, M. (2025). Adversarial machine learning: A taxonomy and terminology of attacks and mitigations. *NIST*. https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-2e2025.pdf

Xia, G., Chen, J., Yu, C., & Ma, J. A. (2023). Poisoning attacks in federated learning: A survey. *IEEE*. https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=10024252

Appendix: “Primary Data Collection Method”

Interview/Survey Questionnaire

Research Title: "AI for Security and Security for AI: A Strategic Evaluation in the Financial Services Sector."

Target Participants: Cybersecurity Directors, Chief Risk Officers, Compliance Officers, Heads of AI/Digital Strategy, IT Governance Directors, Big Data Directors, AI Model Directors, Technology Officers, Architects, Engineers, etc.

About:

This dissertation studies the critical, dual relationship between Artificial Intelligence and Security. The interview consists of 12 core questions, 5 optional qualitative questions, 3 budget questions and will last approximately 10 minutes. It is performed in the context of a post-graduate degree at the Hellenic Open University.

Spyridon Dafalias

Important Notice

You can skip budget questions **4.1A**, **4.1B**, **4.1C** if you are working in a Technology Vendor, because they do not apply to your organization.

Privacy Notice

Please be assured that no personal or identifiable information regarding you or your organization will be collected during this survey. All responses are strictly anonymous. The only data gathered will be your direct answers to the questionnaire, which will be securely stored, protected from unauthorized access, and used exclusively for statistical processing and analysis as part of this dissertation study. By completing and submitting this survey, you consent to your anonymized data being used solely for academic purposes.

Section 1: Introduction, Control Variables & Demographics

1.1A Respondent Role (Select one):

☐ Cybersecurity Director

☐ AI Director

☐ Technology Officer

☐ Risk Officer

☐ Compliance Officer

☐ IT Architect

☐ Security Architect

☐ Other: __

1.1B Years of experience:

……..

1.2A Company/Institution Type:

☐ Banking Services

☐ FinTech

☐ Financial Services

☐ Technology Vendor

☐ Other: __

1.2B Primary Region of Operation:

☐ Greece

☐ Europe

☐ Global

1.3 AI Deployment Maturity Stage (Categorical): How would you classify your organization's current AI deployment?

☐ Pilot: Exploring proofs of concept; no critical production.

☐ Limited Production: Deployed in specific, non-core silos.

☐ Enterprise-wide: Scaled across multiple business units.

☐ Strategic Transformation: AI is foundational to the core business model.

Section 2: Pillar 1 - "AI for Security" (Adoption & Strategic Benefits)

2.1 Objective Metric (Automation):

"Approximately what percentage of your threat detection or fraud analysis is currently automated by AI/ML models?" ____ **%**

2.2 Strategic Value & Competitive Advantage (Likert Scale):

Rate your agreement with the following statements on a scale of 1 to 5: (1 = Strongly Disagree, 2 = Disagree, 3 = Neutral, 4 = Agree, 5 = Strongly Agree)

A. ROI: "Our deployment of AI-powered security tools has delivered a highly measurable and mission-critical Return on Investment (ROI)."

Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree

B. Advantage: "Our use of AI for security serves as a significant market differentiator against our competitors."

Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree

C. Business Case: "It is currently straightforward to justify increased budget requests for AI security tools to the Board/CFO."

Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree

(Qualitative Question)

Looking at these scores, how specifically does this AI advantage manifest? Are you reducing operational costs, or increasing customer trust through fewer false positives?

2.3 Disaggregated Implementation Barriers:

Rate the severity of the following internal barriers preventing optimal AI security leverage: (1 = Not a barrier, 2 = Minor barrier, 3 = Moderate barrier, 4 = Severe barrier, 5 = Critical blocker)

A. Data quality and fragmentation limitations:

Not a barrier	Minor barrier	Moderate Barrier	Severe Barrier	Critical Blocker

B. Legacy IT infrastructure incompatibility:

Not a barrier	Minor barrier	Moderate Barrier	Severe Barrier	Critical Blocker

C. Lack of specialized AI/Security talent:

Not a barrier	Minor barrier	Moderate Barrier	Severe Barrier	Critical Blocker

D. Regulatory uncertainty / Compliance fears:

Not a barrier	Minor barrier	Moderate Barrier	Severe Barrier	Critical Blocker

E. Cultural resistance to automated decisions:

Not a barrier	Minor barrier	Moderate Barrier	Severe Barrier	Critical Blocker

F. Budget constraints / Capital allocation:

Not a barrier	Minor barrier	Moderate Barrier	Severe Barrier	Critical Blocker

(Qualitative Question)

If you rated one of the above as your critical blocker, can you elaborate on why this specific barrier is so difficult to overcome?

Section 3: Pillar 2 - "Security for AI" (Strategic Risks & Vulnerability)

3.1 Objective Metric (Validation):

"Approximately what percentage of your production AI models undergo a formal, documented security validation process (e.g., adversarial stress testing) prior to deployment?" ____ **%**

3.2 Disaggregated AI Risk Exposure:

Rate the perceived potential business impact of the following AI-specific attack vectors on your organization: (1 = Negligible Impact, 2 = Low Impact, 3 = Moderate Impact, 4 = High Impact, 5 = Catastrophic Impact)

A. Model Poisoning (manipulating training data):

Negligible Impact	Low Impact	Moderate Impact	High Impact	Catastrophic I.

B. Model Evasion (bypassing AI filters):

Negligible Impact	Low Impact	Moderate Impact	High Impact	Catastrophic I.

C. Model Theft / Intellectual Property extraction:

Negligible Impact	Low Impact	Moderate Impact	High Impact	Catastrophic I.

D. Adversarial Inputs (e.g., prompt injection):

Negligible Impact	Low Impact	Moderate Impact	High Impact	Catastrophic I.

E. Insider Misuse / Negligence of AI tools:

Negligible Impact	Low Impact	Moderate Impact	High Impact	Catastrophic I.

F. Regulatory Non-Compliance due to AI errors:

Negligible Impact	Low Impact	Moderate Impact	High Impact	Catastrophic I.

(Qualitative Question)

Given these ratings, what specific vulnerability scenario keeps you up at night?

3.3 Governance, Opacity, and Regulation:

Rate your agreement with the following statements (1 = Strongly Disagree to 5 = Strongly Agree):

A. Governance: "Our organization possesses an audit-ready, mature governance framework specifically tailored for AI risks."

Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree

B. Opacity: "I am fully confident in our ability to explain the logic of our 'Black Box' AI decisions to external regulators."

Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree

C. Integration: "AI-specific vulnerabilities are seamlessly integrated into our traditional Enterprise Risk Management (ERM) system."

Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree

D. Regulatory Pressure: "Impending regulatory pressure (e.g., EU AI Act) is the primary driver for our current investments in AI governance."

Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree

(Qualitative Question)

Who ultimately owns the risk of an insecure or biased AI model in your firm, and how is that accountability enforced?

Section 4: Strategic Alignment & Resource Allocation

Important Notice:

You can skip questions 4.1A, 4.1B, 4.1C if you are working in a Technology Vendor, because they do not apply to your organization.

4.1A Objective Metric (Budget Allocation):

"Of your total budget dedicated to AI & Security, roughly what percentage (%) is allocated to 'Deploying AI for defense'?"

……

4.1B Objective Metric (Budget Allocation):

"Of your total budget dedicated to AI & Security, roughly what percentage (%) is allocated to 'Securing the AI models themselves'?"

……

4.1C Objective Metric (Budget Allocation):

"Of your total budget dedicated to AI & Security, roughly what percentage (%) is allocated to other things (ex. AI-related staff training, etc.)?"

……

4.2 Alignment & Future Actions

Rate your agreement with the following statements (1 = Strongly Disagree to 5 = Strongly Agree):

A. Cross-Team Alignment: "Our AI Data Science teams and Cybersecurity teams operate in full partnership from the initial design phase."

Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree

B. Future Preparedness: "Our organization is proactively prepared to defend against the next wave of AI-generated threats (e.g., deepfakes, autonomous malware)

Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree

(Qualitative Question)

Looking at your budget allocation and team alignment, what major strategic shift is required in your organization over the next 3 years to close the gap between AI adoption and AI security?

End of Research.

Thank you for your contribution.

Author's Statement:

I hereby expressly declare that, according to the article 8 of Law 1559/1986, this dissertation is solely the product of my personal work, does not infringe any intellectual property, personality and personal data rights of third parties, does not contain works/contributions from third parties for which the permission of the authors/beneficiaries is required, is not the product of partial or total plagiarism, and that the sources used are limited to the literature references alone and meet the rules of scientific citations.