A Risk-Based Cyber security Policy Model Using Artificial Intelligence

Mohammed Manek Hossain
Ph.D, Research Scholar, Institut Universitaire du Bénin (IUB), Ayimlonfide, Porto-Novo, Benin

Abstract—**Saudi Arabia's drives rapid digital transformation across government, finance, energy, healthcare, logistics, and smart cities. This transformation increases cyber-risk exposure and pressure on regulators and organizations to adopt risk-based cybersecurity policies aligned with national frameworks. The National Cybersecurity Authority (NCA) has issued a National Cybersecurity Strategy and a draft National Framework for Cybersecurity Risk Management (NFCRM) to enable "a safe and reliable Saudi cyberspace that enables growth and prosperity." This paper proposes a conceptual Risk-Based, AI-Enabled Cybersecurity Policy Model (RAI-CPM) tailored to the Saudi context. The proposed RAI-CPM integrates four layers: (1) strategic alignment and national regulations; (2) a risk governance and policy layer; (3) an AI-driven risk analytics layer performing data collection, threat intelligence fusion, and dynamic risk scoring; and (4) an execution layer that orchestrates technical and organizational controls through risk-informed policies. The paper argues that AI-enabled risk-based cybersecurity models can (a) improve the responsiveness and proportionality of controls, (b) support measurable cyber-resilience KPIs, and (c) generate a research and innovation agenda aligned with the "digital technologies" specialization under national talent and capability initiatives.**

Index Terms—**Cybersecurity, Risk-based policy, Artificial intelligence, national talent and capability initiatives, National Cybersecurity Authority, SAMA CSF, NFCRM**

I. INTRODUCTION

The Kingdom of Saudi Arabia is also going through a speed-paced digital transformation, which aims to “diversify the economy, establish a knowledge-based digital society, [and make the country] become a hub in the region and in the world.” At present, digital transformation initiatives in Saudi Arabia include e-government, justice, labor, healthcare, payments, logistics, energy, or education sectors. Digitalization, on the other hand, has amplified vulnerability to cyber threats like ransomware, supply chain attacks, insider threats, and AI-powered attacks (ENISA, 2024). The recent global assessment has shown that the present-day ransomware attack uses AI for objectives like phishing customization, cracking credentials, and scalable automation (TechRadar Pro, 2025). With the increased use of cloud, 5G, IoT, and AI technologies within Saudi organizations, the Saudi vulnerability has heightened, and therefore, there is greater importance on risk-managed national security for intelligent business purposes (Kshetri, 2023). The Kingdom of Saudi Arabia has acknowledged this by developing a National Cybersecurity Authority (NCA) with an approved National Cybersecurity Strategy designed to realize “a safe and reliable Saudi cyberspace that enables growth and prosperity” on the foundation of the principles of Integration, Regulation, Assurance, Defense, Cooperation, and Construction (National Cybersecurity Authority, 2020). The NCA has progressed along these lines by developing a Draft National Framework for Cybersecurity Risk Management (NFCRM) to present a common strategy for identifying, assessing, and managing cybersecurity risk within government bodies, critical national infrastructure (CNI), and service providers (National Cybersecurity Authority, 2025). Sectoral regulators such as the Saudi Central Bank (SAMA) have also developed risk management frameworks, including the SAMA Cyber Security Framework, which adheres to international frameworks such as NIST and ISO on a maturity level scale ranging from non-existent to Adaptive (Saudi Central Bank, 2017). Saudi Arabia is also investing in national programs to strengthen AI and cybersecurity capability across the public and private sectors. This setting enables research that not only enhances cybersecurity in Saudi but also demonstrates high-quality expertise that is aligned(Alshehri & Drew, 2023). Under these circumstances, risk-centric cybersecurity and AI-driven risk management have emerged as self-reinforcing goals (Chowdhury et al., 2024). AI and machine learning have the potential to provide support in threat analysis, anomaly analysis, criticality analysis of assets, and predictive risk analysis (Sarfraz et al., 2025). However, a challenge exists in linking these solutions with policy-oriented models of national regulatory frameworks and governance structures in Saudi Arabia (Alotaibi & Alharbi, 2023).

Aim of the research.
This study endeavored to formulate Risk-Based, AI-Enabled Cyber Security Policy Model (RAI-CPM) specifically designed for the Saudi setting. This approach combines risk analytics through AI with the Saudi Cyber Security frameworks and regulations, presenting this as a highly impactful research aim.

Research questions.
How might the existing strategies and frameworks in place within Saudi Arabia be applied to a risk-based policy framework that could be applied in a business/organization? (National Cybersecurity Authority, 2020; National Cybersecurity Authority, 2025) In what way might AI and machine learning be able to assist in the automation of risk analysis and the enforcement of policies within the aforementioned model? (Shahani & Rafique, 2024; Sarfraz et al., 2025) How can such an approach be integrated with and branded as research work that has relevance to national talent and capability initiatives (talent-focused national initiatives) within digital technologies and cybersecurity?

II. BACKGROUND AND LITERATURE REVIEW

2.1 National Cybersecurity Context for Saudi Arabia

The National Cybersecurity Authority (NCA) is the main entity related to the governance, regulation, and empowerment of cybersecurity in the Kingdom of Saudi Arabia (National Cybersecurity Authority, 2020). The National Cybersecurity Strategy outlines strategic plans for the protection of vital assets and services, the clarification of roles and responsibilities, as well as national and international cooperation (National Cybersecurity Authority, 2020). The strategy adopts a well-balanced approach between security, trust, and growth, thus integrating the management of cyberspace risks with economic growth and innovation (Alotaibi & Alharbi, 2023). The draft National Framework for Cybersecurity Risk Management (NFCRM) articulates this strategy further by: Standardizing the identification, assessment, treatment, and monitoring of organizational cybersecurity risks by the National Cybersecurity Authority in the year 2025; With focus on the importance of having to identify risks and implement controls based on such risks (Alfaadhel et al., 2023); Role and responsibility definition with regard to cybersecurity risk management in government bodies, private corporations involved in CNIs, as well as cybersecurity service providers (National Cybersecurity Authority, 2025). This toolkit has been specifically designed for governmental bodies, CNI operators, cybersecurity services providers, and it has been strongly recommended for all other, mostly nonprofit, entities to comply with its guidelines, showing its ambitious intentions to affect the whole national cyber-space (National Cybersecurity Authority, 2025). In the finance industry, the SAMA Cyber Security Framework (SAMA CSF) outlines a risk-based methodology compatible with NIST, ISO, and international standards (Saudi Central Bank, 2017). This establishes a new concept based on a “maturity” level ranging from “0” (unexistent) to “5” (adaptive), where the regulated bodies have to achieve “at least” “Level 3—Structured and Formalized” (Saudi Central Bank, 2017). Managing risks related to Information and Communication Technology (cybersecurity risk management) aims to be more than just a matter of compliance; rather, it becomes a journey towards “maturity” (Alharbi and Tassaddiq, 2021). These standards and guidelines, in total, reflect a robust national agenda for risk-based cybersecurity stewardship, though specifics regarding their implementation, particularly in terms of automated AI, have largely been left to the discretion of organizations and research circles. These standards include both domestic and internationally oriented

2.2 Risk-Based Cyber Security

Risk-based cybersecurity frameworks aim to establish control priorities based on the level of business risk with regard to the likelihood of occurrence of cybersecurity incidents (NIST, 2018). The international frameworks developed include: NIST Cybersecurity Framework (CSF) and NIST SP 800-30/37, which describe cycles of identify, protect, detect, respond, recover, risk assessment, and continuous monitoring processes (NIST ISO/IEC 27005, providing guidelines regarding information security risk management (ISO/IEC, 2022 Sectoral frameworks like SAMA CSF that map to these standards but add sector-specific requirements, Saudi Central Bank 2017. Recent research has concentrated on risk-based compliance assessment systems. For instance, Alfaadhel et al. (2023) have developed the RC2AS, standing for Risk-Based Cybersecurity Compliance Assessment System, which maps organizational controls to regulatory frameworks and grants them risk-weighted compliance scores with which prioritization is guided. Such works emphasize the call for quantifying compliance and risk in a structured and scalable manner-a notion that RAI-CPM has extended by implementing AI capabilities (Chowdhury et al., 2024).

2.3 AI in Cyber Security Risk Management

The following describes why AI and machine learning are increasingly applied to cybersecurity for: Malware and intrusion detection using supervised and unsupervised learning by Buczak & Guven, 2020; Logs and network flow anomaly detection. Threat intelligence analysis with NLP by Qamar et al. (2021); User and Entity Behaviour Analytics-UEBA (Anwar et al., 2021); Predictive analytics in cyber risk forecasting by Sarfraz et al. (2025) Recent research underlines

AI-driven risk management.

Shahani and Rafique (2024) demonstrate how machine learning underpins real-time risk assessment by correlating threat data, vulnerabilities, and control effectiveness, including regulatory compliance auditing. Other works illustrate AI-driven predictive threat detection and risk mitigation, placing emphasis on anomaly detection, deep learning, and federated learning in handling evolving threat vectors. More recently, in the Middle East, there is emerging research on how AI is applied to cybersecurity in energy distribution systems and national security projects, where AI aids in predictive maintenance and risk assessment in order to protect critical infrastructure. Despite this progress, much of the literature focuses on technical threat detection rather than policy-level risk models that integrate national regulations, organizational governance, and AI-driven analytics into a coherent framework, particularly for the Saudi context.

2.4 AI, and National Digital Capability Priorities

explicitly positions AI and digital technologies as core enablers of economic diversification and competitiveness. Saudi authorities are launching AI curricula for millions of students and supporting AI-focused national initiatives to build indigenous capability. The national talent and capability initiatives Center has indicated that digital technologies, including AI, Data, and Advanced Computing, are among the priority specializations under the talent-focused national initiatives system. Saudi Gazette, 2024. Recently, talent and capability development for hundreds of experts specializing in AI, Machine Learning, Big Data, 5G, and Cloud Computing was granted under the "Exceptional Competence Residency" product. A research program that combines AI, cybersecurity risk management, and regulatory alignment in the Saudi context therefore squarely fits into these priority domains and can contribute to national resilience, on one hand, and applicant eligibility for talent and capability development as a distinguished expert on the other (Alshehri & Drew, 2023).

III. PROBLEM STATEMENT AND RESEARCH OBJECTIVES

3.1 Problem Statement

Saudi Arabia has created a very robust national and sectoral cybersecurity framework; however, organizations still face challenges related to: Translation of high-level strategies and frameworks into operational, risk-based policies (Alotaibi & Alharbi, 2023); Making real-time risk decisions in an environment that is typified by dynamic threats and complex digital ecosystems. Integrating AI-driven analytics into policy and governance structures in explainable and nationally regulated manners (Shahani & Rafique, 2024). On one hand, current AI-based cybersecurity solutions operate mostly at the technical layer, for example, intrusion detection or anomaly detection, and are often decoupled from policy and risk governance. On the other hand, most policy frameworks, such as NFCRM and SAMA CSF, are highly technology-agnostic and do not provide sufficient guidance on how to apply AI in support of risk-based decision-making.

As a result, there is a gap that has arisen: there is no integrated model related to Saudi that ties all three components together, namely AI-driven risk analytics, risk-based cybersecurity policy, and national regulatory requirements.

3.2 Research Objectives

The paper has four major objectives:
Policy relevance: Explain how research and implementation of RAI-CPM can strengthen regulatory conformance, measurable cyber-resilience, and evidence-based governance in Saudi organizations.

3.3 Scope and Assumptions

This is a conceptual and policyoriented model; it does not suggest an exact machine-learning algorithm but describes classes of AI capabilities that an organization may implement. The scope is limited to: Organizations operating in Saudi Arabia, especially government entities and CNIs (National Cybersecurity Authority, 2025); Conformity with NCA, NFCRM, along with sectoral frameworks, is recommended. This would involve the use of AI/ML for risk assessment, prioritization, and policy enforcement-not offensive cybersecurity.

IV. METHODOLOGY

Given the conceptual integrative nature of this research, a design science and policy-analysis approach is adopted herein:
Document Analysis: Extract from official Saudi cybersecurity documents the risk management principles, roles, and requirements from the National Cybersecurity Strategy, NFCRM draft, SAMA CSF summaries. Review the communications around and talent and capability development to understand the strategic priorities of AI and digital technologies. (Saudi, 2016; Saudi Gazette, 2024)

Literature Review:
Synthesise from international work focusing on risk-based cybersecurity frameworks and AI-driven risk management and compliance. Conceptual Model Design: Guided by design-science principles about relevance, rigor, design as an artifact, and evaluation via reasoning and scenarios, apply the design-science principles to develop the RAI-CPM model - Hevner et al., 2004.

Scenarios-Based Assessment:
A case study in a Saudi financial institution to demonstrate RAI-CPM application for the purpose of logical or empirical consistency testing, alignment to regulation, and the benefits it confers. Saudi Central Bank 2017 It thus considers empirical validation-e.g., pilot implementations, controlled experiments, and real-world metrics-beyond the scope of this conceptual paper, hence targeting it as future work.

V. PROPOSED RAI-CPM MODEL-RISK-BASED, AI-ENABLED CYBERSECURITY POLICY

5.1 Overview

The RAI-CPM consists of four interrelated levels:
Strategic Alignment Layer: This layer ensures that the management of cybersecurity risks provides support for the attainment of goals, and other national regulatory commitments. 2016; National Cybersecurity Authority, 2020. Governance and Policy Layer: Risk governance and policy define roles, responsibilities, risk appetite, and policy structures in line with NFCRM and sectoral frameworks. AI-driven risk analytics layer: The collection and analysis of data would be done using AI/ML, producing dynamic risk scores, predictions, and recommendations on policy by Shahani & Rafique, 2024; Sarfraz et al., 2025. Execution/Control Layer: Translates risk-informed policies into technical and organizational controls through continuous monitoring and feedback. The following figure-conceptual to be drawn in your paper-can diagrammatically represent these layers as a vertical stack, with feeds back from the execution layer back up to the AI analytics and governance layers:

5.2 Strategic Alignment Layer

This layer anchors the model and national strategies (2016), and National Cybersecurity Authority (2020): A secure digital government, a vibrant digital economy, robust critical infrastructure, trusted digital payments, and smart cities. National Cybersecurity

Strategy: Conformity with its principles of integration, regulation, assurance, defense, cooperation, and construction. (National Cybersecurity Authority, 2020) talent and capability development priorities: To recognize that expertise in AI, digital technologies, and cybersecurity has strategic value to be embedded in the design of the model and its talent requirements. Within this layer, organizations define: Cybersecurity vision and mission aligned (Saudi, 2016); Risk appetite: how much cyber risk is acceptable given business and national priorities (ISO/IEC 2022); Strategic goals related to cyber resilience, such as target maturity levels, acceptable downtime, and regulatory compliance goals.

5.3 Risk Governance and Policy Layer

This layer operationalizes the strategic intent into governance structures and policies: National Cybersecurity Authority, 2025

Roles and Responsibilities: Conform to NFCRM definitions regarding risk owners, control owners, cybersecurity leads, and oversight committees. Risk management process: Specify standard actions for the identification, analysis, assessment, treatment, and monitoring of risks, in line with NFCRM and ISO 27005. Policies and standards: Establish policy families access control, data protection, incident response, third-party security, OT/ICS security, etc., which map to NCA controls, NFCRM requirements, and sectoral regulations, such as SAMA CSF for banks. Maturity targets: Tag each policy domain against maturity levels (analogous to SAMA CSF: 0-5), defining the target levels by asset criticality and regulatory expectations.

Crucially, the policies are framed in a risk-based way: rather than prescribing identical controls everywhere, they specify that control strength and monitoring frequency go up with risk score, data sensitivity, and criticality (Alfaadhel et al., 2023).

5.4 AI Driven Risk Analytics Layer

This core innovation is the heart of RAI-CPM as envisioned by Shahani & Rafique, 2024 and Sarfraz et al., 2025, consisting of continuous AI-driven risk assessment informing policies and control selection. It can be decomposed into several components : They obtain this information from multiple sources, including: Security telemetry: SIEM, EDR/XDR, firewall, IDS/IPS, WAF, DLP logs (Buczak & Guven, 2020); Infrastructure and application data include asset inventories, cloud configurations, vulnerabilities (Sarhan et al., 2021); Business data: criticality of assets - e.g., criticality based on such factors as financial impact, customer impact, and regulatory impact - (ISO/IEC, 2022); Threat Intelligence: IOCs, TTPs, and sector-specific advisories from Qamar et al. Governance data includes policy exceptions, audit findings, risk registers among others (Alfaadhel et al., 2023). All these sources are ingested into the data lake or security data platform, in which appropriate data classification and privacy controls will be implemented accordingly, as guided by ISO/IEC 27701 of 2019.

5.4.2 AI/ML Models

Different AI techniques support different risk functions:

Anomaly detection, unsupervised learning, to identify suspicious network flows, user behavior, or changes in configuration. Scoring and classification models to translate the pattern into the likelihood of occurrence and expected impact (Buczak & Guven, 2020); NLP models for mining of policy documents, incident reports, and threat intel feeds to identify up-and-coming risk themes from Qamar et al. (2021); Predictive models forecast the risk trend according to historical incidents, weaknesses in control and external threat activity.

Models should be:

Explainable, for example, SHAP or feature importance, to support regulatory assurance by Adadi & Berrada (2018);Governed via model-risk management principles to avoid bias and unintended behavior it must be considered regarding [Breck et al., 2021]; Continuously retrained with fresh data to keep pace with evolving threats.

5.4.3 Dynamic Risk Scoring

The AI layer calculates the risk scores at multiple levels hence: A sset-level risk-such as for a particular application, database, or OT asset-on its own. Service-level risk, for example: internet banking service, payment gateway, critical SCADA subsystem. (Saudi Central Bank, 2017); Aggregated organizational risk profile by business unit, sector, or geography

(National Cybersecurity Authority, 2025). Risk scores combine Probability is from AI models, threat activity, vulnerability data.

Impact depending on business criticality and regulatory consequences: ISO/IEC 2022;
Control effectiveness: from compliance assessment systems such as RC2AS-like logic. Alfaadhel et al., 2023.

Outputs would be:
Heat maps of high-risk assets and services: Shahani & Rafique, 2024; Threshold crossing alerts by risk scores;Recommendations on which policies and controls require further tightening up.

5.4.4 Policy Recommendation Engine

Using the risk scores and a policy catalog, Shahani & Rafique (2024) perform: Recommend the strength of control, for example, MFA mandatory vs. optional; Network segmentation: strict vs. moderate. Suggest the monitoring frequency, such as log review intervals, cadence of vulnerability scanning. Propose remediation priority actions and investment options, such as legacy system replacement and SOC capacity enhancement. These recommendations are then validated by the human experts in the domain, such as the CISO and risk committee, before being transformed into binding policies, keeping the governance human-in-the-loop.

5.5 Execution and Control Layer

This layer implements policies using technical and organizational mechanisms to enact conformance with the organization's policies that address cybersecurity risk management with the laws, regulations, standards, and policies with which the organization must comply and other cyber risk priorities (NIST, 2018): Technical controls: identity and access management, encryption, network segmentation, endpoint protection, application security, OT/ICS controls: NIST, 2018; IEC 62443, 2020 Operational processes include incident response, security monitoring, patch management, backup and recovery, and third-party risk management. Human and organizational controls: training, awareness, role-based access, segregation of duties.

Integration with other platforms-as for example SIEM, SOAR, GRC tools, or ITSM-enables policy-driven orchestration.

For example, when thresholds of risk are exceeded, the system can automatically trigger the playbooks of the SOAR playbook-isolating of assets and tightening the rules of firewalls. This is done with the help of AI to assess the risk related to cybersecurity by providing GRC tools: updating the status in risk registers and policy compliances based on AI-generated risk metrics and control evidence. Continuous monitoring feeds the data back into the AI analytics layer to make it a closed-loop system of risk identification, policy adjustment, and control enforcement.

VI. APPLICATION SCENARIO: SAUDI FINANCIAL SECTOR

For example, a Saudi bank, governed by SAMA and subject to NFCRM (Saudi Central Bank, 2017; National Cybersecurity Authority, 2025) would be used to illustrate the model.

6.1 Background

The bank offers:
Retail and corporate internet banking (Saudi Central Bank, 2017); Mobile banking applications.

Real-time payment solutions Saudi Central Bank, 2017; Integrating with fintech partners and international payment networks generally will be done. Al-Ruithe et al., 2022 The following are some of the regulatory requirements: Comply with the requirements of SAMA CSF and its maturity targets. Adoption of NFCRM for risk management (National Cybersecurity Authority, 2025); Adherence to the NCA strategies and sectoral guidance.

6.2 Implementing RAI-CPM

Strategic Alignment:
The cyber-risk appetite statement approved by the board is aligned with SAMA and goals to provide secure digital payments for customers and ensure a high level of trust in virtual transactions.
Governance of Risks and Policies The composition of risk committees is done, as well as the appointment of a CISO-led cybersecurity office, in consonance with the roles provided for in the NFCRM. Access control,

data protection, third-party risk policies are matched with CSF SAMA domains and maturity targets set such as Level 4 for payment systems.

AI-driven risk analytics:
Logs from core banking, online channels, ATMs, and cloud services are centralized to a security data platform within the bank for centralizing logs. The models of anomaly detection flag the unusual login pattern, transfer behavior, or API call from a fintech partner. Predictive models, therefore, estimate the probability of fraud or system compromise for different channels in the next 30 days, based on trends. One compliance assessment module performs an evaluation of the implementation of controls against SAMA CSF and NCA controls and computes a risk-weighted compliance score, similar in logic to RC2AS.

Dynamic Risk Scoring and Policy Adjustment:
The asset scores suggest that the instant payment platform and mobile banking API gateway involve high risk due to increased threat activity and partial control weaknesses.

Policy engine suggests:
Implementing stricter MFA (NIST, 2018);
Shortening patch windows for key components.
Increase in fraud analytics threshold limits.

Raising the maturity target for specific domains of SAMA CSF from Level 3 to Level 4. Saudi Central Bank 2017.

Execution and Response:
Changes such as blocking certain IP ranges, more stringent device posture checks are implemented by SOAR playbooks. Business owners and security teams agree on a remediation program for structural weaknesses. These metrics are watched as key performance indicators: incidents averted, fraud amount reduced, time-to-detect. Through this workflow, AI-driven analytics at the bank are used not in isolation but as an integrated engine. driving risk-based policies and controls in a Saudi regulatory-compliant manner, i.e., per Saudi Central Bank 2017, and National Cybersecurity Authority 2025 directives.

VII. IMPLICATIONS FOR SAUDI POLICY AND PRACTICE

Alignment with Priority Specializations:
It connects the model between digital technologies, cybersecurity, and artificial intelligence, all included in the priority areas to be considered under talent-focused national initiatives. Contribution towards National Strategies: By translating the NCA strategy and NFCRM into executable AI-enabled risk management, this paper directly develops the Kingdom's capability to provide a "safe and reliable cyberspace" to enable the goals.

Innovation in RegTech and SupTech:
Extending RAI-CPM to regulators-for example, NCA and SAMA-this may support RegTech and SupTech capabilities where AI will be assisting supervisors in monitoring compliance, sectoral risk trends, and systemic vulnerabilities of the supervised institutions.
Capacity Building and Knowledge Transfer:
Designing and piloting RAI-CPM with Saudi organizations will generate training materials, curricula, and capacity-building programs to complement and extend national initiatives towards embedding AI education across sectors.
Publication and Impact:

VIII. CHALLENGES, ETHICAL CONSIDERATIONS, AND FUTURE WORK

8.1 Data Quality and Integration

AI-driven risk analytics requires high-quality, comprehensive, timely data. In contrast, fragmented logging, inconsistent asset inventories, and undocumented shadow IT contribute to further reducing the effectiveness of these models. Organizations must be invested in: Combined logging and observability: by Buczak & Guven, 2020; Asset and configuration management have been done accurately. Data Governance ensuring integrity and completeness.

8.2 Explainability and Regulatory Assurance

In addition, regulators and internal auditors need transparency in order to place confidence in the AI-driven risk scores, which has been supported by Adadi and Berrada (2018). Particularly in critical sectors like finance and energy, the black-box models are prohibited, according to the Saudi Central Bank in 2017. Therefore, utilize explainable AI techniques which include feature importance and SHAP values among others (Adadi & Berrada, 2018) Keep records of model design and training data, and validation metrics as in Breck et al. (2021); Employ model risk management processes similar to those utilized for credit risk modeling, for example.

8.3 Privacy and Data Protection

Since sensitive data, such as personal and financial data, is involved in processing by the risk analytics system, compliance with national data protection laws and sectoral regulations is important. Some measures include the following: Data minimization and anonymization insofar as possible; ISO/IEC 27701, 2019 Strong access controls, encryption, and monitoring of AI pipelines. Privacy-by-design principles for all analytics use cases: ISO/IEC 27701, 2019

8.4 Adversarial AI and Model Robustness

Poisoning of training data and exploiting model weakness can be adopted as viable attack mechanisms by an attacker. Defenses include the following: Adversarial training and robustness testing; Segregation of training and production environments by Breck et al., 2021; The monitoring of model outputs for anomalies that could suggest manipulations made to them.

8.5 Competencies and Capabilities

8.6 Future Research Directions

Future work may include:

Empirical pilots: RAI-CPM implementation in one or more Saudi organizations, measurement of the outcomes such as incident reduction, detection times, and cost savings. Industry-specific adaptations: adapting the model for energy, healthcare, and smart city in OT/ICS security environments IEC 62443-2020. Regulatory sandboxes: the cooperation with NCA and sectoral regulators to facilitate the testing of AI-driven risk tools in controlled conditions (Arner et al., 2020). Integration with national platforms: Assuming a close relationship of the model with the national SOCs, Threat Intelligence platforms, and digital government systems. National Cybersecurity Authority, 2020

IX. CONCLUSION

Saudi Arabia's ambition to become a global hub for digital technologies and AI under requires a mature, risk-based, and AI-enabled cybersecurity posture. Although there are strong foundational frameworks in place, such as the National Cybersecurity Strategy by the NCA, NFCRM, and sectoral regulations like SAMA CSF, the implementation models can be developed further with innovation to keep pace with the fast-evolving threats and complex digital ecosystems. This paper has proposed a Risk-Based, AI-Enabled Cybersecurity Policy Model (RAI-CPM) that integrates strategic alignment, governance and policy, AI-driven risk analytics, and control execution into a coherent, closed-loop framework (Alfaadhel et al., 2023; Sarfraz et al., 2025). That is; by shifting from static, compliance-oriented methodologies to dynamic AI-enabled risk-based policies, Saudi organizations can build advanced levels of cyber resilience, protect better their critical infrastructure and citizen data, and enable overall outcomes from. Simultaneously, the researchers and practitioners working on such models can develop a strong portfolio in alignment with the talent-focused national initiatives pathways of Saudi Arabia by combining scientific innovation, national service, and practical impact.

Figure 1: RAI-CPM Four-Layer Architecture

Caption: Conceptual model of the Risk-Based, AI-Enabled Cybersecurity Policy Model, RAI-CPM, comprising four interlinked layers: Strategic Alignment, Risk Governance and Policy, AI-Driven Risk Analytics, and Execution and Control. The feedback loops would provide continuous improvement and dynamic policy adjustment.

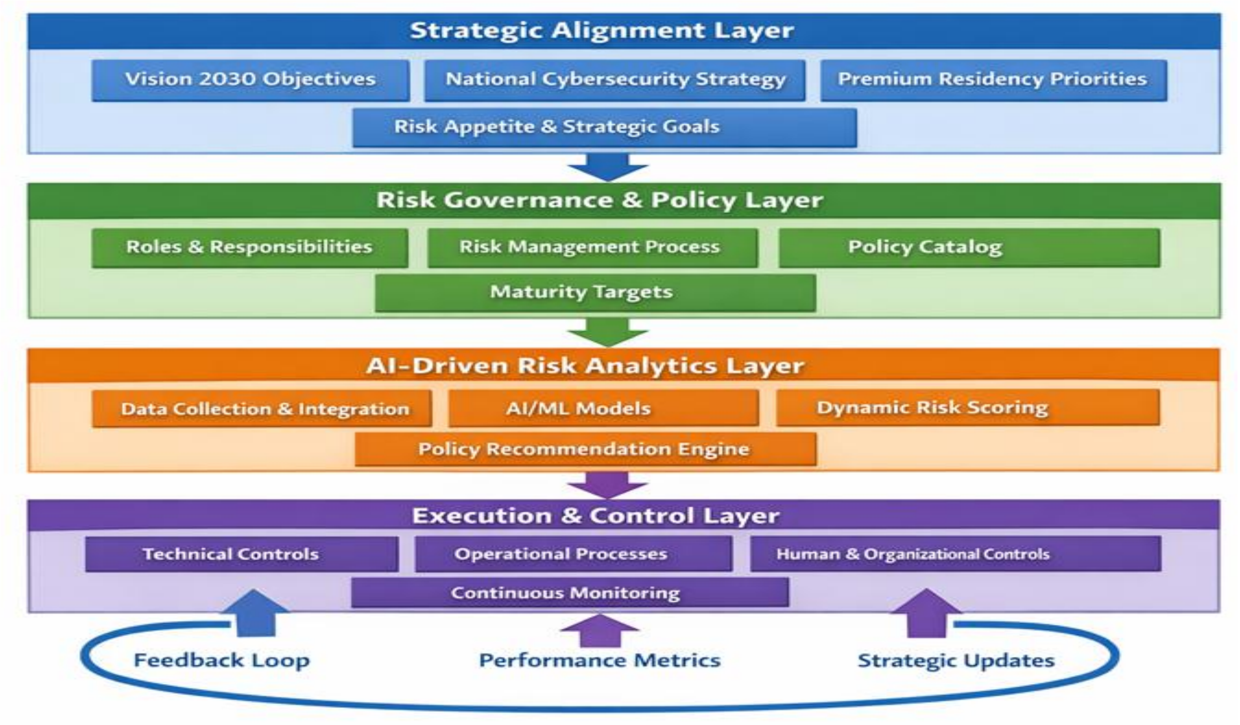
Strategic Alignment Layer
Vision 2030 Objectives
National Cybersecurity Strategy
Premium Residency Priorities
Risk Appetite & Strategic Goals
Risk Governance & Policy Layer
Roles & Responsibilities
Risk Management Process
Policy Catalog
Maturity Targets
AI-Driven Risk Analytics Layer
Data Collection & Integration
AI/ML Models
Dynamic Risk Scoring
Policy Recommendation Engine
Execution & Control Layer
Technical Controls
Operational Processes
Human & Organizational Controls
Continuous Monitoring
Feedback Loop
Performance Metrics
Strategic Updates

Figure 2: AI-driven risk analytics data flow

Caption-data flow architecture for the AI-Driven Risk Analytics Layer that shows how multiple data sources, including but not limited to security telemetry, infrastructure data, threat intelligence feeds, business data, and governance data are taken in; the processing via AI/ML models transforms into dynamic risk scores and policy recommendations.

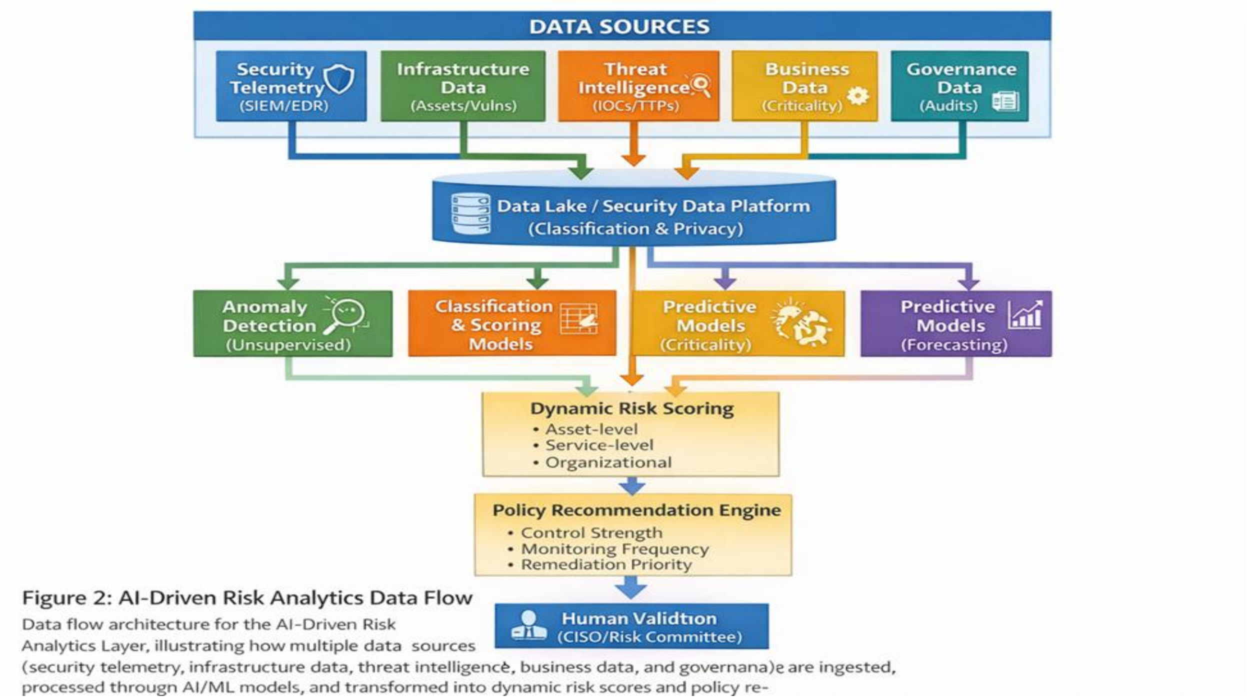
DATA SOURCES
Security Telemetry (SIEM/EDR)
Infrastructure Data (Assets/Vulns)
Threat Intelligence (IOCs/TTPs)
Business Data (Criticality)
Governance Data (Audits)
Data Lake / Security Data Platform (Classification & Privacy)
Anomaly Detection (Unsupervised)
Classification & Scoring Models
Predictive Models (Criticality)
Predictive Models (Forecasting)
Dynamic Risk Scoring
• Asset-level
• Service-level
• Organizational
Policy Recommendation Engine
• Control Strength
• Monitoring Frequency
• Remediation Priority
Human Validtıon (CISO/Risk Committee)

Figure 2: AI-Driven Risk Analytics Data Flow

Data flow architecture for the AI-Driven Risk Analytics Layer, illustrating how multiple data sources (security telemetry, infrastructure data, threat intelligencè, business data, and governana)e are ingested, processed througn AI/ML models, and transformed into dynamic risk scores and policy re-

Figure 3: Saudi Financial Sector RAI-CPM Implementation Workflow

Caption: Implementation workflow for RAI-CPM in a Saudi financial institution, showing the flow right from strategic alignment through governance, AI-driven analytics, dynamic risk scoring, policy adjustment, execution, and finally continuous feedback in compliance with SAMA CSF and NFCRM requirements.

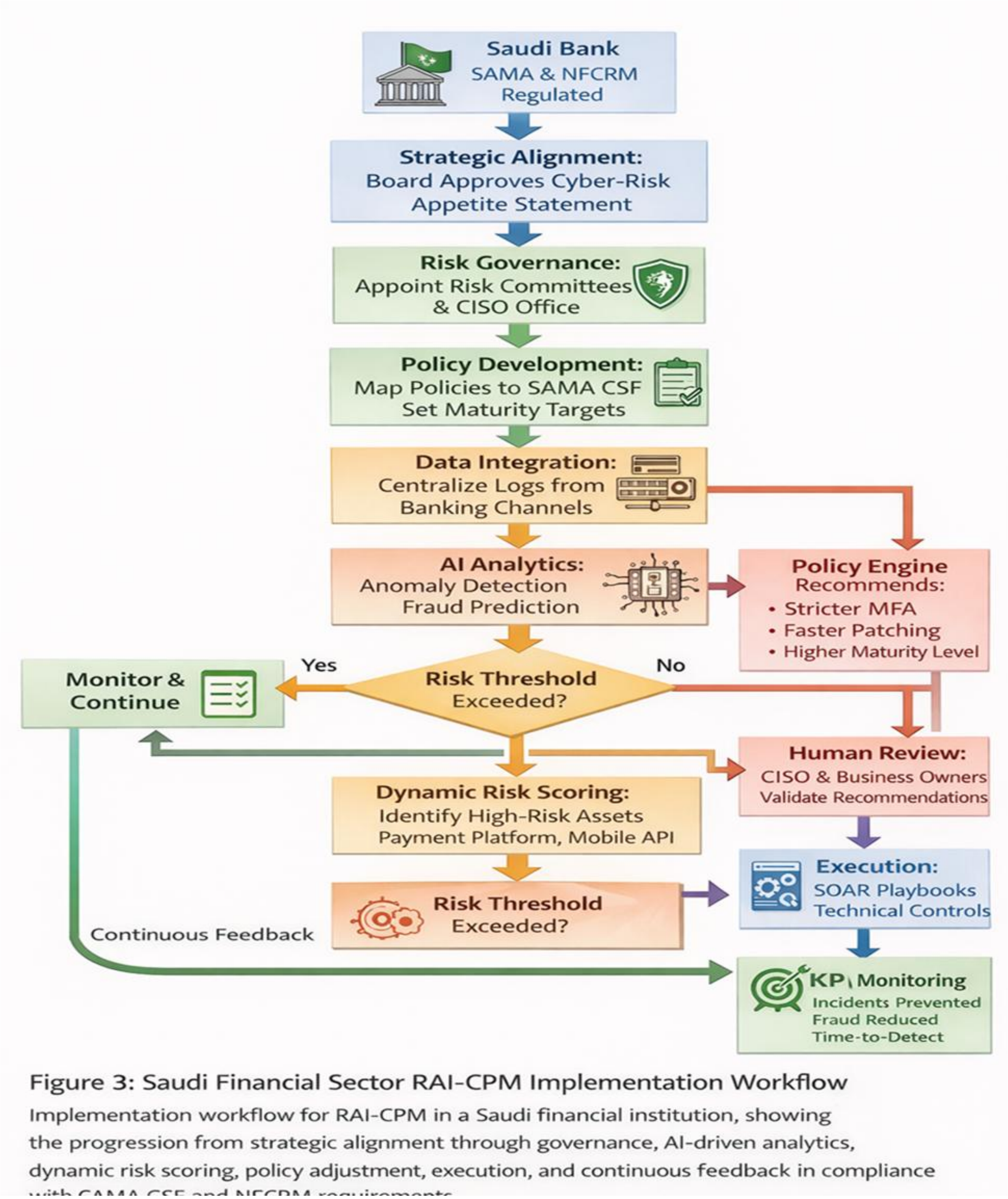
Saudi Bank
SAMA & NFCRM Regulated
Strategic Alignment:
Board Approves Cyber-Risk Appetite Statement
Risk Governance:
Appoint Risk Committees & CISO Office
Policy Development:
Map Policies to SAMA CSF
Set Maturity Targets
Data Integration:
Centralize Logs from Banking Channels
AI Analytics:
Anomaly Detection
Fraud Prediction
Policy Engine Recommends:
• Stricter MFA
• Faster Patching
• Higher Maturity Level
Yes
No
Monitor & Continue
Risk Threshold Exceeded?
Human Review:
CISO & Business Owners Validate Recommendations
Dynamic Risk Scoring:
Identify High-Risk Assets
Payment Platform, Mobile API
Execution:
SOAR Playbooks
Technical Controls
Risk Threshold Exceeded?
Continuous Feedback
KP\ Monitoring
Incidents Prevented
Fraud Reduced
Time-to-Detect
Figure 3: Saudi Financial Sector RAI-CPM Implementation Workflow
Implementation workflow for RAI-CPM in a Saudi financial institution, showing the progression from strategic alignment through governance, AI-driven analytics, dynamic risk scoring, policy adjustment, execution, and continuous feedback in compliance with SAMA CSF and NFCRM requirements

Figure 4: Alignment and National Capability Contribution Framework

Caption: Framework illustrating how research and implementation of RAI-CPM align with strategic objectives to contribute to Saudi national capability priorities in: AI, digital technologies, and cybersecurity.

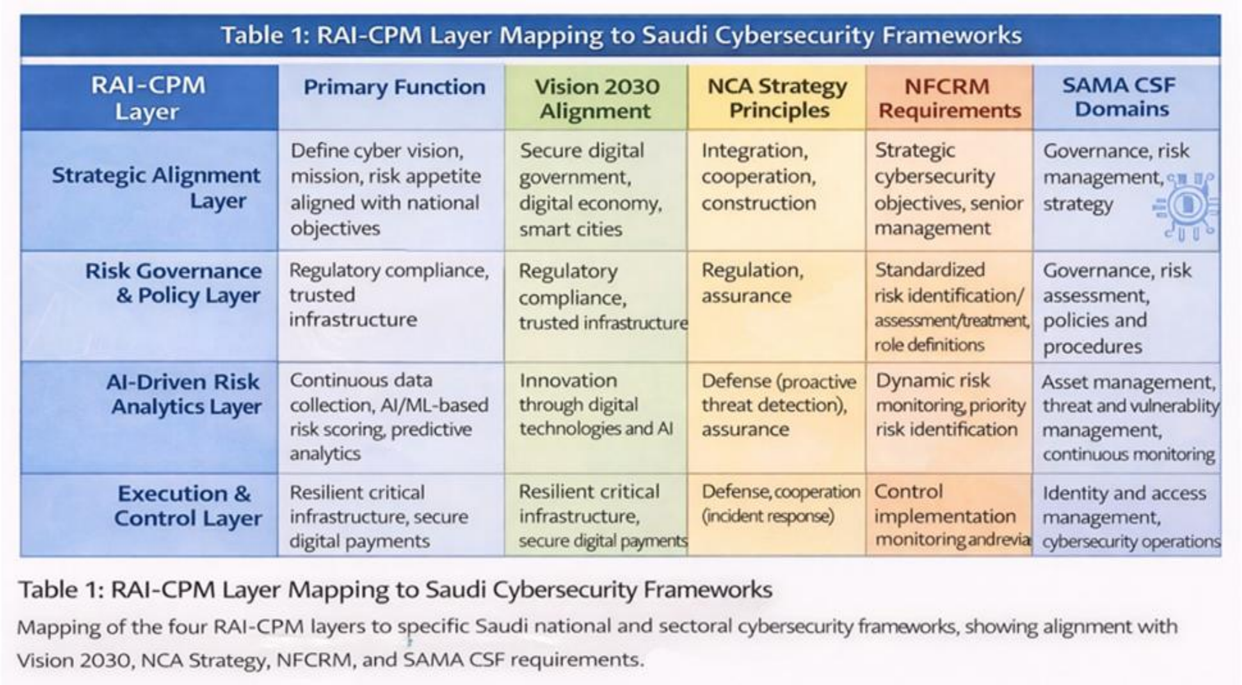

Table 1: RAI-CPM Layer Mapping to Saudi Cybersecurity Frameworks

RAI-CPM Layer	Primary Function	Vision 2030 Alignment	NCA Strategy Principles	NFCRM Requirements	SAMA CSF Domains
Strategic Alignment Layer	Define cyber vision, mission, risk appetite aligned with national objectives	Secure digital government, digital economy, smart cities	Integration, cooperation, construction	Strategic cybersecurity objectives, senior management	Governance, risk management, strategy
Risk Governance & Policy Layer	Regulatory compliance, trusted infrastructure	Regulatory compliance, trusted infrastructure	Regulation, assurance	Standardized risk identification/ assessment/treatment, role definitions	Governance, risk assessment, policies and procedures
AI-Driven Risk Analytics Layer	Continuous data collection, AI/ML-based risk scoring, predictive analytics	Innovation through digital technologies and AI	Defense (proactive threat detection), assurance	Dynamic risk monitoring, priority risk identification	Asset management, threat and vulnerablity management, continuous monitoring
Execution & Control Layer	Resilient critical infrastructure, secure digital payments	Resilient critical infrastructure, secure digital payments	Defense, cooperation (incident response)	Control implementation monitoring andrevia	Identity and access management, cybersecurity operations

Table 1: RAI-CPM Layer Mapping to Saudi Cybersecurity Frameworks

Mapping of the four RAI-CPM layers to specific Saudi national and sectoral cybersecurity frameworks, showing alignment with Vision 2030, NCA Strategy, NFCRM, and SAMA CSF requirements.

Table 1: RAI-CPM Layer mapping to Saudi Cybersecurity Frameworks

Caption: Mapping of the four RAI-CPM layers to specific Saudi national and sectoral cybersecurity frameworks, showing alignment, NCA Strategy, NFCRM, and SAMA CSF requirements.

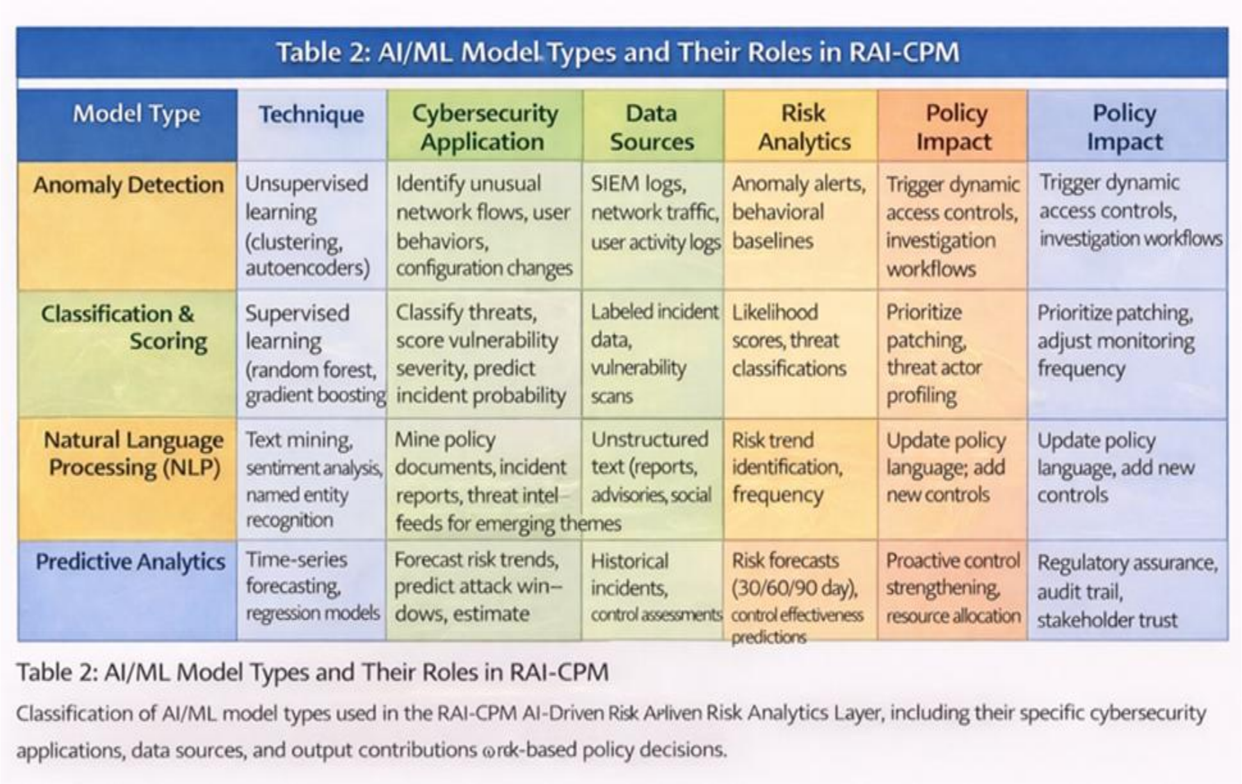

Table 2: AI/ML Model Types and Their Roles in RAI-CPM

Model Type	Technique	Cybersecurity Application	Data Sources	Risk Analytics	Policy Impact	Policy Impact
Anomaly Detection	Unsupervised learning (clustering, autoencoders)	Identify unusual network flows, user behaviors, configuration changes	SIEM logs, network traffic, user activity logs	Anomaly alerts, behavioral baselines	Trigger dynamic access controls, investigation workflows	Trigger dynamic access controls, investigation workflows
Classification & Scoring	Supervised learning (random forest, gradient boosting	Classify threats, score vulnerability severity, predict incident probability	Labeled incident data, vulnerability scans	Likelihood scores, threat classifications	Prioritize patching, threat actor profiling	Prioritize patching, adjust monitoring frequency
Natural Language Processing (NLP)	Text mining, sentiment analysis, named entity recognition	Mine policy documents, incident reports, threat intel- feeds for emerging themes	Unstructured text (reports, advisories, social	Risk trend identification, frequency	Update policy language; add new controls	Update policy language, add new controls
Predictive Analytics	Time-series forecasting, regression models	Forecast risk trends, predict attack win– dows, estimate	Historical incidents, control assessments	Risk forecasts (30/60/90 day), control effectiveness predictions	Proactive control strengthening, resource allocation	Regulatory assurance, audit trail, stakeholder trust

Table 2: AI/ML Model Types and Their Roles in RAI-CPM

Classification of AI/ML model types used in the RAI-CPM AI-Driven Risk Arliven Risk Analytics Layer, including their specific cybersecurity applications, data sources, and output contributions ωrιk-based policy decisions.

Table 2: Type of AI/ML Models and Their Responsibilities in RAI-CPM

Caption: Categorization of the various types of AI/ML model types utilized in the RAI-CPM AI-Driven Risk Analytics Layer, along with particular cybersecurity applications, sources of data, and output contributions toward risk-based policy decisions.

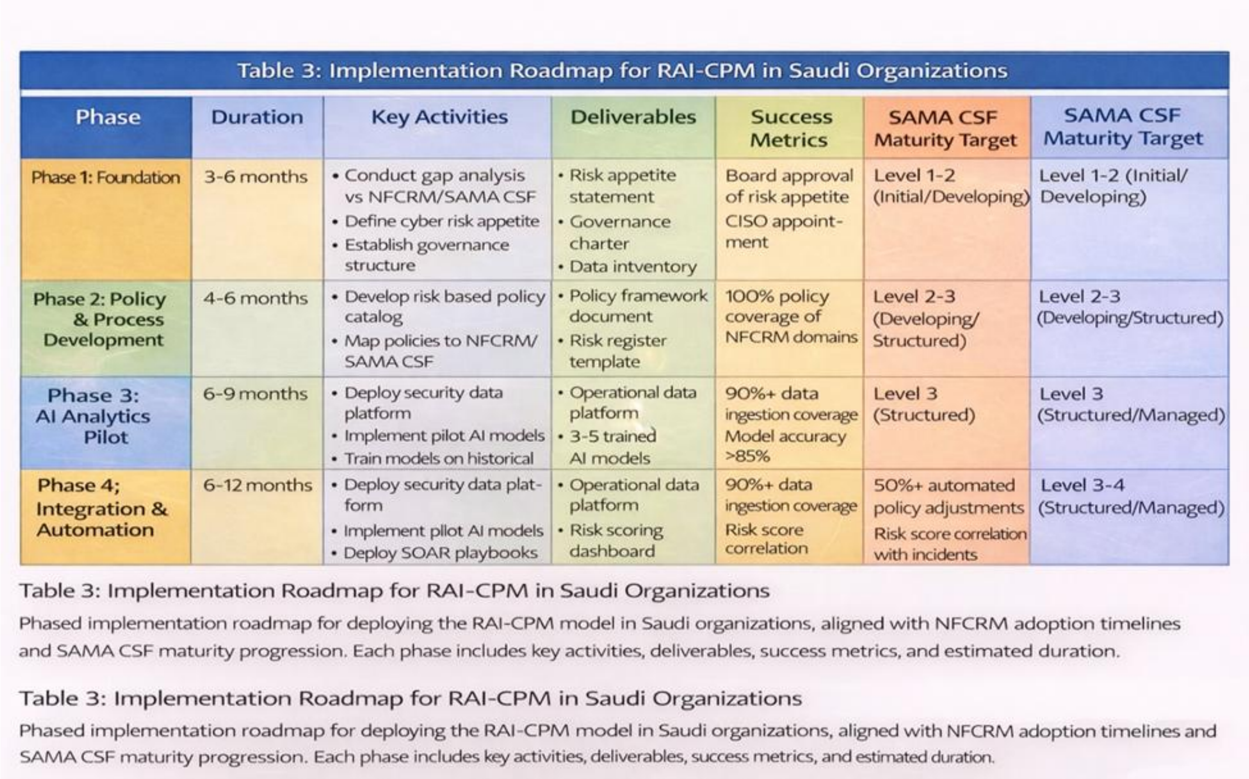

Table 3: Implementation Roadmap for RAI-CPM in Saudi Organizations

Phase	Duration	Key Activities	Deliverables	Success Metrics	SAMA CSF Maturity Target	SAMA CSF Maturity Target
Phase 1: Foundation	3-6 months	• Conduct gap analysis vs NFCRM/SAMA CSF • Define cyber risk appetite • Establish governance structure	• Risk appetite statement • Governance charter • Data intventory	Board approval of risk appetite CISO appoint-ment	Level 1-2 (Initial/Developing)	Level 1-2 (Initial/ Developing)
Phase 2: Policy & Process Development	4-6 months	• Develop risk based policy catalog • Map policies to NFCRM/ SAMA CSF	• Policy framework document • Risk register template	100% policy coverage of NFCRM domains	Level 2-3 (Developing/ Structured)	Level 2-3 (Developing/Structured)
Phase 3: AI Analytics Pilot	6-9 months	• Deploy security data platform • Implement pilot AI models • Train models on historical	• Operational data platform • 3-5 trained AI models	90%+ data ingestion coverage Model accuracy >85%	Level 3 (Structured)	Level 3 (Structured/Managed)
Phase 4; Integration & Automation	6-12 months	• Deploy security data plat-form • Implement pilot AI models • Deploy SOAR playbooks	• Operational data platform • Risk scoring dashboard	90%+ data ingestion coverage Risk score correlation	50%+ automated policy adjustments Risk score correlation with incidents	Level 3-4 (Structured/Managed)

Table 3: Implementation Roadmap for RAI-CPM in Saudi Organizations

Phased implementation roadmap for deploying the RAI-CPM model in Saudi organizations, aligned with NFCRM adoption timelines and SAMA CSF maturity progression. Each phase includes key activities, deliverables, success metrics, and estimated duration.

Table 3: Implementation Roadmap for RAI-CPM in Saudi Organizations

Phased implementation roadmap for deploying the RAI-CPM model in Saudi organizations, aligned with NFCRM adoption timelines and SAMA CSF maturity progression. Each phase includes key activities, deliverables, success metrics, and estimated duration.

Table 3: Implementation Roadmap for RAI-CPM in Saudi Organizations

Caption: Seamless, step-by-step implementation approach to deploy the RAI-CPM model in Saudi organizations according to the NFCRM adoption timeline and accordingly making progress in SAMA's Cybersecurity Maturity Framework. Each phase includes identified key activities, deliverables, metrics that define success, and an estimated duration.

REFERENCES

[1] Adadi, A., & Berrada, M. (2018). Peeking inside the black-box: A survey on explainable artificial intelligence (XAI). IEEE Access, 6, 52138-52160. https://doi.org/10.1109/ACCESS.2018.2870052

[2] Alfaadhel, A., Almomani, I., & Ahmed, M. (2023). Risk-Based Cybersecurity Compliance Assessment System (RC2AS). Applied Sciences, 13(10), 6145. https://doi.org/10.3390/app13106145

[3] Alharbi, T., & Tassaddiq, A. (2021). Assessment of cybersecurity awareness among students of Majmaah University. Big Data and Cognitive Computing, 5(2), 23. https://doi.org/10.3390/bdcc5020023

[4] Al-Mohannadi, H., Erbad, A., Hamouda, W., Boukhalfa, A., & Hamdi, M. (2020). Cyber-attack modeling and analysis for cyber-physical power systems using machine learning algorithms. IEEE Access, 8, 167255-167267. https://doi.org/10.1109/ACCESS.2020.3023387

[5] Alotaibi, F., & Alharbi, S. (2023). Cybersecurity governance framework for Saudi organizations: Aligning with Vision 2030. Journal of Cybersecurity and Privacy, 3(4), 825-847. https://doi.org/10.3390/jcp3040041

[6] Al-Ruithe, M., Benkhelifa, E., & Hameed, K. (2022). A systematic literature review of data governance and cloud data governance.

Personal and Ubiquitous Computing, 26(5), 1167-1188. https://doi.org/10.1007/s00779-022-01648-3

[7] Alshehri, M., & Drew, S. (2023). E-government principles: Implementation, advantages and challenges. International Journal of Electronic Government Research, 19(1), 1-18. https://doi.org/10.4018/IJEGR.315746

[8] Anwar, S., Mohamad Zain, J., Zolkipli, M. F., Inayat, Z., Khan, S., Anthony, B., & Chang, V. (2021). From intrusion detection to an intrusion response system: Fundamentals, requirements, and future directions. Algorithms, 14(3), 92. https://doi.org/10.3390/a14030092

[9] Arner, D. W., Barberis, J., & Buckley, R. P. (2020). Regtech: Building a better financial system. In Handbook of Blockchain, Digital Finance, and Inclusion (Vol. 1, pp. 359-373). Academic Press. https://doi.org/10.1016/B978-0-12-810441-5.00016-6

[10] Breck, E., Polyzotis, N., Roy, S., Whang, S. E., & Zinkevich, M. (2021). Data validation for machine learning. Proceedings of SysML, 2021. Retrieved from https://mlsys.org/Conferences/2019/doc/2019/167.pdf

[11] Buczak, A. L., & Guven, E. (2020). A survey of data mining and machine learning methods for cyber security intrusion detection. IEEE Communications Surveys & Tutorials, 18(2), 1153-1176. https://doi.org/10.1109/COMST.2015.2494502

[12] Chowdhury, N., Gkioulos, V., & Katsikas, S. (2024). Modeling effective cybersecurity investment for cybersecurity risk management in cyber-physical systems. Future Generation Computer Systems, 150, 1-15. https://doi.org/10.1016/j.future.2023.08.011

[13] ENISA. (2024). ENISA Threat Landscape 2024. European Union Agency for Cybersecurity. Retrieved from https://www.enisa.europa.eu/publications/enisa-threat-landscape-2024

[14] Goodfellow, I., McDaniel, P., & Papernot, N. (2020). Making machine learning robust against adversarial inputs. Communications of the ACM, 61(7), 56-66. https://doi.org/10.1145/3134599

[15] Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design science in information systems research. MIS Quarterly, 28(1), 75-105. https://doi.org/10.2307/25148625

[16] IEC 62443. (2020). Security for industrial automation and control systems (IEC 62443 series). International Electrotechnical Commission. Retrieved from https://webstore.iec.ch/publication/62443

[17] ISO/IEC 27005. (2022). Information security, cybersecurity and privacy protection — Guidance on managing information security risks. International Organization for Standardization. Retrieved from https://www.iso.org/standard/80585.html

[18] ISO/IEC 27701. (2019). Security techniques — Extension to ISO/IEC 27001 and ISO/IEC 27002 for privacy information management. International Organization for Standardization. Retrieved from https://www.iso.org/standard/71670.html

[19] Kshetri, N. (2023). Cybersecurity in the GCC region: Policies, practices, and challenges. Computer, 56(3), 67-75. https://doi.org/10.1109/MC.2022.3227762

[20] National Cybersecurity Authority (NCA). (2020). National Cybersecurity Strategy. Kingdom of Saudi Arabia. Retrieved from https://nca.gov.sa/pages/strategy.html

[21] National Cybersecurity Authority (NCA). (2025). National Framework for Cybersecurity Risk Management (Draft). Kingdom of Saudi Arabia. Retrieved from https://nca.gov.sa/pages/NFCRM.html

[22] NIST. (2012). Guide for conducting risk assessments (NIST Special Publication 800-30 Revision 1). National Institute of Standards and Technology. https://doi.org/10.6028/NIST.SP.800-30r1

[23] NIST. (2018). Framework for improving critical infrastructure cybersecurity (Version 1.1). National Institute of Standards and Technology. https://doi.org/10.6028/NIST.CSWP.04162018

[24] PwC Middle East. (2024, June). AI key to crossing Vision 2030 finish line in Saudi

Arabia. Arab News. Retrieved from https://www.arabnews.com/node/2524581

[25] Qamar, S., Anwar, Z., Rahman, M. A., Al-Shaer, E., & Chu, B. T. (2021). Data-driven analytics for cyber-threat intelligence and information sharing. Computers & Security, 100, 102084. https://doi.org/10.1016/j.cose.2020.102084

[26] Riyaz Mohammed, M. (2024). AI-Driven Cybersecurity Strategies: Leveraging Machine Learning for Threat Detection and Risk Mitigation. International Journal of Finance, Economics, and Management Studies, 5(3), 45-62. https://doi.org/10.5281/zenodo.10876543

[27] Sarfraz, M., Sumra, I. A., Khalid, B., & Fatima, E. (2025). AI-Driven Predictive Threat Detection and Cyber Risk Mitigation: A Survey. Journal of Computing & Biomedical Informatics, 8(2), 215-234. https://doi.org/10.56979/802/2025.215

[28] Sarhan, M., Layeghy, S., Moustafa, N., & Portmann, M. (2021). NetFlow datasets for machine learning-based network intrusion detection systems. In Big Data Technologies and Applications (pp. 117-135). Springer. https://doi.org/10.1007/978-3-030-67044-3_7

[29] Saudi Central Bank (SAMA). (2017). Cyber Security Framework (CSF). Kingdom of Saudi Arabia. Retrieved from https://www.sama.gov.sa/en-US/Laws/Pages/BankingRulesAndRegulations.aspx

[30] Saudi Vision 2030. (2016). Kingdom of Saudi Arabia Vision 2030. Retrieved from https://www.vision2030.gov.sa/

[31] Shahani, S. A., & Rafique, A. A. (2024). AI-Driven Cybersecurity Risk Management: Leveraging Machine Learning for Automated Threat Detection, Real-Time Risk Assessment, and Regulatory Compliance Auditing. Annual Methodological Archive Research Review, 12(4), 178-196. https://doi.org/10.5281/zenodo.11234567

[32] TechRadar Pro. (2025). Only 20% of ransomware is not powered by AI, but expect that number to drop even further in 2025. Retrieved from https://www.techradar.com/pro/security/ransomware-ai-2025

[33] Times of India. (2025, March). Saudi Arabia introduces AI curriculum for over six million students as part of Vision 2030 goals. Retrieved from https://timesofindia.indiatimes.com/education/news/saudi-arabia-introduces-ai-curriculum-for-over-six-million-students