ACIG
APPLIED CYBERSECURITY & INTERNET GOVERNANCE

NASK

# AI-Driven Threat Intelligence for National Cybersecurity Governance: A Framework for Adaptive Risk Detection and Policy-Aware Response

**Mohamed Mazloum Salem** | Computer Science, Faculty of Computer and Information Sciences, Egypt | ORCID: 0009-0001-1956-5824

## Abstract

The rapid evolution of cyber threats has exposed critical limitations in traditional, rule-based cybersecurity governance models. Nation-state infrastructures increasingly face adaptive, artificial intelligence (AI)-assisted attacks that outpace static defence mechanisms and policy frameworks. This paper proposes an AI-driven threat intelligence framework designed to enhance national cybersecurity governance through adaptive risk detection, contextual analysis, and policy-aware response orchestration. The framework integrates machine learning (ML)-based anomaly detection, graph-based threat correlation, and governance-aligned decision layers to bridge the gap between technical cybersecurity operations and regulatory oversight. Unlike conventional security information and event management systems, the proposed approach emphasises explainability, institutional accountability, and alignment with national digital governance objectives. The paper presents the conceptual architecture, operational workflow, and governance implications of the framework, demonstrating how AI can support strategic cyber resilience while preserving transparency and policy compliance. The findings contribute to applied cybersecurity research by offering a scalable, governance-centric model suitable for critical infrastructure protection and national cyber defence strategies.

**Received:** 19.12.2025

**Accepted:** 24.06.2026

**Published:** 23.08.2026

**Cite this article as:**
M.M. Salem, "AI-driven threat intelligence for national cybersecurity governance: A framework for adaptive risk detection and policy-aware response," ACIG, vol. 5, no. 1, 2026, doi: 10.60097/ACIG/225022

**Corresponding author:**
Mohamed Mazloum Salem, Computer Science, Faculty of Computer and Information Sciences, Egypt; E-mail: MohamedMazloum@std.mans.edu.eg
0009-0001-1956-5824

**Keywords**
*national security, artificial intelligence, cybersecurity governance, digital policy, threat intelligence*

## 1. Introduction

Cybersecurity has become a central pillar of national security and digital sovereignty. Governments increasingly depend on complex digital infrastructures to support public services, economic systems, and critical national functions. The digitisation of critical infrastructure sectors, including energy grids, financial systems, healthcare networks, and transportation systems, has created unprecedented dependencies on interconnected digital ecosystems. This transformation, while enabling greater efficiency and connectivity, has simultaneously exposed nations to sophisticated cyber threats that transcend traditional boundaries of physical security and territorial jurisdiction.

The contemporary threat landscape is characterised by an exponential increase in both frequency and sophistication of cyber attacks. Nation-state actors, criminal organisations, and hacktivist groups employ increasingly advanced techniques, including machine learning (ML)-powered attack automation, polymorphic malware, and multi-vector campaigns that coordinate across multiple attack surfaces. Advanced persistent threats (APTs) operate with extended dwell times, evading traditional signature-based detection systems. Ransomware campaigns have evolved from opportunistic attacks to sophisticated, targeted operations against critical infrastructure. The emergence of AI-assisted attack techniques represents a paradigm shift, enabling adversaries to automate reconnaissance, generate convincing phishing content, and adapt attack strategies in real time.

Conventional cybersecurity governance models, developed in an era of relatively predictable and manageable threats, rely heavily on predefined policies, manual threat analysis, and static compliance frameworks. While these approaches offer legal and institutional stability, they struggle to respond effectively to rapidly evolving cyber risks that may emerge within hours or days rather than months. The growing disconnects between technical cybersecurity operations (which occur at machine speed) and policy-level decision-making (which operates at institutional speed) creates dangerous blind spots that adversaries can exploit. Security operations centres (SOCs) generate thousands of alerts daily, but without intelligent prioritisation and contextualisation, critical threats may be obscured by noise.

Recent advances in artificial intelligence (AI) and machine learning offer unprecedented opportunities to address these challenges. Machine learning algorithms can process vast volumes of security telemetry data, identifying subtle patterns and anomalies that human analysts might miss. Data-driven analytics enable real-time threat detection, behavioural modelling of normal system operations, and predictive risk assessment based on historical attack patterns. Deep-learning techniques have shown promise in detecting zero-day exploits, while graph-based analytics can uncover coordinated attack campaigns that span multiple organisations or sectors.

However, the integration of AI into national cybersecurity governance faces significant challenges. Many existing AI-based cybersecurity solutions focus narrowly on technical detection accuracy, often neglecting critical governance requirements, such as transparency, accountability, regulatory compliance, and institutional oversight. Black box machine learning models, while effective at detection, may provide limited insight into their decision-making processes, making it difficult for policymakers and legal authorities to understand, validate, or challenge automated threat assessments. Furthermore, AI systems must operate within complex legal and ethical constraints, including privacy protection, data sovereignty, due process, and proportionality in response actions.

This paper argues that AI must be embedded within cybersecurity governance structures, rather than operating as a purely technical layer divorced from policy and institutional considerations. To this end, we propose an AI-driven threat intelligence framework explicitly designed for national cybersecurity governance. The framework integrates adaptive machine learning analytics with policy-aware decision mechanisms, ensuring that automated threat analysis aligns with legal constraints, institutional roles, and strategic decision processes. By bridging the gap between technical cybersecurity capabilities and governance requirements, the framework aims to enhance national cyber resilience while preserving transparency, accountability, and democratic oversight.

### 1.1. Research Questions

To address the identified gap between advanced AI-driven cybersecurity technologies and national governance requirements, this study investigates the following research questions:

RQ1: How can AI-driven threat intelligence improve national cybersecurity governance and strategic decision-making?

RQ2: How can governance requirements, including transparency, accountability, legal compliance, and institutional oversight, be embedded within AI-based threat intelligence architectures?

RQ3: How can policy-aware response mechanisms improve national cyber resilience against sophisticated and rapidly evolving cyber threats?

RQ4: What architectural components and governance mechanisms are required to ensure effective integration between AI-driven analytics and national cybersecurity governance frameworks?

The contributions of this paper are fourfold:

1. Identification and analysis of governance limitations in the existing AI-based cybersecurity systems, highlighting the disconnect between technical capabilities and institutional requirements.
2. Designing of a comprehensive, governance-aware AI threat intelligence framework that integrates adaptive analytics with policy constraints and institutional accountability mechanisms.
3. Detailed specification of the framework's architectural components, machine learning methodologies, and policy alignment mechanisms.
4. Analysis of policy, accountability, ethical, and implementation implications at the national level, including case studies and recommendations for deployment.

The remainder of this paper is structured as follows. Section 2 reviews related work in cybersecurity governance, threat intelligence, and AI-driven security systems. Section 3 examines the limitations of traditional governance models and the evolving role of threat intelligence. Section 4 presents the detailed architecture and design of the proposed framework. Section 5 describes machine learning components and methodologies. Section 6 discusses policy-aware response mechanisms and governance alignment. Section 7 presents implementation considerations and case studies. Section 8 provides a comprehensive discussion of implications and challenges. Section 9 concludes with future research directions.

## 2. Related Work and Literature Review

This section reviews the existing research and frameworks relevant to AI-driven threat intelligence, cybersecurity governance,

and their intersection. The review is organised into five subsections examining: (1) traditional threat intelligence systems and frameworks, (2) AI-driven security analytics, (3) governance frameworks and policy alignment, (4) integration challenges and hybrid approaches, and (5) research gaps and opportunities. This comprehensive review establishes the theoretical foundation and identifies the research gap that this paper addresses.

### 2.1. Threat Intelligence Systems and Frameworks

Threat intelligence has evolved from manual collection and analysis processes to increasingly automated and data-driven approaches. Traditional threat intelligence platforms, such as security information and event management (SIEM) systems, aggregate logs and alerts from multiple sources but often require extensive manual configuration and tuning. The MITRE ATT&CK framework has become a standard taxonomy for describing adversary tactics and techniques, enabling more structured threat analysis and intelligence-sharing. Standards such as STIX/TAXII facilitate structured threat intelligence exchange across organisations. However, such frameworks primarily support manual analysis and lack the adaptive capabilities needed for rapidly evolving threats.

Intelligence-driven security operations centres (ISOCs) represent an organisational approach to integrating threat intelligence into security operations. These centres coordinate between technical security teams and strategic decision makers, but they often face challenges in scaling threat analysis and maintaining situational awareness across complex, distributed infrastructures. The disconnect between technical threat indicators and strategic intelligence requirements remains a persistent challenge.

### 2.2. AI-Driven Security Analytics

Machine learning has been extensively applied to cybersecurity problems, including intrusion detection, malware classification, and anomaly detection. Unsupervised learning techniques, such as clustering and autoencoders, have shown promise in detecting previously unknown threats and zero-day attacks. Supervised learning approaches, trained on labelled datasets of known attacks, achieve high accuracy but struggle with novel attack patterns. Ensemble methods and deep-learning architectures have improved detection rates for complex, multistage attacks.

Graph-based analytics have emerged as a powerful approach for threat correlation and attack campaign detection. By modelling systems, users, and network flows as graphs, these methods can identify patterns of coordinated activity that might be invisible when examining individual events in isolation. Graph neural networks (GNNs) extend this capability by learning complex relationships and temporal patterns from network data.

Explainable AI (XAI) techniques are increasingly important for security applications, particularly in governance contexts where decision transparency is required. Methods such as Local Interpretable Model Agnostic Explanations (LIME) and SHapley Additive exPlanations (SHAP) provide *post hoc* explanations for model predictions, while inherently interpretable models like decision trees offer built-in transparency. However, balancing explainability with model performance remains a challenge, particularly for complex deep-learning architectures.

### 2.3. Governance Frameworks and Policy Alignment

Cybersecurity governance research has explored various institutional arrangements, policy frameworks, and regulatory approaches. The National Institute of Standards and Technology's (NIST) cybersecurity framework (CSF) provides a structured approach to managing cybersecurity risks but focuses primarily on organisational rather than national-level governance. European frameworks, such as the NIS directive and its successor, NIS2, establish requirements for critical infrastructure protection but do not specify technical implementation approaches.

Research on AI governance emphasises the importance of transparency, accountability, fairness, and human oversight. The European Union's (EU) AI Act establishes a risk-based regulatory framework for AI systems, categorising them by risk-level and imposing corresponding requirements. European frameworks, such as the NIS Directive and its successor, NIS2, establish requirements for critical infrastructure protection. High-risk AI systems, including those used for critical infrastructure, must meet stringent requirements for transparency, human oversight, and robustness. However, the translation of these governance principles into specific technical architectures remains an open research question.

The integration of AI systems with legal and policy frameworks presents unique challenges. Automated decision-making systems must respect legal principles, such as due process, proportionality,

and the right to explanation. In the cybersecurity context, automated threat responses must balance effectiveness with legal constraints, including privacy protection, data sovereignty, and jurisdictional boundaries. Research on 'policy-aware' systems has explored formal methods for encoding legal rules and policy constraints into automated decision-making processes, but application to cybersecurity governance remains limited.

### 2.4. Integration Challenges and Hybrid Approaches

Research examining the integration of AI systems with governance frameworks has identified several recurring challenges that complicate the deployment of AI-driven solutions in governance contexts:

5. Temporal and scale mismatches: There is often a fundamental mismatch between the speed and scale at which AI systems operate and the deliberative, oversight-oriented nature of governance processes. AI systems can process vast volumes of data and generate recommendations in real time, while governance processes typically require careful deliberation, stakeholder consultation, and institutional review.
6. Conceptual and linguistic gaps: Technical AI systems and governance structures employ different languages, metrics, and decision-making criteria. Technical systems operate with quantitative metrics, such as accuracy, precision, and recall, while governance structures prioritise qualitative considerations, such as accountability, legitimacy, and public trust. This creates translation challenges when attempting to bridge these domains.
7. Optimisation tensions: There can be inherent tensions between optimisation objectives (e.g. maximising detection accuracy or minimising response time) and governance requirements (e.g. ensuring explainability, fairness, or due process). Optimising purely for technical performance may compromise governance values, while prioritising governance requirements may reduce technical effectiveness.

Some research has explored hybrid approaches that combine automated analysis with human judgement to address these challenges. The concept of 'human in the loop' systems preserve human oversight while leveraging AI for data processing and initial analysis. These systems enable AI to handle routine high-volume tasks while reserving complex or high stakes decisions for human judgement. However, these approaches must carefully balance efficiency gains

from automation with the value of human judgement, particularly in contexts where decisions have significant consequences for national security or individual rights.

The framework proposed in this paper extends these hybrid approaches by explicitly designing the integration between AI capabilities and governance requirements from the outset, rather than treating governance as a constraint or afterthought. By embedding governance principles into the framework's architecture, the design process can proactively address integration challenges rather than attempting to resolve them retroactively.

### 2.5. Research Gaps and Opportunities

Despite significant advances in both AI-driven security analytics and cybersecurity governance research, there remains a critical gap at their intersection. Most AI-based security systems are designed for organisational contexts and do not address the unique requirements of national-level governance, including multi-stakeholder coordination, legal compliance, democratic accountability, and cross-sector information sharing. These systems may excel at technical detection but struggle to translate capabilities into governance-relevant intelligence or coordinate responses across complex institutional structures.

Similarly, governance frameworks provide policy guidance and institutional arrangements but typically lack technical specifications for integrating advanced AI capabilities while preserving governance principles. The translation from policy requirements to technical implementation is often left to practitioners who may lack expertise in both domains, leading to suboptimal solutions or governance principles being compromised for technical expediency.

Academic research has produced numerous innovative AI techniques for cybersecurity problems, but much of this work focuses on isolated technical challenges (detecting specific attack types, classifying malware, and identifying anomalies) without addressing how these capabilities integrate into broader governance contexts. The practical deployment of AI in national cybersecurity governance requires not just effective detection algorithms but also mechanisms for policy alignment, stakeholder coordination, legal compliance, and democratic accountability.

This paper addresses this gap by proposing a comprehensive framework that explicitly bridges technical AI capabilities with

governance requirements. The framework is designed to operate within national cybersecurity governance structures, supporting both technical security operations and policy-level decision-making while maintaining transparency, accountability, and legal compliance. By embedding governance principles into the framework's architecture from the design stage, rather than attempting to add them as constraints or controls afterward, the framework demonstrates how technical capabilities and governance requirements can be mutually reinforcing.

## 3. Cybersecurity Governance and Threat Intelligence

This section examines the fundamental concepts of cybersecurity governance and threat intelligence, establishing the context for the proposed framework. We first define cybersecurity governance and its key components, then analyse the limitations of traditional governance models, and finally explore the evolving role of threat intelligence in national cybersecurity strategies.

Cybersecurity governance refers to the set of policies, institutional arrangements, and decision-making processes that guide how a state manages cyber risks. Effective governance balances technical security measures with legal, ethical, and political considerations. At the national level, cybersecurity governance involves multiple stakeholders, such as government agencies, critical infrastructure operators, regulatory bodies, law enforcement, and intelligence services. The coordination between these entities requires robust information sharing, standardised procedures, and clear lines of authority and accountability.

### 3.1. Limitations of Traditional Governance Models

Traditional cybersecurity governance frameworks emphasise compliance, risk reporting, and incident response protocols. These models typically operate on periodic risk assessments, manual threat analysis, and predefined response playbooks. Human analysts are responsible for interpreting threat data, correlating events across systems, and escalating incidents to appropriate authorities. While this human-centric approach provides valuable judgement and context, it faces fundamental limitations in the face of modern cyber threats.

As cyber environments grow more complex with cloud computing Internet of Things (IoT) devices, mobile networks, and industrial

control systems creating vast attack surfaces, manual processes become increasingly inadequate. SOCs process millions of events daily, generating thousands of alerts, yet human analysts can typically review only a small fraction of these. The signal-to-noise ratio is often unfavourable, with critical threats obscured by false positives and routine security events.

Key limitations of traditional governance models include the following:

1. Temporal delays in threat response: Manual analysis and decision-making processes can take hours or days, while modern attacks can achieve their objectives within minutes. The time gap between threat detection and policy-level response creates windows of vulnerability that adversaries can exploit.
2. Fragmentation between technical and policy domains: Operational security teams operate with technical metrics and indicators, while policy authorities require strategic intelligence and risk assessments. The translation between these domains is often imperfect, leading to miscommunication and suboptimal decision-making.
3. Limited situational awareness across systems: Traditional governance models struggle to maintain comprehensive situational awareness across interconnected critical infrastructure systems. Threats may manifest across multiple sectors or organisations, but fragmented information systems and organisational boundaries impede holistic threat assessment.
4. Static threat models and response procedures: Predefined threat taxonomies and response playbooks cannot adequately address novel attack patterns or rapidly evolving tactics. When adversaries employ previously unseen techniques, traditional frameworks may fail to recognise or respond appropriately to threats.
5. Scalability constraints: As the volume and velocity of cyber threats increase, traditional governance processes face scalability challenges. Manual analysis does not scale linearly with threat volume, leading to backlogs, delayed responses, and missed threats.

These weaknesses are particularly dangerous for critical infrastructure sectors, such as energy, telecommunications, healthcare, financial services, and transportation. A successful cyber attack on critical infrastructure can have cascading effects across multiple sectors, disrupting essential services, causing economic damage, and potentially threatening public safety. The 2015 cyber attack

on Ukraine's power grid, for example, demonstrated how targeted attacks on critical infrastructure can cause widespread outages affecting hundreds of thousands of people. Traditional governance models, with their delayed response times and limited coordination capabilities, are ill-equipped to prevent or rapidly respond to such sophisticated, coordinated attacks.

### 3.2. Role of Threat Intelligence in National Cybersecurity

Threat intelligence aims to collect, analyse, and contextualise information about potential cyber threats, enabling proactive defence and informed decision-making. At the national level, threat intelligence serves multiple functions: providing early warning of emerging threats, supporting attribution of attacks to specific actors or groups, informing strategic policy decisions, and enabling coordinated responses across government agencies and critical infrastructure sectors.

Modern threat intelligence systems aggregate data from diverse sources, including network logs, endpoint telemetry, intrusion detection systems, honeypots, threat feeds from commercial providers, open-source intelligence, and information-sharing partnerships with other nations or organisations [13]. This data diversity, while valuable, also presents challenges. Data formats vary, sources have different levels of reliability and timeliness, and the volume of data can be overwhelming [15].

Traditional threat intelligence processes rely heavily on human analysts to review and correlate information [7]. Analysts examine indicators of compromise (IoCs), review attack patterns, conduct research on threat actors [29], and produce intelligence reports. However, the volume and velocity of threat data often exceed human analytical capacity [16]. Without advanced analytics, threat intelligence systems generate large volumes of alerts with limited actionable insight. Analysts may spend significant time investigating false positives or routine security events while missing subtle indicators of serious threats.

Artificial intelligence has the potential to transform threat intelligence by automating data correlation, identifying hidden patterns across large datasets, and predicting future attack vectors based on historical patterns and the current trends [9]. Machine learning algorithms can process vast volumes of threat data, recognising patterns that human analysts might miss [18]. Natural language processing (NLP) can extract structured intelligence from

unstructured sources, such as threat actor communications, vulnerability reports, or incident response documentation. Predictive analytics can identify emerging threat trends before they manifest as active campaigns [17].

However, the integration of AI-driven threat intelligence with governance structures remains an unresolved challenge [11, 23]. Technical threat indicators must be transformed into strategic intelligence that informs policy decisions. AI systems must operate within legal and ethical constraints, respecting privacy, data sovereignty, and jurisdictional boundaries [29]. The explainability of AI-driven threat assessments is crucial for enabling human oversight and accountability [12, 19]. Furthermore, the rapid pace of threat evolution requires continuous model adaptation and validation, processes that must be integrated into governance workflows rather than operating independently.

## 4. AI-Driven Threat Intelligence Framework

This section presents the proposed AI-driven threat intelligence framework designed for national cybersecurity governance. The framework addresses the limitations identified in previous sections by integrating adaptive machine learning analytics with policy-aware decision mechanisms. We begin with an architectural overview, followed by detailed descriptions of machine learning components and their applications. Table 1 compares traditional and AI-enhanced threat intelligence approaches across key dimensions relevant to national cybersecurity governance.

### 4.1. Architectural Overview

The framework consists of four interconnected layers, each serving distinct functions while maintaining tight integration to ensure end-to-end governance alignment. The layered architecture enables modular development and maintenance while ensuring that governance principles are embedded throughout the system rather than applied as an afterthought. Figure 1 illustrates the overall architecture and data flow, showing how threat intelligence progresses from raw data sources through processing layers to policy-aware decision support.

#### 4.1.1. Data Ingestion Layer

The data ingestion layer aggregates structured and unstructured data from diverse sources, including national computer emergency response teams (CERTs), critical infrastructure

**Table 1.** compares traditional and AI enhanced threat intelligence approaches across key dimensions relevant to national cybersecurity governance.

| Dimension | Traditional approach | AI enhanced approach |
|---|---|---|
| Data processing | Manual review, limited automation | Automated correlation, pattern recognition |
| Analysis speed | Hours to days for complex threats | Real time to minutes for initial assessment |
| Scalability | Limited by human analyst capacity | Scales with computational resources |
| Pattern recognition | Relies on known indicators | Identifies novel patterns and anomalies |
| Predictive capability | Limited to trend analysis | Advanced forecasting and early warning |
| Explainability | High (human analysis) | Variable (requires explainable AI [XAI] techniques) |
| Policy alignment | Manual translation to policy terms | Automated policy-aware categorisation |
| False positive rate | Lower (human judgement) | Higher (requires tuning and validation) |
| Adaptability | Slow (requires process updates) | Fast (continuous model adaptation) |
| Cost | High (human resources) | Lower (after initial investment) |

operators, international intelligence partners, commercial threat feeds, open-source intelligence, and public incident databases [15]. The layer performs comprehensive data pre-processing including normalisation to standardised schemas, de-duplication to eliminate redundant information, quality assessment to identify incomplete or unreliable data, and initial enrichment through entity resolution and context addition [16].

The layer supports multiple data formats including structured logs (Syslog, JSON, XML), network traffic data (packet captures, flow records), IoCs (IP addresses, domain names, file hashes, email addresses), unstructured threat reports (narrative descriptions, incident reports, and vulnerability advisories), and semi-structured data (MITRE ATT&CK framework mappings [14], and STIX/TAXII threat intelligence [15]). Data format conversion and schema mapping ensure compatibility with downstream processing layers.

Privacy-preserving techniques are implemented throughout the ingestion process [29]. Differential privacy mechanisms add calibrated noise to aggregated statistics, preventing inference of individual records while preserving analytical utility [30]. For cross-organisational data sharing, secure multi-party computation enables threat intelligence aggregation without exposing raw data [29]. Homomorphic encryption allows certain computations to be performed on encrypted data, enabling analysis while maintaining confidentiality [31]. The layer also implements data

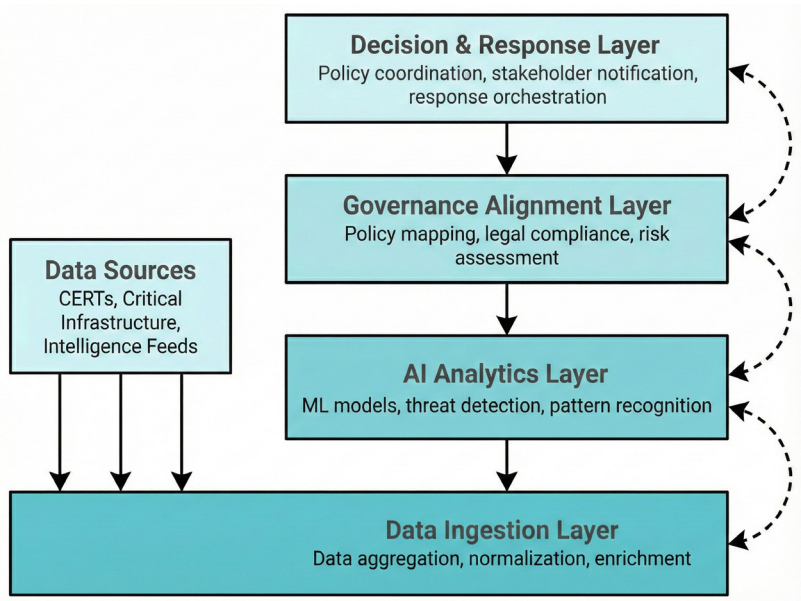


**Figure 1.** Architecture of the proposed AI driven threat intelligence framework for national cybersecurity governance.

quality scoring and reliability weighting, evaluating sources based on historical accuracy, timeliness, completeness, and reputation to ensure high-quality inputs for downstream analysis.

#### 4.1.2. AI Analytics Layer

The AI analytics layer applies multiple machine learning models for anomaly detection, behaviour profiling, threat classification, and predictive analysis. It employs ensemble methods combining supervised, unsupervised, and semi-supervised learning approaches [9]. The layer includes specialised modules for zero-day threat detection [10], attack campaign correlation [11], threat actor attribution [28], and risk scoring. All models incorporate explainability mechanisms to generate interpretable outputs suitable for policy level decision-making [12, 19].

#### 4.1.3. Governance Alignment Layer

The governance alignment layer maps technical threat outputs to policy categories, legal thresholds, and institutional responsibilities [23]. It encodes governance rules, compliance requirements, and jurisdictional boundaries as machine readable constraints [11]. The layer performs threat classification according to national security frameworks [20], assesses legal implications, determines appropriate institutional responses, and ensures proportionality and due process in automated recommendations [32].

#### 4.1.4. Decision and Response Layer

The decision and response layer supports coordinated response actions aligned with national cybersecurity strategies. It provides decision-support tools for policy authorities, generates standardised intelligence reports, coordinates with relevant agencies and critical infrastructure operators [33], and tracks response effectiveness. The layer maintains audit trails for accountability and supports human in the loop oversight for high stakes decisions [24].

This layered architecture ensures that technical insights are transformed into governance-relevant intelligence, with each layer maintaining alignment with policy requirements and institutional structures. The separation of concerns enables modular development, testing, and maintenance while ensuring that governance principles are embedded throughout the system rather than applied as an afterthought. Feedback loops enable continuous learning and adaptation, allowing the system to improve its performance and alignment with governance requirements over time.

### 4.2. Machine Learning Components

The AI analytics layer employs a sophisticated ensemble of machine learning approaches, each optimised for specific threat detection and analysis tasks. The layer integrates multiple model

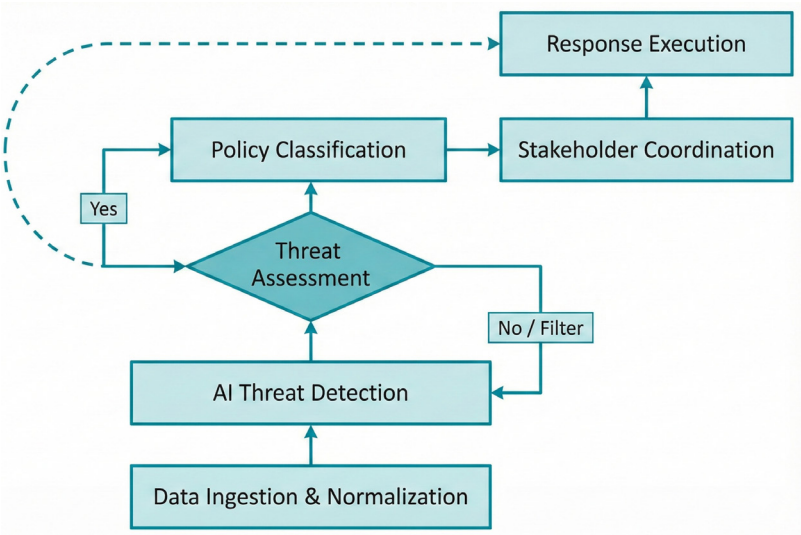


**Figure 2.** Operational workflow of the proposed AI driven threat intelligence framework.

types to leverage their complementary strengths while mitigating individual limitations.

### 4.3. Anomaly Detection and Zero-Day Threat Identification

Unsupervised learning techniques form the foundation for detecting previously unknown threats that lack signature-based indicators. The framework employs several complementary approaches.

Autoencoder-based anomaly detection: Deep autoencoders learn normal behavioural patterns from historical system data, flagging deviations that may indicate novel attack techniques [8]. The models are trained on network traffic patterns, system call sequences, and user behaviour profiles across multiple critical infrastructure sectors.

Isolation forest algorithms: These algorithms efficiently identify outliers in high dimensional feature spaces, detecting rare attack patterns that differ significantly from normal operations [34]. They are particularly effective for detecting stealthy, low-volume attacks that might evade volume-based detection.

One-class support vector machines (SVMs): These models establish boundaries around normal system behaviour, identifying instances that fall outside these boundaries as potential threats. They are well suited for scenarios where attack examples are rare or unavailable.

The framework combines predictions from multiple anomaly detection models using ensemble voting, improving robustness, and reducing false positives. When multiple models independently flag an event as anomalous, confidence in the threat assessment increases significantly.

### 4.4. Graph-Based Threat Correlation

Graph-based analytics enable the identification of coordinated attack campaigns that span multiple systems, organisations, or periods. The framework constructs multi-layered graphs representing:

- Network topology and connectivity patterns.
- Communication relationships between entities (users, systems, and domains).
- Temporal attack sequences and propagation paths.
- Shared IoCs across incidents.

Graph neural networks analyse these structures, learning to identify patterns indicative of coordinated campaigns [11]. For example, simultaneous anomalous activities across geographically distributed systems, when connected through shared infrastructure or known threat actor associations, may signal a coordinated attack campaign. The framework employs attention mechanisms to identify the most relevant connections and temporal patterns. Long Short-Term Memory (LSTM) networks [35] complement graph-based analysis by capturing sequential attack behaviours and temporal dependencies. Table 2 summarises the machine learning models used in the framework and their respective roles.

### 4.5. Explainable AI for Governance

Explainability is essential for maintaining trust and enabling oversight by non-technical decision makers, including policymakers, legal authorities, and senior executives. The framework implements multiple explainability mechanisms:

- Post hoc explanation methods: SHAP values quantify the contribution of each input feature to a model's prediction, enabling analysts to understand which indicators led to a threat assessment [21]. LIME provides interpretable, locally faithful approximations of complex models [19].
- Inherently interpretable models: For critical decision points, the framework employs decision trees or rule-based models that provide transparent, human-readable decision paths [12]. These models trade some accuracy for interpretability, where explainability is paramount [19].
- Narrative generation: Natural language generation techniques automatically produce human-readable explanations of threat

**Table 2.** Summarizes the machine learning models used in the framework and their respective roles.

| Model type | Primary application | Key advantages |
|---|---|---|
| Deep autoencoders | Anomaly detection, zero-day threats | Learns complex normal patterns, scalable |
| Isolation forest | Outlier detection in high dimensions | Fast, handles high dimensional data |
| Graph neural networks (GNNs) | Campaign correlation, attack graphs | Identifies coordinated activities |
| LSTM networks | Temporal pattern recognition | Captures sequential attack behaviours |
| Random forest | Threat classification | Interpretable, robust to overfitting |
| Transformer models | Threat actor attribution | Processes unstructured threat intel |
| SHAP/LIME | Explainability | Provides interpretable predictions |

assessments, translating technical indicators into intelligible narratives suitable for policy-level reporting [23].

The framework maintains a confidence score for each prediction, which is calibrated to reflect both model confidence and reliability of underlying data sources. Low-confidence predictions trigger additional human review, ensuring that high-stakes decisions receive appropriate scrutiny.

## 5. Policy-Aware Response Mechanisms

This section describes the policy-aware response mechanisms that distinguish the proposed framework from purely technical security systems. A defining feature of the framework is its emphasis on policy-aware response, ensuring that automated threat analysis and recommended actions align with legal frameworks, institutional responsibilities, and strategic policy objectives. Unlike conventional security systems, the framework explicitly integrates governance considerations into every stage of threat intelligence processing. We examine governance constraints, privacy mechanisms, proportionality requirements, jurisdictional considerations, and institutional coordination mechanisms.

### 5.1. Governance Constraints and Legal Compliance

Automated threat detection and response recommendations must respect legal and ethical constraints, including privacy protection, data sovereignty, jurisdictional boundaries, proportionality, and due process [23]. The governance alignment layer encodes these constraints as machine readable rules that guide AI-driven recommendations, following policy-aware system design principles [23].

### 5.2. Privacy and Data Protection Mechanisms

The framework implements privacy-preserving techniques throughout the data-processing pipeline. Differential privacy mechanisms add calibrated noise to aggregated statistics, preventing inference of individual records while preserving analytical utility. For sensitive personal data, the framework applies techniques such as:

- Data minimisation: Only necessary attributes are extracted and retained [36].
- Pseudonymisation: Identifiers are replaced with pseudonyms that cannot be reversed without additional information [32].

- Secure multi-party computation: Threat intelligence aggregation across multiple organisations occurs without exposing raw data [29].
- Homomorphic encryption: Certain computations can be performed on encrypted data without decryption [31].

Compliance with regulations, such as the General Data Protection Regulation (GDPR) [36] and national privacy laws, is verified automatically before data processing and threat intelligence-sharing operations.

### 5.3. Proportionality and Due Process

Automated response recommendations must be proportional to the assessed threat severity and respect due process requirements [32]. The framework categorises threats according to severity levels (low, medium, high, and critical) and maps each level to appropriate response actions [23]. Critical actions, such as network disconnection or system isolation, require additional validation and human approval, even for high confidence automated assessments [24, 32].

The framework maintains audit trails of all automated recommendations and actions, enabling retrospective review and accountability [11]. Decision makers can query these trails to understand why specific recommendations were made, supporting both operational learning and legal compliance [12].

### 5.4. Jurisdictional and Sovereignty Considerations

National cybersecurity governance must respect jurisdictional boundaries and data sovereignty requirements [37]. The framework implements geo-fencing mechanisms that prevent threat intelligence data from being processed or stored outside authorised jurisdictions [24]. Cross-border data sharing follows established legal frameworks and treaties, with explicit consent mechanisms and data localisation options where required by law [36, 37].

### 5.5. Institutional Coordination and Stakeholder Management

National cybersecurity governance involves multiple stakeholders, each with distinct roles, responsibilities, and decision-making authorities [33]. The framework supports coordinated decision-making through standardised threat classifications, shared situational awareness, and automated notification and escalation mechanisms [1, 33].

### 5.6. Threat Classification Framework

The framework employs a standardised threat classification system that maps technical indicators to policy relevant categories. Threats are classified along multiple dimensions:

- Severity: Based on potential impact on critical infrastructure, public safety, or national security.
- Confidence: Reflecting the reliability of threat indicators and model predictions.
- Urgency: Indicating the time sensitivity of the required response
- Actor attribution: Nation-state, criminal organisation, hacktivist, insider, or unknown.
- Attack vector: Network-based, physical access, supply chain, social engineering, etc.
- Target sector: Energy, finance, healthcare, telecommunications, government, etc.

This multi-dimensional classification enables appropriate routing to relevant agencies and supports coordinated response planning.

### 5.7. Situational Awareness Dashboard

The framework provides a unified situational awareness dashboard that aggregates threat intelligence across sectors and organisations while respecting information classification and sharing restrictions. The dashboard presents:

- Current threat landscape and active campaigns.
- Sector-specific threat assessments.
- Trends and emerging patterns.
- Response coordination status.
- Risk metrics and trend indicators.

Different stakeholders receive customised views appropriate to their roles and clearance levels. Policy-level decision makers see strategic summaries and trend analyses, while technical security teams access detailed technical indicators and response recommendations.

### 5.8. Automated Escalation and Notification

The framework implements intelligent escalation mechanisms that automatically route threats to appropriate authorities based on severity, sector, and jurisdictional considerations. For example, threats targeting critical energy infrastructure

**Table 3.** Outlines key stakeholders and their roles in the framework's coordination mechanisms.

| Stakeholder | Primary responsibilities |
|---|---|
| National CERT | Threat intelligence aggregation, technical analysis, and incident coordination |
| Sector regulators | Sector-specific oversight, compliance monitoring, and regulatory enforcement |
| Intelligence agencies | Attribution analysis, strategic threat assessment, and classified intelligence integration |
| Critical infrastructure operators | Threat detection, incident reporting, and response implementation |
| Law enforcement | Criminal investigation, attribution support, and legal proceedings |
| Policy authorities | Strategic decision-making, resource allocation, and policy development |
| International partners | Cross-border threat intelligence sharing, coordinated response |

automatically trigger notifications to energy sector regulators, national security agencies, and relevant CERTs, with notifications customised to each recipient's role and information needs.

Escalation protocols ensure that critical threats receive immediate attention while avoiding alert fatigue from routine security events [7]. Machine learning models predict which threats require escalation based on historical patterns and policy rules, reducing manual filtering overhead [8]. Table 3 outlines key stakeholders and their roles in the framework's coordination mechanisms.

## 6. Implementation and Case Studies

This section addresses practical implementation considerations for deploying the framework in national cybersecurity governance contexts. We examine technical infrastructure requirements, organisational prerequisites, and present a detailed case study demonstrating the framework's application in a critical infrastructure protection scenario. Additionally, we discuss performance metrics and evaluation methodologies for assessing framework effectiveness.

### 6.1. Implementation Prerequisites and Requirements

Successful deployment of the framework requires addressing several technical, organisational, and policy prerequisites.

These include data infrastructure, organisational readiness, legal frameworks, and capacity-building initiatives.

### 6.2. Technical Infrastructure Requirements

The framework requires robust technical infrastructure to support data ingestion, processing, storage, and analysis at national scale. Key infrastructure components include:

- High-performance computing resources: Machine learning model training and inference require significant computational capacity, particularly for real-time threat analysis. Cloud computing or dedicated high performance clusters may be necessary depending on data volumes and processing requirements.
- Scalable data storage: Threat intelligence data accumulates rapidly, requiring scalable storage solutions with appropriate security controls and backup mechanisms [16]. Data retention policies must balance analytical needs with privacy and storage cost considerations [29].
- Secure communication channels: Encrypted communication channels are essential for aggregating threat intelligence from multiple sources while protecting sensitive information in transit [37].
- Integration capabilities: The framework must integrate with the existing security systems, including SIEM platforms [16], intrusion detection systems, endpoint protection solutions, and threat intelligence feeds [13]. Standardised Application Programming Interfaces (APIs) and data formats facilitate integration [15].

### 6.3. Organisational and Governance Prerequisites

Beyond technical infrastructure, successful implementation requires appropriate organisational structures and governance mechanisms:

- Clear authority and mandate: The entity deploying the framework must have clear legal authority to collect, process, and share threat intelligence data across sectors and organisations [22].
- Multi-stakeholder coordination mechanisms: Established procedures for coordination between government agencies, critical infrastructure operators, and other stakeholders enable effective threat intelligence-sharing and response coordination [1, 33].
- Trust and information-sharing culture: Organisations must be willing to share threat intelligence data, which may require building trust, demonstrating value, and addressing concerns about information sensitivity and competitive implications [33].

- Skilled personnel: Data scientists, cybersecurity analysts, policy experts, and system administrators with appropriate expertise are needed to operate and maintain the framework [7].

### 6.4. Case Study: Critical Infrastructure Protection Scenario

This case study illustrates how the framework might operate in a real-world scenario involving a coordinated attack on critical energy infrastructure. The scenario demonstrates the framework's ability to detect, analyse, and coordinate response to sophisticated threats.

### 6.5. Scenario Description

A nation-state adversary initiates a multi phase attack targeting energy sector critical infrastructure [27]. The attack begins with spear-phishing campaigns targeting employees of multiple energy companies, followed by lateral movement through networks, reconnaissance of industrial control systems, and preparation for potential disruption operations [4]. The attack spans multiple organisations and occurs over several weeks, making it difficult to detect through traditional, siloed security monitoring.

### 6.6. Framework Operation

During the early phases, the framework's anomaly detection models identify unusual network traffic patterns and authentication anomalies across multiple energy sector organisations [8]. While individually these indicators might be dismissed as benign, graph-based correlation algorithms identify connections indicating a coordinated campaign [11]. The framework correlates indicators, including:

- Similar malware hashes observed at multiple energy companies within a short timeframe
- Shared command and control infrastructure used across different targets
- Coordinated timing of reconnaissance activities
- Geographic clustering of anomalies suggesting targeted regional focus

The governance alignment layer classifies the threat as high severity, high confidence, and targeting critical infrastructure with nation-state characteristics [28]. This classification triggers automated escalation to national CERT, energy sector regulator, national security agencies, and affected organisations [33].

Each stakeholder receives customised intelligence reports: technical teams receive detailed indicators and response recommendations, while policy authorities receive strategic summaries emphasising potential impact and required coordination.

The framework's explainability mechanisms provide clear justifications for threat assessment, enabling decision makers to understand the basis for alerts and make informed response decisions [12, 19]. Policy-aware response recommendations balance immediate containment actions with legal and operational constraints, avoiding disproportionate responses that might disrupt essential services unnecessarily [32].

### 6.7. Outcomes and Lessons

This scenario demonstrates several key framework capabilities:

- Early detection of coordinated threats through pattern recognition across multiple organisations.
- Effective translation of technical indicators into policy-relevant intelligence.
- Coordinated response activation through automated escalation and customised stakeholder communication.
- Balance between automated efficiency and human oversight for critical decisions.

The case study also highlights implementation considerations, including the importance of establishing trust and information-sharing agreements before crises occur, the need for clear escalation and response procedures, and the value of XAI in enabling informed decision-making under time pressure.

### 6.8. Performance Metrics and Evaluation

Evaluating framework effectiveness requires metrics that capture both technical performance and governance objectives.

Technical metrics include:

- Detection accuracy (true positive rate, and false positive rate).
- Time to detection and response.
- Coverage of threat landscape.
- Model prediction confidence calibration.

Governance-focused metrics include:

- Policy compliance rate.
- Stakeholder satisfaction with intelligence quality and timeliness.
- Transparency and explainability scores.
- Response coordination effectiveness.

The framework should be evaluated through controlled testing scenarios, red team exercises [41], and careful monitoring of operational deployments. Continuous improvement requires feedback loops that capture both technical performance data and stakeholder input, enabling iterative refinement of models, policies, and processes [41]. Table 4 presents key performance indicators for framework evaluation.

### 6.9. Proposed Empirical Validation Strategy

Although the proposed framework is conceptual in nature, its effectiveness can be evaluated through a structured empirical validation process. Future implementation should utilise publicly available cybersecurity datasets, such as CICIDS2017, CICIoT2023, UNSW NB15, and CSE CIC IDS2018.

The validation process would involve training and evaluating multiple machines learning models, including Isolation Forest, Autoencoders, Random Forests, and GNNs. Performance should be assessed using accuracy, precision, recall, F1 score, false positive rate, and detection latency.

In addition to technical evaluation, governance-related metrics should be measured, including policy compliance rate, explainability

**Table 4.** presents key performance indicators for framework evaluation.

| Metric category | Specific metric | Target/threshold |
|---|---|---|
| Detection performance | True positive rate<br>false positive rate<br>time to detection | >95% for high severity threats<br><5% to avoid alert fatigue<br><15 min for critical threats |
| Policy compliance | Legal compliance rate<br>Proportionality score | 100% (zero violations)<br>Proportionate responses for 95%+ cases |
| Explainability | Explanation quality score<br>Human review override rate | Mean score > 4/5 from stakeholders<br><10% of automated recommendations |
| Coordination | Stakeholder satisfaction<br>Response coordination time | Mean score > 4/5<br><2 h for multi-stakeholder responses |

quality, stakeholder satisfaction, response coordination effectiveness, and transparency indicators.

Red team exercises and cyber range simulations can further assess framework robustness under realistic attack conditions. Such validation would provide quantitative evidence regarding the framework's ability to support both threat detection and governance-aligned decision-making.

Validation experiments should compare framework performance against traditional SIEM-based monitoring systems and standard threat intelligence workflows. Comparative evaluation should measure detection accuracy, false positive rates, response times, and policy-compliance outcomes under identical operational conditions.

Scenario-based simulations involving ransomware campaigns, APTs, supply chain attacks, and critical infrastructure disruptions can further evaluate framework effectiveness. Such comparative testing would provide evidence regarding the practical advantages and limitations of governance-aware AI-driven threat intelligence architectures.

## 7. Discussion

This section provides a comprehensive analysis of the framework's implications, contributions, and limitations. We compare the proposed framework with the existing approaches, examine implementation challenges and opportunities, and discuss ethical and legal considerations. The discussion situates the framework within the broader context of national cybersecurity governance and identifies directions for future research and development.

### 7.1. Addressing the Governance Technical Gap

The proposed framework addresses a critical gap between technical cybersecurity solutions and governance requirements that has persisted in both research and practice. Traditional security systems excel at technical detection but often operate in isolation from policy and institutional structures. Conversely, governance frameworks provide policy guidance but lack technical specifications for integrating advanced AI capabilities while preserving governance principles, such as transparency, accountability, and legal compliance.

By embedding AI within policy-aware structures from the design stage, the framework demonstrates how technical capabilities and governance requirements can be mutually reinforcing rather than conflicting. The governance alignment layer explicitly translates technical threat indicators into policy-relevant intelligence, ensuring that sophisticated detection capabilities translate into actionable governance insights. This approach contrasts with *post hoc* attempts to add governance controls to purely technical systems, which often result in compromises or inefficiencies.

The framework's emphasis on explainability addresses a fundamental governance requirement: the ability of human decision makers to understand, validate, and if necessary, override automated recommendations. This is particularly important in national security contexts where decisions may have significant legal, political, or operational consequences. The integration of multiple explainability mechanisms from technical feature importance scores to natural language narrative generation enables different types of stakeholders to engage with AI outputs at appropriate levels of detail.

### 7.2. Enhancing National Cyber Resilience

The framework enhances national cyber resilience through several mechanisms. First, by enabling earlier detection of threats through automated pattern recognition and anomaly detection [8, 9], it reduces the time between attack initiation and defensive response. This temporal advantage is crucial, as many sophisticated attacks achieve their objectives within hours or days [4], while traditional manual analysis may take days or weeks [7].

Second, the framework improves situational awareness by aggregating and correlating threat intelligence across multiple sectors and organisations [13]. Individual organisations may observe isolated indicators that seem benign, but the framework can identify patterns indicating coordinated campaigns when viewing the broader landscape [11]. This holistic perspective is essential for defending against sophisticated adversaries who intentionally distribute attacks across multiple targets to avoid detection.

Third, the framework supports coordinated response by standardising threat classifications and automating stakeholder notification and escalation [33]. In complex national security environments with multiple agencies and organisations, coordination delays can significantly reduce response effectiveness. Automated coordination

mechanisms reduce friction while maintaining appropriate oversight and control [24].

However, resilience enhancement is not solely a technical matter. The framework's effectiveness depends on the quality and breadth of data inputs, the trust and cooperation of participating organisations [33], and the maturity of organisational processes and capabilities [6]. Technical capabilities alone cannot compensate for organisational or institutional weaknesses.

### 7.3. Comparison with Existing Approaches

The framework differs from the existing approaches in several key respects. Compared with traditional SIEM systems [16], which primarily focus on log aggregation and rule-based alerting, the framework emphasises adaptive machine learning, threat correlation across organisations, and policy-aware intelligence generation. While SIEMs require extensive manual configuration and tuning, the framework employs self-learning models that adapt to evolving threats.

Compared with commercial threat intelligence platforms, which typically serve individual organisations [13], the framework is explicitly designed for national-level governance, incorporating multi-stakeholder coordination [33], legal compliance mechanisms [22, 36], and policy alignment [23]. Commercial platforms often treat governance requirements as constraints rather than integral design considerations.

Academic research in AI-driven cybersecurity has produced numerous innovative detection techniques [8, 18], but often focuses on isolated technical problems without addressing governance integration. The framework bridges this gap by providing a comprehensive architecture that addresses both technical detection and governance requirements.

### 7.4. Comparative Analysis with Existing Frameworks (Table 5)

The comparative analysis indicates that the existing frameworks typically emphasise either technical detection capabilities or governance guidance. The proposed framework differs by integrating AI analytics, governance alignment, explainability, legal compliance, and coordinated response mechanisms within a unified architecture. This integration enables technical threat intelligence

**Table 5.** Comparative analysis of the existing frameworks and proposed framework.

| Feature | NIST CSF | MITRE ATT&CK-based solutions | Commercial threat intelligence platforms | Proposed framework |
|---|---|---|---|---|
| National governance focus | Partial | No | No | Yes |
| AI-driven analytics | Limited | Limited | Moderate | Advanced |
| Policy-aware decision-making | No | No | Limited | Yes |
| Governance alignment layer | No | No | No | Yes |
| Explainable AI (XAI) integration | Limited | Limited | Moderate | Comprehensive |
| Cross-sector coordination | Partial | No | Limited | Comprehensive |
| Legal compliance mapping | Partial | No | Limited | Comprehensive |
| Automated escalation mechanisms | Limited | No | Moderate | Advanced |
| Human in the loop oversight | Partial | No | Partial | Comprehensive |
| National cyber resilience support | Partial | No | Limited | Comprehensive |

Note: NIST CSF: National Institute of Standards and Technology cybersecurity framework.

to be translated directly into governance-relevant decision support, thereby addressing a gap that remains insufficiently covered in the existing national cybersecurity frameworks.

### 7.5. Implementation Challenges

Despite its potential benefits, the framework faces significant implementation challenges that must be addressed for successful deployment.

### 7.6. Data-Sharing Barriers

Effective threat intelligence sharing requires organisations to overcome traditional barriers, including competitive concerns, legal restrictions, data sensitivity, and lack of trust [33]. Energy companies, for example, may be hesitant to share detailed security data with competitors or even government agencies. Legal frameworks may restrict data sharing across sectors or jurisdictions [37]. The framework addresses these challenges through privacy-preserving techniques [29] and explicit data governance mechanisms, but building trust and establishing appropriate legal frameworks remains a prerequisite that requires sustained effort.

### 7.7. Institutional Resistance to Automation

Organisations may resist increased automation in cybersecurity decision-making due to concerns about job displacement, loss of control, or lack of understanding of AI capabilities and limitations. Resistance may be particularly strong in government agencies where traditional procedures and human judgement have long been the norm [6]. Addressing this requires change management, training, and demonstration of value through pilot programmes and incremental deployment. The framework's emphasis on human in the loop oversight [24] and explainability [12] can help address concerns while enabling gradual adoption.

### 7.8. Adversarial AI and Framework Robustness

The increasing use of AI within cybersecurity governance introduces new attack surfaces that adversaries may exploit. While AI enhances detection accuracy and situational awareness, machine learning models themselves can become targets of manipulation. Therefore, robustness against adversarial attacks must be considered a core design requirement rather than an optional enhancement.

#### 7.8.1. Data Poisoning Attacks

Data poisoning attacks attempt to manipulate training datasets by injecting malicious or misleading samples. Such attacks can degrade model accuracy, bias classification outcomes, or suppress detection of specific threat categories. Within national cybersecurity environments, poisoning attacks may originate from compromised data feeds, malicious insiders, or manipulated threat intelligence sources. Continuous data validation, source reputation scoring, and anomaly screening should therefore be integrated into the data ingestion layer.

#### 7.8.2. Model Evasion Attacks

Adversaries may deliberately craft inputs designed to bypass machine learning detection mechanisms. Malware variants, adversarial network traffic patterns, and modified attack signatures can exploit weaknesses in learned decision boundaries. Ensemble learning approaches and continuous model retraining reduce susceptibility to such attacks by increasing detection diversity and adaptability.

#### 7.8.3. Prompt Injection and Large Language Model (LLM) Manipulation

As LLMs become integrated into cybersecurity workflows, prompt injection attacks represent a growing concern. Malicious

inputs may attempt to manipulate automated reasoning processes, generate inaccurate intelligence assessments, or influence policy recommendations. Defensive measures include prompt filtering, contextual validation, role separation, and human verification of high impact outputs.

#### 7.8.4. Defensive Mechanisms

The proposed framework incorporates multiple protection mechanisms, including adversarial training, continuous monitoring, explainability-driven auditing, model redundancy, and human in the loop validation. Together, these controls improve resilience against manipulation attempts while maintaining operational effectiveness. Robustness testing should be incorporated into routine evaluation cycles to ensure that defensive capabilities evolve alongside emerging attack techniques.

### 7.9. Ethical and Legal Considerations

The framework operates in contexts where ethical and legal considerations are paramount [32]. Automated threat detection may raise concerns about surveillance, privacy infringement, and false positives, leading to unwarranted investigation or action, and potential for abuse. Legal frameworks vary across jurisdictions and may be unclear or evolving [22, 36]. The framework's emphasis on privacy-preserving techniques [29], proportionality, due process, and transparency helps to address these concerns, but legal and ethical review should be an ongoing process rather than a one-time approval.

### 7.10. Resource Requirements

Implementing and operating the framework requires significant resources, including computing infrastructure, skilled personnel, data storage, and ongoing maintenance and development [7]. Smaller nations or organisations with limited resources may face challenges in deploying the full framework. However, the framework's modular architecture enables incremental deployment, starting with core capabilities and expanding over time. Cloud computing and managed services can also help address resource constraints.

### 7.11. Opportunities and Future Directions

Despite challenges, the framework creates opportunities for advancing both technical cybersecurity capabilities and

governance practices. The integration of XAI with governance requirements provides a model for other domains where AI systems must operate within legal and ethical constraints [12, 23]. The framework's emphasis on transparency and accountability may help build public trust in AI applications for security purposes, which has been lacking in some contexts [32].

Future research directions include empirical validation of the framework through pilot deployments, comparative analysis across different national contexts and governance models, and development of standardised metrics and evaluation methodologies. The framework's modular architecture also enables ongoing innovation, as new machine learning techniques, threat detection methods [8, 11], or governance mechanisms [23] can be integrated without requiring fundamental redesign.

The evolving threat landscape, particularly the increasing use of AI by adversaries, creates urgency for adaptive defensive capabilities. The framework provides a foundation for addressing these challenges while maintaining governance principles, but continuous adaptation and improvement will be essential as threats and technologies continue to evolve [10].

## 8. Limitations

Despite its contributions, the proposed framework has several limitations. First, the framework remains conceptual and has not yet been validated through large-scale operational deployment. The effectiveness of the proposed architecture therefore relies on assumptions derived from the existing cybersecurity and AI research.

Second, implementation requires significant computational resources, high-quality threat intelligence feeds, and sustained inter-organisational cooperation. Such requirements may limit adoption within resource-constrained environments.

Third, legal and regulatory requirements vary across jurisdictions. Governance mechanisms proposed in this study may require adaptation to national legal frameworks, institutional structures, and data protection regulations.

Fourth, adversarial machine learning remains an evolving challenge. Although mitigation mechanisms have been incorporated into the framework design, future attack techniques may require additional defensive strategies and continuous model adaptation.

These limitations highlight the need for future empirical evaluation and real-world pilot deployments to validate framework effectiveness across diverse operational environments.

## 9. Response to Research Questions

The findings of this study provide direct responses to the research questions introduced in the introduction.

RQ1 examined how AI-driven threat intelligence can improve national cybersecurity governance and strategic decision-making. The proposed framework demonstrates that AI-driven analytics can enhance governance by enabling automated threat detection, predictive risk assessment, real-time situational awareness, and rapid intelligence generation. These capabilities support more informed and timely strategic decisions at the national level.

RQ2 explored how governance requirements, including transparency, accountability, legal compliance, and institutional oversight, can be embedded within AI-based threat intelligence architectures. The framework addresses this challenge through the governance alignment layer, XAI mechanisms, audit trails, policy-aware response controls, and human in the loop oversight processes.

RQ3 investigated how policy-aware response mechanisms can improve national cyber resilience. The results indicate that integrating governance constraints directly into threat intelligence workflows enables coordinated responses that remain legally compliant, proportionate, and aligned with national cybersecurity objectives while reducing response delays.

RQ4 focused on identifying the architectural and governance components necessary for effective integration between AI-driven analytics and national cybersecurity governance. The proposed architecture demonstrates that successful integration requires coordinated data ingestion, adaptive AI analytics, governance alignment mechanisms, explainability controls, stakeholder coordination capabilities, and decision support systems operating within a unified framework.

## 10. Conclusions

As cyber threats continue to evolve in sophistication, scale, and impact, national cybersecurity governance must move

beyond static, reactive models that struggle to keep pace with rapidly changing threat landscapes. The digitisation of critical infrastructure, the proliferation of interconnected systems, and the emergence of AI-assisted attack techniques create an urgent need for adaptive, intelligent defensive capabilities. However, simply applying advanced AI techniques to cybersecurity problems is insufficient. For national level cybersecurity governance, AI must be embedded within policy-aware structures that preserve transparency, accountability, legal compliance, and democratic oversight.

This paper has presented a comprehensive AI-driven threat intelligence framework explicitly designed to integrate adaptive analytics with policy-aware decision-making for national cybersecurity governance. The framework addresses a critical gap in the existing research and practice: the disconnect between sophisticated technical detection capabilities and the governance requirements essential for national security contexts. By designing governance principles into the framework's architecture from the outset, rather than attempting to retrofit governance controls, the framework demonstrates how technical capabilities and governance requirements can be mutually reinforcing.

The framework's four-layer architecture (data ingestion, AI analytics, governance alignment, and decision/response) ensures that threat intelligence flows from raw data through technical analysis to policy-relevant intelligence and coordinated response, with governance considerations embedded at every stage. The emphasis on XAI enables human decision makers to understand, validate, and when necessary, override automated recommendations, preserving human agency and accountability in high-stakes security decisions. Privacy-preserving techniques, proportionality mechanisms, and legal compliance verification ensure that the framework operates within ethical and legal boundaries.

The framework's contributions extend beyond technical design to addressing broader challenges in cybersecurity governance. By providing standardised threat classification systems, automated stakeholder coordination mechanisms, and shared situational awareness capabilities, the framework supports the multi-stakeholder coordination essential for effective national cybersecurity governance. The emphasis on continuous learning and adaptation acknowledges that both threats and governance requirements evolve over time, requiring systems that can improve their performance and alignment iteratively.

However, the framework's potential can only be realised through careful implementation that addresses the significant challenges discussed in this paper. Data-sharing barriers, institutional resistance to automation, model validation requirements, and resource constraints, all pose obstacles to deployment. Success will require sustained effort to build trust among stakeholders, establish appropriate legal and policy frameworks, develop organisational capabilities, and demonstrate value through pilot programmes and incremental deployment.

The framework also raises important questions for future research and practice. How can we develop standardised metrics for evaluating governance-aligned AI systems in cybersecurity contexts? How can we balance the benefits of automation with the need for human oversight and control? How do different national governance models and legal frameworks affect framework implementation and effectiveness? How can we ensure that AI-driven threat intelligence systems remain robust against adversarial manipulation and maintain effectiveness as threats evolve? These questions point to broader themes in the intersection of AI and governance. As AI systems become increasingly powerful and are deployed in critical domains like cybersecurity, the need for governance-aligned design becomes ever more important. The framework presented here provides one model for how this alignment can be achieved, but continued research and practice are needed to refine approaches and adapt to evolving requirements.

In conclusion, this paper contributes to applied cybersecurity research by offering a scalable, governance-centric model suitable for critical infrastructure protection and national cyber defence strategies. The framework demonstrates how AI can support strategic cyber resilience while preserving transparency, accountability, and policy compliance, essential requirements for democratic governance in the digital age. As nations face increasingly sophisticated cyber threats, frameworks that successfully bridge technical capabilities and governance requirements become essential for maintaining security while preserving democratic values and institutional accountability.

Future work can focus on empirical validation of the framework through controlled testing scenarios and pilot deployments, comparative analysis across different governance contexts and national security models, development of standardised evaluation methodologies, and exploration of how emerging AI techniques can enhance framework capabilities while maintaining governance

alignment. The framework's modular architecture enables ongoing innovation, as new threat-detection methods, machine learning techniques, or governance mechanisms can be integrated to address evolving challenges.

## Funding

This research received no external funding.

## Data Availability Statement

No new datasets were generated or analysed in the current study.

## Conflict of Interest

The author declares no conflict of interest.

## Acknowledgements

The author thanks anonymous reviewers for their valuable comments and constructive suggestions, which significantly improved the quality and clarity of this manuscript.

## References

[1] R. von Solms, J. van Niekerk, "From information security to cyber security," *Computers & Security*, vol. 38, pp. 97–102, 2013, doi: 10.1016/j.cose.2013.04.004.

[2] M. Husak, M. Komarkova, E. Bou-Harb, P. Celeda, "Survey of attack projection, prediction, and forecasting in cyber security," *IEEE Communications Surveys & Tutorials*, vol. 21, no. 1, pp. 640–660, 2019, doi: 10.1109/COMST.2018.2871866.

[3] A. Khraisat, I. Gondal, P. Vamplew, J. Kamruzzaman, "Survey of intrusion detection systems: Techniques, datasets and challenges," *Cybersecurity*, vol. 2, no. 1, pp. 1–22, 2019, doi: 10.1186/s42400-019-0038-7.

[4] R. Sommer, V. Paxson, "Outside the closed world: on using machine learning for network intrusion detection," in Proceedings of the 2010 IEEE Symposium on Security and Privacy, Los Alamitos, CA, USA: IEEE Computer Society, 2010, pp. 305–316.

[5] G. Apruzzese, M. Colajanni, L. Ferretti, A. Guido, M. Marchetti, "On the effectiveness of machine and deep learning for cyber security," in Proceedings of the 10th International Conference on Cyber Conflict (CyCon 2018), Tallinn, Estonia: IEEE, 2018, pp. 371–390, doi: 10.23919/CYCON.2018.8405026.

[6] N. Papernot, P. McDaniel, A. Sinha, M. Wellman, "SoK: Security and Privacy in Machine Learning," in Proceedings of the 2018 IEEE European Symposium on Security and Privacy (EuroS&P), Los Alamitos, CA, USA: IEEE Computer Society, 2018, pp. 399–414, doi: 10.1109/EuroSP.2018.00035.

[7] I. Goodfellow, P. McDaniel, N. Papernot, "Making machine learning robust against adversarial inputs," *Communications of the ACM*, vol. 61, no. 7, pp. 56–66, 2018, doi: 10.1145/3134599.

[8] A. Shostack, *Threat modeling: Designing for security*. Indianapolis, IN: Wiley, 2014.

[9] M. Conti, A. Dehghantanha, K. Franke, S. Watson, "Internet of Things security and forensics: Challenges and opportunities," *Future Generation Computer Systems*, vol. 78, pp. 544–546, 2018, doi: 10.1016/j.future.2017.07.060.

[10] European Union Agency for Cybersecurity (ENISA), *ENISA threat landscape 2024*. Attiki: ENISA, 2024.

[11] National Institute of Standards and Technology (NIST), *Cybersecurity framework (CSF) 2.0*. Gaithersburg, MD: NIST, 2024.

[12] European Parliament and Council, *Regulation (EU) 2024/1689 Laying Down Harmonised Rules on Artificial Intelligence (AI Act)*. Brussels: European Parliament and Council, 2024.

[13] European Parliament and Council, *Directive (EU) 2022/2555 on Measures for a High Common Level of Cybersecurity Across the Union (NIS2 Directive)*. Brussels: European Parliament and Council, 2022.

[14] Cybersecurity and Infrastructure Security Agency (CISA), *Cross sector cybersecurity performance goals*. Washington, DC: CISA, 2024.

[15] Cybersecurity and Infrastructure Security Agency (CISA), *Secure by design principles*. Washington, DC: CISA, 2023.

[16] C. Tankard, "Advanced persistent threats and how to monitor and deter them," *Network Security*, vol. 2011, no. 8, pp. 16–19, 2011, doi: 10.1016/S1353-4858(11)70086-1.

[17] M. Ring, D. Schlor, D. Landes, A. Hotho, "Flow based network traffic generation using generative adversarial networks," *Computers & Security*, vol. 82, pp. 156–172, 2019, doi: 10.1016/j.cose.2018.12.012.

[18] S. Russell, P. Norvig, *Artificial Intelligence: A modern approach*, 4th ed. Hoboken, NJ: Pearson, 2021.

[19] European Union Agency for Cybersecurity (ENISA), *Cybersecurity and resilience of critical sectors in Europe*. Attiki: ENISA, 2023.

[20] World Economic Forum (WEF), *Global cybersecurity outlook 2025*. Geneva: WEF, 2025.

[21] Organization for Economic Cooperation and Development (OECD), *Artificial Intelligence, machine learning and cybersecurity policy frameworks*. Paris: OECD Publishing, 2024.

[22] M. Bishop, *Computer security: Art and science*, 2nd ed. Boston, MA: Addison-Wesley, 2018.

[23] B. Schneier, *Click here to kill everybody: Security and survival in a hyper-connected world*. New York, NY: W.W. Norton, 2018.

[24] European Union Agency for Cybersecurity (ENISA), *Guidelines for Threat Intelligence Sharing and Analysis*. Attiki: ENISA, 2023.

**[25]** MITRE Corporation, *Engenuity, ATT&CK evaluations methodology and results*. McLean, VA, 2024.

**[26]** National Institute of Standards and Technology (NIST), *AI risk management framework (AI RMF 1.0)*. Gaithersburg, MD: NIST, 2023.

**[27]** Organization for Economic Cooperation and Development (OECD), *OECD framework for the classification of AI systems*. Paris: OECD Publishing, 2022.

**[28]** M. U. Iqbal, M. Hammoudeh, S. Zeadally, "Threat intelligence in cybersecurity, A systematic review," IEEE Access, vol. 7, pp. 87008–87031, 2019.

**[29]** C. Dwork, A. Roth, "The Algorithmic Foundations of Differential Privacy," Foundations and Trends in Theoretical Computer Science, vol. 9, no. 3–4, pp. 211–407, 2014. doi: 10.1561/0400000042.

**[30]** C. Dwork, "Differential privacy," *Proceedings of the 33rd International Colloquium on Automata, Languages and Programmin*g, pp. 1–12, 2006.

**[31]** C. Gentry, Fully Homomorphic Encryption Using Ideal Lattices. Stanford University, 2009.

**[32]** European Union, Charter of Fundamental Rights of the European Union, Official Journal of the European Union, 2012.

**[33]** ENISA, Good Practices for Information Sharing and Analysis Centers (ISACs), ENISA, 2022.

**[34]** F. T. Liu, K. M. Ting, Z. H. Zhou, "Isolation forest," *Proceedings of the 8th IEEE International Conference on Data Mining*, pp. 413–422, 2008.

**[35]** S. Hochreiter, J. Schmidhuber, "Long short term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

**[36]** European Union, Regulation (EU) 2016/679 (General Data Protection Regulation), Brussels, 2016.

**[37]** OECD, *Data free flow with trust and cross border data governance*, OECD Publishing.

**[38]** MITRE Engenuity, *ATT&CK evaluations methodology and red team evaluation Framework*, 2024.