

DOI: https://doi.org/10.48009/1_iis_2026_140

Intelligence, information operations, and cyber-physical security in the Middle East, with a focus on Saudi Arabia

May Almousa, *Massachusetts Institute of Technology*, malmousa@mit.edu

Stuart Madnick, *Massachusetts Institute of Technology*, smadnick@mit.edu

Abstract

Artificial intelligence is increasingly reshaping intelligence, information operations, and cyber-physical security in national security competition around the world. As AI-enabled technologies are integrated into IoT platforms, industrial control systems, and smart city infrastructure, they are changing how information is collected, analyzed, and acted upon across civilian and strategic domains. These capabilities expand analytic and operational capacity, but they also create new risks related to governance, resilience, deception, disruption, and cascading failure. Using qualitative content analysis and comparative institutional governance analysis, this paper examines these shifts with attention to regulatory and institutional challenges in the Middle East, using Saudi Arabia as a focal analytical setting. It analyzes how oversight of AI-enabled systems is distributed across cybersecurity authorities, data governance institutions, sector regulators, infrastructure operators, and technology vendors. Institutions such as the National Cybersecurity Authority and the Saudi Data and Artificial Intelligence Authority provide strategic direction through cybersecurity controls and AI governance frameworks. Yet questions remain regarding cross-sector coordination, accountability for AI-enabled decision making, and dependence on global technology supply chains. By tracing Saudi Arabia's evolving regulatory landscape, the paper argues that governance ambiguity can become a source of systemic risk as AI becomes more deeply embedded in intelligence functions and critical infrastructure.

Keywords: AI governance, cybersecurity regulation, national security, information operations, cyber-physical security, Saudi Arabia.

Introduction

Artificial intelligence has become an important structural component of contemporary national security systems. Across intelligence institutions, digital governance platforms, industrial control environments, and information ecosystems, AI-driven systems increasingly mediate how states collect information, assess risk, coordinate responses, and exercise authority. Their integration does not simply improve efficiency; it reshapes the institutional foundations of security governance.

This transformation is unfolding within a rapidly intensifying regional cyber threat environment. Reported data indicate that cyber offensives in the Middle East and North Africa tripled year-on-year in the first quarter of 2024, underscoring the growing frequency and sophistication of attacks targeting government institutions and critical infrastructure (Oxford Analytica, 2024).

At the same time, cyber-physical convergence has altered the relationship between digital systems and material infrastructure. Energy production facilities, transportation networks, telecommunications systems, desalination plants, and smart urban environments now operate through digitally mediated control layers that incorporate automated decision processes. As AI systems become embedded within these infrastructures, vulnerabilities in digital domains can produce direct physical and economic consequences (World Economic Forum, 2023). AI integration therefore expands both capability and risk.

The Middle East provides a particularly salient context for examining these dynamics. Regional states operate within an environment shaped by strategic competition, energy interdependence, and ambitious digital transformation agendas. Governments across the region have incorporated artificial intelligence into national strategies as a tool of economic growth, administrative modernization, and regional positioning (Hendawy & Kumar, 2024). Yet the rapid integration of AI into intelligence systems and critical infrastructure introduces governance challenges that remain unevenly institutionalized.

Saudi Arabia was selected as the focal case because it combines strategic infrastructure significance, rapid AI-enabled modernization, and a developing national governance architecture for cybersecurity, data, and artificial intelligence. Under Vision 2030, the Kingdom has pursued comprehensive digital modernization, including AI-enabled public services, smart urban development, and the digitization of energy and industrial sectors (Vision 2030 Annual Report, 2023; SDAIA, 2020). At the same time, Saudi Arabia's energy, water, transportation, telecommunications, smart city, and public-sector systems make cyber-physical resilience a national and regional security concern. Institutions such as the National Cybersecurity Authority (NCA) and Saudi Data and Artificial Intelligence Authority (SDAIA) provide a strong foundation for examining how AI governance and cybersecurity oversight are structured. However, as AI systems are deployed across ministries, defense institutions, infrastructure operators, sector regulators, state-owned enterprises, telecommunications providers, and commercial technology partners, Saudi Arabia illustrates the governance ambiguity that can emerge when AI-enabled systems cross institutional, sectoral, and technological boundaries.

This paper argues that a central governance challenge of AI integration in national security contexts is governance ambiguity. As AI systems inform intelligence analysis, automate operational decisions, and optimize cyber-physical infrastructure, authority and accountability frequently span institutional boundaries. This ambiguity appears in three related ways. First, responsibility for AI systems is distributed among cybersecurity authorities, data governance bodies, sector regulators, infrastructure operators, and vendors. Second, AI-informed decision processes complicate accountability when algorithmic outputs shape intelligence prioritization, infrastructure responses, or content moderation decisions. Third, reliance on global technology ecosystems embeds national systems within broader supply chains for cloud infrastructure, semiconductors, proprietary models, and vendor-managed software (OECD, 2022). The contribution of this paper is not simply to identify regulatory fragmentation or institutional coordination challenges. Rather, it distinguishes governance ambiguity as a distinct form of systemic risk that emerges when responsibility, accountability, decision authority, and incident leadership become unclear across AI-enabled security and infrastructure systems. This perspective provides a framework for analyzing how formal governance institutions may exist while operational responsibility remains uncertain when AI systems traverse cybersecurity, data governance, sector regulation, critical infrastructure, and vendor-managed technology environments.

This study proceeds in four parts. First, it examines how AI transforms intelligence practices and the broader information environment. Second, it analyzes AI integration within cyber-physical systems. Third, it evaluates Saudi Arabia's institutional and regulatory architecture, including the roles of SDAIA and the

NCA. Fourth, it develops academic and practical implications concerning institutional integration, algorithmic assurance, cross-domain incident governance, and resilience by design.

Literature Review and Theoretical Framework

Artificial Intelligence and National Security: From Capability to Governance

The literature on artificial intelligence and national security has often focused on strategic stability, military applications, autonomous systems, and great-power competition (Black et al., 2024). Earlier literature often emphasized operational capability and strategic competition, while more recent work broadens the discussion to intelligence analysis, cyber operations, information manipulation, and the governance of high-risk AI systems (Black et al., 2024; Brundage et al., 2018; Novelli et al., 2025; SCSP & ASPI, 2024). This shift is central because AI is not only changing what security institutions can do; it also changes how authority, accountability, and oversight are organized.

AI governance research increasingly stresses that high-impact systems require cross-sector coordination, lifecycle monitoring, accountability mechanisms, and risk-based oversight (OECD, 2022; Mäntymäki et al., 2022; Novelli et al., 2025). In national security settings, these requirements become more complex because AI systems often operate across intelligence agencies, cybersecurity authorities, data governance bodies, infrastructure operators, and commercial technology providers. AI therefore functions as a dual-use technology: it can strengthen anomaly detection, predictive analytics, and threat classification, while also expanding attack surfaces through adversarial manipulation, data poisoning, model extraction, and automated deception (Brundage et al., 2018; ENISA, 2020). In this sense, AI does not merely augment security institutions; it restructures the informational and operational environments in which they make decisions (Rahwan et al., 2019).

Cyber-Physical Systems and Systemic Risk

Cyber-physical systems (CPS) provide a second foundation for this analysis. CPS link computational processes to physical infrastructure through sensors, industrial control systems, operational technology, and networked automation. In sectors such as energy, water desalination, transportation, telecommunications, and smart urban systems, digital control layers increasingly mediate physical operations in real time (Radanliev et al., 2021). When AI is embedded in these environments, risk no longer remains confined to information systems. Faulty data, manipulated sensor inputs, model drift, or compromised vendor software can produce operational, economic, and physical consequences.

Systemic cyber risk literature emphasizes that tightly coupled digital systems can propagate failures across sectors (World Economic Forum, 2023). AI intensifies this problem because automated systems may respond to data at machine speed, sometimes without full human interpretation before action is taken. In AI-enabled cyber-physical environments, resilience therefore depends not only on technical cybersecurity controls, but also on institutional coordination, incident leadership, and regulatory interoperability. This is particularly relevant for energy-exporting states such as Saudi Arabia, where disruption to oil production, electricity systems, logistics, or telecommunications may have domestic, regional, and international consequences.

Institutional Fragmentation and Regulatory Ambiguity

Institutional theory helps explain why AI governance becomes difficult when authority is distributed across multiple organizations. Emerging technologies frequently span regulatory jurisdictions, creating coordination problems when responsibility is divided among agencies, regulators, operators, and private vendors (Büthe et al., 2022). AI systems often sit at the intersection of cybersecurity regulation, data

governance, sectoral oversight, intelligence mandates, procurement rules, and international obligations. As a result, no single institution may have complete visibility over the full lifecycle or risk chain of an AI-enabled system.

Marotta et al. (2023) identify patterns, gaps, and areas for improvement in the emerging cybersecurity regulatory environment, suggesting that the growth of cybersecurity regulation does not automatically resolve coordination challenges across organizations, sectors, and jurisdictions. In AI-enabled systems that connect intelligence, infrastructure, and commercial technology providers, these regulatory gaps can become sites of vulnerability. Ambiguity in incident reporting requirements, including uncertainty over timing, definitions, and reporting scope, can complicate effective response and accountability (Marotta & Madnick, 2023). Recent work on cybersecurity regulation reinforces this point. Marotta and Madnick's analysis of emerging cybersecurity regulations identifies recurring features such as incident reporting, risk management, and compliance requirements, suggesting that regulatory expansion does not automatically create regulatory coherence (Marotta & Madnick, 2025a). Although this prior research is not focused specifically on the Middle East, it provides a useful framework for assessing how states allocate authority and manage cross-sector dependencies in AI-enabled environments.

Governance ambiguity also extends beyond domestic regulation. In their analysis of the evolving UN cybercrime treaty, Marotta and Madnick (2025b) argue that AI complicates global cyber governance by raising issues related to cross-border data flows, attribution, digital evidence, and divergent legal standards. These international dimensions reinforce the need for domestic governance structures that can coordinate across legal, technical, and institutional boundaries.

National Control and Supply-Chain Dependencies

AI governance is also shaped by global technology supply chains. Advanced AI systems often depend on cloud and compute infrastructure, proprietary models, vendor-managed software, and external technical support. These dependencies can accelerate modernization, but they also raise questions about data sovereignty, operational continuity, incident reporting, and legal accountability when critical systems are maintained or updated by external providers (ENISA, 2020; OECD, 2022).

For states pursuing rapid digital transformation, national control does not require isolation from global technology markets. It requires visibility into vendors and technical dependencies, contractual safeguards, procurement standards, security assurance mechanisms, and continuity planning. In the context of Saudi Arabia's AI strategy and critical infrastructure modernization, this issue is especially important because the Kingdom's national AI agenda combines domestic ecosystem building, regulatory development, investment, infrastructure, and international partnerships (SDAIA, 2020). The governance question is not whether international partnerships should exist, but how domestic authorities can retain sufficient visibility, auditability, and operational control over systems that support intelligence, information operations, and cyber-physical security.

Synthesis: Tensions in the Literature

The literature also reveals several unresolved tensions that motivate this paper's focus on governance ambiguity. First, the AI governance literature emphasizes transparency, accountability, lifecycle oversight, and risk-based assurance, while cybersecurity governance often prioritizes confidentiality, rapid response, operational continuity, and resilience. These priorities may conflict in national security and critical infrastructure settings, where transparency and auditability must be balanced against secrecy, system availability, and security requirements. Second, institutional theory and cybersecurity regulation literature recognize the risks of fragmented authority and overlapping mandates, but less attention has been given to how such fragmentation becomes operationally consequential when AI systems operate across

informational, cyber, and physical domains. Third, public administration and regulatory governance literature emphasizes interagency coordination, yet AI-enabled infrastructure raises a more specific problem: how responsibility should be assigned when public authorities, infrastructure operators, cloud providers, vendors, and sector regulators share partial control over system behavior and incident response. These tensions motivate the paper's focus on governance ambiguity as a potential source of systemic risk.

Theoretical Contribution: AI and the Architecture of Governance

This paper advances a structural account of AI governance in national security environments. Its central claim is that AI integration can produce governance ambiguity when algorithmic systems operate across institutional, sectoral, and technological boundaries without equivalent integration of oversight mechanisms. This ambiguity can become a form of systemic risk because failures in AI-enabled systems may involve multiple actors, unclear authority, and cascading effects across cyber-physical infrastructure.

The framework distinguishes governance ambiguity from related concepts such as regulatory fragmentation, institutional interoperability, digital governance complexity, and multi-level governance. Regulatory fragmentation refers to the dispersion of rules, standards, and mandates across institutions. Institutional interoperability focuses on the capacity of organizations and technical systems to coordinate, exchange information, and operate across boundaries. Digital governance complexity refers more broadly to the difficulty of governing data, platforms, infrastructures, and digital services across multiple actors and policy domains. Multi-level governance emphasizes interactions among national, sectoral, organizational, and transnational levels. Governance ambiguity, as used in this paper, refers more specifically to uncertainty over responsibility, accountability, decision authority, escalation pathways, and incident leadership when AI-enabled systems operate across several domains at once. The contribution of this framework is therefore not only to observe that governance is fragmented, but to explain how unclear authority can become operationally consequential in AI-enabled intelligence and cyber-physical environments.

Our framework links three bodies of literature: AI and national security, cyber-physical systemic risk, and regulatory fragmentation where the central challenge is not technical capability alone, but the architecture of governance of AI-enabled systems. This perspective frames the analysis of Saudi Arabia, where rapid digital modernization, centralized national AI strategy, and critical infrastructure dependence make the Kingdom an analytically rich case for examining how governance ambiguity emerges and might be reduced.

Research Approach and Methodology

This study uses qualitative content analysis and comparative institutional governance analysis to examine how AI-enabled intelligence, information operations, and cyber-physical security are governed under institutional complexity. Saudi Arabia is used as a focal analytical setting rather than just as a traditional empirical case study. The analysis focuses on how the distribution of responsibility across cybersecurity authorities, data governance institutions, sector regulators, infrastructure operators, and technology vendors can create ambiguity around accountability, incident leadership, and systemic risk.

The study draws on three categories of materials: scholarly literature, publicly available Saudi policy and regulatory documents, and selected policy reports. The scholarly literature provides the conceptual foundation for examining AI governance, national security, cyber-physical security, systemic cyber risk, institutional fragmentation, regulatory ambiguity, and supply-chain dependence. Saudi policy and regulatory materials, particularly those related to Vision 2030, the National Cybersecurity Authority, and the Saudi Data and Artificial Intelligence Authority, are used to analyze the Kingdom's institutional and regulatory landscape. Selected policy reports and secondary analyses are used to situate the Saudi case

within regional cyber threat dynamics, AI adoption, infrastructure resilience, and technology dependence. These materials were selected because they relate directly to the paper's focus on AI-enabled security governance, cyber-physical infrastructure, and cross-sector institutional coordination in Saudi Arabia.

The analysis proceeded by identifying recurring governance themes across these materials, including institutional responsibility, algorithmic assurance, incident governance, cyber-physical resilience, supply-chain dependence, and regulatory ambiguity. These themes were then mapped across the main Saudi governance actors to determine where oversight responsibilities are clearly assigned, where they overlap, and where ambiguity may create systemic risk. The purpose is not to provide a quantitative assessment of Saudi cyber incidents or classified national security processes, but to develop a conceptual and analytical framework for analyzing how governance ambiguity can become operationally significant when AI-enabled systems cross institutional, sectoral, and technological boundaries.

Artificial Intelligence in Intelligence and Information Operations

AI Intelligence and the Restructuring of Analytic Authority

The integration of AI into intelligence workflows marks a structural shift in contemporary security practice. Traditional intelligence cycles relied heavily on human judgment, sequential collection, and hierarchical review. AI systems now support analysts by helping filter large data streams, prioritize information, detect patterns, and flag anomalies before final human interpretation. Machine learning and natural language processing can accelerate all-source intelligence analysis by supporting data triage, synthesis, and pattern detection across large and diverse sources, including open-source and multilingual information environments (SCSP & ASPI, 2024).

This shift redistributes analytic authority. Analysts no longer only interpret information; they also depend on AI systems to determine which signals deserve attention. As Rahwan et al. (2019) argue, algorithmic systems increasingly shape the informational environments in which human decisions are made. In the Middle East, where regional rivalry, cross-border media flows, and cyber operations interact, this transformation is especially significant. AI-enabled monitoring can support the detection of mobilization patterns, disinformation campaigns, and anomalous financial behavior. Yet it also complicates oversight because responsibility for model outputs may be distributed across developers, data curators, deploying agencies, supervisory officials, and commercial providers.

Automation, Predictive Analytics, and Operational Decision-Making

AI now extends beyond analytic support into operational decision-making. Predictive risk scoring systems can guide investigative priorities, automated cyber defense tools may initiate containment protocols, and biometric systems can support identity verification in border, aviation, and large-event security. These tools increase speed and scale, but they also shift decision authority. Automated alerts may trigger action before full human review, while oversight may occur only after implementation.

This creates a governance problem. High-risk AI systems require auditability, documentation, human oversight, and clearly assigned responsibility (OECD, 2022; Mökander, 2023; Novelli et al., 2025). Operational environments, however, often prioritize speed over interpretability. In Saudi Arabia, AI has been integrated into public administration, smart infrastructure, security services, and critical infrastructure management under centralized digital transformation strategies (SDAIA, 2020). While this integration can improve coordination, it also places responsibility within layered institutional frameworks. The NCA establishes cybersecurity controls, while SDAIA provides national AI and data governance direction; sector regulators and infrastructure operators add further oversight responsibilities (NCA, 2019, 2022a, 2022b;

SDAIA, 2024a, 2024b). When AI-driven decisions intersect with cybersecurity rules, data governance requirements, and operational mandates, clear coordination becomes essential.

Generative AI and Information Operations

Generative AI has transformed the information environment by lowering the cost and speed of producing persuasive text, images, audio, and video. These tools can support legitimate communication and analysis, but they can also enable influence operations, impersonation, synthetic media, and automated content amplification (Feuerriegel et al., 2023). In a region shaped by strategic competition and multilingual digital communication, AI-generated content complicates attribution, verification, and public trust. Narratives can move across Arabic and English platforms without distinguishing authentic communication from manipulation.

Saudi Arabia's existing cybersecurity and AI governance instruments provide a foundation for managing part of this risk. The NCA's cybersecurity controls establish requirements for risk management, incident response, and protection of sensitive systems, while SDAIA's national AI strategy and public sector guidance address responsible AI use in government settings (NCA, 2019, 2022a, 2022b; SDAIA, 2020, 2024a, 2024b). Still, there are governance questions that do not fit neatly into a single mandate. Content moderation, digital oversight, disinformation response, telecommunications regulation, and data governance may involve different institutions. As generative AI is embedded in public and commercial platforms, fragmented oversight could weaken the state's ability to respond quickly and consistently.

Attribution, Strategic Stability, and Commercial Dependency

Generative AI also complicates attribution and escalation. Synthetic media may be misattributed to rival actors, while automated content amplification can intensify narratives before verification occurs. In high-stakes environments, such dynamics increase the risk of miscalculation. Existing norms for information operations were not designed for an environment in which synthetic content can be produced, localized, and distributed at scale. Governance responses therefore need to extend beyond detection toward authentication, transparency, institutional communication, and clear responsibility for response.

At the same time, many AI capabilities used in intelligence and information domains depend on commercial technology ecosystems. Cloud providers host data environments, private firms develop natural language processing tools, and external vendors maintain predictive analytics platforms. These partnerships can accelerate adoption, but they also introduce contractual, jurisdictional, and supply chain risks (ENISA, 2020; OECD, 2022). When sensitive workflows depend on externally hosted services or vendor-managed updates, questions arise about data sovereignty, operational control, incident reporting, and continuity during disruption. For Saudi Arabia, whose national AI strategy emphasizes ecosystem development and international collaboration, the issue is not whether commercial partnerships should exist, but how they should be governed to protect national resilience (SDAIA, 2020).

Synthesis: Governance Ambiguity in Intelligence Domains

AI changes intelligence and information operations by reshaping how data are collected, prioritized, interpreted, and used. These systems improve analytic speed and operational coordination, but they also distribute responsibility across agencies, developers, data authorities, cybersecurity bodies, infrastructure operators, and commercial vendors. In the Middle East, and particularly in Saudi Arabia, this creates a governance challenge as AI-enabled intelligence tools operate across cybersecurity, data governance, telecommunications, public administration, and national security domains. Clear auditing, institutional coordination, incident leadership, and supply chain oversight are necessary to reduce systemic risk.

Artificial Intelligence in Cyber-Physical Infrastructure

Cyber-Physical Convergence and AI-Enabled Critical Infrastructure

The integration of AI into cyber-physical systems marks a shift in how critical infrastructure is monitored, governed, and controlled by connecting computational processes with physical operations through sensors, industrial control systems, operational technology environments, and networked automation. In energy grids, oil and gas facilities, water desalination plants, telecommunications networks, transportation systems, and smart cities, digital control layers now mediate physical activity in real time (Radanliev et al., 2021). AI strengthens these systems through predictive maintenance, anomaly detection, resource optimization, and automated response.

This convergence creates a governance setting in which algorithmic systems are embedded directly within the operational core of national infrastructure. Rather than serving only as analytic tools, AI systems increasingly influence decisions affecting energy supply, water distribution, communications reliability, transportation flows, and emergency response. In Saudi Arabia, this transformation aligns with Vision 2030 priorities around digital modernization, smart infrastructure, and data-driven public services (SDAIA, 2020; Vision 2030 Annual Report, 2023). These systems promise efficiency and resilience, but they also concentrate risk within digital architectures that may span several institutions, operators, and vendors.

AI-Specific Vulnerabilities in Operational Technology

Operational technology environments differ from conventional IT systems. They often involve legacy equipment, real-time constraints, safety-critical processes, and limited patching windows. When AI is added to these environments, new vulnerabilities emerge. Manipulated sensor data may distort model outputs without triggering conventional cybersecurity alarms. Poisoned training data may embed hidden weaknesses in predictive systems. Model drift may cause systems to respond poorly to changing operating conditions. Vendor-managed software updates or cloud-based analytics may also introduce supply chain exposure (ENISA, 2020; Radanliev et al., 2021).

These risks matter because AI errors in operational technology can move quickly from data systems into physical processes. In an energy facility, manipulated telemetry could mislead a predictive maintenance model. In a smart city platform, a flawed optimization system could affect traffic, utilities, or emergency response. Small anomalies may cascade across sectors. Traditional cybersecurity controls, such as access management, encryption, and network segmentation, remain necessary, but they may not fully address model validation, adversarial testing, lifecycle monitoring, or accountability for automated decisions.

Systemic Risk and Cascading Failures

Cyber-physical infrastructure is highly interdependent. Energy systems support telecommunications; telecommunications networks support industrial control; transportation and logistics depend on both. Systemic cyber risk arises when disruption in one domain can propagate across others, especially in tightly coupled digital ecosystems (World Economic Forum, 2023). AI adds an algorithmic layer to this problem. Automated systems may propagate erroneous outputs at machine speed, either by taking unnecessary containment actions or by failing to detect malicious manipulation.

In energy-exporting economies such as Saudi Arabia, cyber-physical risk carries strategic significance. Disruption to energy production, electricity distribution, logistics, or telecommunications could affect domestic resilience and broader economic stability. AI governance in cyber-physical infrastructure therefore cannot be treated as a narrow technical issue. It intersects with national security, economic continuity, regulatory coordination, and international technology dependence.

Smart Cities and Centralized Data Architectures

Smart city systems illustrate the promise and risk of AI-enabled infrastructure. Sensors collect data on mobility, utilities, environment, safety, and public services. Centralized platforms aggregate these inputs, while machine learning models optimize resource allocation and operational response. In Saudi Arabia's digital transformation agenda, smart urban projects are positioned as sites of innovation, efficiency, and public sector modernization (SDAIA, 2020). However, the same centralization that enables coordination can also concentrate risk. A breach, model failure, or vendor disruption affecting a centralized smart city platform may influence multiple services at once. Governance becomes more complex because smart city systems may involve municipal authorities, national cybersecurity and data governance bodies, telecommunications regulators, infrastructure operators, and external cloud providers. Without structured coordination, overlapping mandates may slow response to AI-related incidents or obscure responsibility for failure.

Operational Resilience and Assurance

Resilience in AI-enabled cyber-physical systems requires more than technical protection. It requires assurance mechanisms that cover the full lifecycle of AI systems, including testing, documentation, monitoring, human oversight, and incident review. International guidance increasingly emphasizes risk-based governance, lifecycle monitoring, and accountability for high-risk AI systems (OECD, 2022). In Saudi Arabia, the National Cybersecurity Authority provides important baseline controls for critical systems, including cybersecurity and operational technology requirements (NCA, 2019, 2022a, 2022b).

Yet AI-specific issues, such as model robustness, explainability, data poisoning, model drift, and algorithmic accountability, may require closer integration between cybersecurity authorities, data governance bodies, sector regulators, infrastructure operators, and vendors. Effective assurance therefore depends on coordinated audits, adversarial testing, clear documentation, and predefined responsibility for AI-related incidents. Without these mechanisms, technical failures may be amplified by uncertainty over who should investigate, report, or remediate the problem.

Regional Comparison and Strategic Context

Although Saudi Arabia is the focal case, regional dynamics reinforce the importance of AI governance in cyber-physical infrastructure. Gulf Cooperation Council states are pursuing digital transformation strategies that emphasize smart infrastructure, AI-enabled services, and critical infrastructure modernization. Because energy systems, telecommunications networks, logistics corridors, and digital platforms are increasingly interconnected, AI-related vulnerabilities may extend beyond a single sector or national jurisdiction.

Regional coordination on cybersecurity has advanced, but AI-specific governance norms for cyber-physical systems remain less developed. Differences in regulatory maturity, institutional capacity, and incident response structures may create uneven resilience across the region. Strategically, AI-enabled cyber-physical systems represent both economic opportunity and security risk. States that build integrated governance frameworks may strengthen infrastructure reliability and regional confidence, while fragmented oversight may leave critical systems more exposed to disruption.

Synthesis: AI, Infrastructure, and Governance Ambiguity

AI integration into critical infrastructure creates risk because algorithmic decisions can affect physical systems such as energy, water, transportation, telecommunications, and smart city services. In Saudi Arabia, these risks are especially significant because digital transformation projects depend on multiple institutions, infrastructure operators, and technology providers. The central governance challenge is not only technical vulnerability, but unclear responsibility when AI-related failures cross institutional and sectoral boundaries.

Reducing this risk requires coordinated oversight, AI assurance, clear incident leadership, and stronger governance of vendor and supply-chain dependencies.

Saudi Arabia's Institutional and Regulatory Architecture

Saudi Arabia's AI and digital governance landscape combines centralized national strategy with distributed implementation across cybersecurity authorities, data governance institutions, digital government authorities, sector regulators, infrastructure operators, and technology partners. Under Vision 2030, AI is positioned as a strategic tool for economic diversification, public sector modernization, smart infrastructure, and national security (Vision 2030 Annual Report, 2023; SDAIA, 2020). The creation of NCA and SDAIA demonstrates the Kingdom's effort to strengthen national oversight of cybersecurity, data governance, and AI development (NCA, 2019, 2022a, 2022b; SDAIA, 2020).

Beyond NCA and SDAIA, the Digital Government Authority (DGA) is also relevant to Saudi Arabia's AI governance landscape because AI-enabled systems are increasingly embedded in public-sector digital platforms and government services. DGA serves as the national reference for digital government activities and regulates digital government work across government entities, including digital services, platforms, standards, cloud-related digital government work, and compliance (Digital Government Authority, 2023, 2025). In AI-enabled public services, DGA's role intersects with SDAIA's data and AI governance responsibilities and NCA's cybersecurity controls. This creates another area where governance ambiguity may arise unless responsibilities for platform assurance, cybersecurity, data governance, and incident response are clearly coordinated.

As AI-enabled systems are deployed across multiple sectors, governance responsibility becomes more complex. These systems often operate across several institutional domains at once, particularly when AI is embedded in cyber-physical systems, industrial control environments, cloud platforms, and IoT-based infrastructure (Radanliev et al., 2021). For example, an AI system used in an energy facility may involve cybersecurity controls, operational safety standards, data governance requirements, vendor software, cloud infrastructure, and sector regulation. In such cases, no single institution may have full visibility over the entire risk chain, reflecting broader patterns of regulatory fragmentation and systemic cyber risk (Marotta & Madnick, 2025a).

The main governance challenge is therefore not the absence of institutions, but the need for stronger integration among them. Table 1 summarizes the main institutional actors in Saudi Arabia's AI governance landscape, their current roles, and the governance issues that arise as AI systems become embedded in critical infrastructure and national security domains.

Table 1. Institutional Actors and AI Governance Issues in Saudi Arabia

Institution or actor	Current role	AI governance issue
National Cybersecurity Authority	Establishes cybersecurity controls for government entities and critical infrastructure, including Essential Cybersecurity Controls, Operational Technology Cybersecurity Controls, and Critical Systems Cybersecurity Controls (NCA, 2019, 2022a, 2022b)	Existing controls support cyber resilience, but may not fully address AI model validation, algorithmic accountability, or lifecycle AI assurance (Brundage et al., 2020; ENISA, 2020; OECD, 2022)
Saudi Data and Artificial Intelligence Authority	Leads national data and AI strategy, public sector AI guidance, and data	Its AI and data governance role must be closely coordinated with cybersecurity authorities and sector regulators

	governance frameworks (SDAIA, 2020, 2024a, 2024b)	
Digital Government Authority	Regulates digital government activities, platforms, standards, digital government services, cloud-related digital government work, and digital transformation compliance (Digital Government Authority, 2023, 2025)	AI-enabled public-sector platforms may require coordination among DGA, SDAIA, NCA, sector regulators, and vendors. Without clear coordination, responsibility for platform assurance, data governance, cybersecurity, cloud dependency, and incident response may become ambiguous
Sector regulators	Oversee energy, telecommunications, finance, transportation, and other operational sectors	They may not have complete visibility into AI risks that cut across infrastructure, vendors, cloud platforms, and data systems
Infrastructure operators	Manage critical systems such as energy, water, transportation, telecommunications, and smart city platforms	They face operational risks from automated decision systems, AI analytics, vendor tools, and external technology dependencies
Technology vendors and cloud providers	Supply AI models, analytics platforms, cloud services, software updates, and technical support	They introduce supply chain, data sovereignty, transparency, and accountability challenges (ENISA, 2020; OECD, 2022)

As shown in Table 1, AI governance in Saudi Arabia depends on coordination across multiple layers of authority. The NCA provides a foundation for protecting critical systems, SDAIA provides direction for national AI and data governance, DGA shapes the governance of public-sector digital platforms and digital government services. Sector regulators, infrastructure operators, and technology vendors add further layers of responsibility. The analytical issue is not simply that several institutions exist, but that AI-enabled systems may move across different governance logics at once: cybersecurity controls, AI and data governance, digital government platform standards, sector-specific regulation, operational safety, and vendor accountability. This layered structure becomes challenging when AI-related incidents do not fit neatly within one mandate. A disruption in an AI-enabled industrial system, for example, could involve manipulated sensor data, weak model validation, cloud dependency, vendor software updates, and operational safety concerns. Without predefined coordination mechanisms, response may be delayed by uncertainty over institutional leadership, applicable standards, and responsibility for remediation.

The analysis suggests that effective AI governance in Saudi Arabia therefore requires shared risk assessment, coordinated audits, clear incident leadership, and stronger oversight of external technology dependencies. These mechanisms are consistent with research emphasizing lifecycle AI governance, verifiable accountability mechanisms, AI auditing, and institutional coordination in high-risk digital systems (Brundage et al., 2020; Mökander, 2023; Novelli et al., 2025; OECD, 2022). Strengthening these mechanisms would help ensure that AI-enabled modernization improves national resilience rather than creating new sources of systemic exposure.

Implications: Governing AI Under Conditions of Institutional Complexity

This analysis has both academic and practical implications. Academically, the paper contributes to AI governance, cybersecurity regulation, and cyber-physical resilience literature by showing that systemic risk can emerge not only from technical vulnerabilities or fragmented regulation, but also from ambiguity in institutional responsibility. The governance ambiguity framework helps explain why AI-enabled systems that cross cybersecurity, data governance, sector regulation, infrastructure operations, and vendor-managed technology environments may create accountability gaps even when formal institutions exist.

Practically, the Saudi case highlights a governance challenge facing many countries that are developing new AI, cybersecurity, and data governance institutions. As these institutions evolve, overlapping mandates may emerge among cybersecurity authorities, data and AI authorities, sector regulators, infrastructure operators, and private technology providers. Reducing governance ambiguity requires not only the creation of formal institutions, but also mechanisms for shared risk assessment, coordinated audits, cross-domain incident protocols, vendor accountability, and clear escalation procedures before AI-related failures occur.

Building on this academic and practical framing, this section develops four interrelated governance priorities: institutional integration, algorithmic assurance, cross-domain incident governance, and strategic supply chain resilience. Table 2 summarizes these priorities and their governance problems.

Table 2. Policy Priorities for AI Governance Under Institutional Complexity

Policy priority	Governance problem	Practical mechanism
Institutional integration	AI systems cross cybersecurity, data governance, sector regulation, and operational technology domains	Joint oversight, shared risk assessments, coordinated audits, and cross-agency exercises
Algorithmic assurance	Traditional cybersecurity controls do not fully address model behavior, data integrity, drift, or automated decision making	Model validation, adversarial testing, audit trails, human oversight, and lifecycle monitoring
Cross-domain incident governance	AI incidents may involve several authorities, operators, and vendors at once	Incident leadership protocols, evidence sharing rules, escalation criteria, and simulation exercises
Supply chain resilience	Critical AI systems may depend on cloud platforms, proprietary models, vendor software, and semiconductor supply chains	Procurement safeguards, vendor transparency, continuity requirements, data controls, and diversification

Institutional Integration Beyond Formal Mandates

Saudi Arabia's creation of SDAIA and NCA shows that AI and cybersecurity are treated as national priorities. Yet formal mandates alone do not resolve governance strain. AI systems often operate across cybersecurity, data governance, infrastructure operations, and vendor-supplied platforms. Effective governance therefore requires coordination mechanisms that are tested before a crisis occurs.

Joint oversight of AI-enabled critical infrastructure would improve visibility across these domains. Coordinated audits involving SDAIA, NCA, DGA, and sector regulators could reduce gaps where algorithmic integrity falls between cybersecurity compliance and operational safety. Shared risk assessments should also map dependencies across telecommunications networks, cloud services, industrial control systems, vendors, and operational technology environments. This is necessary because AI-enabled infrastructure often functions as an interconnected system in which failures may move across sectors rather than remain isolated (Radanliev et al., 2021).

In the Saudi context, NCA and SDAIA serve as complementary anchors for AI security governance, while the Digital Government Authority (DGA) adds an important digital-government governance layer. NCA focuses on cybersecurity, operational technology, and critical systems controls, while SDAIA focuses on AI governance, data governance, and responsible public-sector AI deployment. DGA is relevant where AI-enabled systems are embedded in digital government services, platforms, cloud-based public-sector systems, and government digital transformation initiatives. Coordination among NCA, SDAIA, DGA, and sector regulators would help translate broad governance principles into practical oversight for AI-enabled infrastructure, especially in energy, smart cities, telecommunications, and public-sector platforms.

A realistic scenario illustrates the need for integration. If an AI-driven predictive maintenance system in an energy facility produces false stability readings after sensor manipulation, the incident may involve cybersecurity authorities, data governance bodies, sector regulators, infrastructure operators, and external vendors. Without agreed coordination pathways, response may be delayed by uncertainty over leadership, evidence sharing, reporting duties, and vendor responsibility. Structured exercises can clarify these roles before a real incident occurs.

Algorithmic Assurance and Lifecycle Governance

AI systems used in intelligence and cyber-physical infrastructure require governance across the full system lifecycle. Cybersecurity controls remain necessary, but they do not fully address risks tied to training data, model behavior, automated outputs, and continuous updates. Research on organizational AI governance, AI auditing, and accountability show that oversight must be built into institutional processes before deployment, not added afterward (Mäntymäki et al., 2022; Mökander, 2023; Novelli et al., 2025).

For high-risk infrastructure sectors, adversarial robustness testing would strengthen assurance practices. Industrial AI systems should be tested against data poisoning, adversarial perturbation, model drift, and abnormal feedback loops. Risks are particularly significant when AI systems interact with physical infrastructure, where faulty outputs can affect equipment, operations, and safety (ENISA, 2020; Radanliev et al., 2021).

AI-assisted decisions should also leave an auditable record, identifying intended use, data sources, testing conditions, known limits, and responsible parties. In intelligence and critical infrastructure settings, this documentation must protect sensitive information, but systems that influence high-consequence decisions should remain reviewable. Human oversight should also be calibrated to risk, with stronger review for systems used in energy operations, border security, cyber defense, and other safety-critical environments.

Cross-Domain Incident Governance

AI-related incidents rarely fit within one institutional mandate. A failure in an AI-enabled energy system may involve cybersecurity controls, corrupted data, vendor updates, cloud dependency, and operational safety at the same time. Clear incident leadership protocols are therefore needed before incidents occur. These protocols should define which authority leads, how evidence is shared, when sector regulators are involved, and how vendor responsibility is assessed (Marotta & Madnick, 2023; NIST, 2023).

Incident classification should also reflect AI-specific risks, including data integrity, model behavior, operational disruption, supply chain exposure, and cascading effects. This approach aligns with AI cybersecurity and systemic cyber-risk research, which emphasizes that failures in interconnected digital systems can spread across organizational and sectoral boundaries (ENISA, 2020). Regular cross-agency simulations would make these protocols more credible. Scenarios involving AI-related disruption in energy systems, smart cities, telecommunications, or public sector platforms would allow institutions to test reporting duties, escalation procedures, and decision authority. The goal is faster response and clearer accountability.

Strategic Supply Chain Resilience

Supply chain resilience does not require isolation from global technology markets. It requires disciplined management of dependence. AI-enabled infrastructure often relies on cloud providers, semiconductor supply chains, proprietary models, vendor-managed software, and external technical support. These partnerships can accelerate modernization, but they also create exposure when critical systems depend too heavily on a small number of external providers (OECD, 2022; ENISA, 2020).

The analysis suggests that AI procurement in Saudi Arabia should be treated as a governance issue, not only a technical acquisition process. Contracts with external providers should include requirements for transparency, security testing, vulnerability reporting, incident cooperation, data access, and continuity of service. For critical infrastructure, procurement standards should also address data localization, cross-border data transfer, vendor update procedures, and exit options if a provider becomes unavailable or strategically risky.

Diversification is equally important. Critical systems should avoid unnecessary dependence on a single vendor, model, cloud platform, or software pipeline. Where external reliance is unavoidable, domestic authorities should retain enough visibility to audit risk, investigate incidents, and maintain continuity during disruption. The goal is resilience with control, not technological isolation.

Resilience by Design

Resilience must be built into institutions, not only technical systems. Saudi Arabia's centralized national strategy and sector-based regulation provide an important foundation, but oversight mechanisms must work when AI systems cross institutional, technical, and sector boundaries. Institutional integration, lifecycle assurance, incident leadership, and supply-chain resilience can help ensure that AI-enabled modernization strengthens national resilience rather than creating new sources of systemic exposure.

Limitations and Future Research

This study has several limitations. First, the analysis relies primarily on publicly available documents, peer-reviewed literature, and secondary policy sources. Because national security governance often involves classified procedures, internal coordination mechanisms, and non-public incident-response processes, the paper cannot fully assess how AI-related governance operates in practice. Second, the study does not include interviews, fieldwork, compliance audits, or direct empirical measurement of institutional performance. As a result, the findings should be understood as conceptual and analytical rather than as evidence of specific governance failures. Third, while Saudi Arabia provides an analytically rich case, the findings are shaped by its specific institutional, strategic, and infrastructural context. The argument may be relevant to other Gulf Cooperation Council (GCC) countries and broader Middle Eastern countries pursuing AI-enabled digital transformation, but it should not be treated as automatically generalizable without further comparative analysis. Future research could apply this framework through interviews, comparative Gulf case studies, sector-specific analysis, or incident-based research on AI-enabled cyber-physical risk.

Conclusion

Artificial intelligence is now on the frontlines of national security. It shapes how states collect intelligence, manage information environments, operate critical infrastructure, and respond to emerging threats. In the Middle East, and particularly in Saudi Arabia, this transformation is unfolding alongside rapid digital modernization, smart infrastructure development, and growing cyber-physical integration.

This paper argued that the central risk is not only technical vulnerability, but governance ambiguity. As AI systems operate across intelligence, cybersecurity, energy, telecommunications, data governance, and vendor ecosystems, authority and accountability become more difficult to assign. Saudi Arabia has made significant progress through institutions such as the National Cybersecurity Authority and the Saudi Data and Artificial Intelligence Authority, as well as cybersecurity controls and AI governance guidelines. Yet formal mandates alone are insufficient when AI systems cross institutional and sectoral boundaries.

By framing governance ambiguity as a source of systemic risk, this paper contributes to debates on AI, national security, and cyber-physical resilience. The Saudi case shows that long-term stability will depend on clearer institutional coordination, defined incident leadership, stronger AI assurance, and oversight mechanisms that evolve with technological change.

Acknowledgement

This research was supported, in part, by the Cybersecurity at MIT Sloan (CAMS) initiative (<https://cams.mit.edu>)

References

- Black, J., Eken, M., Parakilas, J., Dee, S., Ellis, C., Suman-Chauhan, K., Bain, R. J., Fine, H., Aquilino, M. C., Lebet, M., & Palicka, O. (2024). *Strategic competition in the age of AI: Emerging risks and opportunities from military use of artificial intelligence*. RAND Corporation. <https://doi.org/10.7249/RR3295-1>
- Brundage, M., Avin, S., Clark, J., Toner, H., Eckersley, P., Garfinkel, B., Dafoe, A., Scharre, P., Zeitzoff, T., Filar, B., Anderson, H., Roff, H., Allen, G. C., Steinhardt, J., Flynn, C., Ó hÉigearthaigh, S., Beard, S., Belfield, H., Farquhar, S., ... Amodei, D. (2018). *The malicious use of artificial intelligence: Forecasting, prevention, and mitigation*. <https://doi.org/10.17863/CAM.22520>
- Brundage, M., Avin, S., Wang, J., Belfield, H., Krueger, G., Hadfield, G., Khlaaf, H., Yang, J., Toner, H., Fong, R., Maharaj, T., Koh, P. W., Hooker, S., Leung, J., Trask, A., Bluemke, E., Lebensold, J., O'Keefe, C., Koren, M., ... Anderljung, M. (2020). Toward trustworthy AI development: Mechanisms for supporting verifiable claims. <https://arxiv.org/abs/2004.07213>
- Bütthe, T., Djeflal, C., Lütge, C., Maasen, S., & von Ingersleben Seip, N. (2022). Governing AI: Attempting to herd cats? Introduction to the special issue on the governance of artificial intelligence. *Journal of European Public Policy*, 29(11), 1721–1752. <https://doi.org/10.1080/13501763.2022.2126515>
- Digital Government Authority. (2023). Digital government regulatory framework. Government of Saudi Arabia. https://dga.gov.sa/en/regulatory_framework
- Digital Government Authority. (2025). About DGA. Government of Saudi Arabia. <https://dga.gov.sa/en/about-DGA>
- European Union Agency for Cybersecurity. (2020). Artificial intelligence cybersecurity challenges. ENISA. <https://www.enisa.europa.eu/publications/artificial-intelligence-cybersecurity-challenges>
- Feuerriegel, S., DiResta, R., Goldstein, J. A., Kumar, S., Lorenz-Spreen, P., Tomz, M., & Pröllochs, N. (2023). Research can help to tackle AI-generated disinformation. *Nature Human Behaviour*, 7(11), 1818–1821. <https://doi.org/10.1038/s41562-023-01726-2>
- Hendawy, M., & Kumar, N. (2024). *AI in the national: AI strategies of the Arab region*. Arab Reform Initiative. <https://s3.eu-central-1.amazonaws.com/storage.arab-reform.net/ari/2024/11/13112245/2024-08-09-EN-Menna.pdf>

- Mäntymäki, M., Minkkinen, M., Birkstedt, T., & Viljanen, M. (2022). Defining organizational AI governance. *AI and Ethics*, 2, 603–609. <https://doi.org/10.1007/s43681-022-00143-x>
- Marotta, A., Madnick, B. F., & Madnick, S. (2023). Navigating the wave of cybersecurity regulations: A systematic analysis of emerging regulatory developments. *AMCIS 2023 Proceedings*, 12. https://aisel.aisnet.org/amcis2023/sig_sec/sig_sec/12/
- Marotta, A., & Madnick, S. (2023). Regulating cyber incidents: A review of recent reporting requirements. In *Proceedings of the 20th International Conference on Security and Cryptography (SECRYPT)* (pp. 410–416). SciTePress. <https://doi.org/10.5220/0012086000003555>
- Marotta, A., & Madnick, S. (2025a). Analyzing and categorizing emerging cybersecurity regulations. *ACM Computing Surveys*, 58(2), Article 51. <https://doi.org/10.1145/3757318>
- Marotta, A., & Madnick, S. (2025b). The UN cybercrime treaty and AI: Navigating the intersection of technology and global policy. *Issues in Information Systems*, 26(4), 1–16. https://doi.org/10.48009/4_iis_2025_101
- Mökander, J. (2023). Auditing of AI: Legal, ethical and technical approaches. *Digital Society*, 2, Article 49. <https://link.springer.com/article/10.1007/s44206-023-00074-y>
- National Institute of Standards and Technology. (2023). *Artificial intelligence risk management framework (AI RMF 1.0)* (NIST AI 100-1). U.S. Department of Commerce. <https://doi.org/10.6028/NIST.AI.100-1>
- National Cybersecurity Authority. (2019). *Critical systems cybersecurity controls* (CSCC–1:2019). Government of Saudi Arabia. https://cdn.nca.gov.sa/api/files/public/upload/f15af01c-dc59-4281-95e2-03a770655937_Critical-Systems-Cybersecurity-Controls.pdf
- National Cybersecurity Authority. (2022a). *Essential cybersecurity controls (ECC 2:2022)*. Government of Saudi Arabia. <https://nca.gov.sa/en/regulatory-documents/>
- National Cybersecurity Authority. (2022b). *Operational technology cybersecurity controls (OTCC 1:2022)*. Government of Saudi Arabia. <https://nca.gov.sa/en/regulatory-documents/>
- Novelli, C., Hacker, P., Morley, J., Trondal, J., & Floridi, L. (2025). A robust governance for the AI Act: AI Office, AI Board, Scientific Panel, and National Authorities. *European Journal of Risk Regulation*, 16(2), 566–590. <https://doi.org/10.1017/err.2024.57>
- OECD. (2022). *OECD framework for the classification of AI systems* (OECD Digital Economy Papers No. 323). OECD Publishing. <https://doi.org/10.1787/cb6d9eca-en>
- Oxford Analytica. (2024). Cyberattacks on the Middle East will continue to rise. *Emerald Expert Briefings*. <https://doi.org/10.1108/OXAN-DB291422>
- Radanliev, P., De Roure, D., Van Kleek, M., Santos, O., & Ani, U. (2021). Artificial intelligence in cyber-physical systems. *AI & Society*, 36(3), 783–796. <https://doi.org/10.1007/s00146-020-01049-0>
- Rahwan, I., Cebrian, M., Obradovich, N., Bongard, J., Bonnefon, J.-F., Breazeal, C., Crandall, J. W., Christakis, N. A., Couzin, I. D., Jackson, M. O., Jennings, N. R., Kamar, E., Kloumann, I. M.,

- Larochelle, H., Lazer, D., McElreath, R., Mislove, A., Parkes, D. C., Pentland, A., ... Wellman, M. P. (2019). Machine behaviour. *Nature*, 568(7753), 477–486. <https://doi.org/10.1038/s41586-019-1138-y>
- Saudi Data and Artificial Intelligence Authority. (2020). National Strategy for Data & AI: Realizing our best tomorrow: Strategy narrative. https://wp.oecd.ai/app/uploads/2021/12/Saudi_Arabia_National_Strategy_for_Data_and_AI_2020.pdf
- Saudi Data and Artificial Intelligence Authority. (2024a). *General AI guidelines for government*. <https://sdaia.gov.sa/en/SDAIA/about/Files/GenAIGuidelinesForGovernmentENCompressed.pdf>
- Saudi Data and Artificial Intelligence Authority. (2024b). *Generative AI guide for the public sector*. <https://sdaia.gov.sa/en/SDAIA/about/Files/GenerativeAIPublicEN.pdf>
- Special Competitive Studies Project, & Australian Strategic Policy Institute. (2024). *The future of intelligence analysis: AI and human-machine teaming*. Special Competitive Studies Project. <https://www.scpai.org/wp-content/uploads/2024/09/AI-The-Future-of-Intelligence-Analysis-SCSP-ASPI-Report.pdf>
- Vision 2030 Annual Report. (2023). *Vision 2030 annual report 2023*. Kingdom of Saudi Arabia. <https://www.vision2030.gov.sa>
- World Economic Forum. (2023). *Global cybersecurity outlook 2023*. <https://www.weforum.org/publications/global-cybersecurity-outlook-2023/>