مجلس الأمن السيبراني
CYBER SECURITY COUNCIL

United Arab Emirates

NATIONAL CYBER SECURITY POLICY FOR ARTIFICIAL INTELLIGENCE

November 2025 Version 1.1

DISCLAIMER

This document is the sole property of and is approved by the UAE Cyber Security Council.

The UAE Cyber Security Council reserves the right to amend, add, or delete any section of this Policy.

Neither the whole nor part of this document shall be reproduced without the prior written consent of the UAE Cyber Security Council.

VERSION CONTROL

Version	1.0
Date:	September 2025
Prepared by:	CSC
Amendment Content:	First Release

Version	1.1
Date:	November 2025
Prepared by:	CSC
Amendment Content:	Revised the policy structure and updated policy statements to specifically address cybersecurity considerations for artificial intelligence.

Table of Contents

SECTION 1

Introduction

INTRODUCTION

The proliferation and integration of Artificial Intelligence (AI) systems are reshaping economies, ushering in a new era of innovation and transformation across industries. As AI-driven applications become embedded in critical functions and decision-making processes, the need to ensure the security of AI systems has gained unprecedented prominence. With the intricate interplay of data, algorithms, and autonomous decision-making that AI entails, the need to fortify these systems against potential vulnerabilities and unauthorized access has become paramount. In the face of escalating cyber security threats and the potential consequences of AI-related breaches, safeguarding the confidentiality, integrity and availability of AI systems is not only a necessity but a fundamental responsibility.

The Council has established this policy to enhance the cyber security of artificial intelligence, aligned with the UAE's national priority to be a global leader in cyber security; and enhance the cyber security posture of organizations and individuals within the UAE using artificial intelligence.

This policy reinforces principles and objectives of the *UAE Charter for the Development and Use of Artificial Intelligence*, ensuring that AI systems are developed, implemented and managed with strong cyber security and protection measures. It supports the UAE's vision, under the AI Strategy, to become a global hub for responsible and secure AI adoption.

1.1 Purpose

This policy aims to strengthen the secure adoption and use of Artificial Intelligence and Machine Learning (AI/ML) technologies in the UAE by establishing requirements to address known and emerging cyber threats. Broader considerations such as AI ethics, fairness, and trustworthiness fall outside the scope of this policy.

1.2 Scope & Applicability

The applicability of this policy is outlined below:

- The National Cyber Security Policy For Artificial Intelligence applies to Ministries & Federal Authorities and Non-Government Critical Information Infrastructure (CII) Entities.
- Where there exists an Emirate CIIP Program, the applicability assessment, enforcement and monitoring of the National Cyber Security Policy For Artificial Intelligence for Emirate Government Entities and Non-Government CII entities will be carried out by the respective Emirate Lead.
- In the absence of an Emirate CIIP Program, the National Cyber Security Policy For Artificial Intelligence applies to Emirate Government Entities.

**Refer to the National Cyber Security Governance Framework for further details on terminologies.*

1.3 Guiding Principles

The following five guiding principles are laid out to provide decision-makers with foundational elements to drive secure AI adoption, implementation, and operations in the UAE. These principles assist entities to make policy, operational, and procurement decisions in line with the requirements detailed in this document.

Risk Based Approach

- Exercise sound judgement considering their operating environment and risk appetite while adopting the measures outlined in this policy.

Future Ready

- Potential future changes and developments in the relevant field or industry are considered.

Aligned to International Best Practices

- Global good practice frameworks are leveraged to provide security assurance and drive compliance efficiencies.

Transparent and Collaborative

- Open sharing of information, good practices, incident reporting and threat intelligence is encouraged.

Lawful

- Applicable laws, regulations and standards are considered while adopting AI technologies.

1.4 Exception Approval

A policy exception may be granted by the relevant regulator (Emirate or Sector Lead) or the Cyber Security Council (CSC) under special circumstances.

Exceptions will be reviewed on a case-by-case basis, and their approval is not guaranteed.

SECTION 2

Cyber Security Policy for Artificial Intelligence

The following section outlines the policy domains and cyber security requirements applicable to the entities in the UAE.

2.1. AI Cyber Security Governance

Policy Statements	Version	1.1

2.1.1 The entity shall ensure secure design, development, and operation of AI/ML systems by establishing and maintaining comprehensive, enforceable policies and procedures.

2.1.2 The entity shall identify, assess, and mitigate unique cyber security risks associated with AI/ML systems in a systematic and ongoing manner.

2.1.3 The entity shall identify, assess, and manage cyber security risks associated with the supply chain of AI/ML systems, ensuring appropriate controls are implemented to mitigate threats arising from third-party components, services, and dependencies.

2.1.4 The entity shall control and manage changes to AI/ML systems in a way that maintains system security.

2.2. Infrastructure & Application Security

Policy Statements **Version** 1.1

2.2.1 The entity shall maintain an up-to-date inventory of AI/ML assets and apply cyber security controls commensurate with the sensitivity, criticality, and value of each asset.

2.2.2 The entity shall ensure AI/ML systems are securely and consistently configured in accordance with approved security baselines and configuration standards.

2.2.3 The entity shall ensure that security patches and updates for AI/ML systems are identified, tested, and applied in a timely manner to maintain system security and integrity.

2.2.4 The entity shall implement a systematic process to identify, assess, and remediate vulnerabilities in AI/ML systems to enhance their resilience against cyber security threats.

2.2.5 The entity shall ensure that AI/ML applications are securely developed, tested, and maintained throughout their lifecycle in accordance with established security practices.

2.2.6 The entity shall apply robust network security controls to AI/ML systems to ensure confidentiality, integrity, and availability, tailored to the systems' specific needs and characteristics.

2.3 Algorithm Security

Policy Statements **Version** 1.1

2.3.1 The entity shall integrate cyber security into the design and development of AI/ML models from the earliest stages, to proactively mitigate security risks and avoid retrofitting controls after deployment.

2.3.2 The entity shall establish and maintain security measures to protect AI/ML models from potential threats, ensuring their integrity and availability throughout the model lifecycle.

2.3.3 The entity shall ensure the security of the AI/ML inference process by protecting against potential threats and maintaining the integrity and confidentiality of inference data.

2.3.4 The entity shall enforce robust access controls on AI/ML training data to ensure data confidentiality, integrity, and privacy, while supporting accountability and auditability.

2.3.5 The entity shall enforce stringent data protection measures for AI/ML data, both at rest and in motion, to safeguard its confidentiality and integrity.

2.4. Operational Security

Policy Statements **Version** 1.1

2.4.1 The entity shall tailor cyber security measures for AI/ML systems according to their specific application domains, implementing context-appropriate safeguards.

2.4.2 The entity shall ensure the operational resilience of AI/ML systems by incorporating redundancy and failover mechanisms to maintain availability during disruptions.

2.4.3 The entity shall proactively address AI/ML system through business continuity and disaster recovery planning processes and regularly testing restoration capability and continuity scenarios.

2.4.4 The entity shall ensure the security of AI/ML systems through comprehensive testing and validation, providing assurance of their resilience against potential threats.

2.5 Adversarial AI Attacks

Policy Statements	Version	1.1

2.5.1 The entity shall promote awareness and understanding of potential AI/ML attacks, fostering a proactive security culture through education, training, and preparedness for effective incident response.

2.5.2 The entity shall deploy comprehensive defense strategies for AI/ML systems, including proactive security measures, continuous monitoring, regular testing, and robust incident response capabilities.

2.6 AI Monitoring & Response

Policy Statements	Version	1.1

2.6.1 The entity shall establish an AI/ML security analytics framework that utilizes advanced technologies and diverse data sources to enable real-time threat detection, historical analysis, and predictive security insights.

2.6.2 The entity shall establish a robust and efficient incident reporting and management system for AI/ML cyber security incidents that ensures prompt response, compliance with legal and contractual obligations, and supports continuous learning and improvement.

2.6.3 The entity shall establish a robust and comprehensive digital forensics framework tailored to AI/ML cyber security incidents, enabling effective investigation and analysis.

SECTION 3

Implementation Guidance

The following section details the implementation approach for the requirements across the policy domains.

3.1 AI Cyber Security Governance

3.1.1 Cyber Security Policies & Procedures

Implementation Guidance **Version** 1.1

- The entity should develop, implement, and maintain a set of comprehensive policies and procedures that specifically address the unique cyber security considerations associated with AI/ML systems.
- The AI/ML cyber security policy should define roles and responsibilities for security within the AI/ML lifecycle, providing clarity on who is responsible for implementing, monitoring, and enforcing these policies.
- The entity should encourage collaboration between security, development, operations, and business teams involved in AI/ML through shared resources, platforms, or regular interactions to ensure alignment and consistency.
- AI/ML cyber security policies and procedures should be communicated and made accessible to all relevant personnel, ensuring awareness and understanding.
- AI/ML cyber security policies and procedures should be reviewed and updated regularly to address emerging threats, technological advancements, and changes in legal and regulatory requirements relevant to AI/ML.

3.1 AI Cyber Security Governance

3.1.2 Cyber Risk Management for AI/ML

Implementation Guidance **Version** 1.1

- The entity should have a formal cyber security risk management process that specifically addresses the unique cyber security risks associated with AI/ML.
- The AI/ML cyber security risk management process should be integrated into the entity's broader risk management program, ensuring that cyber security risks associated with AI/ML systems are classified, prioritized, and considered as part of the organization's overall risk profile.
- The entity should conduct cyber security risk assessments at regular intervals and at key stages in the AI/ML lifecycle, such as during model design, data collection, training, and deployment.
- The entity should assess and mitigate cyber security risks related to cloud-based AI solutions, particularly those that utilize client data under the End User License Agreement (EULA) for unsupervised learning in the background.
- The entity should regularly review and update the AI/ML cyber security risk management process to address new threats, vulnerabilities, and risk mitigation strategies.

3.1 AI Cyber Security Governance

3.1.3 AI Supply Chain Security

Implementation Guidance **Version** 1.1

- The entity should establish a comprehensive AI supply chain security strategy to manage cyber security risks associated with third-party AI/ML components and services, from sourcing to decommissioning.
- The entity should ensure a clear understanding of its AI/ML supply chain, including the source of components, models, training data, and libraries used in its AI/ML systems.
- The entity should document, develop, and maintain a record of all components used in AI/ML systems, including their sources, models, training data, libraries, versions, and any known vulnerabilities.
- The entity should conduct regular security assessments of its AI/ML supply chain, including vendors, developers, and other third parties contributing to the AI/ML system. The entity may alternatively leverage independent third-party assurance reports (e.g., SOC 2, ISO/IEC 27001) or recognized certifications to inform its assessment.
- The entity should ensure that contractual agreements with vendors and other third parties address cyber security requirements for AI/ML systems.

3.1 AI Cyber Security Governance

3.1.4 Change Management and Reporting

Implementation Guidance **Version** 1.0

- The entity should establish and maintain a change management process for AI/ML systems, ensuring any manual or human-initiated modifications, upgrades, or alterations to models, algorithms, training data, or system configurations are properly documented, reviewed, and approved.
- This process should consider the potential cyber security impacts of changes, including changes that could affect the confidentiality, integrity, or availability of AI/ML systems or the data they process.
- All manual changes should be tested in a secure environment before deployment to identify and address any unexpected impacts or vulnerabilities introduced by the change.
- Where possible, the entity should have a rollback plan in place to restore the system to its previous state in case a change introduces significant issues or risks.

3.2 Infrastructure & Application Security

3.2.1 Asset Management for AI/ML Systems

Implementation Guidance **Version** 1.1

- The entity should establish and maintain an accurate and comprehensive inventory of all AI/ML system components, including hardware, software, data sets, models, and algorithms (experimental and deployment).
- The asset inventory should be updated near real-time to account for changes such as additions, modifications, and retirements of assets.
- The entity should classify all AI/ML assets according to factors such as their criticality to business operations, associated risk, and sensitivity and assign appropriate cyber security controls based on this classification.
- The entity should establish clear ownership and accountability for each AI/ML asset, including responsibility for its security and compliance.
- The entity should have a process to decommission and securely dispose of AI/ML assets when they reach their end of life, ensuring that appropriate data is securely wiped and cannot be recovered.

3.2 Infrastructure & Application Security

3.2.2 Security Configuration Management

Implementation Guidance **Version** 1.1

- The entity should establish and enforce secure configuration baselines for all AI/ML supporting infrastructure and applications.
- The entity should periodically review and update AI/ML secure configuration baselines to ensure continued effectiveness against evolving threats and changes in system architecture or technology.
- The entity should establish procedures for the secure backup and recovery of AI/ML system configurations to minimize downtime and data loss in the event of a security incident or system failure.

3.2 Infrastructure & Application Security

3.2.3 Patch Management

Implementation Guidance **Version** 1.0

- The entity should implement a robust patch management process to promptly address known vulnerabilities in AI/ML infrastructure and applications.
- The entity should regularly monitor for new security patches and updates from manufacturers, vendors, and trusted third-party sources.
- All patches and updates should undergo testing and verification in a controlled environment before deployment in production systems to prevent unintended impacts on system performance and availability.
- The entity should prioritize the application of patches and updates based on the severity of the vulnerabilities they address and the criticality of the affected systems.
- The entity should maintain a complete and up-to-date inventory of all AI/ML systems requiring patch management.

3.2 Infrastructure & Application Security

3.2.4 Vulnerability Management for AI/ML Systems

Implementation Guidance **Version** 1.0

- The entity should implement a robust vulnerability management program to identify, assess, and remediate security vulnerabilities in AI/ML systems.
- The entity should employ an appropriate combination of testing tools, methodologies, and external audits to identify vulnerabilities in AI/ML infrastructure and applications.
- The entity should promptly assess and prioritize identified vulnerabilities based on their potential impact and exploitability and establish a timeline for remediation actions.
- The entity should regularly review and update the vulnerability management program to align with changes in AI/ML technologies, threat landscape, and regulatory changes.

3.2 Infrastructure & Application Security

3.2.5 Application Security

Implementation Guidance **Version** 1.1

- The entity should implement a secure software development lifecycle (SDLC) for AI/ML applications that includes security requirements definition, secure coding practices, security testing, and post-deployment security reviews.
- The entity should perform security testing, including code reviews and static/dynamic analysis, to identify and remediate potential security flaws in AI/ML applications prior to deployment.
- The entity should implement a robust access control mechanism in AI/ML applications to limit the access to AI/ML data and functions based on the principle of least privilege.
- The entity should implement strong encryption and other appropriate security controls to protect data processed and generated by AI/ML applications.
- The entity should provide training to developers on secure coding practices and common security threats in AI/ML application development.
- The entity should maintain an up-to-date inventory of AI/ML applications, including their dependencies, and monitor them for any changes that could introduce cyber security risks.

3.2 Infrastructure & Application Security

3.2.6 Network Security for AI/ML Infrastructure

Implementation Guidance **Version** 1.0

- The entity should establish and enforce a network security policy specifically tailored for AI/ML infrastructure.
- The entity should segment high risk AI/ML systems from the rest of the network to minimize the risk of lateral movement in case of a breach.
- Network traffic to and from AI/ML systems should be securely encrypted and continuously monitored for any signs of intrusion or abnormal behavior.
- The entity should implement firewall rules and Intrusion Detection/Prevention Systems (IDS/IPS), or other security technologies specifically configured for the unique network behaviors and requirements of AI/ML systems.
- The entity should ensure that any remote access to AI/ML infrastructure is secure and controlled, utilizing secure protocols and methods such as multi-factor authentication.
- The entity should routinely conduct vulnerability assessments and penetration tests on the AI/ML network infrastructure to identify and remediate potential security weaknesses.
- The entity should ensure that network security logs for AI/ML infrastructure are stored, protected, and analyzed in accordance with the entity's log management policy.

3.3. Algorithm Security

3.3.1 Security by Design for AI/ML Models

Implementation Guidance **Version** 1.1

- The entity should ensure that cyber security processes are integrated into AI/ML models from the earliest stages of their design and development.
- The entity should perform threat modeling during the design and development of AI/ML models to identify potential cyber security threats and design appropriate mitigating controls.
- The entity should ensure that AI/ML models comply with established cyber security standards and best practices, including data minimization, least privilege, and separation of duties principles.

3.3 Algorithm Security

3.3.2 AI/ML Model Security

Implementation Guidance **Version** 1.1

- The entity should protect AI/ML models from unauthorized access and modification, by implementing measures like secure model storage, stringent access controls, and encryption both at rest and in motion.
- The entity should secure AI/ML models to an appropriate level during their entire lifecycle - from initial training and testing, through validation, to deployment and operational phases, employing strategies such as federated learning, or homomorphic encryption as required.
- The entity should implement safeguards to protect against adversarial attacks, such as adversarial training, robust input validation, and continuous monitoring for anomalous model behavior indicative of potential cyber security incidents.

3.3 Algorithm Security

3.3.3 Inference Security

Implementation Guidance **Version** 1.1

- The entity should implement measures to secure the process of making predictions with AI/ML models, ensuring the integrity, confidentiality, and availability of inference data.
- The entity should adopt strategies to mitigate threats against inference security, such as model inversion attacks, membership inference attacks, and adversarial attacks.
- The entity should deploy defensive measures, including but not limited to robust input validation, to protect against inference attacks aiming to exploit or reverse-engineer AI/ML models.

3.3 Algorithm Security

3.3.4 Access to Training Data Control

Implementation Guidance **Version** 1.0

- The entity should enforce strict access control measures for AI/ML training data, ensuring that only authorized individuals and processes can access this data and only it for its intended purpose.
- Where possible, the entity should implement robust mechanisms to track and log all access to AI/ML training data, supporting auditability and accountability.
- The entity should apply data minimization practices to AI/ML training data, collecting and retaining only the necessary data for training purposes.
- The entity should ensure that AI/ML training data is protected against unauthorized modification, deletion, and exfiltration.
- Anonymization or pseudonymization should be applied when classified data is used for AI/ML training. Additionally, the entity should implement measures such as deleting datasets after training or encrypting data to protect training data.

3.3 Algorithm Security

3.3.5 AI/ML Data Protection at Rest and in Motion

Implementation Guidance **Version** 1.1

- The entity should enforce robust encryption measures for AI/ML data both at rest and in motion to protect against unauthorized access and exfiltration based on the classification of the data.
- The entity should implement secure methods for the transfer of AI/ML data, using only approved and secure protocols based on the classification of the data.
- The entity should enforce strong key management practices for encryption keys used to protect AI/ML data.

3.4 Operational Security

3.4.1 Context-Specific AI/ML Security

Implementation Guidance **Version** 1.1

- The entity should identify and address the unique cyber security requirements of AI/ML systems based on their specific application context (e.g. medical, automotive, industrial, customer service, etc.).
- For safety-critical applications (e.g. self-driving cars, medical diagnostics), the entity should prioritize robustness, reliability, and fail-safe mechanisms in AI/ML design and deployment.
- For privacy-sensitive applications (e.g. personal assistants, customer service bots), the entity should prioritize data protection, privacy, and transparency in AI/ML design and deployment.
- The entity should follow industry-specific guidelines, best practices, and regulations for AI/ML cyber security, as applicable to their application context.
- The entity should provide comprehensive cyber security training to AI/ML operators (if any), equipping them with the necessary knowledge and skills to operate the AI/ML system securely in its application context.

3.4 Operational Security

3.4.2 Fail-Safe & Backup for AI/ML Systems

Implementation Guidance	Version	1.0

- The entity should implement robust fail-safe mechanisms for AI/ML systems to prevent failures and ensure system stability under unexpected conditions.
- The entity should design AI/ML systems, based on their criticality, with failover mechanisms to ensure system availability during disruptions, such as redundant infrastructure deployment in geographically diverse data centers, load balancing, and automatic failover and switching.
- The entity should employ backup measures for AI/ML systems as appropriate, including periodic and secure backup of training data, model parameters, and system configurations.
- The entity should regularly test and validate the effectiveness of fail-safe and backup measures to ensure their readiness in the event of a system failure or other disruptions.

3.4 Operational Security

3.4.3 Continuity Planning for AI/ML Systems

Implementation Guidance **Version** 1.0

- The entity should consider AI/ML systems in business continuity planning and disaster recovery strategies.
- The entity should establish measures to ensure the continuity of AI/ML systems, considering potential disruptions from both technical and non-technical factors, including data corruption, loss of critical resources, and sustained outages.
- In cases where AI/ML systems are critical to operations, redundant systems or alternative processes should be available and ready for immediate deployment.
- The entity should have a defined process to recover AI/ML systems from disruptions and outline key personnel, processes, resources, and critical dependencies necessary for the recovery of AI/ML systems.
- The entity should periodically test the continuity and recovery plan for AIML systems and train relevant staff to ensure they are adequately prepared.

3.4 Operational Security

3.4.4 Testing and Validation of AI/ML Systems

Implementation Guidance **Version** 1.0

- The entity should establish and maintain a robust framework for the testing and validation of AI/ML systems, ensuring their accuracy, reliability, and performance meet defined criteria.
- All AI/ML systems should undergo rigorous testing before deployment to verify their functionality, security, and resilience against potential threats.
- The testing process should consider various scenarios and inputs, including edge cases, to evaluate the AI/ML systems' behavior under different conditions.

3.5 Adversarial AI Attacks

3.5.1 Understanding and Anticipating AI/ML Attacks

Implementation Guidance **Version** 1.1

- The entity should actively work to understand potential AI/ML attacks, including their methods, implications, and possible mitigations, to anticipate and prepare for such events.
- The entity should be aware of and educated about common types of AI/ML attacks, such as adversarial attacks, poisoning attacks, and model inversion attacks.
- The entity should engage in threat intelligence and information sharing with other organizations, research institutions, and authorities to stay updated about the latest trends and developments in AI/ML attacks.
- The entity should implement user training and awareness programs to ensure that staff and stakeholders understand the risks associated with AI/ML systems and are equipped to recognize and respond to potential threats.

3.5 Adversarial AI Attacks

3.5.2 Defending Against AI/ML Attacks

Implementation Guidance **Version** 1.0

- The entity should implement robust defensive techniques to protect AI/ML systems from attacks, including, but not limited to, anomaly detection, adversarial training, defensive distillation, and feature squeezing.
- Security measures should be integrated during the development and deployment of AI/ML models to minimize the potential impact of attacks.
- Regular security testing, including penetration testing and red teaming exercises, should be conducted on AI/ML systems to identify potential weaknesses and implement necessary defenses.
- The entity should continuously monitor the performance and behavior of AI/ML systems to detect and respond to potential anomalies indicative of an attack.
- AI/ML systems should be designed and configured to resist, contain, and recover from attacks, minimizing potential damage and disruption.

3.6 AI Monitoring & Response

3.6.1 AI/ML Security Analytics

Implementation Guidance **Version** 1.1

- The entity should implement advanced security analytics capabilities commensurate with the criticality of the AI/ ML system, to facilitate real-time monitoring, detection, and response to threats targeting AI/ML systems.
- Security analytics should encompass various data sources, including system logs, network traffic, and user activity to provide a holistic view of the security posture of AI/ML systems.
- The entity should employ statistical and behavioral analysis techniques to identify anomalies and potential security incidents in AI/ML systems such as abnormal input patterns that may indicate adversarial attacks, unexpected model behavior, data poisoning, or model extraction attempts.
- The entity should maintain historical security data for AI/ML systems to aid in trend analysis, incident investigation, and the development of predictive security analytics.
- The entity should integrate external threat intelligence feeds with internal data sources to enhance threat detection capabilities.

3.6 AI Monitoring & Response

3.6.2 Incident Reporting and Management for AI/ML

Implementation Guidance **Version** 1.1

- The entity should establish a robust incident reporting and management process for AI/ML security incidents to ensure they are appropriately identified, logged, investigated, and resolved.
- This process should include clear guidelines on identifying and classifying incidents, timely reporting mechanisms, designated roles and responsibilities for incident response, and procedures for post-incident analysis and learning.
- The entity should ensure AI/ML security incidents are reported in a timely manner to all relevant stakeholders, including internal teams, third-party service providers, and regulatory bodies, in line with applicable laws and regulations, and contractual agreements.
- The entity should consider automated response mechanisms for AI/ML security incidents to mitigate the impact of attacks and prevent further escalation such as isolating affected systems, adjusting security controls, or blocking network traffic, based on predefined conditions and incident severity levels.
- In cases where automated response mechanisms may not be suitable, or fully effective, the entity should implement a fail-safe mechanism that allows for timely and informed human intervention.
- The entity should conduct periodic reviews and updates to the automated response rules based on lessons learned from exercises and actual incidents to ensure continuous improvement and alignment with emerging threats.

3.6 AI Monitoring & Response

3.6.3 Digital Forensics for AI/ML Security Incidents

Implementation Guidance	Version	1.1

- The entity should develop a framework for conducting digital forensics in the event of AI/ML security incidents, to gather evidence, determine the cause and impact, and provide insight into how similar incidents can be prevented in the future.
- The entity should work with specialized, trained personnel or external partners for conducting AI/ML forensics, ensuring they have the necessary skills, tools, and access to data.
- The entity should incorporate the forensic insights into the post-incident review process, cyber security strategy and risk management, for enhancing the overall resilience of its AI/ML systems.
- The entity should ensure the forensic process is in line with legal, regulatory, and contractual obligations.

SECTION 4

Policy Implementation

POLICY IMPLEMENTATION

The Cyber Security Council will work with Emirate leads and sector leads to ensure compliance to the requirements of this policy.

Compliance with this policy will be monitored through the UAE Information Assurance (IA) Standard controls. All requirements outlined in this policy are embedded within the UAE IA Standard (Refer to “Appendix E.2 National Artificial Intelligence Security Policy” of the Standard). Entities are expected to report their compliance status via the National Cyber Index Platform.

To bring about the change required to successfully promote cyber security, education, awareness, and communications are needed.

The CSC will continue to engage with various participants in the UAE to enable adoption, implementation and continual monitoring of this policy.

SECTION 5

Performance Monitoring

PERFORMANCE MONITORING

The National Cyber Security Policy on Artificial Intelligence outlines measures for monitoring and evaluating progress towards the following objectives:

- Provide guidance for improvement and taking necessary intervention steps when appropriate.
- Measure the successful implementation of policy requirements by the adopting entities.

SECTION 6

Appendices

10.1 Reference Documents

UAE Policies and Standards

The following UAE policies and standards were referenced when defining these policy statements:

Authority/Body	Document
Minister of State for Artificial Intelligence, Digital Economy & Remote Work Applications Office	UAE National Strategy for Artificial Intelligence
CSC	UAE Information Assurance Standard

National Policies and Frameworks

This policy should be read in conjunction with the following National Policies:

Authority/Body	Document
CSC	National Encryption policy
CSC	National Cloud Security Policy
CSC	Cyber Incident Response Framework
CSC	Cyber Incident Response Plan

10.1 Reference Documents

International Standards

The following table outlines the international sources referenced in this document:

Authority/Body	Document
ENISA	Securing Machine Learning Algorithms
ENISA	AI Cybersecurity Challenges
German Federal Office for Information Security	AI Security Concerns in a Nutshell
MITRE	Adversarial Threat Landscape for Artificial-Intelligence Systems (ATLAS)
NIST	Artificial Intelligence Risk Management Framework (AI RMF 1.0)
Singapore Standards Council	Artificial Intelligence (AI) Security – Guidance for Assessing and Defending Against AI Security Threats – TR 99:2021
UK NCSC	Principles for the Security of Machine Learning

10.2 Abbreviations

Usage	Description
AI	Artificial Intelligence
CII	Critical Information Infrastructure
CIS	Center for Internet Security
CSC	Cyber Security Council
IDS	Intrusion Detection System
IPS	Intrusion Prevention System
ML	Machine Learning
SDLC	Software Development Lifecycle

10.3 Definitions

Term	Description
Adversarial AI/ML	A technique where an attacker tries to manipulate the behavior of AI models by injecting carefully crafted data.
Adversarial Training	A defensive strategy in AI/ML security involving iterative training of models with adversarial examples, enhancing their resilience against adversarial attacks.
Anomaly Detection	A mechanism in AI/ML security that identifies deviations from expected patterns or behaviors within data, alerting to potential security breaches or abnormal activities.
Artificial Intelligence (AI)	AI refers to the simulation of human intelligence processes by computer systems, encompassing tasks such as learning, reasoning, problem-solving, perception, and decision-making.
Attack Surface	The exposed points in an AI/ML system that could potentially be exploited by attackers.
Bias and Fairness	Refers to the presence of systematic and unfair discrimination in AI/ML outcomes, often caused by biased training data or algorithms.
Configuration Baselines	A predefined set of secure configurations for AI/ML systems that provide a stable and secure starting point.
Defensive Distillation	A technique in AI/ML security where a model is trained on softened probabilities from another model, making it more robust against adversarial attacks by reducing the effectiveness of gradient-based attacks.
Feature Squeezing	A defense mechanism in AI/ML security that reduces the complexity of input data by applying compression or filtering techniques, mitigating the impact of adversarial perturbations and enhancing model resilience.
Federated Learning	Training machine learning models across decentralized devices or servers while keeping data localized, thus reducing data privacy risks.

10.3 Definitions

Term	Description
Classification labels	• **Open:** Data that is publicly shared and published online with minimal restrictions. This type of data is freely accessible to anyone, and there are no confidentiality or sensitivity concerns associated with its dissemination. Examples include public datasets, research publications, and published government statistics, public domain art data, job postings, marketing materials, press releases, product brochures, etc. Open data does not include personal identifiable information (PII) that can be used to identify a specific individual, either on its own or when combined with other data including direct identifiers (Full name, Passport number, Emirates ID, Email address, Phone number, Biometric data, Financial data, Medical records, Genetic information etc.) and indirect identifiers (Date of birth, Gender, Geolocation/ IP, Employment/ Education data, Financial transactions, Medical condition/ diagnosis etc.). • **Confidential / Restricted:** Data that is restricted and intended for access only by specific individuals or groups. Unauthorized access to Confidential/Restricted data could harm an entity or individual. This data includes non-sensitive personal information (e.g. name, job title, work email, office phone number, department, employee ID, work location), non-critical entity financial data (e.g. budget estimates, internal forecasts), employee records, financial information (e.g. bank account details, financial statements, tax filings), organizational business plans, internal policies & procedures, and supplier contracts. • **Secret**: Data that requires a higher level of protection due to the significant harm at national, Emirate or sectoral level that could result from unauthorized access or disclosure. Secret data may include sensitive personal and health information (e.g. medical history, biometric data), trade secrets and proprietary algorithms, source code for proprietary software, product design schematics and investment/ legal case strategies. • **Top Secret:** Data that is highly restricted and protected due to its critical importance. Unauthorized access or disclosure of Top Secret data could cause significant harm to national security or organizational integrity. This category includes classified government documents, highly sensitive corporate or investment strategies, critical infrastructure details, intelligence reports, personal (including health and financial) information of high-ranking individuals, highly sensitive foreign policies and national defense plans.

10.3 Definitions

Term	Description
Governance	The controls and practices, and processes that make sure policies are enforced.
Homomorphic Encryption	A cryptographic technique that allows computation on encrypted data without decrypting it, enhancing data privacy.
Machine Learning (ML)	A subset of AI, ML involves the use of algorithms and statistical models to enable computer systems to improve their performance on a specific task through learning from data.
Rollback Plan	A predefined strategy to revert a system to its previous state in case a change introduces significant issues or risks.
Threat Modeling	A methodical process of recognizing and evaluating potential security risks and vulnerabilities in systems. It involves analyzing system elements, data paths, and possible threats to devise effective security measures and protections.
Software Development Life Cycle (SDLC)	The Software Development Life Cycle (SDLC) is a process used by software developers to design, develop, test, and deploy software.

10.4 Roles and Responsibilities

The below table defines the key stakeholders and their respective roles and responsibilities with regards to this policy.

Stakeholder	Roles and Responsibilities
UAE Cyber Security Council (CSC)	As the custodian of this document, CSC shall: • Issue the National Cyber Security Policy For Artificial Intelligence and review the document periodically to ensure its relevance. • Coordinate with relevant stakeholders to disseminate the policy to critical sectors and entities. • Oversee the implementation of the provisions of the regulation to ensure compliance with the policy.
Ministries and Federal Authorities	• Comply with the requirements outlined in the National Cyber Security Policy For Artificial Intelligence. • Implement the Policy's provisions on applicable services. • Exercise due diligence and conduct the appropriate risk assessments outlined in this policy.
Non-Government CII Entities	• Comply with the requirements outlined in the National Cyber Security Policy For Artificial Intelligence. • Implement the Policy's provisions on applicable services. • Exercise due diligence and conduct the appropriate risk assessments outlined in this policy.

10.5 Acknowledgements

The development of the Artificial Intelligence Security Policy has been a collaborative endeavor, and we extend our heartfelt gratitude to all those who have contributed to its success. This document stands as a testament to the dedication, expertise, and commitment of the organizations and individuals whose invaluable input and guidance have shaped the process. Their collaborative efforts have ensured that this standard reflects the highest level of quality, relevance, and applicability, positioning it as a critical tool for advancing AI security practices.

We would also like to extend our heartfelt thanks to the CSC Governance & Policies Program Steering Committee who played a key role in reviewing, approving, and endorsing the Artificial Intelligence Security Policy:

- Ministry of Defence
- Ministry of State for Artificial Intelligence, Digital Economy, and Remote Work Applications
- Abu Dhabi Department of Government Enablement
- Dubai Electronic Security Center
- Sharjah Cyber Security Center
- Department of Digital Ajman
- Smart Umm Al Quwain
- Ras Al Khaimah Electronic Government Authority
- Fujairah e-Government
- Telecommunications and Digital Government Regulatory Authority