NUCLEAR HARMONIZATION & STANDARDIZATION INITIATIVE

Computer Security Considerations for the Use of Artificial Intelligence in Small Modular Reactors

November 2026

Working Paper[1] produced by the NHSI Regulatory Track Nuclear Security Working Group

1. Purpose and Scope

This working paper identifies computer security considerations for artificial intelligence (AI) used in or supporting small modular reactors (SMRs). It supplements existing nuclear computer security guidance by addressing dependencies on data, models, development and deployment environments, infrastructure, and shared services. It also provides consideration for changes in human interaction and developing subject matter expertise.

Some SMR deployment approaches introduce shared digital dependencies through standardized designs, multi-unit operation, centralized services, and remote operations and support. In order to support SMR operability and efficiency, AI base supporting or operation systems are being developed, demonstrated or considered for monitoring, maintenance, engineering, operator support and control. This working paper provides guidelines for protecting SMR systems throughout their entire lifecycle from new risks associated with the use of AI. To help relevant entities work together, this document focuses on three core areas: AI applications in the SMR lifecycle, governance, and AI specific computer security measures.

2. Potential AI Applications for SMRs and security considerations

AI offers opportunities to enhance the design, operation, maintenance, safety and security of SMRs. Table 1 presents potential AI applications across the SMR lifecycle.

Table 1 Potential AI applications in SMR lifecycle

Application	Function or Use
Design and licensing support	Design optimization, AI-enhanced simulation, safety analysis and preparation of licensing material.
Manufacturing inspection	Image analysis and non-destructive examination of components and welds; acceptance testing.
Reactor condition monitoring and anomaly detection	Identify abnormal reactor conditions from pressure, temperature, coolant flow and other operating data.
Predictive maintenance	Estimate remaining useful life, prioritize maintenance and set inspection intervals for plant equipment, including cooling and safety-support equipment.
Collaborative AI Training (Federated Learning)	Enable SMR operators, including competing companies, to collaboratively train AI models for common problems, such as anomaly detection and predictive maintenance, without sharing raw operational or confidential plant data.

[1] This paper is not an IAEA publication. It reflects the consensus opinion of the members of the NHSI Regulatory Track Nuclear Security Working Group. The views expressed do not necessarily reflect those of the IAEA or its Member States.

Application	Function or Use
Digital twins and virtual sensors	Forecast plant behaviour and estimate conditions that are difficult to measure directly using physical and learned models.
Document and operator assistance	Search procedures, event logs, maintenance reports and engineering records; support operating decisions and training using the applicable unit configuration and document revision.
Physical protection surveillance	Recognize and prioritize suspicious activity or unauthorized entry using e.g. images and video under realistic environmental conditions.
Robotic inspection and maintenance	Interpret inspection data and, where authorized, navigate or execute physical inspection and maintenance tasks.
Multi-unit monitoring and coordinated operation	Monitor multiple SMR modules and coordinate shared support resources across unit configurations.
Control and energy coordination	Provide bounded control and optimize shared systems, including load following and other electrical-system support services, and coordination with energy storage and industrial heat users.
Computer security monitoring	Detect and prioritize suspicious activity across plant systems, remote support connections, shared services or sites.
Floating nuclear power plant monitoring	Monitor reactor and seawater-dependent cooling conditions while accounting for platform motion, sea conditions and changes in operating or relocation status.
Nuclear-powered merchant ship propulsion and energy management	Support coordination of reactor power demand, propulsion and electrical loads during maneuvering, port operations and navigation at sea.

When introducing AI into SMRs, unique characteristics distinct from traditional software technology lead to new computer security challenges and vulnerabilities. The key characteristics are as follows:

Data-Driven Dependency: The behaviour of AI models depends not only on their algorithm and implementation, but also on the quality of integrity of the data used for training, validation, and retraining. This creates exposure to unique AI-specific attack vectors, such as data poisoning or adversarial perturbation, that are difficult to detect using traditional security measures.

Lack of Transparency: While traditional programming involves humans directly developing rules in the form of algorithms, AI autonomously learns rules from data provided with input and desired outcomes. As a result, AI models generate results through complex computations and machine learning techniques, making it difficult for humans to intuitively understand or trace the internal decision-making process, called the black-box problem. This lack of transparency makes it difficult for operators to verify the basis of AI decisions in real-time, and complicates analysis in the event of a security incident.

Non-deterministic Characteristics: Traditional software consistently produces pre-defined, identical outputs for identical inputs (deterministic). In contrast, because learning algorithms rely on probabilistic techniques, neural network-based AI inherently produces non-deterministic outputs. While this nondeterministic nature potentially enables autonomy, it also leads to significant disadvantages from a security perspective. These variable responses impede attack detection, obscure the distinction between unintentional system errors and malicious attacks, and obstruct comprehensive verification and validation (V&V).

3. Governance and regulatory compliance for AI supported operations

Governance arrangements for AI-supported systems in SMRs must be integrated into existing State and operator level frameworks for nuclear safety, security and protection of sensitive information.

State-level legal and regulatory framework

Governments should review their legal and regulatory framework to determine whether existing requirements adequately address AI-supported systems. Responsibilities should be clearly assigned to the relevant competent authorities. When responsibilities overlap, effective coordination mechanisms should be established.

The regulatory requirements should consider the national threat assessment or design basis threat, and include incident reporting requirements, supply chain security, and protection of sensitive information. IAEA NSS No. 42-G *Computer Security for Nuclear Security* [1], provides guidance to support state and organizational computer security arrangements.

Operator-level programs and lifecycle management

At the operator level, organizations should address AI-supported systems within their management systems and computer security programmes (CSP). Facility and system computer security risk management (CSRM) should evaluate:

- The functions supported by AI,
- The digital assets associated with these functions, and their data flows, interfaces, model update mechanisms and shared or vendor-provided services.
- The potential consequences of a cyber-attack against these assets,

IAEA NSS No. 17-T (Rev 1) *Computer Security Techniques for Nuclear Facilities* [2], provides guidance for establishing the CSP and managing facility and system CSRM.

AI-supported systems should adopt a “security by design” approach, implementing robust computer security requirements from the conceptual and architectural design, model selection phases, rather than as an afterthought. Where AI supports or interfaces with digital instrumentation and control (I&C) systems, computer security should be addressed throughout the entire I&C system life cycle. This includes configuration management, verification and validation (V&V), security assessment, and system modifications. IAEA NSS No. 33-T *Computer Security for Instrumentation and Control Systems at Nuclear Facilities* [3], provides guidance for managing computer security throughout the I&C system lifecycle.AI-related information should also be assessed for sensitivity and protected according to IAEA NSS No. 23-G *Security of Nuclear Information* [4]. Furthermore, to manage digital dependencies and supply chain security risks effectively, operator programmes, should make use of Software Bill of Materials, and Machine Learning Bills of Materials. These inventories ensure comprehensive tracking of software components, third-party libraries, pretrained models, training datasets, and associated provenance metadata throughout the system lifecycle.

Operational Traceability and Explainability

For AI enabled systems applied in safety and security significant SMR applications, traceability and explainability should be addressed as complementary assurance capabilities extending beyond software-level model interpretation. Traceability should enable the reconstruction of the information chain, from physical sensors and operational technology (OT) acquisition systems to data preprocessing, model inference, and results. This requires appropriate records of data provenance, sensor identification and reliability, calibration

status, timestamps, data integrity and availability, communication latency where relevant, and applicable software, firmware, and model configurations. Depending on the system architecture, independent diagnostic and monitoring mechanisms may provide additional evidence of measurement reliability without interfering with significant facility functions. Furthermore, to support incident response and digital forensics, traceability mechanisms should capture model execution contexts ensuring that AI-driven decisions can be systematically audited, reconstructed, or evaluated post-incident.

Explainability should provide human operators with understandable information about the process conditions, relevant input variables, model limitations, and technical basis supporting an AI-generated recommendation. Consequently, explainable AI should not be limited to feature-attribution algorithms or visualization tools; it should be supported by verifiable operational evidence, time-correlated diagnostic records, and mechanisms enabling operators to assess the reliability and safety implications of AI-supported decisions before taking critical actions.

4. AI Specific Computer Security Measures

There are a variety of computer security considerations that apply to each of the potential AI applications identified in section 2. Not every consideration described below applies to every use case, but most are generally applicable. To protect AI integrations, operators must deploy computer security measures within the defensive computer security architecture (DCSA). These measures must be managed continuously across the system lifecycle.

AI-specific measure	Application to SMRs
Training-data protection	Verify data origin, authenticity, integrity, labels, and transformations; separate training, validation and test datasets; investigate suspicious samples. Protect maintenance histories and reactor simulation datasets against changes that could distort learned relationships.
Model origin and backdoor assessment	Verify the origin, authenticity, and integrity of pretrained models and updates. Evaluate unexpected or trigger-dependent behaviors before deployment. A valid signature establishes the origin, authenticity, and integrity of a model but does not demonstrate that a model is free of malicious behaviors.
Adversarial robustness testing	Test whether manipulated sensor readings, inspection images or operating histories can produce misleading outputs. Include coordinated changes that remain individually plausible and attacks adapted to the implemented defences.
Detection of conditions outside validated use	Monitor input distributions, model performance and uncertainty. Identify unfamiliar plant configurations, transients or environmental conditions, and define when predictions should be withheld or referred for review.
Independent checks on AI outputs	Compare consequential predictions and recommendations with physical constraints, independent measurements or validated engineering calculations. Check virtual-sensor estimates against available physical measurements and identify shared dependencies that could undermine independence.
Prompt-injection and retrieval protection	For generative AI, distinguish trusted instructions from retrieved content, control document versions and unit applicability, and test malicious instructions embedded in procedures or logs. Enforce retrieval and tool permissions outside the model.

Protection against model extraction and information leakage	Limit and monitor model queries and exposed outputs. Test whether responses disclose sensitive training or contextual information. Assess privacy-preserving techniques against required performance and the specific disclosure risks.
Data privacy and distributed model training	In federated learning, multiple organizations collaboratively train models without centrally collecting raw operational data. Federated learning requires organizations to authenticate one another, and protect exchanged parameters and updates using authenticated communication, encryption and secure aggregation where appropriate. Assess the risk posed by malicious participants, compromised aggregators, collusion, data or model poisoning, model inversion and membership inference. Validate aggregated models across SMR configurations and operating conditions before deployment. Federated learning reduces raw-data sharing but does not itself guarantee confidentiality or prevent information leakage. Secure aggregation protects individual updates under defined assumptions; separate measures are needed to address poisoning.
Controlled learning and model changes	Treat retraining, fine-tuning, changes to learning objectives, federated model releases and retrieval-source updates as controlled changes. Revalidate affected behaviour before deployment and prevent unreviewed learning processes from modifying authorized operational models.
Independent limits on AI authority	Enforce approved action limits independently of the model. For control applications, block commands outside justified operating constraints. For document assistants, prevent operational I&C command access.
AI-specific recovery and fleet assurance	Preserve model versions, relevant inputs, outputs and configurations for investigation and recovery. Verify recovery baselines and assess whether common models, federated updates or shared datasets could cause simultaneous failures across units or sites.

5. Conclusion

Artificial Intelligence offers valuable opportunities to enhance the economic viability, operational efficiency, and computer security of Small Modular Reactors. However, these benefits are inextricably linked with risks, making it impossible to leverage one without introducing the other. Consequently, the inherent risks need to be identified, understood, evaluated and treated before AI systems are deployed. This makes realizing the benefits of AI conditional on the simultaneous implementation of robust computer security measures to ensure effective risk treatment.

For SMRs, AI governance should be embedded in the facility management system and in the computer security programme, and supported by facility and system computer security risk management. A graded approach and defence in depth should determine the rigour of measures. Measures should be proportionate to the consequences of compromise and integrated into existing governance and life cycle processes. Data and model governance, configuration management, supplier assurance, verification and validation, monitoring, incident response and controlled fallback are necessary throughout the lifecycle. An AI application should not create an uncontrolled dependency or single point of failure for a safety or security function. If the application is degraded, unavailable or subject to a computer security incident, essential functions and the capability to place the facility in a safe condition should be maintained.

References

[1] INTERNATIONAL ATOMIC ENERGY AGENCY, Computer Security for Nuclear Security, IAEA Nuclear Security Series No. 42-G, IAEA, Vienna (2021)

[2] INTERNATIONAL ATOMIC ENERGY AGENCY, Computer Security Techniques for Nuclear Facilities, IAEA Nuclear Security Series No. 17-T (Rev. 1), IAEA, Vienna (2021)

[3] INTERNATIONAL ATOMIC ENERGY AGENCY, Computer Security of Instrumentation and Control Systems at Nuclear Facilities, IAEA Nuclear Security Series No. 33-T, IAEA, Vienna (2018)

[4] INTERNATIONAL ATOMIC ENERGY AGENCY, Security of Nuclear Information, IAEA Nuclear Security Series No. 23-G, IAEA, Vienna (2015)

IAEA publications relevant to Computer Security of Small Modular Reactors:

Nuclear Security Series:

1. NSS 13, Nuclear Security Recommendations on Physical Protection of Nuclear Material and Nuclear Facilities (INFCIRC/225/Revision 5)
2. NSS 42-G, Computer Security for Nuclear Security
3. NSS 17-T Rev.1, Computer Security Techniques for Nuclear Facilities
4. NSS 33-T, Computer Security for Instrumentation and Control Systems at Nuclear Facilities

Non-serial Publications:

1. Computer Security Incident Response Planning, TDL-005
2. Conducting Computer Security Assessments, TDL-006
3. Computer Security Approaches to Reduce Cyber Risks in the Nuclear Supply Chain, TDL-011

For more information on NHSI SMR Activities: