Journal of Advanced Marine Engineering and Technology, Vol. 49, No. 3, pp. 194~205, 2025
J. Advanced Marine Engineering and Technology (JAMET)
https://doi.org/10.5916/jamet.2025.49.3.194

ISSN 2234-7925 (Print)
ISSN 2765-4796 (Online)
**Original Paper**

# Design of a system-related risk management chain for an intelligent decision support system prototype in Korean nuclear power plants under normal and abnormal conditions

Gwi-Sook Jang† · Seo-Ryong Koo[1]

(Received June 9, 2025 ; Revised June 15, 2025 ; Accepted June 22, 2025)

**Abstract:** An Intelligent Decision Support System (IDSS) is being developed for the Korean APR1400 nuclear power plant (NPP) to enhance decision-making and reduce human error under normal and abnormal conditions. As a non-safety monitoring prototype, the IDSS is currently in development. While traditional Man-Machine Interface Systems in NPPs face risks from software and hardware failures, the integration of deep learning introduces further challenges, particularly those related to open-source software and adversarial AI threats. This study presented a comprehensive system-related risk management chain to address these issues. Risk factors in the IDSS's development and operation were identified using the NUREG/CR-6430 risk assessment methodology. Based on this, a management plan was established, including administrative and technical controls. Administrative measures addressed licensing and cybersecurity for open-source components, while technical controls aimed to protect against cyber intrusions and maintain IDSS performance. The proposed risk management chain is expected to reduce operational losses and improve efficiency in NPP applications. It also supports the safe integration of intelligent systems into nuclear operations. Future efforts will apply the management chain across the IDSS development lifecycle to validate its effectiveness and expand its application to broader intelligent system deployment in NPPs.

**Keywords:** Intelligent decision support system in NPPs, Risk management in AI System in NPPs, Intelligent decision support system prototype

## 1. Introduction

Artificial intelligence (AI)-based decision support systems (DSS) in the nuclear energy field are rapidly advancing, with growing applications in real-time operational monitoring, anomaly detection, predictive maintenance, and digital twin simulations. These systems utilize sophisticated machine learning models to enhance operator situational awareness, detect equipment malfunctions, and enable proactive maintenance planning. Emphasis is placed on human-in-the-loop designs, ensuring that AI supports rather than replaces human decision-making, particularly in safety-critical situations.

While regulatory authorities are carefully developing guidelines to ensure transparency, reliability, and explainability of AI systems, numerous pilot projects and research efforts led by international and national organizations—such as the International Atomic Energy Agency (IAEA), the United States Department of Energy (U.S. DOE), and the Korea Atomic Energy Research Institute (KAERI)—are facilitating the integration of AI into both safety and operational aspects of NPP management **[1]**.

### 1.1 Previous Research

Jang and Koo **[2]** proposed the Intelligent Decision Support System (IDSS) for normal and abnormal conditions, which extends beyond the basic functions of data collection, storage, analysis, and prediction in APR 1400-type Korean Nuclear Power Plants (NPPs). IDSS analyzes and provides critical information to assist control room operators in decision-making. By visualizing information, the system reduces the operator's cognitive load. Additionally, it presents data on system-generated results and

† Corresponding Author (ORCID: http://orcid.org/0009-0002-8594-1477): Principal Researcher, Advanced Instrumentation & Control Research Division, Korea Atomic Energy Research Institute, 111, Daedeok-daero 989beon-gil, Yuseong-gu, Daejeon, Korea, E-mail: gsjang@kaeri.re.kr, Tel: +82-42-868-8626

1 Principal Researcher, Advanced Instrumentation & Control Research Division, Korea Atomic Energy Research Institute, E-mail: srkoo@kaeri.re.kr, Tel: +82-42-868-8831

their underlying rationale, thereby enhancing decision-making and minimizing human errors. The primary objective of developing IDSS for APR 1400-type Korean NPPs is to significantly reduce human error-induced failures and unplanned shutdowns.

Although quantifying the exact reduction in human errors due to IDSS remains challenging, evaluating its effectiveness in preventing past incidents is essential. Also, Jang and Koo **[2]** identified key design challenges in applying the fundamental technology of an IDSS, as well as the necessary design and validation technologies for its implementation. The role of IDSS in Korean NPPs under normal and abnormal conditions is as shown in **Figure 1**.

There has been no prior example of integrating an AI-based system as a subsystem within the MMIS (Man Machine Interface System) of APR 1400-type Korean NPPs. While the risks associated with traditional MMIS primarily involved software defects, hardware failures, and digital common-cause failures, the application of AI technology in IDSS introduced additional risks, particularly those arising from the use of open-source software and adversarial machine learning.

In recent years, AI systems have increasingly been adopted in both critical infrastructure and high-risk domains, including NPPs. This has accelerated research into AI-specific risk management methodologies and frameworks. Yang *et al.* **[3]** proposed a real-time risk management system based on Living PSA (Probabilistic Safety Assessment) using a GO-FLOW modeling technique. Their modular and hierarchical structure enabled rapid model updates and reliability assessments during NPP operation, demonstrating benefits in configuration management and online safety monitoring. While this work enhances risk visibility for operational personnel, it primarily focuses on system reliability metrics rather than AI-induced risk factors. Bogdanov *et al.* **[4]** developed a generalized risk management methodology for AI systems aligned with ISO 31000, ISO/IEC 27005, and NIST Risk Management Framework (RMF). Their work introduces three blueprints for classifying AI system architectures and matching risk control strategies accordingly. While comprehensive, their framework remains largely abstract and lacks application-specific tailoring such as that required in nuclear domains. Lim et al. **[5]** applied the NIST AI Risk Management Framework to a national technology database, introducing a MAP-MEASURE-MANAGE methodology to ensure regulatory compliance and traceability of AI-generated insights. Their work focuses on policy analytics, employing supervised learning and natural language processing tools for large-scale document analysis. Although relevant in terms of demonstrating AI governance principles, their domain and operational constraints are significantly different from safety-critical control systems like IDSS in nuclear power. Giudici *et al.* **[6]** introduced a KAIRI (Key AI Risk Indicators) framework that quantifies four core dimensions of AI trustworthiness—Sustainability, Accuracy, Fairness, and Explainability (S.A.F.E). These are statistically grounded indicators aimed at financial AI applications, but the methodology can be extended to high-risk AI settings by applying objective, reproducible metrics to monitor AI trust over time.

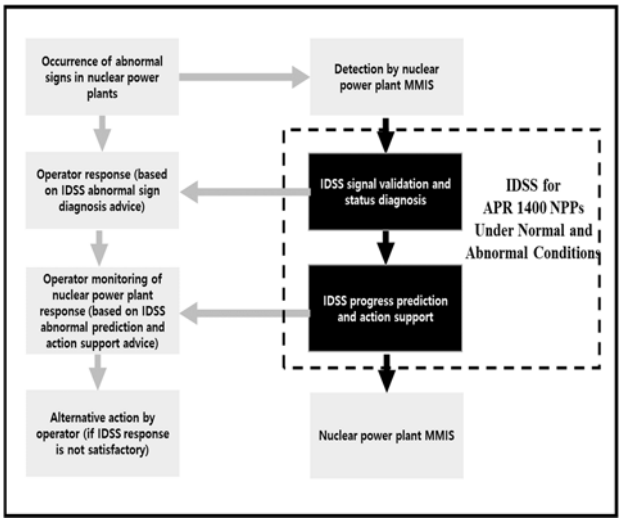


**Figure 1:** Role of the IDSS in Korean NPPs under normal and abnormal conditions

In recent related studies, including the AI RMF, researchers have proposed using quantitative indicators such as loss event frequency, loss magnitude, model confidence scores, and fairness metrics to evaluate risks in AI systems. While these methods enable numeric comparison and automation, they are heavily dependent on large volumes of operational data, which may not be available during early system design phases as in this study. Whereas prior studies either (1) focus on probabilistic safety assessments in NPPs without addressing AI-specific risks, or (2) propose statistical AI risk metrics in finance or government systems without integration into Software Development LifeCycle (SDLC) **[7]** of critical infrastructure. Also, conventional approaches lack continuous traceability and validation mechanisms for managing evolving risks. To solve this, we propose a lifecycle-oriented “risk management chain”. Specifically, we propose a structured risk management chain based on NUREG/CR-6430 **[7]** risk assessment, incorporating both administrative and

technical controls for AI-based IDSS in nuclear applications. The proposed model thus addresses both design-time risk mitigation and operational traceability—bridging a critical gap in AI-enabled NPP safety systems. Jang and Koo **[2]** proposed a prototype of the IDSS for Korean NPPs, emphasizing that potential risk factors should be identified and managed in advance to minimize operational errors. However, their study focused mainly on the design challenges and countermeasures for implementing IDSS, without suggesting a specific framework for managing the identified risks. Therefore, this study builds upon the previous work to qualitatively identify system-related risk factors during IDSS development and operation, and to design a risk management chain to efficiently address them.

### 1.2 Research Methodology

This study presents a system-related risk management methodology specifically designed for the development and operation of an IDSS in NPPs. Based on the NUREG/CR-6430 **[7]** risk assessment matrix, the proposed approach emphasizes lifecycle-based, traceable risk management even in the absence of sufficient operational data. This method fills a methodological gap by addressing the limitations of data-dependent quantitative risk frameworks such as the NIST AI RMF **[8]** and Giudici *et al.* **[6]**, which may be ill-suited for high-assurance domains like NPPs.

The core concept of this methodology is the 'system-related risk management chain,' defined as a structured, recursive, and lifecycle-oriented framework that integrates risk control processes across both development and operational phases. It consists of three interlinked phases: (1) Risk Identification, (2) Countermeasure Implementation, and (3) Validation & Feedback. Unlike conventional quantitative models, this chain prioritizes qualitative assessment and traceability, enabling proactive risk management without requiring extensive runtime metrics.

This chain incorporates both administrative and technical controls, ensuring regulatory compliance while allowing for iterative feedback between observed vulnerabilities and the effectiveness of implemented safeguards. Its integration within the IDSS development lifecycle also enables alignment with design guidelines such as NIST AI RMF **[8]**, Software Bill of Materials (SBOM) requirements **[9]**, Korean guidelines for Trustworthy AI **[10]**, NUREG 0700 **[11]**, and TTAK.KO-10.1497 **[12]**.

As the IDSS prototype is currently under development, this study does not include quantitative metrics of risk reduction. Instead, the effectiveness of the proposed risk management chain is demonstrated conceptually by embedding it into the system architecture, with future work aiming to include measurable indicators suitable for IDSS characteristics. The comparative summary of risk management approaches is as in **Table 1**.

**Table 1:** Comparative summary of risk management approaches

| Comparison Item | System-related Risk Approach (proposed) | Quantitative Risk Approach **[6]** |
|---|---|---|
| Target Domain | High-assurance systems (e.g., NPPs) | General AI systems (e.g., finance, health) |
| Approach Type | Qualitative chain: Identify → Implement → Validate/Feedback | Quantitative metrics: Frequency, Exposure, Confidence |
| Framework Base | NUREG/CR-6430 + Safety Guidelines (e.g., NUREG 0700, TTAK) | NIST AI RMF, Giudici et al. (2024) |
| Verification Method | Traceability through development lifecycle | Numerical simulation and benchmark testing |
| Regulatory Alignment | Strong alignment with NRC, KINS, and Korean AI safety standards | Limited adoption in high-reliability domains |
| Data Dependency | Low: Designed for early-stage use without full datasets | High: Requires operational metrics and historical data |
| Adaptability | Integrates with SBOM, KAIRI, safety modes for future enhancement | Dependent on precise metric definitions and data quality |

## 2. IDSS Prototype

The IDSS is designed using data-driven and deep neural network-based algorithms to enhance robustness and predictive accuracy across various measurement variables encountered under diverse NPP conditions. It predicts plant behavior based on reliable signals, diagnoses the current state, and calculates the remaining reactor trip time, thereby informing operators of the available response window. In single abnormal scenarios, it recommends actions in accordance with procedural guidelines, while in complex situations, it suggests optimal functional recovery strategies **[2]**.

The IDSS operates within a secure development environment (**Figure 2 (a)**)—which includes a nuclear big data platform and a deep learning model platform—as well as a secure operational environment (**Figure 2 (b)**)—comprising IDSS servers and IDSS displays, as illustrated in **Figure 2**. In future implementations, development environment equipment will be located outside the NPP, whereas operational environment equipment will be positioned in the control room and computer room within the auxiliary building of the NPP. The secure development process, conducted externally, involves iterative cycles of data management,

analysis, model training, and performance evaluation using the big data and AI platforms. The resulting AI models and algorithms are distributed as software through containerization, leveraging a microservice architecture. Meanwhile, secure operations within the power plant are facilitated by integrating IDSS display pages into the information display screen layer of the dual IDSS server, which is housed in both the computer room and the main control room control panels.

The IDSS is made up of several key components, each with its own set of capabilities that work together to support real-time operations in a nuclear power setting.

First, there's the Big Data Platform, which acts as the foundation for all data-related tasks. It collects and organizes large volumes of data, processes it to handle missing values or outliers, and labels it for training. It also takes care of balancing uneven datasets, cataloging and integrating different data sources, sampling, selecting relevant features, and storing everything safely. On top of that, it helps with system configuration, validates the data's quality, and even provides tools to visualize it all in a clear, user-friendly way.

Then there's the Artificial Intelligence Platform, which is where the heavy lifting in terms of AI happens. This platform is responsible for training deep learning models and fine-tuning them through hyperparameter adjustments. It tests how well the models perform, checks if they're overfitting or undergeneralizing, and ensures they're robust enough to handle sensor aging or external disturbances. It also includes tools to explain how models arrive at certain decisions—what we call explainable AI (XAI)—and keeps track of different model versions with all their metadata. It manages the models' storage, registration, monitoring, and ensures they can work with other systems seamlessly. At the heart of the system is the IDSS Server, which is where real-time decision-making happens. It validates incoming signals, detects abnormal conditions, predicts how situations may evolve, and suggests actions to prevent potential problems. It uses the trained AI models to do all this, and it feeds the results into the system interface. It also evaluates how trustworthy its own responses are and checks the reliability of the data it's working with.

Lastly, the IDSS Display makes everything accessible to operators. It presents them with clear, actionable information—like alerts, trip forecasts, and recommended procedures—while also showing how confident the system is in its suggestions. If the system notices its own performance slipping, it can trigger a safe mode to ensure continued safety.

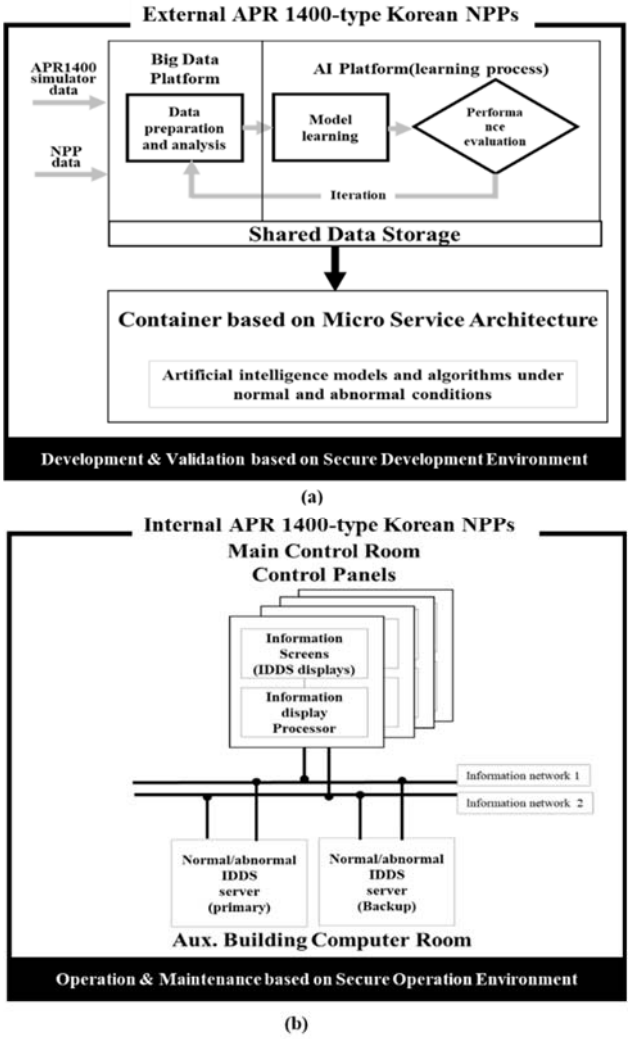


**Figure 2:** IDSS development (a) and operation (b) environment for APR 1400-Type Korean NPPs

## 3. Design of a System-related Risk Management Chain of the IDSS Prototype

The risk management chain for the IDSS prototype aims to minimize operational losses resulting from risk-induced reactor shutdown in NPP operations by systematically identifying, assessing, and managing risks. Additionally, it seeks to enhance operational efficiency and build stakeholder trust to ensure the safety and reliability of NPP operations. The design process of system-related risk management chain of the IDSS prototype is shown in **Figure 3**.

### 3.1 Risk Identification for the IDSS Prototype

According to ISO/IEC 24028 **[13]**, specific AI technology-based system risks must be evaluated from a reliability perspective, with priority given to risks with the most significant impact.

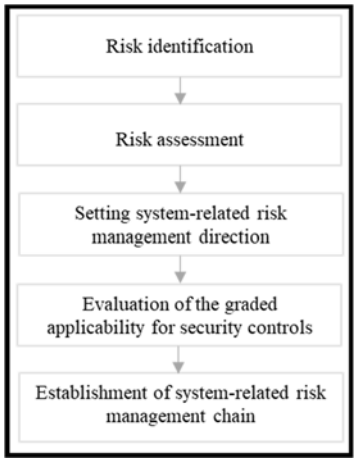


**Figure 3:** The design process of system-related risk management chain of the IDSS prototype

Based on these guidelines and the structure and functions of the IDSS prototype, this study categorizes IDSS risks into four key areas: data, algorithm, model, and system risks.

The management strategies for each category of the IDSS prototype are as follows:

- Data-related risks, such as data quality and security issues, can be mitigated by standardizing data management. This involves creating detailed descriptions of data for learning, testing, and validation, as well as verifying data integrity based on these descriptions.
- Algorithm-and model-related risks can be addressed by ensuring transparency in the model development process. This includes documenting AI model specifications and providing explanations of model information and results. Additionally, defining reliability levels and acceptance criteria for model outputs can enhance trustworthiness. To counteract adversarial machine learning threats, a reliability enhancement model can be applied, thereby improving robustness. Furthermore, the impact of such risks can be minimized by establishing model specification guidelines and conducting model validation accordingly.
- System-related risks can arise from system errors, hacking, vulnerabilities, and compatibility issues with other systems. These risks are closely tied to the use of open-source software, which is essential in implementing AI-based systems. Notably, since this is the first instance of an AI-driven system design incorporating open-source software into the MMIS design of NPPs, evaluating and managing cybersecurity threats is particularly critical. Adversarial machine learning threats associated with open-source software can significantly affect not only system-related risks but also data- and model-related risks.

**Table 2:** Risk impact analysis and assessment

| | Severity Level | | | |
|---|---|---|---|---|
| Frequency Level | Cata-strophic | Critical | Marginal | Negligi-ble |
| Frequent | High | High | High | High |
| Probable | High | High | Medium | Low |
| Occasional | High | High | Medium | Low |
| Remote | High | Medium | Low | Low |
| Improbable | Medium | Low | Low | Low |

## 3.2 Risk Assessment of the IDSS Prototype

ISO/IEC 24028 **[13]** and ISO/IEC 23894 **[14]** individually outline risks that must be considered when evaluating the reliability of AI-based systems. These standards emphasize that management measures should prioritize risks with the greatest impact. During the identification and management strategy development for the IDSS prototype, it is determined that open-source software use and adversarial machine learning threats have a significant impact on data-related, algorithm- and model-related, and system-related risks. Consequently, these risks are classified as high-priority system-related risks in the IDSS prototype through risk impact analysis. The occurrence of risk factors in the IDSS prototype is assessed based on severity levels and occurrence frequency categories (e.g., unlikely, likely, rare), as presented in **Table 2**, referencing NUREG/CR-6430 **[7]**. These assessments reflect the impact of NPP IDSS risk factors, with the most critical risks requiring immediate evaluation and prioritized countermeasures. Since the IDSS prototype is a non-safety monitoring system that supports the main information system of the main control room without involving control actions, its failure has a low impact on the overall safety of the NPP. However, as this marks the first instance of an AI-based system design utilizing open-source software in an NPP, evaluating and managing cybersecurity threats is crucial. Security threats associated with open-source software not only affect data-related, algorithm- and model-related, and system-related risks but may also provide incorrect information to the main control room operator.

Therefore, the security threats associated with open-source software in the development and operation of the IDSS prototype are assessed as occurring occasionally with a critical impact, resulting in a risk impact classification of “high.” Similarly, adversarial machine learning threats due to open-source software are evaluated as remote (infrequent) within the NPP IDSS

development and operation environment but are still considered to have a critical impact, leading to a risk impact classification of "medium." Accordingly, this study prioritizes the management and mitigation of system-related risks in the IDSS prototype, focusing on risks arising from open-source software and adversarial machine learning. This prioritization is based on risk impact analysis and evaluation following NUREG/CR-6430 **[7]**.

### 3.3 Setting System-related Risk Management Direction for the IDSS Prototype

The IDSS prototype faces two major categories of system-related risks: those stemming from open-source software and those related to adversarial machine learning. As open-source components are extensively used to accelerate development and incorporate advanced technologies, potential issues such as licensing conflicts, security vulnerabilities, and compatibility problems are key concerns. To manage these risks, a SBOM is maintained to track all software components, including packages, frameworks, and libraries. This approach helps quickly identify security issues and manage module dependencies. Since specific guidelines for open-source software use in the nuclear sector are not yet established, the IDSS prototype adopts Korean standards TTAK.KO-11.0309 **[15]** and TTAK.KO-11.0322 **[16]** to ensure secure and systematic software use. The list and versions of components are

**Table 3:** SBOM management form for IDSS prototype

| Item | Description |
|---|---|
| Component Supplier | Software, supplier name |
| Copyright holder (Component Author) | Software licenses, the original copyright holder of the software |
| Component Name | Component name |
| Component Version | Component version |
| Unique Identifier | Format ID |
| License Name | Component license |
| vulnerability DB | DB name |
| Relationship | Dependency between software components |
| Release Date | Release date |
| CVE ID | CVE-Year-Serial |
| CVSS Base Score | Base, Impact, Exploitability |
| CVSS Severity | High, Medium, Low, None |

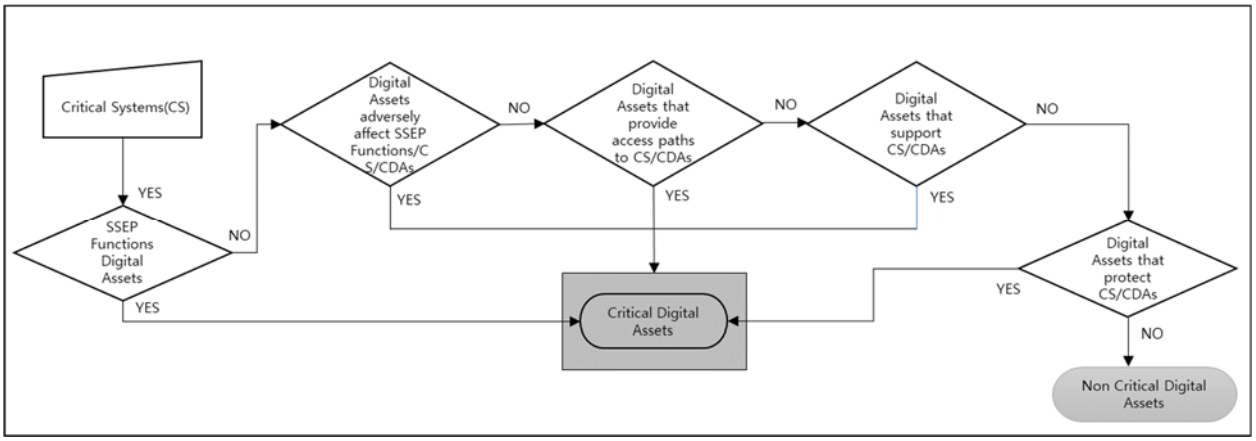


**Figure 4:** Identification procedure of critical system

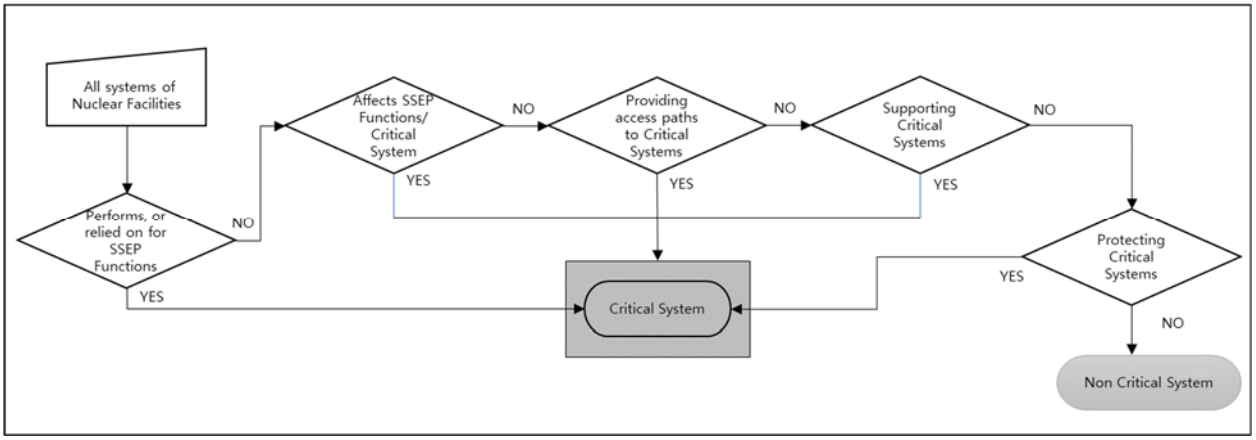


**Figure 5:** Identification procedure for critical digital assets

continuously monitored throughout the development and operational stages to minimize potential risks. The SBOM management form for thje IDSS prototype is as shown in **Table 3**.

Adversarial machine learning poses another serious risk, involving attack techniques such as poisoning, evasion, extraction, and inference attacks—all of which aim to compromise the accuracy or reliability of the system by manipulating training data or model inputs. These threats can lead to critical decision-making errors, making it essential to implement robust countermeasures.

Given that the IDSS prototype serves a vital function in NPPs, it is treated as a high-priority cybersecurity asset. Security strategies are designed to defend against unauthorized access, data manipulation, and other threats, ensuring the core principles of data confidentiality, integrity, and availability are upheld. These strategies follow recognized nuclear cybersecurity standards, such as KINAC/RS-015 **[17]** and NEI 13-10 **[18]**, and are incorporated into each phase of the SDLC as part of a comprehensive system-related risk management plan.

## 3.4 Evaluation of the Graded Applicability for Security Controls

According to the identification procedures for Critical Systems (CS) and Critical Digital Assets (CDA) outlined in KINAC/RS-015 **[16]**, and as illustrated in **Figure 4** and **Figure 5**, the IDSS for APR 1400-type Korean NPPs is classified as both a Critical System (CS) and an indirect Critical Digital Asset (CDA). Based on the graded application of security controls, as determined through the CDA-specific impact assessment and type classification from the NEI 13-10 **[18]** security controls assessment methodology, the IDSS functions as a decision-support system for NPP operators under both normal and abnormal operating conditions. Even if adversarial learning threats due to open-source software usage cause the IDSS to fail, the safety and safety functions of the NPP are not directly affected.

Since the final operational decisions are made by human operators, alternative means are available to compensate for potential system failures. Therefore, the IDSS is classified as an indirect Critical Digital Asset (indirect CDA). To manage potential security risks, the IDSS development process includes the application of a SBOM to identify vulnerabilities, and adopts a baseline set of security controls appropriate for its classification based on NEI 13-10 [18]. The detailed countermeasures for each control method of the IDSS prototype are presented in **Table 4**.

**Table 4:** Detailed countermeasures for each control method of the IDSS prototype

| Control Items | Detailed Items | Detailed Countermeasures |
|---|---|---|
| Vulnerability Assessment | N/A | Not applied due to limitations in the development and validation of IDSS prototype |
| Penetration Assessment | N/A | Not applied due to limitations in the development and validation of IDSS prototype |
| Physical Control | Physical protection zone | Independent execution space configuration |
| | Entry and access control | • Entry and exit only by authorized personnel, entry and exit records and maintenance<br>• Entrance card reader, entry and exit only by authorized personnel and history storage |
| | Physical access control | • Cabinet door lock application<br>• Monitoring through diagnosis and providing results to users<br>• Application software cannot be changed during online use,<br>• Password setting when system booting |
| | User security | • Security guidelines and defense measures education for participants<br>• IDSS development devices are accessed only by authorized personnel through user authentication<br>• Separation of system administrator authority and user authority<br>• User authority setting for restriction of access authority by task |
| Wireless/wired Control (air-gap) | Network security | • Independent network configuration to prevent wired and wireless connections between the operating network and the Internet<br>• Measures to disable wireless communication functions when bringing in devices<br>• Prevention of network access not authorized by the system administrator<br>• Source code and applications must be stored and accessible on a configuration management server independent of the document management system |
| Media Control | Media control | • Seal or disable unused ports<br>• Manage portable storage media according to media security management guidelines<br>• Use passive storage media and prohibit use of mobile devices when using portable media<br>• Block installation and execution of unnecessary applications<br>• Suppress use of unnecessary shared folders |
| Configuration Management | Configuration management | Use the developer document management system to manage the history of development outputs, and perform access control according to authorized authority. |
| Periodic Inspection | Periodic inspection | Perform periodic inspections according to the set diagnosis table |
| Continuous Monitoring of Security Measures | Performance evaluation | Self-diagnosis of the execution cycle must be performed to detect errors that exceed the set execution cycle or are not executed within the execution cycle start time, and the self-diagnosis results must be provided to the user. This is implemented as a safe mode function of the prototype, and is excluded from the management items of this document |

### 3.5 Establishment of System-related Risk Management Chain for the IDSS Prototype

The system-related risk management chain for the IDSS prototype is structured into three core stages: planning, administrative control, and technical control (see **Figure 6**). These stages collectively define a continuous lifecycle for identifying, mitigating, and responding to risks throughout development and operational phases. The process begins with a system-level risk management plan, followed by administrative implementation across the software development lifecycle (SDLC), and concludes with technical controls—particularly safe mode execution—during operational anomalies or failures.

#### 3.5.1 System-related Risk Management Plan

The risk management plan for the IDSS prototype involves three primary activities: risk identification, control item definition, and guideline establishment.

- Risk Identification: Risks are identified through analysis of the software environment and the use of open-source components, with adversarial machine learning highlighted as a major concern. A preliminary SBOM is

  compiled to document these components, aligned with standards such as TTA.KO-11.0309 **[15]** and NEI 13-10 **[18]**.

- Control Item Definition: The IDSS, classified as an indirect Critical Digital Asset (CDA), is not safety-critical but requires risk mitigation due to potential vulnerabilities. Baseline security controls, such as physical, network, wireless, and media controls, are applied. Safe mode is also defined as a technical control for performance failures.
- Guideline Establishment: A detailed risk management plan ensures that administrative and technical controls are documented and implemented throughout the SDLC. This includes record-keeping, evidence generation, and final evaluation reporting (See **Table 5**).

#### 3.5.2 Administrative Control and Response to System-related Risk

Administrative control and response for IDSS prototype risk consist of the system-related risk monitoring and system-related risk management and minimization. Detailed descriptions of each sub-item are provided below:

1) System-related Risk Monitoring: System-related risk monitoring for the IDSS prototype involves managing the SBOM, baseline security control, risk impact analysis, and evaluation to ensure risk-specific management control at each SDLC stage. Detailed descriptions for each item are outlined below.

**Table 5:** Risk management activities

| Stage | Risk Management Activities |
|---|---|
| Concept | • Establishment of a secure development environment |
| Requirement | • Evaluation of the appropriateness of maintaining countermeasures for a secure development environment |
| Design | |
| Implementation | |
| Test | |
| Evaluation | |
| Input | • System-related risk management plan |
| Results | • Creation of system-related risk management implementation data<br>• System security measure management and implementation report (including improvements)<br>• System-related risk management evaluation report |

- SBOM Management: SBOM management is continuously performed throughout the IDSS prototype's SDLC and operation stages, following the performance guidelines for each management control item outlined in the risk management plan. Validation data is generated to ensure compliance. Specifically, based on the confirmed SBOM list from the risk management plan setting stage, the SBOM list, licenses, and security vulnerabilities are compared at each SDLC stage to conduct SBOM configuration management. Implementation data is then generated through ledger management for each management task.
- Baseline Security Control Management: In accordance with the performance guidelines for each management control item in the IDSS prototype's risk management plan, baseline security control items are continuously managed throughout the SDLC and operation phases. Evidence of implementation is generated using the prescribed methods.
- Risk Impact Analysis and Evaluation: The risk impact analysis and evaluation for the IDSS prototype are conducted by assigning severity levels (A/B/C) and occurrence frequency categories (A/B/C/D/E) for each SBOM and baseline security control item. The results are then qualitatively classified as high, medium, or low based on the intersection of severity and occurrence frequency, as shown in **Table 2**.

2) System-related Risk Management and Minimization: The following risk management and minimization activities are performed based on the impact analysis and evaluation results of the risks identified during the system-related risk monitoring phase:

- If the risk impact analysis and evaluation results are classified as high, technical control and response methods for the IDSS system-related risk in the NPP are reviewed to determine feasibility.
- Additional management control methods are implemented as necessary.
- Corrective actions are carried out as follow-up measures after the risk.
- Risk impact analysis, evaluation results, and corresponding responses are documented to enable future comparisons.
- A risk improvement plan is established and implemented through periodic functional inspections and security measure monitoring.

3.5.3 Technical Control and Response to System-related Risk

If a system-related risk causes a failure or performance degradation in the IDSS prototype, despite administrative control measures, the IDSS safe mode is executed. An auditory and visual coding-based warning is issued to the main control room (MCR) operator of the NPP. Additionally, the cause and basis of the issue are provided upon the operator's request. The test for determining safe mode activation consists of three components: manual initiation, automatic application testing, and automatic performance testing during model result generation. The IDSS safe mode display and control method involves adding a dedicated IDSS button to the title bar of the operator display screen in the MCR of APR 1400-type Korean NPPs. This button performs multiple functions, including status display, warning notification, acknowledgment of warnings, and page control, depending on the situation.

In the MCR of an operating NPP, it is uncommon for a single button to perform multiple roles. However, since the IDSS prototype must be additionally installed in the operating NPP, design changes are minimized to account for the limited information display space. As a result, the role of the dedicated IDSS button changes according to the transition situation. The detailed role transitions and operational sequence for the IDSS button are illustrated in **Figure 7**. Additionally, the detailed information display page for IDSS—used for viewing information, conducting tests, accessing logs, and managing the system—is configured as a subpage of the existing operator display in the NPP. During normal IDSS operation, the operator can access the detailed information display page by clicking the IDSS button. The technical control and response sequence for system-related risks in the IDSS prototype consists of six steps, as shown in **Figure 8**. Each step is described in detail below.

- Step 1) Manual start and automatic test for IDSS prototype application: The manual start and automatic test for the IDSS prototype application are defined as check items to be conducted during each NPP operation shift change. These tests are performed at every shift to verify the soundness of the IDSS prototype software. An IDSS detailed information display page is configured on the existing operator display in the NPP MCR, with a "Test" button added to this page. Clicking the button initiates a test case that executes the entire IDSS logic. The accuracy of the IDSS logic is validated by comparing predefined result values with the actual results.
- Step2) Automatic performance test for IDSS model results: When the IDSS prototype is in operation and model results are generated, performance tests are automatically conducted. These tests evaluate factors such as response time and the reliability of the model output.
- Step 3) Determination and recording of IDSS prototype safe mode entry: If the logic checks from the manual start and automatic test in Step 1 fail, or if the automatic performance test (response time or confidence level of the model output) in Step 2 does not meet performance requirements or tolerance criteria, the system is determined to enter safe mode. The determination result and the process are recorded as system logs, which can be searched and viewed through the IDSS detailed information display page.
- Step 4) Notification of IDSS prototype safe mode entry: When it is determined in Step 3 that the IDSS has entered safe mode, the IDSS button on the upper title bar of the existing operator display is displayed as a warning window with a yellow border, and a warning sound is emitted. Upon hearing the warning sound, the operator can click the warning window to acknowledge the warning. The window also functions as a page control button, directing the operator to the IDSS detailed information display page for more information.
- Step 5) Execution of IDSS prototype safe mode at the operator's discretion: In Step 4, after navigating to the IDSS detailed information display page, the MCR operator can choose to stop the IDSS function at their discretion, thereby activating safe mode. Upon clicking the "Function Stop"

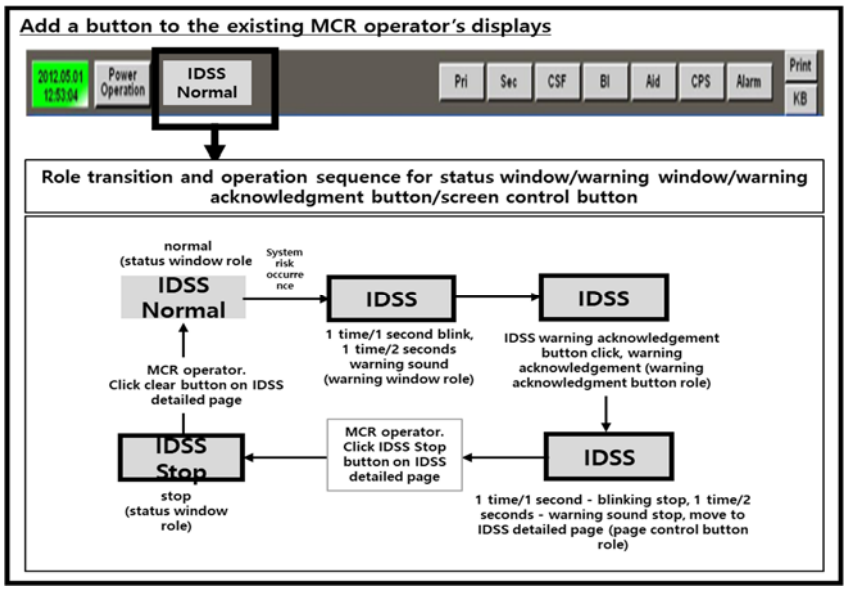


**Figure 7:** Role transitions and operational sequences for NPP's IDSS button

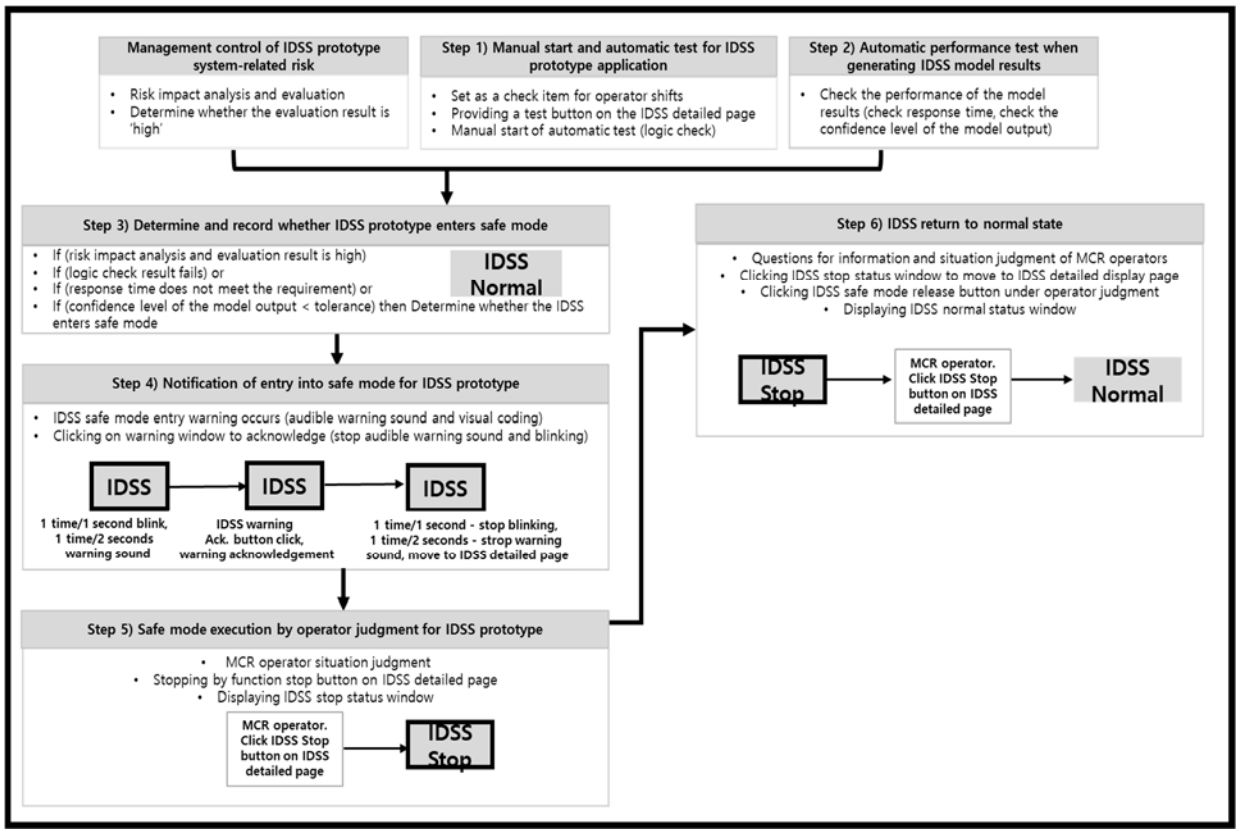


**Figure 8:** Technical control and response sequences to system-related risk

button, the IDSS button on the title bar of the existing operator display changes to a status window indicating that the IDSS has stopped.

- Step 6) Restoration of IDSS to normal state: Restoration of IDSS to normal state: Following Step 5, the MCR operator can click the IDSS stop status window to access the IDSS detailed information display page, where the information used for the safe mode determination can be reviewed. The operator can then restore the IDSS to its normal state by clicking the "Safe Mode Release" button on the detailed information page. The IDSS button on the title bar of the operator display then reverts to its normal state.

## 4. Conclusions and Future Plans

An IDSS for the APR1400-type Korean NPP is being developed to enhance decision-making and reduce human error in the main reactor under both normal and abnormal operating conditions. A prototype is currently in the design and development stage. Since an AI-based IDSS has not yet been integrated as a sub-system of a NPP's MMIS, there is no prior operational experience or established countermeasures for potential risks in IDSS development and operation. Therefore, it is essential to qualitatively identify high-impact risks associated with the intelligent system, formulate appropriate countermeasures, and validate their feasibility through implementation and evaluation across the IDSS prototype's SDLC.

This study designed a system-related risk management chain for the IDSS prototype targeted at APR1400-type Korean NPPs. The management chain addresses AI-specific system-related risks through planning, administrative, and technical controls. The study established the development and operational concept of the IDSS, identified risk factors related to data, algorithms, models, and system elements in the AI-based prototype, and prioritized them through risk evaluation based on NUREG/CR-6430 **[7]**. A comprehensive management plan for development and operational risks of the IDSS prototype was formulated, including administrative control measures to address licensing and cybersecurity risks arising from the use of open-source software, as well as adversarial machine learning threats associated with AI technologies. In addition, technical control measures were designed to respond to external cyber intrusions and performance degradation of the IDSS. However, due to the prototype stage and lack of field deployment, the study acknowledges data-based limitations in validating the effectiveness of the proposed controls. Nevertheless, this approach is expected to mitigate losses caused by system-related risks in NPP IDSSs, improve operational efficiency, and enhance stakeholder trust in the IDSS through a structured management chain applied under both normal and abnormal operating conditions. This marks a significant step forward in the adoption of intelligent systems in NPPs.

In the future, the proposed system-related risk management chain will be systematically applied throughout each stage of the SDLC of the currently developing prototype. Its validity will be rigorously validated through independent third-party qualitative review. Furthermore, the IDSS prototype will be continuously refined to expand the scope of risk management, and empirical evaluations will be conducted using quantitative metrics such as residual risk count, configuration compliance rate, and SBOM-based validation coverage. Finally, the effectiveness of the IDSS will be conclusively validated through human factors tests conducted in a high-fidelity APR1400 nuclear power plant training simulator environment.

## Acknowledgement

This study was supported by the Korean Institute of Energy Technology Evaluation and Planning (KETEP), and the Ministry of Trade, Industry, and Energy (MOTIE) of the Republic of Korea (No. 20224B10100130).

## Author Contributions

Conceptualization, Investigation, Methodology, Writing—Original Draft Preparation, Writing—Review & Editing, G.S.Jang; Supervision, Project Administration, S.R.Koo.

## References

[1] International Atomic Energy Agency (IAEA), "Enhancing nuclear power production with artificial intelligence," IAEA Bulletin, vol. 64, no. 3, 2023.

[2] G. S. Jang, and S. R. Koo, "Design challenges and response plans for intelligent decision support systems for Korean nuclear power plants under normal and abnormal conditions," Progress in Nuclear Energy, vol. 177, 105442, 2024. Available: https://doi.org/10.1016/j.pnucene.2024.105442.

[3] J. Yang, Y. Ming, W. Wenlin, and L. Fengjun, "Online application of a risk management system for risk assessment and monitoring at NPPs," Nuclear Engineering and Design, vol. 305, pp. 200–212, 2016.

[4] D. Bogdanov, P. Etti, L. Kamm, and F. Stomakhin, "Artificial intelligence system risk management methodology based on generalized blueprints," The 16th International Conference on Cyber Conflict (CyCon 2024), NATO CCD-COE, Tallinn, Estonia, pp. 123-140, 2024.

[5] J. S. Lim, S. H. Bae, and T. Kwon, "Applying NIST AI Risk Management Framework: Case study on NTIS database analysis using MAP, MEASURE, MANAGE approaches," Journal of Korean Society of Industrial and Systems Engineering, vol. 47, no. 2, pp. 21–29, 2024.

[6] P. Giudici, M. Centurelli, and S. Turchetta, "Artificial intelligence risk measurement," Expert Systems With

Applications, vol. 235, 121220, 2024. Available: https://doi.org/10.1016/j.eswa.2023.121220.

[7] U.S. Nuclear Regulatory Commission (NRC), Software Safety Hazard Analysis, NUREG/CR-6430, 1996.

[8] National Institute of Standards and Technology (NIST), Artificial Intelligence Risk Management Framework (AI 100-1), 2023. Available: https://nvlpubs.nist.gov/nistpubs/ai/nist.ai.100-1.pdf.

[9] U.S. Department of Energy (DOE), The Minimum Elements for a Software Bill of Materials (SBOM), 2021. Available: https://www.nrc.gov.

[10] Ministry of Science and ICT (MSIT) and Telecommunications Technology Association (TTA), 2023 Guidebook for Development of Trustworthy AI, 2023.

[11] U.S. Nuclear Regulatory Commission (NRC), Human-System Interface Design Review Guidelines, NUREG-0700, Rev. 3, pp. 283-295, 2019. Available: https://www.nrc.gov.

[12] Telecommunications Technology Association (TTA), Requirements for Enhancing the Trustworthiness of Artificial Intelligence Systems, TTAK.KO-10.1497, p. 7, 2023.

[13] International Organization for Standardization (ISO) and International Electrotechnical Commission (IEC), ISO/IEC 24028:2020 – Overview of Trustworthiness in AI, 2020.

[14] International Organization for Standardization (ISO) and International Electrotechnical Commission (IEC), ISO/IEC 23894:2023 – Information Technology – Artificial Intelligence – Guidance on Risk Management, 2023.

[15] Telecommunications Technology Association (TTA), SBOM Attribute Specification for Open-Source Software Supply Chain Management, TTAK.KO-11.0309, 2022. Available: https://tta.or.kr.

[16] Telecommunications Technology Association (TTA), Open Source SBOM Governance Management Guidelines, TTAK.KO-11.0322, 2023. Available: https://tta.or.kr.

[17] Korea Institute of Nuclear Nonproliferation and Control (KINAC), Computer and Information System Security in Nuclear Facilities, KINAC/RS-015, Korea, 2016.

[18] Nuclear Energy Institute (NEI), Cyber Security Control Assessments, NEI 13-10, Rev. 6, 2017.