Risk Management Plan for Intelligent Decision Support System of Korean NPPs

Gwi-sook Jang*, Seo Ryong Koo
Advanced Instrumentation & Control Research Division, Korea Atomic Energy Research Institute, 989-111, Daedeok-daero, Yuseong-gu, Daejeon 305-353, Republic of Korea
E-mail: gsjang@kaeri.re.kr

***Keywords:** risk management of AI system, intelligent decision support system, software bill of material

1. Introduction

The intelligent decision support system (IDSS) facilitates the decision-making process of the main control room operator by monitoring and diagnosing conditions, predicting progress, and providing preventive advice during both normal and abnormal operations of the nuclear power plant (NPP). AI-based decision support systems can pose different risks than traditional nuclear software development and validation. So, it is necessary to recognize the risk factors such as model misidentification, function malfunction, security and privacy issues, which can occur during the implementation and operation of artificial intelligence systems, and to analyze the size of the risk (severity and impact) and to prepare countermeasures.

2. Risk Management Plan for IDSS

2.1 Risk Management Framework of AI

In 2020, U.S. Congress directed NIST (National Institute of Standards and Technology) to create an AI risk management framework based on the National AI Act. As a result, the U.S. Department of Commerce's NIST released the "AI Risk Management Framework (AI RMF) 1.0" in January 2023. The AI RMF is a flexible and systematic framework for effectively managing the risks of AI systems and consists of four capabilities to help identify, assess, and mitigate the risks of AI systems.

The framework enables organizations to maximize the benefits of AI systems while minimizing their negative impacts. The AI RMF consists of four core functions of "govern, map (risk identification), measure, and manage" to create a trustworthy AI system, as shown in Table I, and each function is divided into categories and subcategories. Key features of AI RMF include the following.

- Flexibility: As the risks of AI systems vary, the AI RMF is designed to be flexible so that it can be adapted to the characteristics and circumstances of the organization.
- Systematic: The AI RMF is organized into four functionalities to help you systematically manage the risk of AI systems.
- Measurability: The AI RMF provides a baseline against which risk can be measured, enabling organizations to effectively manage the risk of AI systems.

The AI RMF is a useful framework for any organization adopting or operating IDSS. The AI RMF enables organizations to effectively manage the risks of IDSS and build trusted IDSS.

Table I: AI Risk Management Framework 1.0 (NIST)

Function	Description
Govern	•Defining organizational responsibility and authority for risk management of AI systems. •This enables organizations to establish policies, procedures, and resources to effectively manage the risks of AI systems.
Map (Risk Identification)	•Identifying the potential risks of AI systems. •To do this, organizations assess the AI system's purpose, capabilities, data, environment, etc.
Measure	•Evaluating the risks of AI systems. •To do this, organizations measure the probability and impact of risks.
Manage	•Mitigating the risks of AI systems. •To do this, organizations take actions to eliminate, minimize, or accommodate risks.

2.2 Risk Identification of IDSS

2.2.1 Risk of Using Open-source Software

Open-source software is increasingly being used to develop intelligent systems, and poorly secured open-source software vulnerabilities are highly vulnerable to supply chain attacks. In the design and development phase of IDSS, various open-source software can be used to shorten the development period and flexibly apply the latest technology. The representative issue from the legal aspect are licenses, and the representative issues from the technical aspect are compatibility and security vulnerabilities when using open-source software. When developing IDSS, it is crucial to manage a comprehensive list of software components to mitigate risks such as security vulnerabilities and copyright violations associated with the use of open-source software.

2.2.2 Risk of Adversarial Machine Learning

Adversarial machine learning is a machine learning attack technique that applies disruptive data to deep learning models, causing misclassification and reduced confidence. Types of attacks include poisoning attacks that manipulate the training data set, evasion attacks that fool the system by slightly changing the inputs of the trained model, extraction attacks that extract the model from the inputs and outputs of the model, and inference attacks that identify what data was used for training and exploit weaknesses or biases in the data.

IDSS can be exposed to attacks that intentionally corrupt training data or make minimal changes to input data to cause unexpected results during development or operation. Therefore, it is desirable to review and apply measures to cope with these attacks.

2.3 Software Bill of Material

The Software Bill of Material (SBOM) is a specification containing essential information about the final distributed and utilized software, designed for effective software supply chain management (see Table II). It typically encompasses packaged software, operating systems, and frameworks and libraries used in software development. In instances where challenges arise due to the utilization of open-source libraries, SBOM aids in identifying and managing potential issues arising from module dependencies and security vulnerabilities resulting from frequent module changes. Presently, the U.S. Department of Homeland Security's Cybersecurity Agency and NIST is regulating the management of SBOM. South Korea's National Intelligence Service is also planning to mandate SBOMs for public software.

Table II: SBOM Standard Attributes

Baseline	Attribute
①SBOM Validation Tool Name	ex) LG FOSSLight
②Supplier Name	Component Supplier:
③Author Name	Component Author:
④ Component Name	Component Name:
⑤ version String	Component Version:
⑥Unique Identifier	Format ID:
⑦Component Hash	File Check sum:
⑧License Name	Component License:
⑨License Usage	Dynamic/Static Linking:
⑩Vulnerability DB	Vulnerability DB: NVD
⑪Relationship	Include or Import Component:
⑫Release Date	Release Date:
⑬CVE* ID	CVE-Year-Serial Number:
⑭CVSS* Base Score	Base: , Impact: , Exploitability:
⑮CVSS Severity	High, Medium, Low, None

*CVE: Common Vulnerabilities and Exposure CVSS: Common Vulnerability Scoring System

2.4 Managing Guidelines of Open-source Software

Recently, TTAK.KO-11.0309[2], released in December 2022, titled "SBOM attribute standard for open software supply chain management," was established as a standard through collaboration between the Ministry of Science and ICT and the Software Policy Research Institute (see Table II). Also, TTAK.KO-11.0322[3], released in December 2023, titled "Open-source SBOM governance management guidelines".

In instances where challenges arise due to the utilization of open-source libraries, SBOM aids in identifying and managing potential challenges arising from module dependencies and security vulnerabilities resulting from frequent module changes.

2.5 Risk Management Plan for IDSS

The risk management process and plan for IDSS based on SBOM management and cybersecurity activities are shown in Figure 1 and Table III, respectively.

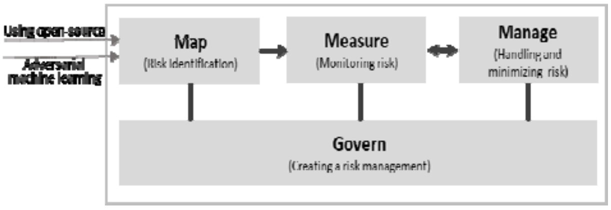
Using open-source
Adversarial machine learning
Map
(Risk Identification)
Measure
(Monitoring risk)
Manage
(Handling and minimizing risk)
Govern
(Creating a risk management)

Fig.1. Risk Management Process for IDSS

Table III: Risk Management Plan for IDSS

AI RMF 1.0 Functions	Risk Management Plan for IDSS
Govern	Define organizational structure, activities and responsibilities
Map (Risk Identification)	•Identify risks (using open-source software use and adversarial machine learning) •Create risk management plan for IDSS -SBOM management plan for IDSS -Cybersecurity assessment plan for IDSS (including development environment security guidelines)
Measure	•Create SBOM management implementation materials in the software lifecycle •Generate implementation materials for cybersecurity (including development environment security) in the software lifecycle •Report on risk management results in the software lifecycle -SBOM management implementation results -Cybersecurity assessment results
Manage	

First, the IDSS risk management plan defines the risk management organization and the responsibilities of each organizational unit for developing and validating the IDSS. The IDSS risk management plan then establishes a management plan for open source software and a cybersecurity assessment plan (including a secure development environment for the software) based on the IDSS risks defined in Section 2.2. Finally, the IDSS risk management plan checks whether the identified IDSS risk items are well implemented in the software lifecycle according to the management plan, takes countermeasures, and documents the results.

Open-source software can be used freely for developing and validating IDSS, but the license is stipulated separately. Therefore, IDSS developers should carefully check the license type and license notice of the open- source component they intend to use to understand their permission and obligations. Changing the open-source library version during IDSS development and validation may lead to incompatibilities with the development environment, language, tools, and other library versions. Therefore, when developing and validating of IDSS, the type and version of the open source library should be selected in consideration of compatibility, including identifying dependencies between libraries. Additionally, security vulnerabilities are sometimes found in open-source libraries used during IDSS development and validation.

To manage licenses, compatibility, and security vulnerabilities associated with open-source software use in NPP IDSS, as described above, select and apply relevant content from TTAK.KO-11.0309. Additionally, based on TTAK.KO-11.0322, continuously monitor open-source libraries for operational and security risks, and prepare a risk response implementation report. A detailed SBOM management process for IDSS is as shown in Table IV.

Table IV: SBOM Management Process for IDSS

Step	Activity	Documentation
Planning	•Utilize open-source software based on SBOM data fields determine whether to utilize open-source software (e.g., whether to implement a feature) •Open-source software licenses •Security vulnerabilities designing SW implementation methods within the scope of compliance with our SBOM usage policy •Determine whether to contract with a vendor •SBOM pre-review	SBOM management plan for IDSS
Implementation	•Open-source software usage list management •Review and reflect compliance with open-source software usage policy	Report on risk management results during implementation step
Evaluation	•SBOM validation through tools for each management area •Check SBOM validation results and violation measures •Management through the system (validation history and validation results)	Report on risk management results during evaluation step
Operation/maintenance	•Fulfillment of obligations for SBOM compliance (collection and notification of SW code to be disclosed, etc.)	Report on risk management results during operation/maintenance step

The defense against adversarial machine learning for IDSS is based on traditional NPP instrumentation and control system cybersecurity planning and implementation. The big data platform of IDSS is also an important security target. Therefore, security goals are needed for data confidentiality, integrity, and availability, along with a system capable of defending against threats like data theft and forgery from unauthorized users. The cybersecurity of IDSS should be based on security measure requirements KINAC/RS-015 [4] and security measure assessment methodology NEI 13-10 [5]. Based on this, cybersecurity evaluation guidelines should be created through an evaluation of the differential applicability of security measures, implemented accordingly, and a result report should be provided. IDSS development and validation should ensure that training data is unbiased and cannot be intentionally compromised, and reliable model validation should be performed to protect the integrity of AI models.

3. Conclusions

AI-based decision support systems can pose different risks than traditional nuclear software development and validation. Therefore, IDSS should have a systematic and measurable process in place to address the risks. The risks for IDSS of Korean NPPs include vulnerabilities due to the use of open-source software and attacks that threaten the confidentiality and integrity of AI training and operation. Open-source software vulnerabilities that have not been secured are highly vulnerable to supply chain attacks. Therefore, there is a growing need to identify which open-source software is used for which vulnerabilities. SBOM is one of the measures to manage and strengthen SW supply chain. Additionally, AI models and learning data can be stolen or abused by malicious users through the theft of training data and functions, or other types of attacks. Therefore, measures should be established to prevent or mitigate these attacks.

So, this paper established a risk management plan based on NIST RMF 1.0 for the use of open-source software and adversarial machine learning issues, which is a major issue in the development and validation of IDSS for Korean NPPs.

Acknowledgements

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korean government (Ministry of Science and ICT) (No. RS-2022-00144150) and the Korean Institute of Energy Technology Evaluation and Planning (KETEP) and the Ministry of Trade, Industry, & Energy (MOTIE) of the Republic of Korea (No. 20224B10100130).

REFERENCES

[1]NIST AI 100-1, AI Risk Management Framework, 2003.

[2]TTAK.KO-11.0309, SBOM Attribute Specification for Open Source Software Supply Chain Management, 2022.

[3]TTAK.KO-11.0322, Open Source SBOM Governance Management Guidelines, 2023.

[4]KINAC/RS-015, Computer and Information System Security in Nuclear Facility, 2016.

[5]NEI 13-10, Rev.6, Cybersecurity Control Assessments, 2017.