Extended Testing Strategies for Enhancing Reliability of an Intelligent Decision Support System Prototype in Nuclear Power Plants

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 : Intelligent decision support system, Computerized operator support system, Testing Strategy

1. Introduction

As the operational complexity of Nuclear Power Plants (NPPs) increases, the introduction of an Intelligent Decision Support System (IDSS) utilizing Artificial Intelligence (AI) has become essential. Traditional systems rely on fixed rule-based logic, which limits their ability to respond to non-linear abnormal situations. AI technology is emerging as a viable alternative to address these limitations. However, the probabilistic nature of AI conflicts with the stringent reliability and security requirements of the nuclear industry. This study proposes an extended testing strategy for the validation of an IDSS prototype designed to diagnose and predict normal and abnormal states of NPPs. To bridge the gap between AI's probabilistic characteristics and the nuclear industry's reliability requirements, this study introduces the separation of development and operational environments, data lifecycle validation, and a confidence-based safety mode transition mechanism. This study aims to compensate for the non-deterministic characteristics of AI-based systems through the four core testing strategies of the proposed IDSS prototype and to establish a system reliability validation methodology essential for the IDSS.

2. Testing limitations of the IDSS Prototype

The IDSS prototype serves as an advisory and pre-warning subsystem designed to support Main Control Room operators. While it is classified as Nuclear Non-Safety and Non-Class 1E and does not perform direct control actions, its potential integration with the APR1400 MMIS necessitates elevated reliability and traceability consistent with nuclear-grade expectations. Traditional validation methods, however, face several critical limitations when applied to AI-driven systems in a nuclear context.

- Conflict between AI probabilism and nuclear reliability: Standard nuclear systems rely on deterministic, fixed rule-based logic. In contrast, the AI-driven IDSS produces probabilistic outputs, which introduces significant validation challenges regarding inference uncertainty and operational acceptability. Conventional functional tests are insufficient to ensure that advisory outputs are conservatively blocked when confidence decreases or system latency exceeds predefined limits.
- Data lifecycle and physical signal integrity : Unlike conventional software where testing focuses on code logic, AI performance is fundamentally dependent on data quality. Traditional testing often overlooks the entire data lifecycle—from acquisition to inference. Specifically, there is a limitation in verifying whether preprocessing operations, such as outlier removal, preserve the physical signal characteristics essential for plant-relevant information.
- Supply chain transparency and cybersecurity : The IDSS relies heavily on open-source AI components, which introduces supply chain vulnerabilities that standard prototype testing does not typically address. Because the prototype environment does not replicate the full plant-wide cybersecurity infrastructure, validating essential controls—such as SBOM(Software Bill of Materials)-based asset transparency and vulnerability management—remains a constraint within conventional testing scopes.
- Constraints on environment separation and data usage : A critical requirement for nuclear safety is the strict separation between development (training) and operational (inference) environments. Standard testing methodologies often fail to functionally enforce or validate the integrity of this boundary, including the prevention of unauthorized artifact injection or runtime modifications. Furthermore, because the use of actual plant operational data is restricted, validation must rely on simulator datasets, requiring sophisticated real-time streaming or replay mechanisms to maintain regulatory compliance.

3. Extended Testing Strategies of IDSS Prototype

To align AI probabilistic behavior with nuclear reliability requirements, extended testing strategies of IDSS prototype was developed. The extended testing strategy integrates the separation of development and operational environments, data lifecycle integrity assurance, supply chain transparency, and probabilistic safety control mechanisms.

3.1 Input layer: Data lifecycle integrity validation

Given that AI performance is fundamentally dependent on data quality, validation extends across the entire data lifecycle—from acquisition to inference. Unlike conventional software testing, this strategy evaluates both data fidelity and model robustness. This lifecycle-oriented approach ensures that the AI model's learning foundation remains physically meaningful and operationally credible.

- Preprocessing operations such as outlier removal and interpolation are quantitatively analyzed to ensure preservation of physical signal characteristics.
- Statistical deviation metrics are calculated to confirm that refinement procedures do not distort plant-relevant information.
- Labeling reliability is evaluated through statistical consistency analysis and expert knowledge comparison. When imbalance mitigation techniques are applied, their influence on diagnostic performance is assessed to prevent overfitting or instability.

3.2 Asset layer: SBOM-based supply chain security control

Modern AI ecosystems rely heavily on open-source software, introducing supply chain vulnerabilities. To mitigate these risks, an SBOM-based security validation mechanism is implemented. The validation process includes:

- Comprehensive software component inventory analysis
- CVE-based vulnerability scanning and severity classification
- License compliance validation
- Risk-level-based mitigation and patch validation

By enforcing component transparency and traceability, the testing strategy reduces hidden dependencies and strengthens cybersecurity governance across the IDSS platform.

3.3 Platform layer: Separation of development and operation environments

A fundamental validation objective of the IDSS prototype is to demonstrate that the separation between the Development Environment and the Operational Environment is not merely architectural, but functionally enforced and verifiable. The Development Environment comprises the Big Data Platform and AI Platform, where data preprocessing, model training, and optimization are conducted. The Operational Environment consists of the IDSS inference server and display system operating within a closed network intended for advisory support to operators. The testing strategy focuses on validating the effectiveness and integrity of this separation boundary. The segregation is verified through the following structured validation mechanisms:

- Physical and logical network isolation validation
  Confirmation that no bidirectional communication channel exists between development and operational networks.
  - Validation that operational systems do not permit outbound training-related traffic.
  - Network scanning and penetration testing to confirm absence of unintended routing paths.
- Unidirectional artifact transfer control validation
  - Validation that only pre-approved, version-controlled model artifacts can be transferred from development to operation.
  - Validation of checksum/hash equivalence before and after transfer.
  - Simulation of unauthorized artifact injection attempts to confirm automatic rejection.
- Runtime immutability enforcement
  - Validation that deployed model artifacts cannot be modified, retrained, or replaced during operational runtime.
  - Validation that dynamic library loading outside the approved SBOM list is blocked.
  - Monitoring of file integrity to detect unauthorized changes.

3.4 Control layer: Intelligent Safety Control Mechanism

To manage inference uncertainty inherent in AI systems, a dual-layer intelligent safety control mechanism is incorporated and validated. First, Explainable AI (XAI) functionality ensures that operators can interpret the rationale behind system recommendations. Visualization of feature importance and decision logic supports transparency and operator trust. Second, a real-time confidence monitoring mechanism continuously evaluates inference reliability. If confidence values fall below predefined thresholds:

- Advisory outputs are automatically blocked
- The system transitions to Safety Mode
- Audible and visual alarms are activated

Safe mode transition behavior is tested under simulated latency spikes, communication disruptions, and internal software anomalies. Recovery validation ensures that normal advisory functions resume only after stability conditions are satisfied. Through this combined strategy, the IDSS maintains advisory functionality while ensuring that probabilistic uncertainty does not compromise nuclear operational safety.

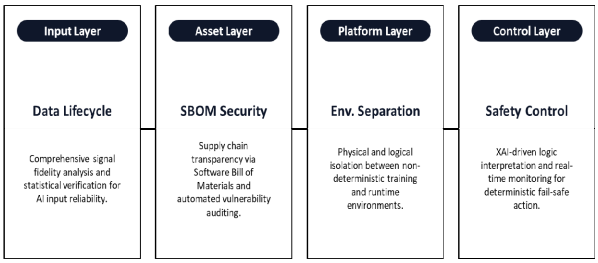

Fig. 1. Extended testing strategies of IDSS prototype

4. Conclusion

This study presented an extended testing strategy for an IDSS prototype intended for application in NPPs. The findings emphasize that conventional prototype validation, which primarily focuses on functional correctness and performance metrics, is insufficient for AI-driven systems operating in safety-critical environments. By identifying the specific limitations of traditional testing—such as the conflict between AI probabilism and nuclear reliability, and the lack of supply chain transparency—this research establishes a more rigorous validation framework. The proposed testing strategy incorporates several core pillars to mitigate these risks:

- Structural Safety Barriers: The enforced separation between adaptive development environments and deterministic operational systems acts as a critical barrier to ensure system stability.
- Integrity Assurance: Through data lifecycle validation and SBOM-based supply chain transparency, the strategy addresses vulnerabilities inherent in open-source AI ecosystems and data dependency.
- Reliability Mechanisms: The implementation of confidence-based safety control ensures that probabilistic uncertainty does not compromise operational safety by providing a safe-mode transition during low-confidence scenarios.

In conclusion, this study demonstrates that the modernization of instrumentation and control systems in NPPs through AI requires supplementary validation strategies tailored to AI-specific risks. The results of this study provide a foundation for establishing a robust system reliability validation methodology, ensuring that the integration of IDSSs aligns with the stringent safety and security requirements of the nuclear industry.

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] G.S.Jang, 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, 2024.
[2] G.S.Jang, S.R.Koo, Functional Requirements of an Intelligent Decision Support System Prototype in Korean NPPs Under Normal and Abnormal Conditions, Transactions of the Korean Nuclear Society Spring Meeting, Jeju, Korea, May 22-23, 2025
[3] G.S.Jang, S.R.Koo, Risk Management Plan for Intelligent Decision Support System of Korean NPPs, Transactions of the Korean Nuclear Society Spring Meeting, Jeju, Korea, May 9-10, 2024