

A Preliminary Study on Identifying the Human Role in the AI Era of Nuclear Power Plants: The Relationship Between AI-Based Support Systems and Human Operators

Yujin Kim^a, Cheolwoong Kim^a, Sinho Kim^a, Joon Eon Yang^a, JinHo Song^a and Sung Joong Kim^{a,b*}

^a Department of Nuclear Engineering, Hanyang University,
222 Wangsimni-ro, Seongdong-gu, Seoul 04763, Republic of Korea

^b Institute of Nano Science and Technology, Hanyang University,
222 Wangsimni-ro, Seongdong-gu, Seoul 04763, Republic of Korea

*Corresponding author: sungjkim@hanyang.ac.kr

***Keywords :** AI-Based Operation Support System, Human Performance, Human-Centered AI, AI Safety Regulation

1. Introduction

The rapid advancement of artificial intelligence (AI) technology is driving a paradigm shift across all industries, and active discussions on AI adoption are underway in the nuclear sector, which requires a high level of reliability and safety. Among these technologies, AI-based operation support systems that assist operators in decision-making within the main control room (MCR) are being actively developed as key technologies. These systems can simultaneously enhance the safety and operational efficiency of nuclear power plants by detecting early signs of anomalies and suggesting appropriate mitigation procedures.

However, nuclear power plants are high-risk, high-reliability systems where a single malfunction or judgment error can lead to an irreversible severe accident. Due to these characteristics, AI support systems applied to the MCR require far stricter safety and reliability standards than those in other fields. Therefore, establishing a clear boundary between the AI's level of autonomy and the human operator's authority to intervene must necessarily precede implementation.

Recent global AI regulatory frameworks, such as the newly enacted Framework Act on Artificial Intelligence in South Korea and the European Union's AI Act, classify technologies that directly affect human life and safety—such as nuclear power and autonomous driving—as "high-risk AI." They legally mandate the establishment of a practical human management and oversight framework throughout the entire operation process.

Nevertheless, AI technology, which is based on probabilistic judgments and inherently possesses "black-box" characteristics with low internal explainability, presents a fundamental difference in perspective from the current nuclear safety regulatory framework, which has traditionally adhered to deterministic methods and conservative evaluations. If such an AI system is deployed in the MCR, numerous cognitive and procedural factors must be considered to reach an optimal decision when the AI's predictions conflict with the skilled experience of human operators. However, a framework to evaluate this dynamic is currently lacking.

Therefore, this study aims to systematically identify the points of role conflict between humans and AI that may arise during the introduction of AI-based operation support systems in the MCR of nuclear power plants and to diagnose the current status and limitations of domestic and international regulatory frameworks responding to these issues. Furthermore, the study proposes a direction for establishing a human-centered control framework to secure the safe acceptability of AI in nuclear power plants.

2. Methods and Regulatory Framework Analysis

To systematically identify the potential issues arising from the adoption of AI systems, this study adopted a methodological approach based on a literature review and framework analysis. As discussed in the introduction, the integration of AI in nuclear power plants inevitably entails high risks. Therefore, this section analyzes the comprehensive regulatory frameworks established by the international community to control AI autonomy, as well as the specific safety regulatory requirements of the nuclear industry. These regulatory and ethical guidelines serve as the analytical framework for evaluating human-machine interactions in the subsequent section.

2.1 Global AI Regulations and Ethical Guidelines

As AI technology spreads throughout society, countries worldwide are establishing regulatory frameworks to prepare for the risks it may pose [1]. A representative example is the European Union's AI Act (Regulation (EU) 2024/1689), which entered into force in 2024. As the world's first comprehensive AI regulation, the EU AI Act mandates AI safety certification protocols and classifies AI according to risk levels: unacceptable, high-risk, limited risk, and minimal risk. Among these, "high-risk AI" explicitly includes AI that poses chemical, biological, radiological, and nuclear risks, as well as AI used in critical infrastructure directly affecting health and safety, such as nuclear power and transportation.

Providers of high-risk AI systems have a duty to identify, analyze, and evaluate foreseeable risks, as well as to conduct real-world testing and post-market surveillance. In particular, the Act mandates "effective

human oversight" to ensure that humans understand the system's operation, address automation bias, and can intervene, interrupt, or override the system when necessary. However, since the EU AI Act is cross-industry legislation, it does not completely replace nuclear safety regulations, necessitating parallel legal reviews specific to nuclear AI.

The Assessment List for Trustworthy AI (ALTAI, 2020), published by the High-Level Expert Group on AI (HLEG) of the European Commission, emphasizes "Human Agency and Oversight" as the first of seven requirements for trustworthy AI [2]. ALTAI categorizes the level of human oversight into three stages: HITL (Human-in-the-Loop), the ability of humans to intervene in every decision cycle of the system; HOTL (Human-on-the-Loop), the ability to monitor the system during operation and intervene in the design phase; and HIC (Human-in-Command), the ability to oversee the overall activity of the AI system and decide whether, when, and how to use it. Furthermore, it specifies special oversight for autonomous systems, adverse effect detection, and a "Stop Button" for safe shutdown as essential requirements.

In the United States, NIST proposed the characteristics of trustworthy AI through the AI Risk Management Framework (AI RMF 1.0, 2023) [3]. To ensure the safety characteristic, it mandates the ability to stop and modify systems that deviate from their intended functions, or the ability for humans to intervene, as an essential condition. It also highlights the "finality of human control" by specifying that human intervention must be included if the AI system cannot detect and correct errors on its own. In addition, the OECD AI Principles (revised in 2024) stipulate that AI should be designed and operated under the principles of being human-centered, explainable, safe, and accountable [4].

Table 1 summarizes the core keywords and concepts derived from these global regulatory frameworks. As illustrated in the table, international standards consistently mandate human agency—such as the ability to intervene, monitor, and safely halt AI systems—as a fundamental requirement for deploying high-risk AI technologies. This international consensus on human-centered oversight provides a crucial foundation for establishing AI regulatory guidelines in the nuclear sector.

Table 1. Core Keywords and Concepts of Global AI Regulations and Guidelines

Document	Keywords	Key Concepts
EU AI Act (2024)	Risk-based Classification	- Classifies nuclear as "High-risk AI"
	Effective Human Oversight	- Mandates authority to intervene, interrupt, or override

ALTAI (HLEG, 2020)	Human Agency and Oversight	- 3 stages of oversight: HITL, HOTL, HIC - Requires a 'Stop Button'
NIST AI RMF 1.0 (2023)	Trustworthy AI Finality of Human Control	- Authority to stop or modify unintended operations - Mandatory intervention for uncorrectable errors
OECD AI Principles (2024)	Human-centered Explainable & Accountable	- Human-centered design and operation - Transparency, safety, and accountability

2.2 Direction of AI Regulatory Frameworks in the Nuclear Industry

Through the *Nuclear Energy Series NR-T-1.26* published in 2025, the IAEA provided high-level guidelines encompassing the AI lifecycle, explainability and reliability, human factors, cybersecurity, and regulatory body readiness, which must be considered when deploying AI applications in the nuclear industry [5]. Based on the considerations outlined in the document, this study conceptualized the six core elements as illustrated in **Figure 1**. These considerations include the AI lifecycle, data governance, explainability and reliability, human factors, cybersecurity, and regulatory body readiness.

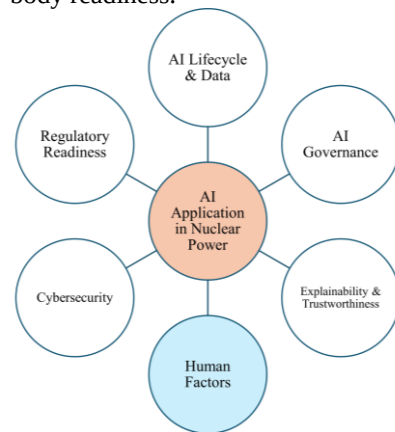


Fig. 1. Key Considerations for Deploying AI Applications in the Nuclear Industry (Summarized and Conceptualized from the IAEA [8])

The U.S. Nuclear Regulatory Commission (NRC) conceptually distinguished automation (pre-defined rules) from autonomy (data learning-based algorithms) in its AI Strategic Plan FY2023-2027 and systemized the

degree of autonomy into five levels [6]. As detailed in **Table 2**, the NRC stated that higher autonomy levels imply lower reliance on human intervention and oversight, thus requiring stricter regulatory scrutiny. The MCR AI operation support system would likely fall under 'Level 2 (Collaboration),' where the algorithm proposes recommendations and the operator verifies and executes them.

Table 2. Condensed AI and Autonomy Levels in Commercial Nuclear Activities [6]

Autonomy Level	Core Characteristics & Human-Machine Role
Level 0: AI Not Used	- No AI integration in systems or processes.
Level 1: Insight (Machine-assisted human)	- Used for optimization and guidance. - No impact on plant safety and control.
Level 2: Collaboration (Machine-augmented human)	- AI provides recommendations (potential safety impact). - Human vets and executes the final decision.
Level 3: Operation (Human-supervised machine)	- AI makes decisions and operates (potential safety impact). - Requires continuous human oversight.
Level 4: Fully Autonomous (No human intervention)	- AI operates, controls, and adapts independently. - No human intervention or oversight required.

Furthermore, in a joint statement of principles by the NRC, the UK's ONR, and Canada's CNSC, practical grades were classified based on the AI's failure significance and autonomy, as illustrated in **Figure 2** [7]. AI-based operation support systems, which do not have direct control capabilities but influence human judgment under time pressure, are classified as 'Grade R2.' This requires independent verification, Human Factors Engineering (HFE), and the ability to revert to existing procedures in the event of failure.

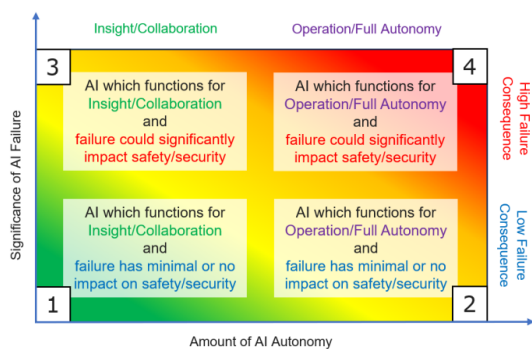


Fig. 2. Categorization of AI based on the Amount of Autonomy and the Significance of AI Failure [7]

In South Korea, although institutions like the Nuclear Safety and Security Commission are discussing AI regulation R&D, an AI regulatory framework specific to the nuclear sector has not yet been established. There is a concern that the locus of responsibility could become ambiguous between the dual regulations of the general "Framework Act on AI" and the "Nuclear Safety Act." Such a regulatory gap could lead to actual decision-making conflicts between on-site operators and AI.

3. Application of AI Operation Support Systems

By applying the regulatory perspectives established in Section 2 to the AI-based operation support system, this section analyzes the comprehensive impacts of AI integration. Specifically, it outlines the anticipated positive effects of AI adoption and subsequently details the specific human-AI role conflicts.

3.1 Positive Effects of AI Operation Support Systems

In the MCR of a nuclear power plant, there are emergency situations such as severe accidents and multiple failure accidents where the operator's cognitive load is maximized. In these situations, operators must interpret vast parameters and select response procedures within a short timeframe. AI operation support systems serve as a powerful auxiliary tool to alleviate this cognitive load. Particularly in severe accident scenarios, they can rapidly predict complex physical phenomena, such as hydrogen concentration changes or the progression of core damage, using machine learning techniques to support proactive mitigation procedures.

According to NUREG-CR-7294, advanced deep learning-based systems support operator decision-making and help establish optimal mitigation plans for plant conditions through accurate real-time predictions [8]. The IAEA also reports that multi-modal interfaces and real-time anomaly detection technologies are effective in lowering operator workload and reducing human errors [9].

3.2 Identification of key impact from AI system towards human (operator)

1. Trust and Reliance: An "automation bias," where AI judgments are accepted uncritically, or conversely, a confusion in decision-making where even correct AI recommendations are distrusted due to deep learning's black-box nature, may occur. This risks leading to inappropriate responses to diagnoses and predictions.

2. Degradation of Situation Awareness: If the AI exclusively handles monitoring and analysis, the operator may experience an "out-of-the-loop" phenomenon, becoming isolated from the system loop. Passive monitoring induces a decrease in the operator's vigilance and leads to complacency. Operators may lose the ability to respond intuitively in the event of an

accident without understanding the system's operating principles.

3. Intensification of Task and Interface Complexity:

When complex AI inference results are displayed alongside massive amounts of data, it can cause information overload during emergencies. AI systems must display not only simple outcome values but also the basis of inference, levels of uncertainty, and elements of Explainable AI (XAI) on the interface, which significantly increases the complexity of measurement and control displays compared to traditional systems.

4. Digital Fatigue and Ambiguity of Legal Responsibility: When an AI's prediction conflicts with an operator's empirical judgment, the need to make a final decision under extreme time pressure maximizes the operator's psychological stress and digital fatigue. In particular, the operation of existing nuclear power plants prioritizes compliance with rule-based procedures and guidelines; if these are legally followed, there is a clear standard that does not hold the operator directly accountable. However, if a problem occurs "despite taking action according to the AI's recommendation," the legal standard regarding whether this can serve as an exemption from regulatory bodies remains ambiguous. If this ambiguity in regulatory responsibility is not resolved, operators will inevitably feel a significant psychological burden in using the system.

Table 3 summarizes the dual impacts of AI integration in the MCR, contrasting the anticipated operational benefits with the corresponding human-system role conflicts.

Table 3. Summary of AI Integration Impacts on Operator Performance in the MCR

Category	Anticipated Benefits	Role Conflicts (Challenges)
Trust & Reliance	• Cognitive Relief • Rapid Prediction	• Automation Bias • Black-Box Distrust
Situation Awareness	• Anomaly Detection • Proactive Mitigation	• Out-of-the-Loop • Vigilance Decrement • Intuition Loss
Interface Complexity	• Condition Diagnosis • Plan Optimization	• Information Overload • XAI Display Burden
Fatigue & Accountability	• Error Reduction • Real-Time Support	• Digital Fatigue • Judgment Conflict • Regulatory Ambiguity

4. Conclusions

The integration of AI technology into the nuclear industry is an irreversible trend, which will create a new

paradigm in operations. For the coexistence of AI and human society, and for a safe nuclear era, securing the "reliability of the technology itself" must be the highest priority.

Similar to the global regulatory stance on AI, the application of AI in the nuclear sector must be grounded in the principle of "Human-Centered AI." When this principle is strictly adhered to, conservative and negative perceptions of AI adoption can be minimized. AI operation support systems, in particular, must remain in a supportive role, and it must be clearly defined that the final situational judgment, operational authority, and corresponding legal responsibility rest with the highly trained human operators.

In conclusion, before the full-scale deployment of AI in the field, legal frameworks must be preemptively established to ensure social consensus and technical safety. In alignment with global AI regulatory frameworks, this study recommends the prompt amendment of domestic AI-related laws reflecting the unique characteristics of the nuclear sector, as well as the establishment of a nuclear-specific AI safety evaluation system.

Acknowledgement

This work was supported by the Innovative Small Modular Reactor Development Agency grant funded by the Korea Government (MCEE) by the Nuclear Safety Research Program through the Korea Foundation Of Nuclear Safety (KoFONS) using the financial resource granted by the Nuclear Safety and Security Commission (NSSC) of the Republic of Korea (No. RS-2024-00404240 and No. RS-2026-25539178).

REFERENCES

- [1] S. Cha, Towards an international regulatory framework for AI safety, Humanities and Social Sciences Communications, Vol.11, p. 506, 2024.
- [2] HLEG on AI, The Assessment List for Trustworthy AI (ALTAI), European Commission, 2020.
- [3] NIST, AI Risk Management Framework (AI RMF 1.0), NIST AI 100-1, 2023.
- [4] OECD, OECD Due Diligence Guidance for Responsible AI, OECD Publishing, 2026.
- [5] IAEA, Nuclear Energy Series NR-T-1.26 – Considerations for Deploying AI Applications in the Nuclear Power Industry, 2025.
- [6] US NRC, NUREG-2261, Artificial Intelligence Strategic Plan, FY2023-2027, ML23132A305, 2023.
- [7] US NRC, UK ONR, and CNSC, Considerations for Developing AI Systems in Nuclear Applications, ML24241A252, 2024.
- [8] US NRC, NUREG/CR-7294, Exploring Advanced Computational Tools and Techniques with AI/ML Operating Nuclear Plants, 2022.
- [9] IAEA, AI for Accelerating Nuclear Applications, Science and Technology (ART-INTweb).