A Regulatory-Perspective Classification Framework for AI Applications in Nuclear Power Plants: A Multi-criteria Watchlist for Regulatory Triage

Hyeong-Ki Shin*
Korea Institute of Nuclear Safety, 62 Gwahak-ro, Yuseong-gu, Daejeon
*Corresponding author: hkshin@kins.re.kr

Keywords: artificial intelligence, nuclear power plant, regulatory classification, autonomy, watchlist, explainable AI

1. Introduction

The application of artificial intelligence (AI) in nuclear power plants (NPPs) has expanded rapidly across monitoring, diagnostics, operator support, design, and operational decision support. International bodies such as the IAEA have proposed technology categories and application areas to describe the expanding use of AI in nuclear power systems [1], and regulatory bodies have also begun examining the implications of AI deployment for nuclear oversight. However, systematic methods for translating the growing body of AI-related nuclear literature into actionable regulatory priorities remain limited. Most existing classification efforts describe what technology is used but provide limited indication of the degree of regulatory attention warranted. This study proposes a regulatory-oriented classification framework based on IAEA AI technology categories and application areas, supplemented by NRC autonomy levels, explainability considerations, and TRL-based response tiers.

2. Methods and Results

2.1 Classification Framework

Each publication is classified along two axes: (L1) AI technology type, based on IAEA NR-T-1.26 Section 2.1, and (L2) nuclear application area, comprising the 12 application categories described in IAEA Section 2.2. Each publication is further assigned an NRC notional AI/autonomy level according to the degree of human involvement in decision-making [2].

Table 1. L1 — AI technology classification

L1 Category	Representative methods / architectures	Source
Shallow Models	Decision Trees, Linear Regression, SVM, Bayesian Models, Shallow/single-hidden-layer ANN*, Genetic Algorithms†, k-Nearest Neighbors†	IAEA §2.1.1 + literature extension
Deep Learning Networks	CNN, deep/multilayer ANN*, GAN*, RNN/LSTM*, Deep Reinforcement Learning†,	IAEA §2.1.2 + literature extension
NLP / LLM	RAG*, Fine-tuning*, Agentic LLM*	IAEA §2.1.3 + literature extension

**Not explicitly named in IAEA NR-T-1.26; included based on established usage within the category. †Not explicitly named in the IAEA §2.1 category, but explicitly used elsewhere in IAEA NR-T-1.26*

Table 2. Reference mapping between L1 technologies and IAEA-based L2 application areas

	Application Area (L2)	Shallow Models	Deep Learning	NLP / LLM
1	Teaching and Learning	-	-	RAG, Fine-tuning, Agentic LLM
2	Monitoring and Diagnostics	Bayesian Models*, SVM	DNN, RNN/LSTM, CNN	-
3	Text Analytics using LLMs	-	-	RAG, Fine-tuning, Agentic LLM
4	Assistance to NDE Inspections	SVM	CNN, DNN*	-
5	Fault Diagnosis and Predictive Maintenance	Bayesian Models, SVM	DNN, RNN/LSTM, CNN	-
6	Operational and Decision Support	Genetic Algorithms*, Bayesian Models	DNN, DRL	Agentic LLM

	Application Area (L2)	Shallow Models	Deep Learning	NLP / LLM
7	Human Performance Optimization Tools for Operators	-	DNN, RNN/LSTM	Fine-tuning, Agentic LLM
8	Sensor State On-line Monitoring	Linear Regression	DNN, RNN/LSTM	-
9	Nuclear Safeguards Applications	Decision Trees*, k-Nearest Neighbors*, Linear Regression*	DNN, CNN	-
10	Operator Actions, Reliability Analysis and PRA	Bayesian Models	DNN, DRL	-
11	Applications for Future Plants	Genetic Algorithms	GAN*, DNN, Deep Reinforcement Learning*	Fine-tuning, Agentic LLM
12	Other Applications	Case-specific	Case-specific	Case-specific

** Explicitly identified in IAEA NR-T-1.26 for the corresponding application area; other mappings are study-defined extensions based on technical applicability and the reviewed literature.*

2.2 Classification Procedure and Watchlist Criteria

Classification follows a stepwise procedure: (1) identify the L1 technology from the stated methods, (2) determine the L2 application from the stated problem context, (3) validate the L1–L2 combination against the reference mapping in Table 2, which combines IAEA-identified examples with study-defined extensions, and (4) assign an autonomy level based on the described human–AI interaction. A publication is flagged for the watchlist if it satisfies any one of the three criteria summarized in Table 3. For screening purposes, L2 #2 and #4–11 were treated as regulatory-significant where AI outputs may affect safety, operations, inspection, human performance, safeguards, or future plant functions.

Table 3. Watchlist criteria (any one triggers a flag)

Criterion	Condition	Rationale
(i) NRC level ≥ 3 × regulatory-significant L2	NRC AI/autonomy level ≥ 3 (machine decision-making supervised by a human or with no human intervention) AND L2 #2 and #4–11 (“regulatory-significant L2”)	NRC autonomy Levels 3–4 shift primary decision-making from humans to the machine, increasing oversight needs in regulatory-significant applications
(ii) Limited explainability × regulatory-significant L2	Non-interpretable or insufficiently explainable AI AND L2 #2 and #4–11	IAEA §5.5 emphasizes that explainability should be commensurate with application criticality and that the meaning of explanation outputs should be understood when supporting safety claims. Limited explainability therefore warrants additional regulatory attention in regulatory-significant applications
(iii) Novel regulatory-relevant technology–application combination	L1–L2 or sub-technique combination not represented in Table 2 and introducing materially different regulatory characteristics	May introduce regulatory issues not covered by existing experience or guidance; algorithmic novelty alone is insufficient.

Publications that meet one or more of the Table 3 criteria are further assessed by Technology Readiness Level (TRL 1–9), summarized in Table 4. TRL is used here to determine the type and the timing of the regulatory response rather than as an independent watchlist trigger. A technology that remains at an early development stage is not yet deployed in an operational setting, whereas a technology approaching or entering operational use warrants greater case-specific regulatory attention. TRL is estimated using an adapted DOE-based screening approach that considers the validation environment, system fidelity, degree of integration, and scale described in each publication. As a practical screening rule, conceptual, analytical, or proof-of-concept studies generally correspond to TRL 1–3; integrated systems or prototypes evaluated in laboratory, simulated, or relevant environments generally correspond to TRL 4–6; and full-scale prototypical systems through actual operational systems generally correspond to TRL 7–9. These groupings are used as study-defined screening bands rather than as a substitute for a formal Technology Readiness Assessment under the U.S. DOE Technology Readiness Assessment Guide [3].

Table 4. TRL-based regulatory response tiers

TRL Band	Response Tier	Regulatory Action
Low (TRL 1–3, concept / analytical or experimental proof-of-concept)	Observation	Early-stage technology; track maturation and emerging regulatory relevance
Medium (TRL 4–6, integrated prototype validation in simulated or relevant environments)	Review Preparation	Deployment is becoming plausible; develop or refine applicable review criteria and guidance
High (TRL 7–9, full-scale prototype to actual operational system)	Priority Regulatory Review	Technology is approaching or has entered operational use; conduct case-specific regulatory assessment

Note: Response tiers and regulatory actions are study-defined and are applied only to literature already flagged by the watchlist criteria.

2.3 Application Results

The framework was applied to five representative NPP AI publications, summarized in Table 5. Hong et al. [4] was classified as Sensor State On-line Monitoring (L2 #8) and Monitoring and Diagnostics (L2 #2) with NRC autonomy Level 2 and was flagged only for limited explainability; its simulator-based validation corresponded to TRL 4–5 and the Review Preparation tier. Zhang et al. [5] was classified as Applications for Future Plants (L2 #11) with NRC Level 3 and was flagged for autonomy and regulatory-relevant novelty; it was assigned TRL 4–5 and the Review Preparation tier. Tunkle et al. [6] was classified as Operational and Decision Support (L2 #6) with NRC Level 4 and triggered all three criteria; its point-kinetics validation corresponded to TRL 3 and the Observation tier. El-Sefy et al. [7] was classified as Monitoring and Diagnostics (L2 #2) with NRC Level 1 and was flagged only for limited explainability; it was assigned TRL 3 and the Observation tier. Shin et al. [8] was classified as Monitoring and Diagnostics (L2 #2) with NRC Level 2 and triggered no watchlist criteria because its explanations were quantitatively evaluated and physically constrained; therefore, no TRL-based response tier was assigned.

Overall, four of the five studies were flagged for different regulatory reasons, while one was not, supporting the use of TRL as a response-tier modifier rather than an independent trigger.

Table 5. Applied classification results for five case studies

Case	L1 Technology	L2 Application	Autonomy	Watchlist Result
Hong et al. (2026) [4]	Deep Learning Networks (LSTM) + Shallow model	Sensor State On-line Monitoring (#8) + Monitoring and Diagnostics (#2)	NRC Level 2	1/3 met (ii) — flagged; TRL 4–5 → *Review Preparation tier*
Zhang et al. (2026) [5]	NLP/LLM (multi-agent)	Applications for Future Plants (#11)	NRC Level 3	2/3 met (i, iii) — flagged; TRL 4–5 → *Review Preparation tier*
Tunkle et al. (2025) [6]	Deep Learning Networks (DRL)	Operational and Decision Support (#6)	NRC Level 4	3/3 met (i, ii, iii) — flagged; TRL 3 → *Observation tier*
El-Sefy et al. (2021) [7]	Shallow Model (single-hidden-layer ANN)	Monitoring and Diagnostics (#2)	NRC Level 1	1/3 met (ii) — flagged; TRL 3 → *Observation tier*
Shin et al. (2025) [8]	Deep Learning Networks (CNN) with XAI	Monitoring and Diagnostics (#2)	NRC Level 2	0/3 met — not flagged; TRL-based response tier not applicable

3. Conclusions

This study demonstrates that combining IAEA-based AI technology/application classification with NRC-based autonomy assessment and explainability-oriented screening can transform routine literature classification into a practical regulatory triage framework. The proposed framework uses a two-stage triage process: a watchlist based on autonomy, explainability, and regulatory-relevant novelty, and a TRL-based response tier applied only to flagged studies. The response tiers distinguish Observation, Review Preparation, and Priority Regulatory Review according to technology maturity and proximity to operational deployment.

Application to five case studies showed that four were flagged for different regulatory reasons, while one was not flagged because its operator-supportive role and physically constrained explainability mitigated the relevant concerns. These results illustrate that AI use alone is not sufficient for watchlist inclusion; instead,

cases are differentiated according to decision authority, explainability, novelty, and maturity.

Because the criteria were evaluated on only five cases, further refinement is needed, particularly in distinguishing genuinely novel regulatory-relevant technology–application combinations from minor algorithmic variations. Future work will expand the literature set across additional technology and application categories, refine aggregation methods for larger-scale analysis, and explore expert-supervised large language model assistance for scalable literature screening.

REFERENCES

[1] International Atomic Energy Agency (IAEA), Considerations for Deploying Artificial Intelligence Applications in the Nuclear Power Industry, IAEA Nuclear Energy Series No. NR-T-1.26, 2025.

[2] U.S. Nuclear Regulatory Commission (NRC), Artificial Intelligence Strategic Plan: Fiscal Years 2023–2027, Rep. NUREG-2261, Office of Nuclear Regulatory Research, Washington, DC (2023).

[3] U.S. Department of Energy, Technology Readiness Assessment Guide, DOE G 413.3-4A, Washington, D.C., 2011.

[4] J.W. Hong, J.H. Park, M.G. Na, Fault identification and signal restoration of sensors based on artificial intelligence at nuclear power plants, Nucl. Eng. Technol. 58 (2026).

[5] E. Zhang, W. Liu, Y.-Y. Xu, C. Ye, Y. Xue, S. Qiang, G. Yuan, T. Liu, M. Zhong, Q. Li, H. Gong, Exploring a multi-agent framework with large language models for nuclear design automation, Energy AI 25 (2026).

[6] L. Tunkle, K. Abdulraheem, L. Lin, M.I. Radaideh, Nuclear microreactor transient and load-following control with deep reinforcement learning, Energy Convers. Manag.: X 27 (2025).

[7] M. El-Sefy, A. Yosri, W. El-Dakhakhni, S. Nagasaki, L. Wiebe, Artificial neural network for predicting nuclear power plant dynamic behaviors, Nucl. Eng. Technol. 53 (2021).

[8] J.H. Shin, J.S. Kang, J.M. Kim, S.J. Lee, Concept of understandable diagnostic cause visualization with explainable AI and multilevel flow modeling, Nucl. Eng. Technol. 57 (2025).