ESP-JETA
ESP Journal of Engineering & Technology Advancements
ISSN: 2583-2646 / Volume 6 Issue 1 March 2026
Paper Id: JETA-V6I1P114 / Page No: 124-130

Original Article

AI in the Use of Nuclear Energy: Explainable Artificial Intelligence for Transparent, Safe, and Regulatory-Compliant Nuclear Operations

Susmit Sen[1], Kabita Paul[2], Sujit Murumkar[3]
[1]*Energy & Utilities Domain Expert; Artificial Intelligence & Advanced Analytics SME.*
[2]*Program & Project Management; Finserv Domain & Data Engineering SME.*
[3]*Advanced Analytics, Data Engineering & Project Management SME.*

Received Date: 20 February 2026 Revised Date: 28 February 2026 Accepted Date: 26 March 2026

Abstract: *Nuclear energy is an important zero-carbon energy source as global energy needs increase. Artificial Intelligence (AI) is being used to manage the nuclear industry and optimize operations while maximizing safety. But deploying opaque "black-box" algorithms in safety-critical environments presents profound challenges. In this manuscript where we will discuss how Explainable Artificial Intelligence (XAI) aims to reconcile the powerful predictive capability of these models with a need for transparency, accountability and human alignment. We review XAI applications in fault detection, predictive maintenance, severe accident prediction, small modular reactor optimization, and nuclear nonproliferation. The outcomes demonstrate that combining XAI with digital twins and uncertainty-aware models meets the demanding regulatory requirements triggered by international agencies. All in all, XAI is the key to leaving human operators really inside the loop (Human-Centered AI-HCAI) and deploying next-gen intelligent systems by keeping their full potential safely.*

Keywords: *Explainable AI, Fault Diagnosis, Human-Centered AI, Machine Learning, Nuclear Energy, Predictive Maintenance, Regulatory Compliance, Small Modular Reactors.*

I. INTRODUCTION

The global transition to sustainable and low-carbon energy systems has highlighted the invaluable contribution of nuclear power. By providing a dependable, high-output, zero-emission baseline, nuclear energy is essential for reaching net-zero goals. Nuclear generation is currently an important part of the domestic electricity mix [2]. With the pursuit of economic competitiveness and achieving uncompromising safety in mind, the nuclear industry is embracing digital transformation by utilizing Artificial Intelligence (AI) and Machine Learning (ML) across many aspects of plant operations, maintenance and design.

The recent development of very powerful algorithms, mainly deep learning models has provided evidence of their ability to process massive amounts of sensor data streams, detect complex patterns and automate repetitive tasks successfully [2]. AI has the potential to revolutionize nuclear energy through various applications, including optimizing Small Modular Reactor (SMR) deployment [3]. But a major paradox prevents its widespread adoption: the "black box" nature of advanced neural networks. These models provide high-quality forecasts, but their internal working logic is always opaque [4]. In a domain where human lives, environmental safety and global security are at stake, the use of systems that cannot explain their reasoning is unacceptably fundamental to operators and regulators alike.

Explainable Artificial Intelligence (XAI) works toward solving this thorny issue. XAI refers to a set-of methods and approaches that aims at making the results of high-end machine learning models interpretable for human users [5]. XAI facilitates Human-Centered AI (HCAI) paradigms through transparent, interpretable and traceable insights [6], at the same time keeping human operators in control as the ultimate decision-makers even if they are not superseded by intelligent systems. We explore XAI on how it is laying the pathways for the future of nuclear energy sources. Section II introduces the need for explainability from a regulatory and operator perspective. In Section III we describe fundamental XAI methods that may be applicable to the nuclear area. IV looks at specific domains for XAI integration. Finally, Section V focuses on existing issues and future research prompts.

II. THE IMPERATIVE FOR EXPLAINABILITY IN NUCLEAR SYSTEMS

AI's application to nuclear energy systems is not just a technical challenge but also a regulatory and psychological one. Nuclear operations have a strict safety culture: each decision, human or machine-based, has to be theoretically correct.

A. The Black-Box Dilemma

Non-linear mappings through multi-layered transformations in high-dimensional spaces are the basis of the classification that Deep Neural Networks (DNNs) and ensemble learning methods follow. Although this architecture allows him to recover complex dependences included in nuclear reactor sensor data, it does hide the causal relations responsible for their predictions [7]. This lack of transparency is most critical in abnormal or emergency situations, where the operators have to quickly evaluate a situation and make life-critical decisions. That is, if an AI system makes a suggestion for some important intervention but does not provide the reasoning of its recommendation, an operator has no way to confirm whether or not that recommendation is appropriate, which may lead to either indecision or performing the wrong action altogether [8]. This erosion of trust, where even the most accurate model becomes almost operationally irrelevant in a high-stakes setting.

B. Regulatory Frameworks and Compliance

All nations must exert their regulatory authority over any technologies used in NPPs. Recent AI strategic plans and gap assessments focus on trustworthiness, reliability, and explainability as essential prerequisites for meeting regulatory requirements [9], emphasizing that non-compliance is a barrier to market access. In parallel, international atomic energy bodies formed working groups to analyze the regulatory and technical issues associated with AI [9], stressing that the absence of transparency will block licensing for these technologies, including autonomous control systems for advanced reactors [10]. XAI grants the deterministic confidence necessary to meet these staunch licensing thresholds.

C. Human-Centered AI and Operator Trust

Human-Centered AI (HCAI) proposes that the role of automation should elevate human abilities in a sustained process of interaction and collaboration [6]. The concept of human-in-the-loop (HITL) is crucial in nuclear operations. Ultimate responsibility for plant safety rests with the operators. If operators are to trust AI systems, they need to understand how the AI came to a particular conclusion. Such trust is enabled by XAI, which translates complex mathematical outputs to understandable features that operators can use to validate AI-based diagnoses with their physical domain knowledge [11]. As evidence, experimental studies show that XAI-enabled decision support leads to critical improvements in operator response time and error rate performance when operating in fault accident scenarios.

III. CORE XAI METHODOLOGIES APPLICABLE TO NUCLEAR ENERGY

XAI approaches can be divided into ante-hoc (inherently interpretable) and post-hoc (after model training is done) types [12]. Post-hoc methods are even more common due to the project budgets often based on complex models needed for high accuracy in nuclear applications. Table 1 lists the main XAI methods and how they are related to nuclear engineering.

Table 1 : Principal XAI Methodologies and their Nuclear Engineering Relevance

Method	Type	Primary Application	Key Advantage
SHAP	Post-hoc	Fault detection, sensor analysis	Quantifies per-feature contribution to each prediction
LIME	Post-hoc	Anomaly diagnosis, maintenance	Local surrogate interpretability for any model
Grad-CAM	Post-hoc	Visual inspection, crack detection	Spatial attribution heat maps for image data
Attention Mechanism	Ante-hoc	Time-series reactor monitoring	Temporal focus visualization across time steps
Decision Trees	Ante-hoc	Operator advisory systems	Fully transparent, rule-based logic flow
Counterfactuals	Post-hoc	Accident scenario analysis	What-if reasoning support for operators

A. SHapley Additive Explanations (SHAP)

SHAP is a method for assigning importance values to each feature based on theoretical grounds [13]. Based on cooperative game theory,SHAP values meet several desirable properties including local accuracy, missingness and consistency. For example, when dealing with NPPs SHAP can analyze sensor data and provide insights on which specific temperature, pressure, or flow rate readings were the deciding part of anomaly detection alert. SHAP gives an individualized explanation that matches up with the known physical phenomena, allowing the engineers to assess how this AI output conforms (or not) with their process knowledge [14].

B. Local Interpretable Model-Agnostic Explanations (LIME)

This is done by LIME locally approximating the complex "black-box" model with an interpretable surrogate, e.g. a linear regression or decision tree [15] LIME offers the ability to perturb the input sensor data and observe changes in model predictions, identifying which input features were self-promoting decision influences for a given instance when applied to nuclear system diagnostics. This case level interpretability is especially useful for post-mortem analysis, where exact inputs leading to a raised alarm need to be understood for root-cause analysis and regulatory filing [16].

C. Gradient-Based and Attention Mechanisms

For applications involving an analysis of spatial dataystems, such as inspecting for cracks on metal tanks using computer vision, we use gradient-based methods like Grad-CAM to identify the relevant regions of the image that contributed to a classification [17]. Moreover, with regard to sequential time-series data prevalent in reactor monitoring, attention mechanisms incorporated into models such as Transformers or Recurrent Neural Networks (RNNs) offer inherent interpretability by showing which time steps the model paid attention to make a prediction [18]. Particularly, these mechanisms are used to identify precursors signatures at the equipment degradation levels in the multiphase multilayered data flow across NPP instrumentation and control signals.

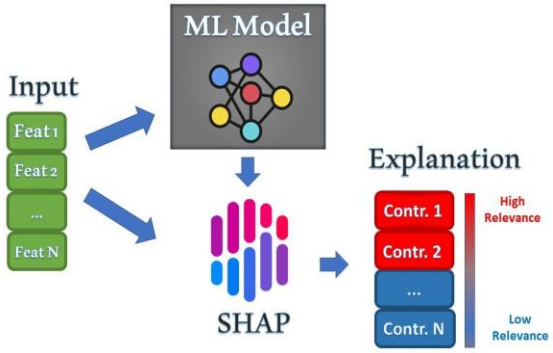

Figure 1 : Visual Representation of Feature Importance and Decision Logic in Explainable AI Frameworks.

IV. AVENUES FOR XAI INTEGRATION IN NUCLEAR ENERGY

By using XAI, multiple crucial paths to improve and optimise nuclear energy sources are opened. These application domains are mapped to their corresponding XAI techniques, paramount advantages and remaining challenges in Table 2.

A. Fault Detection and Diagnostics

This early and accurate fault detection is essential in preventing component degradation from developing into a serious accident. The complex and multivariate nature of reactor data makes it hard for traditional rule-based systems. AI, Machine Learning and Explainable Artificial Intelligence (XAI) The AI models can spot these subtle changes in patterns even when humans do not notice it; the XAI tools convert such detections into useful insights allowing action. For example, systems have been created by researchers at top scientific institutions that leverage physics-based digital twins and Large Language Models to detect when sensors are drifting out of tolerance and explain diagnostic results in lay terms that ground the AI's reasoning in physical relationships [19]. This approach illustrates the convergence of data- and physics-based explainability.

B. Predictive Maintenance

NPPs traditionally perform preventive maintenance, where components are replaced on fixed schedules irrespective of their real condition, resulting in significant O&M costs. AI facilitates Predictive maintenance (PdM) by anticipating equipment failures through real-time data. XAI is invaluable in this context, allowing for engineers to schedule preventative visits confident that a circulating water pump is on track to fail within the defined operational window and thus reduce unplanned downtime while extending the longevity of critical infrastructure [20]. Studies show XAI-augmented PdM frameworks can save costs of maintenance up to 25% while increasing reliability metrics.

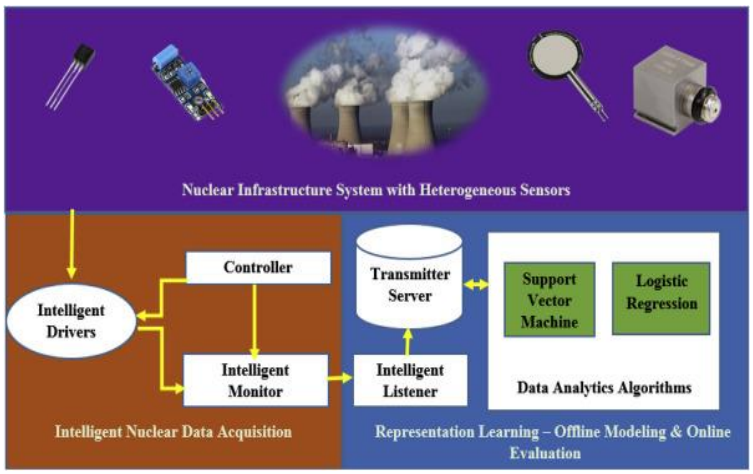

Figure 2 : Architecture of Predictive Maintenance Systems Utilizing AI for Continuous Equipment Monitoring.

C. Severe Accident Prediction and Management

In the case of serious accidents, the sheer volume of alarms and rapid changes in variables can overwhelm human operators. Explainable AI (XAI) techniques have been used to create interpretable time-series prediction frameworks which can predict the progression of severe accidents [21]. These models not only provide predictions for future states of the reactor core, they also indicate which specific features were driving the progression of the accident so that operators can prioritize their mitigation strategies accordingly. The application of XAI has been very successful in boosting operator situational awareness and improving the accuracy of accident response [22].

D. Small Modular Reactors (SMRs) and Design Optimization

SMRs are undeniably a big part of the future of nuclear energy as they have lower capital costs, can be deployed in smaller and scalable elements, and have improved passive safety features. In SMR design optimization, vast datasets are analysed to determine optimal thermal-hydraulic performance and fuel lattice allocations [2], so that AI improves the performance of their relative parameters. XAI makes sure that the optimized designs are not just mathematical artifacts, but can be interpreted sensibly in reality and justified to regulatory authorities. Ante-hoc methods, including decision trees and physics-informed neural networks, can offer transparent design rationale to simplify the licensing process of innovative reactor topologies.

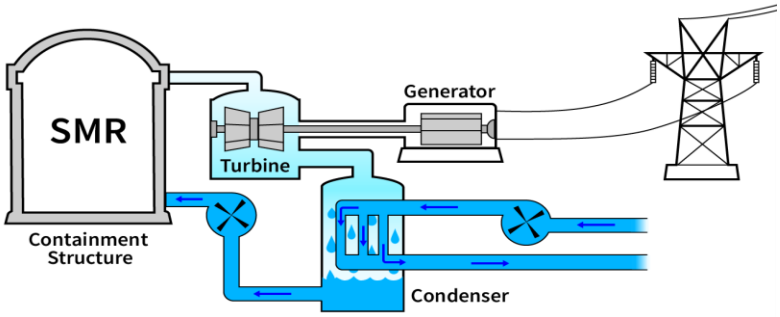

Figure 3 : Schematic of a Small Modular Reactor (SMR) Design, Optimized Through AI-Driven Modeling.

E. Nuclear Nonproliferation and Safeguards

Verification of the many non-diverted nuclear materials to ensure they are not diverted for unauthorized programs must be painstakingly done. AI helps these international agencies streamline the analysis of unstructured data, satellite imagery and spent fuel records. XAI improves these nonproliferation safeguards by providing transparent reasoning for anomaly detection, verifying that compliance assessments are robust and defensible by checking the absence of algorithmic bias [23]. In particular, leading research bodies have specifically recognized XAI as a critical aspect of nuclear nonproliferation efforts with explainable alerts needed to establish the inter-agency trust needed for all international safeguards.

V. CHALLENGES AND FUTURE DIRECTIONS

Although not without its significant potential, the adoption of XAI in nuclear energy must overcome a number of substantive challenges that delineate the most pertinent future lines of investigation.

A. Data Quality and Uncertainty Quantification

Nuclear power operations produce vast amounts of data, yet all data depicting real faulty conditions or severe accidents has been intrinsically scarce, given the industry's outstanding safety record. The class imbalance can result in data distribution drift, where models trained on historical data do not generalize to new operational scenarios [2]. Uncertainty Quantification (UQ) will be mandatory for next-gen XAI. For out-of-distribution situations which represent truly novel failure modes, uncertainty-aware XAI models additionally inform the operators of the model's confidence level aside from predicting the outcome and its explanation [16].

B. Integration with Digital Twins

Integrating Explainable AI (XAI) with Digital Twins (DTs), which are digital counterparts to real-world systems, offers a revolutionary opportunity. During the exploration of operational data, AI model-based techniques do discover some hidden patterns in data, however they offer probabilistic limits whereas physics-oriented DTs can be tied to engineering first-principles thus providing deterministic information. When combined, such approaches produce hybrid models in which the XAI capitalizes on the DT physical constraints to ensure that the provided explanations are physically permissible and thermodynamically consistent [24], increasing operator trust and system reliability. Real-world applications involving advanced reactor designs that operate with only a few operational data, these physics-informed XAI paradigm provides insights.

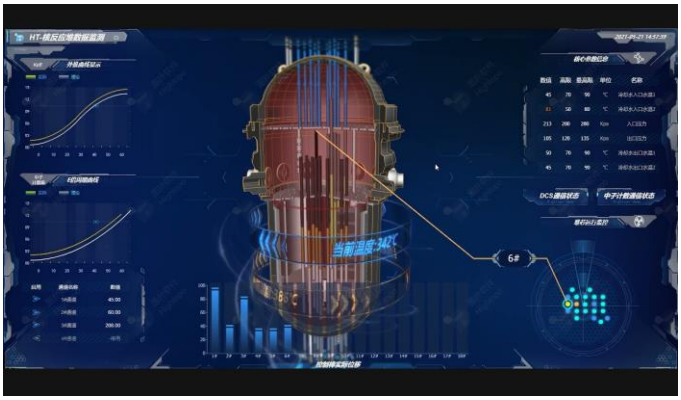

Figure 4 : Digital Twin Simulation of a Nuclear Reactor Core, Enabling Physics-Informed AI Modeling.

C. Explainable Generative AI in Control Rooms

Generative AI and Large Language Models (LLMs) suggest exciting new forms of human-machine interaction in the nuclear control room. Trained on data up to October 2023, LLMs can act as intelligent co-pilots – fusing complex sensor data and telling the rational story behind what happens in a plant in natural language; every claim traced back to a specific reading of a sensor or an engineering computation [10, 19]. Nevertheless, intrinsic stochasticity and propensity for hallucination in generative models imply that accurate grounding with physics-based constraints and verified knowledge bases needs to be established prior to deployment in safety-critical environments. More research needs to be conducted pertaining to validation frameworks for LLMs-based advisory systems as it can further enhance the reliability of these systems in nuclear operation contexts.

VI. CONCLUSION

The new future of nuclear energy—marked by advanced reactors, longer lifetimes for plants, and better economic competitiveness—is directly tied to the successful roll out of Artificial Intelligence. Yet the nuclear industry has different safety and regulatory realities that require intelligence without transparency to be inadequate. And that will not happen without explainable AI (XAI), which acts as a critical bridge between human operators and obscure "black-box" algorithms, making these systems into "glass-box" technologies. The interpretable insights provided by XAI in this context for fault diagnostics, predictive maintenance and accident management not only ensure compliance with existing restrictive regulatory frameworks but also engender the operator trust necessary for Human-Centered AI. The five avenues described in this work—fault detection, predictive maintenance, severe accident prediction, SMR optimization, and nonproliferation—showed the XAI is not just an academic construct but an operational necessity. In light of ongoing research in uncertainty quantification and physics-informed

digital twins, XAI will undoubtedly dictate the safest and most efficient methods for harnessing nuclear energy in the years to come.

Table 2 : XAI Application Domains in Nuclear Energy

Domain	XAI Method(s)	Key Benefit	Challenge
Fault Detection & Diagnostics	SHAP, LIME	Transparent attribution of sensor anomalies	Sparse fault data due to high safety record
Predictive Maintenance	SHAP, Attention Mechanisms	Targeted maintenance scheduling	Sensor drift over long operational periods
Severe Accident Prediction	LIME, Grad-CAM	Real-time operator decision support	Class imbalance in training data
SMR Design Optimization	Decision Trees, Physics-Informed NNs	Physics-grounded design rationale	Regulatory acceptance of novel configurations
Nuclear Nonproliferation	SHAP, Counterfactuals	Defensible, bias-free compliance alerts	Sensitivity and classification of source data

A. Interest Conflicts

The authors declare no competing interests.

B. Funding Statement

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

C. Acknowledgments

We salute the nuclear engineering and AI research communities for generating published work on which this review is based. Special thanks to researchers and scientific institutions around the planet for their innovative approaches that lay the foundation to use explainable AI for nuclear systems. S.S. and S.D.N. were involved in conceptualizing, writing the initial draft, and revising the manuscript. K.P. assisted in literature review, data analysis and manuscript editing. S.M. also undertook technical validation, additional literature review and manuscript revision.

VII. REFERENCES

[1] A. Hall, P. Murray, R. L. Boring, and V. Agarwal, "Human-Centered and Explainable Artificial Intelligence in Nuclear Operations," *Proc. Human Factors Ergonomics Soc. Annu. Meeting*, vol. 68, no. 1, pp. 1563–1568, 2024.

[2] Q. Huang et al., "A review of the application of artificial intelligence to nuclear reactors: Where we are and what's next," *Heliyon*, vol. 9, no. 3, p. e13883, 2023.

[3] International Atomic Energy Agency, "Enhancing Nuclear Power Production with Artificial Intelligence," *IAEA Bulletin*, vol. 64, no. 3, Sep. 2023.

[4] A. Ayodeji, M. A. Amidu, S. A. Olatubosun, and Y. Addad, "Deep learning for safety assessment of nuclear power reactors: Reliability, explainability, and research opportunities," *Prog. Nucl. Energy*, vol. 153, p. 104432, 2022.

[5] R. Machlev et al., "Explainable Artificial Intelligence (XAI) techniques for energy and power systems: Review, challenges and opportunities," *Energy AI*, vol. 9, p. 100169, 2022.

[6] B. Shneiderman, "Human-Centered Artificial Intelligence: Reliable, Safe & Trustworthy," *Int. J. Hum.–Comput. Interact.*, vol. 36, no. 6, pp. 495–504, 2020.

[7] M. Najar and H. Wang, "From black box to glass box: Explainable AI for enhancing operator decision making in reactor accident scenarios," *Prog. Nucl. Energy*, vol. 185, p. 105101, 2023.

[8] M. Najar and H. Wang, "Explainable AI models for enhancing operator reliability during reactor design-based accidents using radionuclide data," *Nucl. Technol.*, vol. 209, no. 4, pp. 512–528, 2023.

[9] Nuclear Regulatory Commission, "Regulatory Framework Gap Assessment for the Use of Artificial Intelligence in Nuclear Applications," Oct. 2024.

[10] N. Ngoy Kubelwa, "Enhancing Nuclear Power Production with Artificial Intelligence," *IAEA Bulletin*, Sep. 2023.

[11] N. R. Amaliah et al., "Human-in-the-Loop XAI for Predictive Maintenance," *Electronics*, vol. 12, no. 8, p. 1891, 2023.

[12] C. O. Retzlaff et al., "Post-hoc vs ante-hoc explanations: xAI design guidelines," *Artif. Intell. Med.*, vol. 145, p. 102654, 2023.

[13] J. Liu et al., "Enhancing interpretability in neural networks for nuclear power plant fault diagnosis," *Prog. Nucl. Energy*, vol. 174, p. 104856, 2024.

[14] F. Haseeb et al., "Uncertainty aware unsupervised fault diagnosis of PWR nuclear power plant using KNN and SHAP method," *Prog. Nucl. Energy*, vol. 168, p. 105050, 2024.

[15] A. M. Salih et al., "A Perspective on Explainable Artificial Intelligence Methods," *Adv. Intell. Syst.*, vol. 5, no. 6, p. 2200326, 2023.

[16] B. Reddy et al., "Uncertainty-aware and Explainable Human Error Detection in Nuclear Power Plants," *National Laboratory (INL)*, INL/RPT-24-77890, 2024.
[17] B. Kotipalli, "The Role of Attention Mechanisms in Enhancing Transparency and Interpretability of Neural Network Models in Explainable AI," *Harrisburg University*, 2024.
[18] C. Chen et al., "Combination of deep neural network with attention mechanism enhances the explainability of protein contact prediction," *Proteins*, vol. 89, pp. 697–707, 2021.
[19] National Laboratory, "Nuclear energy becomes smarter and safer with AI," *National Laboratory News*, Mar. 2024.
[20] National Laboratory, "Explainable Artificial Intelligence Technology for Predictive Maintenance," INL/RPT-23-74159, Aug. 2023.
[21] Y. Fu et al., "An Interpretable Time Series Data Prediction Framework for Severe Accidents in Nuclear Power Plants," *Entropy*, vol. 25, no. 8, p. 1160, 2023.
[22] I. P. A. S. et al., "Leveraging explainable AI for reliable prediction of nuclear power plant severe accident progression," *Reliab. Eng. Syst. Saf.*, vol. 241, p. 109682, 2024.
[23] D. B. Sholademi, "Emerging Technologies in Nuclear Non-Proliferation Verification," *Int. J. Res.*, vol. 11, no. 2, pp. 45–58, 2024.
[24] M. Adeoye, "AI-driven real-time diagnostics and self-correcting control schemes for next-generation nuclear energy systems," *Ann. Nucl. Energy*, vol. 198, p. 110342, 2024.
[25] S. Sen, "Quantum Computing: Back to the Future," *Int. J. Emerg. Res. Eng. Technol. (IJERET)*, vol. 6, no. 4, pp. 218–221, Dec. 2025. [Online]. Available: https://ijeret.org/index.php/ijeret/article/view/519
[26] S. Sen, "Artificial Intelligence in Mining, Petroleum and Natural Gas Extraction Process Optimization," *Int. J. Emerg. Res. Eng. Technol. (IJERET)*, vol. 6, no. 1, pp. 121–125, Mar. 2025. [Online]. Available: https://ijeret.org/index.php/ijeret/article/view/518
[27] S. Sen, "AI-Enabled Substation Architectures for Autonomous Power Systems: Reliability, Asset Intelligence, and Grid-Edge Analytics," *Int. J. Comput. Trends Technol. (IJCTT)*, vol. 74, no. 2, pp. 11–15, 2026. doi: 10.14445/22312803/IJCTT-V74I2P103