Strål
säkerhets
myndigheten

Swedish Radiation Safety Authority

Research

Application of artificial intelligence technologies and machine learning for nuclear power plants: a review

2025:11

Author: Mattias Håkansson, Claire Blackett
Risk Pilot AB
Date: December 2025
Report number: 2025:11
ISSN: 2000-0456
Available at www.ssm.se

Strål
säkerhets
myndigheten
Swedish Radiation Safety Authority

Author: Mattias Håkansson, Claire Blackett
Risk Pilot AB

2025:11

Application of artificial intelligence technologies and machine learning for nuclear power plants: a review

Date: December 2025
Report number: 2025:11
ISSN: 2000-0456
Available at www.stralsakerhetsmyndigheten.se

SSM perspektiv

Bakgrund

Idén om att tillämpa beräkningsintelligens inom kärnkraftsindustrin för olika tillämpningar är inte ny, med tillämpningsexempel från tidigare 1990- och 2000-tal. Med utvecklingen och populariteten av artificiell intelligens (AI) börjar många forskare överväga att tillämpa AI-teknik i kärnkraftverk inom olika områden i ett tidigt skede. AI-teknik skulle potentiellt kunna upptäcka utrustningsfel på ett bättre sätt (genom feldiagnos), minska stress genom att hjälpa operatörer under onormala förhållanden. AI-teknik är särskilt viktig för nästa generations reaktorkonstruktioner, men de flesta tidigare tillämpningar har endast skräddarsytts för specifika detaljerade scenarier. Det är av intresse för SSM att finansiera forskning om AI och maskininlärning för att stödja vidareutveckling och användning av AI- och ML-applikationer i industrin. Det är också viktigt för SSM att lära sig mer om dessa avancerade tekniker till stöd för SSM:s tillsyns- och godkännandeprocesser där AI och ML sannolikt kommer att vara en del av både utformningen och driften av olika kärntekniska anläggningar.

Resultat

Studien bygger främst på en litteraturstudie och deltagande i en konferens om AI. Litteraturgenomgången innehåller exempel och fallstudier av ML-tillämpningar inom olika områden inom kärnkraftsindustrin, såsom säkerhets- och riskanalys, anläggningsdrift och underhåll. Den belyser potentialen hos ML för att förbättra kärnsäkerheten proaktivt. Det understryker också det internationella samfundets stora intresse för att anta och/eller ytterligare etablera nya metoder för att förbättra säkerheten. Men den identifierar också utmaningar som dataproblem och den ”svarta lådan” karaktären hos vissa metoder som måste lösas för att uppnå allmän acceptans för dessa avancerade metoder för kärnsäkerhet.
Det pågående forsknings- och implementeringsarbetet belyser den omvälvande inverkan som dessa tekniker kan ha eller redan har på olika aspekter av kärnkraftverkens drift, inklusive tillämpningsområden som prediktivt underhåll och säkerhetsanalys.
En slutsats från litteraturstudien är att det kanske finns ett behov av en mer harmoniserad klassificering/kategorisering av tillämpningar. Ännu viktigare är att bristen på vägledning om vilken algoritm eller metod som ska användas, dataöverväganden och tillgänglighet, beroende på applikationen, är uppenbar och detta kan vara en utmaning i det framtida antagandet eller övergången till ML-metoder eller för att etablera dem ytterligare.

Relevans

Projektet ger en utökad bild av den nuvarande användningen av AI och ML inom kärnkraftsbranschen. Det finns också många AI/ML FOU-projekt inklusive benchmarkingaktiviteter som beskrivs i litteraturen. SSM är medvetet om den snabba utvecklingen inom detta område och kommer i enlighet med rekommendationerna i rapporten att fortsätta att följa utvecklingen inom AI och ML. Genom att hålla sig informerad om de senaste forskningsframstegen och tillämpningarna kan organisationer inte bara förbli konkurrenskraftiga utan också identifiera och anta innovativa lösningar som stärker säkerheten och minskar risken. Projektet understryker att SSM:s hittills små steg när det gäller användning av AI, i branschen, men även för interna ändamål till stöd för förbättrad ändamålsenlighet och ändamålsenlighet, behöver stärkas.

Behov av ytterligare forskning

Inga specifika ytterligare åtgärder bedöms vara nödvändiga på kort sikt, men SSM behöver följa utvecklingen och sprida kunskap om AI/ML i organisationen inklusive avdelningarna/sektionerna för tillsyn och licensiering av kärntekniska anläggningar.

Projektinformation

Kontaktperson SSM: Per Hellström
Referens: SSM2023-4526 / 4530635

SSM perspective

Background

The idea of applying computational intelligence in nuclear industry for different applications is not recent, with application examples from the earlier 1990s and 2000s. With the development and popularity of Artificial Intelligence (AI), many researchers start to consider applying AI technologies in Nuclear Power Plants in various fields at an early stage. AI technologies could potentially better detect equipment failures (through fault diagnosis), reduce stress by assisting operators in abnormal conditions. AI technology is especially important for next-generation reactor designs, but most previous applications have been tailored for only specific detailed scenarios. It is of interest for SSM to finance research on AI and Machine Learning in order to support further development and use of AI and ML applications in the industry. It is also of importance for SSM to learn more of these advancing technologies in support of SSM oversight and approval processes where AI and ML are likely to be part of both the design and the operation of various nuclear facilities.

Results

The study is based mainly on a literature review and participation in a Topical conference on AI. The literature review includes examples and case studies of ML applications in various areas in the nuclear industry, such as safety and risk analysis, plant operation, and maintenance. It highlights the potential of ML to enhance nuclear safety proactively. It also emphasizes the international community's keen interest in adopting and/or further establishing novel approaches to enhance safety. However, it also identifies challenges such as data issues and the "black box" nature of some methods that need to be solved to achieve general acceptance of these advanced methods for nuclear safety.

The ongoing research and implementation efforts highlight the transformative impact these technologies could have or already has on various aspects of nuclear plant operations, including application areas such as predictive maintenance and safety analysis.

One conclusion from the literature review is that there is perhaps a need for a more harmonized classification/categorization of applications. More importantly, the lack of guidance of which algorithm or method to use, data considerations and availability, depending on the application is evident and this could be a challenge in the future adoption or transition to ML approaches or to establish them further.

Relevance

The project provides an extended view of current use of AI and ML in the nuclear business. There are also many AI/ML R&D projects including benchmark activities that are described in the literature. SSM is aware of the rapid development in this area and as recommended in the report, SSM will maintain continuous monitoring of developments in AI and ML. By staying informed about the latest research advancements and applications, organizations can not only remain competitive but also identify and adopt innovative solutions that strengthen safety and reduce the risk. The project underlines that SSMs so far small steps regarding use of AI, in the industry, but also for internal purposes in support of improved effectiveness and efficiency, need some strengthening.

Need for further research

No specific further actions are deemed necessary in the short term, but SSM need to monitor the development and disseminate knowledge about AI/ML in the organization including the departments / sections for oversight and licensing of nuclear facilities.

Project information

Contact person SSM: Per Hellström
Reference: SSM2023-4526 / 4530635

Application of artificial intelligence technologies and machine learning for nuclear power plants: a review

SAMMANFATTNING

Med utvecklingen och populariteten av artificiell intelligens (AI) har många forskare börjat överväga att tillämpa AI-teknologier inom kärnkraftsindustrin inom olika områden. AI-teknologier och maskininlärning har potentialen att kunna appliceras inom flertalet områden och även inom reaktorsäkerhet. Dock finns det utmaningar att övervinna, speciellt den "black box"-natur som många maskininlärningsmetoder har, för att uppnå ett accepterande av att använda dessa avancerade metoder för reaktorsäkerhet.

Rapporten utforskar, via en litteraturstudie, användningen av artificiell intelligens (AI) och maskininlärning (ML) inom kärnkraftsindustrin, med fokus på säkerhetsanalys. Den betonar att AI-teknologier kan exempelvis detektera avvikelser innan fel i utrustning uppstår och ge beslutstöd till operatörer vid driftstörningar.

AI är den övergripande termen som omfattar olika tekniker för att skapa intelligenta maskiner. ML är en delmängd av AI som fokuserar på att utveckla algoritmer som kan lära sig automatiskt från data. Djupinlärning (DL) är en delmängd av ML som använder neurala nätverk för att lära sig från stora datamängder. ML/DL kan tillämpas inom flera områden, exempelvis systemdesign och -analys, anläggningsdrift och underhåll, samt säkerhets- och riskanalys.

Några av de viktigaste utmaningarna inkluderar:

- Att säkerställa att data som används för att träna ML-modeller är av hög kvalitet och tillgängliga i tillräckliga mängder, vilket är avgörande för att kunna göra korrekta och pålitliga förutsägelser.
- Att noggrant validera och verifiera ML-modeller för att säkerställa att de fungerar korrekt och pålitligt i olika scenarier under olika förhållanden och med olika typer av data.
- Att göra ML-modeller mer förklarbara och transparenta. Detta är särskilt viktigt inom säkerhetsanalyser, där det är avgörande att förstå hur och varför en modell gör vissa förutsägelser.
- Att säkerställa att ML-modeller kan integreras på ett effektivt sätt med operatörer och att de kan användas för att stödja beslutsfattande på ett pålitligt sätt.

Sammanfattningsvis är de viktigaste punkterna följande:

- Det rekommenderas att kärnkraftsindustrin bör fortsätta utforska och implementera AI och ML för att förbättra säkerheten och effektiviteten, och investera i forskning för att övervinna utmaningar.
- Etablera samarbete mellan aktörer och sträva efter standardisering av riktlinjer och säkerhetskrav för AI och ML-applikationer.

AI och ML har stor potential att förbättra säkerheten, genom ökad tillgänglighet och tillförlitlighet inom kärnkraftsindustrin, men det finns också betydande utmaningar som måste övervinnas.

SUMMARY

With the development and popularity of artificial intelligence (AI), many researchers have begun to consider applying AI technologies within the nuclear power industry in various areas. AI technologies and machine learning have the potential to be applied across multiple domains, including nuclear safety. However, there are challenges to overcome, especially the "black box"-nature of many machine learning methods, to achieve acceptance of using these advanced methods for nuclear safety.

The report explores, via a literature study, the use of artificial intelligence (AI) and machine learning (ML) in the nuclear industry, with a focus on safety analysis. It emphasizes that AI technologies can, for example, detect anomalies before equipment failures occur and provide decision support to operators during disturbances.

AI is the overarching term that includes various techniques for creating intelligent machines. ML is a subset of AI that focuses on developing algorithms that can automatically learn from data. Deep learning (DL) is a subset of ML that uses neural networks to learn from large amounts of data. ML/DL can be applied in several areas, such as system design and analysis, plant operation and maintenance, as well as safety and risk analysis.

Some of the key challenges in using AI and ML include:

- Ensuring that the data used to train ML-models is of high quality and available in sufficient quantities, which is crucial for making accurate and reliable predictions.
- Carefully validating and verifying ML-models to ensure that they function correctly and reliably in different scenarios under various conditions and with different types of data.
- Making ML-models more explainable and transparent. This is particularly important in safety analyses, where it is crucial to understand how and why a model makes certain predictions.
- Ensuring that ML-models can be effectively integrated with operators and that they can be used to support decision-making reliably.

In summary, the key points are as follows:

- It is recommended that the nuclear industry should continue to explore and implement AI and ML to improve safety and efficiency and invest in research to overcome challenges.
- Establish collaboration between stakeholders and strive for standardization of guidelines and safety requirements for AI and ML-applications.

AI and ML have great potential to enhance safety by increasing availability and reliability within the nuclear industry, but there are also significant challenges to overcome.

Contents

List of Figures

List of Tables

1 Introduction

1.1 Background

The idea of applying computational intelligence in nuclear industry for different applications is not recent, with application examples from the earlier 1990s and 2000s. [1] With the development and popularity of Artificial Intelligence (AI), many researchers start to consider applying AI technologies in Nuclear Power Plants (NPPs) in various fields at an early stage. AI technologies could potentially better detect equipment failures (through fault diagnosis), reduce human operating pressure by assisting operators in abnormal conditions. AI technology is especially important for next-generation NPP design, but most previous applications have been tailored for only specific detailed scenarios. [2]

AI and Machine Learning (ML) can be used in several ways, for example:

<u>Data analysis</u>
ML algorithms can analyse large amounts of data to identify patterns, trends, and correlations. This can include historical data, real-time data, and external data sources.

<u>Risk prognosis</u>
Predictive ML models can be developed to estimate the probability of a specific event occurring. For example, ML algorithms can analyse historical data on accidents or failures to identify patterns and factors that contribute to those events. These models can then predict the likelihood of similar events happening in the future, allowing for proactive risk management.

<u>Decision support</u>
ML models can be used to develop decision support systems that aid in making risk-informed decisions. These systems can analyse various scenarios and assess their associated risks, allowing decision-makers to choose the optimal course of action. For example, ML algorithms can analyse different risk mitigation strategies and recommend the most effective approach based on historical data and its predictive capabilities.

<u>Anomaly detection</u>
ML can be used to detect anomalies or unusual events that may indicate increased risk. For instance, ML algorithms can analyse sensor data to identify deviations from normal behaviour that may indicate a potential equipment failure.

<u>Topical conference on AI and ML</u>
During the fall 2023, the organisation International Association for Probabilistic Safety Assessment and Management (IAPSAM) arranged a topical virtual conference on *Artificial Intelligence and machine learning* [3]. Included topical areas of interests were the following:

- AI and Machine Learning (ML) to support risk analysis and risk-informed decision-making.
- Automation trustworthiness and transparency for AI-based automation technologies.
- Uncertainty quantification for AI and ML technologies.
- Risk-informed design and regulation of AI and ML technologies.
- Human reliability analysis for a human-machine interface with AI and ML technologies.
- Prognostics and Health Management using AI and ML.
- Digital twins for risk analysis, assessment, and management.
- Interpretability of AI and ML technologies for safety-critical applications.

- AI and ML for condition-based risk assessment.
- AI and ML for safety measures optimization.

This demonstrates that the use of AI and ML for nuclear safety has a clear potential to enhance and/or transform the nuclear reactor safety proactively. The international community seems very keen to adopt novel approaches to enhance safety. However, many applications also identify challenges to overcome, like data issues and the "black box" nature of some methods to reach a general acceptance of using these advanced methods for nuclear safety.

1.2 Artificial Intelligence, Machine Learning and Deep Learning

The hierarchy of Artificial Intelligence (AI), Machine Learning (ML), and Deep Learning (DL), see Figure 1, can be understood as follows:

- Artificial Intelligence (AI): AI is the overarching system that encompasses various techniques and methods to create intelligent machines and systems. Machine Learning and Deep Learning are both subfields of AI.
- Machine Learning (ML): ML is a subset of AI that focuses on developing algorithms with the ability to automatically learn and improve on the basis of data or experience, without being explicitly programmed. There are several types of ML, generally divided into four groups: Unsupervised learning, Semi-supervised learning, Supervised learning, and Reinforcement learning.
- Deep Learning (DL): DL is a subset of ML in which multi-layered neural networks, modelled to work like the human brain, to "learn" from large amounts of data. Within each layer of the neural network, DL algorithms perform calculations and make predictions repeatedly, progressively "learning" and gradually improving the accuracy of the outcome over time. DL is differentiated in that it can process unstructured, unlabelled data.

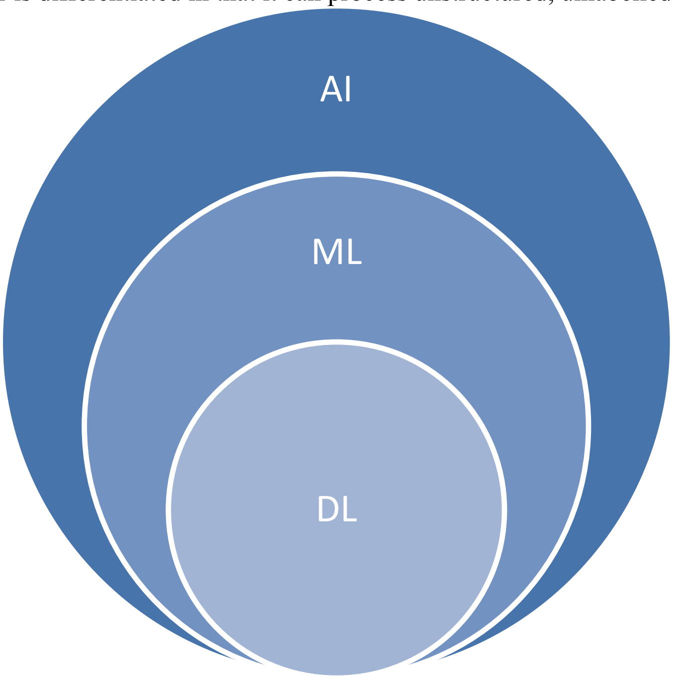

Figure 1 Hierarchy of Artificial Intelligence (AI), Machine Learning (ML), and Deep Learning (DL). [4]

ML can be applied in several types of application fields in the nuclear industry, such as:

- Plant operation and maintenance. For example,
    - Plant degradation modelling, fault, and accident diagnosis and prognosis,
    - Plant operation and maintenance efficiency improvement.
- Nuclear safety and risk analysis. For example,
    - Plant safety assessments including component and system reliability, external events, and severe accidents.
    - Plant security assessments including cybersecurity and physical security.
- Reactor system design and analysis. For example,
    - Reactor thermal hydraulics,
    - Reactor physics,
    - Reactor system performance.

1.3 Scope, Objectives and Limitations

The report is intended to provide a comprehensive overview of the current and potential applications of AI and ML in the nuclear industry, as well as the challenges and opportunities. However, the report also acknowledges the limitations and uncertainties inherent in this emerging, fast-developing, field.

AI and ML can help to increase the quality of both deterministic and probabilistic safety analyses by better identifying inaccuracies, both in terms of documentation and input data. They can also be essential tools for collecting data from other departments, identifying trends, and thereby enabling even more initiative-taking work with risks. In other words, the use of AI and ML could have a major impact on future safety analyses.

The objectives of the project are as follows:

- Identify and evaluate AI and ML within the nuclear industry for safety analysis via a literature review.
- Identify and recommend application areas for AI and ML.
- Analyse what challenges exist with AI and ML and how to solve or minimize these.

Limitations that should be considered when reading this report are:

- This report does not address applications for nuclear security assessments including cybersecurity and physical security.
- ML and DL are rapidly evolving fields that are constantly producing new methods, models, and applications. Therefore, this report is limited to information available when writing.
- The report does not endorse any specific ML technique or vendor over another but rather presents a broad overview of the available options and use cases.
- ML cannot solve all issues related to nuclear safety, plant operation and maintenance. It is an algorithm or model that try to augment human expertise and enable more efficient and effective decision making, but it also requires careful validation, verification, and testing to ensure its reliability and safety.
- ML is highly dependent on the quality and quantity of data that is used to train and evaluate the models. The availability and accessibility of data may vary depending on the type of plant, the regulatory framework, the proprietary interests, and the ethical and privacy concerns. The report does not address these issues in detail but rather presents use cases and their potential.
- ML is influenced by the context and objectives of the plant operators and maintainers, as well as the stakeholders and regulators involved. The report does not account for the

specific needs, preferences, and constraints of each individual plant or organisation, but rather offers general insights and their potential.

- The report also does not address the legal, social, or ethical implications of using ML in nuclear safety, plant operation and maintenance, which may require further investigation.

1.4 Structure of the report

The report consists of eight chapters that cover the following:

- Chapter 1 introduces the motivation and objectives of the report, as well as the scope and limitations of the study.
- Chapter 2 provides an overview of advanced computational tools and techniques that can be used for data-driven modelling and optimization. It presents the main concepts and techniques of ML, such as supervised, unsupervised, and reinforcement learning, and the types of algorithms and models used for different tasks, such as classification, regression, clustering, and optimization. It also introduces common metrics and methods for evaluating the performance and robustness of ML models, such as accuracy and precision.
- Chapter 3 presents the project's literature review. It also presents examples and case studies of ML applications in various aspects of the nuclear industry, such as safety and risk analysis, plant operation and maintenance.
- Chapter 4 discusses the challenges and the future of AI and ML in the nuclear industry. It identifies the main barriers and limitations for applying ML techniques to safety systems, such as data availability and quality, model validation and verification, interpretability, and human-machine interaction.
- Chapter 5 provides and discuss future research directions.
- Chapter 6 concludes the insights from the project.
- Chapter 7 list the references cited in the report.
- Chapter 8 provides categorised tables with reoccurring abbreviations used in the report.

2 An Overview of Machine Learning (ML)

2.1 The Relationship between Statistics and ML

Within nuclear safety, operating experience is collected to be able to estimate different parameters such as component reliabilities, initiating event frequencies, Common Cause Failure (CCF) parameters and to conduct component and system trend analysis. Statistics is normally used to estimate these types of parameters which are used as input in Probabilistic Safety Assessment (PSA).

By looking at the relationship between statistics and ML, it is possible to see where the use of ML could be beneficial. Statistics and ML are linked in terms of methodological principles but are different in their primary goals. Statistics have a focus on inference by modelling the data generation process to formalize understanding, whereas ML concentrates on prediction to identify the best course of actions with no or limited understanding of the underlying mechanism. Statistics is a subfield of mathematics while ML is a subfield of computer science and grew out of AI to focus on learning from data.

Statistical methods have traditionally been used on smaller data sets, in cases where the entire population of data is not known. Advanced ML methods require much more data than the traditional statistical methods but can predict when relationships are more complex. Some of the ML algorithms, e.g., Bayesian Networks (BNs) and Gaussian Processes (GPs), are also popular approaches in statistics. However, ML approaches such as DL sacrifice some degrees of interpretability for predictive power.

Both fields (statistics and ML) involve data analysis but differ in their approaches, goals, and methods. They are similar in the sense that both rely on data to extract insights, both use probability and optimization techniques, and aim to find patterns in data (though their end goals differ).

Traditional statistics aims to find inference, e.g., to understand the relationship between variables and to test hypothesis. It is suitable to use this in case of small datasets since it can yield robust insights or in cases where explainability is of importance, i.e., when clear, interpretable models are required for decision-making.

ML aims to optimize models to obtain as accurate predictions as possible. It works well with high-dimensional data with many variables or unstructured data (e.g., images, text). It requires large data volumes to train complex models. It is ideal for non-linear or unknown variable relationships or when models need to improve automatically over time.

2.2 Overview of ML paradigms

ML can be subdivided into different paradigms of learning, see Figure 2, and the main types are:

- Supervised Learning
- Unsupervised Learning
- Reinforcement Learning

Other paradigms within ML include semi-supervised, rule-based, self-supervised, batch, meta, online, and quantum machine learning. Each of these paradigms has its own advantages and applications, depending on the specific problem and the amount of available labelled data. This is further explored in chapter 3 with examples of applications using these paradigms.

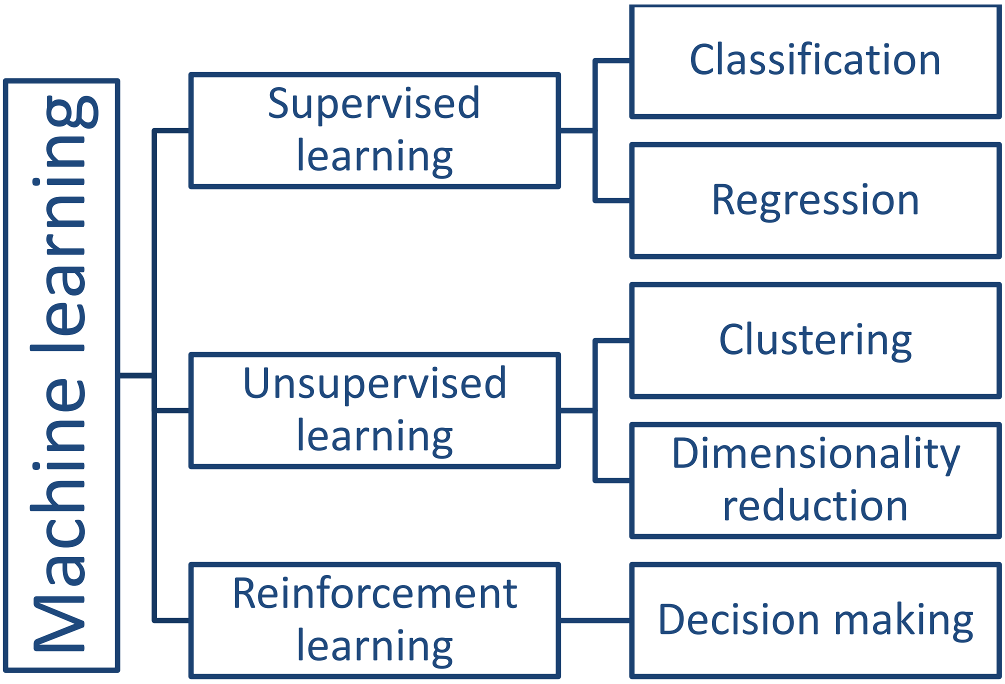

Figure 2 Machine learning paradigms and methods (based on [5]).

2.2.1 Unsupervised Learning

Unsupervised learning is a paradigm in ML where algorithms learn patterns exclusively from unlabelled data. This approach is used to group data based on the underlying hidden features in the data. There are three main approaches in unsupervised learning:

- Clustering: Clustering algorithms group unlabelled data based on their similarities or differences. Common clustering methods include hierarchical clustering, k-means, Gaussian mixture models, DBSCAN (Density-Based Spatial Clustering of Applications with Noise), and OPTICS (Ordering Points To Identify the Clustering Structure). These methods help identify commonalities in the data to create groups.
- Anomaly Detection: Anomaly detection methods identify data points that do not fit into any group or cluster. These methods can help detect outliers or anomalies in the data. Examples of anomaly detection methods include Local Outlier Factor (LOF) and Isolation Forest (IF).
- Learning Latent Variable Models: Latent variable models aim to represent complex data in a lower-dimensional space while preserving the essential characteristics of the original data. These models can be used for dimensionality reduction, feature extraction, and data visualization. For example, Principal Component Analysis (PCA).

In unsupervised learning, all one has is a set of data samples without being told their expected labels (ground truths) for categorical variables, nor the true numeric values for continuous variables. Unsupervised learning methods are promising in many applications due to three major reasons.

1. Labelling a large dataset can be surprisingly expensive and time consuming.
2. Find features that can best represent the data which will be useful for future prediction tasks.

3. Gain insights into the structure of the data, i.e., the understanding of the probability density and subgroups, which can help influence the design for data classification and regression applications.

In summary, unsupervised learning is a powerful tool for extracting insights from data without the need for labelling data.

2.2.2 Supervised Learning

Supervised learning is a type of ML where the algorithm learns from labelled input and output data. The goal of supervised learning is to learn a function that maps input data to output data based on the relationship between the input and output data. There are two main types of supervised learning:

- Classification: In classification tasks, the goal is to predict the categorical class or label of an instance. Some common classification algorithms include Logistic Regression (LR), Decision Trees (DTs), Support Vector Machines (SVMs), and Random Forests (RFs).
- Regression: In regression tasks, the goal is to predict a continuous value. Some common regression algorithms include linear regression, polynomial regression, and DTs.

Supervised learning algorithms learn from labelled data by minimizing the difference between the predicted output and the actual output. This is done by adjusting the model's parameters to minimize the error between the predicted output and the actual output. The performance of a supervised learning algorithm is evaluated using metrics such as accuracy, precision, recall, and F1 score (for classification), and mean absolute error (MAE) and root mean square error (RMSE) (for regression).

Supervised learning implies a training data set that contains the observed values of the variable of interest. The observed values can be either categorical (labels), discrete, or continuous. Supervised learning implies the availability of a labelled training dataset that consists of a set of training samples. In its most common form, each data sample pair has an input feature vector and a desired output value (label). A supervised learning algorithm learns the underlying model (or inferred function) between the input and the output using the training set, and the requirement is that the model should be able to generalize from the training set to unseen data samples.

A wide collection of supervised learning algorithms is available, each with its strengths and weaknesses. The most widely used learning algorithms include Artificial Neural Networks (ANNs), GPs, BNs, SVMs, DTs, RFs, and various models within the subfield of DL such as Convolutional Neural Networks (CNNs).

2.2.3 Reinforcement Learning

Reinforcement learning (RL) is a ML technique that trains software to make sequences of decisions. It is based on the Markov Decision Process (MDP) framework, where an agent takes actions in an environment and receives positive, negative, or zero rewards based on its actions. The agent learns to maximize the cumulative reward over time through a trial-and-error learning process.

Key concepts in RL include the agent (the ML algorithm), the environment (the problem space), actions (steps the agent takes), state (the environment at a given time), and reward (the value associated with taking an action).

There are several types of RL algorithms that can be used to train agents to interact with dynamic environments and maximize rewards. Some of the most common types of reinforcement learning algorithms include:

- Value-based methods: These algorithms estimate the value of each state or state-action pair and use this information to determine the best action to take. Examples of value-based methods include Q-learning and SARSA (State-Action-Reward-State-Action), which is explained below.
- Policy-based methods: These algorithms directly learn the optimal policy, or sequence of actions, to take in a given state. Examples of policy-based methods include "REINFORCE" (Monte Carlo method) and Actor-Critic methods (hybrid method which combines value-based with policy-based methods).
- Model-based methods: These algorithms learn a model of the environment and use this model to plan future actions. Examples of model-based methods include Dyna-Q and Monte Carlo Tree Search.

Q-learning is a type of reinforcement learning algorithm used to train an agent to make decisions in an environment by maximizing the expected cumulative reward. It is a model-free, value-based, off-policy algorithm that learns the optimal action-value function, or Q-value, for each state-action pair. The Q-value represents the expected future reward that an agent can obtain by taking a particular action in a given state. The Q-learning algorithm updates the Q-value iteratively as the agent interacts with the environment and receives rewards.

SARSA is an on-policy reinforcement learning algorithm used to understand the Markov decision process policy. In SARSA, the Q-value is updated taking into account the action performed in the next state, unlike in Q-learning where the action with the highest Q-value in the next state is used to update the Q-table. SARSA is used to learn a policy that balances exploration and exploitation and can be applied in various domains, such as decision making.

2.3 ML Languages and Tools

Different programming languages are associated with the ML algorithms. Python, C++, and R are among the most popular programming languages used for ML, [6]. Each language has its own strengths and weaknesses, making them suitable for different tasks and applications.

- **Python:** It is a fast-growing, general-purpose programming language known for its readability and structure. It has a vast ecosystem of libraries and frameworks for ML and data analysis, such as Pandas, NumPy, and TensorFlow. Python is widely used in the data science community and is suitable for developing ML models.
- **C++:** It is a flexible, object-oriented, mid-level language based on the C programming language. It can directly interact with hardware under real-time constraints and is suitable for parallel computing. C++ is often used in industries that require low-level access to system resources, such as software development and data processing.
- **R:** It is a top choice for many data scientists as a language and environment for statistics, visualization, and data analysis. R has numerous built-in statistical and graphical techniques and can be extended with ML packages. It is widely used in academic research and data analysis.

There are various ML tools, platforms, and software available for data analytics and visualization, such as Python Pandas, NumPy, KNIME, TensorFlow, PyTorch, Accord.net, Google Cloud AutoML, and Jupyter notebooks, [6]. Commercial off-the-shelf software like SAS and MATLAB also offers powerful ML capabilities, with prebuilt functions, extensive toolboxes, and specialized apps for classification, regression, and clustering. Results from these software's are generally trusted.

2.4 Natural Language Processing

Natural Language Processing (NLP) involves a series of techniques and methods to enable computers to understand, process, and generate human language. Depending on the problem, supervised, unsupervised and reinforcement learning methods as well as DL models are used. NLP is an emerging field in the nuclear industry with increasing interest and applications.

Common advanced techniques in NLP involve:

1. Rule-Based Methods: Using predefined linguistic rules.
2. Machine Learning: Using algorithms to learn from data, such as Naive Bayes (NBs), SVMs, and decision trees.
3. Deep Learning: Utilizing neural networks, for example, Recurrent Neural Networks (RNNs), Long Short-Term Memory (LSTMs), and Transformers, e.g., Bidirectional Encoder Representations from Transformers (BERT), Generative Pretrained Transformer (GPT), which are pre-trained language models.

Some of the key NLP techniques are:

1. Text pre-processing. This consist of tokenization (breaking the text into smaller units), lemmatization (reducing words to its base form), normalization (handling abbreviations etc. to standardize the text) etc.
2. Syntactic analysis. This involves determining relationships between words in a sentence, parsing by analysing the grammatical structure of sentences using algorithms.
3. Semantic analysis. This includes for example determining the correct meaning of a word based on context.

3 ML Applications in the Nuclear Industry

3.1 Literature Review

A review of published literature surveys of applications and use cases of AI in different areas in the nuclear industry have been conducted. In the following references, different compilations of list of applications and use cases of AI can be found. This review does not aim to be complete, but more of to present the status and the efforts made so far with ML and DL methods in the nuclear industry.

- *A review of the application of artificial intelligence to nuclear reactors: Where we are and what's next* [5]
  This paper lists previous application of AI in nuclear reactor design optimization, nuclear reactor operation and maintenance (O&M). Applications of AI to nuclear reactor design optimization include nuclear reaction core design, thermal-hydraulic simulation analysis and radiation shielding design. Applications of AI to nuclear reactor O&M include online condition monitoring, fault diagnosis and predictive maintenance. For each application, the used ML method and overall findings are presented.
- *Status of research and development of learning-based approaches in nuclear science and engineering: A review* [7]
  In this paper, popular ML methods are evaluated against different criteria regarding their suitability. It also presents an algorithm selection scheme for nuclear and radiological data criteria.
- *Nuclear Power Plants With Artificial Intelligence in Industry 4.0 Era: Top-Level Design and Current Applications – A Systemic Review* [2]
  This paper categories AI-related nuclear power applications into Physical-Plant-Centred and Human-Operator-Centred technologies and review research works from 7 typical NPP functional scenarios in the recent two decades and covers 106 research papers. The functional scenarios include nuclear fuel management, nuclear data processing, autonomous control for fixed procedure, fault detection and diagnosis, human-machine interaction, emergency alarming and decision-making assistance. Representative AI techniques within the two main categories are summarized.
- *Artificial intelligence in nuclear industry: Chimera or solution?* [1]
  This paper presents applications of ML with presentation of their focus and highlights. It also summarizes research on applications of AI in fuel management, fault diagnosis, transient identification, and accident scenarios. For each application, the used technique, type of application and findings are presented.
- *Exploring Advanced Computational Tools and Techniques with Artificial Intelligence and Machine Learning in Operating Nuclear Plants* [6]
  In this report, a review of applications is summarized in tabular form with presentation of used ML method, data category and the objective of the application. The applications are divided into reactor thermal hydraulics, reactor physics, reactor system performance, plant operation and maintenance, plant cyber security, nuclear safety and risk analysis.
- *Data-Theoretic Approach for Socio-Technical Risk Analysis: Text Mining Licensee Event Reports of U.S. Nuclear Power Plants* [8]
  This paper presents ML techniques for organisational factor in safety/risk analysis and PSA. For each application, the data source, data type, data format, type and sub-type of process and type of technique is presented. One finding was that there are limited studies using text mining approaches for PSA.
- *Probabilistic Safety Assessment and Management (PSAM) 2023 Topical conference on Artificial Intelligence & Risk Analysis for Probabilistic Safety/Security Assessment &*

Management. [3]
Conference proceedings can be found via the reference. A special issue is also planned to be finalized in October 2024.

- *Survey on the Use of Artificial Intelligence in Nuclear Power Plants* [9].
  This PSAM conference paper classifies a survey of applications into diagnosis, prediction, response (i.e., severe accidents), and process (i.e., optimization of design and operation), with the purpose and the used type of learning algorithm for each application.
- *Deep learning for safety assessment of nuclear power reactors: Reliability, explainability, and research opportunities* [10]
  This paper presents state-of-the-art in DL-applications for nuclear safety analysis. Each reviewed application is classified into its field of application, DL method, training data (e.g., CFD, RELAP, MAAP simulation), and predicted parameter (e.g., reactor vessel water level).
- *Possibilities of reinforcement learning for nuclear power plants: Evidence on current applications and beyond* [11]
  This paper focuses on RL applications including different situations such as power startup, collaborative control, and emergency handling. It also discusses possibilities of further application of RL methods and challenges. The authors note that this field is still comparatively blank, and many works can be explored.
- *Application of artificial intelligence technologies and big data computing for nuclear power plants control: a review* [12]
  This paper comprehensively reviews the literature on artificial intelligence technologies and big data, seeking to provide a holistic perspective on their relations and how they can be integrated with nuclear power plants. Further, this review also points out the future opportunities as well as challenges for applying AI and big data computing in the nuclear industry.

In summary, this literature review covers diverse applications and methods of AI in the nuclear industry, emphasizing the potential of ML and DL methods in enhancing reactor design, operations, maintenance, safety, and risk analysis. The large volume of applications distributed over many different fields shows the potential of using such methods. One finding from this review is that there is perhaps a need for a more harmonized classification/categorization of applications. Also, only one of the references above, [7], addresses some guidance of which algorithm or method to use depending on the application.

The applications in the above references cover a total of 335 references. The following subsections presents statistics regarding application area, type of learning and algorithms used. However, it is important to acknowledge that this survey of applications is not complete and may not reflect the full extent of ML usage in the nuclear industry. Due to the rapid development of ML methods in recent years and the diversity of possible applications, as demonstrated in Figure 3, it is challenging to achieve a comprehensive overview of all relevant studies. Moreover, not all applications are open access and thereby available. Therefore, this review should be seen as a snapshot of the current state rather than a definitive map of the ML landscape in the nuclear industry.

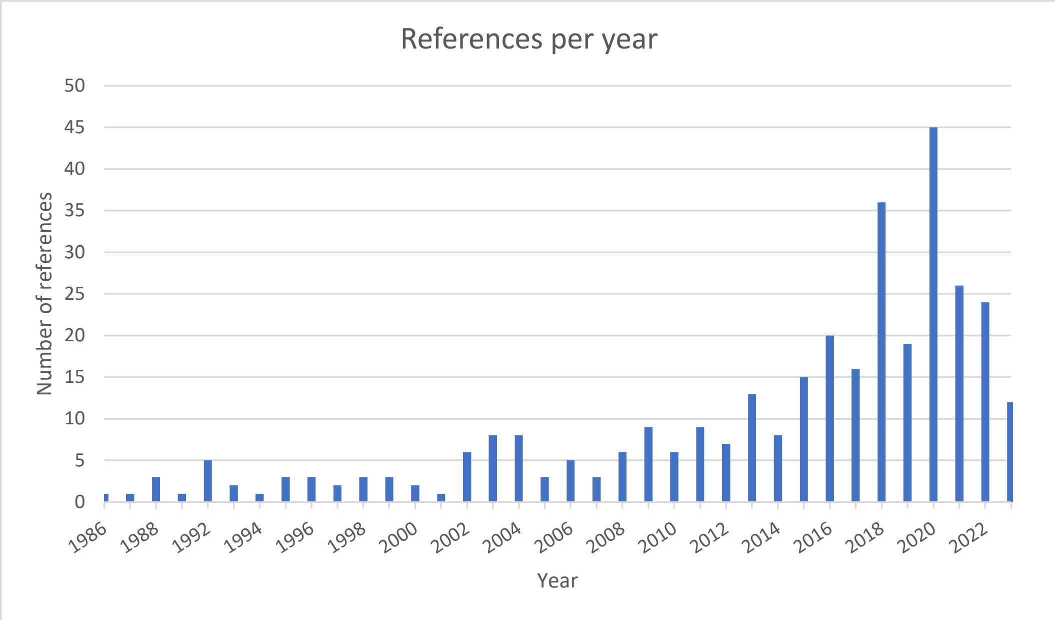

Figure 3 References per year.

The number of available algorithms is large, and it is no easy task to select which algorithm to use in an application case. The algorithm selection for a successful nuclear application depends on two major factors [6]:

1. The nature and objectives of the task, e.g., classification or clustering analysis.
2. Data availability and quality.

In practice, the two above factors can be used to narrow down the searching range, but generic principles can be applied. The best strategy is to evaluate and compare different algorithms. International research and applications with use cases along with extensive experiments with specific physical phenomenon considerations is recommended. The final algorithm(s) should be determined by using values of quantitative metrics, e.g., accuracy, precision, and recall rate, on new datasets.

For assessment of data availability and quality, see further sections 3.1.4 and 3.1.5.

3.1.1 Application Areas

From the literature review, a list of references has been extracted and compiled. Categorizing these applications not only highlights the current areas of focus but also exposes opportunities for further development and integration. Although some overlap between categories exists (as shown in Table 1), this highlights the adaptability and cross-functional capabilities of ML in various nuclear fields.

As seen in Table 1, the most frequently cited application areas are “Physical Plant-centred” and “Reactor Physics”, highlighting the critical role of ML in optimizing plant operations and understanding reactor behaviour. However, the categorisation of “physical plant-centred” is very broad and include an array of applications which overlaps with the more well-defined and detailed application areas. Conversely, the area “Digital twins” demonstrate low grade of adoption, and this area is further discussed in section 4.5.

Table 1 Application areas in the literature review.

Application area	Count
Accident scenarios	22
Diagnostics	41
Digital twins	3
Fuel management	11
Nuclear Safety and Risk Analysis	17
Nuclear Safety and Risk Analysis - Organisational Factors	5
Nuclear Safety and Risk Analysis - PSA	15
Radiation shielding design	4
Reactor Operation and Maintenance	47
Reactor Physics	49
Reactor System Performance	16
Reactor Thermal Hydraulics	23
Transient Identification	10
Physical Plant-centred	52
Human-centred	20
Total	**335**

Previous surveys show that applications of ML have been widely applied to enhance, for example, equipment reliability, reduce radiation exposure to personnel, assist with decision making and optimize maintenance schedule in three major areas: nuclear power plant health and management, nuclear operations and controls, and radiation protection. A list of example cases and algorithms are shown in Table 2, reproduced from [6].

Table 2 Example use cases and algorithms.

Application Area	Use Case	Algorithms
Plant health and management	System behaviour prediction	BNs, NB, ANNs, SVMs
	Severe accident classifications	ANNs, DTs, BNs
	Functional failure of systems	ANNs, Clustering algorithms, e.g., K-means
	Crack detection	CNNs
	Equipment monitoring	CNNs, ANNs, BNs
Nuclear operations and controls	Anomalous event detection	AEs, SVMs, ANNs, DTs
	Unattended operations	DTs, BNs
	Detection and response to degraded or failure conditions	CNNs, ANNs
	Radwaste management	CNNs
Radiation protection	Radionuclide identification	ANNs, SVMs
	Special nuclear material identification	ANNs, GPs, NB, Clustering algorithms

3.1.2 Application Sub-Areas

Table 3 presents further detailed analysis within the category "Nuclear Safety and Risk Analysis". Here, the applications (37 in total, row 5-7 in Table 1) span a wide spectrum of nuclear safety. It is seen that some sub-areas overlap with the other application areas in Table 1.

Table 3 Application sub-areas within “Nuclear Safety and Risk Analysis”.

Application sub-area	
Accident analysis	Other
Cost risk analysis	Performance shaping factors
Data mining	Risk insights in real-time
Dynamic PSA	Risk-informed decision making
Event reports	Scenario identification
Fire hazards	Seismic analysis
Human factors	Simulation data
Integrated DSA-PSA	Transient identification
LOCA detection	

These sub-areas can be grouped and summarized in three main groups, see below.

Risk Analysis and Assessment
A field focused on comprehending and measuring risks and utilizing this knowledge to inform decision-making.

- Cost risk analysis: Estimation of financial risks associated with nuclear safety measures.
- Dynamic PSA: Analysis to evaluate risk in systems where both time-dependent behaviours and interactions between components evolve dynamically.
- Risk insights in real-time: Monitoring risks and gaining insights into them as they happen.
- Risk-informed decision making: Using risk assessments to guide safety-related decisions.
- Integrated DSA-PSA (Deterministic Safety Analysis and Probabilistic Safety Assessment): Combining deterministic and probabilistic models for a more comprehensive safety assessment.

Event and Scenario Analysis
Area that focuses on understanding specific accident or event scenarios.

- Accident analysis: Examination of past accidents or potential future accident scenarios.
- Event reports: Analysing events that have occurred in NPPs.
- Transient identification: Detection and analysis of abnormal or unexpected transient events in NPPs.
- Fire hazards: Analysis and prevention of fire-related risks.
- Seismic analysis: Assessing the impact of seismic events.
- LOCA detection (Loss of Coolant Accident): Detection of potential or actual loss of coolant.
- Generic analyses:
    - Scenario identification: Identification of various potential scenarios that could affect nuclear safety.
    - Data mining: Extracting insights from large datasets to identify trends or patterns.
    - Simulation data: Use of simulation models to analyse scenarios.

Human Factors and Organisational Safety
Area that emphasizes the role of human factors and organisational decision-making in maintaining nuclear safety.

- Human factors: Study of how human behaviour and limitations affect nuclear safety.

- Performance shaping factors: Factors that influence human performance in the context of nuclear safety.

3.1.3 Type of Learning Process and Algorithm/Method

The literature review showed that for the applications, where the type of learning or algorithm were specified (not given in about 16% of the applications), about 53% of the application uses supervised learning.

The most common ML algorithms are listed below, and all of these uses supervised learning. It is also noteworthy that all algorithms, except SVMs, GAs (but can be combined with DL algorithms) and ANNs (depending on their structure), are DL algorithms. Also, GAs are not strictly a ML algorithm but are often used in ML to optimize parameters and neural networks. In general, traditional ML algorithms like SVMs often perform well on structured, smaller datasets, whereas DL methods, particularly DNNs and their variants, are suitable for tasks involving large datasets and high-dimensional data.

- Support Vector Machine (SVM)
  - SVMs are traditional ML algorithms used for both classification and regression tasks. SVMs aim to find the optimal hyperplane that separates data points of different classes by maximizing the margin between them. SVMs are not DL algorithms but are highly effective in smaller datasets or cases where interpretability is important. SVMs are commonly used in tasks like text classification.
- Artificial Neural Network (ANN)
  - In shallow form, ANNs are considered part of traditional ML. However, when ANNs have multiple hidden layers, they become DNNs (see below).
- Deep Neural Network (DNN)
  - DNNs are an extension of ANNs, distinguished by having multiple hidden layers between the input and output layers. These deep architectures allow DNNs to model complex patterns in data.
- Recurrent Neural Network (RNN)
  - RNNs are specialized types of neural networks that are particularly effective in processing sequential data, such as time-series data. Unlike traditional neural networks, RNNs have memory, allowing them to capture temporal dependencies in data.
- Fuzzy Neural Network (FNN)
  - FNNs combine fuzzy logic with neural networks to handle uncertainty and imprecision in data. They are particularly useful in systems where decision-making requires handling vague, imprecise, or noisy input, making them effective in applications for control systems, pattern recognition, and decision support. The integration of fuzzy logic allows FNNs to model complex relationships in data that traditional neural networks might struggle with.
- Genetic Algorithm (GA)
  - Gas are optimization algorithms inspired by the process of natural selection. Unlike most of the other algorithms in this list, GAs do not use a neural network but rather evolve a population of potential solutions to optimize a problem.

3.1.4 Data Selection and Considerations

Data are central to any data-driven AI system's ability to learn. Data can come in structured form (e.g., databases) or unstructured form (e.g., event reports, specification documents, non-destructive testing images and files).

Data processing

Any type of ML model needs data and the data itself need to undergo data processing to be usable. In general, the data go through the following process [13]:

- Data acquisition: Data can be collected through one or multiple sources. The nuclear industry has collected data over decades, but much of this data is stored across separate tools and databases. Thus, resulting in isolated datasets even within a single nuclear power plant. This process also involves assessing the suitability of the data to be included in the model. For example, the data could be biased or may not be broad enough to be representative for its application. Data bias is further addressed below.
- Normalization: Normalization involves adjusting data to a common scale, making different datasets mathematically comparable. Normalization enables the integration of datasets from various plants, ensuring that the combined data (e.g., data from multiple plants) can be effectively used in ML models.
- Data quality checking: Before deploying data in any ML model, it is vital to perform quality checks. The data must be examined for completeness, potential biases, and other factors that could impact its usefulness. Identifying and rectifying data issues, whether by correcting errors or removing faulty data, ensures that the input to the ML model is reliable and robust.
- Data labelling: For applications that need labelled data, i.e., supervised learning ML algorithms, proper data labelling is critical. The nuclear industry presents a wide spectrum when it comes to data labels. For instance, some datasets may be entirely unlabelled, such as process sensor data that lacks information on equipment condition, while other datasets may be fully labelled, like equipment failure that have been classified and reported into a database. Ensuring consistent and accurate labelling across datasets is essential for training ML models.

Data classification

The quality and size of the dataset used in training and validating the ML model is crucial for overall performance. During the development of a ML model, the data is commonly divided into:

1. Training datasets used directly for training the ML model,
2. Test datasets used to monitor and evaluate the training, and
3. Qualification datasets used to evaluate the final performance of the model.

The selection and independence of these datasets have significant effect on the ML model's performance. Data are classified into categories according to their role in the model development. The primary categories include training data, validated data, test data, and production data.

- Training data: Training data constitute the datasets from which the ML algorithm learns by adjusting its model parameters to address the specified task. These datasets may be in the form of time series, images, or texts. Operating experience data typically fulfil this requirement.
- Validated data: Validated data refer to the subset of data that is rigorously examined and confirmed to be reliable for informing algorithmic choices. These data, usually extracted from the same source as the training data. The validation process ensures that the data used in the model development phase accurately reflect the underlying processes, thereby contributing to the robustness of the final model.

- Test data: Test data are primarily used to evaluate the performance of supervised ML models. Since unsupervised models lack labelled data, alternative evaluation techniques are necessary. These may include benchmarking the outputs of different models or incorporating expert human judgments to assess performance. The test data serve as an independent metric for assessing the model's accuracy and generalizability across different scenarios.
- Production data: Production data are those processed by the AI system during its operational phase. This data, typically received in real-time or near-real-time, is critical for supporting ongoing decision-making processes and system monitoring, ensuring that the AI system continues to perform reliably under operational conditions.

Sample size determination (SSD)
The ML model performance is mainly reliant on the size and quality of the data sets used in the ML model training. It is of importance that the available data (historical, simulated, or experimental) is ensured to be adequate for training, validation, and testing of ML models. The required sample size will depend on which requirements the target ML model performance metrics need to fulfil.

Unbalanced data
Significant class imbalance can affect the training process and performance of ML models. Such imbalance may result from biased sampling methodologies, where data are not uniformly collected across all classes, or from labelling errors that incorrectly assign instances from one class to another. This affects the representativeness of the training data, but it also limits the model's ability to generalize effectively to new or rare events.

Data bias
Data bias is a systematic error wherein certain classes are overrepresented relative to others, leading to a training dataset that does not accurately reflect the true underlying population. Consequently, models trained on biased datasets may inherit these imbalances, resulting in skewed or inaccurate predictions.

Beyond the effects of unbalanced data, data bias may also arise from sample bias and measurement bias. Sample bias is where training samples are drawn from a narrow or unrepresentative subset of the population. Measurement bias, which occurs when the conditions under which training data are collected differ significantly from those encountered during actual operation. Additional factors contributing to measurement bias include varying noise levels between training and testing datasets, as well as incorrect labelling in supervised learning contexts, which can significantly affect model performance.

3.1.5 Data Availability in Applications

Data availability can vary widely across different application fields and understanding these differences is necessary for developing a ML model.

From the literature review, about two thirds of the listed applications focus mainly on the type of algorithm used and do not explicitly state the type of data used in the application without further analysis of each reference. Since not all references are open access and thereby available, it is difficult to draw conclusions regarding the data availability. In [6], the data source or the type of training data used is explicitly stated per application and has here been summarized in Figure 4. Here, 67% of the applications (76 references in total) used simulated data.

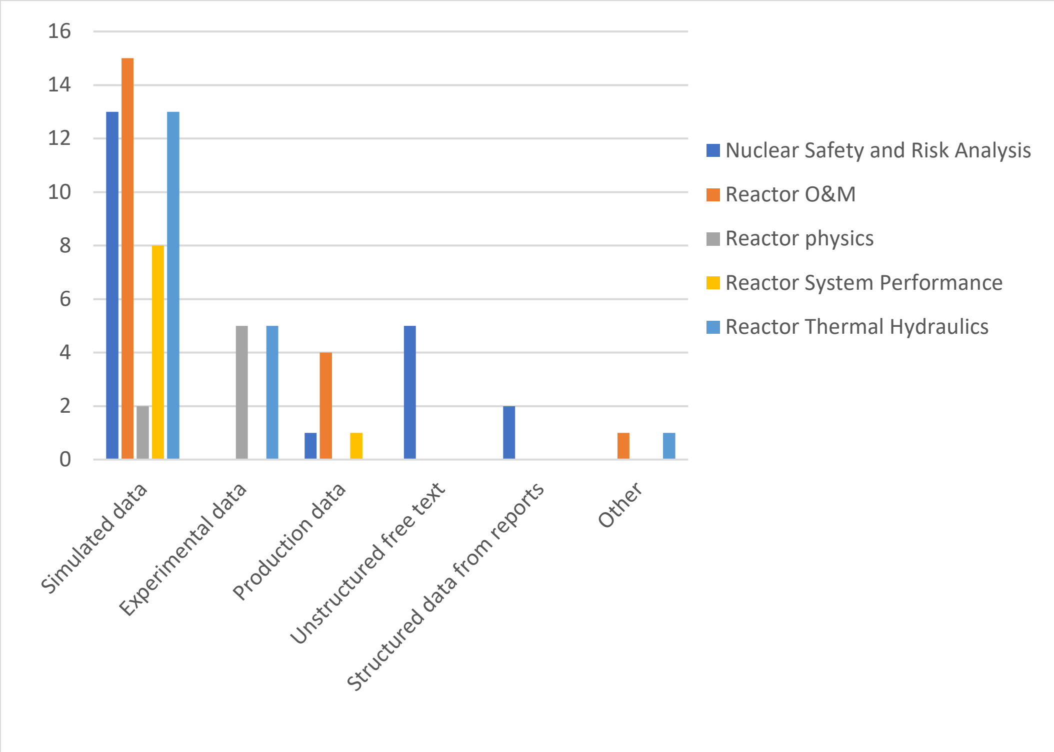

Figure 4 Data classification of applications.

<u>Plant operation (monitoring and process control)</u>
In operational monitoring and process control, data is typically generated continuously by an extensive network of sensors installed throughout a nuclear facility. These sensors produce high-frequency, high-resolution data streams that capture a wide array of operational parameters. Although the amount of data is considerable, most of it reflects normal operating conditions. This means that while there is much data to analyse system performance over time, the occurrence of anomalies or deviations is relatively rare. Consequently, it becomes essential to implement sophisticated anomaly detection techniques that can reliably identify these rare but important events.

<u>Predictive maintenance</u>
For predictive maintenance, the situation is somewhat similar but with its own unique challenges. Equipment monitoring systems accumulate extensive datasets during regular operations. This data volume provides a strong foundation for modelling normal equipment behaviour. However, the actual failure events, which are crucial for training models to predict equipment failures or maintenance needs, occur infrequently. This imbalance between normal and failure data can lead to challenges in model training, requiring the use of techniques such as oversampling or undersampling to accurately predict potential failures. The reliability of these predictive models is vital, as they may impact maintenance planning and operational safety.

<u>Safety and risk assessment</u>
Safety and risk assessment represents another application area where data availability is a significant concern. Safety-related events, such as near-misses or minor incidents, are very rare. This scarcity of real-world event data makes it difficult to build robust models solely based on operating experience, even if both plant-specific and generic (national and international) data are considered. To address this, simulated scenarios often play a key role, providing additional data that mimics potentially hazardous conditions or events. However, the integration of simulated data comes with the challenge of ensuring that these simulated scenarios accurately reflect the complexities of real-world events. The success of safety and risk assessment models, therefore, depends not only on the availability of data but also on the careful validation and calibration of simulation outputs against actual operational data.

Simulation and modelling
Simulation and modelling applications are largely defined by the generation of simulated data. Physics-based simulation codes allow researchers to create detailed models of nuclear processes which can be used for exploring a wide range of scenarios. A key challenge with simulation data is its variability and the potential for discrepancies when compared to real-world observations. It is imperative to continuously validate and adjust simulation models so that they align closely with empirical data, thereby ensuring that the insights derived from these simulations are both accurate and applicable to real operational conditions.

3.2 ML in Nuclear Safety and Risk Analysis

Recent research has developed or applied ML methods for nuclear safety and risk analysis, mainly for PSA of nuclear power plants. In contrast to ML applications in reactor design and system analysis, and plant operations and maintenance, ML applications in nuclear safety and risk analysis are performed on structured data and unstructured free-text data.

ML techniques such as NLP, supervised ML and unsupervised ML are applied to identify free text data and extract implicit information. NLP can be used to extract, from free-text reports, causal relationships between factors that lead to failures. Context analytics and text analytics can support decision-making and PSA activities.

ML applications in nuclear safety and risk analysis are not only directly applied for model development or uncertainty quantification but often embedded in complicated frameworks for different purposes. Particularly, the “black box” nature of ML brings challenges with respect to the trustworthiness and transparency of the results of ML applications and it becomes difficult to meet regulatory requirements.

A review of existing studies was conducted in [8], focusing on applied ML approaches for the PSA of nuclear power plants. Their findings concluded the following:

- There are a limited number of studies using ML to quantify PSA model elements. The application of ML approaches for PSA primarily analysed physical phenomena, where ML was used to cluster the simulation outcomes. In these studies, the data are not historical events but are the results of simulation codes. Consequently, the main challenge is dealing with large volume of data rather than processing heterogeneous data (e.g., varying formats and structures of data).
- Among the PSA-oriented ML studies, most of these efforts used historical event data rather than results of simulation codes. Only a few studies used text mining approaches for PSA. Additional research is needed to compare the performance evaluation of ML techniques for unstructured event data to justify the best selection for PSA.

ML approaches have the potential to improve plant safety and efficiency. Potential application fields within nuclear safety and risk analysis include:

- System, structure, component reliability
- Human reliability analysis
- External event analysis
- Accidental radiological release and monitoring

Each field is further discussed on the following subsections.

ML applications in these fields can provide benefits to both NPP operators and the regulator for plant safety and efficiency. By introducing ML techniques into these fields, potential benefits for plant safety and efficiency include but are not limited to:

- Achievement of a better level of safety by:
    - Removing/reducing failure sources.
    - Developing better failure-preventing strategies.
    - Developing better accident-mitigation strategies.
- Enhancement of safety evaluation techniques by:
    - Expanding safety evaluation scope.
    - Improving safety evaluation accuracy.
- Reduction of human and computational labour cost.

3.2.1 System, Structure, and Component Reliability

Traditionally, System, Structure and Component (SSC) reliability is performed using PSA tools or reliability modelling methods with conventional statistical methodologies, which may have the limitations of inapplicability in some extrapolated conditions and be expensive computationally.

Some efforts prove that ML techniques can be introduced in the analysis, evaluation, and enhancement of SSC reliability by providing an efficient and accurate prediction of SSC failure probability or reliability. By developing surrogate models that may have a better scalability and predictive capability when ML training data is sufficient, ML techniques have the potential to improve SSC reliability analysis and evaluation in plant safety assessments.

Various ML models have been applied to safety-related issues in nuclear power plants, including ANNs, kernel regression models, NLP, unsupervised ML methods such as classification and clustering, and others, [6]. These models can be used for one of the following purposes:

- Data-driven models: Developing a model that learns patterns and relationships from the data without relying on prior physical understanding or models.
- Physics-guided surrogate models: Supporting existing physical models or tools by building a ML model that captures the underlying patterns and relationships in the data.

[6] identifies that the primary technical challenge confronting the integration of ML into SSC reliability analysis concerns the inconsistency between ML training data and the data observed in real-world, full-scale prototypic conditions. ML training data predominantly comprises of numerical simulation data, supplemented by a limited set of available experimental data and/or operational data. The challenge arises from potential scale distortion (i.e., not accurately reflect the range and variability) in the simulated conditions used for generating training data compared to the actual conditions of full-scale systems. Nevertheless, given the extensive past research in this area, the knowledge and empirical correlations derived from past efforts can serve as valuable guides in the development and evaluation of ML models.

3.2.2 Human Reliability

Human Reliability Analysis (HRA) is a systematic technique used to identify, analyse and quantify the human contribution to risk in complex socio-technical systems. Johnson et al. observe that the HRA process “usually involves identifying potential human errors, analysing the causes of those errors, factors that influence these, and determining the likelihood and consequences of those errors”, [14] (pg. 1). In recent years, several studies have been performed to evaluate how AI and ML could be utilised for HRA. Among the potential applications of AI/ML, the most common appear to be (i) data mining, (ii) error classification, (iii) human error prediction and (iv) human performance modelling.

The first HRA method for nuclear power applications was published in 1983, called Technique for Human Error Rate Prediction (THERP, [15]), and several more methods have been developed since then. Despite the maturity of HRA as an analysis technique, one of the most significant challenges for analysts is the lack of empirical data about human error to inform and underpin HRA methods ([16], [17], [18], [19]). It is unsurprising therefore that one of the predicted uses of ML for human factors data analysis is data mining, [20].

[21] noted that diverse HRA data "are helpful for understanding the nature of human errors under a given task context" (pg. 1), and to reduce uncertainty in the calculation of Human Error Probabilities (HEPs). Sources of HRA data typically include "operating experiences (e.g., event investigation reports), observations from full-scope simulators, experiment results using partial-scope simulators, expert judgments and interviews with subject matter experts", [21] (pg. 1-2). [22] observed that "although simulation technology can generate a large amount of data now, accident reports that record real accident scenarios are still essential sources [of human error data]" (pg. 745). Further, [14] noted that "Data from major accident reports have the potential to better capture the interaction between human, machine, and organisational systems, providing additional contexts and scenarios not fully achieved by simulators and expert elicitation", [14] (pg. 2).

Despite this, manual extraction of core information related to human and organisational factors (HOF) is a challenge because it is time consuming and expensive, and because accident report texts from various sources may contain significantly different structures and levels of detail. Further, "manual coding of accident records could bring uncertainties and inefficiencies, especially when many records are available", [22] (pg. 745).

One technique that can be used to mine relevant data from the text of accident reports is NLP. As noted in Section 2.4, NLP can be used to extract causal relationships between factors to understand how HOFs can influence an accident. Reference [20] describes an example approach using "a data mining framework combining with correlation analysis, cluster analysis, and association rule mining for identifying intrinsic correlations among human factors", [20] (pg. 164), shown in Figure 5.

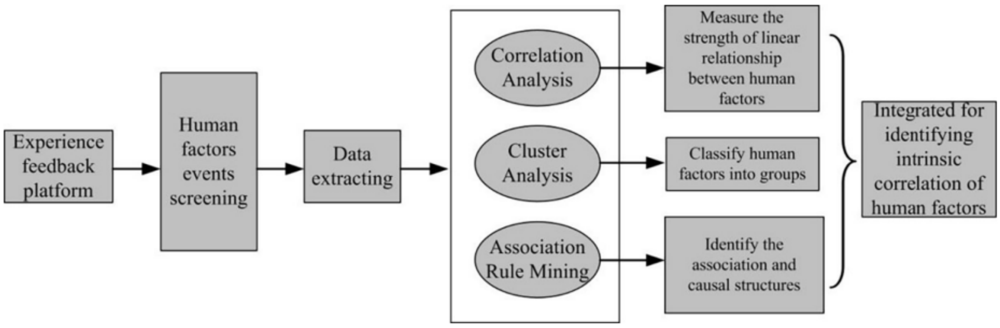

Figure 5 The data mining framework [20].

In this approach, an initial screening analysis of operating event reports is performed to identify those events where human factors was either a root cause and/or a causal factor in the event. A correlation analysis is then performed to identify the strength of the relationship between factors. Cluster analysis is performed to analyse the similarity between root causes and causal factors. Finally, association rule mining is performed to identify associations, frequent patterns, or causal structures from the data sets. Association rule mining can be used to analyse and predict behaviour "to study the associations and causal structure of the influencing factors, which enable us to predict the occurrence of a specific influencing factor based on the occurrences of the other influencing factors", [20] (pg. 167).

The data mining framework technique utilised in reference [20] enabled extraction of the most frequently identified HOF-related root causes and causal factors from the analysed events. Furthermore, this technique allowed for identification of strongly correlated contributing factors, which gives greater insight into both why the events occurred as well as where improvement efforts could be focused for greatest effect.

NLP is a promising approach, especially where it can reduce the workload of the analyst by significantly reducing the amount of time needed to process and extract data from a source report, and potentially identifying latent patterns or relationships that are harder for a person to detect manually. However, there are some limitations. [23] noted that “Factors that make the application of machine learning challenging in this domain are the lack of quality and uniformity of the data, limited depth of information present in the documents, the complexity of the taxonomies applied, lack of uniformity and correctness in the processing and labelling of existing documents, or the presence of low-quality annotated data that may corrupt possible training data”.

Another potential application of ML is as a “virtual human factors classifier”, [24], whereby a text recognition and classification algorithm is used to automatically classify accident reports using a pre-defined human factors taxonomy. The goal is to enlarge a human reliability dataset significantly faster, in about one third of the time it would take to perform this task manually. This method uses data extraction techniques such as “bag-of-words” (BoW), “Term Frequency – Inverse Document Frequency” (TF-IDF) and “word2vec”. BoW extracts strings of words from texts and calculates their frequency of occurrence, to form a vocabulary that allows prediction of which words are likely to occur together. TF-IDF is similar in that it also calculates the frequency of occurrence of words, but (unlike BoW) it assigns higher scores to domain specific words. Word2vec is a kind of neural network that “assumes that words that occur in the same contexts tend to have similar meanings”, [24] (pg. 4).

Once data are extracted from reports, ML techniques are used to classify text features. These methods may be supervised or unsupervised, and the method chosen may be “based on how texts are going to be classified, and if some documents have been previously classified by humans (allowing their use as examples to train the machine)”, [24] (pg. 5). Examples given include “Naïve Bayes” and SVMs, both of which are popular supervised methods. In [24], it is observed that “Naïve Bayes classifiers perform better with missing data, and therefore it might be a good choice to identify human factors interactions in major accidents that are considered rare and uncertain events” whereas “SVM has the potentiality to better capture features interactions and better classify larger documents”, [24] (pg. 8). [24] claims that the virtual human factors classifier has the potential to significantly accelerate the data collection process for analysts and that it represents “an efficient way of expanding existing human reliability databases based on accident reports”, [24] (pg. 23).

Another area of research on how ML can benefit human reliability analysis is the area of human error prediction, although research on this topic is relatively scarce at this time. [25] observed that understanding what is meant by the term “prediction” is in itself a discussion point, stating that “prediction is understood as foreseeing the possibilities for errors” and studies exploring this “approach the problem of real-time prediction of human error in the sense of detecting hidden precursors of forthcoming error in the time series representing temporal evolution of system and/or operator state”, [25] (pg. 173-174).

[25] presents a simple example where ML is used to analyse data on a human balancing a virtual overdamped stick, to predict human control errors that would result in the stick falling. ML is used to train a classifier to distinguish between periods of “normal” and “faulty” stick balancing, but also to identify stick fluctuations and operator actions in “pre-fall” segments from a historical dataset. The goal is to detect the precursors (changes in the operator behaviour or system state) that indicate that the stick might be about to fall, in the seconds before it falls. The study achieved a predictive accuracy of 73-47% and noted that “more advanced data analysis methods (e.g., Bayesian techniques) can also improve accuracy of control error prediction”, [25] (pg. 177).

Research around using ML for human performance modelling appears to be more abundant. This can be used to predict human behaviour and human error probabilities, but also to support the population of databases through generation of synthetic human behaviour data. [26] used a deep-learning model to predict future task performance based on workload estimates and contributing cognitive, physical, visual, auditory and speech components. The authors noted that “a relationship exists between workload and overall task performance; thus, workload information may be used to predict future performance”, [26] (pg. 3). [26] used LSTM neural networks, which predict future time-steps of a sequential data series based on previous time-step information. Experimental studies were performed whereby participants simulated supervision of a remotely piloted aircraft in the NASA MATB-II task environment, and objective and subjective workload measures were collected throughout. The data were used to train a generalised model for the group of participants and then individualised models, with the researchers concluding that the individualised model predictions were more accurate than the generalised model. Although not specifically developed for use in HRA, such research can be extremely valuable for predictive analysis of human reliability in high workload scenarios.

3.2.3 External Events

External events include both natural external events (e.g., earthquakes, high winds, and external flooding) and human-made external events (e.g., airplane crashes, explosions at nearby industrial facilities, and impacts from nearby transportation activities).

These external events normally have wide-area effects that may cause malfunctions of several SSCs at a plant or at several facilities at a site. Specific strategies are essential for preventing and mitigating failures and accidents resulting from these external events, and these strategies vary depending on the design of the NPP and its site.

PSA tools have been widely applied to external event analyses and can provide sufficient information and knowledge for constructing control and management strategies.

External events PSA can also be separated from internal events PSA because it has unique and specialized analysis methods for various kinds of external events. Depending on country-specific modelling practices, external events as well as internal fires may or may not be integrated with the internal events model. In case of using a separate external event PSA model, the construction of it may require extensive computational resources for quantification.

Researchers have suggested the introduction of ML to provide for a more efficient external event analysis in PSA. For example, ML-based methodology for seismic fragility curves estimation using SVMs, and fragility curves based on seismic damage data and numerical simulations by ANNs. [6]

Existing efforts of applying ML in external event analyses have applied various ML and advanced statistical methods, including k-nearest neighbour modelling, mean-iterative neural networks, simple ANNs, DNNs, SVMs, and others for scenario analyses and classification, clustering, and regression trees for identification of external events. [6]

[6] note that the primary technical challenge confronting the integration of ML into external events include the lack of data or knowledge for some rare external events, particularly some combinations of external events.

3.2.4 Accidental Radiological Release and Monitoring

The assessment of accidental radiological releases is of significant importance for nuclear safety, guiding decisions in accident control and management. Once radioactive materials are released into the atmosphere, the source term information becomes inherently unknown and uncontrollable. Therefore, monitoring the dispersion of accidental radiological releases becomes necessary.

ML approaches have been applied to better estimate the release rate, amount, and area of source terms or radioactive materials from NPPs operations and accidents. Various ML methods and data-driven frameworks have found application in this area, encompassing conventional ANNs, GPs, RFs, GAs, and sophisticated DL approaches like CNNs and RNNs. The selection of ML methods is dependent upon the complexity of the database and the latent physics involved. [6]

The Rapid Source Term Prediction (RASTEP) tool, developed by Vysus Group in cooperation with the Swedish Radiation Safety Authority (SSM), is an emergency preparedness tool that uses a probabilistic approach to provide decision support in nuclear emergency situations, [27]. RASTEP utilizes Bayesian Belief Networks (BBNs) to predict potential source terms based on real-time data. By incorporating PSA level 1 and 2 analyses, the tool effectively models various release categories and their associated source terms, allowing for rapid and reliable predictions during a crisis. This probabilistic approach accounts for uncertainty in the data, providing decision-makers with a clearer understanding of potential radiological releases and their impacts. The integration of RASTEP into crisis organizations' operational frameworks enhances situational awareness, enabling informed actions to protect public health and safety during severe accidents.

[6] note that the primary challenge in implementing ML within this field lies in the scarcity of data available for validating ML models and frameworks. However, leveraging high-fidelity simulation data generated through intensive computational processes (e.g., data obtained from simulations that closely emulate the real-world conditions or phenomena) enables the quantification and reduction of uncertainties in ML predictions.

3.2.5 NLP of Operating Experience

In PSA-models, operating experience is used to estimate equipment reliability, unavailability, initiating event (IE) frequencies. For these, classical statistical approaches are used to estimate the failure probabilities/rates and frequencies. However, the estimations are based on time-consuming and resource-heavy manual analysis of the historical data, i.e. the operating experience in terms of failure event reports. With use of NLP, this process could be automated which would save time and resources and as well identify new relationships between failure event reports. These new insights could support decision-making and assist in the identification of anomalous behaviours that might occur in a system, the possible cause-effect relations between events and their potential consequences, [28].

With ML it would be possible to recognize and process both structured data and unstructured data. Structured data would here be data classification using existing or new data fields from existing databases where operating experience are collected. Unstructured data would here be analysed with NLP methods for causal learning, e.g., extracting causal relationships from failure reports. For example, “the cause of the pump failure was due to broken pump shaft”. In this example, the model would identify and couple the words “cause” and “due to”, and thereby identify the relationship between the pump failure and why it failed. With NLP methods it would be possible to extract failure data for parameter estimations.

NLP methods could also be used for other applications to support and optimize NPP operation and maintenance including advanced diagnostic (such as detecting failure causes), prognostic models (such as predicting the remaining useful life) and human reliability analysis (as discussed in section 3.2.2). Examples of applications are:

- [29] describes that the NLP method used free text data (event descriptions) to automatically identify and characterize a Low Power Shutdown (LPSD) initiating event. In the application, a dictionary-informed DL approach was implemented and evaluated. The main challenges identified were that the event descriptions and the terms used in the records are not standardized, and that the training/testing datasets are small.
- [30] developed a similar method for automatic fault detection to build a reliability database based on existing operating experience.

3.3 ML in Reactor System Design and Analysis

ML methods have found extensive application in the design and analysis of reactor systems, encompassing areas such as reactor thermal hydraulics, reactor physics, and reactor system performance.

The predominant data types involved consist of structured experimental data and numerical data produced by simulation codes or simulators. These datasets serve as valuable complements to plant operating experience, allowing the simulation of key parameters and variables in fundamental reactor systems.

ML applications are focused on addressing model and code uncertainties, as well as developing closure models. A closure model refers to a mathematical or computational model used to represent processes or phenomena that may not be fully resolved within the primary modelling framework. ML-based approaches have proven to be instrumental in facilitating the advancement and implementation of thermal-hydraulic and neutronic methods. These methods offer innovative opportunities for dimensionality reduction and the creation of reduced-order models within fluid mechanics or neutronics.

Applications of ML include nuclear reactor design optimization. Reinforcement learning has been applied for various optimization problems within the nuclear industry, especially attributed to the nuclear fuel design. Nuclear fuel design involves two common problems: (1) core optimization and (2) assembly optimization. Core optimization aims at finding the best loading pattern of all assemblies in the core such that the reactor operation is economic and meets safety constraints, whereas assembly optimization aims on finding the optimal material composition and location of all fuel rods in the assembly. Examples of ML applications are:

- Nuclear Power Plant Fuel Optimization, [31]: Reinforcement learning algorithms, including deep reinforcement learning, have been assessed for their potential in optimizing nuclear power plant fuel, presenting a novel approach to solve the loading pattern problem.
- Nuclear Assembly Design Optimization, [32]: Physics-informed reinforcement learning has been employed for the optimization of nuclear fuel assemblies, aiming to improve fuel efficiency, reduce costs, and ensure safety constraints.

While ML algorithms are adept at handling and analysing structured data, a significant technical challenge arises from their inherent "black box" nature. This characteristic introduces a new source of uncertainty, complicating the explanation and trustworthiness of ML techniques, especially in their application to nuclear industry. Additionally, the exceptional capability of ML techniques to capture features from training data may result in overfitting issues in predictions (i.e., a model learns the training data too well, including its noise and outliers, making it perform poorly on new, unseen data).

3.4 ML in Plant Operation and Maintenance

In recent decades, there has been a growing exploration of ML techniques to enhance the support and optimization of nuclear power plant operations and maintenance. The majority of ML applications in this context rely on simulated data due to the scarcity of available plant operating data. Nevertheless, the robust, accurate, and rapid computational capabilities by ML techniques prove highly instructive and valuable. They play a pivotal role in realizing autonomous plant control and management, leading to cost reduction and enhanced reactor resilience (e.g., predicting, identifying, and responding to potential issues).

It is worth noting that most of these ML-aided techniques are specifically developed for safety-significant or safety-related Instrumentation and Control (I&C) applications in NPPs. The licensing process for these applications is subject to stringent regulatory requirements. Therefore, future research should focus on identifying, analysing, and evaluating the trustworthiness, transparency, and robustness of ML-aided techniques.

Applications within operation and maintenance can be sub-divided into two different focus areas: (1) diagnosis and prognosis and (2) optimization of maintenance.

3.4.1 Degradation Modelling

Components in a NPP must endure high-temperature water, stress, vibration, and an intense neutron field. Material degradation within this environment has the potential to result in reduced plant performance or an unplanned shutdown, leading to a loss of power generation. Consequently, modelling degradation and implementing online monitoring are necessary to address issues related to component ageing. This approach allows for precise predictions of failure points or the remaining useful life (RUL) of components, enabling timely maintenance or replacement.

While several models exist for estimating material or component degradation, these models typically come with fixed forms or parameters, restricting their applicability under certain extrapolated conditions. Additionally, conventional methods rely on prior physics knowledge and expertise, often having limited adaptability to learn from extensive measured or simulated data. This highlights the potential for ML methods to create data-driven surrogate models, which can use such data for model improvement.

Depending on the complexity of involved physics, sufficiency of data, and internal dependency of degradation features, different ML methods have been introduced and demonstrated for degradation modelling, including GPs, simple ANNs, DNNs, RNNs, unsupervised learning, and support vector regression, [6].

3.4.2 Fault Diagnosis

Fault Diagnosis (FD) involves identifying abnormal behaviour in a component by analysing specific indicators. In complex systems such as nuclear power plants, pinpointing faults accurately requires collecting data through sensors, processing the data with algorithms, and extracting essential patterns for effective fault identification or classification.

Fault detection, diagnosis and prognosis (FDDP) is extensively implemented in existing NPPs to enhance and ensure the reliability and availability of SSCs. These approaches are physics-based and data-driven and rely heavily on prior physics knowledge and expertise, and the available measured or simulated data. Although these methods do not demand extensive data, they may struggle to precisely predict faults or NPP states in unfamiliar or abnormal conditions.

In contrast, data-driven approaches using ML methods can unravel intricate, nonlinear patterns from large datasets. These approaches exhibit broad applicability and self-improvement capabilities facilitated by adaptable ML models as new data becomes available. The integration of data-driven and physics-based approaches, often termed hybrid physics-guided data-driven approaches, shows promise in addressing knowledge and technical gaps by using the strengths of each approach.

A diverse array of ML methods has been demonstrated, spanning from supervised to unsupervised learning and ranging from techniques like GPs, SVMs, and ANNs to more intricate models such as Deep Neural Networks (DNNs), Convolutional Neural Networks (CNNs), and RNNs. [6]

The primary technical challenges in ML applications for FDDP persist in enhancing their explainability, interpretability, and trustworthiness, especially given the potential impact on the performance of highly safety-related and safety-significant Instrumentation and Control (I&C) systems.

Thus far, FD may be the most active application area of ML techniques in the nuclear industry and considerable progress has been achieved, [5].

3.4.3 Predictive Maintenance

Currently, the nuclear industry adopts preventive maintenance programs for SSCs. Often the preventive maintenance times are based on the average service life of similar equipment, disregarding the specific operating conditions of individual components. Consequently, a conservative stance often leads to excessive maintenance, which results in increased costs but also heightens system risks due to potential human errors during maintenance.

The safety and efficiency of NPP Operation and Maintenance (O&M), is enhanced by the development of models and algorithms for diagnostics and prognostics (i.e., reliability assessment at the current moment and in the future). These Prognostics and Health Management (PHM) techniques use real-time and historical operational data to provide decision support for improved performance, reliability, and maintainability. They are capable of handling high volumes of multi-dimensional data collected by sensors. The PHM techniques aim to shift from an "on-time maintenance strategy" to an "on-demand maintenance strategy" through RUL prediction for key equipment, see Figure 6.

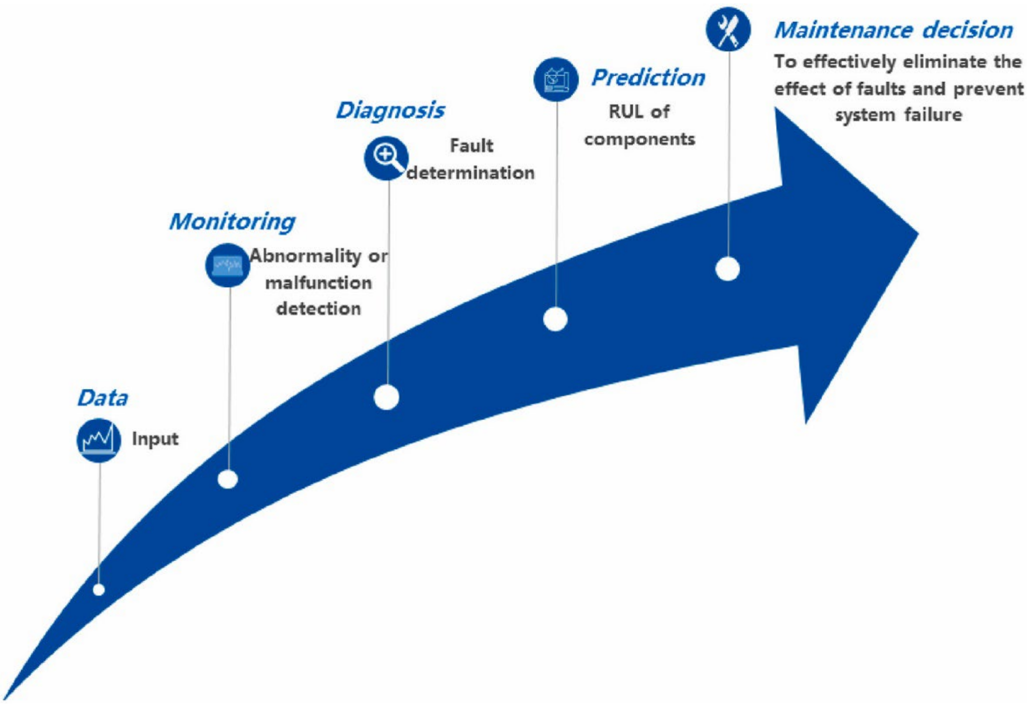

Figure 6 PHM framework [5].

Nuclear researchers are actively working on developing ML-based inference methods that overcome the limitations associated with classical approaches. Inference methods have the ability of reasoning or deducing information beyond what is explicitly stated in the data, whereas classical methods often have limitations in terms of complexity, adaptability, or the ability to handle large datasets. As of now, there are no implementations of these methods ready for practical applications, [33].

An example application for assessment of equipment reliability data is given in reference [34]. In this application, the automatic generation of knowledge is based on a textual element in order to assist system engineers in assessing an asset's historical health performance. The goal is to assist system engineers in the identification of anomalous behaviours, cause–effect relations between events, and their potential consequences, and to support decision-making such as the planning and scheduling of maintenance activities. The application describes how equipment reliability textual data elements are first pre-processed to handle typos, acronyms, and abbreviations, then ML and rule-based algorithms are employed to identify physical entities (e.g., systems, assets, and components) and specific phenomena (e.g., failure or degradation).

A systemic review of ML algorithms for periodic in-service inspection (ISI) using non-destructive examination (NDE) methods, [35], addresses their potential and gaps. The study concluded that the confidence in ML results for NDE relies on careful data selection, model tuning, and robust verification and validation. Using standardized, representative datasets not only improves performance confidence but also enables easier method comparisons. Missing data sets and workflow details limits this further (such as sample size determination, handling unbalanced data and bias).

Research presented in [36] introduce federated-transfer learning (FTL) to scale ML models for condition-based monitoring (CBM) across a component or plant system by combining federated learning (FL) and transfer learning (TL) approaches. FL enables a centralized server to develop an aggregated global CBM model, while the training data are safely and privately distributed on the devices of plant systems, and TL enables application of the developed aggregated model to different but related systems within the same plant site, or to the same system at different plant sites. The study in [36] demonstrated the significance of the FTL approach with use of a multi-kernel adaptive SVM and an ANN, which avoids building exclusive predictive models for each NPP and each system. While there is limited use of FTL in the nuclear industry, the secure, privacy-oriented approach it offers could be advantageous for CBM in NPPs. Given the sensitive nature of nuclear data, FTL's potential to enable collaborative learning without sharing raw data could be a future development in nuclear safety and optimization.

4 Challenges and the Future of AI and ML

4.1 Accelerate Innovation

In the US, the Nuclear Regulatory Commission (NRC) has developed an AI Strategic Plan, [37], covering fiscal years 2023–2027, with vision and goals to continue to improve its skills and capabilities to review and evaluate the application of AI to NRC-regulated activities, maintain awareness of technological innovations, and ensure the safe and secure use of AI.

The AI Strategic Plan, [37], includes five goals: (1) ensure NRC readiness for regulatory decision-making, (2) establish an organisational framework to review AI applications, (3) strengthen and expand AI partnerships, (4) cultivate an AI-proficient workforce, and (5) pursue use cases to build an AI foundation across the NRC.

The first goal aims to establish a robust and flexible AI regulatory framework that provides an objective, sound technical basis upon which regulatory decisions can be made and enforced. Thus, the overall strategy is to continue to keep pace with technological innovations to allow for the safe and secure use of AI in NRC-regulated activities, when appropriate, through existing or new regulatory guidance, rules, inspection procedures, or oversight activities.

The International Atomic Energy Agency (IAEA) has published a document, [33], on how they can assist and help accelerate the development and adoption of ML methodologies for the nuclear industry.

The availability and quality of data is a challenge for several application areas, sometimes because it contains sensitive data. Improving the availability of data sets would enable evaluation/benchmarking of ML techniques. Also, increased sharing and accessibility of data would accelerate the development and adoption of ML methodologies, [33]. Their potential beneficial actions within data and information management systems include:

- Developing and maintaining a library of benchmark datasets for comparing and evaluating performance of various ML algorithms.
- Guiding development and deployment of privacy-preserving methods for data anonymization.
- Creation of a repository facilitating the exchange and sharing of data for representative projects to promote the value of AI.

IAEA also aims to assist the innovation within the field by:

- Facilitating a network with subgroups for each of the main application areas.
- Arranging training workshops, e.g., to improve model transparency and making the ML systems understandable.
- Coordinate research projects. For example, the IAEA has designated the "Center for Science of Information" at Purdue University in USA as the first IAEA Collaborating Centre to support the Agency's activities on AI for nuclear power applications, including reactor design, plant operations, and training and education, [38].
- Review the impact of AI and ML on existing IAEA guidance and standards and provide commentary on their applicability, especially in the safety and security area.
- Develop principles and guidance for AI and ML systems, including best practices. In addition, IAEA topical reports on AI technology and or the practical use of AI in a nuclear power context would be beneficial.

Within the Organisation for Economic Co-operation and Development – Nuclear Energy Agency (OECD-NEA), there is a task force dedicated to "Artificial Intelligence and Machine Learning for Scientific Computing in Nuclear Engineering", [39]. This task force objectives are to provide the following:

- Standardised benchmark exercises with certified experimental data and high-fidelity computational data for the training of AI/ML models.
- Detailed guidelines for applying AI/ML methodologies for supervised, unsupervised, and semi-supervised ML, as well as advanced topics such as deep generative learning and probabilistic ML.
- Proposals towards the development of verification, validation and uncertainty quantification requirements of AI/ML models in nuclear systems based on consensus positions of the task force.
- Guidelines for improving AI/ML trustworthiness through accuracy, robustness. (reproducibility, applicability) and transparency (explainability, interpretability).
- Training opportunities to demonstrate AI/ML principles and practices.
- Demonstrations of the AI/ML guidelines for specific applications.

Overall, these initiatives aim to enhance safety and efficiency in nuclear operations through AI and ML integration. It is clearly seen that many efforts are taken place, and this is an interesting development. These initiatives by well-known organisations are perhaps the key to successfully accelerate the development and adoption of AI and ML methods within the nuclear field.

4.2 Black Box Dilemma and Explainable AI

A major concern with AI systems, particularly those using DL algorithms, is their "black box" nature. The process within a DL model between input and output involves hidden layers where features are encoded through multiple neuron clusters, making it obscure to humans. Figure 7 illustrates the paradox between AI explainability and performance, reproduced from [5]. Models with superior performance often have more complex internal mechanisms and are harder to explain, whereas models with simpler mechanisms, like linear and rule-based models, struggle with complex nonlinear problems. The ideal model would achieve both high performance and high explainability.

The nuclear industry has high requirements in terms of safety and is heavily regulated. Thus, it is crucial to enhance the transparency, robustness, and accountability of AI models. Focusing solely on accuracy is insufficient for the integration of AI. Consequently, researchers must address these broader criteria to advance the field.

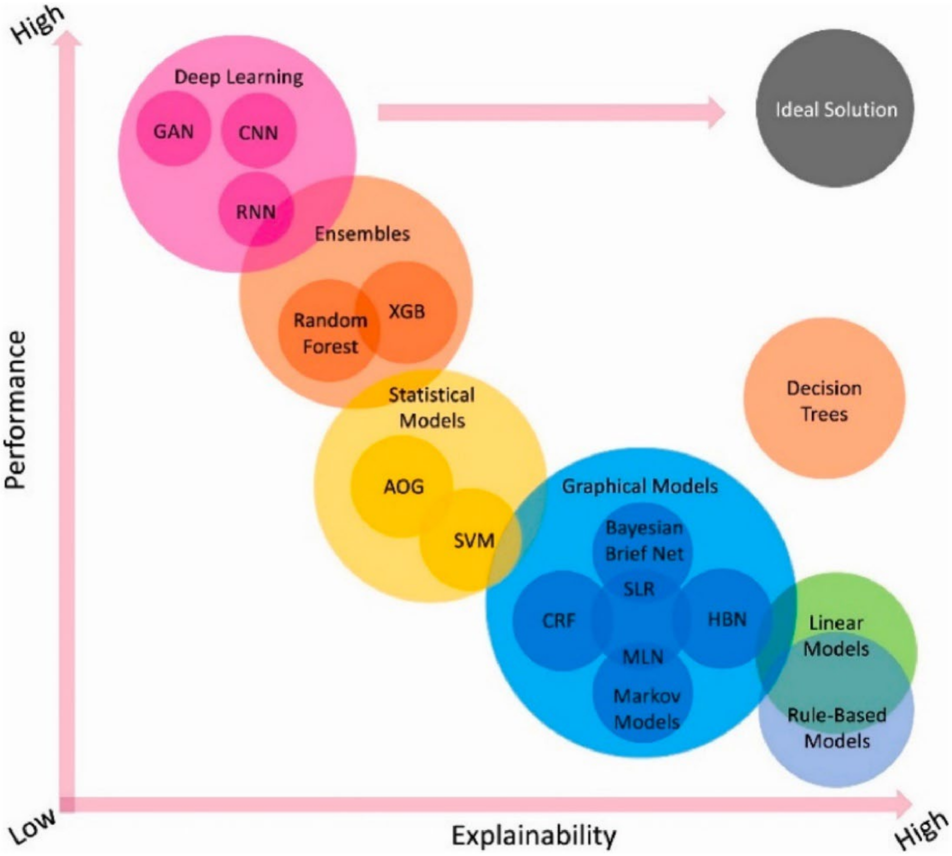

Figure 7 Explainability versus performance.[1]

The existing obstacles that prevent the further fusion of AI and nuclear technologies so that they can be scaled to real-world problems are classified into two categories [5]:

1. Data issues: Insufficient experimental data increases the possibility of data distribution drift and data imbalance.
2. Black-box dilemma: DL methods such as CNNs, RNNs have poor interpretability.

Potential strategies to address the challenge with the black box dilemma:

1. Promoting the use of eXplainable AI (XAI) technologies to enhance the transparency and reliability of AI models used in nuclear applications. XAI aims to make AI systems more interpretable and their decisions more understandable to human users.

2. Better integration of domain knowledge with data-driven approaches. Incorporating expert knowledge and physics-based models can help reduce the reliance on pure "black box" AI and improve the robustness and interpretability of the solutions.

Overcoming the "black box" dilemma is crucial for building trust and ensuring the safe and responsible deployment of AI in critical nuclear operations, where the ability to understand and explain the AI system's decisions is essential.

[1] HBN: Hierarchical Bayesian Networks; SLR: Simple Linear Regression; CRF: Conditional Random Fields; MLN: Markov Logic Network; AOG: Stochastic And-Or-Graphs; XGB: XGBoost; and GAN: Generative Adversarial Network.

The field of explainable AI include a broad spectrum of methods and to support practitioners for choosing an appropriate XAI method for their use-case, some main categories for evaluation of method selection include:

- Scope of interpretability: Global vs local explanations.
- Explanation target: Model understanding vs outcome understanding.
- Data type: Structured versus unstructured data.
- Explanation type: Feature importance, example-based, rule-based, etc.
- Explanation form: Visual, textual, local approximation, etc.
- Model-specific versus model-agnostic (model-generic) methods.

Common strategy for XAI selection includes distinguishing between model-specific and model-agnostic methods, local and global scope of explanations, the data type handled, the form of explanations provided (visual, textual, rules, etc.), and whether they explain the model or its outputs.

Current XAI studies are mostly focusing on classification problems with images or natural language data. However, since AI models in the nuclear industry are applied to various problems other than classification problems, further research is needed to investigate on how to apply XAI for such problems.

4.3 Transparency and Ethics

Transparency in AI is a multi-faceted concept. According to the OECD AI Principles, transparency can refer to (i) *"disclosing when AI is being used"*, (ii) *"enabling people to understand how an AI system is developed, trained, operates, and deployed in the relevant application domain"*, (iii) *"the ability to provide meaningful information and clarity about what information is provided and why"*, and (iv) *"facilitating public, multi-stakeholder discourse and the establishment of dedicated entities, as necessary, to foster general awareness and understanding of AI systems and increase acceptance and trust"*, [40].

According to [41], transparency is tightly linked to the concepts of explainability (see Section 4.2) and openness. In [42], it is noted that *"transparency has been highlighted as one of the key ethical considerations required to build trustworthy AI"* (pg. 1). As noted earlier, the trustworthiness of AI is a key factor when considering its potential use for safety analysis in NPPs. However, when determining how much of the inner workings of AI should be revealed, organisations need to consider the context within which the AI is used. [42] stated that "*While transparency is generally useful in the case of decision-making systems, especially when decisions are being suggested to aid human decision-makers, it isn't entirely clear whether the same is true for all types of AI systems and contexts of application. Additionally, this is transparency of the algorithms alone, outside of its situated context, and excluding the user interactions. Such a specific definition of transparency, arguably, is unlikely to have an effect on trust*". Further, the degree of transparency must be considered to avoid information overload for users, which can actually negatively affect the level of trust that users have in the AI, [42].

The concept of ethical AI refers the safety, privacy, security and the transparency of AI systems, as well as issues such as bias, diversity and privacy preferences, [43]. In fact, ethical AI is considered to be such a significant issue that there are now over 80 ethical AI guidelines and standards, [43], developed by international bodies and expert groups including the European Commission [44], UNESCO[2] [45] and the United Nations [46].

The ethical development and use of AI systems is not straightforward and must consider several factors including fairness, autonomy, anonymity and privacy. Ethical AI is tightly coupled to the concepts of explainability, i.e., transparency and interpretability. Transparency aims to detail how

[2] United Nations Educational, Scientific and Cultural Organization

data is processed and transformed in the ML algorithm. Interpretability focuses on the clarity of the relationship between the input features and the output decision. In simpler terms, while transparency reveals the model's mechanics, interpretability explains how specific inputs lead to particular outcomes.

Although ethical AI guidelines are often concerned with fairness, privacy and safety related to individuals, it is still an important concept for the development of AI/ML systems for safety analysis for NPPs. For example, accountability is an important issue in ethical AI – deciding who is responsible for the AI systems and the decisions they make, and also who is responsible if the system is seen to be causing harm, [43] (pg. 7). Further, explainability of systems is important to ensure that the AI/ML provides adequate justification of the outcomes of the system, and to support trustworthiness in those outcomes, [43].

4.4 Regulatory Requirements and Standards

As mentioned in section 4.1, the NRC anticipates increased use of AI in their regulated activities. Hence, the NRC has formulated a strategic goal to ensure readiness for regulatory decision-making and review of AI applications. AI technologies provide the underlying capability for autonomous systems. Higher autonomy levels indicate less reliance on human intervention or oversight and, therefore, may require greater regulatory scrutiny of the AI system, [37].

The NRC plan to assess whether any regulatory guidance (e.g., regulatory guides or standard review plan sections) or inspection procedures need to be updated or created to clarify the process and procedure for the licensing and oversight of AI in NRC-regulated activities.

The NRC will undertake research to develop an AI framework to determine the approach to assess technical areas such as, risk analysis, explainability, data quality, quantity, applicability, and uncertainty. The NRC will also work with agency stakeholders and the international regulatory community to determine the currently available AI standards and identify the technical areas where gaps may exist. In addition, the NRC will participate with standards development organisations and the international regulatory community (AI Strategic Goal 3) for development of AI standards and guidance documents, [37].

The IAEA has identified that concrete efforts will be needed to develop a roadmap guiding regulatory investigation, research and positioning on the application of AI systems for nuclear power plants, [33].

Enabling acceptance of AI in safety-related applications standards, especially with regard to safety and security, would need existing standards to be adapted, or new standards developed.

The development of AI technologies for safety critical applications could present a challenge to regulators, as many traditional assurance approaches might not be easily applicable.

One possible approach involves the IAEA assessing how AI and ML impact existing guidance and standards, offering insights on their suitability, particularly concerning safety. Regulatory approval is crucial for their implementation in nuclear power plants. However, the lack of a clear pathway for licensing AI applications, especially in safety-critical areas, may hinder their deployment and the potential benefits they offer. Establishing adaptable guidelines addressing common regulatory hurdles for AI and ML would facilitate their development and adoption. Such guidelines should emphasize the advantages of these technologies while ensuring their safe and responsible deployment.

The study in [47] draws attention to the hidden risks often embedded within AI's core decision-making mechanisms and acknowledges the limitations of directly applying IAEA's nuclear safety regulatory framework to AI.

An IAEA Technical Document (TECDOC), *"Safety Implications of the Use of Artificial Intelligence on Nuclear Power Plants"*, is planned to be published in late 2024, [48]. This TECDOC will address identified factors relevant to the use of AI in nuclear safety applications and the potential benefits and challenges of deploying AI.

In general, upcoming regulations would need to address the complete AI/ML lifecycle (in particular, for safety-critical systems), and demonstrate guidance on how to address the range of new vulnerabilities AI/ML systems present in terms of both the data utilised, and the model itself, [49].

The UK regulator, Office for Nuclear Regulation (ONR), plan in the next twelve months to release new guidance on regulating AI for their inspectors. ONR has also a set of AI-related regulatory objectives, see Figure 8, [48]. They will continue to support the safe deployment of AI systems, without compromising their independence. They will continue to work with industry, academia and domestic and international organisations to improve consistency of approach, reduce regulatory uncertainty and achieve common positions on technical matters relating to AI.

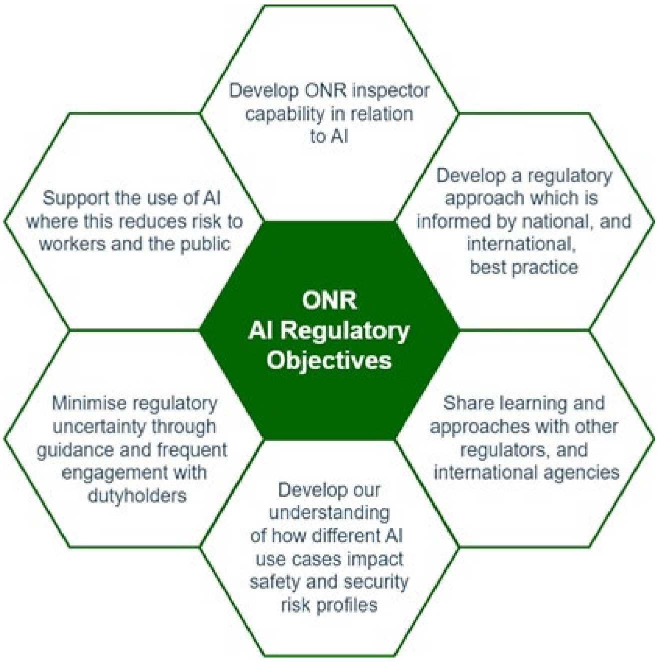

Figure 8 ONR AI regulatory objectives.

Recent collaboration between the regulatory bodies of the US, Canada and the UK have published a report, [50], on their view of AI systems for nuclear applications. The report provides a collection of standards and guidance across regulatory areas. The list contains both existing non-AI and AI-specific standards and guidance relevant to the nuclear industry. The report also concludes that *"the fast pace of AI development means it is unlikely that AI-specific consensus standards for the nuclear domain will be available to support regulatory activities within the near future. In the interim, existing nuclear-specific standards remain a starting point coupled with considering the unique attributes introduced by AI"*.

The overarching goal of these efforts is to ensure a harmonized approach to AI regulation that enhances safety and reliability while ensuring compliance with regulatory frameworks. The regulatory bodies have a vital role in setting and enforcing these regulations, as well as to continuously monitor and assess compliance. The main challenge is that regulations need to keep up with the fast-paced development of AI and simultaneously not hinder the continuous development. To mitigate this challenge, close cooperation between the regulatory body and the utilities is necessary.

Concerning standards, a survey by Idaho National Laboratory (INL) [51], which focuses on nuclear applications, note that "*although AI related laws and regulations have been slow to be enacted, professional organisations have been active in issuing standards for the use of AI*". The Institute of Electrical and Electronics Engineers (IEEE) has and is in the process of publishing several standards that relate to AI systems, see further [51].

In addition to IEEE, the International Organization for Standardization (ISO) has in conjunction with the International Electrotechnical Commission (IEC) also published standards relating to the use of AI, [51]. The nuclear power community has also created a dedicated ISO/IEC subcommittee (JTC 1/SC 42) working on promoting a rapid transfer of AI technologies from pilot studies to wide applications, [33].

[51] presents the following key areas for nuclear regulators and researchers to address:

- **Definitions:** Establish standardized definitions to promote faster evaluation and adoption of new proposals for nuclear applications.
- **System requirements:** Address performance requirements in regulations and standards. Numerous studies have demonstrated impressive evaluation results, such as accuracy, but it remains uncertain if these are adequate for nuclear systems. Additionally, there is a lack of guidance on the necessary transparency and explainability of these systems for implementation.
- **Application requirements:** Applications will likely require unique criteria. Hence, regulators should not only establish benchmark performance metrics for AI in nuclear systems but also identify a range of applications for potential approval.
- **Human factors and training:** Essential to focus on operator interaction, training, and readiness for these systems when implementing AI.

Another technical report[3], IEC TR 63468:2023 [13], overviews AI technologies from a nuclear perspective, and summaries potential AI application scenarios in nuclear facilities. This report was developed by the subcommittee SC 45A for "Instrumentation and control, and electrical power systems". The report identifies proven or potential applications, with the objective to foster better understanding and adoption of AI technologies and to support future standard development. It recommends setting up a new dedicated working group to be responsible for and coordinate standard development efforts. The report notes that "the regulatory framework from nuclear regulators is not yet established" and consequently, the focus is on non-safety related AI applications. Nevertheless, AI technologies can still be applicable in safety applications, given that both the technology and regulations support their use.

4.5 Digital Twins

A Digital Twin (DT) system consists of a physical system (e.g., plant components, sensors, operators), a virtual system (simulations, models), and data relationships between them. The system enables real-time data flow from plant sensors to the virtual model, which can then provide insights for predictive maintenance, anomaly detection, and support for decision making.

DTs can provide insights equivalent to Modelling and Simulation (M&S) but need to learn and provide those insights much faster than the development and uses of M&S. DTs are tightly coupled with operation with the ability to assimilate and adapt to real-time information from the operating environment through continuous learning.

[3] Note that an IEC technical report is entirely informative in nature, and it establishes no requirements.

The DT for nuclear systems has existed for decades in the form of on-line core monitoring systems. Advances and capabilities with ML-based predictive analysis has the potential to further improve decision-making capabilities for reactor operation.

Existing reactors traditionally rely on analogue instrumentation and legacy systems, so current DT systems are typically one-directional, from the physical asset to a digital model, rather than the full, real-time, bi-directional integration envisioned for next-generation systems. However, the increasing use of both digital systems and models for existing reactors provides an increased potential for adopting DT systems within several areas, [52].

- Plant-referenced simulators: These simulators replicate control room environments and simulate plant operations under normal, transient, and accident conditions. While they provide valuable training and design validation, their scope is limited to the fidelity required for operator training and post-event analysis, rather than full-scale, real-time, DTs.
- Condition monitoring and predictive maintenance: Modern upgrades include the installation of advanced sensors (e.g., for vibrations, thermal imaging, acoustics) and enhanced data analytics. Many US utilities have set up monitoring and diagnostic systems that use these sensor data streams (sometimes also augmented with ML) to detect anomalies, predict equipment degradation, and optimize maintenance intervals.

Digital Twins (DT) in the nuclear industry offer significant advantages in terms of operational efficiency, safety, reliability, and regulatory compliance. The technology is expected to see rapid adoption in the nuclear sector over the next 10 years as its benefits become more widely recognized. A special interest has been observed for advanced reactor designs and implement DT systems to optimize such reactors.

The integration of DTs into nuclear operations necessitates a thorough evaluation of current regulations to identify applicable criteria and potential gaps. This ensures that DT applications comply with safety standards and regulatory expectations. The regulatory requirements for DTs will vary depending on their specific applications. Each use case demands a tailored approach to meet relevant regulatory criteria. The regulatory requirements for a DT will be directly dependent upon how it is to be used (i.e., its functionality), which is discussed in [53].

The digital twin concept can be applied to several areas, for example, to support risk-informed safety analyses and PSA. Potential opportunities identified in [54] are:

- **Modelling and simulation to inform safety analyses:**
    - Support risk-informed and deterministic regulatory conclusions with new data and insights.
    - Support decision-making with integrated modelling and simulation.
- **Integration of DT with PSA:**
    - Inform PSA models with internal and external data and analysis and provide dynamic parameter estimation, SSC availability, or risk triplets for advanced reactor SSCs, etc.
    - Upgrade and augment PSAs using DTs (e.g., use of advanced models, physics-based or AI/ML, in DT for identifying novel failure modes, correlations, and dependencies).
    - Estimate human error probabilities and make recommendations for minimizing human errors and their consequences.

Three specific DT challenges when considering ML related to licensing and regulatory activities are identified in [55]. These are:

- Quality/optimum of input data: ML algorithms require significant amount of DT training data to produce reliable results. Gaps can occur when input data is insufficient or when the quality of the data is poor (may lead to false outputs or large uncertainties).
- Identification and selection of appropriate ML algorithms: the wide range of available ML methods and algorithms (as seen in this project's literature review) is a challenge to select the most appropriate algorithms and will be dependent on many factors (e.g., performance, size, complexity, scalability). Also, eliminating algorithm bias is a significant challenge.
- Explainability of the I/O relationships contained within the algorithms: The ability to explain and understand the algorithms relationships will be crucial, both for regulatory purposes but as well as for broader acceptance (i.e., the black-box nature needs to be avoided).

4.6 Natural Language Processing

NLP has the potential to derive actionable information from textual data. This has the potential to enable the nuclear industry to automate many processes, increase efficiency, and reduce costs.

However, open-source dictionaries do not understand the industry-specific language, and therefore is an industry-specific dictionary needed to conduct text mining and apply NLP-based algorithms, [56].

Some challenges that NLP methods are facing concern:

- The variation of phrasing/wording that varies between workers, sites, utilities, and countries. This is especially seen when looking at operating experience over time.
- The most promising NLP models have been developed for the English language, which results in limited applicability for non-English speaking countries.
- Data sharing limitations among utilities and organisations due to security constraints, in order to create larger, high-quality datasets for training NLP models.

Looking ahead, the U.S. Department of Energy's Light Water Reactor Sustainability (LWRS) program is involved with an AI and NLP application, which aims to develop virtual assistants and automated processes, [57].

<u>Large Language Models</u>
Large Language Models (LLMs) are a specific type of deep learning model designed for NLP tasks. They are typically characterized by having a very large number of parameters and being trained on massive datasets, which enables them to generate coherent and contextually relevant text, summarization, question answering, and more.

In the nuclear domain, [58] presents some previous efforts with LLMs, such as NukeBERT and NuclearQA in addition to their own model CurieLM. NukeBERT is a pre-trained language model derived from BERT (Bidirectional Encoder Representations from Transformers), specifically designed for applications within domains that possess limited datasets, such as nuclear science. During evaluations, it has demonstrated substantial performance enhancements compared to the original BERT baseline. NuclearQA is a benchmark tool designed to evaluate LLMs in the nuclear domain. This benchmark revealed a gap in current LLMs' scientific knowledge, pointing to the need for more specialized models. These domain-specific LLMs have demonstrated potential in their respective fields, indicating that fine-tuning LLMs on domain-specific data can be effective.

CurieLM, a LLM specifically tailored for the nuclear domain [58], which was guided by three key principles: ensuring high-quality data, maintaining domain specificity, and involving domain experts in the process.

The dataset included documents published by IAEA, French Institute for Radiation Protection and Nuclear Safety, etc. The study resulted in a model that outperformed the base LLM in correctness, language adaptability, and the provision of concise, relevant, and domain-specific responses. In addition, future research is however needed but also shows the future potential of LLMs in the nuclear domain.

4.7 ML Application Outlook in Sweden

Regarding NLP, there is in Sweden limited awareness and knowledge of NLP and its solutions for domain-specific issues (e.g., operation and maintenance) have not been explored within Nordic NPPs, [59]. One key takeaway in [59] is that the nuclear industry's language is highly specialized and varies significantly between utilities. Additionally, challenges include restricted permissions, limited data access and data quality.

AI efforts differ among Swedish utilities, with one utility notably ahead in its AI implementations compared to the other two, [60]. Some academic research in form of theses has been carried out at the utilities related to AI.

- Björn's thesis [61] explored detecting deviations in sensor data with AI.
- Lindskog and Gunnarsson's thesis [62] focused on predictive maintenance, utilizing ML to predict possible machine damage based on measurements and historical issues.
- Schultz’s thesis [63] investigated trending and condition monitoring to predict malfunctions of valve actuators.
- Sjögren’s thesis [64] explored anomaly detection with ML methods. Here, a data-driven approach using a hierarchical autoencoder framework was studied.

The analysis of component and sensor data remains the most prevalent area for Swedish stakeholders. Additionally, there has been demonstrated interest in predictive maintenance. At the Oskarshamn nuclear power plant a modern tool to harness the plant's data for early fault detection, preventing downtime and enabling better maintenance planning is being used. The tool for anomaly detection uses ML and the plant's process data to monitor operations, identify anomalies, and predict disturbances. This tool has also been evaluated at the other utilities in Sweden in a collaborative effort. Due to the ongoing, experimental nature of the work and the proprietary aspects of the algorithms, no references are available, [64].

In addition to these application and theses, several challenges have been identified and need to be resolved to further increase the applicability of AI/ML. As mentioned above, NLP faces challenges regarding approaches that use domain-specific language. Another area concerns data availability, since ML relies heavily on large datasets to be useful. The extensive historical operating experience of the Swedish nuclear power fleet has the potential to enable models and algorithms to train on such data. However, compared to other countries with nuclear power with much larger fleets, the potential will be limited.

International collaboration and sharing of datasets could bridge these gaps, but data security and data applicability will then be another issue to resolve. [59] identify that collaboration among utilities, stakeholders, and nations for data-driven research will be essential and would greatly benefit the sector when it comes to introducing AI in the nuclear industry.

While the current adoption of ML remains relatively low, it is highly probable that we will see either a rise in its usage or at least pilot implementations of various applications in the near future.

<u>Large language models</u>
AI Sweden, in partnership with WASP and RISE, has launched GPT-SW3, the first LLM designed for Swedish and Nordic languages, [65]. It is offered as an open model for businesses and

organizations to incorporate into their products and services. This model equips Sweden with a vital resource that enhances the capacity to leverage AI without ambiguities surrounding training data, unclear licensing terms, or that models are only accessible as cloud services.

Even though this model is not yet used in the nuclear industry, it is an enabler for future applications and use cases.

5 Future Research Directions

Future work within AI and ML in the nuclear industry represents a promising avenue for enhancing the sector's efficiency, safety, and innovation capacity. As AI and ML techniques continue to evolve, further research and development could yield substantial benefits, as discussed in chapter 4.

Further research could involve conducting benchmarking and comparison of different ML methodologies. This comparative analysis could highlight the strengths and weaknesses of various approaches, providing clearer insights and facilitating the selection of the most suitable methods for the application in question. Also, combining machine learning with traditional PSA and deterministic safety analysis (DSA) methods could offer more robust safety assessments. Research into hybrid models that include domain knowledge and physical laws with data-driven insights is necessary to ensure the reliability of AI in safety-critical systems.

Explainability and trust will be an integral part to aid the implementation of AI and ML. It is critical to ensure that the data-driven models' decisions are transparent and interpretable to build trust with operators and regulators. Further work with this issue is needed, especially for application areas that require regulatory approval. XAI techniques can offer insights into the decision-making process of complex models, making them more suitable for highly regulated environments like nuclear safety.

The use of actual plant data for the training and validation of AI and ML models should be prioritized. While simulators provide valuable data, real-world data can help ensure the accuracy, validity, and reliability. Collaboration could be beneficial to advance the useability of such models.

The application of digital twins could revolutionize predictive maintenance and risk analysis in nuclear facilities. By integrating ML with real-time sensor data and simulations, digital twins can provide highly accurate forecasts of equipment failures or safety risks. Future research should focus on how digital twins can be tailored to nuclear reactors, offering new pathways for nuclear safety.

Future research should also investigate the best ways to integrate human oversight with AI-driven systems. This includes designing interfaces and decision-support systems that enable operators to effectively interact with and oversee AI recommendations. Understanding the dynamics between human decision and AI outputs is critical to ensuring that AI systems in nuclear safety are both reliable and trustworthy.

Overall, further research is needed before the nuclear industry can harness the full potential of AI technologies and ML algorithms. By focusing on these areas, the nuclear industry can leverage the full potential of ML while addressing the unique safety, ethical, and regulatory challenges that this field presents.

6 Conclusions

In conclusion, exploring AI and ML applications in the nuclear industry has shown significant promise as well as notable challenges. The ongoing research and implementation efforts highlight the transformative impact these technologies could have or already has on various aspects of nuclear plant operations, including application areas such as predictive maintenance and safety analysis. This chapter aims to conclude the insights of this project.

Chapter 3 presents the literature review of the project. It includes examples and case studies of ML applications in various areas in the nuclear industry, such as safety and risk analysis, plant operation, and maintenance. The chapter highlights the potential of ML to enhance nuclear safety proactively. It emphasizes the international community's keen interest in adopting and/or further establishing novel approaches to enhance safety. However, it also identifies challenges such as data issues and the "black box" nature of some methods that need to be solved to achieve general acceptance of these advanced methods for nuclear safety.

One conclusion from the literature review is that there is perhaps a need for a more harmonized classification/categorization of applications. More importantly, the lack of guidance of which algorithm or method to use, data considerations and availability, depending on the application is evident and this could be a challenge in the future adoption or transition to ML approaches or to establish them further. Benchmark exercises as studied in [66], where different ML methods and models are trained and validated for a specific application, could be very beneficial and aid the adoption (or establish them further) of AI technologies and ML in the nuclear industry.

Key insights from the review of applications are:

- AI and ML can support risk analysis and risk-informed decision-making.
- AI and ML can enhance event and scenario analyses, i.e., aid the analysis to understand specific accident or event scenarios as well as identify events.
- AI and ML can enhance human reliability analysis by identification of strongly correlated contributing factors, which gives greater insight into both why the events occurred as well as possible improvements.
- AI and ML can be applied in prognostics and health management to predict and manage the health of systems and components.

Chapter 4 discusses the challenges and the future of AI and ML in the nuclear industry. It identifies the main barriers and limitations for applying ML techniques to complex and safety-critical systems, such as data availability and quality, model validation and verification, explainability, and ethics. The interpretability of AI technologies and ML is essential for their application in safety-critical environments.

While ML holds great promise for enhancing nuclear safety through improved decision-making, predictive analysis, and real-time monitoring, several challenges remain that must be addressed to fully realize its potential:

- Data availability and quality: As discussed in sections 3.1.4, 3.1.5 and chapter 4, the availability of high-quality, labelled data is a significant limitation. Many ML algorithms, particularly in supervised learning, rely on extensive datasets that are not always accessible. Efforts are needed to curate robust datasets, standardize data collection, and overcome challenges associated with sensitive or proprietary information.
- Explainability (transparency and interpretability): The "black box" nature of many advanced ML models, especially in DL, poses challenges in highly regulated environments such as nuclear safety. It is critical that ML models provide interpretable results, particularly when their outputs directly impact operational decisions or risk-informed decisions. Integrating XAI techniques could resolve this issue by offering a more transparent decision-making model while ensuring regulatory compliance.

- Regulatory and ethical considerations: Nuclear safety is governed by regulatory frameworks, which must be updated to accommodate the inclusion of AI technologies and ML models. Chapter 4 also highlights the importance of ethical considerations, such as ensuring that AI technologies do not introduce new risks. Collaborative efforts between regulatory bodies and AI developers will be essential to establish.
- Human factors and trust: Lastly, operators and decision-makers must be able to trust AI technologies. Addressing the human factors involved, such as how operators interact with and rely on these, is key to successful implementation. Future work should focus on integrating human oversight with AI technologies, providing operators with transparent, actionable insights without undermining their authority or expertise.

The initiatives by well-known organisations like the NRC and the IAEA are perhaps the key to successfully accelerate the development and adoption of AI and ML methods for the nuclear industry.

Concerning regulatory requirements, many activities are ongoing, and it is expected that this area will be improved in the future and will be able to provide more guidance for adoption of AI and ML. One observation is that, depending on the application area, the level and necessity of requirements to be fulfilled will vary significantly. For example, adoption of ML in fault diagnosis and predictive maintenance will not have the same restriction on transparency compared to if ML is used for safety analysis. Thus, it will be important that the regulatory bodies emphasize this in their work of developing regulatory requirements.

Maintaining continuous monitoring of developments in AI and ML within the nuclear industry is recommended. These technologies are advancing rapidly and offer significant opportunities to enhance safety, efficiency, and maintenance processes. By staying informed about the latest research advancements and applications, organizations can not only remain competitive but also identify and adopt innovative solutions that strengthen safety and reduce the risk.

The use of modern research accelerators that allow for discussion and collaborations in innovation is encouraged. These platforms can provide the infrastructure needed for researchers to share data, tools, and insights, accelerating the pace of innovation. By fostering a collaborative environment, it can be ensured that advancements in AI are quickly translated into practical applications in the nuclear industry. Ultimately, the goal is safe and effective applications of machine learning-based methods in the nuclear industry.

7 References

[1] S. Suman, "Artificial intelligence in nuclear industry: Chimera or solution?," *Journal of Cleaner Production,* 2020.

[2] L. Chao, L. Jiafei, Z. Liming, G. Aicheng, F. Yipeng, Y. Jiangpeng and L. Xiu, "Nuclear Power Plants With Artificial Intelligence in Industry 4.0 Era: Top-Level Design and Current Applications A Systemic Review," *IEEE,* 2020.

[3] PSAM, "PSAM 2023 Conference Proceedings," in *Probabilistic Safety Assessment and Management (PSAM) Topical*, Virtual meeting, 2023.

[4] EPRI, "An Introduction to AI, its Use Cases, and Requirements for the Electric Power Industry," 2019.

[5] H. Qingyu, P. Shinian, D. Jian, Z. Hui, Z. Zhuo, L. Yu, Y. Peng and . , "A review of the application of artificial intelligence to nuclear reactors: Where we are and what's next," *Heliyon,* vol. 9, 2023.

[6] NRC, "Exploring Advanced Computational Tools and Techniques with Artificial Intelligence and Machine Learning in Operating Nuclear Plants," NRC, NUREG-CR-7294, 2022.

[7] G.-F. Mario, H. Kathryn, T. Akira, W. Kent, W. Weng-Keen and Y. Haori, "Status of research and development of learning-based approaches in nucle-ar science and engineering: A review," *Nuclear Engineering and Design,* vol. 359, 2020.

[8] J. Pence, "Data-Theoretic Approach for Socio-Technical Risk Analysis: Text Mining Licensee Event Reports of U.S. Nuclear Power Plants," *Safety Science,* vol. 124, 2020.

[9] S. N. Hyun, S. K. Jung and S. J. Woo, "Survey on the Use of Artificial Intelligence in Nuclear Power Plants," in *Probabilistic Safety Assessment and Management (PSAM) Topical*, Virtual meeting, 2023.

[10] A. Abiodun, A. A. Muritala, A. O. Samuel, A. Yacine and A. Hafiz, "Deep learning for safety assessment of nuclear power reactors: Reliability, explainability, and research opportunities," *Progress in Nuclear Energy,* vol. 151, 2022.

[11] A. Gong, Y. Chen, J. Zhang and X. Li, "Possibilities of reinforcement learning for nuclear power plants: Evidence on current applications and beyond," *Nuclear Engineering and Technology,* vol. 45, no. 6, pp. 1959-1974, 2024.

[12] D. A. Ejigu, Y. Tuo and X. Liu, "Application of artificial intelligence technologies and big data computing for nuclear power plants control: a review," *Frontiers in Nuclear Engineering,* vol. 3, 2024.

[13] IEC, "Nuclear facilities – Instrumentation and control, and electrical power systems – Artificial Intelligence applications," IEC, TR 63468, 2023.

[14] K. Johnson, C. Morais and E. Patelli, "AI Tools for Human Reliability Analysis," in *5th ECCOMAS Thematic Conference on Uncertainty Quantification in Computational Sciences and Engineering*, Athens, Greece, 2023.

[15] A. Swain and H. Guttman, Handbook of human reliability analysis with emphasis on nuclear power plant applications, Sandia National Labs, Albuquerque, NM (USA), 1983.

[16] B. Kirwan, G. Basra and S. Taylor-Adams, "CORE-DATA: A Computerised Human Error Database for Human Reliability Support," in *IEEE Sixth Annual Human Factors Meeting*, Orlando, Florida (USA), 1997.

[17] H. Liao, K. Growth and S. Stevens-Adams, "Challenges in Leveraging Existing Human Performance Data for Quantifying the IDHEAS HRA Method," *Reliability Engineering & Systems Safety,* vol. 144, pp. 159-169, 2015.

[18] Y. Kim and J. Park, "Suggestions of HRA Method Improvement for the Practical Assessment of Human Reliability," *Journal of the Ergonomics Society of Korea,* vol. 37(3), pp. 229-241, 2018.

[19] A. Bye, “Future needs of human reliability analysis: The interaction between new technology, crew roles and performance,” *Safety Science,* vol. 158, 2023.

[20] Y. Zou, Z. Xiao, L. Zhang, E. Zio, J. Liu and H. Jia, “A data mining framework within the Chinese NPPs operating experience feedback system for identifying intrinsic correlations among human factors,” *Annals of Nuclear Energy,* vol. 116, pp. 163-170, 2018.

[21] D.-H. Ham and J. Park, “Use of a big data analysis technique for extracting HRA data from event investigation reports based on the Safety-II concept,” *Reliability Engineering and System Safety,* 2018.

[22] S. Yang, D. Demichela, J. Geng and H. Tab, “Analysis of Human and Organizational Factors Related Accident Reports Based on Natural Language Processing,” *Chemical Engineering Transactions,* vol. 90, pp. 745-750, 2022.

[23] P. Jonk, V. de Vries, R. Wevber, G. Sidiropoulos and E. Kanoulas, “Natural Language Processing of Aviation Occurrence Reports for Safety Management,” in *32nd European Safety and Reliability Conference (ESREL2022)*, Dublin, Ireland, 2022.

[24] C. Morais, K. Yung, K. Johnson, R. Moura, M. Beer and E. Patelli, “Identification of human errors and influencing factors: A machine learning approach,” *Safety Science,* vol. 146, 2022.

[25] I. Zgonnikova, A. Zgonnikov and S. Kanemoto, “Stick must fall: Using machine learning to predict human error in virtual balancing task,” in *IEEE 16th International Conference on Data Mining Workshops*, Barcelona, Spain, 2016.

[26] J. Heard, P. Baskaran and J. Adams, “Predicting task performance for intelligent human-machine interactions,” *Frontiers in Neurorobotics,* vol. 16, 2022.

[27] J. Canepa, A. Riber Marklund, S. Galushin and A. Olsson, “Evaluation of the source term prediction capability of RASTEP against integral response code calculations,” in *Probabilistic Safety Assessment and Management (PSAM)*, 2022.

[28] C. Wang, D. Mandelli and J. Cogliati, “Technical Language Processing of Nuclear Power Plants Equipment Reliability Data,” *Energies,* vol. 17, no. 1785, 2024.

[29] M. Zhegang, X. Fei and Z. Sai, “A New Approach to Identify and Characterize Low-Power Shutdown Initiating Events Using Machine Learning Techniques,” in *Probabilistic Safety Assessment and Management (PSAM) Topical*, Virtual meeting, 2023.

[30] H. Ujita, T. Morimoto, S. Futagami, H. Yamano and K. Kurisaka, “Development of Probabilistic Risk Assessment Methodology Using Artificial Intelligence Technology 2. Automatic Fault Detection Method for Building Reliability Database,” in *Probabilistic Safety Assessment and Management (PSAM) Topical*, Virtual meeting, 2023.

[31] P. Seurin and K. Shirvan, “Assessment of Reinforcement Learning Algorithms for Nuclear Power Plant Fuel Optimization,” *arXiv,* 2023.

[32] . Majdi I. Radaideh, W. Isaac, J. Joshua, T. James J., O. Uuganbayar, R. B. F. Nicholas and S. Koroush, “Physics-informed reinforcement learning optimization of nuclear assembly design,” *Nuclear Engineering and Design,* vol. 372, 2021.

[33] IAEA, “Artificial Intelligence for Accelerating Nuclear Applications, Science and Technology,” IAEA, 2022.

[34] INL, “Technical Language Processing of Nuclear Power Plants Equipment Reliability Data,” *Energies,* no. Special Issue Technological Advancements Enabling Sustainment and Expansion of the Nuclear Industry, 2024.

[35] H. Sun, P. Ramuhalli and R. E. Jacob, “Machine learning for ultrasonic nondestructive examination of welding defects: A systematic review,” *Elsevier,* vol. Ultrasonics 127, 2023.

[36] K. Araseethota Manjunatha, V. Agarwal and H. Palas, “Federated-Transfer Learning for Scalable Condition-based Monitoring of Nuclear Power Plant Components,” in *Probabilistic Safety Assessment and Management PSAM*, 2022.

[37] NRC, “Artificial Intelligence Strategic Plan. Fiscal Years 2023-2027,” NRC, NUREG-2261, 2023.

[38] "IAEA Designates First Collaborating Centre on Artificial Intelligence for Nuclear Power," 20 02 2024. [Online]. Available: https://www.iaea.org/newscenter/news/iaea-designates-first-collaborating-centre-on-artificial-intelligence-for-nuclear-power?_ga=2.59134504.401627378.1725866665-974444624.1725866665.

[39] "Task Force on Artificial Intelligence and Machine Learning for Scientific Computing in Nuclear Engineering," OECD-NEA, 2024. [Online]. Available: https://oecd-nea.org/jcms/pl_77779/task-force-on-artificial-intelligence-and-machine-learning-for-scientific-computing-in-nuclear-engineering. [Accessed 31 March 2024].

[40] "Transparency and explainability (Principle 1.3)," OECD, 2024. [Online]. Available: https://oecd.ai/en/dashboards/ai-principles/P7. [Accessed 31 March 2024].

[41] S. Larsson and . Heintz. F., "Transparency in Artificial Intelligence," *Internet Policy Review,* vol. 9(2), 2020.

[42] K. Haresamudram, S. Larsson and F. Heintz, "Three Levels of AI Transparency," *Computer,* vol. 56(2), pp. 93-100, 2023.

[43] N. Balasubramaniam, M. Kauppinen, S. Kujala and K. Hiekkanen, "Ethical Guidelines for Solving Ethical Issues and Developing AI Systems," *Product-Focused Software Process Improvement (Lecture Notes in Computer Science; Vol. 12562),* pp. 331-346, 2020.

[44] "Ethics Guidelines for Trustworthy AI," European Commission, 2019. [Online]. Available: https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=60419. [Accessed 1 July 2024].

[45] "Recommendation on the Ethics of Artificial Intelligence," UNESCO, 2022. [Online]. Available: https://unesdoc.unesco.org/ark:/48223/pf0000381137. [Accessed 1 July 2024].

[46] "Principles for the Ethical Use of Artificial Intelligence in the United Nations System," United Nations, 2022. [Online]. Available: https://unsceb.org/sites/default/files/2023-03/CEB_2022_2_Add.1%20%28AI%20ethics%20principles%29.pdf. [Accessed 1 July 2024].

[47] S. Cha, "Towards an international regulatory framework for AI safety: lessons from the IAEA's nuclear safety regulations," *Humanities and social sciences communications,* 2024.

[48] ONR, "ONR's pro-innovation approach to AI regulation," ONR, 2024.

[49] Adelard, "The Impact of AI/ML on Nuclear Regulation," Adelard, D/1321/165002/2, 2021.

[50] NRC, ONR, CNSC, "Considerations for Developing Artificial Intelligence Systems in Nuclear Applications," NRC, ONR, CNSC, 2024.

[51] INL, "A Survey of Proposed Standards and Regulations for Artificial Intelligence," INL, INL/CON-24-76067, 2023.

[52] INL, "The State of Technology of Application of Digital Twins," 2021.

[53] ORNL, "Status Report on Regulatory Criteria Applicable to the Use of Digital Twins," ORNL/SPR-2022/2493, 2022 .

[54] INL, NRC, "Regulatory Considerations for Nuclear Energy Application of Digital Twin Technologies," NRC, Letter report TLR-RES/DE/REB-2022-06, 2022.

[55] INL, "Technical Challenges and Gaps in Digital-Twin-Enabling Technologies for Nuclear Reactor Applications," INL/EXT-21-65316, revision 0, 2021.

[56] EPRI, "Quick Insight - Power Industry Dictionary for Text-Mining and Natural Language Processing Application: A Proof of Concept," EPRI, 3002019609, 2020.

[57] "A Gateway to Artificial Intelligence for the Nuclear Industry," American Nuclear Society (ANS), 1.0 June 2023. [Online]. Available: https://www.ans.org/news/article-5001/a-gateway-to-artificial-intelligence-for-the-nuclear-industry/. [Accessed 31 March 2024].

[58] Z. Bouhoun et. al., "CurieLM: Enhancing Large Language Models for Nuclear Domain Applications," *EPJ Web Conf. ,* vol. 302 17006, 2024.

[59] Energiforsk, "Feasibility Study On Artificial Intelligence Technologies In Nuclear Applications," Energiforsk, Report 2023:923, 2023.

[60] F. Kåhrström, "Natural Language Processing for Swedish Nuclear Power Plants," Uppsala University, 2022, Master Thesis.

[61] A. Björn, “Using machine learning to predict power deviations at Forsmark,” Uppsala University, 2021, Master thesis.

[62] J. Lindskog and R. Gunnarsson, “Databearbetning på Ringhals,” Högskolan in Halmstad, 2019, Bachelor thesis.

[63] E. Schultz, “Trending and condition monitoring of valve actuators in the safety system of Forsmark 1 & 2,” Uppsala University, 2020.

[64] S. Sjögren, “Anomaly detection with machine learning methods at Forsmark,” Uppsala University, 2023.

[65] AI Sweden, 16 November 2023. [Online]. Available: https://www.ai.se/en/news/open-release-first-large-nordic-language-model-gpt-sw3.

[66] M. Zubair and Y. Akram, “Enhancement in the safety and reliability of Pressurized Water reactors using Machine Learning approach,” *Annals of Nuclear Energy,* vol. 201, 2024.

8 Abbreviations

Table 4 Machine learning and AI abbreviations.

Acronym	Description
AI	Artificial Intelligence
ANN	Artificial Neural Network
BoW	Bag-of-Words
CNN	Convolutional Neural Network
DL	Deep Learning
DNN	Deep Neural Network
DT	Decision Tree Digital Twin
FNN	Fuzzy Neural Network
GA	Genetic Algorithm
GP	Gaussian Process
LLM	Large Language Model
LSTM	Long Short-Term Memory
ML	Machine Learning
NB	Naive Bayes
NLP	Natural Language Processing
RL	Reinforcement Learning
RNN	Recurrent Neural Network
SARSA	State-Action-Reward-State-Action
SVM	Support Vector Machine
TF-IDF	Term Frequency – Inverse Document Frequency
XAI	Explainable Artificial Intelligence

Table 5 Nuclear and safety assessment abbreviations.

Acronym	Description
FD	Fault Detection
FDDP	Fault Detection, Diagnosis, and Prognosis
HOF	Human and Organisational Factors
HRA	Human Reliability Analysis
NPP	Nuclear Power Plant
PHM	Prognostics and Health Management
PSA	Probabilistic Safety Assessment
RUL	Remaining Useful Life
SSC	Structures, Systems, and Components

Table 6 Organisational abbreviations.

Acronym	Description
IAEA	International Atomic Energy Agency
IEC	International Electrotechnical Commission
IEEE	Institute of Electrical and Electronics Engineers
INL	Idaho National Laboratory
ISO	International Organization for Standardization
NEA	Nuclear Energy Agency
NRC	Nuclear Regulatory Commission (United States)
OECD	Organisation for Economic Co-operation and Development
ONR	Office for Nuclear Regulation (United Kingdom)
PSAM	Probabilistic Safety Assessment and Management

The Swedish Radiation Safety Authority (SSM) works proactively and preventively with nuclear safety, radiation protection, nuclear security, and nuclear non-proliferation to protect people and the environment from the harmful effects of radiation, now and in the future.

You can download our publications from www.stralsakerhetsmyndigheten.se/en/publications. If you need alternative formats such as easy-to-read, Braille or Daisy, contact us by email at registrator@ssm.se.

Strålsäkerhetsmyndigheten
SE-171 16 Stockholm
+46 (0) 8-799 40 00
www.stralsakerhetsmyndigheten.se
registrator@ssm.se

Scandinavian Print Group Sweden AB, 2025