IAEA
International Atomic Energy Agency

IAEA NUCLEAR ENERGY SERIES

No. NR-T-1.26

Considerations for Deploying Artificial Intelligence Applications in the Nuclear Power Industry

TECHNICAL REPORTS

IAEA NUCLEAR ENERGY SERIES PUBLICATIONS

STRUCTURE OF THE IAEA NUCLEAR ENERGY SERIES

Under the terms of Articles III.A.3 and VIII.C of its Statute, the IAEA is authorized to "foster the exchange of scientific and technical information on the peaceful uses of atomic energy". The publications in the **IAEA Nuclear Energy Series** present good practices and advances in technology, as well as practical examples and experience in the areas of nuclear reactors, the nuclear fuel cycle, radioactive waste management and decommissioning, and on general issues relevant to nuclear energy. The **IAEA Nuclear Energy Series** is structured into four levels:

(1) The **Nuclear Energy Basic Principles** publication describes the rationale and vision for the peaceful uses of nuclear energy.
(2) **Nuclear Energy Series Objectives** publications describe what needs to be considered and the specific goals to be achieved in the subject areas at different stages of implementation.
(3) **Nuclear Energy Series Guides and Methodologies** provide high level guidance or methods on how to achieve the objectives related to the various topics and areas involving the peaceful uses of nuclear energy.
(4) **Nuclear Energy Series Technical Reports** provide additional, more detailed information on activities relating to topics explored in the **IAEA Nuclear Energy Series**.

Each publication undergoes internal peer review and is made available to Member States for comment prior to publication.

The IAEA Nuclear Energy Series publications are coded as follows: **NG** — nuclear energy general; **NR** — nuclear reactors (formerly **NP** — nuclear power); **NF** — nuclear fuel cycle; **NW** — radioactive waste management and decommissioning. In addition, the publications are available in English on the IAEA web site:

www.iaea.org/publications

For further information, please contact the IAEA at Vienna International Centre, PO Box 100, 1400 Vienna, Austria.

All users of the IAEA Nuclear Energy Series publications are invited to inform the IAEA of their experience for the purpose of ensuring that they continue to meet user needs. Information may be provided via the IAEA web site, by post, or by email to Official.Mail@iaea.org.

CONSIDERATIONS FOR DEPLOYING ARTIFICIAL INTELLIGENCE APPLICATIONS IN THE NUCLEAR POWER INDUSTRY

The following States are Members of the International Atomic Energy Agency:

AFGHANISTAN
ALBANIA
ALGERIA
ANGOLA
ANTIGUA AND BARBUDA
ARGENTINA
ARMENIA
AUSTRALIA
AUSTRIA
AZERBAIJAN
BAHAMAS, THE
BAHRAIN
BANGLADESH
BARBADOS
BELARUS
BELGIUM
BELIZE
BENIN
BOLIVIA, PLURINATIONAL STATE OF
BOSNIA AND HERZEGOVINA
BOTSWANA
BRAZIL
BRUNEI DARUSSALAM
BULGARIA
BURKINA FASO
BURUNDI
CABO VERDE
CAMBODIA
CAMEROON
CANADA
CENTRAL AFRICAN REPUBLIC
CHAD
CHILE
CHINA
COLOMBIA
COMOROS
CONGO
COOK ISLANDS
COSTA RICA
CÔTE D'IVOIRE
CROATIA
CUBA
CYPRUS
CZECH REPUBLIC
DEMOCRATIC REPUBLIC OF THE CONGO
DENMARK
DJIBOUTI
DOMINICA
DOMINICAN REPUBLIC
ECUADOR
EGYPT
EL SALVADOR
ERITREA
ESTONIA
ESWATINI
ETHIOPIA
FIJI
FINLAND
FRANCE
GABON
GAMBIA, THE
GEORGIA
GERMANY
GHANA
GREECE
GRENADA
GUATEMALA
GUINEA
GUYANA
HAITI
HOLY SEE
HONDURAS
HUNGARY
ICELAND
INDIA
INDONESIA
IRAN, ISLAMIC REPUBLIC OF
IRAQ
IRELAND
ISRAEL
ITALY
JAMAICA
JAPAN
JORDAN
KAZAKHSTAN
KENYA
KOREA, REPUBLIC OF
KUWAIT
KYRGYZSTAN
LAO PEOPLE'S DEMOCRATIC REPUBLIC
LATVIA
LEBANON
LESOTHO
LIBERIA
LIBYA
LIECHTENSTEIN
LITHUANIA
LUXEMBOURG
MADAGASCAR
MALAWI
MALAYSIA
MALI
MALTA
MARSHALL ISLANDS
MAURITANIA
MAURITIUS
MEXICO
MONACO
MONGOLIA
MONTENEGRO
MOROCCO
MOZAMBIQUE
MYANMAR
NAMIBIA
NEPAL
NETHERLANDS, KINGDOM OF THE
NEW ZEALAND
NICARAGUA
NIGER
NIGERIA
NORTH MACEDONIA
NORWAY
OMAN
PAKISTAN
PALAU
PANAMA
PAPUA NEW GUINEA
PARAGUAY
PERU
PHILIPPINES
POLAND
PORTUGAL
QATAR
REPUBLIC OF MOLDOVA
ROMANIA
RUSSIAN FEDERATION
RWANDA
SAINT KITTS AND NEVIS
SAINT LUCIA
SAINT VINCENT AND THE GRENADINES
SAMOA
SAN MARINO
SAUDI ARABIA
SENEGAL
SERBIA
SEYCHELLES
SIERRA LEONE
SINGAPORE
SLOVAKIA
SLOVENIA
SOMALIA
SOUTH AFRICA
SPAIN
SRI LANKA
SUDAN
SWEDEN
SWITZERLAND
SYRIAN ARAB REPUBLIC
TAJIKISTAN
THAILAND
TOGO
TONGA
TRINIDAD AND TOBAGO
TUNISIA
TÜRKİYE
TURKMENISTAN
UGANDA
UKRAINE
UNITED ARAB EMIRATES
UNITED KINGDOM OF GREAT BRITAIN AND NORTHERN IRELAND
UNITED REPUBLIC OF TANZANIA
UNITED STATES OF AMERICA
URUGUAY
UZBEKISTAN
VANUATU
VENEZUELA, BOLIVARIAN REPUBLIC OF
VIET NAM
YEMEN
ZAMBIA
ZIMBABWE

The Agency's Statute was approved on 23 October 1956 by the Conference on the Statute of the IAEA held at United Nations Headquarters, New York; it entered into force on 29 July 1957. The Headquarters of the Agency are situated in Vienna. Its principal objective is "to accelerate and enlarge the contribution of atomic energy to peace, health and prosperity throughout the world''.

IAEA NUCLEAR ENERGY SERIES No. NR-T-1.26

CONSIDERATIONS FOR DEPLOYING ARTIFICIAL INTELLIGENCE APPLICATIONS IN THE NUCLEAR POWER INDUSTRY

INTERNATIONAL ATOMIC ENERGY AGENCY
VIENNA, 2025

COPYRIGHT NOTICE

Publishing Section
International Atomic Energy Agency
Vienna International Centre
PO Box 100
1400 Vienna, Austria
tel.: +43 1 2600 22529 or 22530
email: sales.publications@iaea.org
www.iaea.org/publications

Printed by the IAEA in Austria
September 2025
STI/PUB/2119
https://doi.org/10.61092/iaea.s6uy-wjt8

IAEA Library Cataloguing in Publication Data

Names: International Atomic Energy Agency.
Title: Considerations for deploying artificial intelligence applications in the nuclear power industry / International Atomic Energy Agency.
Description: Vienna : International Atomic Energy Agency, 2025. | Series: IAEA nuclear energy series, ISSN 1995-7807 ; no. NR-T-1.26 | Includes bibliographical references.
Identifiers: IAEAL 25-01774 | ISBN 978-92-0-115525-2 (paperback : alk. paper) | ISBN 978-92-0-115625-9 (pdf) | ISBN 978-92-0-115725-6 (epub)
Subjects: LCSH: Nuclear power plants — Technological innovations. | Nuclear power plants — Automatic control. | Nuclear power plants — Management. | Artificial intelligence.
Classification: UDC 621.039.5:004.8 | STI/PUB/2119

FOREWORD

The IAEA's statutory role is to "seek to accelerate and enlarge the contribution of atomic energy to peace, health and prosperity throughout the world". Among other functions, the IAEA is authorized to "foster the exchange of scientific and technical information on peaceful uses of atomic energy". One way this is achieved is through a range of technical publications including the IAEA Nuclear Energy Series.

The IAEA Nuclear Energy Series comprises publications designed to further the use of nuclear technologies in support of sustainable development, to advance nuclear science and technology, catalyse innovation and build capacity to support the existing and expanded use of nuclear power and nuclear science applications. The publications include information covering all policy, technological and management aspects of the definition and implementation of activities involving the peaceful use of nuclear technology. While the guidance provided in IAEA Nuclear Energy Series publications does not constitute Member States' consensus, it has undergone internal peer review and been made available to Member States for comment prior to publication.

The IAEA safety standards establish fundamental principles, requirements and recommendations to ensure nuclear safety and serve as a global reference for protecting people and the environment from harmful effects of ionizing radiation.

When IAEA Nuclear Energy Series publications address safety, it is ensured that the IAEA safety standards are referred to as the current boundary conditions for the application of nuclear technology.

As artificial intelligence (AI) technologies continue to evolve, they are being used to address a growing number of different applications within the nuclear power industry, facilitated by concurrent technological advancements in sensors, data management, communications and computational capabilities. There is great interest in harnessing AI capabilities throughout the life cycle of nuclear power plants, from design to decommissioning, which is expected to help transform the industry's long term economics and operational efficiency. As the advancements in AI capabilities continue to uncover applications within the nuclear power industry that were previously considered impractical, gaps and challenges need to be addressed to allow AI technologies to transition from the research and development domain to operational use in nuclear facilities. Addressing these gaps and challenges is not straightforward (irrespective of reactor technologies) because many considerations are expected to influence the deployment of AI technologies, such as the historical safety culture of the industry, technical aspects, the need for a valid business value proposition, organization and regulatory readiness, and end user and public acceptance.

Recognizing the need to identify means to support AI technology deployment, the International Network on Innovation to Support Operating Nuclear Power Plants formed a working committee comprising international subject matter experts and advisers from Member States. The working committee solicited input and reviews from representatives of different Member States and engaged with them at technical meetings to develop this publication. The overall objective of this publication is to capture various considerations relating to the design, development and deployment of AI technologies in nuclear power plants. These considerations are expected to provide information to Member States for use in their AI design, development and deployment strategies.

The IAEA is grateful to all the contributors and reviewers listed at the end of this publication, in particular V. Agarwal (United States of America). The IAEA officer responsible for this publication was N. Ngoy Kubelwa of the Division of Nuclear Power.

EDITORAL NOTE

This publication has been edited by the editorial staff of the IAEA to the extent considered necessary for the reader's assistance. It does not address questions of responsibility, legal or otherwise, for acts or omissions on the part of any person.

Although great care has been taken to maintain the accuracy of information contained in this publication, neither the IAEA nor its Member States assume any responsibility for consequences which may arise from its use.

Guidance and recommendations provided here in relation to identified good practices represent experts' opinions but are not made on the basis of a consensus of all Member States.

The use of particular designations of countries or territories does not imply any judgement by the publisher, the IAEA, as to the legal status of such countries or territories, of their authorities and institutions or of the delimitation of their boundaries.

The mention of names of specific companies or products (whether or not indicated as registered) does not imply any intention to infringe proprietary rights, nor should it be construed as an endorsement or recommendation on the part of the IAEA.

The IAEA has no responsibility for the persistence or accuracy of URLs for external or third party Internet web sites referred to in this book and does not guarantee that any content on such web sites is, or will remain, accurate or appropriate.

CONTENTS

1. INTRODUCTION

1.1. BACKGROUND

Artificial intelligence (AI) and machine learning (ML) technologies continue to evolve to address different applications within the nuclear power industry. There is great interest in harnessing AI and ML capabilities throughout the life cycle of a nuclear power plant, from design to decommissioning, which will help transform the industry's long term economics and operational efficiency. Key opportunities for AI and ML application include reactor design, operations and maintenance (O&M) activities, management of materials, operational flexibility, and expanded deployment.

AI and ML have been topics of research and development within the nuclear power industry for several decades, with ebbs and flows in interest. Research and development efforts have ranged from rule based to ML approaches. Classic examples of ML applications for reactor O&M include on-line condition monitoring using different types of artificial neural networks (ANNs); fault diagnosis and anomaly detection using unsupervised, semi-supervised, and supervised ML techniques; and maintenance optimization using condition based monitoring. Technological advancements in the areas of sensors, data management, communications and computational capabilities (from high performance computation to edge computing) have accelerated the development of new AI capabilities that are data driven and physics informed. These AI capabilities include deep learning, natural language processing (NLP) and, more recently, large language models (LLMs).[1] Advancements in AI capabilities continue to find applications within the nuclear power industry that were previously considered impractical. Examples of such applications include text mining using NLP, integration of predictive analytics with dynamic risk assessment, structural health monitoring, and automation of non-destructive examination techniques and manually performed O&M activities.

Continued advancements in AI technologies and their potential adoption (e.g. deployment) in the nuclear power industry could influence a wide range of needs and applications, enabling the industry to achieve better economic performance. The adoption of AI technologies requires a concerted effort from the nuclear power industry to raise the readiness level of AI technologies from the research and development domain to the field. This transition is not straightforward (irrespective of reactor technologies) as multiple considerations are expected to influence the deployment of AI technologies: the historical safety culture of the industry, technical aspects (e.g. data accessibility, quantity and quality; data and AI governance; technical expertise; explainability and trustworthiness of AI applications; and cybersecurity), the need for a valid business value proposition, organizational and regulatory readiness, and end user and public acceptance.

The IAEA, through the International Network on Innovation to Support Operating Nuclear Power Plants, is playing a significant role in increasing collaboration and experience sharing in the field of innovation for the nuclear power industry. AI technologies are one of the innovations within the network scope. For this publication, the IAEA hosted and co-organized a series of meetings where AI subject matter experts, developers and users representing different Member States and organizations (academia, industry, utilities, regulators, and national laboratories) provided input on different aspects of AI innovations, the steps taken to deploy some of the innovations, and the lessons learned. This collective knowledge laid the foundation for capturing considerations for the deployment of AI technologies within the nuclear power industry. These considerations are dynamic and are expected to evolve with AI technologies. They could also be used to inform a deployment strategy within other areas of nuclear science, such as nuclear fusion and nuclear medicine.

[1] Hereafter, the term AI will be used to include ML techniques, unless specified otherwise.

1.2. OBJECTIVE

The objectives of this publication are to:

— Provide an overview of various data driven and physics informed AI and ML applications in the nuclear power industry that have the potential to advance the autonomy level (see Section 1.3.1) when deployed in a nuclear power plant;
— Identify technical, human factors, stakeholder, regulatory, and other considerations (including hardware, software and infrastructure) for enabling broader deployment of AI solutions and end user acceptance;
— Emphasize that different categories of data, their relevance, and governance ensuring data integrity play a vital role in developing an AI technology;
— Promote stakeholder engagement and risk assessment throughout the AI development and deployment process.

Guidance and recommendations provided here in relation to identified good practices represent experts' opinions but are not made on the basis of a consensus of all Member States.

1.3. SCOPE

This publication provides high level considerations to enable the deployment of AI applications for the nuclear power industry. These considerations encompass topics related to data, data management and governance; the design, development, deployment, operation and quality monitoring of an AI technology; and regulatory and stakeholder needs in implementing the AI technology. These considerations are expected to be adapted to specific application needs. AI technology should be used to address the specific needs (problem statement) of a nuclear power plant, and it should be understood that other technologies or approaches might exist that could be employed to address those needs. Therefore, the following questions should be asked:

— Why is AI needed?
— How will it address or solve the problem of interest and how does this differ to other technologies?
— What capabilities of AI make it a better solution than other technologies?
— What additional actions need to be taken to develop, deploy and implement an AI technology?

This publication does not focus on the following:

— Application of rule based techniques as a category of AI approaches. Such techniques are already captured within existing industry practices.
— Safety and security aspects of AI applications, as they will be addressed in a dedicated IAEA publication.

1.3.1. Levels of automation

This section discusses levels of automation, per section 9 of Ref. [1], and outlines parallel notional AI and autonomy levels, as per Artificial Intelligence Strategic Plan: Fiscal Years 2023–2027 [2]. The United States (US) Nuclear Regulatory Commission (NRC) recognizes the differences between automation and autonomy in nuclear applications that could use AI, as presented in Ref. [2].

'Levels of automation' in section 9 of Ref. [1] refers to the extent to which a task is automated: Level 1 refers to fully manual (no automation), and Level 5 refers to fully autonomous operation (human(s) monitor performance and act as backup, if necessary, feasible and permitted). The automation levels described in section 9 of Ref. [1] depend on human operators to provide oversight to monitor plant

performance and intervene when necessary. According to these guidelines, automation technology[2] aims to assist operators rather than replace their daily operational and tactical control responsibilities. The NRC's notional AI and autonomy levels (Level 0 (AI not used) to Level 4 (machine decision making with no human intervention)) as outlined in Ref. [2] (see table 1 of Ref. [2]) are analogous to the levels of automation in section 9 of Ref. [1], except for two differences:

— Autonomy levels[3] specifically call out the use of AI, except for Level 0. Autonomy Levels 1 and 2 are about human decision making assisted and augmented, respectively, by AI; Levels 3 and 4 are about machine decision making supervised by humans and machine decision making with no human intervention, respectively.
— Level 4, denoting 'fully autonomous' (the highest level of autonomy achievable), involves no human intervention in decision making; in comparison, for NUREG-0700 Level 5 (autonomous operation) [1] humans can still intervene.

The automation and autonomy levels both recognize that as technological advancements (including AI) mature and are integrated into nuclear systems for decision making, day-to-day human involvement will be reduced. In this publication, the NRC's notional AI and autonomy levels as outlined in Ref. [2] will be used as a point of reference for the discussion of various aspects of AI technology in the nuclear power industry.

1.3.2. Value proposition

AI is one of the technologies that can be used in different nuclear power plant O&M activities as well as in other aspects of the nuclear fuel cycle. AI presents a value proposition to the nuclear power industry to increase operational efficiency by automating some manually performed tasks; by reducing human errors; by enhancing the reliability of structures, systems and components; by enabling predictive maintenance, outage optimization and preventive maintenance optimization; and even by enhancing nuclear safety. All of these benefits could strengthen nuclear economics in competitive energy markets. Despite the impressive value proposition of AI technologies in different nuclear applications, their adoption in the nuclear power industry is slow and faces several challenges [3]. This slow adoption might be due to lack of sharing of knowledge, considerations and lessons learned among stakeholders. This publication attempts to capture some of these considerations, although not all, as AI is evolving.

1.4. STRUCTURE

This publication is organized as follows:

— Section 2 provides an overview of the applications and capabilities of AI technologies in the nuclear power industry, emphasizing the benefits of these technologies.
— Section 3 presents insights into different considerations for developing life cycle management of AI systems (design, development, deployment, and operation and monitoring), promoting broader end user acceptance.
— Section 4 discusses different categories of data, their relevance and their governance as their integrity plays a vital role in developing an AI technology.
— Section 5 presents further deployment considerations including regulatory oversight and engagement strategies.

---

[2] NUREG-0700 [1] does not refer to any specific automation technology. However, AI could be one of the potential automation technologies.

[3] Autonomy can be achieved without AI, but such approaches are outside the scope of the current publication.

The publication posits that no single section should be tackled in isolation and instead advocates for a holistic review of all sections to support deployment considerations.

2. BENEFITS FOR NUCLEAR POWER PLANTS

AI and ML could have a broad range of possible applications within a nuclear power plant. These applications include (but are not limited to) robotics and maintenance, alarm and signal validation, emergency response, process diagnostics, human–machine interfacing, plant control systems, equipment diagnostics, operation analysis, plant operations and support, probabilistic risk assessment, teaching and learning, and fuel manufacturing. Such applications could provide many benefits to the nuclear power plant, with cost reduction, reduced radiation exposure, and safety improvement being some of the most relevant. Recent advances in AI and associated technologies, such as sensors and computation resources, have improved the maturity and accessibility of AI technologies. However, AI is still an emerging technology, and most proposed applications are still in pilot phases at nuclear power plants.

AI provides the opportunity to automate tasks to focus human effort and enables applications that were previously impossible. Often, this is because the task parameters are inexplicable to the extent required for traditional programming. With ML, the algorithm can be 'trained', if given sufficient data, to find a good solution within the scope of the training data. Even when a task is or can be automated using traditional means, data driven models can often be used to improve performance [4, 5].

Safety and economics are the key issues for the sustainment and modernization of nuclear power and are comprehensively determined by the design, construction, operation and retirement stages of nuclear power plants. Operation is an important stage directly related to economic benefits. O&M accounts for a significant portion of the operating costs of any nuclear power plant. Therefore, there is enormous potential for improving nuclear power plant economics by reducing O&M costs. AI technologies are expected to enhance the decision making and control capabilities of nuclear power plants through data mining of O&M data and the use of innovative solutions. For example, ML can be applied to the analysis of equipment performance data (e.g. temperatures, flow, vibrations) to both inform maintenance frequency and identify impending failures.

In terms of safety, with the application of AI technologies, abnormal situations can be detected early, human error may be reduced, and complex information can be fused to assist with, and possibly make, decisions. In terms of economics, the application of AI technologies can improve the availability of nuclear power plants, reduce comprehensive maintenance costs, reduce dependence on personnel and improve the overall efficiency of O&M.

Through efficient and secure access to the operation and management data of nuclear power plants, the application of AI technology can bring benefits to the safety, availability and economy of nuclear power plants, such as by:

— Increasing automated operation and auxiliary decision making, reducing the workload of nuclear power plant operators;
— Strengthening the monitoring and analysis capability of equipment status, system status and unit status, enhancing the safety of nuclear power plants;
— Providing more efficient and flexible tools for nuclear power plant operation and management personnel.

To advance the use of AI technology, with enduring and practical advantages for nuclear power plants, it is important to delineate the scope of AI application with the goal of enhancing plant safety, reliability and economics. In Sections 2.1–2.4, some current or upcoming applications are reviewed

to provide concrete examples of how AI can be used and how it can provide value within the nuclear power industry.

2.1. GENERIC AREAS OF ARTIFICIAL INTELLIGENCE CAPABILITY

The range of different AI techniques is vast. For many tasks, multiple different approaches may provide a viable solution. As is the case in many fields of engineering, for any given task a preferred approach emerges, usually representing the current state of the art. AI models can be as simple as a linear regression model to predict the thermal losses of a plant based on cooling water temperature or as complex as a deep learning model to extract specific insights from a text document. What these two extreme examples have in common is a set of representative training datasets to teach the model the desired behaviour of the system.

In this section, the main areas of AI capability are reviewed to introduce various possible AI applications in nuclear power plants. Often, the chosen approach significantly impacts the entire process, from the required training data and computing capacity to the available results.

2.1.1. Shallow models

Shallow models (e.g. decision trees, linear regression, support vector machines, Bayesian models) have been implemented in process optimization applications. Their reported advantages include improved process efficiency and early fault or anomaly detection.

Shallow models have traditionally suffered from reliance on hand tuned features (‘feature engineering’). Such reliance limits their usability to domains where the decision can be based on a relatively small number of important features that can be explicated and may result in overly restrictive models.

2.1.2. Deep learning networks

Deep learning networks have shown promise in various machine vision tasks, from inspection data evaluation to fire safety improvement, and as part of robotic tasks. The reported benefits include improvements in the consistency and efficiency of repetitive data evaluation tasks, which due to their nature may be susceptible to human errors.

Convolutional neural networks are being introduced in the light water reactor industry to generate more accurate three dimensional visualizations of core performance by integrating high fidelity simulations with on-line sensor measurements. The combined spatial map exhibits reduced uncertainty in important core performance parameters, such as linear heat rate and burnup, allowing more energy to be obtained from a core loading and thereby increasing revenue.

Deep learning networks have successfully been used in the evaluation of data from non-destructive testing processes. In-service inspections generate vast amounts of data, most of which do not contain indications of flaws. AI models can sift through these data and pinpoint areas of interest for further evaluation by the human inspector, allowing the inspector to focus on anomalous data, reducing potential human factors related errors. Furthermore, the results are available quickly, which helps in scheduling outage activities.

Convolutional and other deep learning models substantially increase model complexity to the point where the internal logic becomes obfuscated and difficult to monitor directly. This limits model validation to performance based and other indirect methods and may limit the use of these models where formal verification is required.

2.1.3. Natural language processing

NLP and the new LLMs present a new opportunity to process and produce textual data. Recently, these models have become widely available and have shown potential in nuclear use cases. Among other capabilities, the models have shown the capacity to process large sets of textual data and can provide efficient ways to extract information from or check compliance with textual guidelines. LLMs can also provide context specific guidance and help support best practices, especially with a changing workforce.

The application of AI to natural language started with extracting keywords from text documents. With recent advancements in NLP and the creation of LLMs, there is a huge opportunity for utilities to harness insights from information within their text documents, databases, pictures, videos and audio files. Advancements in robotics and image processing techniques have also created an opportunity for the industry to leverage these techniques for anomaly detection, visual inspection and process optimization.

NLP is also being used to explain, in terms understandable to humans, how an automated reasoning algorithm arrived at its answer in cases when domain knowledge was included in the reasoning process [6]. In these instances, the user would like to validate how the algorithm got its answer, but in a form more accessible than the mathematical abstraction inside the reasoning engine. The key enabling factor is that the human has an analogous, albeit qualitative, understanding of the information embedded in the data. In this application, NLP transforms the internal mathematical reasoning sequence into a textual description. This approach brings the algorithm to the human rather than the other way around.

LLMs require a staggering level of model complexity, to the point where even complete training from scratch is infeasible for most use cases. Consequently, they entail increasing reliance on pretrained models that are then tuned or prompted to attune to specific issues. This reduces the visibility of the training data, reduces predictability of the model behaviour and may make models more susceptible to abuse through adversarial inputs.

Overall, these AI techniques provide an opportunity to improve and increase the automation of previously difficult or intractable tasks, from process automation to complex multidimensional data evaluation and conversational support. The data driven methods are inherently limited by what can be learned from the training data, and their performance is expected to deteriorate with increasing deviation from the training scope. The extremely high dimensionality of the models enables them to learn complex tasks but also makes them susceptible to highly unexpected or undesired behaviour when confronted with unreasonable or adversarial input. They are best used when representative training can be made available and the inputs and/or outputs properly managed and coupled with industry and application specific validation.

2.2. ARTIFICIAL INTELLIGENCE APPLICATIONS

This section describes some examples of AI applications. Many of the cases shown here are still in the pilot phase.

2.2.1. Teaching and learning

LLMs are used in nuclear teaching and learning activities for both internal employees and external clients and stakeholders. These applications are particularly relevant to the nuclear power industry due to its specific qualification requirements and challenges with knowledge retention due to a transitioning workforce.

The use of AI, and in particular generative AI, could significantly enhance the learning experience and results in the nuclear power industry. AI could also improve the cost efficiency of training activities by reducing the time needed to produce content and to assimilate it. AI can be used to provide a more personalized, effective and engaging learning experience.

Connecting LLMs' capabilities to an existing knowledge base may allow people to find answers to their questions quickly and easily. This will reduce the time and effort required to search for information. Here are some of the main use cases tested so far:

— Summarizing training documentation and providing key take aways;
— Producing ad hoc exams using different kinds of questions (e.g. multiple choice, true–false, fill in the blank);
— Producing new perspectives by relating concepts together or presenting the information in another way;
— Providing specific information for the users' needs (e.g. locating the right insight within a large amount of documentation);
— Adapting responses to the users' needs, for instance simplifying complex topics for beginners or providing more expert descriptions for an advanced learner;
— Providing multilanguage interactions, for instance interaction in Spanish or French on English documentation or vice versa.

2.2.2. Monitoring and diagnostics

With recent advancements in data analytic capabilities, the combination of real-time sensor data and continuous models operating on that data has become more available. Integrating real-time data with a simulated environment enables predictive maintenance, real-time monitoring, and simulation of future states. In addition to sensor updates and physical or data driven models, analysis of the health of an asset can be updated using historical information such as operator logs, engineering inspection and condition assessment reports, and previous root cause analysis [7, 8].

Monitoring and diagnostic analysts work closely with operations and engineering teams to assess the incoming alerts from anomaly detection models. Like other AI applications, the output of the models needs to be verified. These groups are also involved when new AI models are being built to provide their input on what instrumentation signals need to be included in the model and on normal operating values for model training.

The benefits that utilities gain by implementing AI monitoring and diagnostics span multiple categories:

— Safety benefits: Continuous feedback afforded by automated evaluation can improve both response and safety.
— Cost benefits: Detecting anomalies and acting on them before they cause outages or derating can reduce costs.
— Time benefits: Facilitating access to data and trend analysis can save time in troubleshooting and reporting, thereby optimizing preventive maintenance and condition based maintenance.
— Generation benefits: First principles models and data driven models can be used to identify areas of potential generation loss and how to improve them.

A data driven model for predicting moisture carryover in a boiling water reactor was developed using a physics constrained AI technique. Accurate moisture carryover predictions are valuable for commercial boiling water reactor operators because they can be used to adjust operational plans during power cycles to reduce high moisture carryover. This helps prevent increased radiation exposure for on-site personnel and damage to turbine components. This predictive capability is currently used for planning core reloads and for scheduling operations for ongoing cycles. The developers of this technique described the importance of including subject matter expertise on the phenomena being modelled and the importance of understanding the limitations inherent in nuclear datasets, which tend to be 'small data'.

To assess the risk of evolving equipment degradation and related uncertainty, Bayesian inference can be performed to assess the fault posterior distribution using Markov chain Monte Carlo algorithms [9].

During events and emergencies, computer vision could be used to monitor variables at different levels of the reactor building and to undertake real-time sensing of radiation data, enabling expert systems to alert plant operators about potential dangers.

AI can process sensor data for radiation levels and provide continuous radiation monitoring and management. ML algorithms can be used to learn what normal radiation levels are and can alert human operators immediately when anomalies or danger signals are detected. More accurate dose projections can be made using historical data from previous field work and current plant conditions.

2.2.3. Text analytics using large language models

Utilities possess a large volume of information in text format, which includes nuclear power plant condition records, operational experience, work order details, maintenance reports, design and operating manuals for various components and systems, periodic inspection reports, and root cause analysis reports. A large number of research and development reports are also generated by external entities and agencies. The knowledge captured in these documents is useful for areas such as day-to-day O&M, health and safety, assessment of equipment reliability, and asset management. Most of these data sources exist in siloed databases with limited traditional (keyword) search capabilities. As a result, information retrieval from these sources is very labour intensive, causing these rich resources to be heavily underutilized.

Advancements in NLP through deep learning algorithms have resulted in the creation of LLM. Potential applications for LLMs with nuclear power plant documents include insight extraction, summarization, semantic search and report generation. Furthermore, the expansion of cloud computing provides the hardware resources required by LLMs.

Leveraging these LLMs has become a priority in various industries, including electric utilities. LLMs are available through either open source codes or application programming interfaces from various technology providers. Leading utilities are using these technologies to enhance their internal search capabilities for document retrieval and to extract relevant events data and plot them, to gain insights from their previous health and safety related observations and operational experience as well as to summarize large volumes of text information to a human-manageable size.

2.2.4. Assistance to non-destructive evaluation inspections

Non-destructive evaluations are an integral and important part of nuclear power plant operations; multiple components are periodically inspected for fitness for service in relation to known and expected active degradation mechanisms. Although some inspections are performed while the plant is in operation, many are part of the busy schedule of a refuelling outage. Inspections can be performed under challenging environmental conditions, and some can generate large volumes of data for posterior review and analysis. Often, data analysis can involve long reviews that require a great deal of attention to and focus on monotonous data, and fatigue and distractions are significant human factors affecting quality and reliability. Other inspections are performed only at long intervals, and maintaining high proficiency is challenging.

Although the context of non-destructive evaluations varies considerably, they are a prime example of an area that can greatly benefit from the assistance of AI tools. There are multiple ways in which AI may assist non-destructive evaluations; the main two are as follows:

— AI tools may aid the inspector by providing augmented data for real-time evaluation during live inspections. Such assistance is expected to lead to considerable time savings and increased reliability of the inspection.
— AI tools may assist the inspector in post-inspection review and analysis of the data. This may take different forms:

- The AI tool may provide fully automated solutions.
- The AI tool may provide assisted analysis, where it flags regions of interest or potential indications in a larger volume of data that require review by an expert analyst, thus effectively screening the data while maintaining the final decision as a responsibility of the expert analyst.

In either of these cases, the AI tool may help with detection, characterization or both. Additional benefits that the AI tool, having accomplished one or both of those tasks, may provide include automation of clerical tasks such as reporting or verification of inspection parameters. These inspections often consume considerable time of the expert staff; thus, employing these tools could provide considerable benefits. Furthermore, for non-destructive evaluations traditionally performed by inspection vendors, such tools could be valuable in enabling efficient and meaningful site staff oversight of the results that might otherwise be limited by the site staff's limited availability during outages.

Typical or potential benefits that can be obtained from leveraging AI to assist in non-destructive evaluations include the following:

— Faster production of results, allowing utilities more time to address potential issues;
— Reduction in the number of highly qualified personnel that need to be available for the inspection;
— More efficient oversight of vendor activities and results;
— Automation of clerical activities;
— Increase in system reliability, potentially extending the intervals between periodic inspections;
— Increased reliability of inspections.

The non-destructive evaluation activities that can greatly benefit from AI tools include ultrasonic, visual or electromagnetic examinations performed on multiple components of nuclear power plants. These tools will typically employ deep neural networks. Specific examples currently under development, some of which have reached the stage of field trials, include ultrasonic examinations of reactor vessel upper head penetrations (applicable to pressurized water reactor vessels) and dissimilar metal welds [10–13].

2.2.5. Fault diagnosis and predictive maintenance

AI can be used for continuous monitoring of the operating status of nuclear power plant systems and equipment, enabling the rapid detection of potential issues and enhancing equipment health management. Such applications reduce plant downtime caused by system and equipment degradation and failures, hence enhancing plant safety and economy (see also Section 2.2.2).

Existing fault diagnosis processes rely heavily on expert knowledge. There are several data driven and hybrid solutions that can support these processes and enable prompt detection of equipment degradation and failure and automatic identification of causes. Such a diagnosis system learns from operational data and fault cases over time, gradually improving its diagnostic accuracy and forming a more intelligent fault diagnosis system, enabling abnormal situations in nuclear power plants to be detected and responded to earlier.

In addition, through continuous on-line monitoring, health assessment and prediction of equipment lifetime, preventive maintenance can gradually evolve to predictive maintenance, enabling improvement in equipment reliability and maintenance costs. Some utilities are already realizing these benefits through a transition from periodic time based maintenance to condition based maintenance for various plant systems and equipment [14].

2.2.6. Operational and decision support

AI based operational decision support leverages real-time data and prediction models to optimize operational strategies for higher efficiency. With sufficient operational data and previous experience, the system can quickly detect abnormal operating conditions, collect and send all information related to these

abnormalities, and potentially tune the parameters of controllers. This functionality facilitates effective emergency response suggestions for operational personnel. Ultimately, it assists operators in making prompt and accurate decisions, thereby reducing the risk of incidents.

An example of using genetic algorithms to improve fuel cycle performance and to optimize in-core fuel management process is presented in Ref. [15]. Genetic algorithms were used to optimize multiple parameters, including total power peaking factor and fuel cycle length, aiming for economic improvement through longer fuel cycles and uniform power distribution. The study found that the optimized configurations resulted in extended fuel cycles by 118 and 179 days, with slightly altered power peaking factors and higher initial reactivity due to increased fissile material.

An AI based technology that could assist nuclear plant operators to maintain situation awareness and detect faults earlier than would be possible with conventional control room technologies is presented in Ref. [16]. In practice, the process of an operator diagnosing a fault and then adhering to the corresponding paper based procedure to recover from the fault is time consuming and prone to error. Reference [16] describes how AI based solutions can improve plant performance by addressing abnormal events in plants and grids that can lead to transients and present challenges to plant protection systems. Reducing the probability that an incident will lead to an unplanned shutdown has inherent safety and economic benefits. The system was implemented on a full scope simulator, and a human factors assessment was performed to evaluate crew performance; the assessment showed positive results.

In another application, AI algorithms are already in use at several nuclear power plants to support corrective action programmes. AI algorithms can help in evaluating criticality of identified conditions summarized in corrective reports that could impact plant asset performance and in the disposition of the conditions reported in a timely manner. Classification of the severity and disposition of condition reports is a common benefit of AI algorithms; however, the ability to provide aggregate trending of common features across different condition reports is an additional benefit that could also be obtained by AI algorithms. This improves the consistency in annotating the corrective action programme items and reduces the number of hours required for evaluation [17].

Research in power reactor simulators [18] and in the petroleum industry [19] indicates that an advanced alarm system based on expert systems greatly assists the operator by avoiding having an overwhelming number of alarms. Such a system works with alarm groups, guides the operator towards a solution and allows the operator to navigate to the root of the problem. It is important, consistent with international regulations [20], that the operator has a way to navigate through grouped alarms to access the complete list of triggered alarms.

In control rooms, the primary role of AI could be to provide virtual assistance to guide the operator in safely operating the plant. This would result in a collaborative setting, with human decision making augmented by machine (as a virtual assistant), which could significantly reduce manual screening of activities [2]. It may be possible to replace the help manual (the fourth screen) with an AI assistant that uses gestures and NLP techniques.

2.2.7. Human performance optimization tools for operators

Operators play a critical role in the operation of nuclear power plants, and their actions directly impact the safety of the plants. AI technology can be used to improve decision making efficiency and operator response capabilities. Robust algorithms supporting operator actions would first require the adequate interpretation of component faults or performance shifts. Therefore, operator support is ideally a secondary step, after fault diagnosis.

Data driven time series prediction, combined with the operator's current log information and a plant's historical operation data, can provide the operator with auxiliary information on operational risk. AI and extended reality technologies can assist operators in confirming operation steps, identifying the status of equipment, identifying and guiding the log information, and providing objective operational hints.

2.2.8. Sensor state on-line monitoring

AI can be used for the on-line monitoring of sensor state (i.e. on-line estimation of sensor accuracy or response time and diagnosis of the health condition of sensors during plant operation). On-line monitoring can increase the availability of nuclear power plant instrumentation and control systems and detect if one or more sensors degrade or fail. On-line monitoring enables condition based sensor maintenance management and avoids unnecessary calibration, thus reducing maintenance time and the workload and accumulated dose of maintenance personnel.

ANNs have been used to detect, isolate and provide an estimated reading of faulty sensors that are used in monitoring a complex process. Where several measured signals are working to monitor a process, training is performed on an equal number of ANNs. By measuring the error between the real signal and the ANN output, the faulty sensor can be detected. The reading of this faulty sensor can be isolated and an estimated reading sent to the monitoring system. Sensor fault detection is applied to the thermal hydraulic process sensors (core cooling system) of Egypt Training and Research Reactor 2 (ETRR-2) [21].

2.2.9. Nuclear safeguards applications

AI could be applied in various applications related to nuclear safeguards. AI models such as ANNs have been used to determine the plutonium-239 content in spent nuclear fuel based on simulated signatures of the self-indication neutron resonance densitometry [22]. Decision trees, k-nearest neighbours and ANN applications were also tested for ability to identify the percentage of replaced fuel pins in spent nuclear fuel assemblies based on simulated data from different non-destructive assay techniques, namely the partial defect tester, the fork detector and the self-indication neutron resonance densitometry [23]. In addition, ANNs were tested for their capability to identify the presence or absence of individual fuel pins inside spent nuclear fuel assemblies based on simulated measurements of neutron flux and gamma emission rates [24]. The encouraging results indicate the potential for AI in the field of safeguards inspection and the detection of possible diversions in spent nuclear fuel assemblies.

Linear, non-linear and deep learning AI models have been used to predict spent nuclear fuel parameters, such as burnup, initial enrichment, and cooling time, based on simulated signatures such as gamma ray intensities and total Cherenkov light intensities [23–25]. The identification of spent nuclear fuel parameters is a central task for safeguards inspectors to verify the completeness and correctness of operator declarations. Traditionally, this task relies on analysing data from one non-destructive assay instrument at a time; however, the use of AI models has proven advantageous due to their ability to integrate data from different measurement techniques and hence provide more comprehensive results.

AI models have also been considered for safeguards applications in bulk handling facilities. Applying safeguards to such facilities can often be challenging and resource intensive. Emerging technologies in the field of AI and data science hold promise for enhancing the effectiveness and efficiency of nuclear safeguards in this area. ANNs have been tested in the detection of anomalies in the actinide inventories at a plutonium uranium redox extraction reprocessing facility based on a combination of process monitoring data and non-destructive assay measurements [26]. AI models using such input data can complement the traditional safeguards approaches for bulk handling facilities, such as destructive analysis techniques, which are expensive and time consuming and might require on-site analytical laboratories.

AI techniques have also found use in improving the process of characterizing radioactive materials, for example through the combination of measurements (neutron methods, spectrometry, calorimetry) for the purpose of inventorying radioactive materials and reducing uncertainties associated with them [27].

2.2.10. Operator actions, reliability analysis and probabilistic risk assessment

Operator error rate and equipment reliability data in human reliability analysis or probabilistic risk assessment models can be improved using augmented AI to recommend and check for proper implementation of mitigation strategies in the abnormal and emergency operating procedures. The current

methodology for probabilistic risk assessment modelling of operator actions and the actual operator associated underlying performance can be enhanced using AI to confirm proper operator actions and implementation of proceduralized mitigation strategies. AI can play a role in ensuring the operator does not make an error or in identifying that an error occurred so the operations crew can correct the error and ultimately correctly execute mitigation procedures. For a given set of human defined objectives, AI can make predictions, recommendations or decisions influencing real or virtual environments. This will improve the human reliability analysis or probabilistic risk assessment credit given to operator actions and improve actual operator performance. Use of AI technologies can improve operational performance and mitigate operational risk. Improvements will also be reflected in a more accurate probabilistic risk assessment model using AI generated reliability data for human error rates and equipment error rates.

2.2.11. Applications for future plants

Future plants may benefit from many of the same usage patterns as existing plants. However, with future plants, the emphasis is on design processes, and new, sometimes radically different, designs may incorporate new opportunities in a more integrated way. For example, if the full plant is designed with sensors being optimized for automated diagnosis, the diagnosis could be credited when performing risk evaluations throughout the plant's systems. This could lead to reduced overall risk and subsequent design savings.

The new designs could leverage the power of AI and data analytics as one of the tools to enhance equipment reliability, asset management and life cycle planning. New builds could be instrumented according to design requirements and failure mode analyses. The databases could be built in a way in which they are connected to other related databases through data governance principles, and the use of AI would follow a holistic view that can benefit the whole organization, rather than isolated practices.

A challenge for data driven applications for advanced reactors is the lack of operational data. Generating representative operational data for new nuclear power plants is crucial for model training, performance prediction, safety analysis and regulatory compliance. To compensate for the lack of operational data, several techniques can be used. Simulated data from physics based models can emulate the physical properties and behaviours expected in the reactor. Synthetic data generated using techniques like generative adversarial networks can mimic the characteristics of genuine operational data [28]. Virtual data based on similar operational settings, such as data from older plants with similar configurations, can be transformed and normalized to act as a stand-in for the new reactor data or used as the basis for ML techniques.

AI and deep reinforcement learning have been used to optimize the design process for nuclear fuel. Their AI based approach could enhance the configuration of fuel rods in a reactor, extending the life of those rods by about 5% [29].

2.2.12. Other applications

Different AI techniques can be used for intrusion detection in the plant control system, providing support for diagnosis and predictive maintenance systems. Similarly, these techniques can be used to detect intruders in an infrastructure network.

The benefits of AI could extend to reactor simulation. For example, it is possible to train a neural network using data generated from calculations with neutronics codes and then use this neural network as a substitute for the neutronic module in a plant simulator for operator training.

Real-time plant performance data can be highlighted by the AI technology; the operator can then either confirm the proper execution of mitigation strategies or identify that the strategy is not being effective as executed. This will allow the operator to identify ineffective strategies and facilitate the selection of alternative strategies.

AI can provide insight and training strategies to address weaknesses in operator performance. Using information on operator performance in actual and simulator environments, AI can reveal potential

training programme improvements, procedure enhancements and main control room human–machine interface improvements to enhance the performance of operator actions. The dynamic feedback will allow the AI model to continue to be refined.

AI can be trained on historical data to provide more accurate outage and schedule management. More accurate and realistic schedules provide several potential cost savings, including better management of replacement power purchases, improved planning for supplemental outage staffing and better use of outage resources. Additional cost savings may result from more accurate identification of supplemental workforce requirements and the timing of additional resources. For example, one utility leverages NLP and historical outage schedules to predict schedule logic ties and optimize future outage schedules, enabling a reduction in human scheduling errors and reallocation of outage work management resources to higher value tasks [30].

AI can be used to optimize the processes for storage, transportation and disposal of low-level nuclear waste. It might be used to coordinate shipments to ensure that transportation and shipping radiation levels are within allowable limits.

Industry oversight organizations are leveraging AI algorithms to improve accuracy in predicting plant performance using industry data. For example, the World Association of Nuclear Operators Atlanta Centre has developed models to determine nuclear power plant performance and estimate the assessment score of a nuclear power plant using a variety of data, such as performance indicators, operating experience reports, assessments, regulatory information and performance trends [31].

A model for predicting critical on-line parameters based on ingesting large amounts of historical plant data was developed to help optimize the fuel reload designs of boiling water reactors. The model uses neural networks to correct the error between off-line predictions for thermal limits and eigenvalues (reactivity) and the actual observed on-line values. Accurate predictions enable significant fuel savings, minimize power derating, and minimize earlier-than-planned coastdowns, while securing operation within the safety limits defined in the reactor licensing basis [32, 33].

2.3. BUSINESS VALUE

The use of AI in utilities provides value and benefits that vary case by case. It can be challenging to quantify the benefits of early AI adoption despite clearly identifiable performance gains in specific tasks where AI is applied. The local performance gains may fail to translate to tangible benefits for multiple reasons. It might be that the integration of AI with current processes causes friction, or the AI may require full supervision that diminishes the benefits. Also, due to lack of experience with AI, people may discount or fail to appreciate the various second order benefits of AI.

Similar to with other tools and technologies, the value of AI applications normally starts small. Once value is successfully proven, the scale-up application would provide business values that are large enough to justify the cost of implementation.

Some noted value from industry during field trials of AI technology include the following [12, 14, 30, 34]:

— Providing results (even if preliminary) faster than is otherwise possible, allowing more time to prepare for or address potential issues;
— Providing early detection of plant issues, resulting in increased power production;
— Providing valuable help in managing workforce challenges such as availability and maintaining sufficient proficiency for seldomly performed activities;
— Reducing the number of highly qualified individuals required on the site, leading to decreased costs and helping with workforce availability.

Other potential benefits include:

— Increased operational efficiency: Potential benefits in this category include allowing efficient oversight and faster issue identification.
— Increased operational reliability: AI tools can help with known human factors issues, such as those associated with fatigue and distraction.
— Enhanced health and safety: AI can be used to enhance health and safety for employees and contractors by learning from previous events and providing insights from historical incident reports. These insights can be extracted through numerical AI or through LLM techniques. These algorithms can highlight important trends and identify hidden trends that are hard for humans to capture due to the large number of data points. As with other AI tools, human oversight is needed to review the AI model outputs before implementing findings. Other information, such as observations, if they are collected, can be used to identify trends as a precursor to health and safety incidents.
— Increased collaboration capability: Visualization of AI results and recommendations can support efficient communication between different members of the team to accelerate decision making.
— Enhanced financial and operational performance: Financial and operational performance generally trend together. AI could help utilities move from a labour-centric preventive maintenance strategy to a technology driven predictive maintenance strategy. AI can also be leveraged to minimize the repetitive activities organizations do on a daily basis. Enhancements in supply chain, IT services, and financial and performance reporting are a few examples in which AI and automation techniques would save a lot of time. The application of robotics with integrated AI, when implemented correctly, would enhance operational and financial performance. Performing routine visual inspections, performing work in high radiation dose areas to minimize staff exposure time (this ties back to the health and safety benefits as well), and deploying robots to capture picture, video and audio data as needed are just a few examples of how the application of AI can enhance operational and financial performance.

2.4. LESSONS LEARNED

Although experience with deployed AI solutions is still limited, some common success factors can be identified and lessons learned; these factors characterize many of the application examples.

There is currently quite high interest in AI and the new opportunities offered by this technology. Experience has shown that it is important to facilitate sufficient communication among stakeholders to explicate the expected benefits and limitations of an AI application and to focus on possible concerns relevant to the application.

The role of subject matter expertise and domain knowledge in conjunction with AI expertise is emphasized as a key success factor in several applications. Experience gained from engineering systems shows that AI methods are made more robust and reliable if domain knowledge (i.e. human expertise) is embedded in the algorithm. Many engineering problems in the light water reactor industry are high value but are resistant to solution by traditional engineering methods (e.g. Ref. [5]). AI coupled with subject matter expertise can provide a solution to these otherwise intractable problems. The involvement of subject matter experts has proven decisive in developing successful models with high reliability. This involvement is important both in the preparation of high quality training data during model development and in the evaluation of model performance in specific context of the application.

Experience has shown the importance of building the models and applications incrementally and providing the users opportunity to build trust in the models. Reporting recent results at industry events and showing applicability in field trials provide utility personnel, inspection vendors and regulator representatives the opportunity to witness and monitor the development. Potential issues can be raised and resolved prior to adoption. While focus is often on the ‘technical readiness level’ of the proposed solution, it is also important to consider the ‘human readiness level’ (i.e. to make sure that the operators

using the system are comfortable with it and maintain effective coordination with the AI solution). As the models are being qualified for field application, the role of industry guidance increases. Especially with a new technology, best practices and industry level guidance, such as EPRI's materials reliability programme documents [35] and European Network for Inspection and Qualification practices [36], help individual licensees adopt the new technologies in a safe and controlled way.

3. LIFE CYCLE MANAGEMENT OF ARTIFICIAL INTELLIGENCE SYSTEMS

3.1. INTRODUCTION

Current and anticipated uses of AI in the nuclear power industry extend beyond the targeted data analysis of defined scope datasets. The latter cases have a short shelf life and likely rely on a small number of stakeholders. An approach is hence needed to address the following aspects of the development of AI applications:

— The application development may need to integrate input from interdisciplinary teams of technical specialists, AI developers and conventional software developers.
— The application may result in a source code base of moderate or large size.
— The outcome of the project may be needed by many stakeholders.
— The developed application may undergo regulatory scrutiny, as it may produce results of safety or operational significance.
— The developed application will likely have a long shelf life and require maintenance, version control (e.g. see IAEA guidance on configuration management [37]) and documentation.

Life cycle framing enables AI applications to be envisioned in a scalable manner. Furthermore, the life cycle approach ensures that AI applications are designed to meet end user needs, are developed according to design intent, can be kept operable during their intended lifetime and consider regulatory needs during the initiation phase [37]. Life cycle framing is consistent with systems engineering practices typically adopted in nuclear applications, which are outlined in accepted industry standards [38, 39].

The remainder of this section aims to identify topics that may warrant careful consideration as part of the AI life cycle and to explain their significance. The section also provides a high level overview of some approaches and tools that address the identified topics, to serve as an illustration. The following is not intended as a standard specification or AI development methodology but as a set of principles based on the prevalent state of knowledge within the nuclear power industry at the time of writing.

3.2. DESIGN

3.2.1. Artificial intelligence problem statement

As part of the problem statement, it is often necessary to convert a problem or task into a set of AI based requirements. In some cases, this is a simple process that replicates a human decision (e.g. deciding on whether a crack exists in an image). However, in many scenarios, problem statement formulation can involve a complex process of interpretation of domain-specific requirements and translation into an AI problem statement. For example, ongoing research efforts are examining means of introducing automation into compliance verification in plant inspection activities. Inspection procedures are

extensively prescriptive and provide detailed requirements. Converting such requirements into a set of well defined data driven decisions and the associated metrics requires careful review and analysis of the involvement of a subject matter expert who has experience in both the inspection procedure and AI.

Applying AI approaches requires reframing the problem into the desired outcomes of the AI solution, based on the need for:

— Classification of labelled data (e.g. fault classification);
— Prediction of a response variable as a function of predictor variables (e.g. remaining useful life of a component as a function of monitored parameters);
— Insights on any inherent clusters or groupings in the data.

3.2.2. Technical basis

A technical basis for the use of a particular AI methodology to address a given problem is a prerequisite to implementation within an AI application. The nature of a given AI problem statement determines what prior technical basis may exist to allow the use of AI tools.

Some AI problem statements are generic and mature, having a strong technical basis with widely available generic solutions. Current development examples primarily consist of confirming problem–solution fit and integrating the software elements into a working application. Some examples of AI problem statements with generic solutions are:

— Image recognition: Pretrained deep neural networks such as AlexNet are available for these applications.
— NLP: Tools such as LLMs can be used to address the majority of applications involving the analysis, processing and generation of textual information.

However, some AI problem statements are application specific and lack a generic solution. A specific AI methodology needs to be devised to address these problems. The development of the technical basis then becomes part of the development process. Some examples of AI problem statements requiring the development of a technical basis are:

— Developing data driven surrogate models that can be used instead of computationally expensive codes (e.g. as a substitute for neutronics or thermal hydraulic models);
— Fault identification or classification and remaining useful life estimation of nuclear power plant equipment;
— Fault tolerant approach to substitute faulty sensor readings using ANNs in a nuclear power plant.

3.2.3. Model selection

Selecting an appropriate AI modelling approach is a critical step in the solution design process. Poor model selection directly affects the quality and accuracy of the solution. Several non-technical factors may affect model selection. Model complexity could determine the data required; the in-house algorithm development needs; and the computational resources needed for training, testing and running the model. Therefore, factors such as available training data and computational resources could dictate the choice of the model. The process for selecting a suitable model can be broadly classified into two categories (Sections 3.2.3.1 and 3.2.3.2).

3.2.3.1. Application-driven model selection

Application-driven model selection considers the aspects associated with the problem to be solved by the model. The purpose and objective of applying the model are the primary application driven factors

affecting model selection. For instance, the purpose of a model could be to improve the efficiency of maintenance practices at a nuclear power plant, and the corresponding objective of the model could be to detect anomalies indicating component degradation, which could inform predictive maintenance. Some models are known to be more suitable for anomaly detection applications than others and could be chosen for this specific application. The second application driven factor in model selection could be the amount and type of data associated with the application. Several aspects of a model, such as training, testing and application, are affected by the data used. Some of the application driven attributes associated with data that can inform model selection are quantity, the uncertainty distribution, the historical versus real-time nature, the data speed handled by the model, and heterogeneity.

3.2.3.2. Performance based model selection

Performance based approaches rely on developing metrics to quantify the suitability of model predictions. Some commonly used metrics that are applied for statistical model selection are Akaike's information criterion [40], the Bayesian information criterion [41], the minimum distance length [42] and *k*-fold cross-validation [43]. Several quantitative and qualitative metrics that provide measures of model performance are discussed in Section 3.3 and can be considered for model selection.

3.2.4. Goal alignment

Goal alignment refers to the degree to which the AI's training objective matches the actual task it is intended to perform. In the context of nuclear power plants, goal alignment is crucial. The AI model should be trained with a clear understanding of the specific objectives and requirements of the plant, and the operating objectives should be translated into an appropriate training method. Misalignment can lead to incorrect or unsafe decisions or unexpected behaviour.

Achieving goal alignment involves careful selection of training data, loss functions, objective functions and performance metrics to reflect the task's real world goals and safety considerations. Effort should be made to ensure that the objective of the designed system and the objective of the training are in alignment.

3.2.5. Static and dynamic models

AI models can be categorized as static or dynamic. Static models are created and frozen for direct use. Static models reduce the risk and impact of model changes. The benefit of using static models is that the qualification and validation effort is front loaded, likely only needed during initial development. Unless a new version of the model is used, this qualification remains valid.

Dynamic models are used when there is an expectancy of process change to the point where the AI model cannot adapt and needs to retrain and retune. For example, the use of AI in system control is challenged by the ageing of the controlled process, and the model needs to retrain to accommodate emergent conditions. Dynamic models are more challenging to use in nuclear applications because they change in time and are therefore onerous to qualify. Changes can affect the structure or architecture of the model. They may also alter model parameters. From a qualification perspective, dynamic models require the development of continuous model validation approaches or the adoption of an incremental retraining and validation approach during operation.

3.2.6. Risk reduction

Like any model, AI models are expected to fail. A model that achieves 100% accuracy is not practically feasible and usually indicates a problem with the testing dataset (e.g. either the dataset is too limited and not representative of all system states or it is biased to mirror the training and validation datasets). Some methods exist to reduce the risk of failure in specific applications. One relies on biasing

the validation metrics toward the undesired output. For example, if fire detection is used by a camera video feed, a false positive (fire is flagged when there is no fire) is much more tolerated than a false negative (a fire occurs and is not detected). Another approach relies on biasing the loss function or threshold used in the training and classification process. It is possible to add a penalty for a specific false classification or prediction to convince the model that it is more important to detect a certain class than another. However, this improvement of one class or prediction often results in the degradation of others.

3.2.7. Human factors considerations

AI applications are likely to involve human users during their life cycle. Therefore, human factors need to be considered in the design throughout their life cycle. For instance, if AI is used for an operator support system or an automatic system, the interaction between operators and the AI based system should be carefully designed.

Systems may fail when the design does not consider the interaction between operators and system. Many issues related to human performance in the interaction, especially with the automation, have been reported, including the following [44]:

— Out-of-the-loop unfamiliarity: a reduced operator ability to detect automation failure and to resume manual control [45]. This issue often happens in a highly automated (or autonomous) system in which the operator is removed from directly performing the control.
— Clumsy automation: a situation where the automation makes easy tasks easier and hard tasks harder. This can occur when easy tasks are automated and hard tasks are left to the operator.
— Mode errors: a type of mistake in which the operator takes an action when the automation is in the wrong mode. Such an error can occur based on the operator's incorrect interpretation of the system state or based on the expectation that the system would be in a particular mode due to operator instruction.
— Inappropriate trust: instances including those of misuse and disuse by the operator. Misuse means the operator over-relies on the automation and therefore fails to notice that the automation has failed and to intervene. Disuse means that the operator does not use the automation because they have insufficient trust in its reliability.
— Inadequate training and skill loss: situations in which the automation may eliminate opportunities for the operator to maintain skills by doing the job.

More issues and explanations are presented in Ref. [44].

One approach to address these issues is human centred AI. Human centred AI attempts to build on user observation, operator engagement, usability testing, iterative modification and continuous evaluation of human performance [46]. Human centred AI aims to augment or enhance human performance. To be human centred, an AI application should address fairness, accountability, interpretability and transparency [47]. To do this, Xu suggested a framework of human centred AI that consists of three components: ethically aligned design, technical enhancement and human factors design [48]. The ethically aligned design means that AI solutions should avoid discrimination, maintain fairness and justice, and not replace humans. Technical enhancement means that AI technology should be enhanced to reflect the depth characterized by human intelligence. Lastly, human factors design should be adopted to ensure that AI solutions are explainable by, comprehensible to, useful to and usable by human operators.

Cooperative AI (another concept for design of the interaction between operators and system) places more emphasis on the collaboration between human and AI, while human centred AI is more concerned with improving human performance. AI research has focused on improving the individual intelligence of agents and algorithms. Cooperative AI, however, focuses on improving social intelligence, which is the ability of groups to effectively cooperate to solve the problems they face [49]. It covers a broad range of

cooperation: AI–AI, human–AI, and human–human with AI support to facilitate the cooperation. Dafoe et al. suggested four key capabilities for cooperation [50]:

— Understanding: the capacity to consider the outcomes of one's actions and anticipate others' behaviours, beliefs and preferences;
— Communication: the skill of clearly and reliably conveying information to others, helping to interpret actions, intentions and preferences;
— Commitment: the ability to make dependable promises when necessary to support cooperation;
— Norms and institutions: the shared social frameworks — such as common beliefs or rules — that strengthen understanding, communication and commitment.

3.2.7.1. User acceptance

Introducing AI into the nuclear power industry requires careful consideration to address concerns and build trust among practitioners and users. Strategies AI practitioners may employ to make users more comfortable with AI in the nuclear power industry include:

— Transparency and explainability: providing clear explanations of how AI algorithms work and their decision making processes. Transparency aids users to understand the rationale behind AI recommendations. Using interpretable models and avoiding 'black box' approaches may make it easier for users to trust AI outcomes. AI, like any new system, requires new user interfaces to ensure key information is visualized in a manner that users can readily understand.
— Robust validation and testing: conducting rigorous validation and testing processes to enhance the reliability and accuracy of AI models. This can help demonstrate the effectiveness of AI systems and address concerns about their performance. The integrated system validation approach entails user testing of a technology as part of overall integrations such that a variety of use contexts are explored before deployment [51].
— Collaborative development: involving experts and stakeholders throughout the development process. Such collaboration helps the AI solutions align to end user needs, making practitioners more comfortable with the technology.
— Education and training: providing comprehensive training programmes to educate users about AI technology, its capabilities and its limitations. Such training empowers practitioners to work effectively with AI and reduces anxiety associated with the unknown.
— Risk assessment and mitigation: conducting thorough risk assessments to identify potential challenges and vulnerabilities associated with AI implementation and developing strategies to mitigate risks and support the safe and secure use of AI technologies in nuclear applications. These processes may help increase user acceptance. Mitigation includes adequate training of users to fall back to manual operations in the event of degraded AI functionality. AI may introduce new failure modes beyond those found in conventional nuclear power plants. It may be necessary to identify new critical safety functions and important human actions [52] in support of the risk assessment.
— Incremental implementation: introducing AI technologies gradually and in controlled environments. Such an approach allows practitioners to observe and understand the impact of AI on specific tasks, making the adoption process more manageable and potentially less intimidating for the user. For new systems, it may not be possible to phase in new features before deployment. In such cases, it is useful to have users thoroughly test increasing AI functionality such as increasing levels of automation [53].
— User involvement in decision making: involving end users throughout the decision making process of AI technologies, soliciting feedback, addressing concerns and incorporating user input. This undertaking may enhance the acceptance of AI.

— Incorporation of ethical considerations: emphasizing the ethical use of AI and clearly communicating the ethical principles guiding AI development and deployment. These processes may help reassure users that AI technologies are adhering to ethical standards.
— Continuous monitoring and improvement: implementing robust monitoring mechanisms to track the performance of AI systems over time. Regularly updating and improving AI models based on feedback demonstrates to the user a commitment to ongoing improvement and enhancement.

Addressing these considerations will typically require applying sound human factors engineering principles and practices throughout the AI system's life cycle (i.e. from conception through retirement). Human factors engineering programmes based in systems engineering frameworks (e.g. see the models described in Ref. [54] and IAEA Nuclear Energy Series No. NR-T-2.14 [38]) are designed to provide integrated, interdisciplinary support for addressing human performance considerations throughout the system life cycle and thereby can facilitate the successful integration of AI technologies.

3.2.8. Designing for the cloud

The main differences when designing AI applications for the cloud stem from the consideration that data and models would be maintained off-premises. The advantage of a cloud based design lies in the fact that the infrastructure is maintained by third parties and can scale to meet the need: the memory and computing power allocated in a cloud based software solution increase or decrease with traffic, which is useful for managing variable workloads and guaranteeing high availability. Additional specificities of cloud hosting include accessibility from anywhere and security managed by cloud providers. The cost effectiveness of the cloud is frequently highlighted as a benefit; however, this can differ significantly depending on the specific details of the application, making it important to conduct assessments tailored to each case. Some of the considerations for cloud based AI implementation for the nuclear power industry (e.g. infrastructure requirements, costs, risk assessment, reliability) are discussed in Ref. [55].

Currently, cloud software is generally based on containers, which are isolated, portable environments that can be managed by orchestration platforms such as Kubernetes. Containers are used by software developers to run their code locally in an environment that contains the same content (operating systems, libraries, settings) as the one where their code will be executed in the application. Several software elements may have different requirements and are often run within different containers in a single application.

These technologies are useful to data scientists in the nuclear power industry because they enable better reproducibility. The cloud infrastructure relies more on code and less on manual procedures to limit human mistakes: infrastructure is considered as code. The infrastructure design itself is part of the design of the cloud based AI application but could be adjusted more easily during the development phase if necessary.

However, the benefits associated with cloud computing as a means of enabling the use of AI applications within a nuclear facility should be evaluated against cybersecurity considerations. While appropriate cybersecurity architectures may serve to reduce vulnerabilities, organizations would still need to navigate transitioning from a status quo in which important computer systems are housed in local (and sometimes isolated) networks to one in which cloud infrastructure would be integral. Moreover, the role in which AI is used may also influence such decisions; for example, the monitoring of systems via the use of cloud based AI applications may be more acceptable than the operational control of those same systems via cloud infrastructure.

3.2.9. Robustness and resilience

AI models face significant deployment challenges in the face of adversarial threats that actively seek to undermine their performance. In past decades, physical models were closely guarded secrets, but with open source learning and the wide availability of effective AI algorithms, adversaries can mimic

physical models of critical infrastructure by simply observing and relaying the data passing through the system to train their own models [33]. Furthermore, adversaries can detect and exploit weaknesses in the model by exploiting various correlations among the data to send false signals to systems and misguide them. Specifically, with the generative adversarial network framework, it is possible to generate synthetic data closely approximating the underlying distribution of true data, and this task is considerably simpler for an adversary with domain knowledge, as is the case with state sponsored and advanced persistent threats. For example, an adversary may alter the steam flow rate in a nuclear reactor and falsify the sensor measurements with synthetic data to bypass detectors and make it appear as if the system is under normal operation.

Significant work has been done in recent years to improve the robustness of AI in the face of adversarial acts (e.g. robust encoding, sparse representations, preprocessing of input data to remove adversarial inputs) [56, 57]. However, parallel advancements on the adversarial side include membership inference attacks [58], which may be constructed to decipher the training data used to create the model. Poorly trained models that overfit the training data are especially prone to such attacks.

Achieving robust and resilient intelligence requires AI agents to exhibit common sense through learning causation, intuitive physics and the ability to reason. Common sense is a difficult concept to quantify, let alone express mathematically, and may be loosely defined as the aspects of intelligence that most humans take for granted or conclude subconsciously without much thought. Currently, AI is very far from this goal, and even large systems with hundreds of billions of parameters [59] do not appear to show signs of artificial general intelligence or common sense.

3.2.10. Cybersecurity principles

Despite the numerous potential applications of AI for nuclear power, there are computer security concerns that have to be addressed. Conversely, AI also has the potential to improve the computer security of nuclear power facilities.

The overall objective of computer security for nuclear facilities, as defined in IAEA Nuclear Security Series No. 17-T (Rev. 1) [60], is to ensure that cyber-enabled adversaries do not compromise nuclear safety, nuclear security, and nuclear material accounting and control functions through an attack on a supporting computer based system.

The use of AI can result in new vulnerabilities in computer based systems at nuclear facilities, including those performing functions related to nuclear safety, nuclear security. and nuclear material accounting and control. A malicious actor could seek to exploit such vulnerabilities to compromise the functions performed by AI, which could lead or contribute to a nuclear security event. An incomplete summary of AI specific threats and vulnerabilities is presented in the following list. A knowledge base of known adversarial tactics and techniques related to AI systems can be found in the Mitre ATLAS (Adversarial Threat Landscape for Artificial-Intelligence Systems) framework [61]. In computer security, protection relates to ensuring the confidentiality, integrity and availability of computer based systems and information.

— Confidentiality: the property of information not being made available or disclosed to unauthorized individuals, entities or processes [62]. In some cases, artefacts that are associated with AI could contain sensitive information (section 1.1 of Ref. [63]). Such artefacts include AI models and training data, for example. An adversary may attempt to steal this information in various ways, including through ‘conventional’ cyber-attacks that seek to exfiltrate data and through attacks wherein the adversary performs targeted queries of an AI model to gain insights about the (potentially sensitive) data that was used to train it.
— Integrity: the property of accuracy and completeness of information [62]. Ensuring the integrity of AI is of the utmost importance: critical decisions that are related to nuclear security and safety could be informed by predictions made by AI models. Several forms of attack that are specific to AI target integrity. Attacks have been demonstrated that allow ‘backdoors’ to be introduced into AI models

during model training; these backdoors elicit prescribed output from a model when nefarious input is provided. This is a form of model poisoning. A related form of attack is 'adversarial examples', wherein an adversary provides inputs to a model that are manipulated — imperceptibly to the human eye in images, for example — to cause a model to make a targeted or untargeted misprediction. These attacks have been demonstrated, for example, for AI that can be applied to security surveillance systems [64] and computer antivirus software [65].

— Availability: the property of a system being accessible and usable upon demand by an authorized entity [62]. Like non-AI systems, systems that make use of AI technology are susceptible to attacks that can compromise their availability. This often takes the form of 'denial of service attacks', wherein voluminous service requests are made, causing a target to become overloaded and unavailable for legitimate use. In the case of systems that incorporate AI, this type of attack could manifest as large volumes of requests to classify input data or inputs that are intentionally crafted to cause excessive resource use, causing the systems that host the classifier to exhaust computational resources. Because many AI applications require specialized computing hardware, such as a graphics processing unit, it may not be possible to readily scale resources to mitigate these kinds of attack.

In the same way that nuclear facilities are considering the use of AI to make their operations more effective and efficient, adversaries are doing the same. This nefarious use of AI can take many forms and may include the use of AI to support the rapid development of malware capabilities, learning the behaviour of a target facility to enable more effective engagement, the generation of more realistic deep fakes to subvert surveillance systems and generate uncertainty, and the automation of malicious activities. The result could be an adversary that is more effective and that can operate at larger scales with lower costs. The consequence for nuclear facilities and the nuclear sector is a potentially significant shift in the nature of the risk that needs to be understood and addressed.

Conversely, the security community has invested in uses of AI to improve computer security. For example, AI models have been used extensively to support the detection and classification of malicious behaviour that manifests in network or host data. These models are trained on datasets that represent either normal behaviour (for anomaly detection) or malicious behaviour (for malicious behaviour classification). A significant challenge that the nuclear sector faces is a shortage of skilled computer security professionals. To help alleviate this problem, AI is being applied to automate security tasks.

Despite the new computer security challenges that AI introduces, there is a significant body of research and guidance that can be used to manage the risk. This research and guidance can be broadly categorized into approaches to improve the security and robustness of AI itself and techniques that seek to ensure the computer security of AI in the context of a system. For example, there are techniques that can be applied to training models that are used for AI in an adversarial setting and that have the purpose of making AI robust against adversarial attacks [66]. Federated learning is an approach to enabling privacy-preserving ML, which could be applied in this context to reduce the risk of disclosing sensitive information (i.e. to protect confidentiality) [67].

Many computer security best practices and guidelines can be used to protect AI systems. For example, Ref. [60] advocates the use of a graded approach to computer security. Using this approach implies that AI systems that support higher criticality functions should attract stronger security requirements; this could entail limiting the use of AI due to the computer security risk and the consequences associated with compromise, as well as applying stronger security controls to systems as part of a facility's defensive computer security architecture (sections 4.67–4.82 of Ref. [60]). This could, for example, involve implementing strong computer security measures [68] to systems that execute AI models, which would support more critical facility functions, with the objective of preserving the correct performance of those functions. These measures should be technical, organizational and physical in nature.

In summary, there are computer security considerations that need to be understood in relation to the use of AI at nuclear power facilities; these can include new vulnerabilities and adversarial uses of the technology that change the nature of nuclear security risk assessments. There are, additionally, uses of AI to support computer security. There are significant volumes of best practice, standards and guidance from

the IAEA and other organizations that can be applied. These are important and should be applied using a graded approach. There are open computer security questions about the use of AI for nuclear power facilities, including fundamental research questions about the vulnerabilities that AI applications introduce and how they can be mitigated, along with questions regarding how computer security assessments should be made for regulatory approval of systems that use these technologies.

3.3. DEVELOPMENT

3.3.1. Software quality assurance

The goal of quality assurance (QA) standards and programmes for software is to set up systematic and controlled processes for aspects of software that are likely to affect its performance. Such areas are typically the procurement, development and use of software and digital systems.

Software in nuclear safety, safety related or safety significant applications is typically developed according to defined software QA standards, as part of overarching QA programmes. AI applications with safety significance are likely to fall within the scope of such software QA standards. Traditional software QA standards may be insufficient to address the needs of AI systems, which are application specific and based on considerations such as whether the AI system relies on a static or a dynamic model. However, some guidance exists for possible implementation of software QA programmes for AI systems [69].

3.3.2. Qualification

Software for nuclear applications requires qualification to verify that it is fit for use. Qualification is an integral part of the development life cycle and is intended to assess the technical basis, verify the implementation and validate the performance of software against 'real world' data. Software qualification requires a systematic and documented approach, which is typically defined a priori.

If an AI application is anticipated to have operational or safety significance, software qualification may be a good process for reducing the likelihood of failure. Pursuing such an effort is expected to provide clarity to contributors regarding performance expectations and would lead to enhanced end user satisfaction.

3.3.3. Repeatability

AI models are repeatable if, under the same training set-up, the same result is reached under the same evaluation criteria. Some AI approaches (e.g. deep learning) incorporate different sources of randomness, such as initialization of random values in neural network weights and random selection of which data are used for training and which for testing and hyperparameter tuning. Introducing different levels of randomness during training is essential for building robust models but can be a problem when trying to reproduce the results achieved.

Repeatability considerations may also independently apply at the inference stage. During use, a repeatable model is expected to produce the same output given an invariant input. Repeatability within training and during use are independent attributes of the model. The application requirements will dictate which aspects of repeatability are desired.

3.3.4. Open source versus proprietary software

AI solution development involves several choices related to the use of proprietary or open source tools. Such choices may relate to:

— Programming languages;
— Software libraries and toolboxes;

— Training data (a more detailed discussion on open source considerations for data can be found in Section 4.2.4);
— The AI algorithm.

For each of those items, there are open source and proprietary alternatives. At a high level, the advantages and disadvantages of these alternatives can be summarized as follows:

— The use of open source libraries is low cost and enables access to a vast amount of predeveloped software. However, open source libraries can create version control challenges and lead to uncertainty around verifiability and quality.
— The use of proprietary tools provides QA traceability, which is preferred (and in many cases is a requirement) in nuclear applications. However, proprietary software often involves high costs and a small developer community.

3.3.5. Internal versus external development

Nuclear organizations may choose to develop software in-house or engage external partners to conduct software development activities on their behalf. Software proponents may have no prior AI software development experience or no software development experience at all. In such cases, the software proponent may involve a partner to provide the required expertise. In such cases, the software proponent should ensure that:

— There is a clearly stated AI problem statement or there is a preliminary scoping phase planned to obtain one.
— The availability and characteristics of the training data are known a priori (covered in Section 4 in further detail), such as:
  - Quantity and frequency of availability;
  - Format and mode of access;
  - Labelling;
  - Adequate coverage of the variance in the possible sample space to be usable;
  - Category (e.g. artificial, laboratory, field).
— Performance requirements are clearly stated.
— Any applicable software QA requirements can be met, whether under the proponent's QA programme or by a member of the supply chain.

3.3.6. Development process

Agile development is an iterative and flexible approach that emphasizes collaboration, customer feedback and incremental progress. Regular product releases are encouraged so as to deliver value to users sooner and get their feedback on new features. Adjustments are possible throughout the project, and the priorities can change. In contrast, waterfall software development is a more linear approach where the project is made up of phases, each starting when the previous one is completed. Customer input is mainly gathered at the beginning of the project, and the product is delivered all at once after a lengthy development phase, with possible discrepancies between what is delivered in the end and the true needs of the users (which may differ from the specifications written at the beginning of a project).

When developing an AI solution, the agile development process gives access to the end user's reactions to the AI results, making it easier to know whether further tuning is required or if additional features are needed to better understand the results. However, some time-consuming algorithm developments may require longer development cycles before being presented to avoid disappointment for the user.

3.3.7. Training algorithms and metrics

Training metrics for model development and optimization, as well as model performance assessment, are essential in the development of AI models. Performance metrics help data scientists evaluate and improve their models during the training process. As part of developing a supervised or unsupervised model, labelled or unlabelled data are used to build and train the model and then assess the prediction performance. The difference between the actual and predicted values is defined as the loss and represents how far a model's prediction differentiates from its true value. This process of optimizing or reducing the loss is a process that occurs through iteratively making model predictions and using a variety of metrics to assess the model performance and to continue learning until the model parameters result in overall model stability and model convergence.

Some common training approaches used in this learning process include:

— Loss function: This approach measures how well the model is performing by quantifying the error between predicted and actual values. Lower values indicate better performance. The function selection depends on the type of model being trained. For example, classification and regression models may rely on different types of loss function.
— Training loss and validation loss: This process tracks the loss on the training data and a separate validation dataset. It helps in identifying overfitting (when validation loss increases while training loss decreases) and aids in model selection.
— Learning rate curves: Use of these curves shows how the learning rate impacts the loss function, which helps in choosing an appropriate learning rate for training.

AI performance metrics that can be used to evaluate a model during development and testing are generally divided into two categories: (a) regression metrics, which are used for models that make a numerical value prediction, and (b) classification metrics, which are used for models that predict discrete values, such as if something belongs to a category. Regression metrics include:

— Mean squared error: This metric represents the average of the squared differences between the actual and predicted values and can be impacted by large outliers in the data due to the squaring function.
— Mean absolute error: This metric represents the average of the difference between the actual and predicted values and can be used to illustrate how much a model deviates from observed values.
— R squared: This metric measures the variance of a model compared to the actual data and can be used to show if the model is capturing the variance and if the model is fitting to the observed data.

Classification metrics include:

— Accuracy: Accuracy is the ratio of correctly predicted instances to total instances. It provides a general measure of model performance but can be misleading when dealing with imbalanced datasets.
— False positives (type I error): False positives occur when the model incorrectly predicts a positive outcome when it should have been negative.
— False negatives (type II error): False negatives occur when the model incorrectly predicts a negative outcome when it should have been positive.
— Recall (sensitivity): Recall measures the ability of the model to correctly identify positive instances out of all actual positive instances. It is particularly important when the cost of missing a positive instance (i.e. giving a false negative) is high.
— Precision: Precision measures the ability of the model to correctly identify positive instances out of all instances it predicted as positive. It is essential when there is a high cost associated with false positives.

— F1 score: This score is the harmonic mean of precision and recall. It provides a balanced measure of both false positives and false negatives and is useful when balancing the trade-off between precision and recall.

Understanding and choosing the right evaluation metrics is critical in AI model development. The choice of metrics should align with the specific goals, the use case and the potential consequences of the model's predictions; however, it is important to consider the trade-offs between accuracy, false positives and false negatives.

3.3.8. Verification and validation

3.3.8.1. Artificial intelligence model verification and validation

AI models are typically developed using training and validation datasets. The validation set consists of data not encountered by the model during the training phase and represents how well the model performs on new data. Typically, this validation set is sampled from the data available and should have a similar distribution to the underlying dataset. While this process provides an indication of model performance with novel data, it may not generalize well to low probability events or occurrences that were not reflected within the training and validation datasets.

The nuclear power industry has performed significant research on verification, validation and uncertainty quantification of computer algorithms. These traditional techniques may or may not be robust enough to enable the validation of AI enabled systems. The scientific and industrial community is researching verification and validation methodologies for AI-driven systems and seeking to establish the degree of certainty required for anticipated applications.

Validation techniques can also indicate limits of operability for the AI system. For example, if validation indicates the AI system is not robust over a certain range of operation, restrictions can be placed on the system to inhibit operation or provide an indication to operators to take control.

3.3.8.2. Artificial intelligence integrated system verification and validation

As noted in Sections 1 and 2, AI enabled systems in nuclear power applications will include, in addition to computer software and hardware, human users and, in many cases, interfaces to plant instrumentation and control. An AI enabled system, when considered as a system of software, hardware and humans operating in the context of other nuclear power plant systems (i.e. a system of systems), will require validation efforts beyond the validation of the AI models. Such validation efforts, often referred to as ‘integrated system validation’, involve performance based testing using representative system users in a realistic operational setting [54, 70]. Integrated system validation testing can be useful in determining whether human performance considerations, such as those identified in Section 3.2.7, have been effectively addressed such that performance of the integrated system that encompasses the AI reliably achieves the system’s objectives. Given the nature of human performance considerations associated with AI (e.g. transparency and explainability of the AI, calibration of user trust in the AI) and given that AI enabled systems may be used in contexts where alternative or competing sources of information to users will be available, it is important to understand, prior to deployment, whether and how AI enabled systems are actually used in their operational contexts.

3.4. DEPLOYMENT

3.4.1. Deployment environment

The deployment environments for AI models and their associated preprocessing or postprocessing scripts can differ substantially depending on factors like the hosting environment, computing hardware and platform.

The hosting or computing environment may refer to the public or private cloud, on-premises computing, or 'edge' computing. The computing hardware may encompass devices that process on central processing units, graphics processing units, AI accelerators or neural processing units, field programmable gate arrays, or application specific integrated circuits. Additionally, the hardware may be fixed or mobile, such as on a drone or a quadrupedal robot. Deployment platforms or runtime environments may be 'bare metal', virtual machines or containers, with or without orchestration systems. Many combinations of hosting environments, physical hardware and platforms are possible, and these lists are not exhaustive.

In certain applications, it may be beneficial to train a model in one setting and then deploy it in another. For example, the model might be trained with cloud resources and deployed at the 'edge'. Further, the model may be optimized between the training and deployment stages using knowledge distillation, quantization, layer fusion or other model specific techniques due to deployment hardware limitations or for latency or throughput optimizations.

The process of selecting the most suitable deployment environment for a specific application might necessitate a thorough understanding of different restrictions governed by internal policies and external regulations. A variety of questions might need to be addressed, which may pertain to:

— Data or model residency requirements;
— Whether the data or model is subject to export control regulations;
— Data or model governance;
— Network restrictions, including potential requirements for air gaps.

The selection of a deployment option might be influenced by certain requirements or the necessity to optimize various factors. These requirements or factors may include:

— Availability or a high availability requirement;
— Latency;
— Utilization;
— Throughput;
— Hardware or subscription costs;
— Power consumption;
— Footprint or weight;
— Stability with respect to data corruption.

These factors may impact the attainable performance, economics and feasibility of a particular AI application.

3.4.2. Process or change management

The challenges of implementing AI within an organization extend beyond the technical and human factors considerations impacting work processes, job functions and organizational culture that are common across industries. Many technological advancements throughout history, such as the invention of the steam engine or the Internet, have resulted in automation and digitization in ways that have disrupted workforces [71]. Therefore, to ensure investments in AI technology and applications are successful, practitioners should consider pairing their deployments with change management strategies. Process

and organizational change management is an important aspect in the general delivery of new technology solutions within the nuclear power industry; particular additional considerations and challenges need to be planned for when implementing AI applications.

While AI driven applications are a comparatively new technology, organizational change management is a well studied field. It can benefit a practitioner, when considering the implementation of an AI application, to follow an established change management model to better identify and address the important considerations of change applicable to their area of practice. Established frameworks include the following:

— Kotter's eight step change model: a methodical approach for implementing successful organizational change, emphasizing the need for a sense of urgency, strong leadership, a clear vision and engagement of a broad base of stakeholders to drive and sustain change efforts [72];
— Lewin's change model: a three stage theory of change involving an 'unfreeze' stage to prepare for the change, a 'change' stage where the transition occurs, and a 'refreeze' stage to stabilize and integrate the new state as the norm [73];
— The ADKAR model: a framework that includes five essential elements — awareness, desire, knowledge, ability and reinforcement — with a step-by-step approach that emphasizes individual and human aspects of change [74].

Regardless of the change management framework or model that one decides to follow, it is important that change management be treated proactively as part of AI solution delivery and integrated into the AI life cycle to inform design, development, deployment and risk management practices as required. Therefore, assessment of the current state of organizational proficiency is fundamental for understanding readiness for AI. For example, stakeholder analysis might reveal low proficiency in the technology or a preference for interpretable AI models. In this case, developers may elect to design or select algorithms that may be less complex but offer better explainability, providing greater clarity and trust to end users. Some additional common considerations are establishment of a clear vision, identification and analysis of stakeholders, establishment of feedback mechanisms, and monitoring of the AI application's performance. Common causes of resistance to change when implementing AI are job replacement and workforce reduction, lack of technology proficiency among staff and managers, lack of trust in AI systems, and lack of trust in leadership decision making around AI technologies [75]. However, a change management plan should be tailored to the specifics of both the application and the organization and scaled as appropriate based on risk and effort. Finally, the deployment of AI solutions within the nuclear power industry demands a change management strategy that is as dynamic as the technologies themselves. Recognizing the iterative and evolutionary nature of AI technologies, change management is best when integrated into each phase of the AI life cycle while also remaining agile and responsive to ongoing feedback and technological advancements. This strategic agility ensures that change initiatives both address the current state and are anticipatory of future developments, underpinning the long term success of AI applications.

3.4.3. Risks

In 2022, the US Department of Energy released the AI risk management playbook [76], a reference guide for AI risk identification and mitigation pathways to support responsible and trustworthy AI development and use. This guide references more than 140 identified risks distributed among different groups.

The following are some of the inherent risks generally associated with AI applications:

— Over- and under-reliance on AI tools: Over-reliance on AI tools may lead to unintended results, including loss of expertise in the field. Under-reliance can render the tool useless and negate any potential benefits it brings. Inconsistent use of AI tools may also be an issue. Mitigation strategies

are often similar for both cases and can include training and designing the usage scenario in such a way that it maintains the desired level of human agency.

— Unintended or undetected model extrapolation: AI tools can be exposed to input data beyond the characteristics of the data the AI was trained on. In such cases, the response of the AI model can be erroneous. Mitigation strategies include introducing checks on the input data to attempt to detect out-of-scope conditions, monitoring input data for data drifts, and monitoring model performance to detect performance degradation.
— Data reliability and availability: The quality and quantity of data has an impact on the performance of AI tools. Inadequacy in data reliability or availability can impact development and deployment. Mitigation strategies could include developing data ecosystem and governance approaches to establish that there is no erroneous or spoofed data.
— Deployment risks: Once an AI application is deployed, issues can arise from misapplication of the algorithms, which generally results from any of the following:
    - Lack of understanding of the tool;
    - Failure to maintain currency with the real-time process due to ageing, reconfiguration or other changes;
    - Inadvertent use of the algorithm beyond its design scope;
    - Lack of trust due to underperformance or lack of transparency of an AI application.

  Mitigation strategies may include training, human factors considerations and performance monitoring to address potential deployment concerns.
— Introduction of new vulnerabilities: The integration of AI in legacy infrastructure may cause new vulnerabilities. Mitigation strategies may include infrastructure monitoring using tools such as vulnerability assessment and penetration testing.

Additionally, the risk of using AI solutions should always be weighed against the associated risk of current or alternate solutions.

3.4.4. Parallel and gradual deployment

Parallel deployment refers to the initial and temporary deployment of a system alongside currently adopted and accepted solutions. In this time window, the existing solution is still the valid or official one. Gradual deployment consists of incrementally enabling different elements of the AI solution, for example a gradual increase in the level of autonomy of the solution.

Parallel deployment brings several benefits:

— It allows for relevant field testing of the AI solution in a safe environment. Any failures would have no consequences. Conversely, differences in the outputs may also bring to light shortcomings in the existing practices that are mitigated by the AI solution.
— It helps build trust in and acceptance of the technology.
— It provides an opportunity under relevant circumstances for the staff to become familiar with and learn how to properly interact with the AI tool in anticipation of a potential transition.
— It assists in identifying potential issues in the system's human–AI interface.

An AI solution will typically require special considerations during development to enable parallel and gradual deployment. Such special considerations typically drive design decisions, from data input to model architecture and outputs, and therefore may even limit the AI solution. However, the potential benefits can justify such performance trade-offs, especially since they tend to considerably facilitate the adoption and acceptance of the technology. AI assistance to ultrasonic non-destructive evaluations is one example where such benefits have been observed.

3.4.5. Workforce training

Workforce training is a potential method to minimize human performance related issues when deploying AI solutions. Users would be trained on how to use and interact with the AI tool and how to properly interpret its outputs. It is valuable for users to be informed about anticipated failure modes applicable to the AI tools they interface with and how to recognize them. AI is not a replacement for domain expertise; users should still have sufficient proficiency in the applicable domain.

Training on AI itself is not necessarily an effective way to mitigate issues with any particular application. The tool should provide outputs that are relevant to the applicable domain of operation, are immediately understandable to staff with expertise in the area and are not overly dependent on AI terms or concepts.

3.5. MAINTENANCE AND QUALITY MONITORING

Implementation of AI within organizations operating nuclear facilities benefits from a team of subject matter experts from affected programmes along with technology implementers. Subject matter experts should work together to develop AI solutions that are both feasible and technically correct to minimize adverse impacts on quality and compliance. AI learning requires extensive and accurate input to develop models to meet business needs. Subject matter experts can provide or identify this training data and provide validation of model results. In the same spirit, experts in technology solutions should be consulted to integrate the model into existing solutions or create new solutions.

A cross-discipline review of AI performance enhances the probability of success. Engaging experts in the technical process will help achieve the desired results. Early recognition of declines is more likely when knowledgeable individuals are directly involved in the performance reviews of model outcomes. Without these reviews, the viability of using an AI system for critical decisions is challenged, which could have a negative impact on quality and compliance.

Continued efforts by subject matter experts have to be formalized into the implementation strategy for AI integration. Lack of validation and reinforcement of model outcomes may lead to a decline or stagnation in accuracy as time elapses after implementation. For example, a model trained to classify condition reports in terms of priority and severity may need to be updated as nuclear power plant staff turns over and differences in writing styles emerge. Initial training data may no longer accurately categorize future condition reports.

Evaluating performance periodically by testing is necessary to ensure results are still as desired. Depending on the level of degradation, an entirely new model may need to be trained to prevent correction attempts that lead to overfitting. This degradation should be identifiable through statistical analysis of data, dependent on the application of the AI. Applications centred around natural language may be harder to assess than those used for analysing data points collected from instruments.

Models that self-train run the risk of overfitting their data through a reinforcement loop. In nuclear applications, this could lead to undesired outcomes and should involve a human component. This may be the role of a QA group or individual departments that are responsible for the process that is being automated.

4. DATA CONSIDERATIONS

Data play a vital role in AI applications. AI applications depend on data to learn and operate, and their success largely relies on the degree to which the related data are accurate, complete, consistent and relevant. Regardless of the quality of an AI algorithm, its results can ultimately be unreliable when using inadequate data.

Typically, data supporting the development of AI applications can be divided into three groups based on their use:

— Training data: data used for training the AI model. These data are usually the majority of the available data. For supervised learning, these data typically consist of curated input–output pairs.
— Validation data: data used in combination with the training data with the purposes of tuning the untrainable design parameters of the model (called 'hyperparameters'). As an example, the depth of a decision tree is not a trainable parameter in the model; instead, it is set as a design parameter. Validation data enable assessment of the proper choice of such parameters, as well as assessment of the robustness of the model. Validation data are also crucial for avoiding overfitting, which happens when a model learns the training data too well and performs poorly on new data. By testing the model on unseen data, validation data help check if the model generalizes well to unseen data that it was not trained on. This process is essential for ensuring that the model is reliable and performs well on both seen and unseen data, reflecting its practical utility and robustness.
— Testing data: data used to test the model after it has been trained. This dataset should be completely independent from the training and validation datasets. Any data or information leakage from the training or validation datasets to the test dataset can compromise model performance assessment, biasing it towards more favourable results.

All three of these data groups should have similar characteristics (e.g. scope, distribution) and should be relevant to the desired application. The proportion of the total available data separated into each of these groups can vary. Typically, the training dataset uses the largest portion of the available data, and the validation and testing datasets are of similar size. Standard ISO/IEC 8183:2023 [77] provides a discussion on these data subsets. As discussed in Section 4.2, different data sources vary in adequacy across each of these groups.

Figure 1 illustrates the use of data through the life cycle of AI applications. During development, data from different sources are used for model training and testing; during operation, the trained AI models use field data to make predictions or recommend decisions. In applications involving dynamic models, AI models continue to learn and evolve over time by leveraging new field data for further training through a feedback loop.

The correctness and reasonableness of AI outputs are highly dependent on the accuracy, relevance and overall quality of the input data. Data cleaning is a significant and integral step in the data process as it promotes the removal of inconsistencies, inaccuracies and outliers, thereby enhancing the reliability and effectiveness of AI models. Reference [62] demonstrates how faulty training data can affect the model.

During the life cycle of AI applications, data go through many processes, including generation, transfer, cleaning, transformation and use. As shown in Fig. 1, data for the development of an AI model for a nuclear power plant or other domain application can be collected from various sources (e.g. field, laboratory, artificial, unvetted). These data can then be directly used or processed (e.g. transformed) to be suitable for the training and validation of the AI model.

Given the importance of data for AI applications and the overall life cycle shown in Fig. 1, this section discusses the following key topics:

— Data sources: Supporting data can (and often will) come from several sources, which can have their own limitations and considerations and can be more adequate for one stage of use than another. The different potential data sources to support AI model development are identified and discussed in Section 4.2.
— Data permissions: Data are usually proprietary, and proper use permissions are required. This is discussed in Section 4.3.
— Data fitness for usage: The overall characteristics of the underlying data are crucial for AI applications. Some of these characteristics are discussed in Section 4.4.

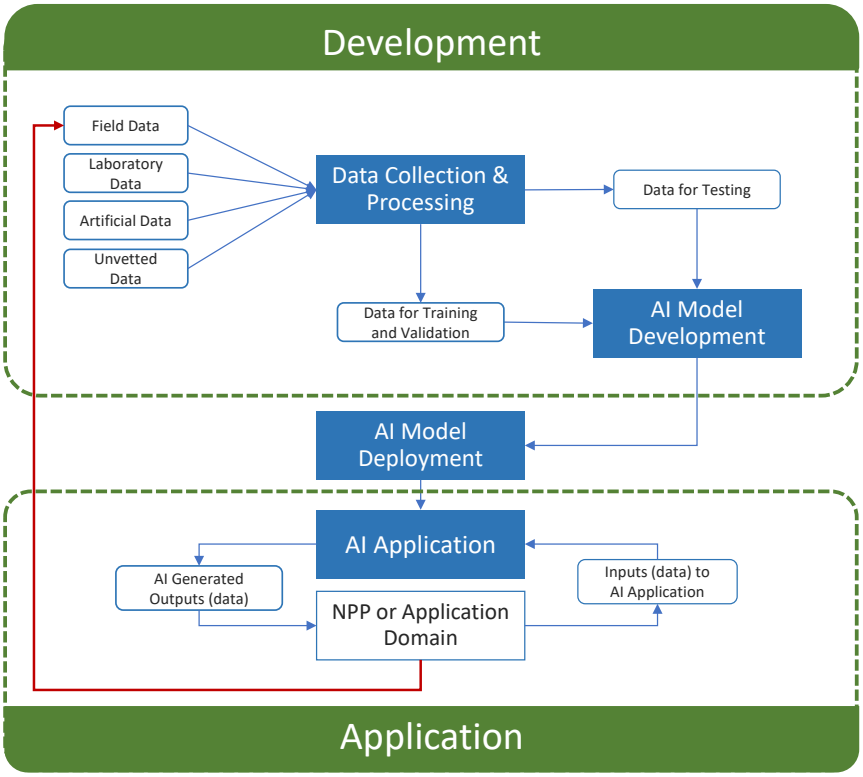


FIG. 1. Data usage through the life cycle of AI applications. NPP: nuclear power plant.

— Data management practices: Section 4.5 contains considerations related to the overall management of data through their life cycle, from data records to quality monitoring and audits.
— Data sharing: Having representative data in sufficient quantity is paramount for the successful application of AI solutions. However, for certain applications of interest to nuclear power plants, it is unlikely that any one single plant would have sufficient relevant data to enable a successful solution; therefore, data sharing will be of critical importance. Section 4.6 discusses the benefits and challenges of data sharing.

4.1. DATA LIFE CYCLE FRAMEWORK

A data life cycle framework is crucial in AI systems for several reasons:

— Data QA: The framework ensures that the data used for training and inference are accurate, reliable and representative. A well managed data life cycle helps identify and rectify issues such as missing values, outliers and biases, enhancing the quality of ML models.
— Model training efficiency: The framework streamlines the process of preparing and feeding data into AI models. Effective data collection and preprocessing during the data life cycle contribute to more efficient and faster model training, saving time and resources.
— Model generalization: The framework aids in creating models that generalize well to unseen data. By incorporating diverse and representative datasets during training, models become more robust and capable of making accurate predictions in real world scenarios.
— Scalability: As efficient handling of large datasets is essential in AI systems, the data life cycle framework supports the scalability of ML models by providing mechanisms to manage and process increasing volumes of data.

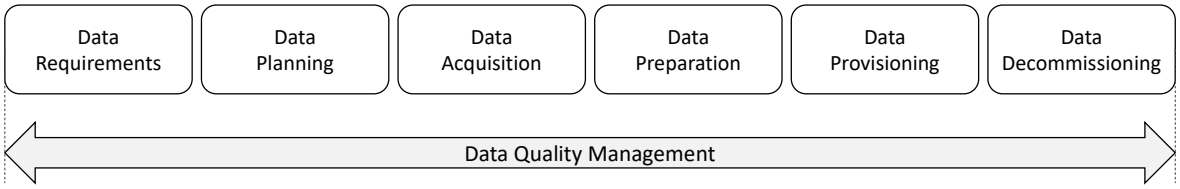


FIG. 2. Data quality management framework in ISO/IEC 5259 [78].

— Adaptability to change: As data patterns evolve over time and models need to adapt to these changes, the data life cycle framework, when integrated with regular model maintenance, allows for continual improvement and adaptation, ensuring that models remain relevant and effective.
— Decision making confidence: Reliable data throughout the life cycle provide confidence in the decision making process. Stakeholders can trust the insights generated by AI systems when they know that the data used are of high quality and have undergone thorough processing.
— Compliance and governance: Adhering to data life cycle best practices facilitates compliance with regulations and governance standards. It ensures that sensitive information is handled appropriately, mitigating risks associated with data integrity, privacy and security.
— Resource optimization: An organized data life cycle minimizes inefficiencies in data management. This optimization is crucial in resource intensive AI projects where effective use of computational and human resources is essential for success.
— Continual improvement: A data life cycle framework supports a feedback loop, allowing organizations to learn from the performance of deployed models. Continuous monitoring and assessment lead to iterative improvements, ensuring that AI systems stay relevant and effective over time.

A data life cycle framework helps ensure that appropriate considerations related to data governance, quality and security, as well as system utility, are taken into account at each data life cycle stage, including, as applicable, planning, acquisition or creation, development, use, maintenance and decommissioning. Figure 2 illustrates stages that data may go through to support an AI system, per ISO/IEC 5259 [78]. Standard ISO/IEC 8183:2023 [77] defines the stages and identifies associated actions for data processing throughout the AI system life cycle.

4.2. DATA SOURCES

As mentioned, data supporting AI can, and often will, come from various sources. This section identifies the main potential sources of data, providing definitions and important considerations for the use of these data sources in support of AI applications.

Table 1 summarizes the use of each kind of data source in training, validation (in the context of model development, rather than validation for use) and testing. This table is intended to provide a relative ranking and identify what justifications should be provided if the ideal data source is not available (as is often the case). Additionally, the application itself may impact the choice or applicability of different data sources, and regulatory bodies may impose restrictions on the types of data that are allowed.

4.2.1. Field data

Field data are collected directly from sensors or human input within the operational environment of a nuclear facility. These are the only type of data seen by the AI model during operation.

TABLE 1. PREFERRED USE AND CONSIDERATIONS FOR DIFFERENT POTENTIAL DATA SOURCES

Data source	Most suited for			Comment
	Training	Validation	Testing	
Field	✓	✓	✓	Preferred type of data for testing, if available.
Laboratory	✓	✓	?	Best option for testing when field data are unavailable. Technical reasoning should be provided discussing the data's relevance and applicability to the intended field application, identifying any assumptions or known limitations.
Artificial	✓	✓	✗	While appropriate for early model development, these data are not ideal for testing. If used for testing in the absence of field or laboratory data, a stronger justification of applicability and relevance is necessary. In addition, reasonable assurance should be obtained that the model is not responding to artefacts in the data resulting from the data generation process.
Unvetted	?	?	?	Such data can potentially be used for any model stage, depending on the source, but require thorough review for accuracy, adequacy, provenance, and so on.

✓: Preferred use.
?: Some justification is necessary.
✗: Typically not ideal; stronger justification required.

Considerations for the use of these data include:

— Can be critical data for real-time monitoring, predictive maintenance and operational decision making.
— Should be validated rigorously to ensure accuracy and safe decision making, especially when using sensor data.
— When used for near real-time inference and decision making, on-line sensor data should be monitored for calibration drift or periodically checked for calibration.
— Typically require data stewards who are experts in the associated system and component being monitored via instrumentation.
— Can be used at any stage. It is essential that these are the main type of data at final testing. Even if they are not the only type of data used in testing, performance should also be assessed exclusively on the field portion of the test data.

4.2.2. Laboratory data

Laboratory data are data gathered in controlled settings to conduct experiments and physical simulations. Examples include data from inspections on laboratory specimens.

Considerations for the use of these data include:

— In training and validating AI models, laboratory data should be used in conjunction with field data and known physical models for testing.
— Laboratory data can support initial selection and development of AI models.

— Laboratory data are often clean and well structured, making these data ideal for the initial training phases of AI models; however, testing in real world settings is still required.
— Care should be taken to ensure that laboratory data adequately represent the complexities of field data.
— Laboratory data can augment datasets that are difficult to collect in real operational settings. For example, in-service failure data are typically rare for critical operating equipment. Laboratory environments can be used to intentionally fail components to collect data across various failure modes when building reliability and predictive maintenance models. In these situations, the data may still lack the operational nuances that only field data can provide.

4.2.3. Artificial data

Field and laboratory data are real, resulting from physical measurements. Data may also be generated artificially. There are several methods to generate artificial data, including:

— Data generation through computational or physical models that simulate real world scenarios; these typically include a physical model of the phenomena in question. Examples include data generated by analytical or finite element method models.
— Data generation through AI methods, such as generative adversarial network or physics informed ML models.
— Data generation by manipulation of real data (field or laboratory). Examples include manipulation of data to alter relevant characteristics (e.g. size or location of a target defect) or to expand the scope of the data beyond what is available as field or laboratory data while still leveraging real datasets. Further examples can be found in Refs [79, 80].

Considerations for the use of these data include:

— Data augmentation can support scenarios with insufficient data quantity or distributions.
— Artificial data should be validated against real world data where possible and with rigour commensurate with the risk associated with the use of AI model outputs.
— Artificial data are useful for initial AI model training and validation due to cost effectiveness but generally should not be solely relied on for testing.
— Special care should be taken to ensure that artificial data adequately represent the complexities of real world data.
— Artificial data could be used for stress testing AI models under extreme conditions that may be unsafe or impractical to replicate in the real world.
— The data lineage (e.g. models, simulations, assumptions, methodology, transformations) used to generate artificial data should be well documented and understood.
— Artificial data may not be suitable for operational decision making.

4.2.4. Unvetted data

Data may also come from unvetted sources and be of any of the discussed types. Open source data are the key example: while publicly available (although still potentially subject to licence agreements or intellectual property (IP) protection), the validity, accuracy, provenance and other characteristics of open source data are often unknown or untraceable.

Another challenge associated with the use of open source resources is that they may compromise the independence between training, validation and testing groups. Because open source datasets and models may overlap, it may be challenging to claim independence when performing validation or testing of AI models. Two approaches can be adopted to overcome this. The first relates to ensuring independence by limiting the testing datasets to a prescreened set of data that is qualified for benchmarking. However, if

those benchmarks are published, model developers could use them in training, potentially compromising the validation independence and biasing the model performance assessment. Another approach is to accept the lack of independence and design the validation process to assume a certain level of dependence; tools can be created to determine the level of dependence of datasets.

Considerations for the use of these data include:

— Open source data can supplement existing datasets. For example, the use of public weather datasets may be useful in the performance of environmental impact modelling.
— The data quality, reliability and relevance of unvetted data should be validated rigorously, as well as compliance with licensing and data usage policies.
— Adequate use between training, validation and testing will depend on the dataset and needs to be assessed. Ensuring adequate independence between training, validation and testing sets may be challenging.

4.2.5. Adversarial data

Adversarial data are data intentionally used with malicious purposes, during either model training or inference. Such data are engineered and introduced by external agents with the malicious purpose to thwart model performance in some way.

Because adversarial data can be used for malicious purposes, developers may use adversarial data during the training process to make the system more robust against adversarial attacks.

4.3. DATA PERMISSIONS

Regardless of their source, data are usually proprietary, and therefore it is necessary to ascertain the proper permissions to use them. Furthermore, certain types of data, such as personally identifiable information, are sensitive and protected by laws that need to be observed by the data management process.

There are two primary types of proprietary data:

(1) Publicly available information: Information that is made available to the public but is not in the public domain (e.g. third party open-sourced data) is still subject to IP protection (e.g. copyright) and its use likely requires compliance with relevant IP licence terms (e.g. open source licences). In some instances, these IP licences may include restrictions that could pose risks to a business. For example, certain open source licences may require a licensee to restrict use of the open-sourced data to internal use only or require that any proprietary data used with the open-sourced data be made available to the public. Orphaned works are works that are available to the public and subject to copyright protection, but for whom the owner is unknown. Risk is higher when using works from unknown sources as owners can resurface and claim rights to the data used.
(2) Confidential information:
    — Information subject to safeguarding or dissemination controls pursuant to and consistent with applicable law, regulations and government-wide policies, such as export control requirements.
    — Other sensitive information derived internally from an organization or received from third parties that is subject to confidentiality treatment:
        • Technical and business sensitive information should be handled appropriately to avoid inadvertent disclosure of trade secrets or other confidential information.
        • Any proprietary information from vendors, contractors or other such third parties working with an organization that is provided under a non-disclosure agreement or similar restriction needs to be treated with confidentiality.

Practical considerations within the context of data permissions as applied to AI model development and usage include:

— Training a model: Generally, if any data are used to train a customized model that should not have been used, the model needs to be retrained. Therefore, a strong data quality management programme and strong data governance need to be exercised to ensure that the data used to train the models are permissible to use.
— Prompts or other inputs: Users of all AI tools should be familiar with relevant data policies of AI based service providers and internal policies of a business. Discretion should be exercised when providing prompts or queries (e.g. inputs) into an AI interface to receive responses (e.g. outputs). In some cases, businesses can reduce risk by operating under a secure and private dedicated environment when using proprietary data as inputs.

When dealing with proprietary or sensitive data, considerations for use include:

— Identifying third party information that is not owned or licensed to the user;
— Obtaining prior consent or a licence from the copyright owner of the information for the intended use;
— Applying data anonymization to remove unnecessary sensitive information that may be present, while preserving the required data properties;
— If data anonymization is performed, reevaluating the modified dataset to ensure that it can be used for the intended purpose without adverse effects;
— Ensuring that the IP indemnity obligations from the service provider of AI tools or model owner (e.g. for copyright infringement claims) are met.

4.4. DATA FITNESS FOR USAGE

The overall quality of the data supporting AI applications is crucial. The characteristics that make a dataset suitable or fit for use in a given application, and thus of quality, are case specific. Specific metrics or methods to quantify these attributes will also vary between different applications. A discussion of such metrics or methods is outside of the scope of this document, but a list of guiding questions is included that can help assess or characterize each attribute as well as clarify its intent.

These considerations apply equally and individually to each of the three subsets of data defined earlier: training, validation and testing data.

4.4.1. Data quantity

Data quantity is the volume of data available for AI applications. The appropriate volume size is case specific and requires an assessment for each application.

Expected impact on AI models:

— Insufficient data is not representative of real world operations and scenarios.
— Having a sufficient volume of data enhances the robustness and ability of AI models to generalize.

Considerations for assessment:

— Assess if the volume of data is adequate for training a reliable model for its intended use.
— Consider alternative strategies, such as data augmentation or synthetic data generation, to mitigate the risk of data scarcity.

— Assess if the data quantity aligns with the complexity of the process or phenomena being modelled.
— Determine if there are any regulatory requirements regarding the minimum amount of data required for validation.

4.4.2. Data relevance

Data relevance refers to how well the dataset is applicable to and describes the desired application scenario.

Expected impact on AI models:

— Data relevance directly affects the adequacy of the models for the intended application.
— Significant impact on performance is expected if a model has been trained on data whose characteristics differ from the data at deployment.

Considerations for assessment:

— Determine whether the data accurately reflect the expected conditions at deployment for the intended application.
— Determine whether the data encompass all expected scenarios during deployment.
— Review which, if any, simplifications or assumptions are inherently included in the data.

4.4.3. Data distribution

The data distribution is the way data points are spread across a range, different categories or classes.

Expected impact on AI models:

— An imbalanced data distribution or a training or test distribution that differs from the real world target distribution can introduce bias into AI models, leading to unfair or inaccurate output.
— An imbalanced dataset where some outcomes are much more likely than others may lead to an AI model that is biased towards the more likely outcomes.
— Ensuring balanced data distributions is critical to the fairness and accuracy of AI model outcomes.

Considerations for assessment:

— Determine whether the data are representative of the different categories the AI model aims to cover and whether the distribution has been quantitatively evaluated.
— Determine if techniques such as resampling are or can be used to balance the data distribution.
— Assess the importance of each class or category in the context of the AI model's task.
— Determine whether the data distribution reflects real world conditions or if it is skewed.
— Assess the risk of the real world distribution changing over time, what impact this would have on model reliability and how this can be monitored.

4.4.4. Data scope

Data scope is the range and context in which the data are applicable.

Expected impact on AI models:

— Data that are too narrow or broad in scope can lead to AI models that are either over-specialized or too generalized for the intended use.
— Properly scoped data ensure that the model and the resources required to train, test and deploy the model are both effective and efficient in their intended applications.

Considerations for assessment:

— Determine whether the data cover all necessary conditions and scenarios for the model's intended application and whether they cover any conditions or scenarios that are irrelevant for the intended application.
— Consider whether there are specific operational contexts where the data scope may need to be adjusted.

4.4.5. Data bias

Data bias is systematic errors in data that can lead to unfair, skewed or poor quality outcomes.
Expected impact on AI models:

— Data bias introduces performance and reliability issues in AI models. For example, in monitoring and diagnostic applications a biased signal can lead to incorrect decision making, masking issues or excessive false calls.
— Addressing bias is important for the model's reliability as well as for maintaining stakeholder trust.

Considerations for assessment:

— Assess if there are known or potential sources of bias in the data.
— Identify what steps are being taken to identify, mitigate or correct data bias.
— Assess whether the data are collected from diverse and representative sources.
— Assess whether there are blind spots in the data that could introduce unintentional bias.

4.4.6. Susceptibility to errors

A dataset's susceptibility to errors is the likelihood of inaccuracies or mistakes being present in the data and is another measure of data quality.
Expected impact on AI models:

— Undetected errors in training data can propagate through the model, leading to inaccurate or unreliable results.
— Erroneous data can cause misallocation of resources by triggering unnecessary interventions, or maintenance activities in the example of predictive maintenance. This highlights the importance of validating model outputs where possible.

Considerations for assessment:

— Review and identify whether there are any data quality and validation checks in place to catch errors in data collection or entry.
— Determine if there is clear ownership and accountability for data sources being used in AI model applications and for correction of errors.
— Assess if data lineage is documented for auditability.
— Assess whether there is a system in place for continual monitoring and correction of errors and if there are automated alert mechanisms for identifying and reporting data anomalies or errors.
— Assess the impact of potential errors on the AI model's performance and safety considerations.

4.4.7. Data alignment

A key consideration is the alignment of the data used in model development (training, validation and testing data) with the target application scenario. Characteristics such as data relevance, distribution and scope for the testing data define and bind the range of applicability of the model with the performance as demonstrated during testing. No assumptions should be made based on the model performance on data that extrapolate from that range. Therefore, it is important to seek alignment between the characteristics of the data used in model development and those expected during deployment of the target application.

4.4.8. Data quality management

As mentioned, the specific required characteristics of a dataset are application dependent. Once those characteristics have been defined, the quality of any given dataset for the desired purpose can be assessed against them.

For data quality management in an AI project, standard ISO/IEC 5259-1 [78] suggests a data quality management framework for the data life cycle. The data life cycle consists of data requirements, data planning, data acquisition, data preparation, data provisioning and data decommissioning, as shown in Fig. 2.

This life cycle in ISO/IEC 5259-1 [78] is an instantiation of ISO/IEC 8183 [77]. The data quality management framework provides the process for determining, accessing and improving the quality of datasets for use in AI applications. The framework includes the following elements:

— Data quality model: a defined set of data quality characteristics (such as those discussed in Section 4.4) that provides a framework for specifying data quality requirements and evaluating data quality;
— Data quality measures: means of evaluating each data quality characteristic in the data quality model;
— Data quality assessment: means of assessing whether a dataset meets its needs and requirements;
— Data quality improvement: means of transforming data to improve the dataset's quality to the extent that it meets the needs and requirements of the organization;
— Data quality reporting: means of publishing data quality reports to determine the root cause of the poor performance of an AI model and for transparency and explainability of the AI.

It is relevant to consider the data security of the data underlying AI applications and how they may be distinctively affected. The primary goals of data security involve confidentiality, integrity and availability, and each can be threatened in a different way at different stages of data processing, as indicated in Table 2.

Threats to the confidentiality and availability of AI supporting data are like those to any other data (e.g. encryption, authentication, duplication) and thus will not be discussed here. AI applications, however, are also subject to specific threats to their data integrity that warrant discussion.

TABLE 2. EXAMPLE THREATS TO DATA SECURITY AT DIFFERENT STAGES OF THE DATA LIFE CYCLE

	Confidentiality	Integrity	Availability
Data storage	Disclosure	Tampering	Ransomware
Data transmission	Eavesdropping	Falsification	Transmission interruption
Data processing	Reverse engineering	Data poisoning	Overloaded

TABLE 3. THREATS TO DATA INTEGRITY IN AI APPLICATIONS

Threat	Presented by
Data drift	● Operational changes in the nuclear power plant (even if expected) ● Natural changes over time
Data corruption	● Accidental or malicious human activity ● Logical errors during data transfer, processing and storage ● Physical compromise to the device hardware or data disk ● Sensor malfunction
Data poisoning	● Malware, viruses or cyber-attacks ● Accidental or malicious human activity

The US National Institute of Standards and Technology defines data integrity as "the property that data has not been altered in an unauthorized manner" [81]. Data integrity in the case of AI application to nuclear power plants can similarly be defined as the property that data maintain their quality throughout the life cycle of the AI application and are not deteriorated by an unauthorized agent or process, where the latter includes the inadvertent consequences of changes in the operational characteristics of the power plant.

Some of the main threats to data integrity for an AI application include:

— Data drift: Unexpected noises contained in the data or undocumented changes to the data structure can cause the AI model to draw inaccurate or incorrect conclusions.
— Data corruption: If the data are corrupted or unusable, the AI model may produce unstable outputs or fail to work.
— Data poisoning: Malicious adversarial data can be injected into the training dataset of AI applications, causing the AI model to learn from the poisoned data. The AI model will be compromised and make untrusted decisions when faced with new data. This can occur at either the training or inference stages.

Table 3 lists some ways through which each of these threats can present themselves. It is noteworthy that data drift can be inadvertently caused by normal operation of the plant without any intended adverse action. While the causes of the other threats are not specific to AI applications and may already be minimized by existing security and other measures, the causes of data drift require specific measures to be implemented, such as specialized monitoring.

4.5. DATA MANAGEMENT PRACTICES

The management of data integrity is necessary to ensure that data, irrespective of how it was generated or its format, is properly recorded, processed, retained and used in a manner that ensures a complete, consistent and accurate record throughout the data life cycle. This management, called data governance, should be carried out based on the organization's internal standards, policies, rules and processes. The data governance provides the following [78]:

— A set of guiding principles established by an organization to actively manage and improve data quality;
— Decision making structures and accountabilities through which those assigned data quality responsibilities are held to account;
— Organizational roles and responsibilities to ensure data quality through repeatable processes.

Some specific considerations for data management include:

— Metadata management:
  - Maintenance of metadata ensures data traceability and understandability.
  - Metadata management can play an important role in facilitating data discovery for AI models.
  - Practices like a business glossary and data catalogue should be implemented for inventory and classification of metadata.
  - Clear ownership and stewardship need to be defined across data domains and classes.
— Data lineage tracking:
  - Data lineage should be systematically mapped to provide a clear audit trail from source to consumption. This will help in debugging model and data issues and verifying transformations.
  - The transformations and preprocessing steps applied to the data should be recorded, and data manipulation should be properly logged and given a version number for traceability.
— Data records:
  - A record should be kept of all data that impact AI model decisions, particularly those affecting safety and regulatory compliance. For example, models used to drive condition based maintenance should have training data and model parameters clearly documented in their records.
  - Records need to be stored in secure environments to ensure integrity.
— Data life cycle management:
  - Data archival and retention guidelines should be implemented. For example, raw sensor data from a reactor system used in model training and inference may need to be stored indefinitely to explain potential failures and gaps in model performance.
  - Data should be regularly backed up to prevent loss.
  - Secure, encrypted and redundant storage solutions should be used to safeguard data against loss, corruption and unauthorized access.
  - Clear procedures should be outlined for safely decommissioning and deleting data that are no longer needed. This process should meet regulatory requirements, and all data lineage information should be updated to reflect the decommissioned data.
— Data quality monitoring:
  - Continuous monitoring systems should be implemented to identify data quality degradation over time.
  - Quality attribute metrics like clarity, accuracy, consistency and completeness should be used for monitoring.
  - Validation rules, cross-references with trusted data sources and periodic reviews by subject matter experts should be implemented as required.
  - Automated alerts should be used for data quality issues, including threshold breaches, anomaly detection and other pattern recognition techniques that could indicate quality degradation, corruption or tampering.
  - Quality monitoring and checks should be implemented at different stages of the data life cycle and automated where possible.
  - A log of identified data quality issues should be maintained, along with status and resolution steps to support issue resolutions and longer-term trending and quality improvement efforts.
  - A feedback mechanism should be established with data stewards and business subject matter experts to continually improve data quality.
— Need for audits:
  - Periodic audits should be conducted for compliance with data governance policies and standards. Audits can be conducted across the full data life cycle and/or on AI models.
  - Audit logs and reports should be stored securely and trends in results identified to support future investigation and analysis.

— Data leakage:
    - It is important to carefully partition data into training, validation and testing sets while ensuring there is no data overlap between them. Sampling techniques that maintain the distribution of classes across different sets may be useful.

4.6. DATA SHARING

Data are fundamental to successful AI solutions. Field data are arguably the most important of the data sources discussed as they are the only data that enable relevant and meaningful testing and assessment of the application. In the case of nuclear power plants, these data reside with the nuclear operators, who often lack the expertise and resources to develop AI solutions on their own. This scenario makes data sharing a crucial enabler for the successful deployment of AI solutions across the nuclear power industry. In the context of this document, data sharing refers to nuclear operators (as the data owners) sharing relevant field data with AI solution developers or other operators in a combined effort or with the AI community at large to support the successful development and deployment of AI solutions of interest.

Despite its importance, a culture and practice of data sharing within the nuclear power industry has been difficult to achieve to the level necessary to support the successful development and deployment of AI tools at large. The reasons for the reluctance in sharing data are understandable, as often the relevant data are sensitive and sharing them can lead to adverse impacts to the data owner for diverse reasons. Nonetheless, the benefits of sharing could outweigh the associated risk, especially if the necessary care is taken. Sections 4.6.1 and 4.6.2 discuss the value of sharing data and some of the practices that can be used to overcome associated concerns.

4.6.1. Value and importance

Some of the benefits brought about by data sharing include the following:

— Data sharing can foster a culture of collective problem solving and innovation, leading to more robust AI models and analytics tools. It is often the case that nuclear power operators are rich in real world data but lack the specialized skills and tools to perform advanced research and implementation of AI models using these data. Conversely, academia and private firms that have these specialized resources often do not have access to these rich real world data. Collaboration between these types of parties therefore supports the advancement of innovation as well as results for the nuclear power operators.
— Data sharing can help in the formation of industry-wide standards and best practices, which is particularly important in the regulated nuclear energy sector.
— While not specifically an AI related benefit, data sharing can also support regulatory compliance and improve the efficiency of regulator audits through a reduction in the resources and effort required during discovery and data collection. This is an important consideration for activities that require regulatory oversight. Although not necessarily (or initially) required, it is in the operator's best interest to share as much data and related information as possible with the regulators; this will not only facilitate and expedite the review by minimizing additional requests but also help foster trust and transparency.
— For some applications of common interest across the nuclear fleet, a single operator may not have sufficient relevant data to support the successful development and deployment of an AI tool. Therefore, data sharing in a collaborative effort may help realize the envisioned benefits.
— Data representativeness has been identified as a common key characteristic of the datasets for successful deployment of AI solutions. Through data sharing, nuclear power operators can guarantee that their assets are represented in the underlying data feeding AI models, thus increasing the likelihood of successful model deployment in their plants.

— Data sharing can support the creation of relevant benchmark sets that enable the industry to assess the performance of proposed solutions on meaningful data in a safe environment and before committing significant resources for deployment. Such datasets would allow the industry to easily evaluate solutions from different providers against common, well understood and well documented scenarios. They would also enable solution providers to assess their solution against relevant data that are otherwise typically unavailable to them.

4.6.2. Confidentiality and sensitivity concerns

Some practices that can aid in addressing or minimizing data sensitivity concerns when sharing data include:

— Addressing confidentiality concerns by anonymizing or pseudonymizing data prior to sharing.
— Using data masking techniques to hide specific sensitive attributes.
— Pursuing licences, such as export licences, through regulators to share restricted or controlled data as required by local laws, regulations and operating licences.
— Making data sharing agreements between parties that specify the terms of use of the data and the rights and responsibilities of each party, and so on.
— Ensuring data sharing respects ethical considerations like individual privacy.
— Adhering to data sovereignty laws.
— Securing data for sharing from assets that are no longer in operation (e.g. decommissioned plants, retired components) while the data are still accessible; helping to identify the existence of such data; and supporting activities where the data can be accessed in a safe and controlled manner. One example is field trials, where data can be accessed locally and used to inform and test models without having to leave the site.

The applicable concerns and obstacles to data sharing will often be specific to the application and data in question and, at times, can be easily addressed. Identification of the specific concerns for a given case and review of applicable mitigation strategies may reveal acceptable solutions that enable utilities to realize the value of data sharing for mutual benefit.

5. FURTHER CONSIDERATIONS

Previous sections have delved into various aspects of AI, notably the benefits AI may present, AI system life cycle management and the importance of data. This section seeks to provide detail on some considerations that have been touched on previously but merit further elaboration.

Risk assessment, integral to all nuclear operations, is fundamental for AI. From the start, those seeking to develop and deploy AI should be asking the following:

— What could go wrong?
— What are the consequences of failure?
— How can negative consequences be mitigated or precluded?
— Does the application need guardrails to limit use?
— Is defence in depth available if the system fails?

Addressing these questions, at a minimum, at the outset of developing an AI system is best practice.

Many nuclear regulators use a risk informed approach to regulating, and such an approach is adaptable to the regulation of AI enabled applications. Areas where this approach may be appropriate

include the training and competence of both humans and AI, periodic safety reviews focused on AI, security, and ethical considerations. Early and frequent regulator engagement on these and other aspects of AI pre- and post-deployment may assist in minimizing surprises on the part of either party.

Explainability, although perhaps elusive, is a critical factor for AI. As the level of accuracy of AI models increases, generally there will be a decrease in the level of explainability, as highly accurate models — especially deep learning models — tend to be more complex and less explainable. The balance between accuracy and explainability should be considered in relation to the desired requirements of the nuclear power plant systems in which the AI is to be incorporated.

Finally, engaging with regulators is prudent throughout the life cycle of the AI application. Early and frequent engagement assists in fostering a 'no surprise' outcome for both parties and may aid in obtaining a positive regulatory response to the application and a safe deployment.

5.1. REGULATOR PREPARATION FOR ARTIFICIAL INTELLIGENCE

Currently, regulatory requirements around the use of AI in nuclear related activities are nascent. However, some national nuclear regulators are engaging with industry stakeholders, including AI developers, operators, end users and international regulatory bodies, to gather insights on AI technologies and applications. These efforts are helping regulators to stay informed about the latest developments and to develop regulatory approaches towards AI enabled applications that are relevant and fit for purpose. Regulators are also establishing forums for outreach to stakeholders to maintain awareness of industry AI applications, obtain feedback on regulatory needs and offer updates on regulatory efforts related to AI. The following examples serve to illustrate those efforts.

5.1.1. Canadian Nuclear Safety Commission, United Kingdom Office for Nuclear Regulation and United States Nuclear Regulatory Commission principles paper

The Canadian Nuclear Safety Commission, the United Kingdom Office for Nuclear Regulation (UK ONR) and the US NRC have a long history of collaboration on regulatory matters. While each regulator has undertaken work separately to understand and characterize the potential safety benefits and risks of deploying AI in nuclear activities, in 2023 they began a cooperative project to publish a trilateral white paper, or principles paper, on this topic.

Published simultaneously on all three regulators' external web sites in the fall of 2024 [82], the paper describes approaches to the oversight of nuclear activities, as well as materials, using AI in Canada, the United Kingdom of Great Britain and Northern Ireland, and the United States of America. The paper explains common objectives and how consideration of areas important to the effective oversight of AI used in nuclear activities across all three countries is possible. It is intended that the considerations outlined will encourage beneficial uses of AI and clarify the challenges arising from such fast developing technologies and the principles applied to regulating them.

The paper touches on the following subjects (not ordered according to importance):

— Country specific regulatory philosophies and perspectives;
— AI use cases (high level categories);
— Use of existing engineering principles for safety and security systems;
— Human and organizational factors;
— AI architecture in nuclear applications;
— AI life cycle management;
— Documentation of AI safety and security.

The paper defines AI as a range of technologies that learn from data or experiences to perform tasks otherwise requiring human intelligence. It also describes key considerations when deploying

AI while maintaining safe and secure operation of nuclear facilities. The three regulatory bodies also recognize the importance of thoughtfully considering these principles when developing, reviewing and deploying AI systems.

5.1.2. United Kingdom Office for Nuclear Regulation and United Kingdom Environment Agency report — pilot of a regulatory sandbox in the nuclear sector

Nuclear regulators are employing innovative approaches to conduct preliminary assessment of how best to allow for the deployment of innovation itself into nuclear regulated activities. One such example is the use of 'regulatory sandboxing' by the United Kingdom Office for Nuclear Regulation in coordination with the United Kingdom Environment Agency.

The sandboxing exercise, also called a RegLab, involved workshops prior to the actual sandboxing exercise. These pre-meeting workshops helped participants define the problem/challenge statements used in the consideration of the two mock safety, security and environment case structures. These key aspects were then used to develop four deep dive topics for each of the two AI applications, which were then explored at the regulatory sandboxing sessions [83].

The report on the effort stated that the benefits of the AI application should be clearly articulated and that how the benefit compared to existing technologies should be taken into consideration. The risks associated with the deployment of AI need to be characterized, understood and mitigated if need be. The report also made the following recommendations:

— Deploy AI gradually and in a phased manner to build confidence and experience.
— Evaluate whether a principles based approach to regulation is preferred, to take into account differing considerations for each potential application of AI.
— Understand the limitations of training data, especially as they apply to deployment with real data.
— Understand the importance of human factors in the AI life cycle.
— Establish an AI safety culture to complement and support the conventional safety culture.

5.1.3. United States Nuclear Regulatory Commission — Artificial Intelligence Strategic Plan: Fiscal Years 2023–2027

The NRC developed an AI strategic plan [2] to proactively prepare for future uses of AI applied to NRC-regulated activities. The plan includes five goals:

— Ensure the NRC is prepared for efficient regulatory decision making on AI applications.
— Establish an organizational framework to coordinate AI activities.
— Expand domestic and international AI partnerships.
— Cultivate a workforce proficient with AI.
— Build foundational knowledge through developing use cases.

A supporting project plan [84] was also developed to outline tasks needed to meet those goals. Some of the tasks in the project plan include:

— Establishing an AI steering committee and an AI community of practice;
— Performing a regulatory framework applicability assessment to evaluate the existing framework applied to the oversight of AI enabled applications;
— Maintaining awareness of potential use cases.

The NRC has hosted a series of public workshops to provide updates on progress related to agency AI activities [85]. These workshops provide a forum for information exchanges on industry use cases and

viewpoints on regulatory needs. The NRC also maintains a web site to provide additional updates related to AI topics [86].

The chair of the NRC tasked staff to perform a review of how AI could be applied to internal NRC licensing and oversight processes [87]. Offices provided use cases, which were reviewed by relevant staff from throughout the agency. The staff response [88] reported common themes among the proposed use cases and provided next steps on leveraging AI for internal use.

5.2. HUMAN FACTORS SAFETY CONSIDERATIONS

Section 1 discussed representative levels of human oversight for an AI system or, alternatively, how much autonomy the AI has in controlling the underlying system. These levels can range from AI providing insights on collected data for a human analyst to review up to full autonomous control by AI of a system with little to no human oversight. The role of the human in the operation of the AI system should be defined, as well as how the human should interact with the system and the appropriate level of training needed to interact with the AI, because the interaction over time between human and AI system could also influence the implementation strategy if more control is granted to the system as more experience is gained with its use. The level of human oversight of the system could play a role in its acceptance but may not be the only factor. The overall risk of the system encompasses both the risk level of the application and how the AI is deployed.

Prior to deployment, and from the start of the development of an AI system, the level of oversight and autonomy should be determined. The level of oversight can impact safety, risk, and ethical and regulatory considerations. While the interplay between human oversight and AI autonomy has many possible degrees of variance, the following describes, at a high level, three ways of viewing the relationship:

— Human-in-the-loop systems: Some AI systems operate with a high degree of oversight, often referred to as human-in-the-loop systems. In these cases, humans are actively involved in the decision making process and can evaluate the adequacy of the AI response.
— Human-on-the-loop systems: AI systems where humans are involved in monitoring and supervising but not actively making decisions during operation are referred to as human-on-the-loop systems. These systems provide a human operator a level of oversight and control, if necessary, while allowing the AI to make decisions with more freedom.
— Human-out-of-the-loop systems: AI systems that operate with minimal or no human intervention can be considered human-out-of-the-loop systems. These systems may be designed to be fully autonomous, or the AI may be granted autonomy as more experience is gained and human oversight is reduced.

Striking the right balance between autonomy and oversight of an AI system is crucial. Too much autonomy may lead to unintended consequences, while too much oversight can hinder the potential benefits of AI. The level of oversight should be adjusted to suit the specific goals and risks of a given application. Regardless of the level of autonomy, transparency and accountability mechanisms are essential. These mechanisms should include tracking and explaining the decisions made by AI systems, providing avenues for recourse when errors occur and conducting audits to ensure compliance with ethical and legal standards.

For systems where humans are integral to the operation of the AI, the performance of the user should be considered when evaluating the performance of the AI system. The role of the human in the operation of the system should also be defined. This is important when a user is providing oversight of an AI system as well as when an AI is providing recommendations to a user. For instance, if a human is in the loop to provide a backstop for the performance of the AI system but accepts all recommendations from the AI, the human is not serving their function as a backstop. The operator of the system should receive the appropriate level of training to use the system and monitor both the performance of the underlying

physical system and the AI application to be able to judge when the AI is functioning appropriately. The operator should understand what they should be monitoring and when they should provide oversight, as well as what they should do if oversight is required.

Leveraging human oversight may be an effective strategy to support the deployment of autonomous systems. A system intended to be fully autonomous may initially be deployed to offer recommendations to a human operator and run concurrently with traditional operations. As more experience is gained with the AI system, the AI can be given more responsibilities with human oversight. If the AI performs successfully with human oversight, it could transition to a fully autonomous system. If used, this deployment strategy should be documented and metrics should be specified to evaluate if the system is performing adequately under the different thresholds of oversight.

Regardless of the level of oversight, the performance of the AI should be assessed throughout the lifetime of the system. If performance degradation is observed, the model should be appropriately modified, for example through retraining, updating algorithms, or altering the level of autonomy.

5.3. RISK ASSESSMENT

Because AI may play a crucial role in enhancing nuclear power plant safety and operational efficiency at nuclear power plants, it is important to address the identification and documentation of risk in the context of AI. Consideration of potential issues, such as vulnerabilities, behaviour attributes, consequences of failure and AI response, can support AI development and deployment. The following questions may aid the identification of, assessment of, and strategy for addressing the potential risks in using an AI system.

— What could go wrong?
  - Vulnerabilities: Vulnerabilities in AI systems at nuclear power plants may arise from various sources, such as errors in data input, software bugs or external cyber-attacks. These vulnerabilities can compromise the integrity, reliability and safety of AI systems. Vulnerabilities in AI may manifest as incorrect predictions, misclassifications or other deviations from expected behaviour. For example, a vulnerability in a predictive maintenance AI system might lead to missed detection of a critical equipment failure. It is important to demonstrate that behaviour attributes are met and vulnerabilities are absent.
  - Behaviour attributes: Understanding the behaviour attributes of AI is vital. AI systems can exhibit unexpected behaviours due to data anomalies, model drift or adversarial attacks. It is important to monitor and validate AI systems continuously to detect any abnormal behaviour.

— What are the consequences of failure?
  - What happens if the application fails? The consequences of AI failure in a nuclear power plant can be severe. AI systems are often used for tasks like equipment health monitoring and anomaly detection. If an AI system fails to perform its duties correctly, it can lead to safety incidents, equipment failures or disruptions in plant operations.
  - How does AI respond to vulnerabilities? AI systems have to be designed with robustness and fault tolerance in mind. It could be beneficial to include mechanisms for self-assessment, self-correction and adaptation. For example, if an AI system detects that its performance is degrading or that vulnerabilities are present, it could trigger an alert, notify operators and potentially switch to a safe mode of operation. Additionally, AI systems can be equipped with cybersecurity measures to defend against external threats. AI should be designed to supply reliable exit mechanisms for users, should they be needed, to alleviate, avoid or escape from the results of a vulnerability.

To mitigate these risks and consequences, nuclear power plant operators and AI developers can take several measures:

— Comprehensive testing: Thoroughly test AI systems under various conditions, including edge cases, to identify vulnerabilities and ensure that the systems behave as expected.
— Redundancy: Implement redundancy in critical AI systems to ensure that backup mechanisms are in place if a primary system fails.
— Regular auditing and monitoring: Continuously audit and monitor the AI system's performance and behaviour, looking for signs of vulnerabilities or deviations from expected behaviour.
— Cybersecurity measures: Strengthen cybersecurity to protect AI systems from external threats and ensure the integrity and confidentiality of data.
— Training and education: Train and educate personnel on AI systems, their limitations, how to respond to potential issues, and the underlying physical system or process to be able to evaluate the performance of the AI.
— Mitigation plans: Develop comprehensive mitigation plans that outline procedures for handling AI failures or vulnerabilities to minimize their impact on plant operations and safety.

The use of AI at nuclear power plants can provide significant benefits but also brings the need for robust risk assessment, monitoring and mitigation strategies. Identifying vulnerabilities and understanding AI's behaviour attributes are essential to the safe and reliable operation of AI systems in critical environments.

To mitigate or preclude negative consequences in the context of AI at nuclear power plants, it is crucial to be able to monitor the performance of an AI system and have contingency plans if performance degrades. This includes having guardrails to limit AI actions and employing defence in depth strategies. Guardrails are supporting systems that can be applied to an AI system to place constraints on its performance and address potential risks. These systems could be physical or digital and could include:

— Constraints and thresholds: implementing constraints and/or thresholds on the AI system to prevent it from making decisions, or taking actions, that fall outside safe operational boundaries. For instance, an AI system for reactor control should be constrained by strict limits to prevent it from making decisions that could lead to a meltdown.
— Human oversight: maintaining human oversight and control over critical decisions. While AI can provide recommendations and automate processes, there could be a human operator who can intervene and override AI decisions, if necessary. This human–AI collaboration can act as an additional layer of protection. Additionally, an administrative control could be placed on a system to dictate when it could be used.
— Auditing and validation: regularly auditing and validating the AI's behaviour against predefined safety rules and guidelines. If the AI system starts to behave in an unexpected manner or approaches predefined boundaries, it should trigger an alert for human intervention.

Is mitigation available if the system fails?

— Redundancy: Implement redundancy not only in the AI systems but also in the entire control and safety infrastructure. This means having backup systems that can take over in the case of an AI system failure. For example, if an AI system used for emergency shutdown fails, there should be a redundant manual or automated backup system in place.
— Isolation and containment: Ensure that AI systems are isolated from safety critical systems, such as reactor control. This separation prevents a failure in the AI system from directly affecting the core operation of the plant.

— Mitigation plans: Develop comprehensive mitigation plans that outline step-by-step procedures for addressing AI system failures or vulnerabilities. This includes clear roles and responsibilities for human operators and established protocols to restore safe operation.
— Cybersecurity measures: Implement robust cybersecurity measures to protect AI systems from external threats. Regularly update and patch software, use firewalls and employ intrusion detection systems to detect and respond to potential attacks.
— Training and drills: Conduct regular training and simulation drills to prepare plant personnel to respond to AI system failures. This includes understanding the procedures, knowing how to switch to manual control and managing communication during crises.

Mitigating or precluding negative consequences in the context of AI at nuclear power plants involves implementing guardrails to limit AI actions, ensuring human oversight and having a mitigation strategy in place. Mitigation strategies could include redundancy, isolation, emergency response plans, cybersecurity measures, and training to handle AI system failures effectively while maintaining the safety and integrity of nuclear power plant operations.

5.4. GRADED AND RISK INFORMED APPROACHES

Graded or risk informed regulation involves tailoring regulatory requirements based on the level of risk associated with specific activities or systems. When applied to the oversight of AI in the nuclear power industry, it implies a systematic and flexible approach to ensuring safety and security while accommodating advancements in technology. Elements the regulator may consider when applying a graded or risk informed regulatory approach may include the following:

— Risk assessment:
  • Identification of risks: The regulator may conduct a thorough risk assessment to identify potential risks associated with the use of AI in nuclear applications. This includes understanding the consequences of AI failure, potential vulnerabilities and potential impacts on safety and security.
  • Categorization of systems: The regulator may classify AI systems based on their significance to safety and security. High risk systems may include those involved in critical decision making or control processes, while lower risk systems may have less direct impact on safety.
— Graded approach:
  • Tailoring of regulatory requirements: The regulator may apply a graded approach by tailoring regulatory requirements based on the assessed risk levels. High risk AI applications may be subject to more stringent regulations, while lower risk applications may have more flexible requirements.
  • Scalable oversight: The regulator may implement a regulatory framework that allows for scalable oversight. Critical AI applications may undergo more extensive regulatory scrutiny, while less critical applications may have streamlined regulatory processes.
— Performance based regulation:
  • Focus on outcomes: Some regulators may adopt a performance based regulation approach that emphasizes achieving safety and security outcomes rather than prescribing specific technology or methods. This allows for flexibility in incorporating new technologies, like AI, while maintaining a focus on overall risk reduction.
  • Continual improvement: Regulators are also likely to encourage continual improvement in safety and security practices by setting performance goals that can be adapted as technology and industry practices evolve.

— Training and competence:
    - Consideration of human factors: Regulators should recognize the importance of human factors in AI integration and ensure that the personnel responsible for AI systems are adequately trained to understand, operate and address potential issues associated with AI technologies.
    - Competency assessments: Regulators may develop mechanisms for assessing the competence of personnel involved in AI related activities. Such mechanisms may include training programmes, certification processes and ongoing competency assessments to ensure a high level of expertise.
— Periodic safety reviews: Regulators may request periodic safety reviews and assessments to re-evaluate the risk associated with AI technologies, keeping the use of AI aligned with the evolving state of technology and industry best practices.
— Security and ethical considerations:
    - Cybersecurity requirements: Regulators may seek evidence of the integration of cybersecurity requirements into the regulatory framework to address potential vulnerabilities associated with AI systems.
    - Ethical standards: Regulators may incorporate ethical considerations into the AI framework, emphasizing responsible AI development and deployment, which may include addressing issues such as bias, transparency and accountability in the use of AI in decision making processes.

Public communication is an element some regulators emphasize, and they may seek evidence of transparent communication with the public regarding AI in the nuclear power industry. It is possible the regulator may wish to communicate regulatory decisions, safety assessments and risk mitigation strategies relevant to AI to build public trust and confidence in the use of AI technologies in the nuclear power industry. Regulators may seek to encourage public input in the regulatory process, particularly when it comes to high impact AI applications. Soliciting public perspectives may assist regulators to consider a diverse range of views and concerns, contributing to a more comprehensive regulatory framework.

By integrating these elements into a graded approach or risk informed regulatory framework, nuclear regulators can effectively oversee the use of AI in the industry, balancing innovation with safety and security requirements. This approach enables the responsible integration of AI technologies while minimizing risks and ensuring the continued safe operation of nuclear facilities.

5.5. EXPLAINABILITY

Explainability is a critical factor when it comes to using AI at nuclear power plants as the techniques provide information to the user on why the system performs the chosen action. Different AI techniques have different levels of explainability, and it is important for the developer to determine if the technique chosen for an AI solution has a level of explainability commensurate to the application.

Trade-offs exist between model selection and model explainability. Black box models, such as deep neural networks, are highly accurate but challenging to interpret. They make decisions based on complex, nonlinear relationships within the data, which are not readily explainable. In situations where transparency is critical, black box models may be less suitable. Explainable models, like decision trees or linear regression, provide human understandable insights into why a particular decision was made. These models are interpretable and can provide insights into the key features and factors driving their predictions.

There is often a trade-off between accuracy and explainability. Highly accurate models, especially deep learning models, tend to be more complex and less interpretable but achieve higher levels of accuracy. Achieving a high degree of explainability might result in a sacrifice in predictive accuracy. Finding the right balance is essential. In safety critical applications like nuclear power plants, explainability is usually prioritized over predictive accuracy. It is essential to balance the accuracy of AI models with their explainability. The model developer should identify the appropriate modelling framework and what advantageous features that framework provides relative to the level of explainability achievable.

How explainability techniques work and what the output of those techniques implies about the performance of the AI systems should be understood when using the techniques to make a safety claim about the performance of the system. Deliberate choices between accuracy and explainability, selection of suitable model types, alignment with the task, and employment of various methods for understanding and interpreting AI decisions are important factors when considering the development and deployment of an AI system. The specific choice of methods should align with the critical requirements of the nuclear plant's operation and safety.

5.6. REGULATORY ENGAGEMENT

Engaging with regulators when applying AI in nuclear power plants is a crucial step to ensuring compliance with safety and operational standards. It is essential to approach regulators in a timely manner and to address key considerations. Considerations regarding whether the regulator should be engaged about the use of an AI system include the following:

— Does the application affect the licensing basis? It is important to approach the regulator, even if the use of AI is determined to not affect the licensing basis. A proactive approach helps ensure transparency and regulatory awareness. Although engaging the regulator may not result in a change to the licensing basis, it could lead to discussions or guidelines that clarify regulatory expectations for AI applications. The use of AI, even if not affecting the licensing basis, could potentially result in negative inspection findings if it leads to deviations from established safety procedures or standards. Therefore, even in applications that do not affect the licensing basis, communication with the regulator is good practice.
— Does the application affect the safety of the unit? Should the AI application have the potential to affect the safety of the unit, it is imperative to engage with the regulator. Any changes that impact safety need to undergo rigorous regulatory review and approval.
— Does the application affect the safety classification of systems? If an AI system is judged to have an impact on the safety of a nuclear power plant, the underlying supporting infrastructure may require a safety assessment to understand how the AI with safety significance may function if a supporting system fails. For example, if a sensor is used as input to an AI system that is judged to be safety significant, the sensor should be examined to understand how its performance affects the system. If the performance of this sensor is judged to be important to the functioning of the AI system, the safety classification of that sensor should be assessed because the sensor's output may have changed relative to how it was traditionally used.

When approaching a regulator on the deployment of an AI application, it would be advantageous to describe the implementation strategy of the system. Consideration of the implementation strategy early in the development process can assist in producing a fieldable system. Topics of engagement may include the following:

— Supporting infrastructure: Describing the infrastructure supporting the AI application, including hardware, software and cybersecurity measures, ensures that the supporting infrastructure is robust, secure and compliant with regulatory requirements.
— User training: Detailing the training programme for personnel who will interact with the AI system ensures training aligns with regulatory standards and that operators are proficient in using the AI tool effectively and safely.
— Life cycle: Providing a comprehensive description of the AI application's life cycle, including data collection, model development and validation processes, assists in addressing aspects like continuous training versus locked models, performance monitoring and retraining.

— Continuously training versus locked model: It is good practice to explain whether the AI model is continuously updated or if it remains static (locked) and to describe the rationale for the use of this type of model and any safety implications.
— Performance monitoring: Outlining how performance is monitored (including metrics and thresholds) is essential, as is indicating how deviations from acceptable performance are addressed. If performance degradation becomes unacceptable, retraining may be necessary. However, this retraining process should not affect the licensing basis of the tool unless there are significant changes to its functionality or its safety impact.
— Risk assessment and mitigation: This describes the approach used to conduct risk assessment and the means of mitigating any identified risks.
— Data:
    • Characterization of data during the life cycle of the AI system: When detailing how data used for AI model training and operation are characterized during the AI system's life cycle, it is important to address ageing or failing sensors, data quality and any data preprocessing or cleansing. Describing how the AI system reacts to data stream interruptions or failures helps achieve a higher level of confidence in the overall system.
    • Data sharing with regulator versus characterization of data: Defining whether data are shared with the regulator or if data characterization and relevant information are provided instead helps identify how the regulator can access data if necessary for assessments and clearly communicates to the regulator the AI tool's applicability, its limitations and how it aligns with regulatory standards and requirements.
— Regulatory engagement:
    • Collaboration with stakeholders: Regulators are actively engaging with industry stakeholders, including AI developers, operators and end users, to gather insights into AI technologies and applications. Collaborative efforts are helping regulators stay informed about the latest developments and ensure that regulatory approaches towards AI remain relevant and fit for purpose.
    • Feedback mechanisms: Regulators are establishing feedback mechanisms to receive input from industry participants regarding the effectiveness of regulatory regimes, and this dialogue assists regulators in staying agile and responsive to emerging challenges and opportunities posed by AI.

In conclusion, engaging with the regulator with a well prepared and transparent approach may aid in achieving a positive safety finding.

6. SUMMARY AND PATH FORWARD

This publication presented the benefits of AI along with lessons learned and high level considerations that could enable its deployment across several applications in the nuclear power industry. There is significant interest in the new opportunities offered by AI technologies. The experience and examples discussed in this publication show that it is important to facilitate communication among stakeholders to explicate the expected benefits and limitations of an AI application. Through the examples presented in this publication, the role of subject matter experts, utility personnel, regulator representatives and other stakeholders, along with an incremental deployment strategy, is emphasized in building trust in AI and in deploying AI technologies with high reliability.

The AI life cycle discussion brought forward the need to involve multiple stakeholders, to engage interdisciplinary teams to develop a solution, and to consider special infrastructure needs (integrating with legacy plant systems) that might require long term monitoring and maintenance. The degree to which an

AI life cycle should be formalized and adhered to depends on the safety, operational and business risks posed by the application. Parties interested in the deployment of AI should perform a thorough technical, economic and risk assessment before initiating AI development and deployment initiatives. As with other technology deployment efforts in the nuclear power industry, it is expected that a deliberate, planned and methodical deployment of AI applications will yield the highest likelihood of success.

The relevance and heterogeneity of data are important aspects that need to be considered across the AI life cycle. As noted in this publication, the success of AI applications relies largely on the characteristics and proper use of the underlying data. Data quality is key; in particular, it is paramount to have good alignment between the characteristics of the data used in model development and those of the data in the target application scenario. Data management practices are often needed throughout the application's life to maintain quality. While some of these practices are similar to those employed for non-AI applications, specific actions may be needed to address vulnerabilities specific to AI. The publication also recognizes the value of data sharing to enable the development of AI solutions.

The strides regulators and other stakeholders are making to fully understand the implications of AI by identifying potential risks and investing in risk informed frameworks are outlined in this publication. For successful deployment and future operation of AI systems, all elements of the AI, including explainability and its interfaces with the nuclear power plant, as well as the role of a human operator, should be characterized. Staff need to be trained, and supporting hardware and software have to be in place, as needed, to allow the AI to fulfil its intended function. The level of autonomy of the application, particularly in safety critical applications, may drive discussion with the regulator and may dictate the degree of scrutiny brought to the specific AI implementation.

In summary, this publication has discussed different and important aspects of AI technologies in the nuclear power industry. Some progress has been made with respect to the deployment of AI technologies in nuclear power plants to achieve economic and operational benefits. Still, there are several challenges and opportunities to be realized for successful and safe usage of AI across the nuclear power industry.

REFERENCES

[1] NUCLEAR REGULATORY COMMISSION, Human-System Interface Design Review Guidelines, Revision 3, Rep. NUREG-0700, Office of Nuclear Regulatory Research, Washington, DC (2020).

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

[3] HALL, A., AGARWAL, V., Barriers to adopting artificial intelligence and machine-learning technologies in nuclear power, Prog. Nucl. Energy **175** (2024) 105295.

[4] ANTONELLO, F., BUONGIORNO, J., ZIO, E., Physics informed neural networks for surrogate modeling of accidental scenarios in nuclear power plants, Nucl. Eng. Technol. **55** 9 (2023) 3409.

[5] SUN, L., GAO, H., PAN, S., WANG, J.-X., Surrogate modeling for fluid flows based on physics-constrained deep learning without simulation data, Comput. Methods Appl. Mech. Eng. **361** (2020) 112732.

[6] GURRAPU, S., KULKARNI, A., HUANG, L., LOURENTZOU, I., BATARSEH, F.A., Rationalization for explainable NLP: A survey, Front. Artif. Intell. **6** (2023).

[7] MANJUNATHA, K.A., AGARWAL, V., Multi-kernel based adaptive support vector machine for scalable predictive maintenance, Proc. Annu. Progn. Health Manage. Soc. Conf. **14** 1 (2022).

[8] RAMUHALLI, P., WALKER, C.M., AGARWAL, V., LYBECK, N., TAYLOR, M., "Nuclear Power Prognostic Model Assessment for Component Health Monitoring" Proc. 12th Nuclear Plant Instrumentation, Control and Human-Machine Interface Technologies Conf. (2021) 976–986.

[9] EL-HALYM, H.A.A., MAHMOUD, I.I., HABIB, S., Proposed hardware architectures of particle filter for object tracking, EURASIP J. Adv. Signal Process. **2012** (2012) 1.

[10] ELECTRIC POWER RESEARCH INSTITUTE, AI-Assisted Analysis of Ultrasonic Inspections, Rep. 3002023718, EPRI, Palo Alto, CA (2022).

[11] ELECTRIC POWER RESEARCH INSTITUTE, Quick Insight Brief: Using Artificial Intelligence to Maximize the Benefits of Drones for Nuclear Power Plants, Rep. 3002023930, EPRI, Palo Alto, CA (2022).
[12] ELECTRIC POWER RESEARCH INSTITUTE, Quick Insight Brief: Leveraging Artificial Intelligence for Nondestructive Evaluation, Rep. 3002021074, EPRI, Palo Alto, CA (2021).
[13] ELECTRIC POWER RESEARCH INSTITUTE, Automated Analysis of Remote Visual Inspection of Containment Buildings, Rep. 3002018419, EPRI, Palo Alto, CA (2020).
[14] ZARZECZNY, W., et al., "Ontario Power Generation's monitoring and diagnostic centre", Proc. 39th Annual CNS Conf. and 43rd Annual CNS/CNA Student Conf., Ottawa, ON (2019).
[15] PONCIROLI, R., MOISEYTSEVA, V., DAVE, A.J., NGUYEN, T.N., VILIM, R.B., Design and Prototyping of Advanced Control Systems for Advanced Reactors Operating in the Future Electric Grid, Rep. ANL/NSE-23/48, Argonne National Laboratory, Lemont, IL (2023).
[16] VILIM, R., et al., "Computerized operator support system and human performance in the control room", Proc. 10th Intl. Topical Mtg on Nuclear Plant Instrumentation, Control, and Human-Machine Interface Technologies, San Francisco, CA (2017).
[17] ELECTRIC POWER RESEARCH INSTITUTE, Automating Corrective Action Programs in the Nuclear Industry, Rep. 3002023821, EPRI, Palo Alto, CA (2022).
[18] GRIET, M., VENTURINI, V., FLURY, C., "Sistema avanzado de alarmas para el reactor RA6C", Centro Atómico Bariloche, Comisión Nacional de Energía Atómica, Argentina (2016), https://inis.iaea.org/collection/NCLCollectionStore/_Public/51/100/51100291.pdf.
[19] ENGINEERING EQUIPMENT AND MATERIALS USERS' ASSOCIATION, Alarm Systems — A Guide to Design, Management and Procurement, EEMUA publication 191, EEMUA, London (2013).
[20] SANCHEZ-PI, N., LEME, L.A.P., GARCIA, A.C.B., Intelligent agents for alarm management in petroleum ambient, J. Intell. Fuzzy Syst. **28** 1 (2015) 43.
[21] ELNOKITY, O., MAHMOUD, I.I., REFAI, M.K., FARAHAT, H.M., ANN based sensor faults detection, isolation, and reading estimates – SFDIRE: Applied in a nuclear process, Ann. Nucl. Energy **49** (2012) 131.
[22] BORELLA, A., ROSSA, R., ZAIOUN, H., Determination of ^{239}Pu content in spent fuel with the SINRD technique by using artificial and natural neural networks, ESARDA Bull. **58** (2019).
[23] MISHRA, V., BRANGER, E., ELTER, Z., GRAPE, S., JANSSON, P., "Comparison of different supervised machine learning algorithms to predict PWR spent fuel parameters", Institute of Nuclear Materials Management, Mount Laure, NJ (2021).
[24] AL-DBISSI, M., ROSSA, R., BORELLA, A., PÁZSIT, I., VINAI, P., Identification of diversions in spent PWR fuel assemblies by PDET signatures using artificial neural networks (ANNs), Ann. Nucl. Energy **193** (2023).
[25] GRAPE, S., BRANGER, E., ELTER, Z., PÖDER BALKESTÅHL, L., Determination of spent nuclear fuel parameters using modelled signatures from non-destructive assay and random forest regression, Nucl. Instrum. Methods Phys. Res. Sect. A **969** (2020).
[26] SHOMAN, N., CIPITI, B., "Advances in machine learning for safeguarding a PUREX reprocessing facility", SAND2020-5394C, Sandia National Laboratories, New Mexico (2020).
[27] LALOY, E., et al., Probabilistic radwaste characterization: Findings of a multi-method multi-mockup exercise using interpolation-based surrogate efficiencies, Ann. Nucl. Energy **194** (2023) 110065.
[28] PARK, N., et al., "Data synthesis based on generative adversarial networks", arXiv (2018), https://doi.org/10.48550/arXiv.1806.03384
[29] RADAIDEH, M.I., et al., Physics-informed reinforcement learning optimization of nuclear assembly design, Nucl. Eng. Des. **372** (2021) 110966.
[30] HRYNIEWICKI, M.K., STRAPP, J., "Improving Nuclear Unit Outage Scheduling with Artificial Intelligence", Power Engineering (2019), https://www.power-eng.com/nuclear/improving-nuclear-unit-outage-scheduling-with-artificial-intelligence
[31] INSTITUTE OF NUCLEAR POWER OPERATORS, Performance Continuum Manual, Rev. 3.1, INPO, Atlanta, GA (2023).
[32] GRUENWALD, J.T., NISTOR, J., TUSAR, J., Powering our nuclear fleet with artificial intelligence, Nucl. News **65** 2 (2022) 50, https://www.ans.org/news/article-3633/powering-our-nuclear-fleet-with-artificial-intelligence
[33] HONNEY, T., "AI tools deployed at two constellation nuclear plants", Nuclear Engineering International (2024), https://www.neimagazine.com/news/ai-tools-deployed-at-two-constellation-nuclear-plants-11640288

[34] ELECTRIC POWER RESEARCH INSTITUTE, AI Tool Developed by EPRI Significantly Cuts Analysis Time in U.S. Nuclear Plant Field Trial: Data Analysis Takes Four Hours Compared to Four Days Without Artificial Intelligence, Rep. 3002025510, EPRI, Palo Alto, CA (2022).
[35] ELECTRIC POWER RESEARCH INSTITUTE, Materials Reliability Program: Guideline for Nondestructive Examination of Reactor Vessel Upper Head Penetrations, Revision 1 (MRP-384), Rep. 3002017288, EPRI, Palo Alto, CA (2019).
[36] VIRKKUNEN, I., et al., ENIQ Recommended Practice 13: Qualification of Non-Destructive Testing Systems That Make Use of Machine Learning, Rep. 65, European Network for Inspection and Qualification (2021).
[37] INTERNATIONAL ATOMIC ENERGY AGENCY, Configuration Management in Nuclear Power Plants, IAEA-TECDOC-1335, IAEA, Vienna (2003).
[38] INTERNATIONAL ATOMIC ENERGY AGENCY, Introduction to Systems Engineering for the Instrumentation and Control of Nuclear Facilities, IAEA Nuclear Energy Series No. NR-T-2.14, IAEA, Vienna (2022).
[39] INTERNATIONAL ORGANIZATION FOR STANDARDIZATION, INTERNATIONAL ELECTROTECHNICAL COMMISSION, INSTITUTE OF ELECTRICAL AND ELECTRONICS ENGINEERS, Systems and Software Engineering — System Lifecycle Processes, ISO/IEC/IEEE 15288:2023, ISO, Geneva (2023).
[40] AKAIKE, H., "Akaike's Information Criterion", International Encyclopedia of Statistical Science (LOVRIC, M., Ed), Springer, Berlin (2011) 25.
[41] NEATH, A.A., CAVANAUGH, J.E., "The Bayesian information criterion: Background, derivation, and applications", Wiley Interdiscip. Rev. Comput. Stat. **4** 2 (2012) 199.
[42] HANSEN, M.H., YU, B., Model selection and the principle of minimum description length, J. Am. Stat. Assoc. **96** 454 (2001) 746.
[43] FUSHIKI, T., Estimation of prediction error by using K-fold cross-validation, Stat. Comput. **21** 2 (2011) 137.
[44] LEE, J.D., "Human factors and ergonomics in automation design", Handbook of Human Factors and Ergonomics, 3rd edn, John Wiley & Sons, Hoboken, NJ (2006).
[45] ENDSLEY, M., KIRIS, E., The out-of-the-loop performance problem and level of control in automation, Hum. Factors **37** 2 (1995) 381.
[46] SHNEIDERMAN, B., Human-Centered AI, Oxford University Press, Oxford (2021).
[47] RIEDL, M., Human-centered artificial intelligence and machine learning, Hum. Behav. Emerging Technol. **1** 1 (2019) 33.
[48] XU, W., Toward human-centered AI: A perspective from human-computer interaction, Interactions **26** 4 (2019) 42.
[49] DAFOE, A., et al., "Open problems in cooperative AI", arXiv (2020), https://doi.org/10.48550/arXiv.2012.08630
[50] DAFOE, A., et al., "Cooperative AI: Machines must learn to find common ground", Nature **593** 7857 (2021) 33.
[51] INSTITUTE OF ELECTRICAL AND ELECTRONICS ENGINEERS, Guide for Human Factors Engineering for the Validation of System Designs and Integrated Systems Operations at Nuclear Facilities, IEEE Std 2411-2021, IEEE, New York (2021).
[52] NUCLEAR REGULATORY COMMISSION, Guidance for the Review of Changes to Human Actions, Rev. 1., Rep. NUREG-1764, Office of Nuclear Regulatory Research, Washington, DC (2007).
[53] ALBERTI, A.L., AGARWAL, V., GUTOWSKA, I., PALMER, C.J., DE OLIVEIRA, C.R.E., Automation levels for nuclear reactor operations: A revised perspective, Prog. Nucl. Energy **157** (2023) 104559.
[54] NUCLEAR REGULATORY COMMISSION, Human Factors Engineering Program Review Model, Rev. 3, Rep. NUREG-0711, Office of Nuclear Reactor Regulation, Washington, DC (2012).
[55] WALKER, C.M., AGARWAL, V., APPIAH, R., "Development of a scalable, risk-informed, predictive maintenance cloud based strategy at nuclear power plants", Proc. 13th Nuclear Plant Instrumentation, Control and Human-Machine Interface Technologies Conf., Knoxville, TN (2023).
[56] JONES, E., JIA, R., RAGHUNATHAN, A., LIANG, P., "Robust encodings: A framework for combating adversarial typos", Proc. 58th Annual Mtg Association for Computational Linguistics, Virtual event (2020) 2752–2765.
[57] GOPALAKRISHNAN, S., MARZI, Z., MADHOW, U., PEDARSANI, R., "Combating adversarial attacks using sparse representations", 6th Int. Conf. Learning Representations — Workshop Track Proceedings, Vancouver, BC (2018).
[58] SHOKRI, R., STRONATI, M., SONG, C., SHMATIKOV, V., "Membership inference attacks against machine learning models", Proc., 2017 IEEE Symposium on Security and Privacy, IEEE (2017) 3–18.
[59] BROWN, T., MANN, B., RYDER, N., SUBBIAH, M., KAPLAN, J., "Language models are few-shot learners", Proc. Adv. Neural Inf. Process. Syst. **33** (2020).

[60] INTERNATIONAL ATOMIC ENERGY AGENCY, Computer Security Techniques for Nuclear Facilities, IAEA Nuclear Security Series No. 17-T (Rev. 1), IAEA, Vienna (2021).
[61] Mitre ATLAS,
https://atlas.mitre.org
[62] HINES, J., GARVEY, J., GARVEY, D., SEIBERT, R., Technical Review of On-line Monitoring Techniques for Performance Assessment: Limiting Case Studies, Rep. NUREG/CR-6895 Volume 3, Oak Ridge National Laboratory, TN (2008).
[63] INTERNATIONAL ATOMIC ENERGY AGENCY, Security of Nuclear Information, IAEA Nuclear Security Series No. 23-G, IAEA, Vienna (2015).
[64] METZEN, J.H., KUMAR, M.C., BROX T., FISCHER, V., "Universal adversarial perturbations against semantic image segmentation", Proc., 2017 IEEE Int. Conf. Computer Vision (2017) 2774–2783.
[65] HU, W., TAN, Y., "Generating adversarial malware examples for black-box attacks based on GAN", arXiv (2017),
https://doi.org/10.48550/arXiv.1702.05983
[66] BAI, T., LUO, J., ZHAO, J., WEN, B., WANG, Q., "Recent advances in adversarial training for adversarial robustness", Proc., Int. Joint Conf. Artificial Intelligence (2021).
[67] YANG, Q., LIU, Y., CHEN, T., TONG, Y., "Federated machine learning: Concept and applications", ACM Trans. Intell. Syst. Technol. **10** 2 (2019) 1–19.
[68] INTERNATIONAL ELECTROTECHNICAL COMMISSION, Nuclear Power Plants — Instrumentation, Control and Electrical Power Systems — Security Controls, IEC 63096:2020, IEC, Geneva (2020).
[69] INTERNATIONAL ORGANIZATION FOR STANDARDIZATION, INTERNATIONAL ELECTROTECHNICAL COMMISSION, Software Engineering — Systems and Software Quality Requirements and Evaluation (SQuaRE) — Quality Model for AI Systems, ISO/IEC 25059:2023, ISO/IEC, Geneva (2023).
[70] INSTITUTE OF ELECTRICAL AND ELECTRONICS ENGINEERS, IEEE Guide for Human Factors Engineering for the Validation of System Designs and Integrated Systems Operations at Nuclear Facilities, IEEE, New York (2022).
[71] AGBAJI, D., LUND, B., MANNURU, N.R., "Perceptions of the fourth industrial revolution and artificial intelligence impact on society", arXiv (2023),
https://doi.org/10.48550/arXiv.2308.02030
[72] KOTTER, J.P., Leading Change, Harvard Business School Press, Boston, MA (1996).
[73] LEWIN, K., "Group decision and social change", Readings in Social Psychology (MACCOBY, E.E., NEWCOMB, T.M., HARTLEY, E.L., Eds), Holt, Rinehart & Winston, New York (1958) 197–211.
[74] HIATT, J., ADKAR: A Model for Change in Business, Government and Our Community, Prosci Learning Center Publications (2006).
[75] DALE CARNEGIE & ASSOCIATES, Preparing People for Success with Generative AI (2023),
https://www.dalecarnegie.com/en/resources
[76] UNITED STATES DEPARTMENT OF ENERGY, AI Risk Management Playbook (2022),
https://www.energy.gov/ai/doe-ai-risk-management-playbook-airmp
[77] INTERNATIONAL ORGANIZATION FOR STANDARDIZATION, INTERNATIONAL ELECTROTECHNICAL COMMISSION, Information Technology — Artificial Intelligence — Data Life Cycle Framework, ISO/IEC 8183:2023, ISO/IEC, Geneva (2023).
[78] INTERNATIONAL ORGANIZATION FOR STANDARDIZATION, INTERNATIONAL ELECTROTECHNICAL COMMISSION, Artificial Intelligence —Data Quality for Analytics and Machine Learning, Part 1: Overview, Terminology and Examples, ISO/IEC 5259-1:2024, ISO/IEC, Geneva (2024).
[79] TYYSTJÄRVI, T., VIRKKUNEN, I., FRIDOLF, P., ROSELL, A., BARSOUM, Z., Automated defect detection in digital radiography of aerospace welds using deep learning, Weld. World **66** 4 (2022) 643.
[80] ELECTRIC POWER RESEARCH INSTITUTE, AI-Assisted Analysis of Ultrasonic Inspections, Rep. 3002023718, EPRI, Palo Alto, CA (2022).
[81] NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY, Recommendation for Key Management, NIST Special Publication 800-57 Part 1 – General, NIST, Gaithersburg, MD (2020).
[82] CANADIAN NUCLEAR SAFETY COMMISSION, UNITED KINGDOM OFFICE FOR NUCLEAR REGULATION, NUCLEAR REGULATORY COMMISSION, Considerations for Developing Artificial Intelligence Systems in Nuclear Applications, CANUKUS (2024).
[83] UNITED KINGDOM OFFICE FOR NUCLEAR REGULATION, "Outcomes of nuclear AI regulatory sandbox pilot published" (2023),
https://www.onr.org.uk/news/all-news/2023/11/outcomes-of-nuclear-ai-regulatory-sandbox-pilot-published

[84] NUCLEAR REGULATORY COMMISSION, Project Plan for the U.S. Nuclear Regulatory Commission Artificial Intelligence Strategic Plan Fiscal Years 2023–2027, Rev. 0, United States NRC, Washington, DC (2023).
[85] NUCLEAR REGULATORY COMMISSION, "Data science and artificial intelligence regulatory applications workshops" (2024), https://www.nrc.gov/public-involve/conference-symposia/data-science-ai-reg-workshops.html
[86] NUCLEAR REGULATORY COMMISSION, "Artificial intelligence" (2024), https://www.nrc.gov/about-nrc/plans-performance/artificial-intelligence.html
[87] NUCLEAR REGULATORY COMMISSION, Advancing Use of Artificial Intelligence at the U.S. Nuclear Regulatory Commission, Rep. ML23303A143, United States NRC, Washington, DC (2023).
[88] NUCLEAR REGULATORY COMMISSION, Advancing Use of Artificial Intelligence at the U.S. Nuclear Regulatory Commission, Rep. SECY-24-0035, United States NRC, Washington, DC (2024).

Annex

CHINA NUCLEAR POWER ENGINEERING COMPANY FRAMEWORK OF INTELLIGENT NUCLEAR POWER PLANTS

From a long term perspective, the China Nuclear Power Engineering Company has considered coordinating the following to develop intelligent nuclear power plants:

— Artificial intelligence (AI) applications: choosing appropriate intelligent algorithms (sometimes combining traditional analysis and intelligent computing) to develop AI applications that match the functional requirements of nuclear power plant users;
— Data: fully mining the existing data of nuclear power plants and generating additional effective data to ensure the development and validation of AI applications;
— Platform: planning overall platform development, with high scalability, accommodating various intelligent applications, effectively promoting high data integration and reducing project deployment costs.

A–1. FUNCTIONAL OBJECTIVES OF INTELLIGENT NUCLEAR POWER PLANTS

In the near to medium term, the application of AI technology can contribute to the following functional objectives:

— Intelligent operation: reducing the workload of operator monitoring and avoiding human errors through real-time monitoring and early warning of plant operating status; reducing the workload of on-site operations and maintenance personnel and improving efficiency through intelligent monitoring and analysis of equipment performance; optimizing the execution cycle of regular tests by accurately understanding the equipment status.
— Intelligent equipment management and maintenance: using technologies such as equipment status monitoring, fault diagnosis and remaining useful life prediction to ascertain the operational status of structures, systems and components, to make condition based maintenance decisions, to provide early warning before equipment failures, and to use structures, systems and components more economically.
— Intelligent hazard prevention: improving the detection and identification capabilities of hazardous factors in and around nuclear power plants and relying less on manual inspections.
— Intelligent emergency response: providing nuclear emergency decision making support based on a massive amount of information (nuclear accidents have wide range and long duration of impacts) in a short period of time.

A–2. SYSTEM ARCHITECTURE OF INTELLIGENT NUCLEAR POWER PLANTS

To accommodate intelligent applications, a new, reliable and flexible system architecture is proposed for providing nuclear power plant operational data. The architecture does not affect the normal operation of the plant, and the data are used in a secure and efficient way.

As shown in Fig. A–1, operators control the nuclear power plant in the main control room and the distributed control system executes the commands. The intelligent operation assistance system receives and analyses the plant data from the distributed control system. Data analytics are presented to operators

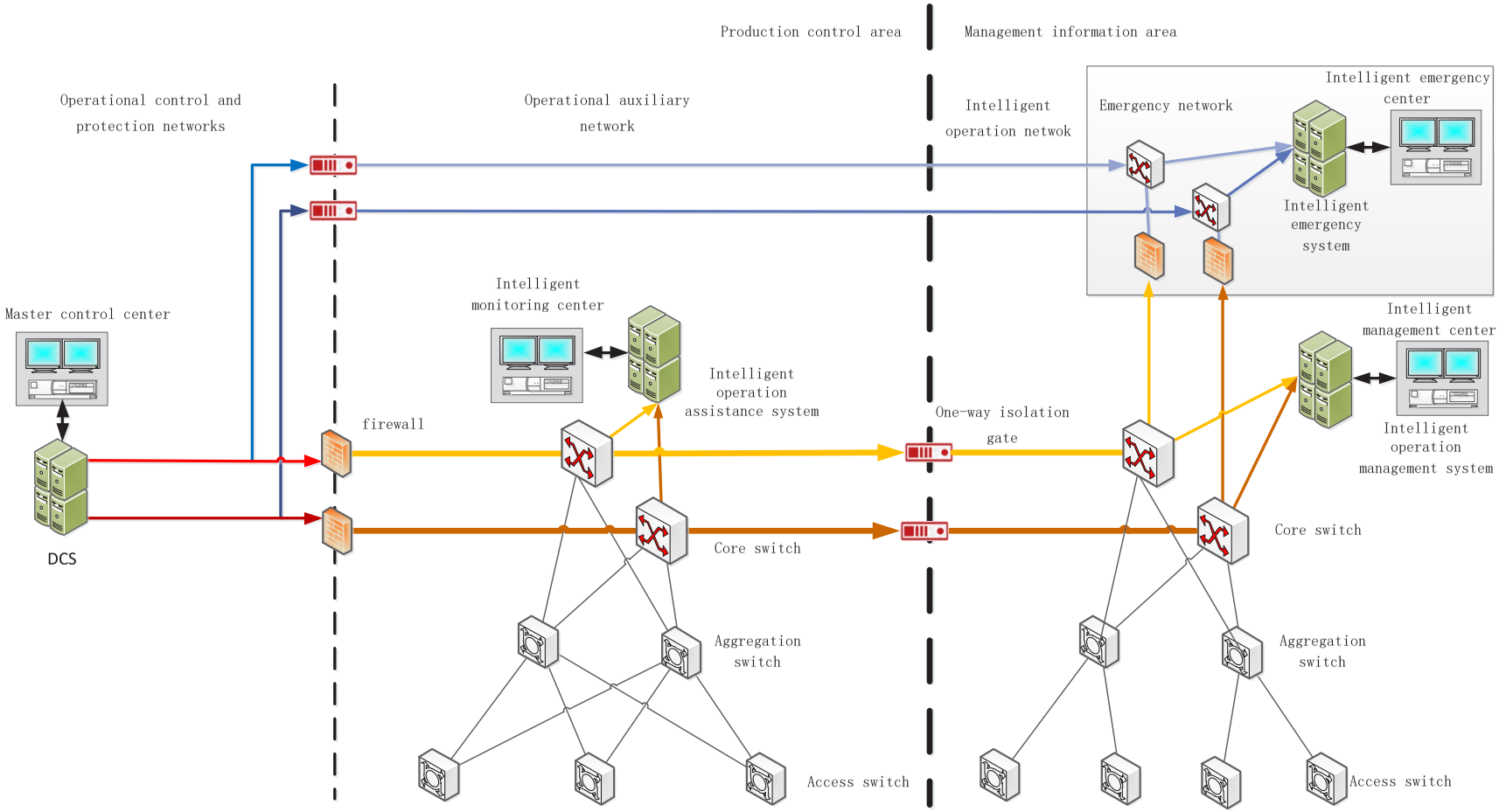


FIG. A–1. Example of possible architecture of intelligent nuclear power plant.

through the intelligent monitoring centre and assist operators in making operational decisions. The intelligent emergency system receives plant data in one direction to execute emergency related functions, and data analytics are used by the intelligent emergency centre to make decisions. The intelligent operation management system deployed in the management information area receives plant data from the production control area in one direction. This system architecture and data transmission method can reduce investment in platform construction and, later, operation and maintenance.

A–3. PLATFORM FOR AI APPLICATIONS

An integrated information platform is proposed for AI applications and would be a core part of the infrastructure for intelligent nuclear power plants. It would provide unified data collection or access services, resources for applications (computing power, storage, memory, etc.), network security protection and other functions. The platform features a streamlined architecture that efficiently channels plant data to a wide range of applications. It would also be able to efficiently deploy various AI applications and comprehensively improve the plant operation. The platform is capable of further processing unit-level operational and management data, aggregating them, and then transmitting the information to the plant-level management system to assist with plant management.

A–4. AI APPLICATIONS

The China Nuclear Power Engineering Company has developed various AI applications for the health management of equipment and systems, decision support for operations and maintenance, and other actions. Some of the typical applications were discussed in Section 2.2. The following applications have been deployed in commercial nuclear power plants and have generated positive outcomes:

— Internet of Things based smart construction site. This application features earthwork real-time monitoring and management, anti-collision monitoring of lifting equipment, integrated management of Internet of Things devices, real-time monitoring of operations in confined space, safety hazard monitoring, construction efficiency monitoring, and intelligent security inspection.
— Intelligent decision making system (Fuxi intelligent decision making system). This system is used for five scenarios: customized design of nuclear power systems, intelligent monitoring and decision making for equipment maintenance, intelligent patrol inspection in main control room during plant operation, expert collaborative decision making, and intelligent decision making in emergency evacuation planning.

A–5. CHALLENGES

Challenges still exist for developing an intelligent nuclear power plant. Some of the AI applications need a large amount of operational data to train models and validate their outcomes, and the lack of data or feasible simulations makes it difficult to generate trustworthy models. The validation of some AI applications is still unachievable before being delivered to utilities. The platform of AI applications should have an upgradable structure and be able to accommodate future, more advanced AI applications. The regulatory framework for applying AI in nuclear power plants is still being established.

GLOSSARY

air gapping. The process of isolating a computer or network from other computing devices and networks, including the Internet, to enhance cybersecurity.

autonomy. The degree to which the artificial intelligence interacts with humans, with full autonomy being the level at which there is no human interaction.

bare metal deployment. The process of installing software directly onto physical machines, without using virtualization tools such as hypervisors.

container (software). A self-contained, portable unit consisting of an application bundled together with all its necessary dependencies so it can run reliably in any environment.

convolutional neural network. A network that learns by itself via increasing levels of filters.

data bias. Systematic errors or skewed distributions in the data that make the data favour one class or category over another.

data corruption. Errors that are unintentionally introduced during writing, reading, storage, transmission or processing of data that may render the data unreadable or otherwise compromised.

data distribution. The way data points are spread across the relevant ranges, categories or classes.

data drift. Unexpected, undocumented or uncontrolled changes over time in data characteristics, such as data distribution.

data governance. The overall practices and processes for suitable data management, including provisions for how data are properly recorded, processed, retained, used and distributed throughout their life cycle.

data integrity. The property of data maintaining their quality throughout their life cycle, not deteriorating due to unauthorized agents or processes (intentional or unintentional).

data leakage. A failure in proper data partitioning where data or information about data used in training and validation are intentionally or unintentionally included in test sets.

data poisoning. The intentional injection of malicious data into artificial intelligence or machine learning models either during training and development or during inference at deployment.

data quality. The overall condition of the combined set of relevant data characteristics that make a dataset suitable for a given application.

data scope. The context of and range covered by a dataset.

data security. The overall practices and effects of ensuring the maintenance of proper data integrity, confidentiality and availability.

data sharing. The practice of data owners making data available to a broader community.

deep learning. A subset of machine learning in which there are multiple layers in the neural network.

dynamic models. Applications that include a feedback loop where the deployed model continuously learns from current data and changes as required. In this case, the current data are actively and continuously being included in the training of the model, which is periodically and automatically updated based on the new information.

edge computing. A computing paradigm that brings computation close to the location where it is needed. This may be advantageous with regards to latency and network bandwidth requirements.

explainability. The ability of a human to understand and trust the results of the output of an artificial intelligence agent and its expected impact and biases.

field data. Data collected from sensors or humans within the operation environment of a nuclear facility.

guardrails. Constraints, either physical or digital, placed on a system that define the operating regime of or bound the performance of the artificial intelligence system.

hypervisor. A software layer to manage and allocate resources to virtual machines.

inference. The process of using a trained artificial intelligence model to make predictions or decisions on new, unseen data.

laboratory data. Data gathered in controlled settings to conduct experiments and physical simulations.

natural language processing. A subfield of artificial intelligence focused on the artificial intelligence agent's ability to understand and communicate in human language.

on-premises computing. The deployment of computing or other related services on the physical premises of an organization.

open source data. Publicly available data.

private cloud (computing). A type of computing where services are provided on private shared IT resources and can often allow for on-demand use.

public cloud (computing). A type of computing in which an external service provider provides the resource, such as computing or other related resources, over the Internet. Many pricing schemes are available, including subscription, on demand, or pay per use.

virtual machine. A software based emulation of physical computers, allowing the handling of computing hardware as shared pools. A virtual machine allows multiple users to share the same hardware, with each virtual machine behaving as an isolated system with its own processing power, memory, network interface and storage.

ABBREVIATIONS

AI	artificial intelligence
ANN	artificial neural network
IP	intellectual property
LLM	large language model
ML	machine learning
NLP	natural language processing
NRC	Nuclear Regulatory Commission
O&M	operations and maintenance
QA	quality assurance

CONTRIBUTORS TO DRAFTING AND REVIEW

Abdel-Khalik, H.	Purdue University, United States of America
Agarwal, V.	Idaho National Laboratory, United States of America
Al-dbissi, M.	Chalmers University of Technology, Sweden
Al-Rashdan, A.Y.	Idaho National Laboratory, United States of America
Andrachek, J.	Pressurized Water Reactor Owners Group, United States of America
Aversano, G.	Kinectrics, Canada
Ayalasomayajula, S.	Nuclear Promise X, Canada
Betancourt, L.	Nuclear Regulatory Commission, United States of America
Boring, R.	Idaho National Laboratory, United States of America
Briquez, B.	Westinghouse, Spain
Busquim, R.	International Atomic Energy Agency
Cancila, D.	Commissariat à l'énergie atomique et aux énergies alternatives, France
Carter, C.E.	Utilities Service Alliance, United States of America
Chu, J.	China Nuclear Power Engineering, China
Comeaux, K.	Institute of Nuclear Power Operations, United States of America
Cox, N.	NuScale Power, United States of America
Deng, S.	China Nuclear Power Engineering, China
Dennis, M.	Nuclear Regulatory Commission, United States of America
Desaulniers, D.	Nuclear Regulatory Commission, United States of America
El Bouzidi, S.	Canadian Nuclear Laboratories Ltd, Canada
Foster-Roman, D.	Ontario Power Generation, Canada
Gadallah, I.M.	Egyptian Atomic Energy Authority, Egypt
Golightly, C.	Energy Northwest, United States of America
Gruenwald, J.T.	Blue Wave AI Labs, United States of America
Guerra-O'Hanlon, S.	Adelard, NCC Group, United Kingdom
Hall, M.	Evergy, United States of America
Hathaway, A.	Nuclear Regulatory Commission, United States of America
Hewes, M.	International Atomic Energy Agency
Highland, B.D.	Energy Northwest, United States of America

Kim, J.	Chosun University, Republic of Korea
Lagarde, J.	Metroscope, France
Lee, K.	Canadian Nuclear Safety Commission, Canada
Lenci, G.	Metroscope Inc., United States of America
Li, J.	Tsinghua University, China
Li, M.	China Nuclear Power Engineering, China
Li, W.	China Nuclear Power Engineering, China
Lynde, J.	Pressurized Water Reactor Owners Group, United States of America
Movassat, M.	Ontario Power Generation, Canada
Murray, I.	World Association of Nuclear Operators, United Kingdom
Nangia, B.	Nuclear Promise X, Canada
Nieto, L.A.	Comisión Nacional de Energía Atómica, Argentina
Pereira de M. Martins, G.	Electronuclear, Brazil
Powell, M.	Pressurized Water Reactor Owners Group, United States of America
Priestman, K.	Nuclear Promise X, Canada
Prinja, N.	Prinja and Partners, United Kingdom
Seuaciuc-Osorio, T.	Electric Power Research Institute, United States of America
Seymour, J.	Nuclear Regulatory Commission, United States of America
Sladek, J.	Canadian Nuclear Safety Commission, Canada
Smith, P.	Lancaster University, United Kingdom
Venturini, V.	Comisión Nacional de Energía Atómica, Argentina
Vilim, R.	Argonne National Laboratory, United States of America
Virkkunen, I.	Aalto University, Finland
Walker, C.M.	Idaho National Laboratory, United States of America
Xu, S.	China Nuclear Power Engineering, China
Yadav, V.	Idaho National Laboratory, United States of America
Yao, W.	China Nuclear Power Engineering, China
Yu, D.	China Nuclear Power Engineering, China
Zeng, Z.C.	Canadian Nuclear Safety Commission, Canada
Zhang, L.	China Nuclear Power Engineering, China

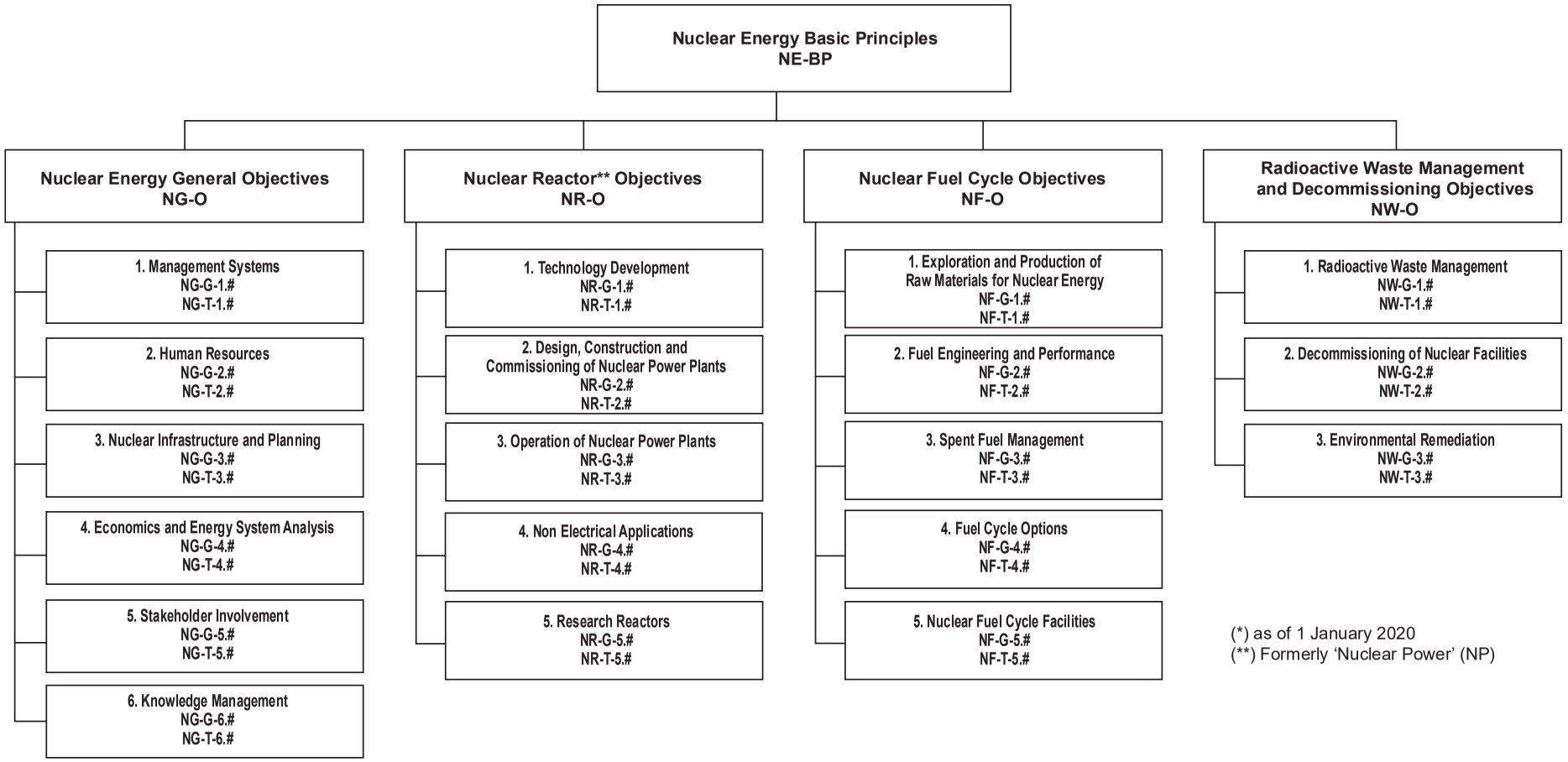
Consultants Meetings

Vienna, Austria: 5–7 July 2023, 4–8 December 2023, 29 July–2 August 2024

Technical Meeting

Rockville, Maryland, United States of America: 18–21 March 2024

Structure of the IAEA Nuclear Energy Series*



(*) as of 1 January 2020
(**) Formerly 'Nuclear Power' (NP)

Key

BP: Basic Principles
O: Objectives
G: Guides and Methodologies
T: Technical Reports
Nos 1–6: Topic designations
#: Guide or Report number

Examples

NG-G-3.1: Nuclear Energy General (**NG**), Guides and Methodologies (**G**), Nuclear Infrastructure and Planning (topic **3**), **#1**
NR-T-5.4: Nuclear Reactors (**NR**), Technical Report (**T**), Research Reactors (topic **5**), **#4**
NF-T-3.6: Nuclear Fuel (**NF**), Technical Report (**T**), Spent Fuel Management (topic **3**), **#6**
NW-G-1.1: Radioactive Waste Management and Decommissioning (**NW**), Guides and Methodologies (**G**), Radioactive Waste Management (topic **1**) **#1**

IAEA
International Atomic Energy Agency

CONTACT IAEA PUBLISHING

Feedback on IAEA publications may be given via the on-line form available at:
www.iaea.org/publications/feedback

This form may also be used to report safety issues or environmental queries concerning IAEA publications.

Alternatively, contact IAEA Publishing:

Publishing Section
International Atomic Energy Agency
Vienna International Centre, PO Box 100, 1400 Vienna, Austria
Telephone: +43 1 2600 22529 or 22530
Email: sales.publications@iaea.org
www.iaea.org/publications

Priced and unpriced IAEA publications may be ordered directly from the IAEA.

ORDERING LOCALLY

Priced IAEA publications may be purchased from regional distributors and from major local booksellers.