INL/RPT-24-79471

4th Big Data for Nuclear Power Plants Workshop 2023

Workshop Report

JULY 2024

Nancy Lybeck, Cody Walker, William Chuirazzi, Mathew Swisher, Katya Le Blanc, Matthew Anderson, Vaibhav Yadav, Samuel Bays

Idaho National Laboratory

Carol Smidts, Xiaoxu Diao, Marat Khafizov, Leonardo Moraes, Praneeth Kandlakunta

The Ohio State University

Idaho National Laboratory

Battelle Energy Alliance manages INL for the U.S. Department of Energy's Office of Nuclear Energy

Page intentionally left blank

INL/RPT-24-79471

4th Big Data for Nuclear Power Plants Workshop 2023

Workshop Report

Nancy Lybeck, Cody Walker, William Chuirazzi, Mathew Swisher, Katya Le Blanc, Matthew Anderson, Vaibhav Yadav, Samuel Bays
Idaho National Laboratory
Carol Smidts, Xiaoxu Diao, Marat Khafizov, Leonardo Moraes, Praneeth Kandlakunta
The Onio State University

July 2024

Idaho National Laboratory
Idaho Falls, Idaho 83415

http://www.inl.gov

Prepared for the
U.S. Department of Energy
Office of Nuclear Energy
Under DOE Idaho Operations Office
Contract DE-AC07-05ID14517

Page intentionally left blank

ABSTRACT

The Ohio State University and Idaho National Laboratory organized the 4th Big Data for Nuclear Power Plants Workshop, held in November 2023 in Columbus, Ohio. The workshop topics were chosen to better elucidate the challenges and gaps that must be addressed to maximize data's impact on the nuclear industry, as well as the associated applications and risks. Discussions centered around six specific application areas: operation and maintenance, machine learning (ML) in nuclear materials and advanced manufacturing (AM), cybersecurity, high-performance computing (HPC) and massive computation, big data and digital twins (DTs), and nuclear non-proliferation.

A diverse range of opportunities, challenges, and risks were identified within these six focus areas. Some common themes emerged, such as the importance of data integrity, quality, coverage, privacy, and traceability. Big data and artificial intelligence (AI)/ML tools can be leveraged to reduce costs, optimize human tasking, and reduce human error across various application areas. But for the nuclear industry to benefit from big data and advanced analytical capabilities, certain challenges and risks must be addressed, such as data privacy, model reliability, and computational resource availability. Learning from other industries that have already successfully implemented big data and AI/ML technologies (e.g., the aerospace industry) can aid the nuclear industry in successfully integrating them as well.

Page intentionally left blank

CONTENTS

FIGURES

TABLES

Page intentionally left blank

ACRONYMS

3S	safety, security, and safeguards
AI	artificial intelligence
AM	advanced manufacturing
DOE	U.S. Department of Energy
DT	digital twin
GAN	generative adversarial networks
HPC	high-performance computing
I&C	instrumentation and control
LLM	large language models
M&S	modeling and simulation
MUF	material unaccounted for
ML	machine learning
NLP	natural language processing
NPP	nuclear power plant
NRC	U.S. Nuclear Regulatory Commission
SMR	small modular reactor

Page intentionally left blank

4th Big Data for Nuclear Power Plants Workshop 2023

Workshop Report

1. INTRODUCTION

"Big data," a ubiquitous term first popularized in the 1990s, has evolved to include not just large amounts of data, but also complex data of different types, time scales, and origins.[a] Nuclear power plants (NPPs) contain many potential sources of data, including data historians (e.g., the PI server), work logs, and licensee event reports. While the amount of data for a single plant is not huge in relative terms, the complexity of the data pushes them into the big data realm.

Big data technologies have seen significant development in the past few years, and an increasing number of nuclear utilities, vendors, and research organizations are researching potential applications of intelligent data use in nuclear power, due to their significant impacts on plant economics and safety.

The Big Data for Nuclear Power Plants workshop was established in 2017 to better elucidate the potential application of big data technologies to NPPs. While the first workshop mostly focused on big data technology applications in the field of plant operation and maintenance, the scope progressively expanded to include additional applications in the areas of cybersecurity, materials discovery, etc. A timeline of the workshop's various offerings is given in Figure 1.

The Big Data for Nuclear Power Plants workshops for 2017, 2018, and 2019 identified the significant potential held by big data technologies in terms of optimizing plant maintenance and component inventories, and generated a broader array of roles pertaining to big data analytics in nuclear power systems. In the 2023 workshop, emphasis was placed on data science, machine learning (ML), and artificial intelligence (AI) applications used to support six specific nuclear engineering domains: operation and maintenance, high-performance computing (HPC) and massive computation, nuclear materials and advanced manufacturing (AM), cyber and physical security, digital twins (DTs), and nuclear non-proliferation.

This 2-day workshop featured presentations and discussions by researchers from academia, industry, and government, the goal being to identify challenges and gaps that must be addressed to maximize the impact of data on the nuclear industry.

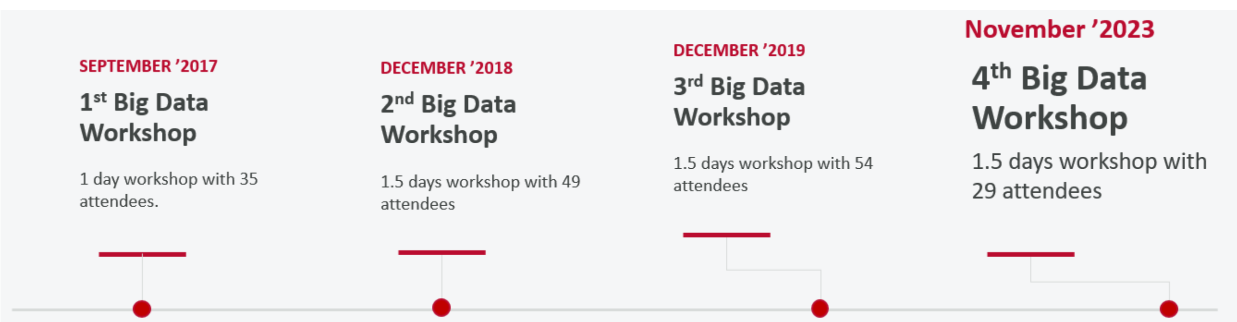

Figure 1. Big Data for Nuclear Power Plants workshop timeline.

The split of attendees for the 4th Big Data for Nuclear Power Plants Workshop is shown in Table 1. A complete list of participants is given in Appendix C.

[a] Lohr, Steve. "The Origins of 'Big Data': An Etymological Detective Story." *The New York Times*, February 1, 2013. https://archive.nytimes.com/bits.blogs.nytimes.com/2013/02/01/the-origins-of-big-data-an-etymological-detective-story/

Table 1. Split of Attendees.

Sector	Number of Attendees	Represented Institutions
Academia	12	The Ohio State University Colorado State University New York University
Laboratory	11	Idaho National Laboratory Naval Nuclear Laboratory
Industry & Government	6	U.S. Department of Energy U.S. Nuclear Regulatory Commission Electric Power Research Institute Blue Wave AI Labs I&C Operative

2. Workshop Structure

The agenda for the 4th Big Data for Nuclear Power Plants Workshop is provided in Appendix B. The workshop included keynote presentations from the U.S. Department of Energy (DOE) and U.S. Nuclear Regulatory Commission (NRC) that covered past, present, and future DOE/NRC programs related to big data technologies, AI, and ML. Additionally, six breakout sessions were held on different topical areas: operation and maintenance, ML in nuclear materials and AM, cybersecurity, HPC and massive computation, big data and DTs, and nuclear non-proliferation. Prior to the breakout sessions, session chairs for the topical areas gave brief introductory presentations on the topics to be covered.

Session chairs were instructed to—to the degree possible while remaining meaningful—cover the following questions with the session attendees:

1. What does big data mean for this topical session area?
2. What are the possible opportunities for big data and ML in this topical session area?
3. What are the low-hanging fruit?
4. What are the long-term applications of big data and ML?
5. What are the challenges to using these approaches within the opportunity areas identified in 2?
6. What needs to happen to resolve the challenges identified in 5?
7. What are the risks involved in the use of big data and ML techniques in this topical session area?
8. What needs to happen to resolve these risks?

After each discussion, the chairs summarized the discussion contents. These summaries are provided throughout the remainder of this document.

3. Session I – Operation & Maintenance

Description: This session focused on the application of automation and optimization techniques for operating and maintaining current NPPs, small modular reactors (SMRs), microreactors, and fission batteries. Automation carries the potential to significantly reduce operation and maintenance staff workload as well as human errors. For SMRs, microreactors, and fission batteries, significant opportunities exist for optimizing operation and maintenance activities so as to improve the efficiency and safety of nuclear systems. For improved efficiency of current NPPs, optimization can also be performed for outage scheduling, inventory management, and fuel usage. Among the topics considered were applicability, performance targets, managing uncertainty, data sharing, interpretability and explainability, and risk issues related to the use of big data and AI/ML.

Chairs: Cody Walker (INL), Carol Smidts (OSU)

This session discussed potential applications of big data, AI, and ML as they relate to the operation and maintenance of NPPs. Participants were asked to consider opportunities, challenges and risks.

Opportunities

Opportunities for big data technologies and the associated AI/ML algorithms can be found in different areas. Specifically, big data technologies offer the possibility to:

Reduce Human Errors. Big data technologies enable collection, analyses, and extraction of a breadth and wealth of information not easily accessible a priori by human personnel involved in maintenance and operations. For example, applications may include the ability to extract information from maintenance records and derive lessons learned in near real time—lessons that can then be fed back to the plant's personnel to help correct future actions. Big data technologies enable consideration of heterogenous data and thus allow for merging information from various sources (visual, text, numeric) in such a way that the information extracted is richer than any information that can be provided from a single source.

Reduce Labor Costs. Processes that were once performed manually can now be automated and the information extracted automatically. Replacing inspector rounds through the use of drones is one such example of this. The collected information can then be processed using big data technologies. Another example lies in automating the rounds performed as part of physical security to determine whether the boundary sensors are functional. The corresponding data can be acquired via big data technologies.

Preemptively prepare for big data analyses in future reactors. The potential exists to preemptively prepare for the application of big data technologies in new reactors. For instance, to allow for easy processing of text-based records, one could move from semi-structured data to more structured data. Sensors can also be more easily added into the reactor environment, as the capabilities will exist to capture and possibly analyze the data in real time.

Move humans toward what they are best at. While humans are, in essence, multi-modality sensors, they are hampered in radioactive environments and difficult-to-access locations. However, one of their major strengths is their ability to interpret and understand complex scenarios. By using big data technologies, we enable humans to capitalize on their strengths, thus reducing their workloads. It is important to note that this benefit assumes that the obtained results lend themselves to interpretation and that there is no loss of situational awareness. In essence, this represents moving toward a paradigm in which the human functions as a leader and the machine functions as a delegated sensor.

Challenges

Such opportunities entail challenges to applying big data technologies to plant operation and maintenance. These challenges are discussed below.

Change Aversion. The nuclear field tends to be change averse, and much of this aversion can be related to culture and long-held proven practices and habits.

Regulation. With nuclear energy being a regulated field, changes are accompanied by the need for regulatory approval. A systematic analysis of the regulatory space and how it may block the progression of big data technologies should be considered. At the same time, early interactions with the regulator regarding any planned changes can make the process more efficient and effective and remove possible roadblocks, reducing the necessary timeframe for obtaining approvals. Additionally, one should realize that the data afforded by big data technologies can be of benefit to the plants in terms of helping meet regulations, informing the regulator, and providing evidence of the need for regulatory changes. These aspects should be considered and studied. Some automation of activities may be covered by NRC Code of Federal Regulations 50.59: Changes, Tests, and Experiments. Since big data technologies are rapidly evolving, they require an adaptive regulatory framework, as well as technologies already proven reliable.

Risks

The risks discussed are listed below. Some pertain to the AI technologies and ML algorithms typically used in conjunction with big data technologies.

Data Space Coverage. Big data technologies allow for the processing of large quantities of data and may give the illusion that the datasets cover all possible conditions and outcomes. This is typically not the case, and it is a known fact that associated AI algorithms are poor extrapolators and may be highly sensitive to minor variations in the data.

No Accountability. The quantity of data processed, their heterogeneous nature, and the degree and method of aggregation easily lead to a loss of traceability between a proposed lesson learned and the actual dataset used to derive said lesson. The conditions under which a lesson holds true should be made explicit. This knowledge is difficult to transfer. The process by which a lesson was derived should also be made explicit and explainable to those making related decisions. Post-hoc methods can be used in support of this type of objective.

Data Privacy and Proprietary Issues. The variety of data accessible through big data technologies is significant. It includes video streams acquired through mobile units such as drones or compilations of text/email messages. Thus, big data technologies raise the issue of defining boundaries and managing the privacy of operation and maintenance personnel. This in turn necessitates the secure management of data storage platforms. Traceability of the data sources/platforms being used must be ensured. The same issues apply to the proprietary nature of the data being collected, as well as the need to preserve the nature of proprietary data. Both are stumbling blocks to the use of big data technologies. Furthermore, the risk escalates when considering possible SMR deployment outside the United States. Possibilities pertaining to the development of test beds for various big data technologies have been discussed as a way to bypass the issue of proprietary data. However, these may not be representative of actual plant data. The process of anonymizing data was evoked by the associated risk of the data not retaining key meaningful information. A process and a business case may need to be created to study these issues. The possibility of creating a database under the lock and key of a federal agency (e.g., DOE) was evoked as well.

4. Session II – Machine Learning in Nuclear Materials and Advanced Manufacturing

Description: Big data analytics tools are finding applications in AM and materials science for current NPPs, SMRs, microreactors, and fission batteries. Examples include the use of image processing tools to identify flaws for qualification of the additive manufacturing process, microstructure characterization via electron microscopy and x-ray imaging tools, ML algorithms to accelerate predictions of material property behavior in extreme environments, and machine-learned interatomic potentials for atomistic simulations of microstructure evolution.

Chairs: William Chuirazzi (INL), Marat Khafizov (OSU), Mathew Swisher (INL)

In this session, the discussions focused on big data analytics and ML in nuclear material and AM applications. Both nuclear materials and AM techniques are vital to the design and implementation of new commercial reactors. Nuclear materials must be designed, constructed, and rigorously tested before they can be approved for use in commercial NPPs. This requires different types of data obtained throughout the entire lifecycle of nuclear material design and testing. Maturation of AM techniques has enabled rapid, flexible fabrication of parts with novel geometries. These AM techniques are of particular interest to the nuclear industry, as they can help circumvent long lead times and supply chain delays. However, a large amount of fabrication and performance data must be collected, optimized, and analyzed in order to ensure the proper performance of components. Furthermore, characterization of materials both before and after exposure to service conditions can generate heterogeneous data ranging from property measurements to 3D tomographic images obtained via x-ray and/or electron microscopy.

The goals of this session were (1) to identify big data types/streams in both nuclear materials and AM, (2) examine how to leverage these data in tandem with ML so as to boost efficiency and overcome current challenges, and (3) outline the long-term applications of big data and ML in these areas. Advances in both nuclear materials and AM would improve the efficiency of new NPPs and reduce plant construction and maintenance costs. Paths forward from the current state-of-the-art to the desired long-term goals were also discussed.

While the size of big data in this context can vary by several orders of magnitude, the consensus was that it is typically 50–100 GB per dataset. In AM processes, data streams include in situ fabrication monitoring to ensure the components are made to specifications; post-production examinations to ensure the components meet the acceptance criteria; microscopy, modeling, and experimental results on material performance in operating environments; and modeling efforts related to part fabrication and performance. For nuclear materials, the data include design parameters and specifications, as well as modeling and experimental information on material performance in extreme, reactor-like environments. While heterogenous data may not be readily available, and missing data may also be a concern, in both areas capturing data from simulations and experimental-/engineering-scale tests is of paramount importance.

Advances in ML techniques have opened the door to several opportunities for big data utilization in the areas of AM and nuclear materials. First, ML could be used for data reduction as well as data transfer between researchers, and for general safety and reliability monitoring outside the plant. In modeling applications, ML could be utilized to expedite density functional theory modeling and to create potential energy models for atomistic simulations. It could also be used to select or design materials that provide the desired properties. One of the biggest impact areas for ML methods may be the linking of modeling efforts to experimental data, as well as creating correlations between models and experimental results across length scales, from the atomistic level all the way up to the engineering scale.

The long-term, far-reaching applications of ML in these fields include using natural language processing (NLP) models to review old reports and publications so as to identify where previously reported roadblocks in AM and nuclear materials research can now be overcome thanks to new technological developments. Likewise, these NLP models could be used to identify missing data in existing datasets, fuse together multiple datasets, and scan the literature for images that may be recovered to augment datasets. By examining historical data, NLP models can assist in knowledge transfer. Generative AI models could be used for predictive maintenance in NPPs, as well as in the design of plant components that reflect the desired material properties in a given environment.

Challenges to these long-term goals center around the availability of data on everything from individual reactor components to the performance of different NPP types under a variety of operating conditions. Data sharing between commercial entities and researchers is another challenge, as is the process of validating DT models. While some of these challenges have far-reaching implications, some low-hanging fruit was identified that could help advance the state of the field toward these long-term goals. Sharing data via DTs created for the entire fleet, as well as sharing trained models without sharing the original data, were two suggested pathways for facilitating data sharing among the various stakeholders. Another promising suggestion was to learn from how data analytics is used in other fields (e.g., robotics, aerospace, machining, welding, and even soil compaction) so as to help train and inform ML models. As for missing data, NLP models could be used to search for data that may exist but are considered missing, or be used to create a list of nonexistent data that still need to be generated.

Short-term practices that would bring us closer to achieving these low-hanging fruit involve active learning to update ML models as new data are collected. Additionally, uncertainty quantification of ML models must be performed to determine the level of risk posed by unknowns in a dataset. Borrowing lessons learned from other industries—particularly the aerospace industry—may aid in overcoming some of these recent challenges.

Among the risks associated with these ideas is training on adversarial, bad, incomplete, or missing data. This would ultimately lead to model inaccuracies, severely compromising the models' ability to generate meaningful results. In model-sharing applications, there is also risk of extracting original confidential data from trained models that have been shared. Data must therefore be verified for accuracy and relevancy prior to being accepted for data training, and data security must be the top priority.

Data-driven ML techniques carry the potential to transform not only nuclear material and AM research, development, and implementation, but could also reduce the costs and concept-to-implementation timelines for novel nuclear reactor technologies. Successfully applying ML to these fields could enable breakthroughs such as in situ monitoring and adjustments to part fabrication (through AM techniques), and the reduction or elimination of new parts that do not meet the acceptance criteria. ML could also aid in the design of new nuclear reactor technologies, thus expediting the process of modeling, testing, and implementing nuclear materials, as well as correlating data between modeling and experimental results across length scales so as to gain a comprehensive understanding of material performance in reactor environments. Realizing such potential requires a path forward for data sharing amongst various different commercial and research entities, as well as for validating models so that they can be implemented.

5. Session III – Cybersecurity

Description: Digitalization and computerization of instrumentation and control (I&C) has made NPPs more vulnerable to cyberattacks than ever. Such attacks may have severe implications for plant operation and safety. This session aimed to identify and resolve cybersecurity issues regarding big data and AI/ML applications in the nuclear field, and to apply big data and AI/ML technologies to cybersecurity efforts in a manner that incorporates robust cybersecurity policies, procedures, and practices so as to protect the vital components of current NPPs, SMRs, microreactors, and fission batteries. Topics of interest included but were not limited to confidentiality, integrity, and availability issues in big data and AI/ML applications; cyber threats pertaining to data sharing for AI/ML model training; adversarial uses of AI/ML models, as well as using big data and AI/ML for cybersecurity I&C system functions; identification and assessment of cyber threats; and analyses of I&C systems' attractiveness to potential adversaries, their vulnerabilities, their operating environments, and any consequences that could directly or indirectly result from NPPs becoming compromised.

Chairs: Katya Le Blanc (INL), Carol Smidts (OSU)

This session discussed the use of big data technologies for cybersecurity applications in NPPs. Workshop participants discussed various topics, including defining what constitutes big data and identifying related opportunities and challenges/risks pertaining to NPP cybersecurity.

Each of these issues is discussed, in turn, below.

What data are included in big data?

First, one should define what data are considered relevant to the cybersecurity problem for NPPs. This may include a range of possible data, such as those associated with system and network configurations, vulnerabilities, system logs, procedures, simulators, test beds, test suites, sensor-related information, training data, open-source code, financial information related to all plant personnel, data from social media accounts associated with plant personnel, data that can be associated with attackers (tools, techniques, demographics), data related to insider threats (e.g., incidents), and any outlier (e.g., outlier access data) or anomalous data.

What are the challenges?

Data Diversity. One primary challenge relates to the diversity of the types of data mentioned above. Due to the diversity and extent of the data needs, which even extend outside the plant walls, systems, and networks, data collection efforts may require significant investment and will need to be proven economically viable. These efforts may be use-case specific. Given the amount and diversity of the data outlined, use of reduced-order models for cyber data should be considered.

Game Identification. In essence, plant personnel and attackers are ultimately engaged in a game: namely, the security game. The issue then is to understand the type of game being played. This is made particularly difficult by the fact that the game itself may not be stationary and its nature may change over time. To understand the nature of this game, one must repeatedly expose oneself to the adversary in order to develop an appropriate immune response, be proactive rather than reactive, and learn to think like an attacker.

Special Needs for ML Algorithms. ML algorithms apparently require specifically dedicated blue and red team approaches.

What are some of the opportunities?

Big data technologies can be used to collect data on an opponent or to validate cybersecurity in real time (or almost real time). They can be utilized to enforce practices such as the verification of canonical security data, the enforcing of multi-factor authentication, or the use of digital signatures.

These technologies may also afford the data necessary to enforce cyber defense practices such as cyber-resilience, adaptive reinforcement learning, or adaptive/online learning. They may also be used to populate cybersecurity simulators with interactive training capabilities.

Big data technologies unlock the potential to capture the necessary data to inform critical functions.

They can also be used to monitor "imprudent" behavior.

What are some of the risks?

Centralization. Risks are entailed by any centralization of data such as those employable in managing NPP cybersecurity.

Privacy and Trust. Any data capture that can be traced to an individual will certainly raise a flag that a privacy-related issue may have occurred.

6. Session IV – High-Performance Computing and Massive Computation

Description: HPC is the practice of aggregating computing to deliver much higher performance than one could get out of a typical desktop computer or workstation. The goal of HPC is to solve large problems in science and engineering. Although HPC has progressed remarkably over the past couple of decades, in recent years that progress has been achieved through greatly increased hardware complexity thanks to the rise of multicore and manycore processors, and this is affecting application developers' ability to achieve the full potential of these systems. In other words, HPC systems are becoming more and more complex and the hardware is exposing massive parallelism at all levels, making full utilization of these resources a challenge. In this session, participants addressed challenges and opportunities in several areas of HPC that support big data at the university, company, and national laboratory scale and that are relevant to current NPPs, SMRs, microreactors, and fission batteries. What hardware should be purchased (central vs. graphics processing units)? What unique challenges do big data pose to hardware? How do we balance the needs of engineering-level computation and data analytics? Can high-fidelity modeling and simulation (M&S) support big data analytics? How can high-performance data analytics improve prediction accuracy? How do smaller companies and universities with less access to computational resources utilize data analytics? Would HPC lead to better results or greater utilization of big data? These and other equally important questions were identified and examined, and fruitful avenues to improve the handling of big data through HPC were proposed.

This session deliberated on prospective applications of big data and AI/ML, alongside the implementation of HPC to support these efforts within the context of NPPs. Participants were asked to examine conceivable opportunities, challenges, and associated risks linked to incorporating these technologies into this field.

Chairs: Matthew Anderson (INL), Leonardo Moraes (OSU)

Opportunities

Exploring the potential for augmenting the utilization of high-fidelity models. Reliance on low-fidelity models for nuclear reactor systems corresponds to a high potential for excessive over-conservatism in their design, and this has proven an impediment to their economic competitiveness relative to other energy sources. Addressing this challenge so as to develop the most economical designs and operation plans for both existing and advanced reactor systems requires complex and computationally demanding high-fidelity models. HPC is pivotal for facilitating the resolution of such models, enabling the generation of synthetic data to serve as a foundational basis for generative AI applications, thus enhancing the efficiency of high-fidelity models. One example discussed was the selection of fidelity within multiphysics codes in the M&S domain. Depending on the characteristics of the medium, low-fidelity models can be employed with commendable accuracy. Introducing a tool that, by analyzing problem parameters, autonomously selects appropriate models and parameters would substantially enhance the efficiency of the computational code. Another noteworthy example is the generation of initial guesses for solutions. This is particularly pertinent in regard to numerical techniques founded on iterative processes aimed at progressively refining the solution to a given problem. Initiating this iterative process via an initial guess proximate to the actual solution markedly improves program efficiency.

Use of physics-informed generative adversarial networks. Integration of domain-specific physical laws and principles into the framework of generative adversarial networks (GANs) presents a promising avenue for overcoming the lack of publicly available datasets. Physics-informed GANs carry the potential to generate synthetic data that adhere to the underlying physics of the system, ensuring that the generated samples possess realistic physical characteristics. This approach not only overcomes data scarcity but also offers a means to augment training datasets for ML models in the context of NPPs, where obtaining large datasets can be made challenging due to safety and security concerns. By enabling development of robust and accurate models for tasks such as anomaly detection, system optimization, and predictive maintenance, physics-informed GANs help enhance the safety and efficiency of nuclear power operations.

Possibility of adopting internal chatbots. Internal chatbots powered by open-source large language models (LLMs) offer significant opportunities for NPPs. These AI-driven conversational agents could potentially greatly enhance communication and operational efficiency within the intricate, safety-critical environments of nuclear facilities. By incorporating knowledge specific to nuclear engineering, these chatbots can provide instantaneous and precise responses to inquiries, aid in troubleshooting, facilitate information retrieval and the ingesting of documentation, and generate reports, summaries, etc. They can also contribute to the ongoing learning and training of personnel, thus ensuring adherence to safety protocols and best practices.

Possibilities concerning traditional AI. These established algorithms, including rule-based systems, expert systems, and classical ML techniques, represent a reliable solution for addressing operational challenges, including via fault diagnoses, process optimization, and predictive maintenance. The efficacy of these approaches is evident in situations when explicit rules and patterns govern system behavior, ultimately contributing to the safety and efficiency of nuclear operations.

Moreover, traditional AI can enhance the decision-making process by providing interpretable and explainable results—something of paramount importance in safety-critical environments. Comprehensive analysis of possible applications of traditional AI in NPPs reveals a plethora of opportunities, ranging from enhancing meshing capabilities in M&S to attack prevention, code writing, anomaly detection, post-irradiation examination, quality of service for network traffic, neural networks for cross sections, radiation spectroscopy, and environmental response to unforeseen incidents (reinforcement learning).

Challenges

Limited resources for training LLMs. Training LLMs is a complex task, especially when resources are limited. Challenges such as restricted computational power, limited memory, and insufficient access to diverse training data make it difficult to effectively explore extensive model architectures and datasets. Moreover, the high computational costs of LLM training demand substantial financial investments. These obstacles can impede the model's ability to effectively understand a wide range of linguistic nuances.

Lack of a data curation policy. The absence of a comprehensive policy for managing data poses significant challenges and risks to effective digital information handling. Without a structured policy, research groups and organizations may face inconsistent data quality, thus hindering reliability and impeding informed decision making. Additionally, the lack of standardized data documentation and metadata practices can make it challenging to understand and interpret datasets, ultimately limiting their long-term usefulness. Furthermore, without a policy, data become vulnerable to security breaches and unauthorized access, potentially compromising confidentiality and integrity. In terms of risks, insufficient data curation can undermine the reproducibility of research findings, limiting result verifiability and impeding scientific progress. Moreover, without a well-defined policy, the potential for data loss or corruption increases, putting at risk the preservation of valuable information crucial for future analyses and insights. Establishing a robust data curation policy is essential for addressing these challenges and ensuring the integrity, accessibility, and usability of digital assets.

Use of cloud services to provide the computational resources needed to address HPC resource limitations. Using cloud services to overcome the limited availability of HPC resources in NPPs entails certain challenges that may make it impractical in certain situations. Even though cloud platforms provide flexibility and scalability, the considerable costs of setting up HPC capabilities on demand can be exorbitant, especially for resource-intensive applications and prolonged computational workloads. Moreover, expenses related to data transfer and storage can add to the overall operational costs. Concerns about data security, privacy, and compliance also arise given the stringent regulations that dictate data handling and storage practices.

Difficulty in hiring and retaining data scientists. Recruiting and retaining data scientists in the field of nuclear engineering entail significant challenges as well. The shortage of professionals with the specialized skills required for nuclear engineering and advanced data analytics makes suitable candidates difficult to find. This field requires expertise in both nuclear and data science, representing a unique pool of qualified individuals. Adding to this challenge is the fact that the attractiveness of data science roles in more mainstream industries makes it hard for universities and nuclear engineering organizations to attract top-tier talent. And the dynamic and competitive job market further increases the difficulty of retaining skilled professionals in this field. To tackle these challenges, targeted initiatives are needed to cultivate a talent pipeline of personnel with qualifications in both nuclear and data science. Competitive compensation structures and a stimulating work environment that highlights the meaningful contributions that data scientists can make in the nuclear engineering sector are also crucial considerations.

Risks

Regulation being unable to resolve the risks of AI. Regulatory frameworks have a long way to go when it comes to dealing with the challenges brought on by AI. These inefficiencies can introduce significant risks and lead to a lack of clarity on ethical standards, accountability, and transparency in AI applications. This confusion can result in unchecked deployment of AI systems, potentially leading to detrimental consequences such as biased decision making, privacy infringements, and insufficient safeguards against unintended AI behaviors. Moreover, inconsistent or insufficient regulatory measures can impede innovation and hinder responsible development and deployment of AI technologies. Lack of global harmonization in AI regulations can also contribute to regulatory arbitrage, with entities opting for jurisdictions with lax oversight, thus exacerbating the risks associated with unregulated or poorly regulated AI systems. It is crucial to establish robust and adaptive regulatory frameworks that balance innovation with ethical considerations so as to ensure responsible and accountable use of AI technologies across a diverse range of applications.

Model/data poisoning. ML models, though powerful, can be vulnerable to malicious attacks. Model and data poisoning techniques can inject false or misleading information into the training data or manipulate the model inputs in subtle ways, leading to skewed predictions, security breaches, and unauthorized access. Adversaries can exploit vulnerabilities in ML systems, creating biased models that behave in unexpected ways and undermine the reliability of decision-making processes. To mitigate these risks, ML systems must be equipped with robust defenses such as anomaly detection and secure data handling practices.

Inadequate understanding among AI users. Utilization of AI technologies without comprehensively understanding their intricacies poses significant risks. Individuals who lack familiarity with AI may encounter difficulties in interpreting results, thus leading to misguided reliance on automated systems. Misconceptions regarding the capabilities and limitations of AI may result in overdependence on flawed models, potentially leading to significant financial, ethical, or legal consequences. Furthermore, the lack of comprehension may impede effective oversight, making it challenging to identify and correct biases, errors, or unintended consequences entailed by AI applications.

Challenges inherent in proprietary data utilization. Utilization of proprietary data entails inherent risks that mainly center around issues of privacy, security, and limited transparency. Organizations that rely on proprietary datasets face significant challenges in maintaining the confidentiality and integrity of sensitive information, potentially exposing individuals to privacy infringements. Moreover, the lack of transparency in proprietary data may lead to a dearth of comprehension as to the data's origins, collection methods, and potential biases, thereby hindering thorough assessments and responsible usage. Security breaches also pose a significant concern, as unauthorized access to proprietary datasets could result in severe consequences such as data manipulation or misuse.

7. Session V – Big Data and Digital Twins

Description: DTs are now being actively developed and used in many stages of the NPP lifecycle, including for design, licensing, construction, operations, oversight, monitoring, and maintenance. A recent NRC report identified several challenges and gaps associated with nuclear applications of DTs, including real-time integration of sensor data with DTs and the use of traditional M&S tools as data-informed models. Data quality, quantity, applicability, and uncertainty were identified as key technical considerations in regulatory decision making in NRC's AI Strategic Plan of 2023. This session focused on considerations, opportunities, challenges, and gaps associated with integrating big data with DTs in nuclear applications.

Chairs: Vaibhav Yadav (INL), Xiaoxu Diao (OSU)

State of the Art

DTs in the realm of big data signify a paradigm shift in how data are managed and utilized. They embody a unique duality, acting as both consumers and generators of data. As consumers, DTs assimilate vast amounts of information, characterized by the "3 v's" of big data: volume, velocity, and veracity. Volume deals with the large quantities of data processed, and velocity underscores the need for rapid processing, which is essential for real-time integration. In fact, the velocity at which DTs operate is critical, as they must often function at faster-than-real-time speeds in order to effectively predict and simulate scenarios. The veracity (or trustworthiness and diversity of the data) adds further complexity, requiring DTs to handle heterogeneous data sources reliably. Additionally, the metadata associated with each DT provides vital context, thus enhancing the accuracy and applicability of the data. The outcomes or outputs from DTs, ranging from predictive analytics to operational optimizations, hinge on these data characteristics. Crucially, the use cases of DTs dictate the optimum resolution of data in both temporal and physical scales, ensuring that the DT provides meaningful and actionable insights. This balance of resolution is pivotal in tailoring DTs to specific needs, thereby highlighting their adaptability and the transformative potential they hold in harnessing big data.

Opportunities

Integration of big data and ML into DTs presents a host of opportunities that could revolutionize their functionality and impact. Generally, a DT requires synergy between big data and ML in order to achieve its full potential. By harnessing vast amounts of data and applying advanced ML techniques, DTs can achieve more accurate and nuanced predictions. This accuracy is particularly crucial for scenarios in which DTs must operate at faster-than-real-time speeds, a feat made possible through ML-based surrogates that can effectively replace or augment traditional physics-based M&S.

Moreover, the convergence of big data and ML within DTs introduces unique challenges and opportunities in terms of verification and validation as well as uncertainty quantification. Typically, ML models are trained on M&S-generated data that are then validated against real-world outcomes. However, when these ML models are embedded within a DT that utilizes real-time sensor data, the dynamics change significantly. This shift necessitates novel approaches for seamlessly integrating M&S, ML, and real-time data within the DT environment. The continuous feedback loop between real-time data and ML models within DTs not only enhances predictive accuracy but also presents new challenges in ensuring model reliability and validity. This complexity underscores the transformative potential of big data and ML in DTs, paving the way for advanced applications that can dynamically adapt to real-world conditions and deliver insightful predictions and analyses.

In the context of implementing DTs in the nuclear sector, identifying the low-hanging fruit involves several strategic steps. First, it is crucial to pinpoint the DT's primary purpose in a given nuclear application, which could range from enhancing operational efficiency and safety to performing predictive maintenance and emergency response planning. Once the core objective is established, the next step is to identify the essential elements and technologies needed for the primary DT. This includes not only the hardware and software components but also the data analytics and ML tools integral to DT functionality.

A key aspect in this process is the thorough examination and validation of existing data prior to incorporating them into the DT. Ensuring the accuracy, relevance, and quality of these data is fundamental to the effective operation of the DT. Additionally, understanding the current state of the art in DT applications in other industries can lead to valuable insights. Non-nuclear sectors may have developed advanced practices in DT implementation, and these can potentially be adapted or modified for nuclear applications.

Finally, it is vital to identify DT-enabling technologies already in use in other domains, and to assess their potential applicability to nuclear DTs. This may involve leveraging existing software platforms, data-processing techniques, or simulation models. By focusing on these areas, the nuclear industry can swiftly and effectively harness the potential of DTs, thus capitalizing on the low-hanging fruit in order to achieve significant operational and safety benefits.

Long-Term Applications

Kinetic DTs represent a dynamic, on-demand approach in the realm of DT applications. Such DTs are invoked only when necessary, thus ensuring efficient use of resources. In the context of NPPs, this could manifest in model predictive controls aimed at operational optimization. For instance, a kinetic DT could provide recommendations on reducing refueling frequencies and accounting for real-life constraints such as procedural, budgetary, physical, and temporal factors.

Autonomous controls and operations represent another frontier for DT applications. By integrating DTs with robotic actuation, systems can achieve higher levels of autonomy, thus improving efficiency and reducing human error levels. This integration paves the way for advanced applications such as root cause analysis, with DTs being able to dissect incidents so as to identify underlying causes, thereby informing preventive measures.

Finally, the scope of DTs also extends to *security aspects*. By integrating cyber and physical security scenarios into DTs, future security measures can be made more robust and responsive. This integration allows for the simulation and analysis of potential security breaches, thus enabling proactive measures and rapid response strategies. In summary, the kinetic DT concept, with its focus on demand-driven activation and integration with various technologies and data sources, carries immense potential for transforming operations across different sectors, particularly in NPPs.

Challenges

One primary challenge is to recognize the trade-offs inherent in DT design and operation. For instance, a DT that operates at a high speed may sacrifice a degree of accuracy, necessitating a careful balance based on the application requirements.

Furthermore, ensuring the reliability and validity of DT-generated data and predictions is paramount, especially in high-stakes environments such as NPPs. This involves not only rigorous testing and validation processes, but also continuous monitoring and updating of the DT so as to reflect the latest data and insights.

Cybersecurity presents a unique set of challenges in the context of DTs. It is crucial that the extra cybersecurity dimensions specific to DTs, such as data levels and tracking, be understood and addressed. This requires a comprehensive approach that encompasses not only the DT itself but also its data sources and the communication channels used.

Another challenge lies in determining the optimum update frequency for DTs. This frequency must align with the update intervals for any enabling technologies, as well as the real-time dynamics of the system being modeled. A delicate balance must be struck to ensure that the DT remains relevant and accurate without overwhelming the system with constant updates.

Finally, it is also crucial to identify the balance between the different DT operational modes: real-time, periodic, or triggered/queried. Each mode has its advantages and is suited for different scenarios. Real-time DTs are essential for immediate decision-making processes, periodic DTs can be useful for routine monitoring, and triggered or queried DTs can be effective for specific on-demand analyses. Understanding and selecting the appropriate operational mode based on the needs of the application is key to overcoming challenges and maximizing the potential of DTs in various opportunity areas.

Risks

One significant risk is the potential for improper assessment and diagnosis by the DT, particularly given the public perception of AI in the nuclear sector. To mitigate this, it is crucial that DTs possess the capability to learn from and identify incorrect assessments, thus enhancing their accuracy and reliability over time.

Another risk involves the trustworthiness of DTs, especially when they make recommendations based on scenarios not covered in their training data. To address this, maintaining a human-in-the-loop approach is essential. This involves having human oversight and input in critical decision-making processes, ensuring that the DT's recommendations are validated and interpreted correctly.

Data and information overload is a further concern. The sheer volume of data processed by DTs can lead to challenges in efficiently extracting relevant insights. Implementing robust data management and analysis strategies can help filter out noise so that pertinent information can be focused on.

Lastly, the risk of loss or gaps in data and inputs can significantly impact DT effectiveness. Ensuring continuous and comprehensive data collection, along with backups and redundancy measures, is vital for maintaining the integrity and continuity of DT operations.

8. Session VI – Nuclear Non-Proliferation

Description: Per the Nuclear Non-Proliferation Treaty, the International Atomic Energy Agency is tasked with promoting peaceful use of nuclear energy technologies, in addition to deterring and detecting the spread of nuclear weapons—as mandated by State signatories of the treaty. Big data technologies can facilitate new scientific discoveries in physics, chemistry, and biology so as to advance security solutions for current NPPs, SMRs, microreactors, and fission batteries. This session focused on big data technologies that can be reused in nuclear non-proliferation applications, including for supporting the detection of undeclared plutonium and uranium, for use in ultra-sensitive nuclear measurement systems, for nuclear explosion detection, and for legal and regulatory analyses.

Chairs: Samuel E. Bays (INL), Praneeth Kandlakunta (OSU)

State of the Art

The combined topics of nuclear safety, security, and safeguards (3S) represent a very broad domain space. Each individual topic typically has its own datasets and analysis frameworks for identifying risks.

Each also has its own field of study, usually requiring significant specialization for analysts to be able to draw meaningful conclusions and posit risk mitigation strategies. From the public's perspective, the most perceptible risk factor is radiation exposure at the nuclear facility's property boundary.

Nuclear safety is intrinsic to nuclear reactor designs. Final design and operating approval is only granted by government regulatory authorities if the design has been reviewed and validated against experience and meets public expectations. Ex- and in-core process and radiation monitoring data are routinely collected to confirm safe operation of the approved design. Non-reactor nuclear facilities also collect significant amounts of process data to confirm safe and normal operation. However, safety does not end with design and process monitoring. Typically, safe operation holistically involves human factors that lead to defense in depth and situational awareness. These human-based activities, in tandem with acceptable design and process monitoring, serve to reduce the risks down to acceptable levels. The challenge is for the facility operational management and the regulatory authority to know when safety awareness is insufficient to predict radiological events.

Nuclear facility security safeguards nuclear materials from theft or sabotage by sub-national adversaries. Nuclear security planning and the monitoring of nuclear material and equipment is the primary deterrent to theft and sabotage. Successful security planning involves minimizing the amount of time between detecting and responding to malicious acts. Significant surveillance, intrusion detection, personnel access restrictions, and trustworthiness verification of workers are all important aspects of security planning, together accounting for a significant share of nuclear facility operating costs. Training professional security staff is time intensive and typically retroactive to past threats. Furthermore, false-positive indications from surveillance or information collection unduly complicates operations. Evolving cyber threats or cyber-enabled physical threats pose a constantly evolving risk to physical security.

Safeguarding nuclear materials and facilities from diversion and undeclared use is a human-capital-intensive activity performed by the International Atomic Energy Agency. In the advent of significant market growth of advanced reactors and fuel cycle facilities, nuclear material safeguards will rely much more heavily on remote monitoring methods. Currently, remotely operated cameras, motion sensors, and radiation detectors are used to monitor the movement of special nuclear material. Nuclear material transfers also require extensive records management, accounting, and inventory databases. These records must be verified via in-field inspection. The inspection process, records management, and data interpretation are all very human-resource intensive.

The large volumes of digitized data generated by the 3S processes could be collected more efficiently if supported by ML/AI.

Opportunities

Current methods leverage existing intrinsic datasets. These data typically pertain to the nuclear material itself. Other extrinsic datasets that are not typically part of normal verification activities could be leveraged for gauging the resiliency of nuclear facilities in responding to 3S events. Nuclear facilities collect large amounts of data for safety, quality, and workforce management. These extrinsic datasets include information on special chemicals needed to control the acidity and impurities of working fluids, anticontamination clothing, specialized tools, resin beds, and air filters. They also include data pertaining to releases from nuclear facilities, including low-level waste, liquid effluents found in the facility sump and cooling ponds, and airborne effluents found downwind of nuclear facilities.

Extrinsic datasets can also characterize the human factors pertaining to the facility's workforce, such as personal dosimetry, human resource data, and days-not-worked due to occupational injury. Examples of human resource data are demographics, salary, and educational background. Examples of occupational data are slips, trips, and falls. These types of extrinsic datasets, when integrated with intrinsic facility monitoring datasets, can lead to a better-informed model of the state of 3S at a given nuclear facility.

The ability of ML/AI algorithms to combine diverse datasets and infer insightful conclusions about the system status is well demonstrated. ML/AI can also combine the experience and knowledge of many domain experts into a unified set. In combining diversified large datasets characterized by domain experts, ML/AI can generate new insights likely not previously obtainable given the siloed domain spaces (i.e., each of the 3S's individually).

Challenges

The main challenge with holistic approaches enabled by ML/AI is timely detection of unexpected changes in the status of radiological or special nuclear materials. Nuclear materials arrive and depart from nuclear facilities in discrete quantities. The challenge is that, after arrival, many nuclear materials (e.g., actinide fuel) or supporting materials (e.g., anticontamination clothing) are proportioned into smaller items—or even bulk quantities—that are more difficult (or nearly impossible) to track as a discrete quantum. Similarly, disposal of certain nuclear and non-nuclear materials is difficult to track at the item level. These challenges can lead to the production of undetectable quantities of unaccounted-for materials. This is especially true for unirradiated and uncontaminated materials, which lack a detectable associated radiation field.

Path Forward

Failure of 3S systems triggers a prompt and sometimes catastrophic event. Such events are typically preceded by gradual losses in situational awareness. Event precursors can manifest as equipment degradation, aggregated human performance failures, and perceived loss of deterrent by adversaries. An ML/AI-enabled assessment tool that combines intrinsic datasets with extrinsic ones can provide facility operators with quantitative and explainable measures of system state. Such explainability measures can be used to enhance maintenance and corrective actions. Validated ML/AI protocols can build confidence in decision making. Objective decision making reduces risk, improves operation efficiency, and makes regulatory oversight more effective.

Investment in multi-domain expertise pertaining to data analytics, nuclear engineering, physical security, cybersecurity, nuclear materials safeguards, and facility operations is necessary for building facility holistic-data tools. Facility design should include sensor placement and data acquisition plans. Such development requires a highly trained workforce and trustworthiness-verified performers. Clear discussions must be held between tool developers and stakeholders so that the critical functions of the nuclear facility are understood. Such information and insights gathered in a single location or possessed by a specialized team heighten cyber awareness.

Future proposal calls should consider both intrinsic and extrinsic datasets in the creation of ML/AI tools. Such data already exist at DOE and commercial nuclear facilities. The research community should not wait until the creation of advanced reactor and fuel cycle facilities to validate new ML/AI protocols.

9. Summary

Clearly, the opportunities, challenges, and risks identified in each of these six focus areas are quite diverse; however, consistent patterns do emerge (Figure 2, Appendix A). The integrity, quality, coverage, privacy, and traceability of data are common themes among all the topic areas. While this is hardly a surprise in a workshop on big data, it does point to a need to consider best practices for data management—including security and privacy—if the goal is to develop AI/ML tools based on those data.

Figure 2. Word cloud based on the discussion summaries for all the sessions.

Across the various application areas, opportunities exist to leverage big data and AI/ML tools in order to reduce costs, optimize human tasking, and reduce human error. AI/ML models can be trained on augmented datasets that incorporate less traditional data sources (e.g., high-fidelity synthesized data) to improve the fidelity and optimize the efficiency of the models. NLP could assist in knowledge transfer, including populating simulators for the purpose of training. Ultimately, these technologies are anticipated to reduce the costs and timelines of concept-to-implementation development of novel nuclear reactor technologies.

Common challenges were identified in the breakout sessions. In addition to the data challenges—including the ostensibly competing issues of privacy and traceability—model reliability and validity must be established for each application, including characterizing the sensitivity of the models to parameters of interest. Additionally, the availability of computational resources to train LLMs for specific use cases must be considered.

Among the common risks identified across the various application areas are the inherent risks in data centralization, including potential security breaches and loss of information (or its context). Model inaccuracy as a result of missing, bad, or incomplete training data is also a problem. Furthermore, there may be an insufficient number of experts possessing knowledge in all three needed areas: data science, the nuclear industry, and the specialized application domain.

There are a wealth of opportunities via which the nuclear industry can benefit from big data and advanced analytic capabilities, but taking advantage of these opportunities entails certain challenges and risks. Addressing these challenges and risks is essential for successfully implementing big data technologies in the nuclear industry, which can and should seek to benefit from lessons learned in other industries that have already been leveraging big data and AI/ML for many years (e.g., the aerospace industry).

The technical and research gaps identified throughout this workshop represent challenges that must be addressed to facilitate adoption and maximize the impact of big data and AI/ML technologies by the nuclear industry. Addressing these challenges will have a positive impact on the nuclear industry as a whole. The highest priority challenges identified during the workshop are as follows:

1. **Data curation policy:** Training and validation of models for each specific nuclear research domain requires access to curated comprehensive datasets—preferably with digital identifiers. A coordinated data curation policy must be developed for nuclear research to ensure long-term preservation of (and access to) experimental data and metadata, material characterization data, etc. The policy should include data from commercial entities, potentially entailing additional needs pertaining to data protection and anonymization.
2. **Data warehouse:** A capability should be developed to store and make publicly available comprehensive data for NPPs—from the component level to the plant level—across various operating conditions.
3. **Data collaboration:** Capabilities should be established for enhanced data sharing between commercial entities and researchers, thus fostering innovation and collaborative progress.
4. **Advanced manufacturing data:** Reliable and confidential data sources should be established for training ML models and validating DT models in AM processes.
5. **Error reduction:** Automation carries the potential to significantly reduce operation and maintenance staff workloads, as well as human errors. However, it may also increase the potential for human errors due to complacency or lack of situational awareness. A methodology should be developed to determine how big data and AI/ML can be optimally used in light of reducing human error.
6. **Change aversion:** Industry has a reputation for being averse to changes such as the adoption of new technologies (e.g., AI/ML). Understanding why this aversion exists and developing strategies to overcome it are crucial to harnessing the benefits of big data and AI/ML.
7. **Game identification:** The interaction between plant staff and potential intruders is akin to a strategic contest—a security game. This game and its characteristics shift over time, making it challenging to comprehend the specific dynamics involved. Grasping the essence of this ever-changing game requires consistent engagement with the opposition. This aids in cultivating a suitable defense mechanism, encourages a proactive rather than a reactionary stance, and cultivates an understanding of the attacker's mindset in defense strategies.
8. **Bulk accountancy:** Bulk-accountancy-safeguarded facilities such as pebble-bed reactors, molten-salt reactors, enrichment facilities, and reprocessing facilities face challenges in identifying material unaccounted for (MUF). These facilities collect large amounts of data that could be analyzed using ML/AI techniques that reduce the amount of MUF or better explain the causes of MUF at bulk accountancy facilities.
9. **Tool integration:** Integration of traditional M&S tools with AI advances that were achieved independent of M&S efforts offers significant potential benefits. One example is the use of AI/ML models as efficient surrogates for high-fidelity simulation models, eliminating some of the over-conservatism introduced by relying on low-fidelity models. Another example is using hybrid models to integrate known physical laws and principles with the computationally efficient AI/ML models.
10. **DT validation:** DTs face unique challenges in the area of validation and uncertainty quantification. Integration of M&S, ML, and real-time data within a DT environment that enhances predictive accuracy also creates unique challenges in ensuring model reliability and validity. Use of big data and

ML in DTs paves the way for advanced applications that can dynamically adapt to real-world conditions and deliver insightful predictions and analyses. The question of how best to leverage big data—both from simulations and physical systems—to validate these complex DTs is critical to their acceptance and adoption in the nuclear industry.

11. **Integration of DTs:** There have been several DT use case demonstrations and applications in the recent past. One major challenge in fully capitalizing on the benefits of DTs for NPP applications is the integration of various DTs to create virtual models of complex systems. Integration of multiple DTs poses several major challenges associated with data transfer, communication between different models (e.g., ML communicating with physics-based models), etc.
12. **Real-time AI within DT:** One of the strongest benefits of DTs is their ability to provide real-time analytics and outcomes. When AI is implemented as part of a DT, it is critical that the integrated AI models can be run in real-time. This poses challenges, including computational requirements, real-time data transfer and storage, and real-time interfacing of the AI models with other models.

10. Acknowledgement

The workshop was partially funded by the Battelle Energy Alliance National University Consortium.

Page intentionally left blank

Appendix A

Workshop Summary

The report is summarized as follows. The report was studied and the most frequently used concepts and terms in the sessions were identified. Qualitative coding was used to synthesize these results. A total of six sessions (i.e., operation and maintenance, nuclear materials and advanced manufacturing, cybersecurity, high-performance computing and massive computation, digital twins, and nuclear non-proliferation) were considered, and the discussions from these sessions were summarized into the following four broad categories:

1. What is big data?
2. What opportunities are associated with big data?
3. What are the challenges of using big data?
4. What risks are associated with big data?

It should also be noted that generic challenges and risks associated with each topic area were identified in addition to domain-/technical-area-specific ones.

1. What is Big Data?

Operation and Maintenance Data

Nuclear Materials and Advanced Manufacturing Data

- Big data is sized at 50–100 GB per dataset
- AM data streams include:
  - In situ fabrication monitoring data
  - Post-production examinations data
  - Microscopy, modeling, and experimental results on material performance in operating environment data
  - Modeling efforts on part fabrication and performance data
- Nuclear materials data sources include:
  - Design parameters and specifications
  - Modeling and experimental data on material performance in extreme, reactor-like environments

Cybersecurity Data

- System configurations
- Network configurations
- Vulnerabilities
- Systems' logs
- Procedures
- Simulators
- Test beds
- Test suites

- Sensor-related information
- Training data
- Open-source code
- Financial information related to all plant personnel
- Data from social media accounts associated with plant personnel
- Data that can be associated with attackers
  - Tools
  - Techniques
  - Demographics
- Data related to insider threats
  - Incidents
- Outlier data
  - Outlier access data
  - Anomalous data

High-Performance Computing & Massive Computation Data

Digital Twin Data

- As data consumers
  - High-volume data
  - High-velocity data
  - High-diversity data
  - High-veracity data improves through associated metadata
  - Temporal and physical scales are dependent on use cases for meaningful and actionable insights
- As data generators

Nuclear Non-proliferation Data

- For nuclear safety
- For nuclear reactors
  - Reactor design for initial approval
  - For confirmation of safe and normal operation
    - Ex- and in-core process data
    - Ex- and in-core radiation data
- For non-reactor nuclear facilities
  - Confirmation of safe and normal operation
  - Design and process monitoring
  - Human factors for defense in depth and situational awareness
- For nuclear facility security
  - Security planning data
    - Surveillance data
    - Intrusion detection data

        - Personnel access restrictions data
        - Trustworthiness verification of workers data
    - Monitoring of nuclear material and equipment
    - Training of professional security staff retroactive to past threats
    - False positives complicate operation
    - Evolving cyber threats or cyber-physical threats
- For nuclear safeguards
    - Data on the movement of special nuclear material includes:
        - Data from remotely operated cameras
        - Data from motion sensors
        - Data from radiation detectors
        - Data from records
        - Data from accounting
        - Data from inventory databases
        - Data from in-field inspections

2. Opportunities

Operation and Maintenance

- Human errors
- Labor costs
- Human responsibilities in interpreting and understanding complex scenarios
- AI capabilities in future reactors

Nuclear Materials and Advanced Manufacturing

- Efficiency of new NPPs
- Costs of NPP construction and maintenance
- Data reduction and data transfer between researchers
- Safety and reliability monitoring outside the plant
- Density functional theory
- Potential energy models for atomistic simulations
- Selecting or designing a material to fulfill the desired material properties
- Connecting modeling efforts to experimental data
- Correlations between models and experimental results across length scales
- Roadblocks to AM and nuclear materials (long term)
- Identifying missing data (long term)
- Images to augment existing material datasets (long term)
- Fusing together multiple material datasets (long term)
- Historical data to support knowledge transfer
- Predictive maintenance

- Design components with the desired materials properties under a given environment

Cybersecurity

- Opponent characterization
- Imprudent behavior characterization
- Real-time cybersecurity validation
- Support for sophisticated defenses
- Simulators with interactive training capabilities

High-Performance Computing & Massive Computation

- Utilization of high-fidelity models to reduce over-conservatism
- Generate synthetic data to train ML models
- Generate synthetic data that adhere to the laws of physics and can replace the data from limited publicly available datasets
- Generate synthetic data to overcome data scarcity or dataset limitations imposed by safety or security concerns
- Automatically select models and parameters based on problem parameters (HPCMC3)
- Generate initial guesses for solutions generated by iterative methods
- Train chatbots that can facilitate information retrieval, ingest documentation, and generate reports and summaries
- Train chatbots that can facilitate learning and the training of personnel, thus ensuring adherence to safety protocols and the implementation of better practices
- Identify the plethora of explainable, reliable applications of traditional AI

Digital Twins

- Reaching their full potential
- More accurate and nuanced predictions
- Operate faster than real time
- Allow for dynamic adjustments
- Offer advanced applications that can dynamically adapt to real-world conditions and deliver insightful predictions and analyses
- Require novel approaches to integrate ML, M&S, and real-time data
- Data processing techniques may exist that are leveragable for DTs
- Operational efficiency and safety
- Predictive maintenance
- Emergency response planning
- Different objectives
- Kinetic digital twin (i.e., on demand)
- Autonomous controls and operations

- Security applications

Nuclear Non-proliferation

- Extrinsic data beyond material data (e.g., low-level waste, liquid effluent in the facility sump and cooling pumps, and airborne effluents downwind of nuclear facilities).
- Extrinsic data to characterize human factors (e.g., days not worked, slips, and trips).
- Better-informed model
- Combines diverse datasets from siloed domain spaces

3. Challenges

a) General Challenges

- Data
  - Availability (of data on everything from plant components to the plants themselves under a variety of operating conditions)
  - Sharing (between commercial entities and researchers)
  - Diversity
  - Curation policy
  - Reliability, validity, timeliness
  - Security
  - Update frequency
  - Small quantities
  - Without detectable radiation field
- Acceptance
  - Nuclear culture is averse to change
  - Barrier of regulatory approval

b) Domain-specific Technical Challenges

Digital Twins

- Validating digital twin models
- Trade-offs in accuracy
- DT operational modes (real time, queried, etc.)
- Real-time assessment of model reliability and validity

Cybersecurity

- Game identification

Nuclear Non-proliferation

- Tracking small quantities
- Tracking quantities without detectable radiation field

Resources

- Computational
  - Limited resources to train large language models (power, memory)

  - Prohibitive costs associated with using cloud services to provide computational resources needed to address HPC resource limitations
  - Data sharing medium security
- People
  - Difficulty in hiring and retaining data scientists
  - Red and blue teams specifically dedicated to ML algorithms

4. Risks

- Regulation
  - Regulatory entanglement
  - Disharmonized regulations
  - Regulation as a barrier to innovation
- Privacy Violation regarding Individuals and Companies
  - Privacy infringement
  - Extracting confidential original data from models
  - Crossing the boundaries of individuals' privacy and companies' property restrictions
  - Compromising individuals' privacy
  - Security breaches in proprietary datasets, leading to data manipulation and misuse
- Data
  - Quality
    - Data and information overload
    - Data gaps (interruption)
    - Poor coverage of the data space, leading to poor extrapolations
    - Bad, incomplete, missing, or adversarial data, leading to inaccurate models
  - Security
    - Security breaches in proprietary datasets, leading to data manipulation and misuse
    - Adversary access to centralized data
    - Data poisoning
- Models
  - Biased models
  - Inaccurate models
- Outcomes
  - Improper assessment and diagnosis (improper models)
  - Poor extrapolations
  - Biased decision making
  - Unethical behavior
  - Unaccountable behavior
  - Unintended behavior
  - No accountability
- People
  - Users unfamiliar with AI capabilities and limitations

Appendix B

Agenda

Day 1 (November 7th)

8:00 am – 8:30 am	**Check-in and Continental Breakfast** ***Room:*** ***Hayden Room***
8:30 am – 8:45 am	**Welcome and Introduction, Remarks by Carol Smidts (OSU) and Nancy Lybeck (INL)** ***Room:*** ***Scioto Room***
8:45 am – 9:15 am	**Presentation: Office of Nuclear Energy Overview of Big-Data Activities** **Presenter:** Daniel Nichols (DOE) ***Room:*** ***Scioto Room***
9:15 am – 9:50 am	**Intro Presentations – Operation & Maintenance, Nuclear Materials and Advanced Manufacturing** **Title:** Big Data for Operation and Maintenance: Research and Applications **Presenters:** Cody Walker (INL), Carol Smidts (OSU) **Title:** Big Data in Post Irradiation Examination Experiments and Modeling **Presenters:** William Chuirazzi (INL), Mathew Swisher (INL) ***Room:*** ***Scioto Room***
9:50 am – 10:15 am	**Coffee Break** ***Room:*** ***Hayden Room***
10:15 am – 11:15 am	**Session 1 – Operation & Maintenance** **Chairs:** Carol Smidts (OSU), Cody Walker (INL) ***Room:*** ***Scioto Room***
10:15 am – 11:15 am	**Session 2 – Machine Learning in Nuclear Materials and Advanced Manufacturing** **Chairs:** William Chuirazzi (INL), Marat Khafizov (OSU), Mathew Swisher (INL) ***Room:*** ***Abbey Room***
11:15 am – 11:30 am	**Coffee Break** ***Room:*** ***Hayden Room***
11:30 am – 12:00 pm	**Report Out** ***Room:*** ***Scioto Room***
12:00 pm – 1:30 pm	**Lunch on your own**
1:30 pm – 2:05 pm	**Intro Presentations – Cybersecurity, High Performance Computing and Massive Computation** **Title:** Cybersecurity for Large-scale Data-driven Applications for Nuclear Power: Challenges and Opportunities **Presenters:** Katya LeBlanc (INL) **Title:** Enabling Scientific Machine Learning in MOOSE Using Libtorch **Presenter:** Matthew Anderson (INL) ***Room:*** ***Scioto Room***
2:05 pm – 3:05 pm	**Session 3 – Cybersecurity** **Chairs:** Carol Smidts (OSU), Katya LeBlanc (INL) ***Room:*** ***Scioto Room***

2:05 pm – 3:05 pm	**Session 4 – High-Performance Computing & Massive Computation** **Chairs:** Leonardo Moraes (OSU), Matthew Anderson (INL) ***Room: Abbey Room***
3:05 pm – 3:20 pm	**Coffee Break** ***Room: Hayden Room***
3:20 pm – 3:50 pm	**Report Out** ***Room: Scioto Room***
3:50 pm – 5:05 pm	**Demos/Posters/Presentations** **Title:** Toward Tactically Resilient Operational Technology for NPPs **Presenter:** Quanyan Zhu (NYU) **Title:** Explainable AI and ML for Cyber Resiliency of Nuclear Power Plants **Presenter:** Rakesh Podder (CSU) **Room: Scioto Room**
6:00 pm	**Dinner: (No host)**

Day 2 (November 8th)

8:00 am – 8:30 am	**Continental Breakfast and Introduction to Day 2** ***Room: Hayden Room***
8:30 am – 9:00 am	**Presentation: Artificial Intelligence Preparedness – A Regulatory Perspective** **Presenter:** Matthew Dennis (NRC) ***Room: Scioto Room***
9:00 am – 9:35 am	**Intro Presentations – Digital Twins, Nuclear Non-Proliferation** **Title:** Challenges and Gaps with Big Data for Nuclear Digital Twins **Presenter:** Vaibhav Yadav (INL) **Title:** Digital Engineering, Artificial Intelligence and Their Role in Safety, Safeguards, and Security by Design **Presenter:** Samuel Bays (INL) ***Room: Scioto Room***
9:35 am – 10:00 am	**Coffee Break** ***Room: Hayden Room***
10:00 am – 11:00 am	**Session 5 – Digital Twins** **Chairs:** Vaibhav Yadav (INL), Xiaoxu Diao (OSU) ***Room: Scioto Room***
10:00 am – 11:00 am	**Session 6 — Nonproliferation** **Chairs:** Samuel Bays (INL), Praneeth Kandlakunta (OSU) ***Room: Abbey Room***
11:00 am – 11:15 am	**Coffee Break** ***Room: Hayden Room***
11:15 am – 11:45 am	**Report Out** ***Room: Scioto Room***
11:45 am – 12:00 pm	**Closeout**

Appendix C

List of Participants

Name	Institution	Role
Prof. Tunc Aldemir	The Ohio State University	participant
Dr. Matthew Anderson	Idaho National Laboratory	session chair, presenter
Dr. Samuel E. Bays	Idaho National Laboratory	Session chair, presenter
Dr. William Chuirazzi	Idaho National Laboratory	session chair, presenter
Mr. Atitarn Dechasuravanit	The Ohio State University	participant
Mr. Matthew Dennis	U.S. Nuclear Regulatory Commission	keynote speaker
Dr. Xiaoxu Diao	The Ohio State University	session chair, workshop organizer
Prof. Praneeth Kandlakunta	The Ohio State University	session chair
Prof. Marat Khafizov	The Ohio State University	session chair
Dr. Sean Kil	Electric Power Research Institute	participant
Dr. Katya Le Blanc	Idaho National Laboratory	session chair, presenter
Dr. Leonard Lucas	Naval Nuclear Laboratory	participant
Dr. Nancy Lybeck	Idaho National Laboratory	organizer
Mr. Ryan Marcum	I&C Operative	participant
Dr. Leonardo Moraes	The Ohio State University	session chair
Dr. Daniel Nichols	U.S. Department of Energy	keynote speaker
Dr. Jonathan Nistor	Blue Wave AI Labs	participant
Mr. Rakesh Podder	Colorado State University	poster presenter
Mr. David Raab	Naval Nuclear Lab	participant
Mr. Md Ragib Rownak	The Ohio State University	participant
Prof. Carol Smidts	The Ohio State University	organizer, session chair, presenter
Dr. Mathew Swisher	Idaho National Laboratory	session chair, presenter
Mr. Pavan Kumar Vaddi	The Ohio State University	participant
Mr. Anthony Valiaveedu	U.S. Nuclear Regulatory Commission	participant
Dr. Cody Walker	Idaho National Laboratory	session chair, presenter
Dr. Vaibhav Yadav	Idaho National Laboratory	session chair, presenter
Mr. Vinicius Zanardo Rodrigues	The Ohio State University	participant
Prof. Quanyan Zhu	New York University	poster presenter