INL/RPT-23-74696
Revision 0

Light Water Reactor Sustainability Program

Assessment of Cloud-based Applications Enabling a Scalable Risk-informed Predictive Maintenance Strategy

September 2023

U.S. Department of Energy

Office of Nuclear Energy

INL/RPT-23-74696
Revision 0

Assessment of Cloud-based Applications Enabling a Scalable Risk-informed Predictive Maintenance Strategy

Cody Walker and Vivek Agarwal
Idaho National Laboratory

Tom Gruenwald and Jonathan Nistor
Blue Wave AI Labs

Pradeep Ramuhalli and Michael Muhheim
Oak Ridge National Laboratory

September 2023

Idaho National Laboratory
Light Water Reactor Sustainability
Idaho Falls, Idaho 83415

http://lwrs.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 current light-water reactor fleet uses time-based maintenance strategies to achieve high-capacity factors. But to make nuclear more competitive in the energy market, these reactors could utilize emerging artificial intelligence (AI) and cloud computing technologies to enable a cost-effective, predictive maintenance (PdM) strategy. This report examines the capabilities, feasibility, and regulatory concerns of cloud computing in relation to meeting nuclear industry needs.

The technical viability of cloud computing was analyzed using data from a boiling-water reactor's safety relief valve (SRV). The models were hosted on three different systems: a local personal computer, Idaho National Laboratory's high-performance computer (HPC) system, and Microsoft Azure. The data were loaded and processed, and two types of models were trained in an A/B fashion. Based on the speed at which these actions were completed, it was determined that cloud computing affords adequate computing resources. Additionally, the computing power can scale with the demanded load.

To enable cloud computing in the existing fleet, additional sensors, networks, and other requirements must be implemented to ensure a smooth transition from current maintenance strategies. However, the benefit is that the plants no longer need to manage their own servers, software, cybersecurity, and information technology (IT) support staff for in-house data analytics purpose. Many of these features can be offloaded to the cloud provider. A comprehensive analysis was completed, revealing the current annual cost of operating to be more expensive than using cloud computing resources.

Currently, the regulatory framework does not explicitly address AI applications, but the Nuclear Regulatory Commission (NRC) and other regulatory bodies are working to provide guidance to address gaps, rather than implementing new regulations to address the use of AI and machine learning (ML). But since many nuclear AI applications focus on non-safety-related components (e.g., balance-of-plant components), they will likely require little or no regulatory restrictions or needed approvals. Demonstrating how AI can improve the maintenance and operation of these non-safety-related systems seems the likely path forward for implementing AI and cloud computing resources inside nuclear power plants (NPPs).

Page intentionally left blank

ACKNOWLEDGEMENTS

This report was made possible through funding from the U.S. Department of Energy (DOE)'s Light Water Reactor Sustainability Program. We are grateful to Jason Tokey of DOE, and Bruce P. Hallbert and Craig A. Primer of Idaho National Laboratory (INL) for championing this effort. We thank John Shaver of INL for the technical editing of this report.

Page intentionally left blank

CONTENTS

FIGURES

TABLES

ACRONYMS

AI	artificial intelligence
CBRS	Citizens Broadband Radio Service
CPU	computer processing unit
DBU	Databrick Unit
DAS	distributed antenna system
FNN	feed-forward neural network
GPU	graphics processing unit
HPC	high-performance computing
IoT	Internet of Things
INL	Idaho National Laboratory
IT	information technology
LSTM	long short-term memory network
LTE	long-term evolution
ML	machine learning
MPN	Mobile Private Network
M&D	monitoring and diagnostic
NaN	not-a-number
NN	neural network
NPP	nuclear power plant
O&M	operations and maintenance
PdM	predictive maintenance
SRV	safety relief valve
V&V	verification and validation
VM	virtual machine

Page intentionally left blank

1. INTRODUCTION AND BACKGROUND

Operations and maintenance (O&M) activities are key to ensuring the availability and reliability of energy generated by nuclear power plants (NPPs) [1],[2]. O&M costs—including activities such as inspection, calibration, testing, and replacement—are among the major non-capital costs that contribute to the overall operation costs of NPPs. The three main maintenance strategies for ensuring availability, reliability, and safety are: (1) time-based periodic maintenance (referred to as preventive), (2) failure-based maintenance (referred to as corrective), and (3) condition-based maintenance (referred to as predictive). Over the years, the nuclear fleet has relied on time- and failure-based maintenance strategies for enabling their structures, systems, and components to achieve high-capacity factors. This has led to higher operating costs, presenting the existing fleet of light-water reactors with long-term economic sustainability challenges in the current energy market.

An ongoing R&D project entitled the Technology Enabled Risk-Informed Maintenance Strategy, conducted under the U.S. Department of Energy's Light Water Reactor Sustainability Program, is developing a well-constructed, scalable, risk-informed predictive maintenance (PdM) approach [3]. Such an approach requires computing resources and the availability of databases at the plant site or remote monitoring and diagnostic (M&D) center. Thus, an information technology (IT) team or M&D IT team is required at the plant site to ensure that these computing resources and databases are always operating and available, have updated hardware and software, and present no cybersecurity concerns.

In recent years, cloud computing has emerged as a dominant technology by virtue of its low cost, ability to host applications on various types of virtual infrastructures, and computing and storage adaptability. Cloud computing can be a cost-effective alternative to onsite storage and diagnostics, and even for M&D centers. To efficiently utilize signal processing algorithms for fault diagnoses, enhanced real-time data-driven machine learning (ML) and artificial intelligence (AI) techniques should be employed. Implementing a real-time PdM approach using advanced AI models can reduce costs associated with periodic equipment repairs, labor, and unnecessary outages [4]. The NPP industry should seek to leverage cloud computing's elastic processing power and high-availability computing resources by advantageously shifting data storage locations away from computers located at plant sites and over to cloud servers with integrated security management. Additional savings can be realized via the cloud service provider's pay-as-you-go method, which enables users to pay only for the specific resources required for the task at hand.

Cloud computing refers to utility-based computing resources accessed over the internet. On the cloud platform, virtualization techniques permit multiple simulated environments and resources to be generated by a single physical hardware system, using a type of software referred to as a hypervisor. The hypervisor provides an interface to various resources hosted on physical hardware systems and distributes them appropriately to secure environments known as virtual machines (VMs). Services, applications, and infrastructure resources on the cloud are available, with resource pooling and rapid elasticity (i.e., automatic scaling of dynamic resources), to users on-demand through network access. Subscribing to cloud platforms requires licensing agreements/management, security compliance, certificates, and the meeting of any regulatory measures required by the cloud provider. The advantages and disadvantages of various cloud platforms are discussed in [5].

The National Institute of Standards and Technology's Special Publication 800-145 describes cloud computing as "a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction" [6]. The cloud model in that document consists of five essential characteristics, three service models, and four deployment models. For details, see [7].

The research presented in this report is an extension of prior research activities performed to develop a cloud-based application for enabling a scalable risk-informed PdM strategy at NPPs regarding the implementation of digital monitoring at NPPs [6,7]. In the present research, Light Water Reactor Sustainability Program researchers at Idaho National Laboratory (INL) collaborated with Oak Ridge National Laboratory and Blue Wave AI Labs to demonstrate the technical feasibility of cloud computing in solving a safety relief valve (SRV) prognostic problem. Three different resources (i.e., a local desktop, INL's high-performance computing [HPC] system, and Microsoft Azure) were compared in regard to this problem. The Oak Ridge National Laboratory research team evaluated the potential regulatory and security concerns associated with using cloud computing for nuclear power applications, whereas the Blue Wave AI Labs team studied the economic feasibility of enabling NPPs to utilize cloud computing.

Being already available at INL, the Microsoft Azure cloud platform was utilized in the context of this report to assess the technical feasibility of applying cloud computing to the SRV prognostic problem. However, the lessons learned are expected to remain relevant for other cloud-based computing applications as well, with few differences. The present research takes advantage of the proposed NPP cloud-based high-level architecture depicted in Figure 1 [6,7]. This architecture is briefly described below; for additional details, see [6,7].

NPP site wireless sensor network: The left-hand side of Figure 1 (i.e., "local plant site") illustrates the local NPP site connectivity that supports various network types among sensors and actuators, the Internet of Things (IoT) Hub gateway, and VM servers. For local wireless connectivity, a distributed antenna system (DAS) long-term evolution (LTE) combines wireless amplifiers and fiber optic cables to distribute wireless signals to antennas over a wide frequency range (kHz to GHz). Large volumes of sensor data are continuously collected at the plant site and transmitted to the cloud via a Wi-Fi router and the IoT Hub gateway point, ensuring a high data transmission rate with bidirectional communication. An edge device can be employed to enhance the data processing activities (e.g., data cleaning and feature extraction) prior to sending the data to the cloud resources. The VMs and data servers at the plant site are accessed from the cloud, using a virtual private network connection in order to maintain data security. Currently, some data sources are still recorded periodically by hand, but deployment of new wireless sensors can replace these infrequent, route-based measurements with frequent and reliable sensor measurements. Continuous collection and monitoring of sensor data, as enabled by wireless sensor monitoring, will ensure that equipment remains healthy and continues to cost-effectively operate within the acceptable limits. Thus, there is a need to ensure that the wireless network can support these new sensors and their corresponding frequency ranges.

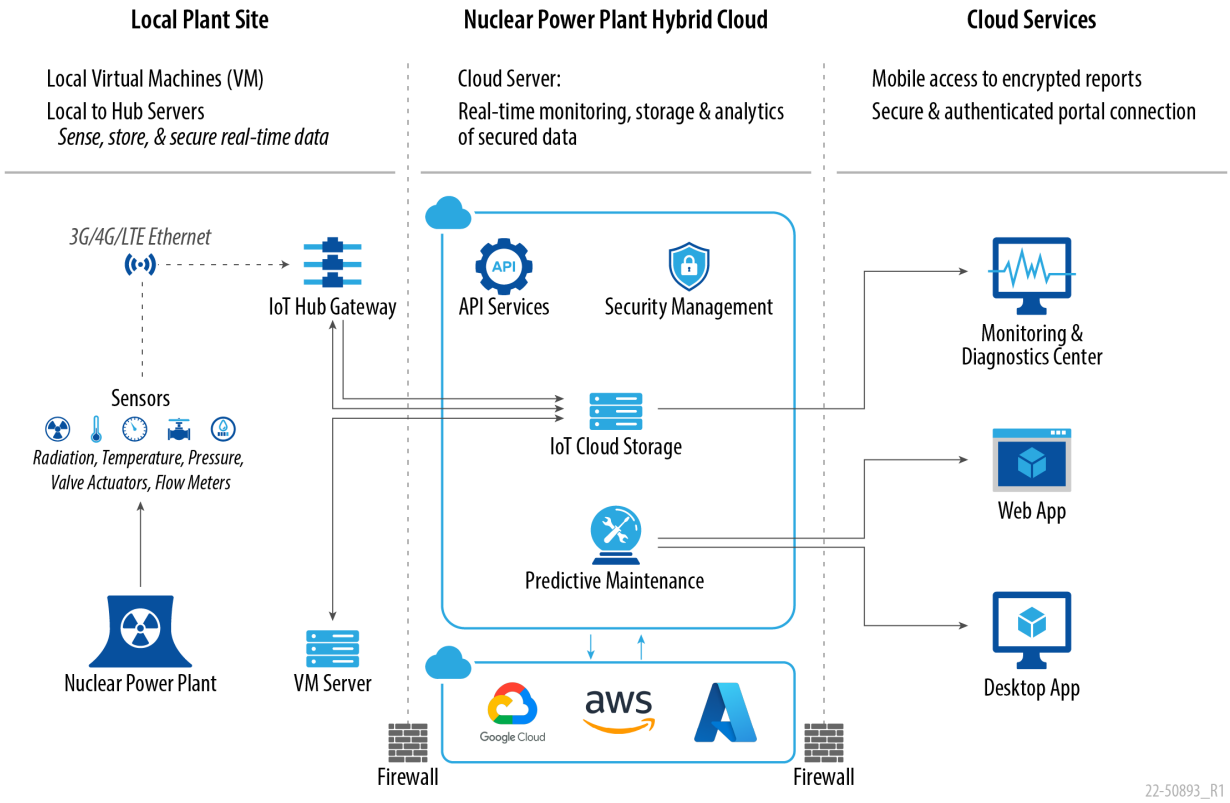

Figure 1. Proposed high-level architecture of the hybrid cloud [6,7].

NPP cloud network: The second aspect of the architecture proposed in Figure 1 encompasses further aspects of real-time data processing and analytics, storage, and security management. On the cloud platform, a series of applications can, as needed, be integrated into the central IoT Hub infrastructure adaptors. Data routing and authentication of all incoming messages from various sensor devices and locations is performed by the Event or IoT Hub gateway prior to conducting further data processing and AI data analytics. Web applications are also hosted by container and Kubernetes services for real-time visualization of archived historical data.

NPP cloud services: The third aspect of the proposed architecture encompasses the sharing of analysis reports and data visualizations via a web browser or portal. M&D reports are also made available to authorized plant and technical staff, and administrative management functions are made accessible to front-end users. Through services such as Azure, GE Predix, and AVEVA, cloud service providers have shown that data can be shared and demanded services provided via secure database servers.

The remainder of this report is organized as follows. Chapter 2 provides a brief overview of AI and some of its challenges. Chapter 3 covers the viability of cloud computing for nuclear PdM applications. Chapter 4 overviews the cloud computing economics, and Chapter 5 covers regulatory development related to cloud computing. Chapter 6 concludes the report by summarizing the findings and the path forward.

2. Overview of AI

Before determining what regulations (and guidance) would be applicable to a system that uses AI to make informed decisions, AI and its potential uses must first be understood.

AI is an umbrella term that refers to any computer algorithm that makes decisions intended to mimic or replace those made by humans. AI represents a wide array of computational techniques that enable computers and robots to solve complex, seemingly abstract problems previously only solvable through human cognition. The term AI encompasses many types of learning, including ML, natural language processing, deep learning, and data science [8]. The European Union Aviation Safety Agency defines AI as "any technology that appears to emulate the performance of a human." ML refers to the ability of computer systems to improve their performance through exposure to data, without having to follow explicitly programmed instructions. Deep learning, a subset of ML, has emerged in recent years as a result of the advent of deeper NNs.

AI's usefulness lies in its ability to find patterns and recommend actions based on inputs from diagnostics, virtual sensors, intelligent control, aging management, preventive maintenance, anomaly detection, etc.

The workflow for an AI system is very straightforward and is comprised of three major components:

1. Input selection, curation, and data wrangling
2. AI processing
3. Output postprocessing.

As an example, if an AI system were used to identify NPP transients, it would take inputs from plant instrumentation and control (I&C) systems, perform analysis and inference, and generate outputs that predict the type of transients the plant will experience. Driving the AI system workflow is an AI model designed and developed by human developers.

2.1 Data Processing and Specifications

Data are central to any data-driven AI system that learns the underlying models that represent input variables' relationships to the desired output. Data can be structured (e.g., relational databases in NPP data historians) or unstructured (e.g., event reports, specification documents, and nondestructive testing images and files).

The biggest hurdle in data science is always the data. Models are only ever as good as the data. In most cases, the data infrastructure for currently operating NPPs was built before AI or ML had even been conceptualized. As a result, the data from operating plants are siloed, incomplete, and filled with errors that must be dealt with before a ML model can be developed. About 45% of ML development is spent collecting, cleaning, and organizing the data so that they can be used in the model [9]. Often data undergo multiple processing steps before they can be used in an AI system, as per International Standards Organization [ISO]/International Electrotechnical Commission [IEC] 22989 [10].

2.2 Autonomy and Automation

AI is often applied to systems that can control physical actuators or trigger online actions. When AI comes into contact with the everyday world, issues pertaining to autonomy, automation, and human-machine teaming arise. Autonomy refers to a system's ability to operate and adapt to changing circumstances, without the need for human control. The different levels of autonomy are addressed in ORNL/SPR-2023/3072 [10].

2.3 Challenges

AI applications will produce unique challenges—not only in terms of how it is used, but also from a regulatory perspective. Because of the uniqueness of AI characteristics as compared to typical digital I&C

systems in NPPs, specific considerations must be made for AI:

- Autonomy – Higher autonomy levels indicate less reliance on human intervention or oversight and may therefore require greater regulatory scrutiny of the AI system.
- Cybersecurity and data security – As more data are generated, the incentives for stealing or modifying those data equally increase. The AI models themselves may need to be secured against unauthorized access and modification.
- Trustworthiness – AI applications have generated certain challenges stemming from their high complexity and low level of reliability. To address these, licensees must demonstrate the trustworthiness of their AI.
- Transparency – AI, particularly when using deep NNs, operates as a black box. The input and output of the system are observable, but the computational process leading from one to the other is difficult for humans to comprehend. It is particularly difficult for humans to understand what such a system has learned, and thus how it might react to input data that differ from those used during the training phase. The lack of transparency and explainability of these systems leads to a fundamental problem of predictability. A ML-based system might fail in ways unthinkable to humans, as engineers do not fully understand its inner workings. The lack of transparency is problematic from a regulatory standpoint, as it complicates the task of identifying the cause of a problem and attributing responsibility when something goes wrong. This type of AI may increase regulatory uncertainty.
- Bias in Decision Making – ISO/IEC TR 24027:2021, Information technology —AI — Bias can exist in AI systems and AI-aided decision-making processes. Measurement techniques and methods for assessing bias are described to address and treat bias-related vulnerabilities. All AI system lifecycle phases are in scope, including but not limited to data collection, training, continual learning, design, testing, evaluation, and use.
- AI verification and validation (V&V) – As in the application of digital I&C systems for the nuclear industry, V&V will play a central role in enabling the use of AI systems. However, because of the black-box nature of some AI systems, special considerations must be made in their V&V processes, particularly for scenarios in which only a subset of the AI system can be verified, validated, or both.

3. VIABILITY OF CLOUD COMPUTING FOR NUCLEAR

Cloud computing could generate cost savings by replacing many onsite features (e.g., IT personnel, cybersecurity experts, data historians, and onsite diagnostic centers). Many features such as cybersecurity and software updates are handled locally, when they could easily be offloaded to external services. Microsoft Azure Government cloud has already established expectations for security levels and uptime rates, and these expectations are continually updated as needed.

3.1 Description of the Problem and Machine Learning Architecture

To demonstrate the technical feasibility of cloud computing, a SRV prognostic problem was solved using three different resources: a local desktop, INL's HPC system, and Microsoft Azure—all of which were compared with each other in terms of ML model accuracy and training time.

A SRV is a nonpowered component that relieves excess pressure from steam lines or pressure vessels. Figure 2 illustrates a three-stage SRV design. During the SRV's operation, four sensor measurements were being recorded:

4. One thermocouple (in the second stage) that measured the temperature of the valve body
5. One downstream thermocouple that measured the temperature of the valve discharge
6. Two redundant thermocouples in the pilot stage.

Temperature data are relevant to SRV operation and health. Data from each thermocouple were recorded every second throughout a 14-month collection period. After downsampling to once per minute, the data totaled 46.12 Mb.

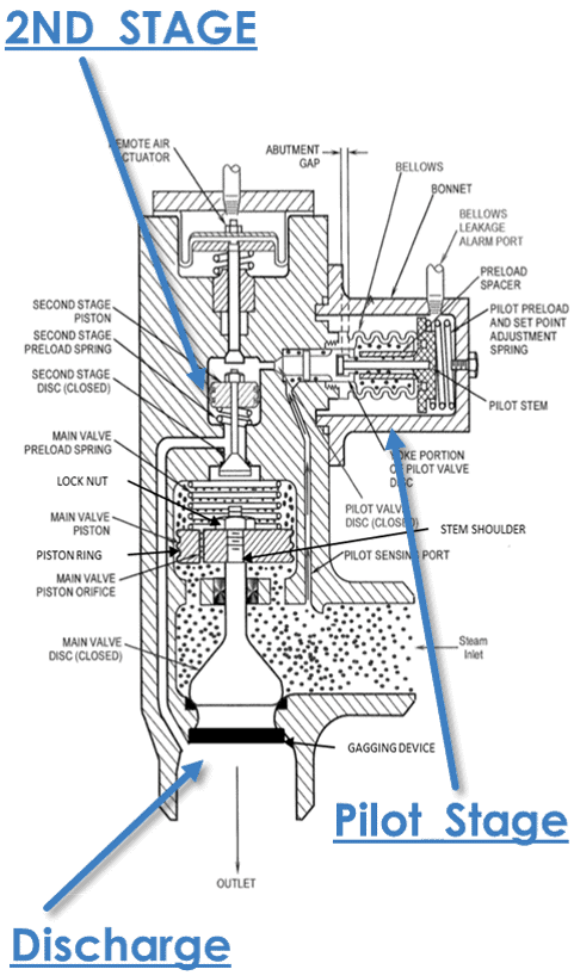

Figure 2. SRV diagram marking the sensor placements [11].

The goal of this study was to make forward predictions about each recorded variable. The models would then be updated with new information to see how continuously updating the model in an A/B fashion could help improve model performance as the system changes over time (see **Figure 3**).

In our approach, input data are first sent (either in a continuous or batch fashion) to be preprocessed. The preprocessed data are then sent to a pretrained ML model (Model A) for making diagnoses and predictions. Model B is a clone of Model A and is continuously updated with new data. Once Model B is verified as being stable and performing better than Model A, Model B replaces Model A as the pretrained model, and the process of training a new model begins anew.

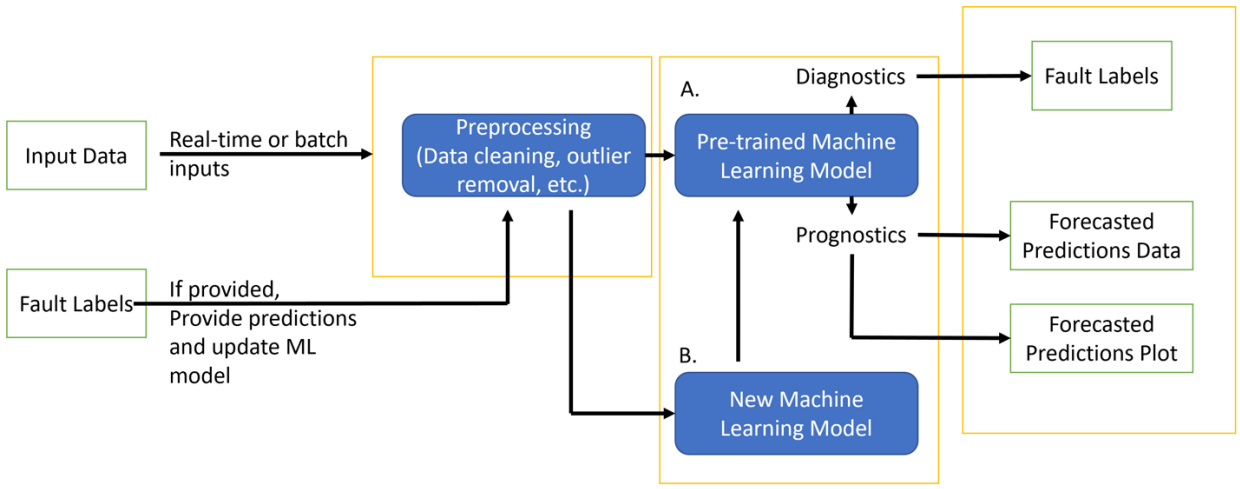

Figure 3. Model A/B testing.

Two types of ML models were used to make the forward predictions: a classical feed-forward neural network (FNN) and long short-term memory (LSTM) network. In FNNs, which are composed of several layers (i.e., the input layer, 64 neuron layer, 32 neuron layer, and dense output layer), only forward transmission of the data occurs. On the other hand, a LSTM network is a recurrent neural network (NN), so certain information is transferred from both the input and output layers. This recurrent nature enables long-term dependencies in time series data to be learned [12]. The architecture used in this study consisted of five layers (i.e., the input layer, 124 LSTM layer, 64 LSTM layer, 8 LSTM layer, and dense output layer). Each model's initial training entailed a maximum of 10 epochs, with a patience of 3 epochs when considering the validation loss. The procedure in this document encompassed several stages, each of which was timed:

1. Loading data
2. Preprocessing inputs
3. Training/testing Model A
4. Training/testing Model B

We also analyzed how long the data took to upload from the data source to Azure. This data ingress measure can determine which would be technically feasible: continuous or batch uploading—an important choice in determining which systems (i.e., safety or non-safety-related) could potentially be effectively monitored by cloud computing. For the present research, raw SRV data from four .mat files (one file per variable) were loaded into an Azure table.

Dat was then moved into a singular Pandas DataFrame for preprocessing. Preprocessing consists of various steps for conditioning the data before inputting them into ML models, including not-a-number (NaN) removal, outlier removal, downsampling, thresholding, smoothing, and standardization. NaN removal consists of using interpolation to replace values missing due to sensor or recording errors. Outlier removal identifies outliers that are four standard deviations away from the data and replaces them with interpolated values. These data are recorded every second, but as temperature data are unlikely to significantly change during such a short time interval, the data were downsampled to once per minute to significantly reduce the amount of total data required during training. Thresholding was applied to some of our signals, such as one in which a long-form, unrealistic value was recorded (such errors are typically the result of a sensor fault or connection issue). Thresholding afforded a means of identifying and replacing this persistent outlier. A median filter with a rolling window was applied to smooth the signal and remove noise throughout the preprocessing. Lastly, the data were standardized to a mean of 0 and a standard deviation of 1 for improved ML model performance.

The cloud computing resource used for this research was Microsoft Azure. However, other cloud computing platforms such as Amazon Web Services and Google Cloud Platform offer similar services. Within Microsoft Azure, Azure Databricks was used for hosting the data and training the models. Azure Databricks is an open analytics platform for creating, deploying, distributing, and maintaining enterprise-level data, analytics, and AI solutions [13]. It provides several useful features for building data engineering workflows, including workflow scheduling, security management, data discovery, computational management, ML, and source control with Git.

In demonstrating the technological viability of cloud computing, the total amount of time to ingress the data (in batch fashion, in our case), train the model, evaluate the component state, and egress the results is vital in regard to implementation feasibility. The whole process also goes hand-in-hand with how much lead time is required when reacting to a diagnosis or prediction made by the ML model. For example, if maintenance personnel require 3 days to respond to a damaged motor (e.g., find replacement parts and plan and perform the maintenance), the diagnostic model must alert maintenance and diagnostic personnel of the required maintenance at least 3 days prior to the occurrence of an incipient fault. Thus, the times required to ingress the most recent data, train a model (if needed), evaluate the data, and egress the results were recorded alongside the accuracy of the models themselves.

For comparative results, Azure resources were contrasted to both a local desktop and INL'S HPC system (see **Table 1** for the specifics). In this case, the Azure resources consisted of a general-purpose kernel for making cost-effective calculations. If required by the demand, Azure's computational performance can be increased by using more expensive nodes.

Table *1*. Comparison of computing resources.

	Local	HPC (CPU)	HPC (GPU)	Microsoft Azure
Processor	Intel CITM i7-9700 CPU @ 3.00 GHz	2 Intel Xeon 8268 CPUs @ 2.90 GHz 24 cores per CPU	NVidia Tesla V100 32 GPU	28–112 Gb Memory 8–32 Cores
Installed RAM	32 Gb	8 Gb RAM per core	32 GB RAM	14 Gb Memory

3.2 Cloud Computing Results

Microsoft Azure offers flexible resources that can be scaled up or down according to demand. For this test, the demand was rather low, as we were training on a single valve with 14 months' worth of data. However, as demand increases when monitoring multiple systems and components, the expense and processing power will increase in correlation to the application needs. The Microsoft Azure results, shown in Table 2, can be improved by selecting different Azure resources. The resources selected for the present research consisted of a cost-effective, general-purpose cluster that featured 14 Gb RAM with 4 cores per worker (based on 2–8 workers). The worker role inside Azure run applications and service level tasks. The number of workers automatically scales based on load. Azure's pay-as-you-go approach means that only the actual amount of computer usage must be paid for, including data storage, networking, and data processing. A Databrick Unit (DBU) is a normalized unit of processing power that Azure uses for pricing purposes. This particular cluster cost 2–7 DBU/h. To put that into perspective, the cost of data storage and training for 10 ML models for the FNN test was about $7 for the entire training process.

Table 2 shows how quickly each resource was able to load, process, and train the various models. Loading was the one area in which Azure seemed to come up short in comparison to the local desktop and INL HPC system. The loading process involved uploading to a table, then reading from said table into the Pandas DataFrame. This process could perhaps be streamlined using other Azure features such as Azure Storage, Azure SQL Database, Azure Cosmos DB, or a combination thereof. Uploading the 1.3Gb of raw data into Azure's table format took 4 minutes and 8 seconds to complete, while converting from the table

to a Pandas DataFrame took an average of 14.67 seconds. Each resource took roughly the same amount of time to preprocess the data, with the local resources performing slightly better (at 34.89 seconds). The loading/preprocessing times were identical across all HPC resources, and so are not shown.

Table 2. Comparison of how quickly each computing resource completed certain tasks.

Cloud Computer Speed Testing (in seconds)					
Test	Local	HPC (GPU)	HPC (CPU)	HPC (multi-CPU)	Azure
Loading Data	3.26 ± 0.10	2.87 ± 0.02	N/A	N/A	14.67 ± 0.29
Preprocess	34.89 ± 0.10	37.94 ± 0.02	N/A	N/A	38.00 ± 0.29
Train FNN 1	8.09 ± 2.01	27.57 ± 5.59	7.18 ± 1.97	11.14 ± 2.51	7.12 ± 2.14
Update FNN 2	0.97 ± 0.06	2.62 ± 0.2	1.04 ± 0.027	—	1.07 ± 0.03
Train LSTM 1	103.67	89.62	126.68	162.47	96.07
Update LSTM 2	8.78	7.76	14.5	—	8.72

More interesting results were found while training the FNN and LSTM models, due to the size and time span of the SRV data. For training the original FNN (i.e., FNN 1), the quickest times were recorded by Azure and the HPC single-unit computer processing unit (CPU), which produced average training times of 7.12 and 7.18 seconds, respectively. There was a slight increase in time with the multithreaded CPU (due to the overhead of communication between CPUs), and a significant increase when using a graphics processing unit (GPU). The GPU is slower than the CPU in this situation, as it can be quite computationally expensive to call the GPU, copy the data to it, perform the calculation/training, and retrieve the data. When the operation has few parameters—as in our case, featuring just four predictor variables—the overhead of calling the GPU kernel outweighs the benefits. As the input matrix increases in size (as is the case when training the LSTM), we see the benefit of using a GPU over a CPU. The LSTM is a recurrent NN, meaning that current predictions rely on past predictions in a recurrent fashion. Due to this recurrent nature, multithreading CPUs are not expected to speed up the process, as there is no way to process the training in parallel and thus the increased training time is due to the overhead of communication between multiple CPUs for no real gain. The HPC GPU was the fastest at training the LSTM models, as the aforementioned recurrent nature enabled more calculations, which the GPU excels at handling. When updating Model 1 with Test 1 data to create Model 2, each model used 4 epochs. Each resource and model type saw a substantial speed increase when working with a pretrained model on a smaller dataset.

Table 3 shows the performance of each model. For the SRV problem, the LSTM results were comparable to the FNN results, perhaps due to the size or (lack of) complexity of the SRV data. However, in each case, Model 1 performed better during Test A than during Test B, most likely because of the temporal proximity of training data to the testing data (with closer being better). Figure 4 shows where the splits for the training, validation, Test 1, and Test 2 data reside, as well as giving the LSTM models' predictions. A large seasonal component is reflected in one of the temperature variables, as seen in Figure 4.

Table 3. Comparison of how well each model did at making forward predictions about the SRV variables.

A/B Model Testing		
Test	FNN (RMSE)	LSTM (RMSE)
Model 1 Test A	1.361 ± 0.679	1.411
Model 1 Test B	2.559 ± 1.264	2.792
Model 2 Test B	0.593 ± 0.300	0.504

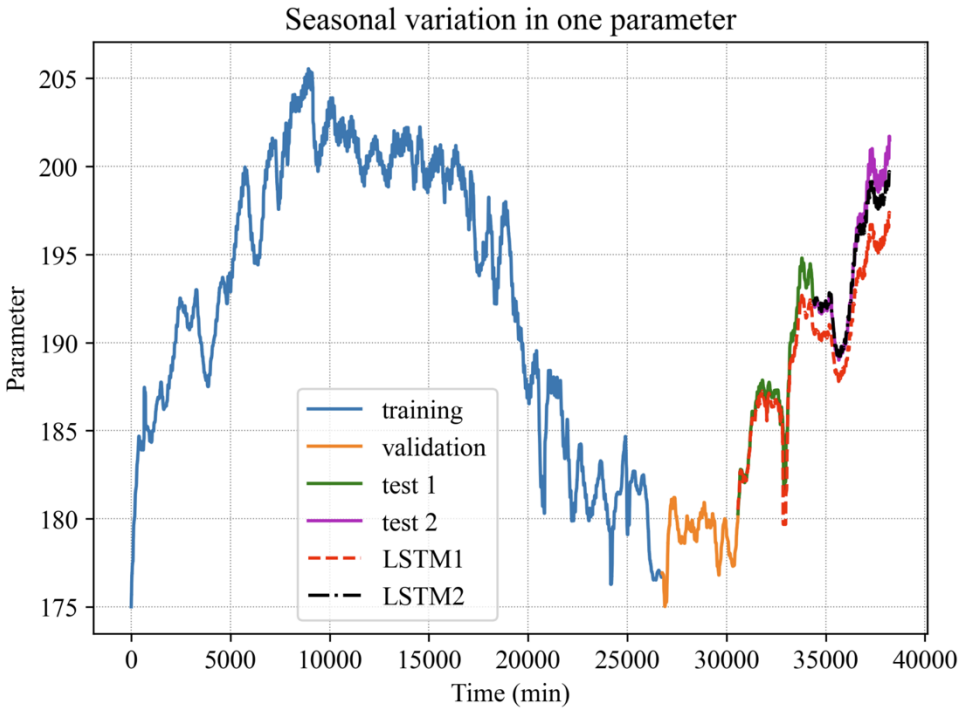

Figure 4. This SRV variable reflects a large seasonal component.

To better illustrate model performance, **Figure 5** focuses solely on the testing portion of the data. LSTM 1 degrades in performance over time. There is very little error at around time step 31,000; however, a bias slowly forms. When LSTM 1 is updated with the data from Test 1 so as to create LSTM 2, the extent of the bias is mitigated. This shows the importance of continually updating models by using the most recently measured data, as relationships within the system may change over time.

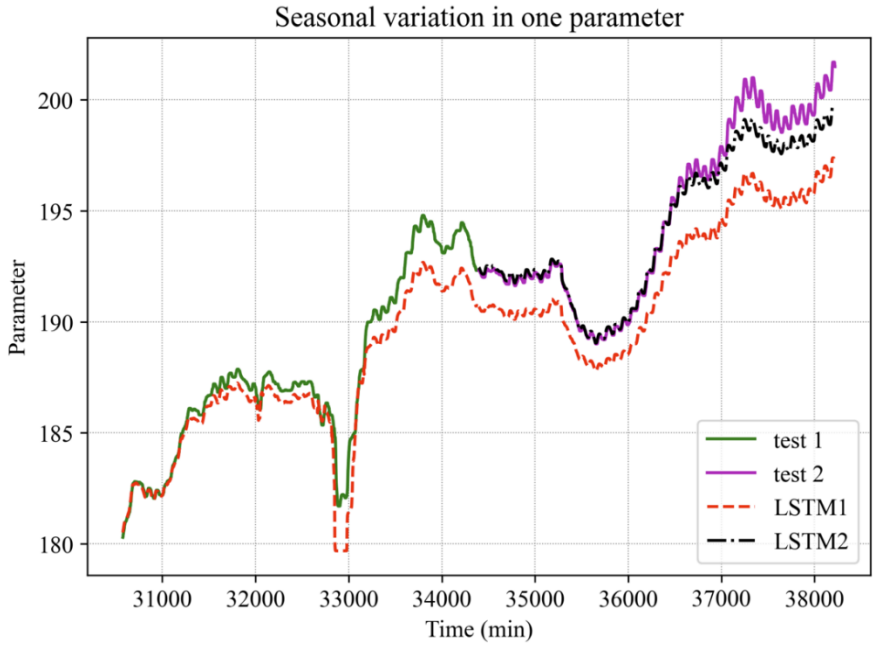

Figure 5. Models were used to predict future values of the variable. Updating the model with more recent data improved its performance.

From a technological perspective, the Azure resources compare favorably with other existing resources. The advantages lay in the scalability of Azure resources and the pay-as-you-go strategy. Once an initial pretrained model is created, it can be continually updated to ensure accuracy as plant systems age and degrade. Transfer learning is an additional potential benefit if multiple utilities move to the same cloud service. At the moment, transfer learning is not a viable solution for improving models, as plants locally host data and are concerned with maintaining their privacy. If all their data are located on the cloud, there may arise an opportunity to anonymize the data for joint collaboration or even create a federated learning model by developing siloed models and aggregating those into a master model to improve model performance across the nuclear fleet.

4. CLOUD COMPUTING ECONOMICS

To analyze the financial implications of moving non-safety-related data and PdM applications to the cloud, we must first determine the baseline (or present) cost of hosting the data at the utility's IT infrastructure. Here, the components of this cost will be the IT staff, infrastructure, physical computers, network equipment, software acquisition and maintenance or subscription costs, back-up costs, and long-term storage costs. There is also considerable manual overhead in terms of physically reading many of the sensors, as well as inherent inefficiency in gathering data for maintenance analyses and activities. All these costs will be included the base case. For the target architecture, a new set of wireless sensors will be purchased. Those costs will be included as well.

Cloud costs were determined based on recent tests by Blue Wave AI Labs on the AWS GovCloud. The costs for Azure are assumed to be similar. The costs of hosting and periodic training on the data will be included, though the costs associated with the initial model creation will not, as that is expected to occur offline.

Network costs will be calculated by comparing DAS implementations for the plant side against Mobile Private Network (MPN) costs. MPNs are inherently cheaper and more flexible than DAS networks, thanks to the advent of a new spectrum not controlled by the traditional carriers. Specifically, the Citizens Broadband Radio Service (CBRS), which is the unlicensed spectrum in the United States that can be leveraged for private 4G or 5G 4G LTE networks, consists of 150 MHz of spectrum in the 3.5 GHz band. For the MPN case, not only are the initial costs lower, but very expensive mobile charges are avoided as a result of there being no direct bulk connectivity to the wireless carrier.

4.1 Plant Side

For plant-side PdM (see Table 4), the largest driver of current base-case expense is IT staff. IT staff numbers are assumed for three shifts. The next biggest expense stems from various inefficiencies in collecting and analyzing data. Many sensor readings must be taken manually by staff who physically walk the facility, record readings, and enter them into the maintenance computers. However, this is not their only responsibility, so only a fraction of their time—and thus expense—is put toward the annual cost.

Table 4. Current base costs for PdM.

Hardware	Number of Items Saved	Item Cost	Total Cost	Annual Costs	Monthly Costs	Assumptions
Servers	18	$4,500.00	$81,000.00	$16,200.00	$1,350.00	Replaced every 5 years
Network Elements (e.g., routers)	30	$800.00	$24,000.00	$4,800.00	$400.00	Replaced every 5 years
Software						
Commercial Software	Base Cost	$200,000.00	$200,000.00	$30,000.00	$2,500.00	Maintenance contracts
Purpose-built Software	Base Cost	$500,000.00	$500,000.00	$75,000.00	$6,250.00	Contract programming
IT Support Staff Average Salary	17	$150,000.00	$2,550,000.00	$2,550,000.00	$212,500.00	—
Offsite Backup	—	—	—	$3,600.00	$300.00	—
Cybersecurity	—	—	—	$20,000.00	$1,666.67	—
Operational Staff	35	$150,000.00	$5,250,000.00	$262,500.00	$21,875.00	Fraction of their time
Manual Sensor Reading	8	$85,000.00	$680,000.00	$680,000.00	$56,666.67	Headcount for manual sensor reading/ recording
Facilities Costs						
Electricity	—	—	—	$300,000.00	$25,000.00	Yearly cost of electricity
Total	—	—	—	$3,942,100	$328,508	—

The following section covers the estimated plant-side upgrades to enable cloud connectivity. This includes any onsite servers or wiring necessary to enable seamless connectivity, but also includes any new sensors or wireless technologies needed to enable wireless sensor monitoring. This section computes the costs of purchasing new sensors to implement total wireless data gathering. It also compares the cost and efficacy of DAS networks as opposed to CBRS-based private mobile networks.

4.1.1 Wireless Sensors

Wireless sensors are making rapid advances. They have both wireless and Bluetooth connectivity, and can be implemented independent of a wireless carrier, significantly reducing connectivity fees. Some have onboard processing to generate an alarm whenever sensors detect abnormal behavior, as determined by simple classification software constantly running on the processor. As data accumulates, it generates a "normal-not-normal" model. Additionally, it transmits all collected data to the cloud. These devices measure a host of elements such as temperature and acceleration. They are priced similarly to other wireless sensors but have the added capability of onboard processing and programmability. A range of sensors were surveyed in [7] for improving monitoring through an entire plant system, as shown in Table 5. These sensors include a high-grade category reserved for reactor-related sensors that require specific grading and hardening. This type of sensor is not required through the entire plant. For this report, the sensors will be for the balance of plant are assumed to be low/mid-grade. These sensors cost about $500 each and could easily be run through the commercial dedication process. Assuming that approximately 600 sensors are required and get replaced after 5 years, the annualized cost over the sensors' lifetimes would be approximately $136,000.00 per year (see Table 5). Ongoing maintenance of sensors, including the resetting of components and the replacement of batteries and devices, is estimated at half of one salary.

Table 5. Fixed costs for a range of wireless sensors [14,15].

Sensor Grade	Number of sensors	Price per unit	Subtotal	Acquisition Annualized Over Lifetime
Low grade	70	$300	$21,000	$42,000
Low/mid	120	$500	$60,000	$96,000
Mid-range	220	$800	$176,000	$242,000
Mid/high	120	$1,200	$144,000	$180,000
High grade	70	$2,500	$175,000	$196,000
Total	600	-	$576,000	$756,000

Table 6. Wireless sensor costs.

Item	Acquisition Costs	Acquisition Annualized Over Lifetime
Sensors (600)	$300,000.00	$60,000.00
Annual Maintenance (0.5 head count)	N/A	$76,000.00
Total	$300,000.00	$136,000.00

4.1.2 Network Aggregation Equipment

Signals from wireless sensors will have to be aggregated, formatted, and transmitted to the cloud, where the information will then be parsed and stored in the appropriate locations in the PdM database. Some wireless sensors can connect directly to the wireless network without requiring significant processing. We will take a conservative approach in which sensor data may be aggregated in different locations, and in which several routers are required to ingest the sensor data and forward them to the DAS or private mobile network. Maintenance is estimated at 10% of the acquisition cost. Per **Table 6**, the total cost for network aggregation equipment and maintenance is approximately $15,000.00 per year.

Table 7. Network aggregation equipment costs.

Item	Acquisition Cost	Annualized Cost
Network Aggregation Equipment	$50,000.00	$10,000.00
Maintenance Cost	N/A	$5,000.00
Total	$50,000.00	$15,000.00

4.1.3 In-building Network

Some type of wireless network must be installed to provide access points for receiving wireless data from sensors.

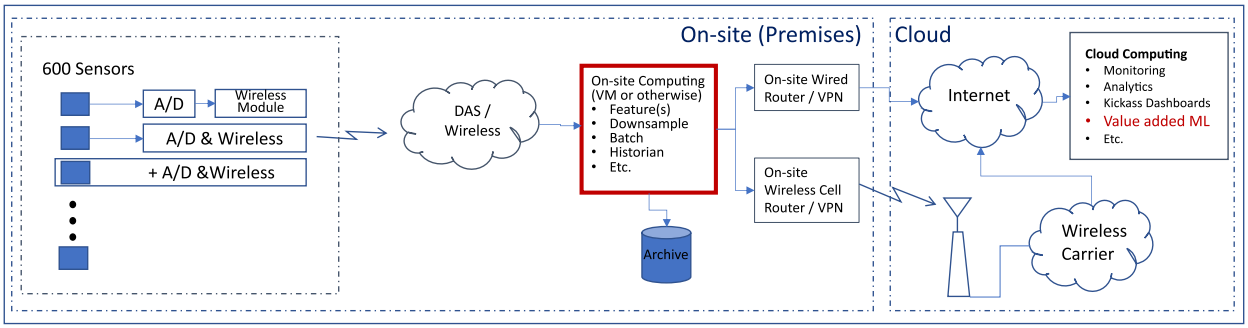

Figure 6. Example of a DAS network.

Figure 6 shows the entire network, with a DAS used as the networking system. The DAS must connect to a mobile carrier network so that the data can be transferred from the nuclear facility to the cloud via the carrier's mobile network. The currently implemented DAS consumes a great deal of power, whereas newer-generation DASs are far less power hungry. Some are even server-based and software configurable. Table 7 shows the typical annualized cost for a DAS to be approximately $222,628 per year.

Table 8. Estimated costs for a DAS system.

	Acquisition Cost	Annualized Cost	Assumption
DAS	$1,031,480.00	$103,148.00	10-Year life
Installation	$100,000.00	$10,000.00	—
Maintenance Cost	N/A	$103,480.00	10% of Acquisition
Carrier Charges	N/A	$6,000.00	100 mb/sec connection at $500/month
Total	$1,131,480.00	$222,628.00	—

4.1.4 Mobile Private Network

Figure 7 shows the base architecture for a MPN. This particular implementation is from LEMKO, a network equipment provider whose system is presently being trialed at the Dresden nuclear plant.

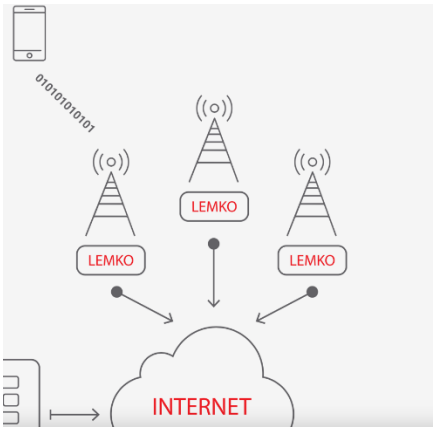

Figure 7. Example of an MPN, using LEMKO as the network equipment provider.

The base architecture provides a full LTE network that can aggregate traffic and, via the internet, send data to the cloud. The MPNs are typically 4G LTE, due to 4G LTE being solid, well-tested, and cheaper than alternatives such as 5G. Typical MPN network installations cost $100,000–$500,000, so a mid-price point of $250,000 will be used for the in-building network below.

4.2 Cloud Side

This section covers cloud-side expenses. There are three types of loads on the cloud: hosting costs, retraining costs, and storage costs.

Hosting costs encompass the cost of storing and hosting applications on the cloud. This load consists of using trained applications that are run at user request. Blue Wave AI Labs has extensive experience in hosting and supporting applications on the cloud. Generally, the cost is under $275 per month per application, but for added conservatism a cost of $500/month was used instead (see **Table 8**). The hosting costs are estimated to total $150,000 annually.

Table 9. Estimated hosting costs.

	Number of Applications	$/month	Total/month	Annual Cost
Model Hosting Cost	25	$500.00	$12,500.00	$150,000.00

Retraining costs are accrued by periodically retraining ML models when new data become available. This can be somewhat expensive. Retraining can potentially be performed by the application owner on other platforms; however, there are good reasons to do the retraining on the cloud. For example, the data already reside on the cloud, and maintaining export control requirements and security is more straightforward on the cloud. Three applications of differing complexity were retrained and the costs summarized in **Table 10**. It is assumed that 50 pieces of equipment will need to be supported. The average yearly cost to update and maintain models is $28,152 ± $5,080.

Table 10. Estimating costs for retraining and maintaining models with the most recent data.

Model Maintenance Cost	Per Hour	Number of Hours	Cost	Pieces of Equipment	Total Cost
App1 Retrain + Analysis	$3.67	8	$29.38	50	$1,468.80
App1–Debug - Models	$3.67	10	$36.72	50	$1,836.00
App1 Experiments	$3.67	25	$91.80	50	$4,590.00
Total Retrain	—	—	—	—	$7,894.80
App2 Experiments	$3.67	100	—	50	$18,360.00

Model Maintenance Cost	Per Hour	Number of Hours	Cost	Pieces of Equipment	Total Cost
App 2 Retrain + Analysis	$3.67	12	$44.06	50	$2,203.20
App2 New Site	$3.67	20	$73.44	50	$3,672.00
Total Retrain	—	—	—	—	$24,235.20
App3 Master Models	$3.67	40	$146.88	50	$7,344.00
App3 Finetune + Analysis	$3.67	10	$36.72	50	$1,836.00
App3 From Scratch Models	$3.67	15	$55.08	50	$2,754.00
App3–Debug - Models	$3.67	20	$73.44	50	$3,672.00
App3 Experiments	$3.67	200	$734.40	50	$36,720.00
Total Retrain	—	—	—	—	$52,326.00
	—	—	—	Average	$28,152.00

Storage Costs. Storage costs are relatively inexpensive and easy to calculate. Quotes from cloud providers come to $527 per year for 500 GB.

4.2.1 Direct Cloud Costs

Table 11 shows the total direct cloud costs of hosting and maintaining 25 applications along with their associated data. We assume model retraining is done on the cloud, but not model development. The cost of data storage may be underestimated, as a great deal of historical data could potentially be brought over to the cloud in addition to the data associated with each application. We also assume that it takes 10 IT personnel to provide continual upkeep and completes the model retrainings.

Table 11. Estimated direct cloud costs.

Direct Cloud Costs	Cost	Number of Applications	Total
Storage of 500 Gb/year	$567.00	25	$ 14,175.00
Total Model Retraining	$28,152.00	25	$703,800.00
Application Hosting	$6,000.00	25	$150,000.00
IT Personnel	$150,000.00	10	$1,500,000.00
—	—	Total	$2,367,975.00

4.2.2 Total Cloud Costs

The total cost of using the cloud to host maintenance data and PdM applications is calculated by combining the direct costs with the cost of installing the proper equipment in the facility in order to collect and transport the data to the cloud (see Table 12). This includes the sensor costs, data aggregation equipment, and in-building network. In addition to the costs already discussed will be a one-time labor cost for moving data and applications to the cloud. Experiments by Blue Wave AI Labs show it takes 10–15 staff days to move a single application. Including cleanup of the site facilities, we assumed approximately 5 staff years to perform the transition at a one-time cost of $600,000.

Table *12*. Total cloud installation costs.

Total Cloud Costs	Initial Costs	Annual Costs
Sensors	$300,000.00	$136.00
In-building Network	$1,131,480.00	$176,278.00
Network Aggregation Equipment	$50,000.00	$15,000.00
Installation and Cloud Set-up	$600,000.00	$60,000.00
Total	$2,081,480.00	$251,414.00

The finalized cost comparison in **Table 13** shows that using cloud resources is economically viable. After the initial capital costs of about $2 million, there is an annual savings of around $1.3 million. These savings are primarily through the reduction of the staff require to manually check the sensors and IT personnel required to maintain the cyber security. These savings do not consider any savings that could potentially occur with the deployment of additional sensors as they add to the overall situational awareness of the system.

Table *13*. Cost comparison of current onsite costs vs. cloud implementation.

Cost Comparison	Current Onsite Cost	Cloud Cost
Installation Cost	N/A	$2,081,480.00
Annualized Cost	$3,942,100.00	$2,619,389.00

5. REGULATORY REQUIREMENTS FOR AI

Regardless of how data are stored and shared, licensees must still be able to protect safety-related and important-to-safety functions. To do so, licensees must analyze digital computer and communication systems/networks, identify which assets require protection against cyberattacks, and establish, implement, and maintain a cybersecurity program for protecting those assets. The regulatory requirements for a cloud-based server are addressed in INL/RPT-22-70543 [7].

AI is transforming many fields, and the nuclear industry is eager to expand its use and capabilities. Though recent advances in AI have enabled many more possibilities, how it is used, its autonomy, and the needed regulatory perspective on its uses will slow its migration into the nuclear arena [16]. This review identifies the regulatory requirements related to using AI for evaluating data stored on cloud servers. As some technical aspects of AI are still evolving, and the regulatory framework for nuclear regulators has not yet been established, this report instead focuses on how AI could be used to evaluate data stored on a cloud-based server.

In terms of regulating AI applications in the United States, the NRC is still mostly in an information-collection phase, seeking input from stakeholders and holding workshops to better understand the technology and its potential nuclear applications [17,18].

When considering AI as applied to NPP O&M, three specific challenges arise in regard to licensing and regulatory activities [19,20]:

1. Quality/optimum input data. Data in digital form may be insufficient to employ ML and AI algorithms, and poor-quality data may lead to false output relationships or large uncertainties. Both data quality and data sufficiency are equally required.

2. Identification and selection of appropriate AI algorithms. The specific application will lead to a short list of suitable algorithms, and selecting the appropriate one will depend on factors such as the desired performance, size, and complexity of the training data; the scalability of the algorithm; and deployment and business case criteria (software implementation, legacy solutions, cost, etc.). Eliminating bias—or proving that this approach is systematic and fits the specific application—is a significant challenge.
3. Explainability of the AI algorithms. Explaining the performance of the algorithms is equally important, with the analyst or regulator being able to clearly recognize from the output whether the model predictions are more accurate than random chance, or if a human could achieve better results.

5.1 Regulatory Framework Outside the Nuclear Industry

Both the United States and European Union have identified a need to ensure better conditions for AI development and use. While the European Union has proposed a regulatory framework for this purpose [21,22], the United States has laid out policy considerations that should guide regulatory and non-regulatory approaches to AI developed and deployed outside the federal government. Of relevance to nuclear power regulations is the fact that the Federal Aviation Administration and National Highway Traffic Safety Administration are working to establish nimble and flexible frameworks that ensure safety while encouraging innovation. Thus, this approach implements a risk/benefit assessment to the regulation of AI-enabled products by assessing the reduction in risk that could result with AI, alongside those aspects of risk it might increase.

On a national scale, the National Science and Technology Council's Committee on Technology determined that long-term concerns over super-intelligent general AI should have little impact on current policy [2016] [23]. The National Science and Technology Council assessed policy requirements for self-driving vehicles, drones, etc., and the consensus of the commenters, as part of the White House Future of Artificial Intelligence Initiative, was that broad regulation of AI research or practices would be inadvisable at this time.

The consensus seems to be that new regulations may not be needed, and that the solution is perhaps to develop guidance documents that encompass AI.

5.1.1 Activities Related to AI

Standards Development Organizations and industry are moving forward with AI guidelines. Many of their efforts have focused on the ethics of AI, and many companies and governments are working to create definitions, policies, and regulations around these ethics.

To provide a means of compliance for certifying AI in safety-critical aeronautical systems, SAE International reviewed existing standards and performed a gap analysis to understand how and why existing standards cannot be reliably used [24]. Among the main gaps identified, requirements traceability, mapping of ML model functions and parameters between aerospace engineering concerns, and the application or lack of verification methods suitable for datasets raised many concerns. The identified gaps highlight the fact that a data-driven paradigm for AI may not be adequately addressed by existing standards. The SAE committee noted that:

> *"Extending this licensing procedure [of training and extensive testing for pilots and air traffic control] to autonomous software would lead to an analogous system of gained trust. Certification would be eventually attained through extensive, though not comprehensive, demonstration of knowledge and skill by the advanced software systems."*

OMB M-21-06 [Nov. 2020] [25] lays out policy considerations that should guide regulatory and nonregulatory approaches to AI applications developed and deployed outside the federal government. Although federal agencies currently use AI in many different ways to perform their missions, government use of AI is outside the scope of this memorandum. Though this memorandum uses the AI definition that was recently codified in statute, it focuses on "narrow" (i.e., weak) AI, which extends beyond advanced conventional computing in order to learn and perform domain-specific or specialized tasks by extracting information from datasets or other structured/unstructured sources of information. The memorandum states the following:

> *"When developing regulatory and non-regulatory approaches, agencies should pursue performance-based and flexible approaches that are technology neutral and that do not impose mandates on companies that would harm innovation. Rigid, design-based regulations that attempt to prescribe the technical specifications of AI applications will in most cases be impractical and ineffective."*

5.1.2 Examples of Use Cases

Automated vehicles (e.g., self-driving cars) and AI-equipped unmanned aircraft systems are currently relevant examples of the regulatory challenges presented by AI-enabled products. To evolve the relevant regulations, the U.S. Department of Transportation is using an approach based on building up expertise within the department, creating safe spaces and test beds for experimentation, and working with industry and civil society to evolve performance-based regulations that will enable more uses as evidence of safe operation accumulates.

For the interested reader, ISO/IEC TR 24030:2021 [26] provides a collection of use cases of AI applications in a variety of domains. In total, 132 AI use cases were submitted by experts between July 2018 and the end of November 2019.

5.2 Regulatory Framework in the Nuclear Industry

The current regulatory framework for nuclear energy in the United States focuses on the use of defense-in-depth measures to provide reasonable assurance of safety. Generally, such measures include periodic inspection and testing of safety-significant components, with preventive maintenance being performed on a time-based schedule to ensure component operability. AI can be used to maintain the same level of safety while optimizing preventive maintenance schedules.

The deployment of AI technology may require additional considerations and requirements (e.g., specific documentary evidence of performance and specific forms of technical data) prior to its acceptance for safety-significant components. Use of AI technologies on balance-of-plant components or components not considered safety significant will likely require little or no regulatory restrictions or needed approvals.

The United States is not alone in establishing a regulatory framework for the technical aspects of AI. An IEC working group on AI stated that "the regulatory framework from nuclear regulators is not yet established," and recognized that uses in safety-related applications will require further guidance from a regulatory perspective [27].

Any regulatory requirements and associated guidance will depend on how the AI is used. Although NPP operation was a primary driver behind the nuclear industry's leveraging of AI, the broader nuclear scientific and professional community rapidly adopted AI, as well. AI is now used in reactor design, fuel optimization, intelligent control, preventive maintenance, aging management, nondestructive testing, physical protection, cybersecurity, and many other related fields.

Regulatory treatment of AI will depend not only on whether it is used to support safety systems but also on its embedded functionality. The functional roles for AI can be grouped into the following main categories, which are generally consistent with NRC characterizations of the various automation levels:

(1) non-control functionality (advisory), (2) control functionality (shared), and (3) communications functionality (generally advisory).

- In the non-control functionality category, the AI's role would be to provide information or advice to the operator, but not to directly affect the plant or its operation. The main issues involve the quality, correctness, and fidelity of the information provided to the operator (i.e., the trustworthiness of the data) and the potential risk to plant safety as a result of deploying and relying on these techniques. Thus, a review might address whether proper performance, determination of uncertainty, and transparency of the basis for the information have been demonstrated. Consequently, the depth of the review would depend on the impact of erroneous or uncertain performance on safety and human reliability.
- Regarding the control functionality, the degree to which responsibility for actions is shared between the operator and the AI must be considered. This could range from the boundaries of manual control, with the AI advising onexpected responses to potential control actions, to autonomous control using an embedded AI for predictive control and/or automatic adaptation. The extreme end of autonomy is not anticipated as a near-term application for AI. The range of control functionality introduced to AI would lead to more rigorous regulatory review and a greater need for evidence of the system's safety impact. Plants with high degrees of passive safety are seemingly good candidates for implementing AI.
- Regarding the communications functionality, the level of regulatory review would depend on the safety significance of the data being transmitted. If vital communication of safety-related data is involved, the communications functionality afforded by the AI would necessarily be subject to safety or safety-related review. This would include independence, isolation, reliability, fault-tolerance/accommodation, etc. If the communications functionality is solely advisory or nonvital, the review would be similar to that for other non-control functionalities.

5.2.1 General Requirements for I&C

The regulatory framework determined by the NRC is intended to protect public health and safety. The NRC's mission is to ensure safe use of radioactive materials for beneficial civilian purposes, while still protecting people and the environment. The NRC regulates commercial NPPs and other uses of nuclear materials through licensing, inspection, and enforcement of its requirements. At a high level, no system malfunction or failure can prevent/block a safety action or initiate a challenge to that system.

At the highest conceptual level, I&C systems in NPPs can be categorized as either safety (protection) or non-safety (control) systems. If the AI is to provide a protection system function, it must meet the requirements of a safety system. The control system objectives are to maintain the controlled variables within the prescribed operating ranges, and the effects of operation or failure of these control systems are bounded by the accident analyses in Chapter 15 of the safety analysis report [28].

Relevant regulatory criteria for an I&C system using AI will vary depending on its use. For a cloud-based system, the AI is not expected to provide any control functions, but simply to provide data to users.

5.2.2 Regulatory Requirements for AI

Identifying the regulations that might apply to AI (based on its application) may enable a determination as to whether existing regulations are sufficient and can be adapted or considered to accommodate AIs for advanced reactors, or whether new regulations and guidance are needed.

A cloud-based AI data storage system would use AI to identify patterns and recommend actions, based on inputs from diagnostics, virtual sensors, intelligent control, aging management, preventive maintenance, anomaly detection, etc. This type of tool would not provide any control capabilities.

Largely, the AI tools can be broadly grouped into two categories: anomaly detection and ML. Anomaly detection monitors live data or computed results to identify instances of data that are

inconsistent with the previously defined statistical norm. Such tools can warn operators and designers of anomalies otherwise imperceptible to human observers. In practice, anomalies are often data spikes that reflect significant deviations from the expected values [29]. Actuation of an alarm when an anomaly is significant enough puts the operator (and plant) in a reactive state. Use of AI, noting the changes in state, would allow operators to be proactive.

In principle, ML is akin to human learning in that the software (human) is taught to detect a pattern. An algorithm is given data, then allowed to train itself to find patterns in those data. Many different algorithms are used for AI. Westinghouse recently developed a tool that evaluates over 10 regression-based AI algorithms to find trends in the data and then select the optimal algorithm based on data-driven modeling validation metrics [29]. These functional patterns can then augment anomaly detection, as well as prognosticate future behavior, based on the historical (or simulated) data. These types of predictive capabilities are very useful in determining the remaining useful life of a component or structure, the long-term behavior of a system (maintenance related), and pathways for system optimization.

As noted, the NRC issued its strategic plan for AI in NUREG-2261 [30]. This plan encompasses five goals: (1) ensure NRC readiness for regulatory decision making, (2) establish an organizational framework for reviewing AI applications, (3) strengthen and expand AI partnerships, (4) cultivate an AI-proficient workforce, and (5) pursue use cases to build an AI foundation across the NRC. The overall goal of this strategic plan is to ensure continued staff readiness to review and evaluate AI applications effectively and efficiently.

Several methods for reaping the benefits of PdM and condition monitoring are currently in use [31], including clustering algorithms for anomaly detection and Gaussian approaches for correcting instrument error based on adjacent or physically redundant data. Within this same scope, Metroscope employs a Bayesian method to find root causes by leveraging physics models, expert knowledge, and operating experience. In addition, autocorrelation methods are also being explored for nondestructive examination applications.

6. SUMMARY AND CONCLUSION

During the cost comparison, certain assumptions were made to make the analysis as conservative as possible. Most notably, the DAS implementation assumed for the in-building network is considerably more expensive that the 4G LTE MPN. In the nuclear industry, MPN is becoming the preferred network type, thanks to its cost and flexibility advantages. This is evidenced by the implementation of the LEMKO 4G LTE network at the Dresden plant.

The cloud offers other advantages besides cost. First, due to the virtual nature of the hosting hardware, capacity can be “flexed up” as needed. As shown in this document, Microsoft Azure offers competitive tools that can scale with load. Additionally, back-up is transparent. Moreover, the cybersecurity resources of Microsoft or Amazon are brought to bear. Individual plants in the base case are responsible for their own cybersecurity, whereas plants using cloud services can lean on the cloud provider for security measures.

AI is being implemented throughout industry and its usage in NPPs will increase. The current regulatory framework does not explicitly address AI or autonomous control, but if AI is applied to nuclear operations such as communicating with a system at a plant, it must meet the requirements of a digital I&C system. Currently, all use of AI at NPPs is focused on non-safety related applications. Cloud-based data storage would be concerned with using the data to find patterns and to recommend actions based on inputs from diagnostics, virtual sensors, intelligent control, aging management, preventive maintenance, anomaly detection, etc. Such use is similar to existing narrow AI applications such as language translation, self-driving vehicles, image recognition, virtual assistants, self-driving cars, AI-powered web searches, recommendation engines, and intelligent spam filters. For this type of use case, no new regulations or guidance would be necessary. The NRC and other regulatory bodies are working to provide to address gaps, rather than introducing new regulations to address the use of AI and ML. This approach seems the best way to encourage development without adding regulatory uncertainty. However, the use of

AI technologies on balance-of-plant components or components not considered safety significant will likely require little or no regulatory restrictions or needed approvals. Demonstrating how AI can improve the maintenance and operation of these non-safety-related systems seems the likely path forward for implementing AI and cloud computing resources inside NPPs.

7. REFERENCES

1. Ngarayana, I. W., T.-M. Dung Do, K. Murakami, and M. Suzuki. 2019. "Nuclear Power Plant Maintenance Optimisation: Models, Methods & Strategies." Journal of Physics Conference Series 1198(2):022005. https://doi.org/10.1088/1742-6596/1198/2/022005.
2. IAEA. 2015. "Plant Life Management Models for Long Term Operation of Nuclear Power Plants." IAEA Nuclear Energy Series, no. NP-T-3.18, IAEA, Vienna (2015).
3. Agarwal, V., et al. 2021. "Scalable Technologies Achieving Risk-Informed Condition-Based Predictive Maintenance Enhancing the Economic Performance of Operating Nuclear Power Plants." INL/EXT-21-64168, Rev. 0, Idaho National Laboratory. https://doi.org/10.2172/1894498.
4. IAEA. 2022. "Artificial Intelligence for Accelerating Nuclear Applications, Science and Technology." Non serial Publications, IAEA, Vienna (2022). https://www.iaea.org/publications/15198/artificial-intelligence-for-accelerating-nuclear-applications-science-and-technology.
5. Zhang, S. et al. 2022. "Practical Adoption of Cloud Computing in Power Systems—Drivers, Challenges, Guidance, and Real-World Use Cases." IEEE Transactions on Smart Grid 13(3):2390–2411. https://doi.org/10.1109/tsg.2022.3148978.
6. Walker, C., R. Appiah, and V. Agarwal. 2023. "Development of a Scalable, Risk-informed, Predictive Maintenance Cloud based Strategy at Nuclear Power Plants." 13th Nuclear Plant Instrumentation, Control and Human-Machine Interface Technologies Conference, Knoxville, TN. July 17-21, 2023.
7. Appiah, R., C. Walker, V. Agarwal, J. Nistor, T. Gruenwald, M. Muhlheim, and P. Ramuhalli. 2022. "Development of a Cloud-based Application to Enable a Scalable Risk-informed Predictive Maintenance Strategy at Nuclear Power Plants." INL/RPT-22-70543, Rev. 0, Idaho National Laboratory, Idaho Falls, US.
8. European Union Aviation Safety Agency (EASA). 2021."EASA Concept Paper: First usable guidance for Level 1 machine learning applications."
9. 2020 State of Data Science. https://www.anaconda.com/resources/whitepapers/state-of-data-science-2020/.
10. ISO/IEC 22989:2022. "Information technology — Artificial intelligence — Artificial intelligence concepts and terminology."
11. BlueWave AI Labs. Characterizing Chaotic Systems (Chattering Valves in Nuclear Reactors). [Online] August 2023. https://www.bluewaveailabs.com/characterizing-chaotic-systems-chattering-valves-in-nuclear-reactors/.
12. Chandra, R., S. Goyal, and R. Gupta. 2021. "Evaluation of Deep Learning Models for Multi-Step Ahead Time Series Prediction." IEEE Access, 9:83105–83123. https://doi.org/10.1109/ACCESS.2021.3085085.
13. Microsoft Azure. 2023. "What is Azure Databricks? - Azure Databricks." Accessed August 2023. https://learn.microsoft.com/en-us/azure/databricks/introduction/.
14. National Control Devices. n.d. "NCD Store." NCD. Accessed December 7, 2022. https://www.ncd.io
15. PCB PIEZOTRONICS. n.d. "PCB PIEZOTRONICS." PCB PIEZOTRONICS. Accessed December 7, 2022. https://www.pcb.com

16. Muhlheim, M. D., P. Ramuhalli, A. Huning, A. Guler, and A. Saxena. 2023. "Status Report on Regulatory Criteria Applicable to the Use of Artificial Intelligence (AI) and Machine Learning (ML)." ORNL/SPR-2023/3072.
17. NRC. 2020. "Virtual Workshop on Digital Twin Applications for Advanced Nuclear Technologies." Virtual Workshop, December 1–4, 2020, Adams Accession No. ML20314A108, (December 2020).
18. NRC. 2021. "Enabling Technologies for Digital Twin Applications for Advanced Reactors and Plant Modernization." Virtual Workshop, September 14–16, 2021, Adams Accession No. ML21228A082 (September 2021).
19. Yadav V., et al. 2021. "The State of Technology of Application of Digital Twins." Letter Report: TLR/RES-DE-REB-2021-01.
20. Sun, H., P. Ramuhalli and R. Jacob. "Machine Learning for NDE Literature Review." Submitted to Ultrasonics.
21. News European Parliament. 2023. "Artificial intelligence: threats and opportunities." News European Parliament Updated: 20-06-2023 - 09:08, Created: 23-09-2020 - 14:36. https://www.europarl.europa.eu/news/en/headlines/society/20200918STO87404/artificial-intelligence-threats-and-opportunities.
22. News European Parliament. 2023. "EU AI Act: first regulation on artificial intelligence." 4-06-2023 - 14:06 Created: 08-06-2023 - 11:40. https://www.europarl.europa.eu/news/en/headlines/society/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence.
23. Executive Office of the President National Science and Technology Council Committee on Technology. 2016. "Preparing for The Future of Artificial Intelligence." National Science and Technology Council.
24. AIR6988. 2021. "Artificial Intelligence in Aeronautical Systems: Statement of Concerns." SAE International.
25. Executive Office of The President, Office of Management and Budget. 2020. "Guidance for Regulation of Artificial Intelligence Applications." Washington, D.C. 20503, November 17, 2020.
26. ISO/IEC TR 24030:2021. "Information technology — Artificial intelligence (AI) — Use cases."
27. IEC TR-63468. 2023. "Nuclear facilities – Instrumentation and control, and electrical power systems – Artificial Intelligence applications."
28. NUREG-0800. "U.S. Nuclear Regulatory Commission Standard Review Plan. 7.7 Controls Systems." NRC ADAMS Accession No. ML070670042.
29. Banyay, G. A., Westinghouse. 2021. "Role of Artificial Intelligence Tools in U.S. Commercial Nuclear Power Operations, May 31, 2021. [NRC ADAMS Accession No. ML21202A180]." Letter to US NRC.
30. NUREG-2261. "Artificial Intelligence Strategic Plan, Fiscal Years 2023-2027, May 2023." NRC ADAMS Accession No. ML23132A305.
31. Peters, G., Framatome Inc. 2021. "Role of Artificial Intelligence Tools in U.S. Commercial Nuclear Power Operations, May 21, 2021. [NRC-2021-0048], [NRC ADAMS Accession No. ML21153A056]. Letter to US NRC.