

Light Water Reactor Sustainability Program

Scalable Methods to Automate Manual Work Management Activities Using Artificial Intelligence

Ahmad Al Rashdan, Kellen Giraud, John Mapes, Brian Wilcken, and Jeremy Mohon
Idaho National Laboratory
Patrick Murray
EQRPI Inc.



August 2024

U.S. Department of Energy
Office of Nuclear Energy

DISCLAIMER

This information was prepared as an account of work sponsored by an agency of the U.S. Government. Neither the U.S. Government nor any agency thereof, nor any of their employees, makes any warranty, expressed or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness, of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. References herein to any specific commercial product, process, or service by trade name, trade mark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the U.S. Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the U.S. Government or any agency thereof.

Light Water Reactor Sustainability Program

**Scalable Methods to Automate Manual Work Management Activities
Using Artificial Intelligence**

**Ahmad Al Rashdan, Kellen Giraud, John Mapes, Brian Wilcken, and Jeremy
Mohon
Idaho National Laboratory
Patrick Murray
EQRPI Inc.**

August 2024

**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

EXECUTIVE SUMMARY

In nuclear utilities, work is generally initiated by three sources: unexpected events that need to be addressed to maintain plant operations, updates or upgrades that improve plant performance or prevent potential issues, and periodic preventative maintenance activities, such as inspections or servicing of structures, systems, and components to monitor degradation and minimize unexpected outage risk. Thousands of work requests are generated annually in a nuclear power plant. Once a work need is identified, work preparation is initiated and managed through a highly proceduralized process (known as work management) involving decisions and actions by various plant organizations.

The work management process is both time-consuming and costly. The process includes several phases: screening unexpected events, scoping to define the nature of the work, planning and assembling a work package, scheduling work and resources, executing the work, and documenting findings. Various plant organizations—including but not limited to maintenance, engineering, and operations—may be involved depending on the specific impact of the work on plant operations. Each organization utilizes its own data sources and brings different expertise and perspectives to the process.

Due to the length and complexity of the work management process and the involvement of multiple organizations, routine tasks can incur significant expenses, often reaching tens of thousands of U.S. dollars per work package due to extensive staff involvement across various phases. With thousands of work packages executed annually per plant, these costs can total tens of millions of U.S. dollars per year, making work management a major contributor to the high operational expenses of nuclear power plants.

To address the high costs of work management, this effort focused on determining ways to automate or semi-automate the process using artificial intelligence (AI) in its various forms, such as natural-language processing or machine learning. AI is a critical tool for automation due to its ability to handle complex and varied decisions by understanding context similarly to humans.

To identify areas where automation may be effective in improving work management processes, the entire work management process was mapped to specific activities. To facilitate automation, the work management process was broken down into specific, clearly defined steps performed by specific staff that can be represented by decisions or actions. Each phase—screening, scoping, planning, scheduling, execution, and post-execution—was divided into numerous decisions and actions. Human time estimates were assigned to each decision and action. The goal was to identify those steps that would most likely benefit from integrating AI into the process. The analysis showed that most of the time-consuming steps were concentrated in the planning phase, which alone accounts for over 42% of the total cost.

To demonstrate the potential for AI in automating the work process, several methods and tools (e.g., topic modeling, regression, similarity analysis) from Machine Intelligence for Review and Analysis of Condition Logs and Entries (MIRACLE) were customized or expanded. These methods and tools were grouped into a novel list of scalable AI functions (e.g., classification, estimation, guidance) that were then assigned to the decisions and actions of the work

management process. Example demonstrations of these AI functions and the technology readiness of each function were evaluated using both existing and newly developed methods and tools. The analysis revealed several challenges in implementing AI technologies, such as issues with sparse labels, inconsistent data, and prompt sensitivity in classification. Addressing these challenges is crucial for improving the reliability and efficiency of AI functions, ultimately enhancing their effectiveness in real-world scenarios. Several of these challenges are addressed in this research and proposed solutions are provided. Other challenges are identified and are active areas of research within the AI community. Regardless of the specific methods or tools used—whether developed or yet to be developed—AI functions were applied to automate each decision and action.

Using the AI functions and their mapping to work management decisions and actions, this research concludes that significant cost savings can be achieved through targeted improvements. This analysis shows a potential cost savings of \$14.5M per nuclear power plant unit per year. These cost savings are derived by using AI to automate 131 steps in six phases of the work management process:

Phase	Current Cost Per Unit Per Year	AI Cost Saving Per Unit Per Year	Ratio of Saving
Screening	\$1.53M	\$0.88M	57.6%
Scoping	\$2.50M	\$2.13M	85.2%
Planning	\$7.79M	\$7.72M	99.1%
Scheduling	\$1.96M	\$1.32M	67.2%
Execution	\$4.45M	\$1.78M	40.1%
Post-execution	\$0.16M	\$0.14M	84.4%

Despite being the most expensive phase of the work management process, planning was found to be almost fully automatable. Execution, the second-most expensive phase, shows much lower potential for automation with AI.

By quantifying the cost savings for each decision or action and linking it to the associated AI functions, this research identifies the most promising functions for cost reduction. Review, generation, structuring, comparison, summarization, and supervised classification each resulted in estimated cost savings over \$2.7M per unit per year, with the review function offering the highest savings at around \$4.0M per unit per year. Meanwhile unsupervised classification and estimation achieved cost savings between \$1.1M and \$1.2M per unit per year. Guidance and insertion achieved cost savings of \$0.85M per unit per year. All these functions were demonstrated using work management examples.

Although review, generation, and structuring show the greatest potential for cost savings, the AI methods and tools associated with them are not yet fully developed and validated or ready for deployment due to several discussed challenges. While solutions were proposed and/or demonstrated, further in-depth research is needed to advance these functions for industry use. In contrast, the methods and tools for comparison and classification, despite offering slightly lower cost savings, are more developed and ready for implementation. In this

effort, several innovative methods were developed for these functions to effectively address their deployment challenges and facilitate their broader use.

Page intentionally left blank

ACKNOWLEDGMENTS

The authors wish to thank the Light Water Reactor Sustainability Program for funding this work and acknowledge its leadership support in making this effort a success. They also wish to thank the industry partners who provided valuable data and insights throughout this effort. Finally, the authors extend their gratitude to ScottMadden for their support in validating the work management time estimates.

This research made use of the resources of the High-Performance Computing Center at Idaho National Laboratory, which is supported by the Office of Nuclear Energy of the U.S. Department of Energy and the Nuclear Science User Facilities under contract no. DE-AC07-05ID14517.

Page intentionally left blank

CONTENTS

EXECUTIVE SUMMARY	iii
ACKNOWLEDGMENTS	vii
ACRONYMS	xvi
1. INTRODUCTION	1
1.1 Overview of the Work Management Process.....	2
1.2 Approach.....	4
1.3 Report Structure	5
2. AI METHODS AND TOOLS USED FOR TEXT ANALYSIS	6
2.1 Text Preparation.....	6
2.1.1 Rules-Based Processing	6
2.1.2 Tokenization	7
2.1.3 Vectorization.....	8
2.1.4 Embedding	9
2.2 Supervised Learning	10
2.2.1 Regression.....	10
2.3 Unsupervised Learning	12
2.3.1 Clustering.....	12
2.3.2 Dimensionality Reduction.....	12
2.3.3 Generative AI.....	12
2.3.4 Similarity Analysis.....	15
2.3.5 Topic Modeling.....	15
2.4 Semisupervised Learning	16
3. APPLICABLE AI FUNCTIONS IN TEXT ANALYSIS AND EXAMPLE DEMONSTRATIONS	16
3.1 Classification (Supervised)	16
3.1.1 Tools Overview.....	17
3.1.2 Example Application.....	17
3.1.3 Technology Readiness	19
3.2 Classification (Unsupervised)	19
3.2.1 Tools Overview.....	19
3.2.2 Example Application.....	19
3.2.3 Technology Readiness	20
3.3 Comparison	20
3.3.1 Tools Overview.....	21
3.3.2 Example Application.....	21
3.3.3 Technology Readiness	21
3.4 Dialogue (Chatbots)	22
3.4.1 Tools Overview.....	22
3.4.2 Example Application.....	22
3.4.3 Technology Readiness	22
3.5 Estimation	23

3.5.1	Tools Overview.....	23
3.5.2	Example Application.....	23
3.5.3	Technology Readiness	24
3.6	Generation.....	24
3.6.1	Tools Overview.....	25
3.6.2	Example Application.....	25
3.6.3	Technology Readiness	26
3.7	Guidance	26
3.7.1	Tools Overview.....	26
3.7.2	Example Application.....	26
3.7.3	Technology Readiness	27
3.8	Insertion	28
3.8.1	Tools Overview.....	28
3.8.2	Example Application.....	28
3.8.3	Technology Readiness	30
3.9	Review	30
3.9.1	Tools Overview.....	30
3.9.2	Example Application.....	30
3.9.3	Technology Readiness	31
3.10	Structuring.....	32
3.10.1	Tools Overview.....	32
3.10.2	Example Application.....	32
3.10.3	Technology Readiness	33
3.11	Summarization	33
3.11.1	Tools Overview.....	33
3.11.2	Example Application.....	34
3.11.3	Technology Readiness	34
3.12	Visualization	35
3.12.1	Tools Overview.....	35
3.12.2	Example Application.....	35
3.12.3	Technology Readiness	35
4.	MAPPING THE WORK PROCESS TO AI FUNCTIONS	37
4.1	Work Mapping Approach	37
4.2	Assumptions.....	38
4.3	Initiation.....	40
4.4	Screening.....	41
4.5	Scoping.....	41
4.6	Planning	44
4.7	Scheduling.....	48
4.8	Execution	50
4.9	Post-Execution	50
5.	POTENTIAL COST SAVING IN WORK MANAGEMENT.....	53
5.1	Screening.....	53
5.2	Scoping.....	55

5.3	Planning	55
5.4	Scheduling.....	55
5.5	Execution	58
5.6	Post-Execution	58
5.7	Summary of Cost Savings in Work Management.....	59
6.	POTENTIAL COST SAVING OF FUNCTIONS	60
6.1	Screening.....	60
6.2	Scoping.....	61
6.3	Planning	62
6.4	Scheduling.....	63
6.5	Execution	63
6.6	Post-Execution	64
6.7	Summary of Cost Savings of Functions.....	64
7.	FUNCTION CHALLENGES AND POTENTIAL METHODS	65
7.1	Supervised Classification	66
7.1.1	Sparse Labels	66
7.1.2	Consistency of Labels Overtime or Across Utilities.....	69
7.1.3	Missing or Highly Imbalanced Labels Data.....	70
7.2	Unsupervised Classification.....	70
7.2.1	Iterations Time for New Labels	71
7.2.2	Prompt Sensitivity.....	72
7.3	Supervised and Unsupervised Classification	74
7.3.1	Short Text.....	74
7.3.2	Input Quality	76
7.4	Comparison	81
7.4.1	Different Sizes of Text.....	81
7.4.2	Suitability of Embedding Model.....	82
7.4.3	Finding Similarities and Differences	82
7.5	Dialogue.....	82
7.5.1	Prompt Sensitivity.....	83
7.5.2	Hallucination.....	83
7.5.3	Limited Insight.....	83
7.6	Estimation	83
7.6.1	Missing Context or Mixed Context.....	84
7.7	Generation.....	87
7.7.1	Irrelevant Text.....	88
7.8	Guidance	88
7.8.1	Confusion.....	88
7.9	Insertion	88
7.9.1	Generic Words	89
7.10	Review	89
7.10.1	Complex Structure and Confounding Context.....	89
7.11	Structuring.....	89
7.11.1	Complex Structure and Dispersed Low-Quality Data	89

7.12	Summarization	89
7.12.1	Misleading the Model	90
7.13	Visualization	90
7.13.1	Data Volume and High Dimensionality	90
7.14	Broader Challenges	90
7.14.1	Obsolescence.....	90
7.14.2	Scalability	90
7.14.3	Availability and Cost of Computational Power	90
7.14.4	Memory Constraints.....	91
8.	CONCLUSIONS	91
9.	REFERENCES	93
Appendix A	Mapping of Screening	100
Appendix B	Mapping of Scoping	132
Appendix C	Mapping of Planning	170
Appendix D	Mapping of Scheduling	198
Appendix E	Mapping of Execution	218
Appendix F	Mapping of Post-Execution	245

FIGURES

Figure 1.	MIRACLE initial mission to assist the CR screening process.	2
Figure 2.	An overview of the work management process.....	3
Figure 3.	MIRACLE expanded mission to automate the work management process.	5
Figure 4.	Visual abstract of the report denoting the scope of each section.....	6
Figure 5.	The RAG process.	15
Figure 6.	Actual vs. predicted completion time in days.....	24
Figure 7.	Insertion accuracy for one, two, and three words.....	30
Figure 8.	UMAP dimensionality reduction for topics generated by a CorEx model.	36
Figure 9:	Histograms of cosine similarity values for records tagged as CAQ, NCAQ, or SCAQ.....	69
Figure 10.	Histogram of the writing quality of CRs evaluated using the junk probability detector.	78
Figure 11.	Histogram of the writing quality of CRs evaluated using BART.....	81
Figure 12.	Actual vs. predicted completion time in days after including additional fields.....	84
Figure 13.	Histogram of CR lengths and regression performance using a sliding window.....	85

Figure 14. Model performance as a function of token count for 511 CRs with fewer than 100 tokens each	87
Figure 15. Actual vs. predicted completion time in days after CR size elaboration.....	87
Figure 16. Flowchart of the screening steps that are discussed in Appendix A.....	100
Figure 17. Flowchart of the scoping steps that are discussed in Appendix B.....	132
Figure 18. Flowchart of the planning steps that are discussed in Appendix C.	170
Figure 19. Flowchart of the scheduling steps that are discussed in Appendix D.....	198
Figure 20. Flowchart of the work execution steps that are discussed in Appendix E.....	218
Figure 21. Flowchart of the work post-execution steps that are discussed in Appendix F.....	245

TABLES

Table 1. The different types of generative AI methods considered in this research.	13
Table 2. Summary of performance results for supervised classification of CAQ.....	18
Table 3. Summary of performance results for unsupervised classification of three events.	20
Table 4. Pairwise cosine similarity of three conditions using three different models.....	21
Table 5. AI guidance of a user who is typing a CR.	27
Table 6. AI insertion for a user while typing a CR.	29
Table 7. Effect of the inclusion of reference material on F1 score.	31
Table 8. Keywords and phrases used by the CorEx to describe fire hazards.....	33
Table 9. Confusion matrix for the LLM and CorEx decisions on whether text relates to fire.....	33
Table 10. Example of the current decisions and conditional actions table.	38
Table 11. Example of the automated decisions and conditional actions table.....	38
Table 12. Estimated potential time saved in minutes per CR for each screening step by each AI function.....	42
Table 13. Estimated potential time saved in minutes per WO for each scoping step by each AI function.....	45
Table 14. Estimated potential time saved in minutes per WP for each planning step by each AI function.....	47
Table 15. Estimated potential time saved in minutes per ST for each scheduling step by each AI function.....	49
Table 16. Estimated potential time saved in minutes per WP for each work execution step by each AI function.....	51
Table 17. Estimated potential time saved in minutes per WP for each post-execution step by each AI function.....	52
Table 18. Reference metrics and scaling factors for each work management phase.	53
Table 19. Current costs and potential cost savings by using AI for every step of screening.....	54

Table 20. Current costs and potential cost savings by using AI for every step of scoping.....	56
Table 21. Current costs and potential cost savings by using AI for every step of planning.	57
Table 22. Current costs and potential cost savings by using AI for every step of scheduling.....	58
Table 23. Current costs and potential cost savings by using AI for every step of execution.....	59
Table 24. Current costs and potential cost savings by using AI for every step of post-execution.....	59
Table 25. Summary of current costs and potential cost savings per phase.	60
Table 26. Cost savings in screening by each AI function.	61
Table 27. Cost savings in scoping by each AI function.	62
Table 28. Cost savings in planning by each AI function.	62
Table 29. Cost savings in scheduling by each AI function.	63
Table 30. Cost savings in execution by each AI function.....	64
Table 31. Cost savings in post-execution by each AI function.....	64
Table 32. Cost savings for each AI function across all phases.	65
Table 33. Summary of performance results for supervised classification of SCAQ.	68
Table 34. Correlation of data consistency metric and model performance.....	70
Table 35. Top 15 n-grams relating to the bigram “valve_stem”.....	72
Table 36. Summary of performance results for the autoregressive methods with and without prompt engineering.....	73
Table 37. Summary of performance results for unsupervised classification of three events after word- and context-enhancement.....	76
Table 38. Probability of bigrams occurrence using TF-IDF.....	77
Table 39. Autoencoder reconstruction of example text and the cosine similarity of the original and reconstructed text.....	80
Table 40. Impact of text elaboration on model performance and computational time.	86
Table 41. Impact of text elaboration on model performance for 511 CRs with fewer than 100 tokens each.	86

Page intentionally left blank

ACRONYMS

AI	artificial intelligence
ALARA	as low as reasonably achievable
BoW	bag of words
CAP	corrective action program
CAPCo	corrective action program coordinator
CAQ	condition adverse to quality
CO	clearance order
CorEx	correlation explanation
CR	condition report
EC	engineering change
EQ	environmental qualification
ESF	Engineered safety feature
FME	foreign material exclusion
FTE	full-time equivalent
GAN	generative adversarial network
HP	human performance
HR	human resources
INL	Idaho National Laboratory
IT	information technology
k-NN	k-nearest neighbors
k>NNR	k-nearest neighbors regression
LDA	latent Dirichlet allocation
LLM	large language model
LWRS	Light Water Reactor Sustainability (Program)
MEL	master equipment list
MIRACLE	Machine Intelligence for Review and Analysis of Condition Logs and Entries
NA	not applicable
NCAQ	not a condition adverse to quality
NEI	Nuclear Energy Institute
NLP	natural-language processing
NN	neural network
NPP	nuclear power plant
OE	operating experience

PI	performance improvement
PM	preventive maintenance
PMT	post-maintenance testing
POS	part-of-speech
PRA	probabilistic risk assessment
RAG	retrieval-augmented generation
RCIC	reactor-core isolation cooling
RHR	residual-heat removal
RNN	recurrent neural network
RP	radiation protection
RPA	radiological protected area
RWP	radiological work permit
SCAQ	significant condition adverse to quality
SME	subject-matter expert
SRO	senior reactor operator
SSC	structure, system, and component
ST	schedule task
TF-IDF	term frequency-inverse document frequency
UMAP	Uniform Manifold Approximation and Projection
WO	work order
WP	work package
WR	work request

Scalable Methods to Automate Manual Work Management Activities Using Artificial Intelligence

1. INTRODUCTION

The fundamental task of staff at a nuclear power plant (NPP) is to maintain and operate the power plant to safely generate electricity in the most cost-effective way possible. Though nuclear energy provides several advantages over other energy sources, NPPs have been economically challenged to compete with other sources of energy in terms of cost per megawatt-hour of power. The current fleet of NPPs in the United States (U.S.) is labor dependent and still frequently relies on manual processes. While some automation efforts have been successful across the industry, in general migration toward automation has been slow due to the industry's unique regulatory nature. As a result, operations and maintenance costs have remained high, while other industries have leveraged technology development and reduced their operations and maintenance costs.

In nuclear utilities, work execution is managed by a process that is highly proceduralized by a set of decisions and actions made by people from several organizations in the plant. Those decisions and actions are enabled by various forms of information in the form of documents, such as procedures, manuals, data sheets, etc. This process is often referred to as the *work management process*. The work management process is very time-consuming and therefore expensive. Routine tasks can incur substantial costs that often reach beyond the actual cost of the job execution due to extensive staff involvement in several phases of the work evolution. A utility has conducted a recent study and identified that a simple task in NPPs can cost upward of tens of thousands of U.S. dollars (USD) due to the overhead associated with the work execution. With thousands of work orders (WOs) per plant executed annually, these costs can accumulate to tens of millions of USD per year for each plant, highlighting how work management is a major factor in the high operational expenses of NPPs.

The Light Water Reactor Sustainability (LWRS) Program has recognized the potential for automation to reduce the cost of the work management process and proposed several technologies in previous efforts. In 2016, Al Rashdan et al. (2016) identified 50 key advanced functions to implement in the work process and conducted an industry survey to verify the research findings on the importance of those functions. Although these functions were considered ambitious at the time due to the lack of existing technology, the rapid advancements in artificial intelligence (AI) in recent years have enabled the realization of many of these functions. In 2018 the LWRS Program began developing AI methods to enable automated work processes. The first challenge to achieve this objective was data integration, which was researched in 2019 (Al Rashdan et al. 2019). Research into the use of AI to analyze work management data commenced in 2019 (Al Rashdan et al. 2020), and research focused on condition reports (CRs) followed that effort (Al Rashdan et al. 2021).

Given recent advancements in the use of AI for text analysis, the nuclear industry leveraged several existing methods and tools to create application-specific solutions. Those methods and tools were generic in nature—that is, they were designed to be used in various functions. Therefore, users frequently found that they resulted in suboptimal performance, eroding their trust and resulting in the industry abandoning those methods and tools after the pilot stage. This motivated the LWRS Program to research novel methods and tools and create new approaches to achieve optimal performance, with the aim of outperforming a human counterpart. This resulted in a novel technology called Machine Intelligence for Review and Analysis of Condition Logs and Entries (MIRACLE). MIRACLE (Figure 1) included several novel technologies—AI methods to analyze a plant's CRs (as highlighted in NEI 2018)—and consequently won an R&D 100 Award in 2022. Additionally, MIRACLE aimed to explore new applications the industry had not explored. The nuclear industry recognized the potential for this research, and data from the corrective action programs (CAPs) of 39 reactors operating in the U.S. were shared with Idaho National Laboratory (INL). MIRACLE was made available for licensing and is currently

being used by a utility as part of the utility's condition screening tool. Since 2021, researchers have had exponential success finding new applications for MIRACLE, including, using its tools for inspection, risk analysis, outage optimization, performance improvement (PI), among other tasks. A key application for MIRACLE is work management, which is discussed in this report.

Considering the cost associated with the work management process, it is desired to determine a means to migrate it to an automated or at least semi-automated model in which the decisions and actions are made by a machine using AI. AI is a critical tool for automation mainly because the decisions for work management vary in complexity and nature, a challenge that AI can tackle, given its ability to understand context like a human. The goal of this work is to identify and develop effective AI solutions to reduce workforce costs and streamline the work management process.

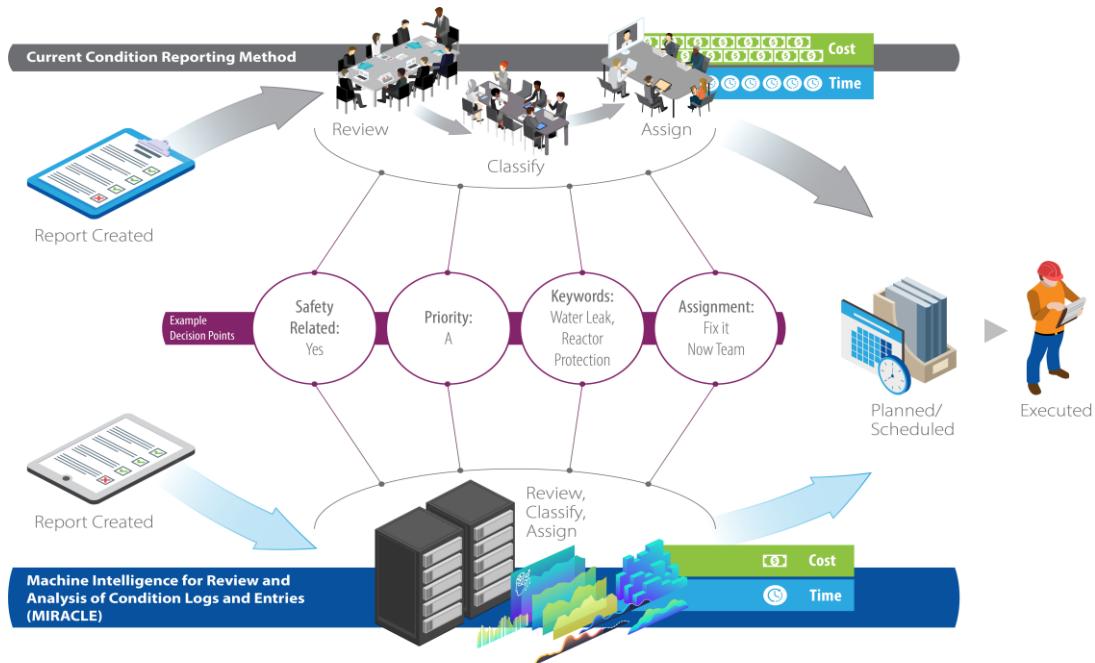


Figure 1. MIRACLE initial mission to assist the CR screening process.

1.1 Overview of the Work Management Process

There are generally three sources that initiate work in a nuclear utility:

1. Corrective maintenance: an unexpected event that needs to be rectified to sustain plant operations and described in a CR
2. Project: an update, upgrade, or change that needs to be implemented to improve plant performance or prevent a foreseen issue
3. Preventive maintenance (PM): activities that could be in the form of inspecting or servicing a structure, system, and component (SSC) to ensure its degradation is tracked and outage risk is minimized.

Once the work need is identified, a work request (WR) is created and work commences, passing through several phases, each leading to different outcomes (see Figure 2). The main work phases are screening of unexpected events in the plant, scoping to characterize the nature of the needed work, planning the work and assembling the work package (WP), scheduling the work and creating schedule

tasks (STs) for the specific work activities at specific times and assigning the needed resources, executing the actual work, then documenting the findings and insights post-execution.

The execution of the work requires that all relevant organizations review their daily schedule and execute the activities assigned to that day. During the work, each organization completes its assigned part of the WP and provides progress or completion indications for each step of its dedicated procedure.

Once a WO is initiated, if it relates to equipment it goes through the scoping process. Scoping is the process of characterizing attributes of the WO such as priority, criticality, and whether the planned work will have any operational impact. This process also determines the type of resources needed (in-house or contract), risk and risk-mitigation strategies, and any other work scope implications. During scoping, the WO is broken down into one or more WO activities (also referred to as sub-WOs or tasks in the industry). For example, associated work scopes that may require the creation of separate WO activities include erecting (and dismantling after the work) scaffolding to access the equipment, removing (and reinstalling) insulation installed on the equipment, and establishing foreign material exclusion (FME) controls as required. Each WO activity is a substep or task of the main scope. After the WO is completed, a WP is created as part of the planning process. The WP converts the WO requirements into a set of procedures, instructions, manuals, and other related data sheets or documents. This is performed by the planner. The WP could be updated from an existing WP if one exists for similar work, or a new WP could be created if needed. After the WP is completed, it needs to be assigned to staff to execute the work. The scheduling team determines when the work will be executed within the week. Schedulers consider all the resources that may be needed in terms of staff, parts (and lead time to get them), and risk constraints (provided by the planner), among other factors discussed in this report. Once the work is scheduled, it is “frozen,” and the staff are assigned to the WP task for the dedicated time. After the work is completed, actions are taken to ensure that the quality of work is satisfactory, and any lessons learned are logged.

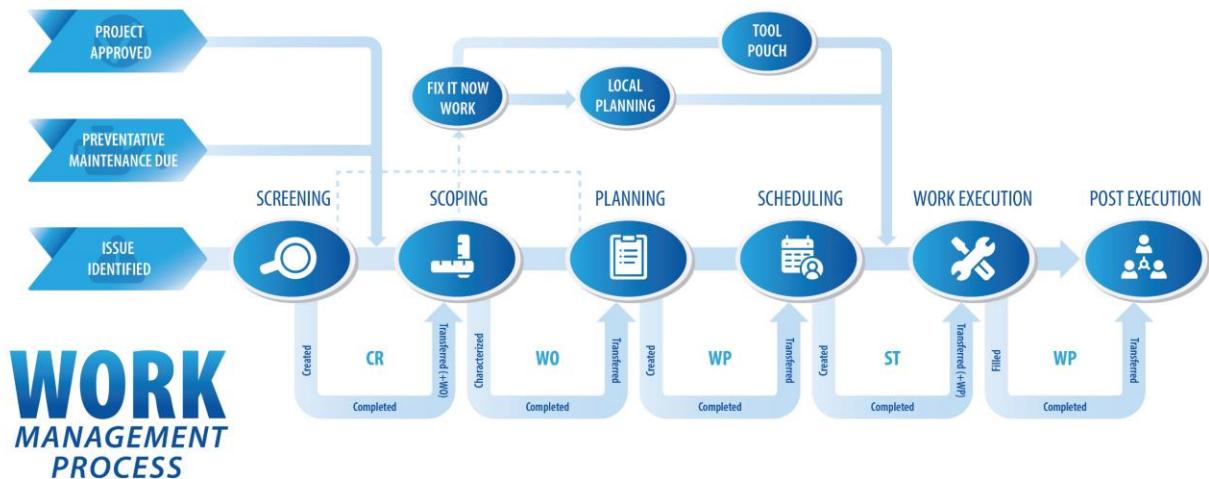


Figure 2. An overview of the work management process.

Throughout this process, various organizations in the plant beyond maintenance, engineering, or operations could be involved. For example, chemistry could be involved if the work impacts plant chemistry, radiation protection (RP) could be involved to ensure low exposure levels to humans are sustained and contamination is prevented, or security could be involved if a barrier needs to be broken. Each organization uses its own data sources, has different knowledge and skill sets, and considers the plant from a different perspective. The details of the work management process are discussed in Section 4 and Appendix A-F.

1.2 Approach

To introduce a systematic approach to automation, the MIRACLE methods and tools were expanded in this effort and grouped into the functions (for an application) they perform (Figure 3). The functions were designed to be scalable, enabling them to automate the diverse forms of human decision-making and action-taking activities. The key functions included supervised and unsupervised classification for categorizing and labeling operational data, text comparison for identifying similar text or context, and dialogue systems for real-time human-AI interactions. The functions also included estimation for predicting continuous outcomes based on historical data patterns, as well as generation for creating new content through AI-driven innovation. Additional functions discussed are guidance for text recommendations, insertion for facilitating new information, review for assessing unstructured documents, structuring for data organization and extraction, summarization for concise overviews of lengthy texts, and visualization for making complex data more comprehensible.

Most of the MIRACLE methods and tools would benefit more than one function. For example, nonlinear regression (using a neural network, or NN) can be used for classification of text, for estimation of an outcome, or for automated text insertion. Using functions instead of a specific method or tool in automating the work management process provides flexibility for users in deciding what method or tool to use, provided they serve the targeted function. The function is not fixed to a certain method or tool; instead, it is focused on how methods and tools can be used rather than on which is used.

Each work management phase (screening, scoping, planning, scheduling, execution, and post-execution) consists of a series of steps that rely heavily on staff from various organizations to review documents, including procedures, manuals, and data sheets, and to make decisions. The approach of this report is to identify the key decisions and actions involved in each of these steps, resulting in over 130 identified decisions and actions (Appendix A-F). Each decision and action was assigned a frequency and an estimate of the human time required to execute it. To automate specific steps of the work management process, the AI functions were assigned to each identified decision and action. This approach allowed for estimating the potential cost savings that could be achieved by introducing AI into the process.

Given that the cost savings for each step is quantified, and the associated AI function is linked, those functions can be aggregated to determine which ones have the most promising cost savings. This was used to focus this research and can be used to focus future research on the functions that have a high cost-saving impact. Within those functions, several challenges were identified and research into methods to overcome them was conducted. For example, sparse labels of significant safety events in NPPs (due to their rare occurrence) would limit AI's ability to understand how and when to classify an event as a significant safety event. Several solutions are proposed in this research to tackle such challenges.

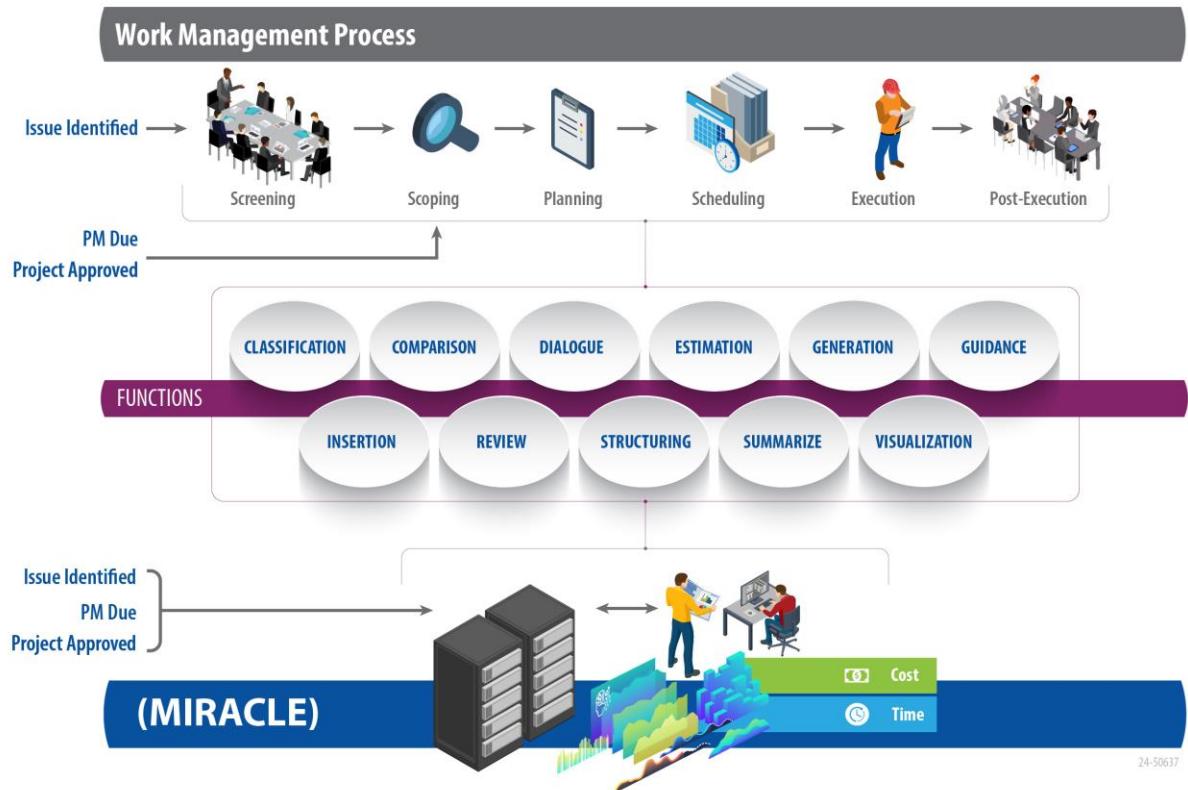


Figure 3. MIRACLE expanded mission to automate the work management process.

1.3 Report Structure

The report layout is summarized in Figure 4. The following Section 2 introduces the scalable AI methods and tools for any form of text analysis, covering text preprocessing and supervised, unsupervised, and semisupervised approaches. Section 3 explains how these methods and tools are applied to achieve the outlined AI functions, with example demonstrations using real utility data. Section 4 provides an overview of the work management mapping approach, outlines the assumptions made, and discusses the main phases. Appendices A-F break down each phase into numerous specific steps currently performed by humans, detailing the decisions and actions taken and the time each step requires, as well as how these steps could be automated using the discussed AI functions. Section 5 presents an estimate of the aggregated potential cost savings from automating each phase. Based on these savings, Section 6 discusses the cost savings that can be achieved by each AI function. Finally, Section 7 discusses the challenges of deploying these functions and presents research into overcoming these challenges.

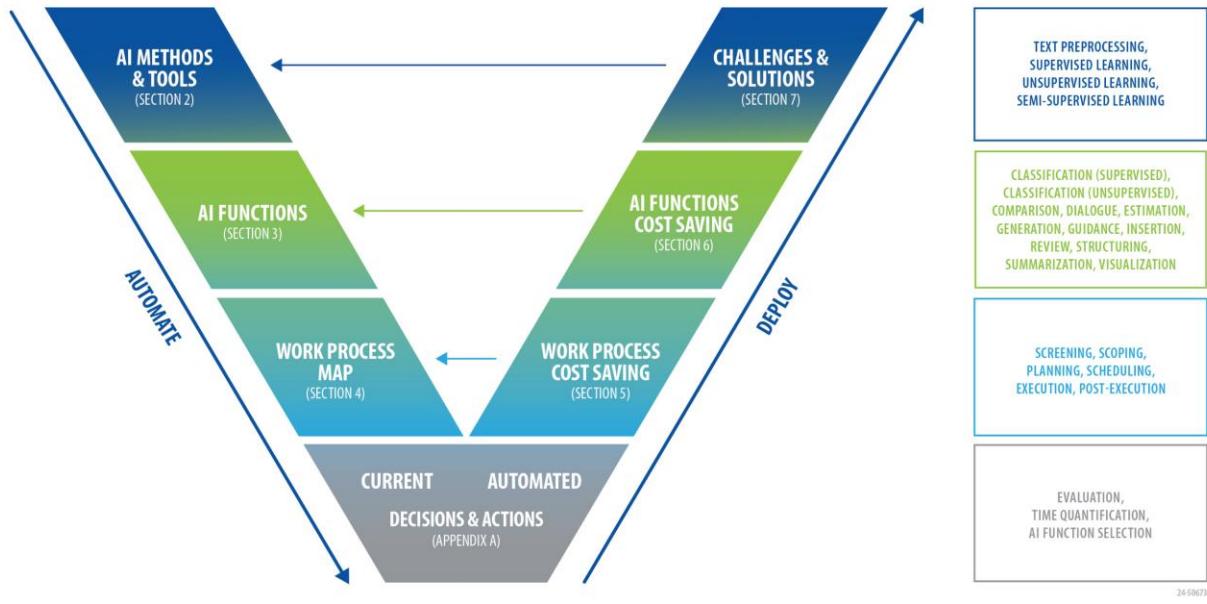


Figure 4. Visual abstract of the report denoting the scope of each section.

2. AI METHODS AND TOOLS USED FOR TEXT ANALYSIS

AI for text analysis is a research-intensive area with several new methods and tools being created every day. However, the vast majority can be categorized into specific groups that describe their general purpose. In this effort, the methods and tools used are grouped for text preparation and modeling, which is broken down into supervised, unsupervised, and semisupervised learning. This section summarizes the AI methods and tools in the context of their purpose. Examples of the usage of these methods and tools in specific applications are provided in Section 3. Section 7 examines the challenges associated with using these methods and tools at scale to achieve the targeted AI functions, along with research and demonstrations into potential solutions.

2.1 Text Preparation

Most data preparation solutions are ready-to-use tools that have been demonstrated and used across several industries. In this section, the commonly used tools are introduced with a short overview on how they work. Details on the underlying methods can be found in the listed references.

2.1.1 Rules-Based Processing

Free-form text presents a challenge to analyze since it encompasses words that span the entire English language. Making it even more difficult, free-form text often includes abbreviations, industry-specific jargon, and—very often—misspellings. Before text could be applied to any AI method, it had to undergo a detailed data curation and be reduced to a form that could be fed into a model. This activity removes text discrepancies and errors that could impact the results, such as by correcting common misspellings, and refines and simplifies the text for efficient processing, such as through spelling out common acronyms, lemmatization, etc. Specifically, data preparation can include any of the following:

- Determining which columns are to be used (i.e., go through each text field to decide which are useful out of potentially hundreds of columns for some utilities). This involves interacting with each utility subject-matter expert (SME).

- Eliminating invalid values (e.g., NaN in the text fields). The use of such abbreviations is very common by utilities but inconsistent across the industry, necessitating consistency in how they are used in AI.
- Removing all meaningless nonalphanumeric characters in text fields (e.g., `(\|/(\|)<|\>-*)+`) and replacing those that do have meaning (e.g., &).
- Removing duplicate records. In some cases, this results in a significant reduction in the amount of text.
- Removing very short records, since they are meaningless (e.g., Canceled, Error, reviewer_comment).
- Spelling out all contractions (e.g., don't → do not).
- Spelling out acronyms and abbreviations (e.g., convert "PM" to "preventative maintenance," "Bkr" to "breaker," and "Crit" to "criticality").
- Fixing common spelling errors via automatic/manual rules-based processes.
- Fixing words split apart by unwarranted spaces (e.g., "cl eared early amps ears ean amp ear adding ip"). This type of artifact often exists due to data migration from one tool to another within a nuclear utility.
- In some cases, removing employee names (e.g., Doe, John). For this work, a list of about 116,000 first, last, and surnames for men and women from countries around the world was compiled. Names pollute the vocabulary with high-frequency noise. Such noise can reduce the usefulness of the term frequency-inverse document frequency (TF-IDF, discussed in a later section).
- Removing frequently occurring meaningless text (e.g., "Operator Input:").
- Lemmatization, in which words are reduced to their most basic forms in the context of the other tokens surrounding them (e.g., convert to singular form, change past to present tense).
- Creating n-grams of meaningful commonly used words in a specific context (e.g., "high-pressure coolant injection" becomes one n-gram). Frequently occurring patterns of one, two, or three tokens—called unigrams, bigrams, and trigrams (referred to as n-grams), respectively—are identified.

The preprocessing steps that are necessary depend on the data, the method used, and the intended application. Once the relevant steps are applied, the text is reduced to its foundational format. This often results in a more robust use of AI methods to extract the relative meaning and context of each text field so that a machine can make an automated decision.

Text preprocessing often relies on the user defining a set of predefined rules or patterns. In its simplistic form, a find and replace tool or query is used to search for specific text or a pattern of words and replacing it with alternative values. While the basic form of this approach is straightforward, it can be used to handle more complex tasks with advanced rules, especially when inconsistent data is processed. For example, rule-based processing can be used to detect names and codes (for procedures, manuals, etc.).

2.1.2 Tokenization

Tokenization is the process of converting text into smaller units known as *tokens*. This step is used in natural-language processing (NLP) to convert raw text into a format (i.e., numbers) that a machine can analyze. The tokenized value for a word has no real meaning in a linguistic sense; it is simply a number chosen to represent the given word. Tokens can represent sentences, words, parts of words, or characters. For example, "This is an example" can be presented as one token, or it can become "This," "is," "an," "example," "." if each word is a token, or "T," "h," "i," "s," "i," "s," "a," "n," "e," "x," "a," "m," "p," "l," "e," "." if each character is a token. The most used form of tokenization is word based. Converting the tokens into numbers is performed by assigning numbers to all words in a training dataset, then simply

converting the tokens to their assigned numbers. For example, the tokens “This,” “is,” “an,” “example,” and “.” would become [1, 2, 3, 4, 5]. More details on tokenization can be found in Grefenstette (1999).

2.1.3 Vectorization

Text vectorization is the process of turning words, sentences, or whole documents into a vector of numbers in a space with many dimensions. This step is necessary because vectors can be mathematically analyzed. Similar text ends up with vectors that are close to each other. Using “This is an example,” if a corpus^a contains 100,000 unique words, then the “This is an example” sentence could be converted to a vector with the size 1x100,000, with 1 assigned to the words that appear in “This is an example” and 0 otherwise. This method is called *hot-encoding* and results in sparse vectors (vectors with mostly zero values). The resulting vector might look like this:

this	and	is	person	...	example	an
1	0	1	0		1	1

Other methods include bag of words (BoW; Zhang, Jin, and Zhou 2010) and TF-IDF (Ramos 2003), both of which are common techniques used in NLP for vectorization. BoW is a process that counts the occurrences of each word in textual data and groups words as bags of words based on their frequency.

TF-IDF adds importance to BoW by comparing the prevalence of words in an individual document against their prevalence in a corpus. TF-IDF executes two main steps: The term frequency (TF) measures how frequently a term (word) appears in a document, and words that appear more often are considered more important. Inverse document frequency (IDF) aims to regulate the TF output by removing words with less meaning, such as “and.” It evaluates how rare a term is across all documents in a corpus (not just a single document), and those rare words are given higher weight since they are more specific to the context than words such as “and.” This process ensures that words that are used frequently within a specific document but rarely across the entire corpus are given a high TF-IDF score. Using the same example, “This is an example,” this process could result in:

this	and	is	person	...	example	an
0.2	0	0.2	0		0.4	0.25

High numbers indicate the word is both frequent in a document and rare across the corpus. This explains why “example” has higher value than “an,” for example, indicating it is more informative. One key benefit of vectorization is that similar text will result in spatially close vectors. For example, “This is an example” and “An example is provided” would be denoted as similar because they have nonzero values for the common words “example,” “an,” and “is.”

2.1.3.1 Dictionary

A *dictionary* includes predefined, predeveloped sets of correlated words and is curated over decades of language research and available for direct use. A human or SME can also provide a dictionary that reflects the field-specific use of words. A dictionary can be used to assist vectorization by converting each word into a set of representative words. This approach ensures targeted semantic analysis and domain-specific knowledge inclusion. Dictionaries provide a structured framework that enhances specific interpretability, allowing for tailored text-processing tasks. For example, if the word “system” is used in a

^a *Corpus* is used in language process to refer to a large set of texts that often represents all text available for a problem.

sentence, it can result in a token for “equipment” since they both refer to similar or interchangeable concepts in the nuclear industry.

2.1.4 Embedding

Despite the ability of vectorization to detect text similarities, it is sensitive to the use of words. For example, if “This is an example” is compared to “A demonstration is presented,” vectorization is unable to determine that they both provide essentially the same information because “is” is the only overlapping word. This highlights the need for methods to incorporate context, which can be achieved by using embedding. Text embedding is the process of turning words, phrases, or whole documents into numerical vectors that have context relationships. These vectors show how words are related in meaning and how they are used in sentences. This process allows machines to understand and process natural language more effectively for tasks like text classification.

Embedding can be considered fundamentally different from vectorization because it reduces data dimensionality to force words with similar context to be grouped into single vectors. In essence, embedding extracts from the text the features that matter. This is different from the previously discussed tokens vectorization, which tends to result in high-dimensional, sparse vectors. The first step in embedding is to choose a desired reduced dimension; this is the embedding vector size. For example, if a corpus contains 100,000 unique words and it is desired to reduce all the vocabulary to 1,000 dimensions (i.e., all contextually similar words in that 100,000-word vector are represented by 1,000 dimensions only), a matrix of $1,000 \times 100,000$ is generated. Using this matrix and an optimizer, it is possible to use a corpus (with billions of words but only 100,000 unique ones) to tune how the matrix would convert any document^b into 1,000 dimensional embedding vectors. Every document can therefore be converted into a vector of 1,000 dimensional embeddings that represent the document context.

To train the embedding matrix, there are two choices available. One uses the available corpus. For example, if the work management data is available, the text can be used to develop the embedding matrix. The second uses nonrelevant text that is available (e.g., that of Wikipedia or WWW). The first approach is very application-specific aware, while the second encompasses general knowledge.

Creating the correlation of words can be performed with statistical methods like latent Dirichlet allocation (LDA; Tong and Zhang 2016), positive pointwise mutual information (Schneider 2005), and hyperspace analogue to language (Azzopardi, Girolami, and Crowe 2005). These methods analyze the co-occurrence and distribution patterns of words in a corpus and generate vector representations that capture the semantic and contextual relationships between words.

Although a matrix is used in this section to explain embedding, it is also possible to implement embeddings using nonlinear models, such as a NN. The underlying concept is the same—that is, a model is presented, and an optimizer is used to train it to generate embeddings using a training dataset. The training process could involve attempting to predict a word or sentence used in a corpus, for example. Those nonlinear models often contain billions of tunable parameters and are therefore capable of detecting complex relationships among words (i.e., capture broader context). Such models are referred to as large language models (LLMs). An example of an LLM that was trained on preexisting data is Google’s BERT (Bidirectional Encoder Representations from Transformers), which was trained on Wikipedia and books. LLMs are ubiquitous NNs that can be used for many tasks, such as classification, question answering, summarization, translation, etc.

Word2Vec (Rong 2014) is an example of an LLM that can be trained for a specific corpus. It can be trained on nuclear-specific text to understand the word analogies in the nuclear field. Word2Vec is a NN that reconstructs linguistic contexts for each word. It converts every word into vectors in a continuous

^b A document is usually used in the context of NLP to a single piece of text that serves as a unit of analysis or processing.

space in which similar words have similar vectors. Word2Vec has a proven ability to form analogies and produce reasonable word similarities, (Church 2017).

If several of the mentioned encoding methods are used, the context of a word, sentence, or document can be significantly enhanced. By integrating embeddings, the underlying and hidden relationships can be revealed, improving any text-related analysis.

2.1.4.1 Transfer Learning for Embedding

Instead of using a pretrained LLM or training one for a specific corpus, it is possible to combine the benefits of pretrained models with those trained for a specific dataset. This is referred to as *transfer learning*. Transfer learning analyzes large sets of text (such as Wikipedia), learning patterns and relationships from it. The transferred models of this effort can be fine-tuned to focus a model on specific smaller-domain text. For example, a general LLM might consider the word “coolant” to relate strongly to vehicles, but when refined with NPP data it might associate it with reactor coolant systems. This enables modes to be biased towards the desired applications.

One example of an LLM that was used in this research is Falcon (Almazrouei et al. 2023). This model was used for this work because it can be downloaded and run locally, which is important given the sensitive nature of the data used. Falcon was trained on Wikipedia to infer context-sensitive document similarities by comparing pairs of Wikipedia articles. For instance, a 2016 Wikipedia article should be more similar to the 2023 Wikipedia article of the same name than any other article.

MiniLM is another LLM that was used locally in this research. It can be fine-tuned for specific fields too (Wang et al. 2020). However, its benefit is that it is smaller in size (hence the use of “mini”) and therefore enables faster deployment and requires lighter computational demand to train.

2.2 Supervised Learning

Supervised learning establishes empirical relationships between sets of inputs and specific outputs. These models are trained on labeled data, unlike unsupervised models, which do not use labeled data. Labeled data consists of pairs for which each input is associated with a specific output. An optimizer adjusts the model parameters to improve its performance in predicting the output given the input. This process is known as training.

A key aspect of supervised models is the need for labels. Using a large corpus, such as Wikipedia, to train a model does not involve labeling in the supervised learning sense, as the text is not explicitly tagged with output labels. For example, if the task is to classify text related to work scope, a label like “Chemistry” might be assigned to text that implicitly discusses chemistry. This label enables the model to classify the text into categories such as “Chemistry” or “Not Chemistry,” even if the term “Chemistry” is not explicitly mentioned. In contrast, labels in supervised learning are explicitly assigned to text or data points and are used as output targets.

2.2.1 Regression

Regression is a suite of methods within supervised learning that approximate the relationship between inputs and outputs. These methods can be based on linear or nonlinear models and may produce outputs that are categorical or continuous.

Linear regression methods approximate the relationship between inputs and outputs using a linear equation. They estimate the output as a weighted sum of the inputs, where the weights (parameters) are adjusted during the learning or training process. Ridge regression adds regularization to improve robustness, especially in the presence of multicollinearity. It has been used for decades and is considered valuable by applied statisticians (McDonald 2009). It follows the law of parsimony—that is, it seeks the simplest answer by trying to reduce the number of words that are relevant to the classification (Hastie 2020; Ng 2004). Logistic regression is another linear method used for classification that estimates

probabilities for each class. It consists of two main components: first, a linear combination of inputs weighted by coefficients, and second, a sigmoid function, which transforms this linear output into a probability value between zero and one for classification.

Another method that can be used in a linear manner is Bayesian regression. This method extends regression by treating model parameters as distributions rather than fixed values, incorporating uncertainty into the predictions.

Support vector regression is another regression method that can be used in a linear (or nonlinear) manner. It aims to find a function that approximates the relationship between input and output while ensuring that the deviation between the predicted and actual outputs remains within a specified margin—that is, it aims to ensure that a certain separation between the actual output and the regressed output is not exceeded (the error stays within a margin). Support vector regression balances the trade-off between fitting the training data and maintaining generalizability.

Nonlinear regression methods are designed to capture complex relationships between inputs and outputs that cannot be effectively represented by linear equations. These models can reveal hidden patterns and interactions within data. NNs are prominent examples of a nonlinear model. NNs consist of interconnected nodes, or neurons, in which each neuron applies a nonlinear function to its inputs. By stacking multiple layers of neurons, NNs can model high-order relationships through complex mathematical formulas. During training, the weights associated with each neuron's inputs are adjusted using input text. There are various types of NNs, including recurrent NNs (RNNs). RNNs are designed to handle sequential data by incorporating information from previous time steps, allowing the model to capture temporal dependencies, such as the sequence of words in a sentence.

A simpler and commonly used type of nonlinear model involves analyzing distances between data points in a reduced-dimensional space. The k-nearest neighbors (k-NN) algorithm (Cover and Hart 1967) makes predictions based on the closest k data points to a given query point using a distance metric such as Euclidean distance (distance between two points in Euclidean space). In the context of text, if “This is an example” and “This is a demonstration” are converted into a two-dimensional cluster of points, k-NN would locate those two sentences to be close to each other since the words “example” and “demonstration” are used interchangeably in a corpus, and therefore it would assign them the same decision. For continuous regression tasks (k-NNR), the average value of the k-NN is computed rather than assigning each data point to a specific class. For instance, if a new sentence like “This is a great example” is presented, the algorithm computes the distance between this sentence and the other two sentences. The algorithm then predicts a decision value based on the average of the values of its nearest neighbors, reflecting how closely the new sentence aligns with the known data points.

Another class of nonlinear models use decision trees. Decision trees are models that recursively partition and tune the partition using the data based on certain criteria. Each partition’s weight is tuned to enhance the model’s performance. A key advantage of decision trees is their interpretability and their ability to handle nonlinear relationships in data. One common method of decision trees is CatBoost (Dorogush, Ershov, and Gulin 2018). CatBoost is a gradient-boosting algorithm, specifically optimized for categorical variables. It builds an ensemble of decision trees and uses gradient-boosting techniques to sequentially improve the accuracy of decision trees. Using the example “This is a great example,” if the aim is to determine whether this is a positive or negative statement, sentences are split based on the presence of specific keywords. A sentence containing “great” might be classified as positive, whereas one with “awful” might be classified as negative. As the decision tree is trained on thousands of examples of labeled positive and negative sentences, the decision nodes adjust based on the patterns in the training data to provide the final classification. Decision trees also incorporate regularization methods to mitigate overfitting (i.e., preventing the tree from growing too complex).

2.2.1.1 Transfer Learning

Instead of creating a model from scratch in regression, it is possible to perform transfer learning on an LLM beyond the text-preprocessing tool discussed in Section 2.1.4.1. The LLM can be integrated with a linear or nonlinear model to perform a function. The LLM and the attached model parameters could be tuned simultaneously to perform the same function that a standalone linear or nonlinear model would accomplish. This enables the integrated mode to use the domain knowledge that is encoded in the LLM.

2.3 Unsupervised Learning

Unsupervised learning identifies patterns or structures within text without relying on labeled data. Unlike supervised models, which are trained on data where each input is associated with a specific output label, unsupervised models work with unlabeled data and focus instead on finding underlying relationships or groupings. The creation of embeddings discussed earlier is a good example of an unsupervised model used to understand text correlations.

2.3.1 Clustering

Clustering relies on finding relationships between data (e.g., words) by grouping similar occurrences (text that shares common context) into clusters. This approach uses the co-occurrence of data to extract patterns based on these relationships. An example of such a method is k-means clustering (Ahmed, Seraj, and Islam 2020), which organizes data into a user-specified number of groups (clusters) based on similarity. K-means clustering works by initially selecting k centroids (center points) and then iteratively adjusting these centroids to minimize the distances (e.g., Euclidian distance) between each data point and the nearest centroid. This process is repeated until the centroids converge, meaning they no longer move significantly. For example, k-means might group the two sentences “This is an example” and “This is a demonstration” into one cluster because they are similar, while placing the sentence “The sky is blue” into a different cluster due to its distinct content.

2.3.2 Dimensionality Reduction

Dimensionality reduction aims to reduce the number of variables (typically high-dimensional word vectors) under consideration by obtaining a set of key text—that is, by keeping important text only. Dimensionality reduction serves multiple purposes. It can be used to create embeddings (discussed earlier) and aids in isolating the most relevant information to improve model computational performance in tasks such as text classification. It is also a powerful tool in removing noise. Additionally, it can help humans visualize high-dimensional text data, enabling them to understand data relationships and patterns.

Principal component analysis (Abdi and Williams 2010) is a common technique used for linear dimensionality reduction. Principal component analysis works by identifying the orthogonal principal components along which the data information is contained. It reduces high-dimensional data into a lower-dimensional format while sustaining key variability.

Uniform Manifold Approximation and Projection (UMAP; McInnes et al. 2018) is a nonlinear method often used for visualization. It reduces data dimensions while preserving the manifold shape of the data. This means that clusters close together in the UMAP space have higher similarity than those farther apart. The orientation of the clusters is also important. Orthogonal clusters that are close together use similar words, but in the context of other dissimilar words.

2.3.3 Generative AI

Generative AI refers to the various approaches used to create new text out of some input text. The various approaches and tools used within this category are summarized in Table 1 and will be discussed in this section. Except for generative adversarial networks (GANs), all the methods listed use decoders, so it is necessary to introduce decoders ahead of the methods discussion.

Decoders are tools for converting an abstract form of text (such as embeddings) into a readable text output. Decoding could be developed using an RNN, for example, and trained using various techniques. For example, greedy decoding would use the training text to select the most likely next token at each step, aiming for the highest probability sequence. This technique assumes that the model is not aware of the future words and simply aims to select the most likely token based on the current state. It uses the training text to tune the NN until the model generates the word that was provided in the training text. In this method, if the prompt is “the valve is,” the model might generate “leaking” as the next word because it has the highest probability based on the training data. The model repeats this process, choosing the word with the highest probability each time, resulting in a sequence like “The valve is leaking water.”

Table 1. The different types of generative AI methods considered in this research.

Model	Autoregressive Models	Encoders-Decoders	Dialogue Models	Generative Adversarial Networks (GANs)	Retrieval-Augmented Generation (RAG)
Objective	Reconstruct text based on its knowledge	Reconstruct text in targeted manner based on training	Interact with users in a coherent manner	Generate new data out of examples of historical data	Searches through large set of documents to create answers from the documents
How it works	Predict the next word using a regression model	Predict the text using the whole representation of the input	Trained on generic questions and answers	Iterate over fake data until it looks like (but is not) the real data	Finds text related to the query and formats it in an answer to the query
Strength	Mimic human text generation	Clean up and augment data	Improve response coherence	Can create novel data	Can mine large sets of existing data

2.3.3.1 Autoregressive Models

Autoregressive models are methods used for generating sequential data by applying regression models to text to determine what the next word will be based on the previous text (e.g., set of words). An autoregressive model relies on training the regression model using a large corpus of data. The core principle behind them is the use of decoders that generate outputs sequentially. This sequential generation is achieved by making a prediction based on the previously correlated text. The models create a probability distribution of the next element in a sequence, based on the context provided by preceding elements, and generate realistic and coherent sequences of data.

The well-known ChatGPT is an example of an autoregressive model. ChatGPT’s model operates by processing sequential data. It predicts the next word in a sentence or conversation by analyzing all previous words, performing complex correlation of text. This capability allows an autoregressive model to not only generate fluent text but to also mimic the style and coherence of the corpus text data it was trained on.

2.3.3.2 Encoders-Decoders

Encoder-decoder models are designed to convert input text into an output text that may vary in content and length based on the application. They consist of two parts. The first part is the encoder (possibly using NNs), which aims to convert the original text into a reduced-order representation, often

contextually focused, called the latent space. The second part is the decoder; it takes the encoded text and generates the output text, which can be similar to or different from the input text, depending on the task. Encoder-decoder models are trained on large corpora of text to learn how to effectively map input text to target output text and can be trained in both supervised and unsupervised manners. Seq2Seq (Liu et al. 2018) models are examples of encoder-decoder architecture that use RNNs, for example, and they are designed for tasks where the input and output are sequences of text. Transformers are another type of encoder-decoder architecture that rely on self-attention mechanisms to handle long-range dependencies and parallelize processing and are therefore more sophisticated than traditional RNN-based Seq2Seq models.

Autoencoders are a specialized form of the encoder-decoder architecture where the input and output sequences are the same, and the goal is to reconstruct the input data rather than to transform it into a different format. Autoencoders are also unsupervised as they use the actual text to optimize the encoder and decoder parameters to minimize reconstruction error. Variational autoencoders introduced by Kingma and Welling (2013), are a common type of text autoencoder that rely on creating probabilities (i.e., Bayesian inference) in the latent space. Denoising autoencoders (Vincent et al. 2008) are a common type of text autoencoder that are trained by using a known input text that has several words removed. The goal of the training is to teach the model to recreate the original text without those words included in the input (i.e., masking those words).

2.3.3.3 Dialogue Generation Models

Dialogue generation models are methods aimed at producing contextually suitable responses to requests or questions in conversations. They can be developed in a supervised manner where a machine is fed specific requests and answers to learn how to reply. More commonly they are used in an unsupervised manner, where they are trained on a large dataset of text that includes dialogue and find correlations between requests and responses on their own. An autoencoder can be used in which the input is the request (e.g., a question) and the response comes from the decoder.

Dialogue generation models can be developed by a variety of techniques aimed at generating coherent and contextually relevant responses in conversational AI systems (e.g., chatbots). For example, Seq2Seq models can map input dialogues to output responses using RNN. Reinforcement learning techniques (Li et al. 2016) can also be used. They rely on user feedback, encouraging the user to generate good answers.

2.3.3.4 Generative Adversarial Networks (GANs)

GANs are methods used to generate novel text that resembles some provided examples. GANs do not use decoders. Instead, they consist of two main components: a generator and a discriminator. The generator generates text, attempting to mimic some provided text examples. The discriminator, on the other hand, is trained to distinguish between the provided text examples and the text produced by the generator. The generator aims to fool the discriminator by adjusting the text to look real, while the discriminator aims to catch the generator's attempts to fool it and decide whether the text is real or not. This is the reason why this method is called “adversarial” (i.e., the components compete). When the generator can create indistinguishable text, it is considered a success, and that text can be used to resemble real text. TextGAN (Zhang et al. 2017) is an example of a text GAN that modifies the conventional GAN architecture to accommodate the discrete nature of textual data by employing a Seq2Seq model as the generator. The discriminator evaluates the entire sequences generated by the generator, rather than individual tokens, for the overall model to effectively learn and generate coherent text sequences.

2.3.3.5 Retrieval-Augmented Generation (RAG)

RAG (Lewis et al. 2020) is an approach that uses similarity analysis to compare text provided as input to a large set of documents to determine which of the documents most resemble the text, then it

summarizes all those relevant documents in one response. RAG acts like a librarian that can mine large sets of documents to provide a concise summary focused on the provided input.

The RAG process involves multiple steps, which are laid out in Figure 5. Proceeding from left to right, the user poses an inquiry or question to the RAG system. The user's inquiry is then encoded into a text embedding using an LLM that has been pretrained for the purpose of determining the similarity between the user's question and the entries in a previously assembled vector database that contains the text embeddings of the domain-specific data. The similarity comparison itself is performed using the cosine similarity algorithm (discussed in the next section). Once the set of similar documents has been retrieved, they are presented to a different LLM that has been fine-tuned for conversation (an instruct model). The instruct model then summarizes the documents and provides the user with the summaries. Decoders are subsequently employed to convert these document embeddings into human-readable text, focusing on capturing the core context.

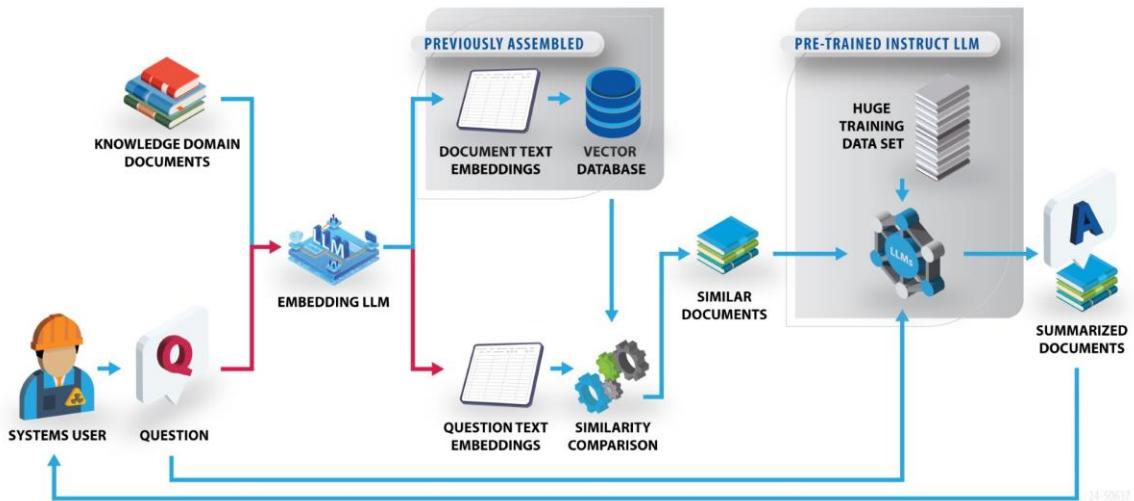


Figure 5. The RAG process.

2.3.4 Similarity Analysis

Similarity analysis is an approach that compares text to another set of texts to determine if they are literally or contextually similar. This approach relies on extracting the key embeddings vector of one input text and comparing it to the vector of another input text. The distance between these multidimensional embedding vectors can be used as an indicator of similarity (e.g., Euclidean distance for relatively small-dimensionality data). Similar text would result in vectors that are close to each other. For large data dimensionality, the angle between the vectors can be used to determine similarity instead (e.g., cosine similarity; Rahutomo, Kitasuka, and Aritsugi 2012).

2.3.5 Topic Modeling

Topic modeling is an approach to extract the main theme (i.e., the topic) of text by extracting the text's context and consolidating the underlying information into a set of keywords. The approach relies on the occurrence of correlated words that hint at the theme. The co-occurrence of multiple correlated words increases the method's confidence in the theme of the text. The models can also be fully unsupervised, by which the text is mined for correlated words and consolidated into a few keywords, or semisupervised, in which a subset of words is provided for a topic and the text is tested against them. If the text words resemble the topic words, the text can be assigned to that topic.

LDA is a topic modeling method that uses the underlying assumption that a collection of documents can be approximated by a set of latent topics. It produces a set of concepts related to words that frequently occur within a given corpus and accomplishes this by emphasizing the weight of high-frequency words that occur together. Anchored correlation explanation (CorEx; Gallagher et al. 2017), a state-of-the-art technique for topic modeling, leverages domain knowledge via an information-theoretic approach that does not rely on any generative assumptions about latent topics. CorEx assumes that correlations between documents can be explained by sets of mutually informative words, and CorEx involves similar assumptions about the exchangeability of words in documents. The CorEx algorithm was built to also accept domain-specific guidance, which then motivates its ability to discover highly informative topics. Whereas LDA considers the frequency of words in each document, CorEx uses TF-IDF document embeddings. Another well-known transformer-based topic model called BERTopic (Grootendorst 2022) can also be used to optimize the guided topic modeling for standard topics. Guided topic modeling means that the targeted topics are known before the model is run.

2.4 Semisupervised Learning

Though either supervised or unsupervised methods and tools were discussed in previous sections, a combination of both is often used. Those hybrid unsupervised and supervised learning methods (i.e., semisupervised methods) combine the strengths of both approaches to address limitations inherent in each method. They leverage the strength of unsupervised learning in discovering underlying text patterns and context extraction and the supervised methods ability to better model the training data. Several examples of these methods are discussed in Section 7.

3. APPLICABLE AI FUNCTIONS IN TEXT ANALYSIS AND EXAMPLE DEMONSTRATIONS

Having introduced the methods and tools used for text analysis, this section elucidates their application to specific functions. While some of these methods and tools are directly applicable to these functions, others may necessitate customization or the integration of multiple approaches to achieve the desired outcomes. This is essential to demonstrate how some tools can have multiple applications. The primary focus of this section is to explore these functions in detail and provide illustrative examples of some of the methods and tools that are used to serve a function in nuclear work management contexts. These examples will demonstrate the effectiveness and suitability of these functions. The maturation of each technology is discussed in each of the discussed sections. This exploration serves as a precursor to discussing how these AI capabilities can streamline and automate work management processes. The challenges associated with these methods are not addressed in this section; instead, common challenges and innovative solutions are examined in Section 7.

The functions targeted in this section are specific to work management automation. For example, sentiment analysis, which is a function of AI, is not discussed in this work since it is mostly used for social media data analysis and is not relevant to this effort.

3.1 Classification (Supervised)

Classification is the process of categorizing an outcome of text into a predefined set of categories. Supervised classification is a specific approach where a machine is trained to categorize texts using existing labeled examples. These labeled examples allow the machine to learn patterns in the text and replicate the decisions. The classification can be either binary (e.g., yes or no for safety classification) or multiclass (e.g., priority assignment with numbers between 1 and 10). The underlying approach for both types is similar, differing mainly in the final output layer of the model used.

3.1.1 Tools Overview

To be able to perform classification, the machine first aims to group words that have similar meaning or context in numerical forms through the tokenization and embedding processes mentioned in Section 2.1. This enables the machine to become more context focused rather than correlating singular occurrences of words when making a decision.

Classification can be achieved through a regression model that generates a probability, confidence, or likelihood. In this method, a threshold is set, and any value exceeding it is marked as belonging to a specific class. Alternatively, the model can directly produce the classification outcome in a discrete form, rather than as a continuous regression value. Linear regression models (Section 2.2.1) are among the simplest and most interpretable types of models used in supervised classification. They make predictions based on a linear correlation of input embeddings. Nonlinear models (Section 2.2.1) can capture nonlinear relationships between embeddings and the decisions and are usually more capable in complex decision-making processes.

3.1.2 Example Application

When an employee identifies a condition in writing and submits it to the CAP system, the resulting CR goes through a screening process to determine the importance of the condition; thus, the safety significance is assigned to the condition among several other decisions that are discussed in the screening process (Section 4.4). In this section the experimentation and implementation of classification capabilities for CRs in NPPs is demonstrated. The classification of CRs includes multiclass classification, for example: not a condition adverse to quality (NCAQ), a condition adverse to quality (CAQ), or a significant CAQ (SCAQ). In this example, the first two are targeted, and a binary classifier is sought.

CR datasets sourced from five utilities were used to train several text classification AI models to identify the one that performed the CR classification task the best. This training dataset encompasses multiple timeframes and facilities with their own internal processes. The dataset comprises 879,485 CRs. Each CR was labeled by the respective utilities as either a CAQ or NCAQ condition.

If a classifier is not used and all the CRs are classified as CAQ, an F1^c score of 31.1% can be achieved. This is because the data are imbalanced and the majority of CRs are NCAQ. Naturally the F1 score alone is not a good indicator; both recall and precision are needed. In this case classifying all CRs as CAQ resulted in a precision of 0 but a recall of 100. The TF-IDF algorithm was used next to convert documents into tokens, as discussed in Section 2.1.2. A ridge regression algorithm (Section 2.2.1) was used to fit the TF-IDF text embeddings to the CAQ/NCAQ labels for each CR. The results are shown in Table 2. Adding context awareness, Word2Vec (Section 2.1.4) was used to expand on the words before they were converted to numerical values. Two forms of large language models—Falcon and MiniLM (Section 2.1.4.1) were also used to generate more sophisticated embeddings to increase context awareness. This further degraded the performance. Next, nonlinear models were tested (CatBoost and k-NNR). Neither yielded improvements in performance.

The results for this specific example indicate that the linear model was the most successful. This is a common finding for the research community. Nonlinear models, despite their ability to model complex relationships, can be unnecessary and can overfit the data. The use of linear and nonlinear models is really a question of data linearity. Linear data are better modeled by linear models.

^c F1 score is a common performance metric that is used in AI.

Table 2. Summary of performance results for supervised classification of CAQ.

Method	F1 Score (%)	Precision/ Recall (%)	Confusion Matrix			
			Predicted			
Baseline	31.1	0.0/100.0	—	—	Predicted	
			—	—	NCAQ	CAQ
			Actual	NCAQ	0	717,544
				CAQ	0	161,941
TF-IDF + Ridge Regression	82.1	81.6/82.7	—	—	Predicted	
			—	—	NCAQ	CAQ
			Actual	NCAQ	687,348	30,196
				CAQ	28,048	133,898
Word Vectors + TF-IDF + Ridge	75.2	74.5/75.9	—	—	Predicted	
			—	—	NCAQ	CAQ
			Actual	NCAQ	675,529	42,015
				CAQ	38,970	122,971
Falcon Embeddings + Ridge	70.0	67.6/72.5	—	—	Predicted	
			—	—	NCAQ	CAQ
			Actual	NCAQ	661,268	56,276
				CAQ	44,544	117,379
MiniLM Embeddings + Ridge	65.7	61.4/70.7	—	—	Predicted	
			—	—	NCAQ	CAQ
			Actual	NCAQ	645,418	72,126
				CAQ	47,375	114,566
Falcon Embeddings + CatBoost	68.1	66.0/70.5	—	—	Predicted	
			—	—	NCAQ	CAQ
			Actual	NCAQ	658,695	58,849
				CAQ	47,827	114,114
MiniLM Embeddings + CatBoost	73.6	71.3/76.0	—	—	Predicted	
			—	—	NCAQ	CAQ
			Actual	NCAQ	668,012	49,532
				CAQ	38,821	123,120
Falcon Embeddings + k-NNR	63.9	63.1/64.7	—	—	Predicted	
			—	—	NCAQ	CAQ
			Actual	NCAQ	674,572	42,972
				CAQ	67,326	94,615
MiniLM Embeddings + k-NNR	75.7	76.5/74.8	—	—	Predicted	
			—	—	NCAQ	CAQ
			Actual	NCAQ	689,849	27,695
				CAQ	48,391	113,550

3.1.3 Technology Readiness

As demonstrated in this section, there are several methods that can be applied to automated supervised classification. Several embedding models exist, and models are continuously expanding the number of parameters (i.e., context) they can understand. Also, several linear and nonlinear models are approaching the maximum performance value that is possible, given the quality of the provided data. It is important to note that the performance of the models mentioned herein is not indicative of their highest performance potential, but rather what they could achieve given the quality of the data used (as is described in Section 7).

Summary of Technology Readiness: Ready but the context awareness capabilities of the tools should continue to be improved.

3.2 Classification (Unsupervised)

Unlike supervised classification, unsupervised classification is used for data that is not labeled. Instead, unsupervised classification automatically identifies groups in the text and establishes clusters that can be assigned a certain label. The user can control the number of groups the data is broken into. This parameter can be adjusted to fit the specific needs of the analysis. The groups can be binary or multiclass. The approach for both types is similar, differing mainly in the number of clusters the algorithm aims to identify. For example, in a binary classification scenario, the algorithm aims to divide the data into two distinct clusters (e.g., Yes or No clusters). In contrast, a multiclass classification scenario involves dividing the data into more than two clusters, each representing a different category or theme.

3.2.1 Tools Overview

Unsupervised classification can leverage various techniques to effectively group and categorize data. Clustering algorithms such as k-means are the most common approach for identifying natural groupings within the text based on similarities in a k number of groups. Topic modeling can also be used to identify topics or themes within a collection of texts for classification. On the generative front, autoregressive models such as ChatGPT use broader context to fill in the text based on word-to-word prediction. They can be used to answer direct questions on how likely it would be for a text to fall into one category based on the data they trained on and similarity of the text to the training data.

3.2.2 Example Application

In this example demonstration, logs that contained few words describing any operator-taken actions were mined for shutdown, failure, and startup for reliability purposes. Those logs were not labeled as any of those events. Therefore, they were classified using unsupervised models. Like CRs, they are free-form text but have no labels for what the nature of the action is. Since the data is not labeled, 100 operator logs were manually labeled by an SME and used as the truth to compare against.

In this example, three sets of methods were used. Clustering was performed using k-means first, in which the text was either classified as one of those classes or not. This resulted in three classifiers, each comprising two clusters. The results are shown in Table 3 and show no significant improvement over a random classifier.

Topic modeling was also used by CorEx. The single keywords “start,” “shutdown,” and “inoperable” were used as seed words to generate the results. This proved to be more effective than using a more complex list of keywords, possibly because CorEx finds the rest of the important keywords. CorEx outputted the probability that a given operator log belonged to a particular class. The probability threshold chosen for the classification was guided by the need to balance precision and recall to produce an optimum F1 score for each task.

Table 3. Summary of performance results for unsupervised classification of three events.

Method	Shutdown F1 Score (%)	Startup F1 Score (%)	Failure F1 Score (%)	Mean (%)
Random Baseline	66.6	66.6	66.6	66.6
Clustering Methods				
K-Means	66.6	67.5	66.6	66.9
Topic Modeling				
CorEx	81.4	76.6	68.5	75.5
Autoregressive Methods				
Generic Ratings Dialogue Using Vicuna	82.4	71.0	70.1	74.5

Next, an autoregressive model was used. Vicuna 1.3 (Zheng et al. 2024) was asked to provide responses that varied from 1 to 10 in terms of confidence that an operator log was related to a certain type of event.

Prompt:

“Please only give a number between 1 and 10 as your response. A shutdown score of 1 means definitely not an equipment shutdown. A shutdown score of 10 definitely represents an equipment shutdown. In this nuclear power plant operator log what is the shutdown score only: ‘[OP LOG]’”

Response:

“1”... “10”

Various thresholds were tested for the rating, and the best results are shown in Table 3. A similar prompt was used for failure and startup operator-log entry classification. Despite the larger model and the detailed prompt that was provided, the results showed minor improvement, except for startup where it seems it was very sensitive to the prompt used.

Both generative-AI-based models and topic-modeling models performed relatively similarly. Most of the misclassifications using generative AI were due to its difficulty understanding where such events would fall. For example, the models identified a breaker trip as a failure because it indicates a failure-caused safeguard, yet the human disagreed. The right answer is very subjective and specific to the type of classifier used.

3.2.3 Technology Readiness

Unlike with supervised classification, there are fewer approaches that can apply to automated unsupervised classification. Current methods are being explored to enable generative AI models to understand the criteria of classification better. Also, prompt engineering was identified as a key challenge for the deployment of autoregressive models. Different prompts generated different results, emphasizing the need for proper prompt engineering.

Summary of Technology Readiness: Ready but there is a need to continue to improve the incorporation of domain-specific knowledge decisions and prompt engineering.

3.3 Comparison

Text comparison aims to determine whether one text resembles another within a corpus, and it is commonly used to replicate decision-making based on the assumption that similar inputs yield similar outputs, or to indicate that the text is contained in another text. This comparison can be performed either

literally or contextually. Additionally, the texts being compared do not need to be of similar length, as the focus is on the underlying context, which is independent of the document size.

3.3.1 Tools Overview

The most common methods for text comparison use similarity analysis techniques to evaluate how closely two text elements are related by examining their embeddings. These embeddings are projected into a high-dimensional space, allowing the similarity analysis to quantify the closeness of the texts based on their vector representations. Clustering is another effective approach, grouping similar texts into the same cluster based on their embeddings; texts within the same cluster are considered related, while those in different clusters are viewed as dissimilar. Dimensionality-reduction techniques further aid in text comparison by reducing the high-dimensional space of embeddings to a lower-dimensional space while preserving the relative distances between similar texts. Unlike similarity analysis, which operates in high-dimensional space, dimensionality reduction simplifies comparisons by positioning similar texts closer together in a reduced space.

3.3.2 Example Application

In this example, three CR entries were compared against each other using three similarity analysis methods. The aim was to determine if the algorithms can detect that the CRs are similar, and hence should have the same CAQ/NCAQ classification (discussed earlier). The three CRs are:

Handcrafted CAQ: “*There was a steam leak coming from a check valve near the reactor core isolation cooling system. This leak caused damage to piping and resulted in a boric acid spill. There may need to be revisions to technical specifications.*”

Sanitized CAQ: “*During the walkdown, dry white boron was observed near [AREA]. This is not reportable, and operability is maintained. The presence of dry white boric acid suggests an inactive leak with minimal or no corrosion. There is reasonable assurance that the leak did not affect the flow of the available emergency core cooling system. Operability remains intact.*”

Sanitized NCAQ: “*Half of the high bay lights in [Room] are out, leaving one end particularly dark. Please replace the faulty lights.*”

The results shown in Table 4 indicate that the handcrafted and CAQ text were deemed similar, while the one that was randomly selected and is NCAQ was deemed different.

Table 4. Pairwise cosine similarity of three conditions using three different models.

Condition 1	Condition 2	TF-IDF Cosine Similarity (%)	MiniLM Cosine Similarity (%)	Falcon Cosine Similarity (%)
Handcrafted CAQ	Sanitized CAQ	7.9	29.0	67.8
Handcrafted CAQ	Sanitized NCAQ	1.8	-3.2	25.7
Sanitized CAQ	Sanitized NCAQ	0.0	-2.6	27.5

3.3.3 Technology Readiness

Similarity analysis tools can identify text similarities both literally and contextually. When using TF-IDF or vectorization alone, the focus is on literal similarity. In contrast, embedding methods emphasize contextual similarity. These methods have primarily been demonstrated for short texts, and similarity analysis for large documents is still under development. It is anticipated that there will be issues related to focus with large documents.

Summary of Technology Readiness: Ready to use for short text. Needs to be demonstrated for large documents.

3.4 Dialogue (Chatbots)

Dialogue systems are increasingly utilized across various industries because of their ability to simulate human conversational skills. These systems are designed to interact with users in a seamless and natural manner, leveraging underlying AI models, such as LLMs, to convey knowledge. They are built to understand and interpret text inputs, maintain conversational context, and generate responses that are both coherent and contextually appropriate.

3.4.1 Tools Overview

Dialogue systems are primarily developed using generative AI methods, as discussed in Section 2.3.3. Autoregressive models can be employed due to their training on data that includes question-answer pairs. Encoder-decoder architectures are also effective when trained to handle specific types of text responses. While less common, similarity analysis can be used to train models to identify questions and provide corresponding answers based on the similarity of the new questions to the ones used in training.

3.4.2 Example Application

Interacting with a machine is a common tool for every chatbot (leveraging autoregressive models). Therefore, it is desired in this section to demonstrate when a chatbot fails, mainly due to hallucination. AI hallucinations occur when a chatbot creates outputs that are illogical or entirely fabricated. To give an example of a hallucination the Microsoft’s Phi-3 mini-LLM was used with reference to the Nuclear Energy Institute NEI-99-02 procedure (NEI 2023):

Prompt:

“Log: ‘Reset panel BKR [EQUIP ID] [LOCATION] per [CODE] BKR tripped due to welding in the area.’ ... Please explain if you are sure that this equipment failure is covered under NEI-99-02, don’t make assumptions.”

Response:

“...Regarding NEI-99-02, this standard covers the Unavailability Reporting and Investigation (URI) for nuclear power plants. The log entry describes a failure of a breaker, which is a monitored component. Therefore, this equipment failure is covered under NEI-99-02.”

The hallucination is that URI is not an acronym for “Unavailability Reporting and Investigation.” Instead, URI is an acronym for “Unreliability Index” in NEI-99-02. The use of this incorrect acronym it casts doubt on the truthfulness of the model’s responses. Furthermore, the statement “don’t make assumptions” seems to have had no effect on the incorrectness of its assumption regarding the acronym URI.

AI hallucinations arise from several factors (Ji et al., 2023). Subjective bias, or innate divergence, occurs when a language model is trained to be engaging rather than strictly factual. Nonmemorized learning, or parametric knowledge bias, happens when the model generalizes from training data instead of using specific texts. Imperfect representation learning involves the model misinterpreting or fabricating facts due to its flawed comprehension of patterns in the data. Longer outputs are more susceptible to hallucinations due to exposure bias. Additionally, defective training data, or heuristic data collection, results from using unchecked datasets, given the immense volume of text involved.

3.4.3 Technology Readiness

It has been demonstrated that autoregressive models can answer questions that they have been trained on. However, it remains questionable if they can accurately answer specific field questions that are beyond what is typically used in training the models. Also, hallucination is one of the main limitations of dialogue systems since it erodes human trust.

Summary of Technology Readiness: Ready for simple dialogue but unknown for complex conversations and tasking. More research is needed to reduce hallucination.

3.5 Estimation

Estimation is the process of modeling the relationship between text inputs and a continuous response variable. This process starts by training an AI model in a supervised manner, where the model learns from a dataset consisting of input text and corresponding continuous responses. The goal is to uncover patterns and correlations within historical data that allow the model to predict continuous outcomes for new unseen text. The key distinction between estimation and classification lies in the nature of the predictions: estimation predicts continuous values, such as duration; classification, on the other hand, categorizes inputs into discrete class labels.

3.5.1 Tools Overview

Estimation in AI using text data as input can be achieved through supervised learning methods, which can be either linear or nonlinear. The core of these models involves assigning weights to each word or its numerical representation, such as tokens or embeddings. In linear models, each word embedding is assigned a weight, and these weighted values are summed to produce a final prediction. In contrast, nonlinear models, such as NNs, handle embeddings and their weights in a more complex manner. These models process embeddings through multiple layers of operations, enabling the models to learn the intricate dependencies and interactions between words. This approach allows nonlinear models to capture more nuanced and sophisticated patterns in text data, potentially leading to more accurate predictions.

3.5.2 Example Application

In the work management process the vast majority of estimation applications are used to predict the duration or time needed to complete certain tasks. Therefore, a similar use case was targeted for this example. This example estimated the time to complete a CR (i.e., the origination date of a CR until the completion date of a CR). This time to completion is measured in days and is needed for planning purposes (i.e., to estimate how much time it typically takes to resolve the issue). The training dataset comprised about 78,000 CRs from a utility. Each CR was labeled by the utility with a CR time to close.

A linear model was used for this example. TF-IDF was used with ridge regression to predict the number of days from the origination of a CR until it was completed. Five-fold cross-validation was performed.^d The correlation between the actual and predicted time should fit a line with a slope of unity. However, as the points do not follow a clean linear distribution, the resulting line might not exactly fit the desired shape, and the noise would reduce the line fitness. R^2 is a statistic that directly measures the predictions of a curve against its known values for goodness of fit (Kasuya 2019). It normally ranges from 0 to 100%, with 100% indicating a perfect fit.

Figure 6 shows the actual and predicted time. Log transformation was used given the variation in values in both axes. For the percentile interval a sliding window of the predicted duration of 1,000 CRs was used. The text predicts CR duration with an R^2 between ~30% and ~45% in the first several years. Then the model degrades during the latter years with some time periods being no better than random chance. The average R^2 value is 33%. This mainly indicates that the text that exists in the CRs is not sufficient to predict the time it takes to complete the event. However, these results could still indicate that the machine may perform better than a human. Earlier efforts demonstrated that the machine was able to generate better results than the human despite its poor performance (Lawrence et al. 2023). Both the human and machine suffered from their inability to understand the context and background of the text.

^d Cross-validation is a technique in which a subset of the data is omitted during the training process and used for validation. This omitted subset is then included in the training, and a different subset is left out for validation. This process is repeated multiple times to ensure that every part of the data is used both for training and validation, providing a robust assessment of model performance.

For example, one condition could be resolved in weeks and another similar one might take years because of the unavailability of explicitly mentioned background in the text for a machine to learn.

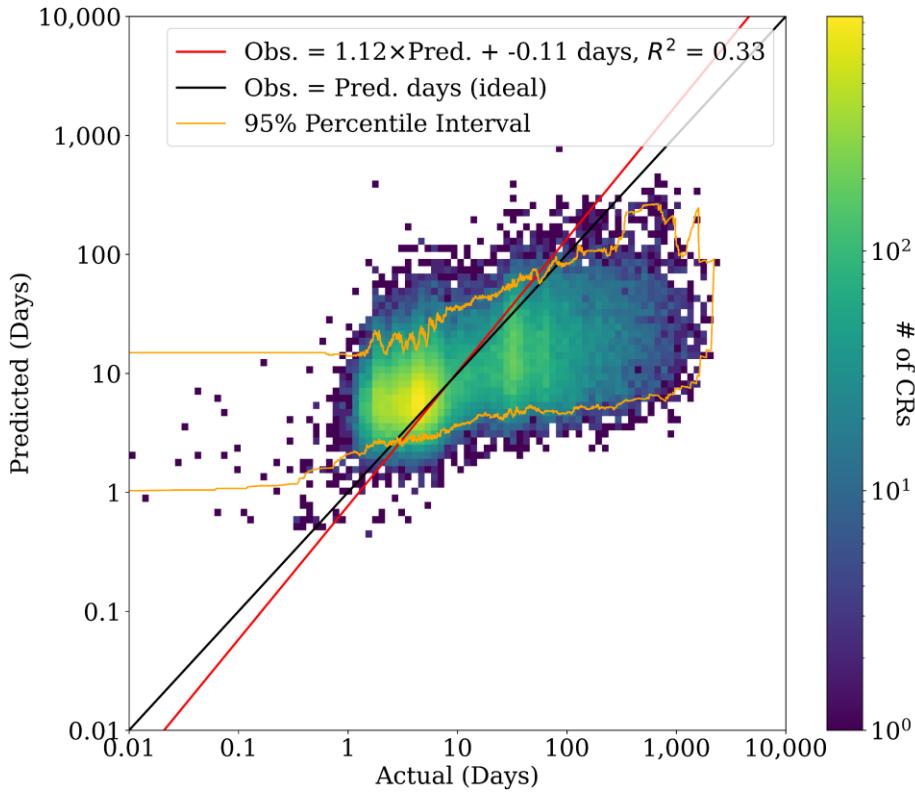


Figure 6. Actual vs. predicted completion time in days.^e

3.5.3 Technology Readiness

As demonstrated in this section, various methods can be applied for estimation. While the example used a simple linear model, other approaches, including nonlinear or more complex models, can also be utilized. Regardless of the model chosen, regression models are often sensitive to the text provided. Missing background information and additional variables that are not explicitly mentioned can complicate the development of effective regression models.

Summary of Technology Readiness: Needs a means to improve the regression performance for relatively short text or text with missing context.

3.6 Generation

Generation is a process in which an AI model creates new text that it has not explicitly encountered during its training. This process involves the model generating documents based on patterns learned from a broad set of documents used to train LLMs, as well as any additional specific documents it was later trained on. By combining knowledge presented to it in the form of language elements, the model creates new content that is contextually relevant, while remaining contextually relevant and natural.

^e The 95% percentile interval is shown in orange. The fitted line is given in red. The ideal line is shown in black. The intensity of the cells in the grid is based on a logarithmic scale. When there are no CRs at a point in a grid, the color of the cell is white.

3.6.1 Tools Overview

Most of the generative tools discussed in Section 2.3.3 can be used to create content, but innovating and generating new content that does not resemble the training dataset is generally associated with GANs. Because RAG is primarily used for summarizing a provided set of documents and autoregressive models are typically limited to the scope of their training data, both can be used to create new text, but with a focus on specific content. Encoder-decoder models can also generate new content, but their ability to do so effectively depends on them being trained to produce content in a specific manner.

3.6.2 Example Application

Generating new documents in the work management process could be associated with creating new procedures or instructions for work that have not been done before, reports on findings, or notifications that are machine generated to inform the human of an event. For this example, more than 1,600 work procedures provided by a utility were used. It has not yet been demonstrated that GANs can generate such documents, and this is considered an active area of research for text generation. Therefore, the example in this section instead used RAG, demonstrating its ability (or inability) to perform this function.

In this example, it was desired to generate text describing the risk of arc chuting during breakers inspections. Several out of the more than 1,600 procedures include those that reduce such risks, but there is no explicit mention of the risk. RAG was used with the embeddings focused on pages, since documents are too long to embed and could miss the context of specific parts of the procedure. Using word token embeddings produced a huge vector database with nearly 27 billion parameters, a computationally expensive model, but on the other hand it enabled a much finer level of control over how much focus to place on any individual excerpt of a document when trying to retrieve excerpts that match a user's question.

Using page embeddings, the retrieved excerpts were aggregated and presented to the RAG instruct model with the goal for the model being to produce a cohesive summary that captured the details and semantic intent of the excerpts. The experimental procedure relied on a relatively small embedding model (moka-ai/m3e-base; Wang, Sun, and He 2023) optimized for text similarity comparisons and a conversational instruct model (Mistral-7B LLM; Jiang et al. 2023) for generating the facsimile representations of the pages. The cosine similarity algorithm was then used to find the page with the highest overall cosine similarity to all the other pages in the vector database.

Prompt:

“What is the risk of arc chutes when conducting breakers inspections?”

Response^f:

“Based on the provided document, the risk of arc chutes during breaker inspections is related to the potential exposure to asbestos when handling and removing the arc chutes. Asbestos is a known health hazard and should be handled in accordance with established safety protocols and regulations. Therefore, personnel should follow the procedures outlined in the document to minimize the risk of exposure.”

The response presents two perspectives: the utility-specific one relates to asbestos, a material used to overcome arc chuting in breakers, and a generic one on why asbestos is dangerous. Considering the documents the model used to generate its response, the first perspective demonstrates that the LLM created new insight because this content wasn't explicitly mentioned in the utility's procedure documents.

Testing the model's ability to innovate, two tests were conducted. The first related to creating a procedure that has nothing to do with nuclear work—that is, to replace a vehicle's brake rotor.

^f A subset of the response is presented to preserve the data privacy of the utility.

Prompt:

“How to replace a vehicle brake rotor?”

Response:

“Based on the given question about replacing a vehicle brake rotor, I cannot directly answer that using the provided document as it does not contain any information related to vehicle brake systems or rotor replacement procedures. The document focuses on the replacement of RHR Pump Motors in a nuclear power plant. However, I can provide some general information about brake rotor replacement. The process of replacing a brake rotor involves the following steps:

1. Loosen the lug nuts on the wheel.
2. Remove the wheel.
3. Remove the brake caliper or pads.
4. Remove the old rotor.
5. Inspect the hub for damage.
6. Install the new rotor.
7. Reinstall the brake pads and caliper.
8. Torque the lug nuts to the manufacturer’s specifications.
- ... ”

This example demonstrates that RAG innovated but used the knowledge incorporated in the LLM to answer the questions, and it indicated that before answering. If RAG was pushed further to answer a fictional question like *“How to change the brake rotor of a spaceship?”* it would not be able to provide an answer, which would be the expected outcome.

3.6.3 Technology Readiness

RAG showed some potential for innovation but was constrained by the knowledge provided. Generating novel text using GANs is an active area of research. Generating a totally new document requires teaching the machine to think like a human and innovate beyond what it has learned. This ability has not yet been demonstrated with text.

Summary of Technology Readiness: Need to develop models that can innovate beyond the knowledge-specific training dataset.

3.7 Guidance

One of the key challenges with data input by humans is that it often lacks enough context because the staff assumes the reader would understand the context without the need to explicitly mention it. AI can be run on the fly to determine if the text inserted by a human is sufficient or proactively ask for more information if the user text is not clear.

3.7.1 Tools Overview

There are two means to accomplish guidance: the first is to run supervised models in real time while the text is being typed to generate classification and confidence. If the confidence is low, the user is asked to provide more information. The second approach is unsupervised and relies on models, such as autoregressive ones, to compare text against what the machine expected and ask for clarification when needed.

3.7.2 Example Application

Using the previously discussed CAQ/NCAQ-supervised classifier, an example could be created where users input text and the algorithm provides feedback on the classification, along with suggestions

to add more text or generate insights in real time. The iterative process is demonstrated in Table 5. Initially, the classifier finds the text as indicating an NCAQ event but requests additional information. Because a crucial field such as “equipment details” is missing, the model recommends adding this information to enhance confidence. After it is included, confidence in the NCAQ decreases significantly, suggesting safety implications, and the model leans toward a CAQ classification. The model continues requesting more text. As more text is added, the machine runs classifiers to decide if this text has an impact on the plant. As users provide more context, the AI model continues to reassess the text’s impact on the plant. This iterative process repeats until the AI model confidently determines the necessary decisions in the screening process.

Table 5. AI guidance of a user who is typing a CR.

Text	CAQ/CAQ Classification	Confidence (%)	Recommendation
Valve packing is leaking.	NCAQ	29.1	Add equipment/ location information
Valve packing is leaking for valve 45CDDF224, RCIC storage tank level instrument standpipe vent.	NCAQ	2.9	Add more detail
Valve packing is leaking for valve 45CDDF224, RCIC storage tank level instrument standpipe vent. The leak is slowly seeping at a rate of 3 drops per minute.	CAQ	7.0	Add information about impacts of the issue
Valve packing is leaking for valve 45CDDF224, RCIC storage tank level instrument standpipe vent. The leak is slowly seeping at a rate of 3 drops per minute. There is a potential for the leak to cause corrosion of the piping near the valve. The valve is causing a puddle of contaminated water on the floor.	CAQ	4.5	Add information about operational impacts of the issue
Valve packing is leaking for valve 45CDDF224, RCIC storage tank level instrument standpipe vent. The leak is slowly seeping at a rate of 3 drops per minute. There is a potential for the leak to cause corrosion of the piping near the valve. The valve is causing a puddle of contaminated water on the floor. There are no operational impacts for this condition, since there is adequate inventory of RCIC and there is no impact on the RCIC system pressure boundary.	CAQ	5.1	Stop

3.7.3 Technology Readiness

The supervised approach demonstrated in this example is usually achievable provided that computational power exists to run the text against the models on the fly. The unsupervised approach has not been demonstrated but is explored in Section 7, mainly for detecting junk text.

Summary of Technology Readiness: More research into unsupervised methods to guide users in inserting input is needed.

3.8 Insertion

Insertion is similar to the guidance approach except the user is presented with more text to describe the user input as the user is typing text. This is a more proactive approach; instead of just informing the user that the text is insufficient, the AI model helps the user describe the input and asks the user to mainly validate the generated input.

3.8.1 Tools Overview

Predictive text insertion can be used to predict an arbitrary number of words beyond the words in a user's input. Given that this is purely a generative model (i.e., new text is generated that relates to the provided text), generative AI models can be used for this purpose. Autoregressive models are the easiest to use, but they might not have the specific knowledge of the inserted text and thus end up hallucinating (i.e., generating text that does not relate to the provided text). RAG can be used if it knows the source of the information that the user text relates to. It can generate assistive text describing what the user inserted if the text is contained in the RAG's training data. GANs can be used if the input being used is of a new context that the model has not seen before.

3.8.2 Example Application

In this example, it was desired to demonstrate how AI can add words to the user input. The experiment involved single-, double-, and triple-word lookahead predictions. A sampling of nearly 600 NPP CRs was synthesized using. These synthetic CRs were then used to test the generative AI to see if it could correctly fill in the words that had been redacted from each report.

Several examples of text were used (Table 6). Mistral-7B was given guidance that it would be used as a text-completion assistant, and that its responses ought to be given in the context of the day-to-day goings-on at an NPP. By providing the model with this context, it was possible to coax its output so that it could supply word-insertion predictions with varying degrees of success depending on how many words forward it was trying to predict. Table 6 shows some of the texts the LLM was trying to predict. In this case, the generative AI was trying to predict three words that had been redacted from the end of the original report in the leftmost column. The center column includes the report text after the last three words had been removed. The rightmost column shows the results after the generative AI attempted to predict the removed words. The **green text** indicates words that were correctly predicted by the generative AI. The **red text** indicates words of the generative AI that differed from the original. It is important to note that just because some of the predicted words differed from the original does not necessarily mean that the prediction is not usable; it could be that the same context was generated. On the other hand, row 5 shows a case where there is a subtle difference in meaning between the intended words and the predicted words. “Backup Power Supply” system could be referring to something completely different than the predicted text’s “Backup Emergency Generator” system.

A metric was prepared to examine how successfully the generative AI predicted the redacted words. To calculate the metric, context-sensitive text embeddings had to be generated for the original text and the text-completion versions. The similarity between the predicted tokens versus the actual tokens was measured using cosine similarity. Just the end tokens were used in the similarity comparison and not the entire set of tokens for all the original text. By doing this, the context associated with the beginning tokens that were identical in the original and predicted text would not unnecessarily saturate the cosine similarities. After all, the desire was to focus on the similarities between the end tokens (target and predicted), which would automatically inherit the context of all beginning tokens anyway because of the context captured by the embedding model.

Histogram plots of the cosine similarity scores for each case of single-, double-, and triple-word predictions were prepared and are shown in Figure 7. The histogram plots were overlayed on the same plot to make it easy to compare performance. It should come as no surprise that the performance results

for single-word prediction significantly exceeded the results for double- and triple-word predictions, with nearly 23% of the CR examples having a perfect 100% similarity between the target and predicted words, and nearly 60% of the predictions having an accuracy of 70% or above. Triple-word prediction was much less successful, with most predictions scoring at 60% or less. Therefore, when the generative AI is coached with some domain-specific context, it is possible to get reasonable single-word lookahead predictions. However, looking much beyond single words is much less likely to produce usable results.

Table 6. AI insertion for a user while typing a CR.

Original Text	Redacted Text	With Predicted Text
A potential issue has been identified in the Feedwater System, where the expected water level is significantly lower than normal .	A potential issue has been identified in the Feedwater System, where the expected water level is significantly	A potential issue has been identified in the Feedwater System, where the expected water level is significantly lower than normal .
Unusual levels of radiation have been detected within the Containment Building .	Unusual levels of radiation have been detected within	Unusual levels of radiation have been detected within the Containment Building .
An unexpected and potentially critical issue has been identified in the Emergency Core Cooling System, with the primary pump not operating as intended .	An unexpected and potentially critical issue has been identified in the Emergency Core Cooling System, with the primary pump not	An unexpected and potentially critical issue has been identified in the Emergency Core Cooling System, with the primary pump not functioning as expected .
An unanticipated deviation from normal operating parameters has been observed in the Steam Turbine, necessitating a thorough investigation and potential corrective action .	An unanticipated deviation from normal operating parameters has been observed in the Steam Turbine, necessitating a thorough investigation and	An unanticipated deviation from normal operating parameters has been observed in the Steam Turbine, necessitating a thorough investigation and immediate corrective action .
An anomaly has been detected in the functioning of the Backup Power Supply system .	An anomaly has been detected in the functioning of the Backup	An anomaly has been detected in the functioning of the Backup Emergency Generator System .
An anomaly has been detected in the Main Steam System, with pressure levels deviating from normal operational standards .	An anomaly has been detected in the Main Steam System, with pressure levels deviating from	An anomaly has been detected in the Main Steam System, with pressure levels deviating from normal expected values .

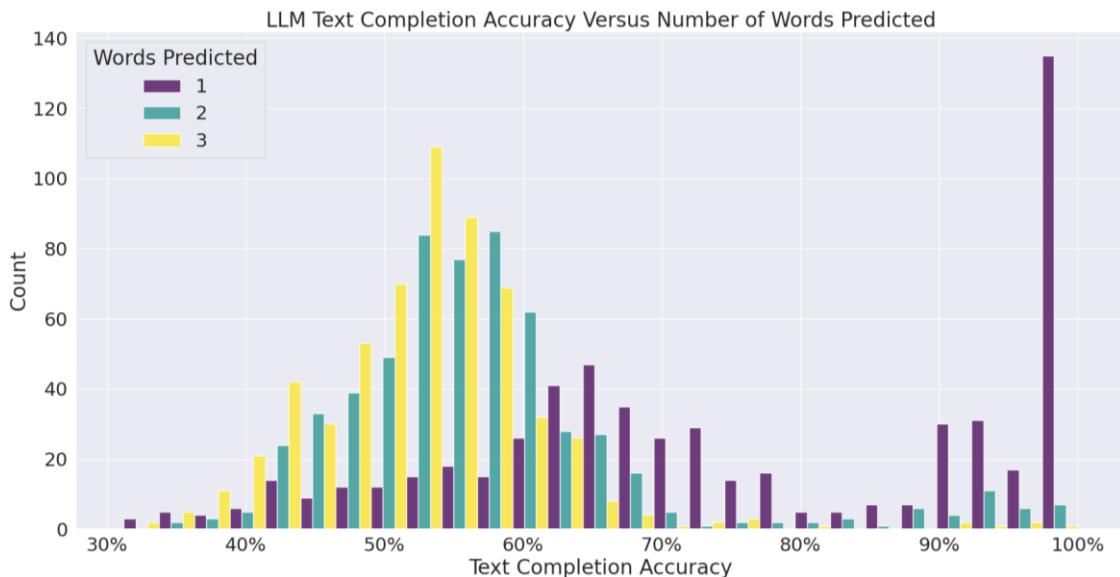


Figure 7. Insertion accuracy for one, two, and three words.

3.8.3 Technology Readiness

Autoregressive models have been widely used in chat engines and have shown their ability to provide more context, but they suffer from hallucinating, especially when the specific knowledge is not part of their training. Model refinement can be performed to train a model on a specific field, but its performance has yet to be demonstrated in this research. However, hallucination is less of an issue for short text prediction. The use of RAG for knowledge-specific uses cases or GAN for unseen text is yet to be demonstrated.

Summary of Technology Readiness: Ready if an autoregressive model is used for short text or specific knowledge is not needed. The use of autoregressive, RAG or GAN methods for knowledge-specific applications is not demonstrated.

3.9 Review

Review is the process of comparing some text against a criterion to indicate if compliance or misalignment is found and presented in a self-explanatory format. For example, some text might need to be tested against a procedure to determine what actions need to be taken. Review is often used when there are several decisions to be made and creating a single model to make all of them is not feasible.

3.9.1 Tools Overview

The two most promising tools for review are autoregressive models, because they incorporate the general ability to understand contextual text and can be fed explicit instructions from the reference procedure or guidance, and RAG, since it operates in the same manner but with a focus on specific documents. Encoder-decoders can be used if enough examples can be used to train the models.

3.9.2 Example Application

For this example, the decision to mark failures in log entries, discussed in Section 3.2.2, is used again. In this section's example, it was desired for an AI model to make the decision after reviewing guidance on how to make the decision. An autoregressive model was used and instructed to use the help of NEI-99-02, herein referred to as the "guideline." One of the purposes of the guideline is to show how to identify the failures of monitored equipment that are directly used for the computation of reliability. The information of the guideline was needed because the autoregressive model did not know where the

boundaries of the equipment were for determining if a failure occurred on a specific piece of equipment. Nor was it able to understand the definition of failure for specific pieces of equipment.

The definitions from pages F-14 to F-16 of the guideline were used to define whether a failure occurred for eight different types of equipment and failure combinations. Also, the equipment boundaries per page G-33 of the guideline were provided. The following are some of the examples of equipment and failure combinations:

Sample Failure Definition: “*Pump failure to run - Given that it has successfully started and run for an hour, a failure of a pump to run/operate. Exclude post-maintenance tests, unless the cause of failure was independent of the maintenance performed. Include all failures that result from a non-PMT demand following return to service. If a PMT failure occurs following return to service and was caused by the maintenance activity, then this failure is excluded and the train, during the period from the completion of the maintenance activity to the declaration of return to service, is counted as unavailable.*”

Sample Component Boundary: “*Solenoid Operated Valves - The valve boundary includes the valve body, the operator, the supply breaker (both power and control) or fuse and its associated control circuit (relay contacts for normally auto-actuated components, control board switches for normally operator actuated components).*”

This information was provided to Microsoft’s Phi-3 mini-LLM (Abdin, M., et al., 2024) because the text contained less than 8,000 tokens (around 5,000 words). The same 100 text entries from the NPP operator logs discussed in Section 3.2.2 were used. The autoregressive model was used to predict if this SME-labeled data concerns an equipment failure by answering yes or no only. The F1 score of the failure classification is given in Table 7.

These two sources of information did not seem to improve the results. Instead, they degraded the model performance. The several pages of definitions and boundaries created a complex set of instructions that a human can follow with significant effort; however, it is hypothesized that this is beyond generative AI’s capabilities at the time of this writing. It is also possible that not all the necessary procedures, guidelines, etc. were provided for generative AI to be able to classify an entry correctly.

Table 7. Effect of the inclusion of reference material on F1 score.

Reference Material Included	F1 Score (%)
SME-Created Prompt	76.27
SME-Created Prompt + NEI 99-02 Rev. 8 Component Boundaries	74.80
SME-Created Prompt + NEI 99-02 Rev. 8 Failure Definitions	75.55
SME-Created Prompt + NEI 99-02 Rev. 8 Component Boundaries + Failure Definitions	75.68

3.9.3 Technology Readiness

From the demonstration, it was concluded that autoregressive models may not yet be able to understand complex instructions, such as those from references and standards, at the same level as humans. RAG has not been demonstrated for review. Encoder-decoders have not been demonstrated for a specific type of decision.

Summary of Technology Readiness: Need to determine the means to improve an autoregressive model’s ability to understand and follow instruction or guidance, and to demonstrate RAG or encoder-decoder use for review.

3.10 Structuring

The vast majority of data in the nuclear industry is in an unstructured format (i.e., documents, tables, figures). Mining these data sources to acquire data and structuring the data in a format so that data from one source can be combined with that of a different utility, or even so that data from multiple utilities can be combined, is a challenging human effort. AI could resolve this challenge.

3.10.1 Tools Overview

Both supervised and unsupervised models can be used for this purpose. Supervised models can be used if the text follows a somewhat similar structure. For example, tables can be extracted by training a machine to recognize table entries (such as the tags of equipment) so it can mine documents that relate to that entry. However, since most data in the nuclear industry is nonstructured (follows various patterns) within the utility and across utilities, unsupervised models have much more potential. Methods like topic modeling can provide broader information about the theme of the information to be used for mining. Similarity analysis methods can perform a similar function if the data in the unstructured documents have a certain structure that is sought. Generative AI models are considered the most promising ones as they can understand unclear patterns. For example, if safety precaution data related to a certain type of work are required, RAG can be used to extract that specific information if it is readily available but hard to find in a large set of documents. GANs can be used if examples of the needed data are present to teach the AI to find or even generate similar data from a large set of structured data. Autoregressive models can create data in a structure that resembles what the machine has seen before.

3.10.2 Example Application

An experiment was conducted to identify whether a collection of more than 1,600 work procedure documents include a fire protection aspect (i.e., a fire hazard is relevant). This could enable a new fire protection procedure to be developed if it resembles the provided procedures.

An autoregressive method using the Mistral-7B LLM was used. This model is limited to 4,096 tokens, which is about two pages of text, so procedures that are more than two pages long were not used (i.e., the full procedures were not used). The hypothesis was that the LLM should be able to establish the intent and meaning of the document by examining the first two to three pages. This seemed reasonable since most procedure documents are quite explicit about the purpose of the document early on, and hazards are usually presented in the first part of a procedure document due to their importance.

Because the LLM relies on carefully constructed textual prompts to guide its responses, several experiments were conducted with different prompt variants before deciding on one that resulted in reasonable performance. The LLM was asked to decide whether the procedure pages had anything to do with fire protection programs or fire protection equipment or could describe a fire hazard.

Since autoregressive models are used in an unsupervised manner, another method (topic modeling using an advanced form of CorEx) was used to benchmark the results. The CorEx model used the keywords and phrases listed in Table 8 to describe the fire hazard involvement in the procedure.

The results, presented in Table 9, show great agreement between the two methods in identifying fire-related procedures, but several disagreements were noticed too. In addition to a yes/no response, the LLM was asked to provide two to three sentences to explain its decision. In some cases, the reasons it provided were fascinating. For instance, in one case that the LLM identified a procedure as fire related, but CorEx did not, the LLM admitted that the document did not directly describe details about fire protection, but that its contents could be construed as describing fire-related procedures based on context. The document in question described excavation and trenching activities in the context of ensuring that natural gas lines are avoided, etc., an obvious potential cause of fire if sufficient caution is not taken, and yet the topic model could not possibly have detected this due to the absence of relevant terminology.

Table 8. Keywords and phrases used by the CorEx to describe fire hazards.

combustible	fire panel	hot work
detector	fire protection	hydrant
fire	fire protection program	impairment
fire brigade	fire pump	national fire protection association
fire detection	fire watch	smoke detector
fire detector	fire zone	sprinkler
fire drill	halon	sprinkler system
fire impairment	hose	zone
fire marshal	hose station	

Table 9. Confusion matrix for the LLM and CorEx decisions on whether text relates to fire.

		LLM Predicted	
		Fire	Not Fire
CorEx Prediction	Fire	27	30
	Not Fire	25	1,573

3.10.3 Technology Readiness

While the example application demonstrates one approach to structuring data, it represents a small portion of the types of unstructured data that could exist. This function is in early stages of development and requires additional research into what types of unstructured data exist and the means to structure each. For example, asking an AI model to analyze a figure or table to provide knowledge or perform a complex task was not demonstrated in this effort and remains an area for future research.

Summary of Technology Readiness: Need more research into the various formats of unstructured data and the means to structure each.

3.11 Summarization

Summarization is often needed to mine a large set of documents to determine some key information that a user needs to present. It saves users the time they would need to search through a large set of documents; for example, technical specification manuals could be mined to address a specific question about an action in the work process. Summarization can also be used to present to users the key information of a document before they read it, or to conduct a literature review on a topic.

3.11.1 Tools Overview

For summarization, models must be able to ingest the new documents they are being asked to summarize. Autoregressive models can be used in two ways: if the document to be summarized is small, it can be directly fed to the model, and if it is large, the model can be retrained with the newly added documents to make them part of the model's knowledge. If an autoregressive model is not provided the full extent of documents it has been tasked to summarize, problems can present themselves in the form of hallucination, where the model might give suboptimal summaries of documents that contain extrapolations of context or semantics that confuse rather than assist understanding. Instead, RAG can be used because it can summarize the documents into a size that can be fed into an autoregressive model. RAG can therefore be provided given domain-specific data to review. Encoders-decoders can be used if the summarization process is specific and examples can be used to train the model to perform the process in a specific manner.

3.11.2 Example Application

The same experiment conducted in Section 3.6.2 for generation was used for summarization to demonstrate how the RAG process can be used to reliably produce summaries of pages sourced from nuclear industry documents. The hypothesis was that if a page of text shared much commonality with many other pages of text in a set of pages known to RAG, then RAG ought to be able to generate a close facsimile of that page using just the context of the similar pages. In this example, the procedure to perform breaker inspection was targeted again with a focus on arc chuting. The same prompt that was used in the previous experiment was used again, but one of the answers resulted in a summary of the procedure:

Prompt:

“What is the risk of arc chutes when conducting breakers inspections?”

Response^g:

“During the inspection and maintenance of arc chutes, personnel should follow the outlined procedure carefully to minimize the risk of arc flash incidents. The procedure advises caution to keep the arc chutes dry and to tilt them back carefully to avoid causing the breaker to tip over. The arc chutes should be cleaned using a HEPA filtered vacuum cleaner and a mild solvent or water if necessary. Inspecting the arc chutes for broken or cracked ceramic parts and erosion of ceramics is also important to ensure their continued safe operation.”

This summary represents several pages of instructions summarized well in a few sentences. An autoregressive model (Mistral-7B LLM) was used to generate a response that was compared to the RAG response to evaluate its performance.

Prompt:

“Provide a procedure for protection of arc chutes when conducting breakers inspections in one paragraph.”

Response:

“To protect arc chutes during circuit breaker inspections, begin by thoroughly reviewing the breaker’s service manual and ensuring all personnel are trained and certified in electrical safety procedures. De-energize the circuit and implement lockout/tagout (LOTO) procedures to prevent accidental re-energization. Equip inspectors with appropriate personal protective equipment (PPE) such as insulated gloves and safety glasses. Conduct a visual inspection of the arc chute for damage or contamination, clean it carefully with a dry cloth or vacuum, and check its alignment. Perform insulation resistance and operational testing according to manufacturer guidelines. Document all findings and report any significant issues for further action. After reassembling the circuit breaker and removing LOTO devices, restore power and conduct a final operational test. Review safety practices post-inspection and schedule any necessary follow-up maintenance.”

This response demonstrates that autoregressive models provide generic instructions, given they are not trained on breaker inspections.

3.11.3 Technology Readiness

RAG is well suited for creating summaries due to its ability to summarize large documents without the need to retrain LLMs, unlike autoregressive models. For smaller references, any of those methods can be used, but larger documents require either RAG or the retraining of autoregressive models. It has not been demonstrated that encoder-decoder models are effective for specific summarization tasks.

^g A subset of the response is presented to preserve the data privacy of the utility.

Summary of Technology Readiness: Ready using RAG but not demonstrated using autoregressive models. Targeted summarization using encoder-decoders has yet to be demonstrated too.

3.12 Visualization

Visualization is integral to text analysis, especially when managing extensive and complex corpora of unstructured data. The volume and intricacy of text data can be daunting, but visualization techniques transform this information into more comprehensible and actionable insights. By converting large text datasets into graphical formats, these techniques allow analysts to discern patterns, trends, and relationships that might elude conventional analytical methods.

3.12.1 Tools Overview

Several visualization tools, including heatmaps, word clouds, and network diagrams, can reveal the relationships between different terms or concepts, aiding in the comprehension of large data sets and enhancing transparency in the analytical process. However, these tools depend on underlying data analysis techniques used such as dimensionality reduction, clustering algorithms, and topic modeling to effectively present the data.

3.12.2 Example Application

In this example, topic modeling was used to classify text by specific topics in an unsupervised manner. The result needed to be visual to determine how close the topic clusters were to each other. For this purpose, UMAP was used. An example is presented in Figure 8. The figure shows the UMAP reduction of the topic-modeling score data for the top 30,000+ most informative words out of a vocabulary of 100,000 words total. Each point in the scatter plot represents a single word, and each word is clustered with its topic.^h Broader clusters seem to indicate that the information gain of the topic words experiences greater variance and thus offers a broader capacity for describing a greater number of documents. The orientation of the wispsiness of the clusters could also be meaningful. Though purely hypothetical, it may be that topics in proximity and with a similar orientation use words that have similar conceptual meanings in similar contexts. In essence, these topics have a positive correlation, whereas topics that are in proximity and have an orthogonal orientation use words with similar meanings to describe very different ideas.

3.12.3 Technology Readiness

While this section provides a single example for demonstration, numerous proven visualization tools have been employed across various research studies. Visualization typically involves methods to reduce data dimensionality while maintaining a focus on the process of interest. For instance, reducing data to two dimensions for visualization can reveal different clusters, which may vary depending on the dimensionality reduction techniques used.

Summary of Technology Readiness: Ready but more research is needed to determine the optimal means to transform the data for visualization.

^h Each topic cluster has a numeric suffix appended to its title to indicate whether multiple clusters exist.

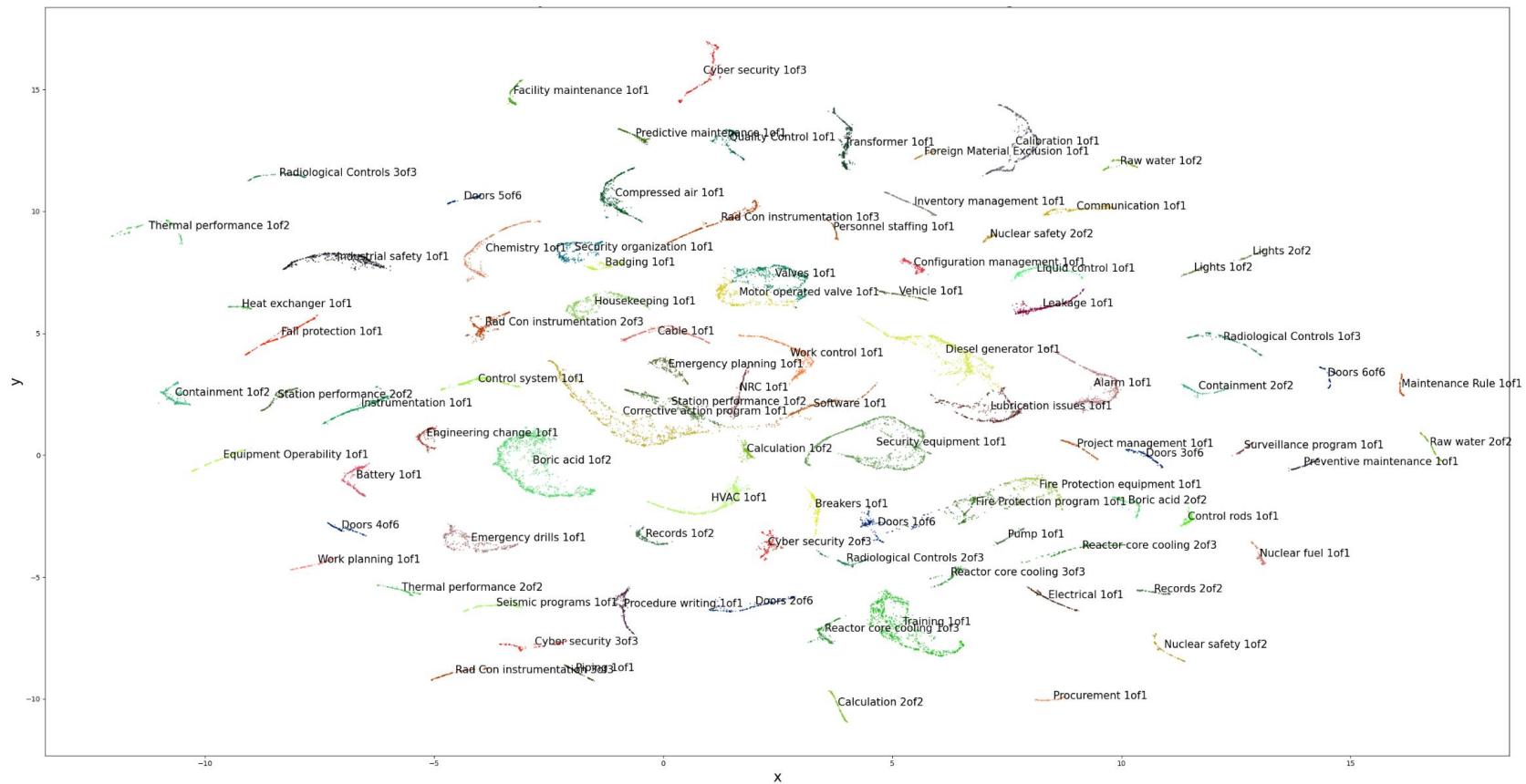


Figure 8. UMAP dimensionality reduction for topics generated by a CorEx model.

4. MAPPING THE WORK PROCESS TO AI FUNCTIONS

An overview of the work management process was presented in Figure 2. This section provides a detailed explanation of the phases shown in the figure, the approach used to automate the steps in each phase, and the cost savings achieved through automation in each phase and by each AI function. Several conclusions are also made from the mapping of the work management process and the assignment of AI functions.

4.1 Work Mapping Approach

Each step in the work processes detailed in Appendices A-F commences with an introduction for context to describe what the step entails and why it is performed. Each step is categorized into a decision, an action, or both. A decision is formulated into a clear question while an action is formed as a command. Depending on the decision outcome, each decision is broken into two possible actions.

The step is broken into two subsections: the first focuses on how it is currently done by the staff and titled “Current Decisions and Conditional Actions” if a decision is made or “Current Unconditional Actions” if an action is taken. The second focuses on how it can be automated by AI and is titled “Automated Decisions and Conditional Actions” if a decision is made or “Automated Unconditional Actions” if an action is only taken.

In considering the human process first, the specific steps in making the decision or taking the action are explained. The time staff takes is quantified in minutes full-time equivalent (FTE), and the organization taking the action is listed. This information is presented in a table format.

The frequency of each decision or action is listed. The frequency is normalized against a CR, WO, WP, or ST depending on where in the process the step is taken. This variation is because the phases target different outcomes. However, the outcomes can be scaled to each other as discussed later in Section 5.

The frequency depends on the previous step in the flow. For example, if a decision outcome is to perform the next step and this happens 50% of the time the decision is made, the frequency of the next step is 50% of the frequency of the current step. As steps are incremented, the frequency tends to drop to the point where some decisions or actions occur very rarely. Those rare decisions or actions are not listed herein.

Using the frequency and time needed, it is possible to estimate how much time each of the decisions and actions take per CR, WO, WP, or ST. These estimates are listed at the bottom of the table. Those numbers might look insignificant; however, considering the thousands of occurrences in the plant, the numbers can scale up a significant amount.

An example of the decision and condition actions table is shown in Table 10. In this example, the first step of the screening process, “Issue Identification and Reporting,” is used. The objective of this step is to determine whether the condition has been reported before so that it is not logged again. In the manual approach, if the condition relates to equipment, the staff inspects the equipment or searches through existing CRs to determine if the observation is already in the screening system. If the condition is already in the system, the staff may add comments if they believe additional insight could help. If the condition is not found, staff writes a new CR. This process can be performed by any member of plant staff. It takes 3 minutes to make the decision about whether the condition is already there and 5 minutes to log a new one if one does not exist.

The decisions and actions are also reflected in the automated approach. Each decision and action is assigned an AI function from Section 3 to indicate methods that can achieve the desired function to replace the human process. AI functions were selected for steps based on a systematic process and SME experience in using AI for process automation. While some subjectivity is involved in selecting AI functions for work steps, the overall trends and conclusions of this report are valid and not based on

individual AI function assignment decisions. For each function, the data source used for training the AI is listed, if applicable. Additionally, the input data, the model used, and the AI output are listed. Using the same example, Table 11 shows that three options exist to replace the human decision with the machine decision. In the first one, the previous CRs can be compared to the one being logged. Therefore, the function of the AI is to perform a comparison. The method to achieve this could be, for example, similarity analysis as discussed in Section 3.3. If the AI finds a similar CR, it is flagged as already logged and can possibly even provide the similar text. If the CR was found, the AI can elaborate on the event provided by a user if any text is provided. If the condition is not found in the system by the AI, it can assist the user by inserting new text and providing guidance as soon as the user starts typing. This reduces the 5 minutes required by the user significantly.

A similar process for mapping each phase step (except initiation since it is considered out of the work management scope) is provided in Appendices A-F.

Table 10. Example of the current decisions and conditional actions table.

Decision	Effort Description	Discipline	Time in Minutes	Frequency Per CR
Process	Check if the condition was previously identified. e.g., there is a tag on the equipment or is in the condition reporting system	Any	3	150%
If yes	Add comments to the existing condition if needed. (End of Process)	Any	2	50%
If no	Write a CR.	Any	5	100%
Total (minutes FTE per CR)				10.5

Table 11. Example of the automated decisions and conditional actions table.

Decision	AI Description			
	Training/Reference Input	Use Input	Functions	Output
Process	Historical CRs	Identified and reported condition text	Comparison	Yes/No (and previous CR)
If yes	Previous CR	Identified and reported condition text	Generation	Modified CR text
If no	Historical CRs	Identified and reported condition text	Guidance	New CR text
	Historical CRs	Identified and reported condition text	Insertion	New CR text

4.2 Assumptions

The analysis presented in Appendices A-F relies on several key assumptions due to the variability in processes across different nuclear power plants. This section summarizes these key assumptions:

Variation in Estimates: Time estimates and step frequency estimates can vary by utility and plant. The used estimates have been generated based on the best available information from industry experts. The organization of decision-making may also differ. While there is some uncertainty in individual estimates, on average the overall trends and conclusions are an accurate representation of the work management process.

Precision of Estimates: Estimates are rounded to the nearest minute in most cases. Fractions of a minute are included for some steps where warranted.

Sequence of Steps: The sequence of steps in the appendices may differ from actual processes. However, this does not impact the report's conclusions, as cost savings are not dependent on the order of steps.

Repetition of Steps: Very often a step could result in previous steps being repeated, such as when work needs to be rescoped. The estimates presented in this report accounted for those repeated steps, so they are not accounted for separately.

Depth of Mapping: The steps outlined in the appendices may not encompass every step performed. Achieving complete mapping fidelity would require excessive effort. The report focuses on identifying critical AI functions, and mapping most steps is sufficient for the report's objectives.

Scope of Work Management: The focus of this work is on work management. Tangential tasks that support work management but fall outside of it are not mapped. For instance, if operations needs to configure the plant to perform the work, this configuration is not considered part of the work management process.

Outlier Scenarios: Outlier scenarios, which are rare or exceptional cases, are not considered in the steps decisions and actions. The analysis focuses on typical scenarios to maintain practicality and relevance.

Online vs Outage: This work focuses on online work management in a plant, given it represents most of the work performed in NPPs. Outage work management follows a similar approach, and future work could expand to consider the similarities and differences for outage work.

Maximum Automation Credit: For some steps, a tool is used to expedite the human review process without fully automating it. For this work, the exact time saved through automation for these partially automated steps was not quantified and thus each step was assumed to be fully automated for purposes of time savings calculation. As a result, the estimates provided represent the maximum possible cost savings. The only exception is for guidance and insertion, where it is assumed that a human will still spend half the time performing the work if these functions are implemented.

Data Sources: Decisions and estimates are based on the current format of data sources. Future technological advancements or process changes may introduce new labels and alter tools and methods, potentially affecting decision-making.

Crediting Functions: AI Functions can be combined or used independently for a given step. In the cost savings estimates, it was assumed that each function could be used separately, meaning each was credited independently regardless of what other solutions exist.

Subjective Functions: The two AI functions that were discussed in this report and not mapped to work management decisions and actions are dialogue and visualization. These functions are highly subjective

and vary based on human needs. Some utilities may require staff to interact with AI and visualize data in specific ways, while others may need less interaction and visualization. Consequently, these functions were excluded from this mapping effort.

4.3 Initiation

The phases of work in a nuclear power plant starts with three types of initiators: (1) a condition occurring in the plant, (2) modifications or upgrades sponsored by the projects team, and (3) periodic plant PM activities performed by maintenance. A condition is any off-normal event or configuration associated with an event in the plant that requires the plant staff's attention. There is a variety of ways that an event can be identified, including through staff performing their daily activities, operator rounds, walkthroughs, equipment trending and monitoring via various instrumentation and monitoring methods, or through a self-revealing performance issue (e.g., the equipment stops performing its intended function). A large portion of the events observed relates to SSC or specific equipment. When an event is identified, it is sent to the screening committee that analyzes conditions. This committee decides what priority, impact, or action should be taken (see Section 4.4). A CR (also called action report or issue report) is generated in the process. The outcome of this effort could be a WR, followed by the creation of a WO.

Another initiator of work relates to projects, which are usually associated with modifications to the plant or upgrades. Those stem from long-term plans driven by strategic objectives to sustain the plant operations. Each site has a list of projects that are desired to be completed based on priority needs, such as regulatory commitments, plant improvements, or other corrective actions. As these projects become ready for implementation, WOs are created, and the projects are turned over to work management to identify the best work window for ensuring the project's successful implementation. In most cases, one or more plant modifications are performed during the execution of the project. Prior to this point in a project's life cycle, many hours have been spent—mostly with engineering and projects resources (with other organizations providing guidance and input as needed)—to develop the project to the point that it is ready for scoping. For the purposes of this report, the scope of the project development is beyond this report scope, but a project is considered part of the work management life cycle when it arrives at the scoping phase. One key aspect of this initiator is that it does not go through the events or conditions screening process.

The third initiator of work is PM. It typically consists of a model WO, a defined frequency for performing that model WO, and a due date by when the next PM is to be performed. Every NPP has a scheduling tool that takes defined PM frequencies and due dates and generates or prompts PM WOs in the schedule.

PMs could also go through the screening process if a change is needed. Sometimes an equipment condition is the result of an issue in the PM strategy. Thus, when an equipment condition is identified, an issue may also be identified with the PM strategy. The PM change process is typically completed by the system engineer responsible for the equipment. An issue with an existing PM task could include a frequency issue (performing the PM too often or not often enough) or a scope issue (work performed does too much, not enough, or the wrong tasks).

Some screened work becomes a fix-it-now orders. This process is usually started by maintenance staff and involve routine and low-risk activities that do not need the level of scrutiny that more significant work must go through. Examples of such work could be tightening a nut for a flange. This type of work skips the work process and gets directly scheduled as there is no need for a WP with detailed procedures to be created.

The following sections discuss each of the following phases in detail with focus on how AI could automate the process.

4.4 Screening

Screening in an NPP involves a systematic analysis of events to determine their characteristics, necessary actions, and the path forward. This process begins with the identification and reporting of issues and determining relevance to equipment, followed by evaluating the event criticality and plant impact. Maintenance, operations, engineering, and other relevant organization validations are conducted to ensure that identified issues are addressed effectively. Compensatory actions are established to mitigate risks until permanent solutions can be implemented. The significance of the issue to each organization is determined.

Next, an investigation is conducted to determine if this is issue is considered a rework. For significant work, resources are estimated to help decide whether the work can be done in-house or if a contract is needed. Modifications or field changes are identified. The screening process then analyzes safety significance and impacts related to administrative, radiological, and chemical factors. This is followed by categorizing the severity of the issue, which enables prioritization of resources. The necessary actions for resolving the issues are determined.

Finally, meetings, including prescreeningⁱ and screening sessions, facilitate thorough discussions of the identified issues and the necessary corrective actions. During the screening meetings, relevant stakeholders collaborate to evaluate and finalize decisions on actions needed to address issues effectively. After these meetings, updates on corrective actions and WRs are documented, and maintenance WOs are created when applicable for the next stage of work management.

As discussed in Section 4.1, each step is mapped to specific activities performed by dedicated staff, and AI functions are identified to replace those tasks. The detailed steps and considerations involved in the process are presented in Appendix A. Table 12 summarizes the findings from the analysis of the process in Appendix A and is used to quantify the potential cost savings of the entire process in Section 5 and each AI function in Section 6.

4.5 Scoping

Scoping is a crucial process that involves characterizing the WO by assessing attributes such as priority and criticality and determining the potential operational impact of the planned work to assist in scheduling the work week. This phase ensures that all aspects of the work scope are understood and managed effectively. Specific steps within this phase include merging new WOs into existing ones when applicable and establishing priorities and criticality levels for the tasks at hand. Additionally, the determination is made on whether work should occur during an outage or while the unit is online, along with preliminary scheduling to optimize resource allocation and timelines.

If the work is relevant to ongoing projects, several decisions and actions are taken. Project sequencing and readiness are evaluated, and engineering evaluations are conducted to prepare proper drawings and documentation. At this phase, it is determined whether contractor services are needed for projects. The scoping process creates associated WO activities from each WO, typically resulting in multiple work activities per WO. Resource estimation is performed, and availability is confirmed to facilitate allocation during planning. If support is planned in collaboration with other plants, considerations for plant, fleet, and corporate schedules are also integrated. Additional assessments are then conducted, including considering technical specification requirements to ensure compliance, evaluating whether unit down power is necessary, and analyzing the impacts on plant chemistry.

ⁱ Prescreening meetings are not consistently part of the process in the nuclear industry.

Table 12. Estimated potential time saved in minutes per CR for each screening step by each AI function.

Task Time in Minutes	Classification (Supervised)	Classification (Unsupervised)	Comparison	Estimation	Generation	Guidance	Insertion	Review	Structuring	Summarization
Issue Identification and Reporting	—	—	4.5	—	1	2.5	2.5	—	—	—
Equipment Relevance	0.5	0.5	0.5	—	—	—	—	—	—	—
Criticality	1.5	—	1.5	—	—	—	—	—	—	—
Operability and Impact to Plant	3	—	—	—	—	—	—	3	—	—
Maintenance Validation	—	—	3	—	—	0.075	0.075	—	—	3
Maintenance Compensatory Actions	—	—	—	—	—	—	—	—	—	—
Maintenance Significance	0.75	—	—	—	—	—	—	0.75	—	—
Operations Validation	—	—	0.75	—	—	0.05	0.05	0.75	—	0.75
Operations Compensatory Actions	—	—	—	—	—	—	—	—	—	—
Operations Significance	0.75	—	—	—	—	—	—	0.75	—	—
Engineering Validation	—	—	0.75	—	—	0.025	0.025	—	—	0.75
Engineering Compensatory Actions	—	—	—	—	—	—	—	—	—	—
Engineering Significance	0.75	—	—	—	—	—	—	0.75	—	—
All Others Validation	—	—	0.75	—	—	0.025	0.025	—	—	0.75
Rework	—	—	2.25	—	—	—	—	—	—	—
Resources for Significant Work	—	—	2.25	—	—	—	—	—	2.25	—
Project/In-house for Significant Work	—	—	0.75	—	—	0.005	0.005	—	0.75	—
Modification/Field Change	0.03	—	0.78	—	—	—	—	0.75	0.75	—
Safety Significance	1.5	—	1.5	—	—	—	—	—	—	—
Administrative Impact	1.25	1.25	—	—	—	—	—	1.25	—	—
Radiological Impact	—	2	—	—	—	—	—	2	—	—

Task Time in Minutes												
		Classification (Supervised)		Classification (Unsupervised)			Estimation		Generation		Guidance	
		5		—			—		—		—	
Chemistry Impact		—		—			—		—		—	
Severity	3	—	—	—	—		—	0.05	0.05	3	—	—
Actions Needed	2	—	—	—	—		—	—	—	2	—	—
Prescreening Meeting	—	—	—	—	—		—	—	—	—	—	—
Screening Meeting Preparation	—	—	—	—	—		—	—	—	—	—	—
Screening Meeting	—	—	—	—	—		—	—	—	—	—	—
Condition Report Update and Work Request	—	—	—	—	—		—	—	—	2	—	—
Maintenance Work Order Creation	0.45	—	0.45	—	—		—	—	—	—	—	—

Other considerations such as reactivity effects, coolant system impacts, redundant systems, and risks associated with transients and conflicts with surveillance schedules are assessed to maintain plant integrity. The scoping process also addresses other operational factors, including seasonal impacts, grid instability, and containment integrity. Furthermore, cycle timing evaluations are conducted to ensure alignment with operational schedules. Factors such as relevant work to plant security are also considered. Maintenance rules are verified to uphold safety and compliance. Risk screening is performed. At this point, work management conducts a review to ensure that the work scope is clear and concise. Validation of priority assignments are performed to ensure that all maintenance and operational activities effectively align with the overall objectives of the plant. Work is then ranked by priority to ensure that the most critical tasks are addressed first, and PM grace periods are evaluated to allow flexibility in scheduling should conflicts arise.

As discussed in Section 4.1, each step was mapped to specific activities performed by dedicated staff, and AI functions were identified to replace those tasks. The detailed steps and considerations involved in the process are presented in Appendix B. Table 13 summarizes the findings from the analysis of the process in Appendix B and is used to quantify the potential cost savings of the entire process in Section 5 and each AI function in Section 6.

4.6 Planning

Planning consists of a series of structured steps designed to ensure that all necessary elements are thoroughly addressed in a WP before work execution. Initially, it is determined whether a WP is required. If yes, a planner identifies previous similar WPs to help leverage past experiences and knowledge, in parallel with reserving long lead-time parts. If similar previous WPs are found, they may be updated or a new WP may be created, depending on the situation. If a new WP is created, a thorough review of similar WPs for relevant information is conducted, along with a review of technical information. This helps ensure all necessary details are available for effective planning.

Next, clear and concise work instructions are crafted. This is an essential step to guiding workers during the execution phase. Any relevant operating experiences are incorporated into the WP to make sure mistakes are not repeated. Worker feedback from previous similar work is examined to identify potential areas of improvement or concerns that should be addressed, and human performance (HP) tools are integrated to enhance execution. A clearance order (CO) is requested, if applicable. An evaluation is needed for engineering and operations documentation to confirm that all relevant engineering changes (ECs) and technical specifications are covered. Administrative or technical procedures are also considered.

Steps are taken to ensure that fire protection and industrial safety controls and permits are in place. A WP walkdown is performed to assess the work area and confirm that all preparations are adequate. Radiological controls are also reviewed, along with the ALARA (as low as reasonably achievable) plan, to minimize exposure to workers. The radiological work impact on workers is evaluated to ensure their safety throughout the process. Additionally, worker training and qualification are assessed to ensure that all personnel are adequately prepared. Resources are then assigned to the WP, followed by reviewing equipment qualification requirements. Finally, materials and parts are reserved, and their request is completed before the WP is reviewed and approved.

As discussed in Section 4.1, each step was mapped to specific activities performed by dedicated staff, and AI functions were identified to replace those tasks. The detailed steps and considerations involved in the process are presented in Appendix C. Table 14 summarizes the findings from the analysis of the process in Appendix C and is used to quantify the potential cost savings of the entire process in Section 5 and each AI function in Section 6.

Table 13. Estimated potential time saved in minutes per WO for each scoping step by each AI function.

Task Time in Minutes	Classification (Supervised)	Classification (Unsupervised)	Comparison	Estimation	Generation	Guidance	Insertion	Review	Structuring	Summarization
Merger of Work Orders for Same Issue	—	—	5	—	0.25	—	—	—	—	—
Work Order Priority	2.85	—	—	—	—	—	—	2.85	—	—
Work Order Criticality	2.85	—	2.85	—	—	—	—	2.85	—	—
Outage or Online Work	1.9	—	—	2.5	—	—	—	1.9	—	—
Preliminary Work Schedule	—	—	—	2.5	—	—	—	—	—	—
Project Relevance	—	—	1	—	—	—	—	—	—	—
Project Sequence	—	—	0.6	—	—	—	—	0.6	—	—
Project Week(s) Readiness	—	—	1.2	—	—	—	—	—	—	—
Project Engineering Evaluation Review	—	0.6	—	—	—	—	—	0.6	—	—
Project Need for Contractor Services	0.4	—	—	—	—	—	—	—	0.4	—
Work Order Activities Creation	15	—	—	—	2.8	1.4	1.4	—	—	—
Resources Estimation	3	3	—	3	—	—	—	—	—	—
Resources Availability	—	—	—	—	—	—	—	6	—	—
Plant/Fleet/Corporate Schedule	1	—	—	—	—	—	—	0.9	1	—
Technical Specifications	3	—	3	—	—	—	—	3	—	—
Unit Down Power	3	—	3	—	—	—	—	3	—	—
Chemistry Impact	—	1.5	1.5	—	—	—	—	1.5	—	—
Reactivity Effect	—	3	3	—	—	—	—	3	—	—
Reactor Coolant System Effect	—	1.5	1.5	—	—	—	—	1.5	—	—
Redundant Structures, Systems, and Components Effect	—	—	1	—	—	—	—	—	1	—
Plant Transient	1	—	1	—	—	—	—	1	—	0.9
Surveillance Schedule	—	—	3	—	—	—	—	—	—	—

Task Time in Minutes	Classification (Supervised)	Classification (Unsupervised)	Comparison	Estimation	Generation	Guidance	Insertion	Review	Structuring	Summarization
Seasonal Impacts	—	—	3	—	—	—	—	3	3	—
Grid Instability	—	1	1	—	—	—	—	1	—	—
Containment Integrity/Heat Removal	—	3	3	—	—	—	—	3	—	—
Cycle Timing	—	—	—	—	—	—	—	3	—	—
Plant Security	—	0.5	0.5	—	—	—	—	0.5	—	—
Maintenance Rule	—	—	1.5	1.5	—	—	—	—	—	—
Risk Screening	—	—	—	—	—	—	—	—	—	—
Work Scope Clarity and Conciseness	—	—	—	—	3	0.2	0.2	1.5	—	—
Integrated Risk Screening for Execution Week	—	—	—	—	—	—	—	—	—	—
Priority Assignment Validation	0.25	—	—	—	—	—	—	0.25	—	—
Priority Work Rank	—	—	1.5	1.5	—	—	—	1.5	—	—
Preventive Maintenance Grace Periods	—	—	—	—	—	—	—	—	—	—

Table 14. Estimated potential time saved in minutes per WP for each planning step by each AI function.

Task Time in Minutes	Classification (Supervised)	Classification (Unsupervised)	Comparison	Estimation	Generation	Guidance	Insertion	Review	Structuring	Summarization
Need for Work Package	4	—	—	—	9.9	—	—	—	—	—
Previous Similar Work Packages	—	—	5	—	—	—	—	—	—	—
Long Lead-Time Parts	—	—	—	15	—	—	—	—	15	—
Existing or New Work Package Use	—	—	—	—	5	1.875	1.875	—	—	5
Similar Work Packages Review for Information	—	—	—	—	1.95	—	—	—	1.95	3.75
Needed Technical Information	—	—	15	—	—	—	—	—	15	15
Work Instructions	—	—	—	—	15	7.5	7.5	—	—	—
Incorporate Operating Experience	—	—	10	—	5	—	—	—	10	20
Worker Feedback from Previous Similar Work	—	—	5	—	3	—	—	—	5	10
Human Performance Tools	2	—	—	—	3	—	—	2	2	—
Clearance Order Request	3	—	—	—	4.95	—	—	3	3	—
Engineering and Operations Documentation	—	—	—	—	1.5	—	—	—	—	15
Administrative or Technical Procedures	2	—	—	—	10.05	—	—	2	2	—
Fire Protection and Industrial Safety Controls	1	1	1	—	1	—	—	1	1	—
Fire Protection and Industrial Safety Permits	0.1	—	—	—	1.2	—	—	0.1	0.1	—
Work Package Walkdown	5	—	—	—	3	—	—	—	5	—
Radiological Controls	1	—	1	—	7.5	—	—	—	1	—
ALARA Plan	0.75	—	0.75	—	—	—	—	—	0.75	—
Radiological Work Impact on Workers	—	—	—	2.5	0.2	—	—	—	—	—
Worker Training and Qualification	3	3	—	—	3.75	—	—	3	3	—
Resources Assignment to Work Package	—	—	5	—	—	—	—	—	—	—
Equipment Qualification Requirements	5	—	5	—	—	—	—	1.5	5	—
Parts and Materials	20	—	—	—	—	—	—	—	20	—
Work Package Review and Approval	—	—	—	—	—	—	—	26	—	—

4.7 Scheduling

The scheduling process in work management involves several critical tasks designed to optimize resource allocation and ensure efficient execution of work activities. The process begins with the creation of all necessary STs for the execution week. This initial step establishes the framework for subsequent evaluations and adjustments.

Once the STs are drafted, the next phase assesses whether all STs have appropriate resources assigned. This evaluation ensures alignment between the work planned and the available personnel. Next, an assessment is conducted of available resources against task requirements to identify any gaps or surpluses, which helps clarify the resource situation for the upcoming week.

Addressing any identified resource deficiencies is critical for successful execution. The process includes evaluating the feasibility of using overtime to fill resource gaps, providing a potential solution for shortages. Additionally, the option of employing contractors is explored to supplement the workforce as needed. In some cases, resources may need to be overloaded beyond 100% capacity during the execution week, assuming work can be performed more efficiently. If no solutions work, tasks may be removed from the execution week schedule to alleviate the burden on available resources.

Alternatively, if the work scheduled is less than the available resources, the scheduling team may face the need to add extra normal work or low-priority work to the execution week schedule.

Once the deficiency or surplus is taken care of, the impact of any new tasks on existing scope is evaluated. A logical ordering of existing tasks is then carried out to support the most efficient workflow and optimal resource usage.

After all evaluations and adjustments, the work execution week schedule is finalized and “frozen” to prevent further changes. A certification meeting is held to review and confirm the execution week plan, addressing any remaining concerns before moving forward. The execution week plan is then formally handed over from the cycle planner to the workweek manager to ensure clarity in responsibilities. Finally, a daily work schedule is issued to guide teams throughout the execution week, ensuring that all personnel are aware of their tasks and responsibilities.

As discussed in Section 4.1, each step is mapped to specific activities performed by dedicated staff, and AI functions are identified to replace those tasks. The detailed steps and considerations involved in the process are presented in Appendix D. Table 15 summarizes the findings from the analysis of the process in Appendix D and is used to quantify the potential cost savings of the entire process in Section 5 and each AI function in Section 6.

Table 15. Estimated potential time saved in minutes per ST for each scheduling step by each AI function.

Task Time in Minutes	Classification (Supervised)	Classification (Unsupervised)	Comparison	Estimation	Generation	Guidance	Insertion	Review	Structuring	Summarization
Schedule Tasks Creation	—	—	—	—	5	—	—	—	—	—
Resources Assignment to All Schedule Tasks	—	—	—	—	—	—	—	—	—	—
Deficiency or Surplus of Resources for Execution Week	—	—	—	—	—	—	—	—	—	—
Overtime Use	—	—	—	—	2.5	—	—	—	—	—
Contractors Use	1.5	—	—	—	0.45	—	—	1.5	1.5	—
Resources Overload	—	—	—	4.5	—	—	—	0.3	2.25	—
Work Removal from Execution Week Schedule	—	—	—	—	—	—	—	—	—	—
Work Addition to Execution Week Schedule	0.4	0.4	—	0.4	—	—	—	0.4	—	—
Lower-Priority Work Addition to Execution Week Schedule	0.4	0.4	—	0.4	—	—	—	0.4	—	—
Impact of New Items Screened on Existing Week Scope	—	—	1	—	—	—	—	—	—	—
Logical Order of Schedule Tasks	—	—	—	—	—	—	—	5	—	—
Execution Week Schedule Freeze	—	—	—	—	—	0.5	0.5	—	—	—
Execution week certification meeting	—	—	—	—	—	—	—	—	—	—
Changes from Certification Meeting	—	—	—	—	—	—	—	—	—	—
Turnover of Execution week from Cycle Planner to Workweek Manager	—	—	—	—	—	—	—	—	—	—
Daily Work Schedule Issuance	—	—	—	—	—	—	—	—	—	0.5

4.8 Execution

The work execution process involves a series of critical tasks designed to ensure the effective completion of WPs while maintaining operational safety and efficiency. The process begins with the review of the daily work schedule, which sets the stage for the day's activities.

Key tasks include conducting RP operations and engineering work associated with the WP. RP might need to perform new surveys to accommodate the work scope. Operations might perform tagout and/or configure the plant as needed. Engineering might need to perform inspections, surveillances, and/or troubleshooting. Other organizations may also have predecessor tasks. All teams involved would then update the work schedule to reflect any changes or actions. The predecessor tasks must be completed before maintenance activities can commence.

The maintenance team needs to prepare for execution, which includes confirming that predecessor tasks have been completed, picking up necessary materials and parts from the warehouse, storing them if needed in a dedicated area, and checking out tools and calibrated equipment. The maintenance team also ensures that the chemicals and lubricants needed for the tasks at hand are available.

Communication is critical throughout the process, requiring maintenance personnel to check in with operations and discuss plans before commencing work. If necessary, maintenance will implement FME zones.

Next, work groups perform their assigned tasks, and they update WP notes and statuses to maintain accurate records of progress. The maintenance supervisor must approve the completed work before conducting further updates to the work schedule. Finally, the process concludes with the closeout of the WP and a completion check to ensure all tasks have been properly completed and documented.

As discussed in Section 4.1, each step was mapped to specific activities performed by dedicated staff, and AI functions were identified to replace those tasks. The detailed steps and considerations involved in the process are presented in Appendix E. Table 16 summarizes the findings from the analysis of the process in Appendix E and is used to quantify the potential cost savings of the entire process in Section 5 and each AI function in Section 6.

4.9 Post-Execution

The post-execution process involves several key tasks aimed at evaluating and improving work performance. It begins with gathering execution week performance information through work management systems. These data are then analyzed across all relevant groups to identify trends and areas for improvement.

After the analysis, a post-execution week performance critique meeting is conducted, where team members discuss findings and share insights. Finally, any performance shortfalls are documented to ensure transparency and facilitate future improvements. This structured approach helps organizations learn from each execution cycle and enhances overall operational efficiency.

As discussed in Section 4.1, each step was mapped to specific activities performed by dedicated staff, and AI functions were identified to replace those tasks. The detailed steps and considerations involved in the process are presented in Appendix F. Table 17 summarizes the findings from the analysis of the process in Appendix F and is used to quantify the potential cost savings of the entire process in Section 5 and each AI function in Section 6.

Table 16. Estimated potential time saved in minutes per WP for each work execution step by each AI function.

Task Time in Minutes	Classification (Supervised)	Classification (Unsupervised)	Comparison	Estimation	Generation	Guidance	Insertion	Review	Structuring	Summarization
Daily Work Schedule Review	—	—	—	—	—	—	—	—	—	—
RP Work Associated with the Work Package	—	0.5	—	—	3.75	—	—	0.5	—	—
RP Work Execution Week Survey	1.25	1.25	—	—	—	—	—	1.25	1.25	—
RP Work Schedule Updates	—	—	—	—	—	0.25	0.25	—	—	—
Operations CO Associated with the Work Package	—	1	—	—	—	—	—	1	—	—
Operations Work Schedule Update	—	—	—	—	—	0.2	0.2	—	—	—
Engineering Determination of Work	—	1	—	—	—	—	—	—	—	3
Engineering Work Schedule Update	—	—	—	—	—	0.1	0.1	—	—	—
Other Work Groups Predecessor Tasks Prior to Maintenance	—	1	—	—	—	—	—	—	—	—
Other Groups Work Schedule Update	—	—	—	—	—	0.1	0.1	—	—	—
Maintenance Determination of Work Predecessor Tasks	—	—	—	—	—	—	—	—	—	3
Pick-up of Materials and Parts from Warehouse	—	—	—	—	—	—	—	—	—	—
Temporary Materials and Parts Storage	—	—	—	—	—	—	—	—	—	—
Check-out of Tools for Work	—	2	2	—	—	—	—	—	—	—
Check-out of Calibrated Tools and Equipment	—	2	2	—	—	—	—	—	—	—
Chemicals and Lubricants	—	2	2	—	—	—	—	—	—	—
Maintenance Check-in with Operations	—	—	—	—	—	—	—	—	—	—
Maintenance FME Zone Actions	—	1	—	—	—	—	—	1	—	—
Work Groups Perform Work	—	—	—	—	—	—	—	—	—	—
Work Groups Update of Work Package Notes/Status	—	—	—	—	—	1	1	—	—	—
Maintenance Supervisor Approval of Work	—	—	—	—	3	2.375	2.375	5	—	—
Maintenance Work Schedule Updates	—	—	—	—	—	1	1	—	—	—
Work Package Closeout	—	—	—	—	—	7.5	7.5	—	—	—
Work Package Completion Check	—	—	—	—	1.5	—	—	10	—	—

Table 17. Estimated potential time saved in minutes per WP for each post-execution step by each AI function.

Task Time in Minutes	Classification (Supervised)	Classification (Unsupervised)	Comparison	Estimation	Generation	Guidance	Insertion	Review	Structuring	Summarization
Work Management Evaluation of Week Performance Information	—	—	—	—	2	—	—	—	—	2
Relevant Groups Analysis of Execution Week Performance Information	—	—	—	—	—	—	—	1	—	—
Post-Execution Week Performance Critique Meeting	—	—	—	—	—	—	—	—	—	—
Performance Shortfalls Documentation	—	—	—	—	0.05	—	—	1	—	—

5. POTENTIAL COST SAVING IN WORK MANAGEMENT

In this section, a summary of potential cost savings is developed by quantifying the cost of each step of the work management process and comparing it to the percentage reduction in cost after implementing AI. As discussed in the assumptions, the potential savings of AI, if deployed, is assumed to the maximum extent, i.e., where the manual process is totally abandoned.

To calculate the total time saved, it is necessary to estimate the number of CRs generated and scale them to WOs, WPs, and STs, as each phase in Appendices A-F uses a different scale to quantify effort. A survey of datasets from multiple utilities indicated a conservative estimate of 6,000 CRs per nuclear unit per year. Table 18 presents an estimated ratio of the number of records scaled in comparison to each other. Those are assumed ratios based on discussion with industry. Projects and PM WO numbers were also estimated on the conservative side based on discussion with industry. In this section's analysis, a nuclear plant staff average cost was estimated at \$200 per hour. This enables converting the findings into actual cost saving per year.

Table 18. Reference metrics and scaling factors for each work management phase.

Phase	Metric	Scale to previous metric	Assumed number per unit per year
Screening	CR	Not applicable (NA)	6,000
Scoping	WO	One corrective WO per five CRs	1,100
		Project WOs	400*
		PM WOs	6,000
		Total WOs	7,500
Planning	WP	Four WPs per three WOs	10,000
Scheduling	ST	Two STs per WP	20,000
Execution	WP	One WP to two STs	10,000
Post-Execution	WP		10,000

* A typical WO contains two work activities. However, project WOs often involve multiple activities and require more effort at each step. While the actual number of project WOs is estimated at 120, it is assumed to be 400 to account for the additional work involved in project activities.

5.1 Screening

A summary of screening steps is provided in Table 19. The analysis of each of the listed steps reveals significant insights into current cost thousands of dollars (\$K) and the potential cost-saving benefits of automation (in %). The table shows that most steps can be automated resulting in a maximum potential cost saving of 57.6%. However, certain steps like prescreening meeting, screening meeting preparation, and the actual screening meeting show no potential for savings, since they rely on people coordinating and discussing in person. Specifically, the prescreening and screening meetings currently take 30 minutes each and involve several staff members. Compensatory actions were also found to be non-automatable because they depend on the broader staff experience, which AI models cannot replicate. Considering the overall cost of this phase at \$1.5M and the 57.6% potential cost saving, this results in around \$880K of cost saving.

Table 19. Current costs and potential cost savings by using AI for every step of screening.

Task	Current Cost (in \$K per unit per year)	Maximum AI Cost Reduction (%)
Issue Identification and Reporting	210	76
Equipment Relevance	10	100
Criticality	30	100
Operability and Impact to Plant	60	100
Maintenance Validation	63	98
Maintenance Compensatory Actions	15	0
Maintenance Significance	15	100
Operations Validation	17	94
Operations Compensatory Actions	15	0
Operations Significance	15	100
Engineering Validation	16	97
Engineering Compensatory Actions	15	0
Engineering Significance	15	100
All Others Validation	16	97
Rework	45	100
Resources for Significant Work	45	100
Project/In-house for Significant Work	15	99
Modification/Field Change	16	100
Safety Significance	30	100
Administrative Impact	25	100
Radiological Impact	40	100
Chemistry Impact	100	100
Severity	62	98
Actions Needed	40	100
Prescreening Meeting	120	0
Screening Meeting Preparation	139	0
Screening Meeting	120	0
Condition Report Update and Work Request Notification	205	20
Maintenance Work Order Creation	13	71
Total/Average	1,526	57.6

5.2 Scoping

A summary of scoping steps is listed in Table 20. The table outlines the steps associated with scoping, including the current cost and the potential cost savings when automated with AI functions. The table indicates that most steps can be automated resulting in a maximum potential cost saving of 85.2%. This represents a significant improvement over the previous phase, due to more systematic decision-making and reduced reliance on staff coordination and non-automatable physical activities. Two steps that cannot be automated relate to risk. While AI could potentially aid in risk optimization, this was deemed outside the scope of the current effort and could be a research effort on its own. Also, risk screening is too risky and complex for effective automation. It requires human expertise in risk assessments and extensive knowledge of the plant. Preventive maintenance grace period determination was another step that was not automated by AI because it is simply a lookup from a database table. The highest potential for cost savings is associated with creating work activities. Automating this step alone could result in around \$400K in annual savings. Considering the overall cost of this phase at \$2.5M and the 85.2% potential cost saving, this results in around \$2.1M of cost saving.

5.3 Planning

A summary of planning tasks is provided in Table 21. The table outlines the steps associated with planning, including the current cost and the potential cost savings when automated with AI functions. Except for one step, all steps in the table can be automated by AI resulting in a maximum potential cost saving of 99.1%. This indicates that this entire phase is automatable. However, as discussed in Section 6, this relies heavily on tools to review unstructured text, follow procedures, and create documents. Many of these are not ready for deployment, as discussed in Section 3 and Section 7. The one step that could not be automated in this phase was the actual reservation of parts, since this involves manual data entry into a system, which could be automated by simple data management tools. Considering the overall cost of this phase at \$7.8M and the 99.1% potential cost saving, this results in approximately \$7.7M in cost savings. Many of the listed steps can produce savings in the range of hundreds of thousands of dollars, making this phase a priority for AI deployment, provided the current AI challenges are resolved.

5.4 Scheduling

A summary of scheduling steps is listed in Table 22. The table outlines the steps associated with scheduling, including the current cost and the potential cost savings from automating these steps with AI functions. The table indicates that less steps can be automated than the previous two phases, resulting in a maximum potential cost saving of 67.2%. Many of the steps that were not automated by AI were mainly due to their simplicity and the possibility of automation without AI. For example, verifying resource assignments to all tasks could be done through a simple database query. Identifying deficiencies or surpluses in resources would require a query to confirm that available resources match the number needed for scheduled tasks. The removal of low-priority work from the execution week schedule is enabled through labels from previous steps. Additionally, all meeting steps were not automated due to the need for human interaction, coordination, and discussion. Considering the overall cost of this phase at \$1.96M and the 67.2% potential cost saving, this results in around \$1.3M of cost savings.

Table 20. Current costs and potential cost savings by using AI for every step of scoping.

Task	Current Cost (in \$K per unit per year)	Maximum AI Cost Reduction (%)
Merger of Work Orders for Same Issue	131	100
Work Order Priority	71	100
Work Order Criticality	71	100
Outage or Online Work	110	100
Preliminary Work Schedule	63	100
Project Relevance	25	100
Project Sequence	15	100
Project Week(s) Readiness	30	100
Project Engineering Evaluation Review	15	100
Project Need for Contractor Services	5	100
Work Order Activities Creation	445	92
Resources Estimation	75	100
Resources Availability	188	80
Plant/Fleet/Corporate Schedule	48	100
Technical Specifications	75	100
Unit Down Power	75	100
Chemistry Impact	38	100
Reactivity Effect	75	100
Reactor Coolant System Effect	38	100
Redundant Structures, Systems, and Components Effect	25	100
Plant Transient	48	100
Surveillance Schedule	75	100
Seasonal Impacts	75	100
Grid Instability	25	100
Containment Integrity/Heat Removal	75	100
Cycle Timing	75	100
Plant Security	13	100
Maintenance Rule	45	83
Risk Screening	60	0
Work Scope Clarity and Conciseness	123	96
Integrated Risk Screening for Execution Week	188	0
Priority Assignment Validation	6	100
Priority Work Rank	38	100
Preventive Maintenance Grace Periods	38	0
Total/Average	2,498 ^j	85.2

^j This number might not match the summation of all the numbers above it due to rounding.

Table 21. Current costs and potential cost savings by using AI for every step of planning.

Task	Current Cost (in \$K per unit per year)	Maximum AI Cost Reduction (%)
Need for Work Package	463	100
Previous Similar Work Packages	167	100
Long Lead-Time Parts	567	88
Existing or New Work Package Use	333	100
Similar Work Packages Review for Information	190	100
Needed Technical Information	500	100
Work Instructions	500	100
Incorporate Operating Experience	500	100
Worker Feedback from Previous Similar Work	267	100
Human Performance Tools	167	100
Clearance Order Request	265	100
Engineering and Operations Documentation	550	100
Administrative or Technical Procedures	402	100
Fire Protection and Industrial Safety Controls	67	100
Fire Protection and Industrial Safety Permits	43	100
Work Package Walkdown	267	100
Radiological Controls	283	100
ALARA Plan	25	100
Radiological Work Impact on Workers	90	100
Worker Training and Qualification	225	100
Resources Assignment to Work Package	167	100
Equipment Qualification Requirements	217	100
Parts and Materials	667	100
Work Package Review and Approval	867	100
Total/Average	7,786	99.1

Table 22. Current costs and potential cost savings by using AI for every step of scheduling.

Task	Current Cost (in \$K per unit per year)	Maximum AI Cost Reduction (%)
Schedule Tasks Creation	333	100
Resources Assignment to All Schedule Tasks	67	0
Deficiency or Surplus of Resources for Execution Week	200	0
Overtime Use	217	77
Contractors Use	130	100
Resources Overload	170	100
Work Removal from Execution Week Schedule	50	0
Work Addition to Execution Week Schedule	27	100
Lower-Priority Work Addition to Execution Week Schedule	27	100
Impact of New Items Screened on Existing Execution Week Scope	67	100
Logical Order of Schedule Tasks	333	100
Execution Week Schedule Freeze	67	50
Execution Week Certification Meeting	203	0
Changes from Certification Meeting	7	0
Turnover of Execution Week	33	0
Daily Work Schedule Issuance	33	100
Total/Average	1,964	67.2

5.5 Execution

A summary of work execution steps is listed in Table 23. The table outlines the steps associated with execution, including the current cost and potential savings from automating these steps with AI functions. The table indicates much fewer steps can be automated compared to the previous phases, resulting in a maximum potential cost saving of 40.2%. This phase is the least promising for AI due to the nature of the steps, which involve human physical actions such as retrieving materials and tools, setting up the work environment, performing work, and interacting directly with the schedule and staff. These activities are inherently manual and cannot be easily automated by AI. Despite the lower automation potential, the overall cost of this phase is estimated at \$4.45M, and the 40.2% potential cost saving translates to approximately \$1.79M in savings. Interestingly, this step generates around \$450K more cost savings than the previous scheduling phase, even though that phase had a 67.2% automation potential.

5.6 Post-Execution

A summary of post-execution steps is listed in Table 24. The table outlines the steps associated with post-execution, including the current cost and potential savings from automating these tasks with AI functions. The maximum potential cost saving for this phase is 84.4%. Except for the post-execution week performance critique meeting, which cannot be automated as it requires human interaction and discussion, all other steps can be automated by AI. Since this phase involves review and performance evaluations, all the AI functions used for this step relate to generation, summarization, or review, as will be shown in Section 6. The overall cost of this phase is relatively low, estimated at \$160,000. The 84.4% potential cost saving translates to approximately \$135,000 in savings.

Table 23. Current costs and potential cost savings by using AI for every step of execution.

Task	Current Cost (in \$K per unit per year)	Maximum AI Cost Reduction (%)
Daily Work Schedule Review	333	0
RP Work Associated with the Work Package	142	100
RP Work Execution Week Survey	102	41
RP Work Schedule Updates	17	50
Operations CO Associated with the Work Package	33	100
Operations Work Schedule Update	13	50
Engineering Determination of Work	133	100
Engineering Work Schedule Update	7	50
Other Work Groups Predecessor Tasks Prior to Maintenance	133	25
Other Groups Work Schedule Update	7	50
Maintenance Determination of Work Predecessor Tasks	100	100
Pick-up of Materials and Parts from Warehouse	442	0
Temporary Materials and Parts Storage	200	0
Check-out of Tools for Work	442	15
Check-out of Calibrated Tools and Equipment	317	21
Chemicals and Lubricants	167	40
Maintenance Check-in with Operations	335	0
Maintenance FME Zone Actions	83	40
Work Groups Perform Work	Out of scope	
Work Groups Update of Work Package Notes/Status	67	50
Maintenance Supervisor Approval of Work	425	81
Maintenance Work Schedule Updates	67	50
Work Package Closeout	500	50
Work Package Completion Check	383	100
Total/Average	4,447	40.2

Table 24. Current costs and potential cost savings by using AI for every step of post-execution.

Task	Current Cost (in \$K per unit per year)	Maximum AI Cost Reduction (%)
Work Management Evaluation of Week Performance	67	100
Relevant Groups Analysis of Execution Week Performance	33	100
Post-Execution Week Performance Critique Meeting	25	0
Performance Shortfalls Documentation	35	100
Total/Average	160	84.4

5.7 Summary of Cost Savings in Work Management

In this section, the findings of the previous sections are summarized in Table 25. The data presented highlights the substantial potential for cost savings through the automation of various phases in the work

management process. Each phase demonstrates significant reductions in manual processing time when automated, leading to notable financial benefits for the organization.

The most significant potential savings are observed in the planning phase, where automation could reduce costs by up to \$7.72M. This substantial reduction demonstrates that optimizing this phase can yield transformative financial benefits for the organization. However, as will be shown next, this phase relies heavily on methods and tools for generation, structuring, summarization, and review, which must perform sophisticated functions and are not yet ready for deployment. Scoping follows as the second-highest cost-saving phase, benefiting more from currently available tools. Scheduling and execution, on the other hand, are less automatable. Scheduling primarily involves simple operations that do not require AI, while execution is heavily reliant on physical activities. Additionally, while screening has been a major focus of the industry automation efforts so far, it is evident from the table that it has one of the lowest cost-saving potentials.

Table 25. Summary of current costs and potential cost savings per phase.

Phase	Current Cost per Unit per Year	Maximum AI Cost Saving	Ratio of Saving (%)
Screening	\$1.53M	\$0.88M	57.6
Scoping	\$2.50M	\$2.13M	85.2
Planning	\$7.79M	\$7.72M	99.1
Scheduling	\$1.96M	\$1.32M	67.2
Execution	\$4.45M	\$1.78M	40.1
Post-Execution	\$0.16M	\$0.14M	84.4

6. POTENTIAL COST SAVING OF FUNCTIONS

In this section, a summary of the potential cost savings derived from implementing AI functions is discussed across all steps in the phase. It is again assumed that the deployment of AI for a function would fully replace manual processes, i.e., the estimates reflect the maximum possible cost savings, except for guidance and insertion, which receive half credit (Section 4.2). One key assumption that was mentioned in Section 4.2 is that the cost saving from each function is assumed not to overlap (i.e., two functions can be used in one step, and each is counted as a cost-saving function). While one function is usually used to automate the decision, or action, it was necessary to double-count the redundant methods in the context of quantifying AI function cost saving to provide insight into the full potential of each AI function.

6.1 Screening

A summary of the AI functions for screening is listed in Table 26. The table illustrates the substantial cost-saving potential of various AI functions. It also compares the ratio of savings each function provides relative to others on the list, offering insights into their impact on overall cost reduction. Supervised and unsupervised classification together account for 30% of the total cost saving, amounting to \$485K. This substantial contribution is due to the critical role of classification in categorizing screening records into specific labels. Review also plays a similar role, contributing 27%, by leveraging procedures such as the CAP procedure and mining complex or unstructured documents like technical specifications or implied instructions in manuals to support decision-making. Comparison contribution at 24% is crucial because decisions often depended on references in records that were structured differently. Other functions, including generation (1%), summarization (6%), guidance (3%), insertion (3%), and structuring (5%), offer varying degrees of support but with less overall impact.

One key aspect of the cost savings presented in Table 26 is that summing all the costs saved would exceed the cost-saving figure discussed in Table 19. This is because of the overlapping credit taken for each function, as discussed in the introduction of this section.

Table 26. Cost savings in screening by each AI function.

Task	Contribution of AI function (%)	Maximum AI Cost Saving (\$K per unit per year)
Classification (Supervised)	19	310
Classification (Unsupervised)	11	175
Comparison	24	395
Estimation	0	0
Generation	1	20
Guidance	3	55
Insertion	3	55
Review	27	440
Structuring	5	75
Summarization	6	120

6.2 Scoping

The AI functions for scoping are summarized in Table 27. The table illustrates the substantial cost-saving potential of various AI functions and compares the ratio of savings each function provides relative to others, offering insights into their impact on overall cost reduction. Since scoping relies heavily on labeling to characterize WOs, supervised and unsupervised classification together account for 31% of the total cost saving, which is higher than in screening. This indicates that decision-making in scoping is more structured than screening. Also, given that scoping is a more expensive phase than screening, this amounts to \$1.2M in savings, demonstrating the considerable impact of these ready-to-deploy tools. Review also played a major role in scoping, similar to screening, contributing 27% with a saving of approximately \$1.1M. It focused on procedures such as chemistry procedures, operations procedures, technical specifications, and work management procedures, and sometimes reviewed unstructured documents like system drawings. Comparison, with a 25% contribution and a saving of around \$1M, was valuable for connecting WOs to structured lists of information. Other functions, including estimation (7%), generation (4%), guidance (1%), insertion (1%), structuring (3%), and summarization (1%), again provided varying degrees of support but with less overall impact. Overall, the potential cost savings for scoping appear to be higher than for screening, although the AI functions benefiting both are the same.

Table 27. Cost savings in scoping by each AI function.

Task	Contribution of AI function (%)	Maximum AI Cost Saving (\$K per unit per year)
Classification (Supervised)	22	856
Classification (Unsupervised)	9	353
Comparison	25	954
Estimation	7	275
Generation	4	151
Guidance	1	40
Insertion	1	40
Review	27	1,061
Structuring	3	135
Summarization	1	23

6.3 Planning

The AI functions to automate the planning process are summarized in Table 28. The table illustrates the substantial cost-saving potential of various AI functions and compares the ratio of savings each function provides relative to others, offering insights into their impact on overall cost reduction. A notable shift in AI function contributions is observed in this phase: supervised and unsupervised classification together account for only 12% of the total cost savings, a significant decrease compared to previous phases. In contrast, there is a substantial increase in the contributions from generation (19%), structuring (22%), and summarization (17%). Review also played a significant role in planning, contributing 9% with savings of approximately \$1.3M. This shift is due to planning involving significantly more complex decisions than earlier phases, with a greater reliance on unstructured and complex documents. These functions (generation, structuring, summarization, and review) depend on generative AI methods and tools, which, as discussed in Sections 3 and 7, can be challenging to deploy. However, since the combined cost savings from these four functions amount to more than \$9M, prioritizing research in this area is essential.

Despite a drop in the contribution ratio, comparison, with an 12% contribution, still resulted in significant savings of around \$1.6M due to planning being the most expensive phase. Other functions, including estimation (4%), guidance (2%), and insertion (2%), provided less impact.

Table 28. Cost savings in planning by each AI function.

Task	Contribution of AI function (%)	Maximum AI Cost Saving (\$K per unit per year)
Classification (Supervised)	11	1,561
Classification (Unsupervised)	1	133
Comparison	12	1,592
Estimation	4	583
Generation	19	2,533
Guidance	2	313
Insertion	2	313
Review	9	1,287
Structuring	22	2,993
Summarization	17	2,292

6.4 Scheduling

The AI functions to automate the scheduling process are summarized in Table 29. The table illustrates the substantial cost-saving potential of various AI functions and compares the ratio of savings each function provides relative to others, offering insights into their impact on overall cost reduction. Supervised and unsupervised classification together account for only 11% of the total cost savings, reflecting a continued decrease from previous phases. In contrast, estimation has a notable increase, contributing 18% to the total savings due to its role in time and resource estimation for scheduling. Generation, with a 26% contribution, was found to significantly impact few but costly steps, resulting in the 26% or \$530K savings. Review, contributing 25%, was crucial for optimizing schedules by considering constraints and resources, while structuring made a significant impact at 12% by extracting implied labels from unstructured documents. Other functions, including comparison, guidance, insertion, and summarization, had relatively insignificant cost-saving impacts.

Table 29. Cost savings in scheduling by each AI function.

Task	Contribution of AI function (%)	Maximum AI Cost Saving (\$K per unit per year)
Classification (Supervised)	8	153
Classification (Unsupervised)	3	53
Comparison	3	67
Estimation	18	353
Generation	26	530
Guidance	2	33
Insertion	2	33
Review	25	507
Structuring	12	250
Summarization	2	33

6.5 Execution

The AI functions to automate the work execution process are summarized in Table 30. The table illustrates the substantial cost-saving potential of various AI functions and compares the ratio of savings each function provides relative to others, offering insights into their impact on overall cost reduction. In this phase, which primarily involves physical activities and staff logging information, insertion and guidance are the most promising AI functions in terms of cost-saving, each contributing 12% to the total cost savings. Review played a significant role by mining procedures and making decisions from large datasets, contributing 24%. Unsupervised classification contributed 15%. Generation supported staff in creating novel texts, such as notes on rework needs or radiological impact, contributing 11%. Other functions were less influential: supervised classification contributed only 2%. Comparison and summarization provided minimal support, each contributing 8%, while structuring had a minimal impact at 2%.

Table 30. Cost savings in execution by each AI function.

Task	Contribution of AI function (%)	Maximum AI Cost Saving (\$K per unit per year)
Classification (Supervised)	2	42
Classification (Unsupervised)	15	392
Comparison	8	200
Estimation	0	0
Generation	11	275
Guidance	16	418
Insertion	16	418
Review	24	625
Structuring	2	42
Summarization	8	200

6.6 Post-Execution

The AI functions to automate the post-execution process are summarized in Table 31. The table presents an overview of tasks related to cost savings achieved through automation, specifically highlighting time savings and their corresponding monetary impacts. In this phase, generation, review, and summarization are the primary effective AI functions, each contributing modestly to cost savings. These functions are useful due to the phase's focus on reviewing or summarizing documents about performance and generating insights for future work. Other functions, including classification (supervised and unsupervised), comparison, estimation, guidance, insertion, and structuring, had no impact in this phase.

Table 31. Cost savings in post-execution by each AI function.

Task	Contribution of AI function (%)	Maximum AI Cost Saving (\$K per unit per year)
Classification (Supervised)	0	0
Classification (Unsupervised)	0	0
Comparison	0	0
Estimation	0	0
Generation	34	68
Guidance	0	0
Insertion	0	0
Review	33	67
Structuring	0	0
Summarization	33	67

6.7 Summary of Cost Savings of Functions

In this section, the results from the previous sections are aggregated. Table 32 presents a detailed breakdown of the cost savings associated with various AI functions across different phases of screening, scoping, planning, scheduling, execution, and post-execution. Overall, the table highlights the varying effectiveness of AI functions across these phases, with some excelling in specific tasks and others having

a more modest impact. Notably, planning benefits the most from AI functions due to its high cost. Among the functions, generation (\$3.6M), review (\$4M), structuring (\$3.5M), and summarization (\$2.7) are highlighted for their substantial total cost savings, though they are not yet ready for deployment as discussed in Sections 3 and 7. Comparison, with a total saving of \$3.2M, offers similar cost-saving potential and is ready for use. By prioritizing this function, organizations can enhance efficiency and reduce expenditures soon. The more challenging functions can be the focus of future deployment.

The table also reveals that each AI function excels in specific phases beyond planning. Supervised classification shows substantial contributions in screening and scoping. Unsupervised classification is effective in scoping and execution. Comparison stands out for its significant impact in screening and scoping. Estimation mostly impacts scoping and scheduling. Generation is notable in scheduling and execution. Guidance and insertion are equally impactful in execution. Review contributes significantly across most phases. Structuring excels in scheduling, while summarization is strong in execution. As discussed in the previous sections, the effectiveness of an AI function in each phase depends on the decisions made and the type of data or documents available. This explains why some functions perform exceptionally well in certain phases while having a more modest impact in others.

Table 32. Cost savings for each AI function across all phases.

Task	Cost savings in \$K per unit per year						
	Screening	Scoping	Planning	Scheduling	Execution	Post-Execution	Total ^k
Classification (Supervised)	310	856	1,561	153	42	0	2,922
Classification (Unsupervised)	175	353	133	53	392	0	1,106
Comparison	395	954	1592	67	200	0	3,207
Estimation	0	275	583	353	0	0	1,212
Generation	20	151	2,533	530	275	68	3,578
Guidance	55	40	313	33	418	0	858
Insertion	55	40	313	33	418	0	858
Review	440	1,061	1,287	507	625	67	3,986
Structuring	75	135	2,993	250	42	0	3,495
Summarization	105	23	2,292	33	200	67	2,719

7. FUNCTION CHALLENGES AND POTENTIAL METHODS

In the previous section, the most impactful functions were identified. This section examines the challenges faced as the methods and tools outlined in Section 2 were applied to these functions, as discussed in Section 3, within the context of the work process described in Section 4. The obstacles encountered are analyzed, and the research conducted to address and overcome many of these challenges

^k This number might not match the summation of all the numbers due to rounding.

is presented. Demonstrations of the solutions were included for each challenge when they were conducted.

7.1 Supervised Classification

Supervised classification has several challenges, primarily related to the availability and quality of labels, including issues with labels inconsistency and missing or imbalanced labels. The challenges that arose as supervised classification methods were applied to NPP data drove the research to investigate methods to tackle those challenges.

7.1.1 Sparse Labels

This challenge is specific to supervised methods, since supervised classification relies on labeled data to learn text patterns associated with classification decisions. Having sparse or very few labels has a significant negative impact on the performance of the developed solutions. For example, it is rare for the screening process that was discussed in Section 4.4 to classify events as SCAQ. A utility might have a few dozen such events out of a million CRs, which is not enough to train an AI model. Three methods were explored, each of which is scalable to any classification problem with sparse labels.

Developed Solution 1: Combining Data from More Utilities

Combining data from multiple utilities can have a positive effect if the data are consistent in the means that decisions are being made. This is usually the case in the nuclear industry; however, care should be taken when utilities use different criteria to describe what is assigned to a particular label.

Demonstration of Solution 1: SCAQ Example

In this example, an SCAQ prediction was generated using data from a single utility (containing 6 SCQ CRs) and compared to the results obtained when data from multiple utilities was used (Containing a total of 112 SCAQ CRs). A TF-IDF and logistic regression model were used with five-fold cross-validation in both cases.

The results are shown in Table 33. A baseline, where all CRs are labeled as SCAQ, was generated first to be used a benchmark. The baseline F1 for an SCAQ is 0.01% for a single utility and 0.03% for several utilities. The TF-IDF and logistic regression model were then applied to the two datasets: the one for a single utility and another using several utilities. The recall was set to around 30% to enable visual comparison of the F1 scores and precision results.

The results show that the F1 score for the single utility did not improve due to the low number of CRs. However, the F1 score for several utilities' SCAQ classification using logistic regression was 1.6%. This is a 53-fold increase over the random prediction baseline, indicating the model was able to better learn text patterns that indicate an SCAQ when data from multiple utilities is used. It is reasonable to assume that in this example the labeling process across the different utilities was consistent and therefore improved the SCAQ classification.

Developed Solution 2: Use of Projection through Another Class That Has Sufficient Labels

Often the class with sparse labels is an extension of another class that has labels. For example, if a priority 8 to 10 assignment for a WO is very rare, a regression model can be developed to estimate other more common priorities (0 to 7), and extrapolation can be used to predict what a priority 8 to 10 would be. The same could apply to confidence of a class. For example, because CRs are rarely assigned an SCAQ label, an SCAQ could be assigned to a CAQ classifier if the classifier is very confident it is a CAQ (since SCAQ is a special class of CAQ in extreme conditions).

Demonstration of Solution 2: SCAQ Example

The SCAQ prediction is again targeted in this example. To handle the sparse SCAQ label situation, SCAQ CRs were given a magnitude instead of a class label. An NCAQ had magnitude of 0, a CAQ a

magnitude of 1, and an SCAQ a magnitude of 100. This allowed the AI algorithm to learn from the large amount of CAQ data to use for SCAQ predictions. Ridge regression was used to predict the SCAQ labels. Five-fold cross-validation was used.

To demonstrate the potential for this method, it was benchmarked against a logistic regression model that was trained on the 112 available labels. The logistic regression therefore did not use the CAQ CRs to inform the regression, only the SCAQ CRs. The recall was again set to around 30% to enable visual comparison of the F1 scores and precision results.

The results are shown in Table 33. The single utility model achieved an F1 score of 0.68%, whereas the several utilities grouped together achieved an F1 score of 37.36%. This latter score is a 1,245-fold increase over the random baseline and a 23-fold increase over logistic regression.

Developed Solution 3: Using Context Sensitive Word Embeddings and Similarity

For labels that occur infrequently in a dataset, it is possible to develop a semisupervised model that uses an algorithm to find certain words that occur frequently and the likelihood that they will occur to create a rare-class classification. It is then possible to feed the words that are found into an embedding model to generate context-aware embedding vectors. Any new text can be compared to the embedding vectors, using similarity analysis methods, and if it is similar enough it is classified as positive.

Demonstration of Solution 3: SCAQ Example

The same dataset used in the previous example is used again herein. A relatively small embedding model LLM (moka-ai/m3e-base) was optimized to transform raw text into high-dimensional representations suitable for measuring the similarity of documents against each other using cosine similarity. The CRs were each converted into a 768-dimensional representation. In addition, five-fold cross-validation was performed. Each fold used one or two SCAQ CR embedding vectors from the 112 SCAQ CRs embeddings to compare using cosine similarity against the remaining 110 or 111 SCAQ CRs.

The results are shown in Table 33. With the recall rate set to 30% the F1-score stayed relatively low at 0.05%. This approach suffered from a very high false positive rate, the reason for which can be understood by analyzing the distribution of cosine similarity values for all SCAQ, CAQ, and NCAQ records superimposed on the same histogram. That is, cosine similarity results were calculated for all CRs against each other. The similarity values were then segregated according to the true label for each CR: either SCAQ, CAQ, or NCAQ. The histogram is shown in Figure 9. The figure shows that the similarity values for the SCAQ and CAQ records overlap significantly. Although the mean for the SCAQ values is farther to the right than it is for the CAQ or the NCAQ values, and the standard deviation is narrower, it is still very difficult to distinguish between the similarity values. Therefore, while context-sensitive embeddings can be used with some success to correctly classify a CR as an SCAQ, it is very difficult to do so without sustaining a large penalty in the form of many false positives.

Table 33. Summary of performance results for supervised classification of SCAQ.

Approach	Method	F1 Score (%)	Precision/Recall (%)	Confusion Matrix			
Baseline (single utility)	All CRs are labeled as an SCAQ	0.01	0.0/100.0	—	Predicted		
				—	—	Not SCAQ	SCAQ
				Actual	Not SCAQ	0	83,137
					SCAQ	0	6
Baseline (five utilities)	All CRs are labeled as an SCAQ	0.03	0.0/100	—	Predicted		
				—	—	Not SCAQ	SCAQ
				Actual	Not SCAQ	0	838,980
					SCAQ	0	112
SCAQ-driven classification (single utility)	TF-IDF + logistic regression	0.01	0.003/33.3	—	Predicted		
				—	—	Not SCAQ	SCAQ
				Actual	Not SCAQ	26,768	56,363
					SCAQ	4	2
SCAQ-driven classification (five utilities)	TF-IDF + logistic regression	1.60	0.82/29.46	—	Predicted		
				—	—	Not SCAQ	SCAQ
				Actual	Not SCAQ	835,006	3,974
					SCAQ	79	33
Extrapolation approach (single utility)	TF-IDF + ridge regression	0.68	0.34/33.3	—	Predicted		
				—	—	Not SCAQ	SCAQ
				Actual	Not SCAQ	82,553	578
					SCAQ	4	2
Extrapolation approach (five utilities)	TF-IDF + ridge regression	37.36	48.56/30.35	—	Predicted		
				—	—	Not SCAQ	SCAQ
				Actual	Not SCAQ	838,944	36
					SCAQ	78	34
Cosine similarity of context-sensitive embeddings	Moka embeddings + cosine similarity	0.05	0.03/33.3	—	Predicted		
				—	—	Not SCAQ	SCAQ
				Actual	Not SCAQ	74,728	7,504
					SCAQ	4	2



Figure 9: Histograms of cosine similarity values for records tagged as CAQ, NCAQ, or SCAQ.

7.1.2 Consistency of Labels Overtime or Across Utilities

Adding data from multiple utilities or from various periods of time for one utility is usually perceived as a means for improving model performance; however, adding contradicting data to broader models could worsen model performance. Tribal knowledge at each utility can be critical to the classification process; for example, utility staff often use a particular phrase to describe a situation that is likely to be critical, whereas staff at a different utility use a different phrase to describe the same situation. Some utilities have specific procedures that might not mirror what other utilities are using.

Even in the same utility, a decision-making procedure might change due to relaxed or more stringent policies or regulations. It has been observed that a utility will produce data for one period that are relatively easy to predict, and yet data from a different period are more difficult to predict. This inconsistency can occur gradually over time. For instance, a permanent procedure change can occur that will outdated data older than the change. Alternatively, data points can be inconsistently labeled during a pandemic, an outage, or other circumstance.

Combining models from various periods of time or across utilities could result in improved performance if the utilities follow a similar decision-making process, or in degraded performance if the utilities follow a different decision-making process. Consistency is a critical aspect of data that directly impacts a model's performance., Robie, and Branson (2015) note that "*Incorrectly or inconsistently labeled training data lead[s] to significant prediction errors because the learning algorithm will find an overly complex rule that deviates from the true best classifier.*"

Developed Solution 1: Measuring Consistency Similarity between Text and Decisions

Consistency determination methods can be based on specific models that are used where inconsistent data seems to appear as output outliers. However, an approach was developed in this effort that can be used in an application-generic manner. The idea behind this approach is that if similar data are labeled differently, this implies inconsistency in the decision-making process. Therefore, it is possible to compare label agreement to data similarity using similarity analysis tools to quantify the quality of the data before models are run, and to isolate severe inconsistencies for the user to clean (i.e., elaborate on the text differences or change the classification to match).

Demonstration of Solution 1: CAQ Example

For consistency, the data from five utilities discussed earlier were analyzed. Two durations were used for one of the utility's datasets to highlight decision-making inconsistencies even within the same utility. To evaluate the consistency metric performance, a subset of 4,000 datapoints were used from each utility and balanced at 50% CAQ and 50% NCAQ classification. The consistency was compared against a simple classifier that was discussed earlier (TF-IDF and ridge regression). If more consistent data results in high model performance, this confirms the hypothesis that the proposed approach to consistency measurement effectively provides insight into data quality.

The results are shown in Table 34. Two fictional benchmark cases were added. One in which all the data were labeled randomly; the other had correct 50% CAQ and 50% NCAQ labels. The results confirm the direct proportionality of consistency and model performance (using the simple classifier). This experiment demonstrates that it is possible to assess data label consistency even before any AI model is used for classification. The table shows that the data label consistency for utility 3 changed over time, which resulted in two different model performance results.

Table 34. Correlation of data consistency metric and model performance.

Dataset	Consistency Metric	Testing Model Accuracy	Testing Mode F1 Score
Random labels	50.3	51.7	66.7
Utility 1	58.4	76.9	78.5
Utility 2	61.0	71.9	74.7
Utility 3 duration 1	65.3	85.9	85.7
Utility 3 duration 2	63.6	81.9	82.0
Utility 4	63.6	82.0	82.3
Utility 5	66.0	88.9	88.7
100% consistent data	100.0	100.0	100.0

7.1.3 Missing or Highly Imbalanced Labels Data

Missing data points are problematic, especially in the classification process. For example, if data for a certain classification task have a strong bias toward a certain decision, the resulting AI mode will inherit this bias and generate models that resembles the data labels. Imbalance is another issue. Some utilities have a more severe labels imbalance than others. As an example, a normal imbalance might be 30%/70% CAQ/NCAQ, but some utilities could be at 5%/95%. This results in a dataset that is more difficult to classify because of the severe imbalance. Since this is a data issue, no solutions are proposed herein beyond ensuring that the training data reflect the population of labels, or the results are normalized to account for it.

7.2 Unsupervised Classification

Topic modeling and generative AI autoregressive models are two of the most powerful tools for unsupervised classification (see Section 3.2). While generative AI can be used for generic knowledge classification, topic modeling can be used for challenging classifications that require specific-domain knowledge, such as whether a WO requires a HP evaluation.

Like supervised classification, unsupervised classification also has several challenges; these are mostly associated with training time when topic modeling is used, and prompt sensitivity when generative AI is used. As unsupervised classification methods were applied to NPP data, the challenges that arose drove the research to investigate methods to tackle those challenges.

7.2.1 Iterations Time for New Labels

Topic modeling can be guided by an SME. However, training a single topic can take hours to be able to perform superior unsupervised classification, as the AI model needs to evaluate new correlations of topic words across the whole corpus. This process involves tuning billions of parameters and is therefore a time- and computational-power-consuming process. The training could involve terabits of data to enable the model to learn patterns and structures.

If models are being used to assign predesigned topics to text, this process can happen in the background, and the results are stored in a readily format for users. However, if new topics (custom topics) are being created on the fly and need to be trained, it is not practical for users to perform this online, especially if iterations are expected. This introduces a usability challenge. The user can't wait for hours for each iteration of a new topic.

It is also sometimes the case that certain categories or clusters within a dataset are underrepresented or nonexistent. This occurs when there is insufficient text data available for a specific label, and the model is not able to identify and analyze those topics effectively. Regardless of how many attempts are performed to generate those labels, the data limit the model's ability to understand them as they are nonexistent.

Developed Solution 1: Using Word Correlation as a Preliminary Indicator

A novel approach was developed that involves optimizing the information gained from not only the seed words but also any other words in the vocabulary that improve the capacity of the topic. The approach relies on establishing word relationships regardless of the topic being developed. Those relationships show how words are correlated in the context of each document and in the context of the whole corpus. A correlation matrix is created, and as new topics are created it is used in a rapid manner to indicate whether the words inserted by the SME are contextually correlated. If the words are uncorrelated, the topic-modeling model operates like a search engine. If the words are correlated, they teach the model to find context-aware results.

Demonstration of Solution 1: Events Classification into a Custom Label

In this example, the goal is to demonstrate how correlation matrices can support topic modeling. It is assumed that the SME is interested in finding words that relate to valves. Two reference sets were used: the first represented the correlation of words across the entire corpus of text from an NPP, while the second focused on the correlation of words within each individual document of the same corpus. Using the word “valve_stem” and corpus from an NPP, a summary of 15 highly correlated n-grams was generated (see Table 35). The top row in the table indicates which corpus or document correlations were used to mine the related words that are displayed in the following rows. The rows are ordered according to normalized adjacency scores. This data shows that the adjacency scores for the n-grams “valve” and “valve_stem” are significantly higher (95.46% for corpus and 96.18% for document) than those for “valve_stem” and itself (88.76% for corpus and 86.18% for document). It is also noteworthy that there is a significant overlap between the two sets. That is, three-fourths of the n-grams are shared between the two sets. This shows that either approach to forming the word adjacency matrices, corpus or document, provides similar results. After these additional related words are discovered with relative ease, they can easily be added to the topic as desired, and topic modeling can be performed for a more refined model using those words. This approach reduces the need for iterative topic modeling.

Table 35. Top 15 n-grams relating to the bigram “valve_stem”.

Corpus		Document	
Related N-Gram	Adjacency Score (%)	Related N-Gram	Adjacency Score (%)
valve	95.46	valve	96.18
cv	93.72	cv	95.28
solenoid_valve	92.05	solenoid_valve	91.95
mv	90.76	mv	88.34
motor_operated_valve	89.21	actuator	87.78
open	89.00	air_leak	87.67
actuator	88.93	open	87.45
packing	88.88	packing	87.40
valve_stem	88.76	motor_operated_valve	86.20
stem	88.06	valve_stem	86.18
isolation_valve	87.99	ao	86.08
solenoid	87.61	control_valve	86.06
handwheel	87.59	leak	85.95
closed	87.25	solenoid	85.80
stroke_time	87.17	isolation_valve	85.44

7.2.2 Prompt Sensitivity

Generative AI that relies on prompts are especially sensitive to the prompt format (White et al. 2023). This is because these models generate text based on the patterns present in the prompts. When the text is slightly modified, the model can develop a different output. Also, if the prompt is too vague, the model can become stochastic, leading to different answers every time it runs.

Developed Solution 1: Using Words Correlation as a Preliminary Indicator

The field of research known as prompt engineering focuses on optimizing the prompt to the generative AI to produce accurate decisions that can automate human-level decision-making and reasoning tasks. It is hypothesized that proper prompt engineering can improve unsupervised classification using autoregressive models. A new approach was developed, in which the generative AI was asked to generate tens of prompts that can be used to achieve an ultimate objective. Each prompt is tested against a labeled validation set. The hypothesis is that the machine-generated prompts will contain very effective ones, which can be used independently. Also, the generative AI can provide insights into the decision-making process, thereby aiding in the optimization of future prompts.

Demonstration of Solution 1: Failure Labeling of Text

In Section 3.2.2, the classification of logs into shutdown, startup, or failure was discussed. The same use case is discussed in this example to use the previous results as a benchmark. In this example, the Phi-3 mini-LLM was used to create more than 100 different prompts to identify operator log entries that contain equipment failures. The following is the best-performing prompt, which begins with instructions and then asks whether the log entry indicates a failure.

Prompt:

“To identify a failure in an operator log entry at a nuclear power plant, you should follow these steps:

- 1. Read the entire operator log entry carefully, paying close attention to the details and events described.*
- 2. Look for any specific words or phrases that suggest an incident or issue, such as those provided in the initial list (e.g., failure, trouble, not running, fault, etc.).*
- 3. Identify if any equipment or systems are mentioned in relation to these problem indicators.*
- 4. Determine if the issue resulted in a negative outcome, such as equipment not performing its intended function, requiring troubleshooting and repair, or causing an unplanned or unexpected event.*
- 5. Verify if the failure occurred within the critical 24-hour period mentioned in the definition of unreliability, and if it happened during the previous 12 quarters.*
- 6. Cross-reference the log entry with relevant datasets to ensure accurate evaluation and analysis.*

Based on the details provided in the operator log entry, can you confirm if there was an equipment failure at the nuclear power plant? [OPERATOR LOG ENTRY.]”

Response:

“Yes.” or “No”

Since the data were not labeled, the 100 SME-labeled operator logs that were used in Section 3.2.2 were used again. The results are shown in Table 36 and indicate that an F1 score of 70.5% was achieved. This was deemed slightly better than the best generative AI model that was used for the experiment discussed in Section 3.2.2, which achieved an F1 score of 70.1%.

Table 36. Summary of performance results for the autoregressive methods with and without prompt engineering.

Method	Failure F1 Score (%)
Generic ratings dialogue using Vicuna (From Section 3.2.2)	70.1
First prompt engineered model using Phi-3 mini-LLM	70.5
Second prompt engineered model using Phi-3 mini-LLM	76.3

The model made mistakes on about one-quarter of these operator log entries. The false positives and false negatives were evaluated along with the reason that Phi-3 mini-LLM thought these operator log entries were or were not a failure. The SME observed that a common error is the “unexpectedness” of a failure. An unexpected equipment failure is a crucial criterion for making the decision, so the following prompt was developed by the SME using the LLM insight.

Prompt:

“To identify an equipment failure in an operator log entry at a nuclear power plant, follow these steps:

- 1. Carefully read the entire operator log entry to understand the context and details of the event.*
- 2. Determine if any of the text indicates an equipment failure has occurred.*

*A failure is when an equipment stops working or stops running **unexpectedly**.*

If it was intentionally taken out of service, then that is not a failure.

If there is no clear indication that there is an unexpected failure, then do not classify it as a failure.

3. Consider the context and details surrounding the identified keywords. This may include the time and duration of the failure, any actions taken by the operator, and the impact of the failure on the overall system.

4. After analyzing the log entry, conclude whether an equipment failure occurred by evaluating the information gathered from the keywords and their context.

Now, after learning the steps provided above, please indicate whether the following operator log entry represents an equipment failure.

[OPERATOR LOG ENTRY.]

*Is this operator log talking about an **unexpected failure**, planned removal from service, or not a failure. If this is an **unexpected failure** say “yes,” otherwise, “no.”*

Response:

“Yes.” or “No”

This prompt achieved an F1 score of 76.3%, a significant improvement over the 70.5% achieved using the previous prompt. This confirms that such an automated iterative approach to prompt engineering can increase the ability of unsupervised classification to label text.

7.3 Supervised and Unsupervised Classification

In this section, some generic classification challenges are identified and discussed. The methods introduced herein could apply to both supervised and unsupervised classification. The challenges of short text classification and input quality are addressed, with methods developed to overcome both issues.

7.3.1 Short Text

The problem of short text is common across the work management process. Staff often use short text to describe an action or process, and it usually does not provide enough information for a machine to understand the context, nor does it provide enough features for a classifier to be able to make decisions. For example, moving from a CR to a WO usually involves the short text used in the CR, which the scoping phase inherits. To be able to automate scoping decisions, more context is needed. Short text impacts both supervised and unsupervised classifiers, but it is much more severe in unsupervised functions because context is much more impactful in making decisions without the knowledge introduced by labels.

Developed Solution 1: Hybrid Supervised and Unsupervised Models

Hybrid unsupervised and supervised methods for text classification (often called semisupervised methods) combine the strengths of both approaches to address limitations inherent in each method. These methods leverage the strength of unsupervised learning by discovering underlying text patterns and extracting context then use these patterns and context in supervised learning. While dimensionality reduction methods simplify data by reducing them to key significant pieces of information (i.e., features), making the data easier to classify, given the short nature of text, methods to expand the size of the text can be more beneficial. For example, embedding methods can be used to expand on each word in the text. Generative AI tools (e.g., autoregressive methods) can also be used to expand the context of the text for better classification.

Demonstration of Solution 1: Shutdown, Startup, or Failure Example

In Section 3.2.2, the classification of logs into shutdown, startup, or failure was discussed. Autoregressive models were used to inform the classifier of the criteria to make the decisions. In this section, two methods to improve context awareness are discussed using a subset of the data and the same 100 SME labeled logs for validation. The classification methods were expanded by means of transfer learning, using the Word2Vec methods discussed in Section 2.1.4. Every word was expanded to include 0, 5, 10, 20, 40, or 80 of the most similar word vectors before being fed into a linear regression model. Additionally, autoregressive methods were used to elaborate or expand on the text in the operator logs to determine if this would improve the classification process. Both expanded versions of the operator logs were combined with a supervised approach to place weights on the expanded list of words. The explanation prompt follows:

Prompt:

Make an argument that the nuclear power plant operator log is a shutdown and another argument that it is a startup: [TEXT.]

On the other hand, the elaborate prompt focused on getting more context without considering the classification objective:

Prompt:

Please elaborate on this nuclear power plant operator log: [TEXT.]

For this example, ridge regression was the supervised method that was used as it was the best method identified in Section 3.1.2. The results are shown in Table 37. The first-row results show the baseline in which a supervised method was used with no context (0 words). Comparing this to the results with the expansion of each word to 40 words demonstrate that for startup, a 4.1% F1 score improvement was achieved. Using 20 words resulted in a 6.8% F1 score improvement for failure classification. For shutdown, it was found that autoregressive models (Vicuna in this case) resulted in a 2% F1 score improvement using the elaboration prompt. This example demonstrates that more context usually but not always helps as it sometimes causes confusion for the supervised part of the classifier. The optimal level of context varies depending on the nature of the raw text. If the text is already context rich, more context could worsen the model. If the text is context poor, it is usually beneficial to add context.

Developed Solution 2: Classification by Similar Text Classification

A novel approach was developed in this effort to achieve context-aware topic inference. Instead of making decisions based on a single text entry, each entry was grouped with other entries that resemble the single text entry using similarity analysis. The classification was applied to each entry separately. The label assigned to each text entry was then summed in a weighted manner. The resulting label for a text entry then became the topic with the largest summed weights from the group of similar entries.

Demonstration of Solution 2: Events Keyword Assignment

An example is used herein to demonstrate assigning keywords to CRs. Those keywords are used for tracking the nature of events the plant experiences for performance tracking. An analysis was performed to understand whether the context-aware topic inference approach significantly improved the topics alignment with the text for a given set of documents.

Table 37. Summary of performance results for unsupervised classification of three events after word- and context-enhancement.

Method	Shutdown F1 Score (%)	Startup F1 Score (%)	Failure F1 Score (%)	Mean (%)
Extended Logs Using Transfer Learning				
TF-IDF + word vectors + ridge (0 nearest words)	87.3	90.0	73.6	83.6
TF-IDF + word vectors + ridge (5 nearest words)	85.5	93.2	76.5	85.1
TF-IDF + word vectors + ridge (10 nearest words)	86.0	92.5	80.0	86.2
TF-IDF + word vectors + ridge (20 nearest words)	87.0	93.1	80.4	86.8
TF-IDF + word vectors + ridge (40 nearest words)	86.8	94.1	77.9	86.3
TF-IDF + word vectors + ridge (80 nearest words)	88.5	92.8	76.3	85.9
Extended Logs Using Autoregressive Models				
Vicuna explain + TF-IDF + ridge	88.0	83.8	70.1	80.6
Vicuna elaborate + TF-IDF + ridge	89.3	87.7	68.3	81.8

Since this was an unsupervised approach, a benchmark set of results was needed to evaluate the performance. An SME prepared a set of more than 150 CRs by assigning sets of topics to each of the 119 standard topics that were developed in previous efforts (Al Rashdan et al. 2023a, 2023b). In total, there were 500 topics assigned, so each CR had an average of three topics assigned to it by the industry expert. This set of CRs with topics comprised a ground-truth-testing dataset that could be used to empirically gauge the performance of the context-aware and non-context-aware topic inference approaches.

All the 119 standard topics were equally represented in the testing dataset. This is because some of the topics occur much more frequently than others. For instance, “leakage” occurred very frequently and was assigned to 13 CRs, whereas 24 of the 119 topics were assigned to just two CRs, but none had fewer than two assignments.

The results of this experiment demonstrate a significant improvement in topic modeling. However, it was not possible to quantify the improvement because of the assumption that the human labels were the truth. Considering the results of individual topic assignments demonstrated that the SME struggled to select all the right topics and in the right sequence. However, this solution was tested by partnering with a utility on specific text entries that were challenging in the conventional topic modeling. Significant improvement was indicated but not quantified.

It is worth noting that though this method was developed for unsupervised classification, it can also be used for supervised classification.

7.3.2 Input Quality

Data quality is an essential part of any text-based AI. The methods used for cleaning data vary from simple rules-based search and replace (Section 2.1.1), such as removing stop words (e.g., “the,” “of,” etc.); correcting spelling; stemming (e.g., reducing words to their root form); removing special characters or links; extracting abbreviations; to more intelligent forms focused on identifying junk text or

contradicting text. However, a more complex challenge is removing text that does not fit contextually. For example, some text could be so poorly composed by staff that the point being made is not clear, or some fields might be concatenated to the point that they do not make sense. Those examples confuse a classifier.

This work tackled data quality metrics, an active area of research. However, research into this area will continue to evolve beyond the issues that were observed in the available data. Other datasets are expected to have different types of issues.

Developed Solution 1: Finding Junk Based on Frequency

An approach was developed to use clustering to determine how often a text contains outliers. Outliers were searched for by finding the data points that were most distant from the expected text. For instance, a character bigram might be common (even though it can be an abbreviation or more commonly a part of a word) and less likely to be an outlier. However, another character bigram might be rare and likely to be an outlier. If a text contains enough rare character combinations to exceed a certain threshold, it could be junk text.

Demonstration of Solution 1: CR Text Quality

CRs represent the core of several decisions made in the work management process. Therefore, it is essential that they are high quality. CR text quality is the focus of this experiment. This experiment used a CR dataset that contains around 1M records. To detect for junk text, outliers were located by clustering the words and identifying for the data points that were most distant from those expected for a particular CR. In this example, this was done for all the character bigrams in the CRs for demonstration. If there were many infrequent bigrams in a CR, then it was likely an outlier. If there were common bigrams in the CR, then it was likely not an outlier. The probabilities of the character bigrams are shown in Table 38. The table shows that if the CR contains “ab” it is less likely to be an outlier. However, the character bigram “zy” is rare, and if a CR contains “zy” then it is likely an outlier.

Table 38. Probability of bigrams occurrence using TF-IDF.

Bigram	aa	ab	...	&a	&b	...	zy	zz
Mean TF-IDF	8.7e-5	9.7e-4	...	7.6e-7	4.6e-7	...	1.8e-7	9.6e-6

Developed Solution 2: Finding Junk by Part-of-Speech Anomalous Text

An intelligent junk detector was developed in this effort. It did not focus on words but rather on text that does not make sense to a human. Part-of-speech (POS) tags were used for this junk detection method. POS tags are labels assigned to each word in a sentence to indicate its language function (e.g., noun vs. verb, prepositions, etc.). The tags are created by supervised models, and then used in an unsupervised manner. For example, every text should have a verb and a noun; if it does not it is likely of low quality. If a text does not contain a certain percentage of nouns, it is also likely of low quality.

Demonstration of Solution 2: CR Text Quality

The 1M CR records used in the previous example were evaluated in terms of their proportions of POS words. This process identified several junk texts. For example, “[CODE 1] Ch. B failure” was found to be anomalous because it did not include a verb. Technically the verb “occurred” should occur after the word “failure.” Though the machine identified this text as being of poor quality, it did not mark the CR as junk because a human could still understand the implicit word that was left out. Another example is “asdf asdfasdf asdfasdfasdf asdfasdfasdf.” It is not clear why this text was included in the dataset (it is possible staff were testing the system). A third example is “[Code] Pump oil low.” The implicit word

“was” was left out, therefore this CR contained zero verbs. This approach was able to catch junky CRs for some tasks, but given the typical writing style of the industry several false positives were flagged too.

Solution 3: Junk Probability Score

In this solution, the frequency scoring method (solution 1) was combined with the POS tag scoring method (solution 2). If the score of both methods score was high, then the junk probability score was high. But if the scores were a mixture of the two, then an intermediate score was given.

Demonstration of Solution 3: CR Text Quality

Using the same dataset as the previous example, this process resulted in a distribution of probabilities (Figure 10), where a high value indicates a poorly written CR and a value of 0 would be the ideal case. The figure demonstrates that in terms of proportions, most of the CR text seemed coherent. The histogram’s tail to the right identifies junk, which seems minimal. Examples of this tail are:

“*asdf asdfasdf asdfasdfasdf asdfasdfasdf*”

“*cancel cancel cancel*”

“*DELETE DELETE*

“*DELETE*”

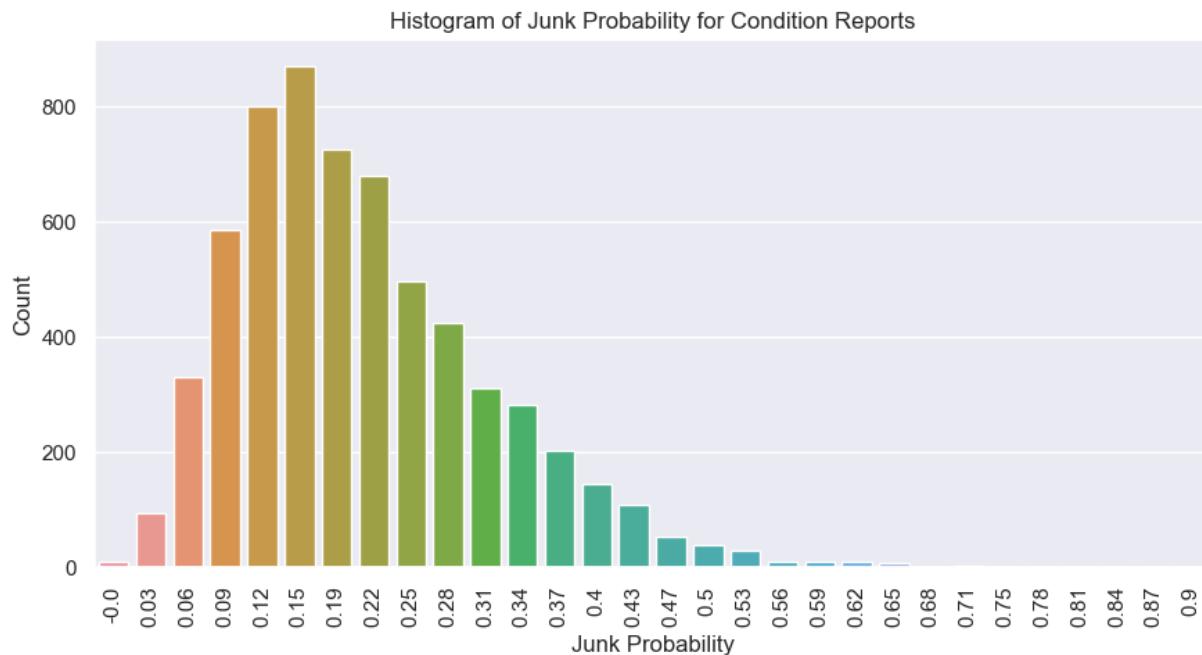


Figure 10. Histogram of the writing quality of CRs evaluated using the junk probability detector.

Developed Solution 4: Similarity to Other Junk Text

Another approach for junk detection that could be used is comparing text against junk text randomly generated by generative AI models. Similarity analysis can be used to compare each text to the random text, and those that are similar enough can be flagged as junky.

Demonstration of Solution 4: CR Example

An experiment was conducted in which an SME came up with several words that have nothing to do with each other (for example, “sky,” “dog,” “car,” “building,” and “burger”). These were combined in one topic. Generative AI was used to generate several variations of such topics, and the resulting topics

were compared against all the CRs. Several junk CRs were detected, but the model missed a lot of the CRs that do not resemble the Junk topic that were created. This method is therefore useful in combination with other methods.

Developed Solution 5: Autoencoders

Autoencoders (discussed in Section 2.3.3.2) are very powerful tools for removing noise, but they can also be used to detect how much noise exists in data. A document can be sent through an autoencoder to get the same output (i.e., it reproduces the document). Once the output is reproduced, the ratio of words lost can indicate the quality of the initial document. The error between the expected output versus the actual output from the network is calculated and used as the metric of the data quality. A poor reconstruction means the model is perplexed and the document could be junk, and a perfect reconstruction indicates the document is not junk.

Demonstration of Solution 5: CR Text Quality

The CRs were processed through a denoising autoencoder called BART. The BART decoder was trained by Meta AI to “(1) corrupt text with an arbitrary noising function, and (2) train a model to reconstruct the original text” (Lewis et al. 2019). For instance, given an input sequence of “Idaho National Laboratory <mask> in nuclear energy,” the target output would be “Idaho National Laboratory is a leader in nuclear energy,” where “<mask>” is replaced with “is a leader.” Examples of the generated output are shown in Table 39. To quantify the performance of the autoencoder, the input was compared to the model’s output by cosine similarity. A limit of 100 tokens was imposed to isolate the length of the CR as a factor. The results show that the cosine similarity decreased when the original text no longer resembled the reconstructed text, and in some cases, this was because the junk CRs could not be reconstructed the same. The machine struggled to make sense of the text like a human would. The similarity of input and output of 100,000 CRs is shown in Figure 11. The histogram shows an expected distribution, with most of the CRs scoring reasonable quality.

Developed Solution 6: Autoregressive Model

This approach relies on generative AI capabilities. An autoregressive model was used for junk detection. The process simply relies on the autoregressive model being clearly asked to identify issues with the data.

Demonstration of Solution 6: CR Text Quality

The same set of CRs used in the previous examples are again used for the experiment herein. The Vicuna LLM model was used to check some of the CRs that were known to be junk and not junk. The following prompt was given with an associated CR that is not junk.

Prompt:

“Is the following nuclear power plant condition report incoherent: ‘[Code1] Tube leak [Code 2] Tube leak [Code 3]’?”

Response:

“No, the condition report is not incoherent. It provides specific information about two tube leaks, one in the [Code 1] and the other in the [Code 2], with the same identifier [Code 3]. This information could be used to track and address the leaks.”

This response indicates that the model hallucinated (i.e., tried to make sense out of unusual text). In another attempt, the following CR was processed.

Prompt:

“Is the following nuclear power plant condition report incoherent: ‘asdf asdfasdf asdfasdfasdf asdfasdfasdf’?”

Response:

“Yes, the condition report is incoherent. It does not provide any meaningful information about the nuclear power plant’s condition.”

This response indicates that the model can determine clearly junky CRs. In this section, six solutions are proposed for analyzing text quality. Typically, a combination of these methods is employed, with most of them likely to successfully identify junky text.

Table 39. Autoencoder reconstruction of example text¹ and the cosine similarity of the original and reconstructed text.

Original Text	Reconstructed Text	Cosine Similarity (%)
asdf asdfasdf asdfasdfasdf asdfasdfasdfasdf	asdf asdfasdfAsdfasdadasdfas dfasdf	16.0
[Code 1] Walkdown - Missing Insulation [Number 1] TB Missing Insulation on [Code 1] @ [Code 2] Photo 8	[Code 1] Walkdown - Missing Insulation [Number 1] Photo 8 Photo 9 Photo 10 Photo 11 Photo 12 Photo 13 Photo 14 Photo 15 Photo 16 Photo 17	29.4
Errors on [Code 1] and [Code 2] On [Code 1] and [Code 3] at coordinates FR-2 are the	Errors on [Code 1] and [Code 2] On 1[Code 1], [Code 2], 2 [Code 4] and 2 [Code 5] at coordinates FR-2 are the	40.4
Ultrasonic M&TE for [Code 1] During performance of WO [Number 1], [Code 1]	Ultrasonic M&TE for [Code 1] During performance of WO [Number 1] WO [Number 1], [Code 2]	49.9
Procedure Enhancement to [Code 1] Section [Number 1] During the performance of [Code 1] Section [Number 1]	Procedure Enhancement to [Code 1] Section [Number 1] During the performance of [Code 2] and the performance thereof, the following procedures were implemented: Section [Number 3] and [Number 4]	60.1
[Code 1] Typo During performance of WO [Code 2], a typo was discovered on STEP 6.3.	[Code 1] Typo During performance of WO [Number 1], during the performance of [Code 3], a typo was discovered on STEP [Number 2].	70.0
Implement [Code 1] for scaffolding stored in Containment during [Code 2]	Implement [Code 1] for scaffolding stored in Containment during [Code 2]	80.0
[Code 1] Outage RP Lessons Learned During the [Code 2] Outage walk downs and inspections performed	[Code 1] Outage RP Lessons Learned During the [Code 2] Outage walk downs and inspections performed	90.0
cancel cancel cancel	cancel cancel cancel	100.0

¹ The text shown in the table was sanitized to preserve the data privacy of the utility

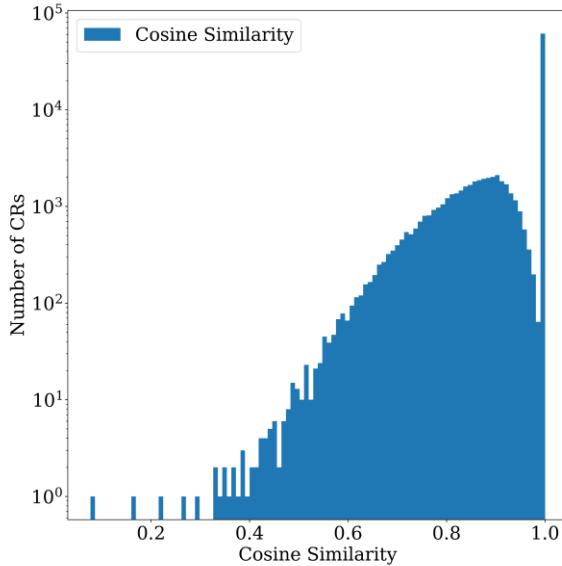


Figure 11. Histogram of the writing quality of CRs evaluated using BART.

7.4 Comparison

Comparison challenges were found to mainly relate to two issues: the comparison of two texts of different size and the use of a proper model for comparison. Solutions to these challenges exist and are discussed herein. Additionally, a third issue—related to extracting the specific similar or differing text—was identified but has not yet been demonstrated.

7.4.1 Different Sizes of Text

Comparison methods are sensitive to the size of the two sets of text that are being compared. If one text is short and the resulting embedding model has a limited context, comparing it to a large document that has several context embeddings might be challenging, if document-based embeddings are used.

Proposed Solution 1: Truncated Text

An approach that has not been demonstrated in the efforts would be to truncate long text into multiple chunks of text that are comparable in size to the short text. This approach involves dividing the text vector into multiple chunks, producing embeddings of each chunk, and then “meshing” these chunks together according to a predetermined amount of overlap. That is, if the embedding model can handle input in sizes up to 512 tokens, and the total length of the text vector is 4,096 tokens, then the text vector could be divided into eight separate chunks, and each chunk could be processed individually. However, since the embedding model is meant to capture context that is dependent on all the tokens in each text, indiscriminately extracting a chunk may mean that essential context in a preceding or following chunk is missed when the chunk embedding is generated. Therefore, a striding approach can be employed in which context from the preceding and following chunks are combined with the current chunk to optimize context extraction. Approaches such as strided chunking (Gong et al. 2020) could be used. The number of tokens included in each stride is typically half the size of the context length, so in the example given, the stride size would be 256 tokens. Therefore, instead of generating embeddings for eight chunks, 15 chunk embeddings would be generated, and the overlapping portions of each chunk would be averaged together to produce a final embedding matrix with 4,096 token-level embeddings.

Proposed Solution 2: Padding and Attention Masking

A very short text vector can be artificially inflated to the required size by appending inert content to the end of the vector. Padding and attention masking are two mechanisms for aiding the effective similarity analysis of short and long texts. Padding adds dummy tokens to shorter texts to force its vector size to match the length of longer texts. Attention masking is then used to focus the model on the desired content, therefore preventing the model from being influenced by the irrelevant padding.

7.4.2 Suitability of Embedding Model

Text comparison relies on an embedding representation of the text that is fed into a model either as an intermediate step in the support of downstream tasks, such as to supply context that is needed for the decoding phase of an autoencoder, or as its final output. Typically, embedding models are a subset of LLMs, and in many cases have a specific goal of using text embeddings that could be different than document comparison. For example, a model designed for context embeddings in classification might not work well for comparison. Therefore, not all embedding models have equal capacity to form usable embeddings for comparison. To be most effective, the model must have been trained specifically to identify similarities between texts.

Proposed Solution 1: Training a Model with Known Similar and Dissimilar Text

For higher accuracy in similarity analysis, it is necessary to consider models that generate embeddings optimized for comparing text. Techniques like Siamese networks (Neculoiu, Versteegh, and Rotaru 2016) can be used. Siamese networks use two or more identical subnetworks (encoders) that share the same weights. They process pairs of texts through these identical encoders to produce embeddings. During training, the model learns to minimize the distance between the embeddings of similar texts (positive pairs) and maximize the distance between embeddings of dissimilar texts (negative pairs). Unlike encoder-decoder models, Siamese networks do not use decoders; they focus solely on producing embeddings for comparison.

A similar concept is used in triplet loss (Novoselov et al. 2018), but instead of two text inputs three are provided: one as the reference or anchor, one that is similar to it (i.e., the positive example), and one that is dissimilar (i.e., the negative example). They use one network for the comparison, which is tuned to minimize the distance between the reference and similar text and maximize the distance to the dissimilar one.

If a domain-specific text or dataset is used, a network that includes a variety of similarity examples can be fine-tuned. This helps the models better represent the domain-specific text to create more accurate embeddings.

7.4.3 Finding Similarities and Differences

Similarity analysis methods typically focus on identifying how similar or dissimilar text pairs are. However, from this analysis, it was found that models often need the actual similarities and differences rather than just evaluating how similar two sets of text are. Preliminary testing demonstrated that generative AI models can assist in this task, but more research is needed in this area and remains for future analysis.

7.5 Dialogue

LLMs designed for question-and-answer (dialogue) tasks can be impacted by several challenges. One such challenge is prompt sensitivity, where the model's responses can vary significantly based on the design of the prompt query. Another issue is hallucination, where the model might incorrectly assume it understands the user's intent when it does not. Additionally, the limited size of prompt inputs can restrict the amount of information exchanged with the LLM. Solutions to these challenges exist and are discussed herein.

7.5.1 Prompt Sensitivity

The sensitivity of dialogue models to the specific structure and vocabulary used in a text prompt is considered one of the main challenges of dialogue models as this prompt text is used to guide their response. Every causal language model contains several internal hyperparameters that can impact the stochasticity and quality of its replies. Depending on the settings of these hyperparameters, the sensitivity of the model to the user's prompt can behave in either more predictable or less predictable ways.

Proposed Solution 1: Dialogue Fine-Tuning

Fine-tuning a dialogue model's hyperparameters with targeted examples can ensure the model will give reasonable responses to varying text prompts that are specifically used during the structured fine-tuning phase. However, depending on the breadth of questions and answers used during the fine-tuning phase, adjustments to the hyperparameters could result in an LLM that responds in terms too generic to be of any legitimate use for a specific knowledge domain. As a result, hyperparameter tuning can be a significant influence when engineering text prompts.

7.5.2 Hallucination

Another major challenge is that a generative AI model can hallucinate to the point that the answer it provides is neither factually accurate nor logically sound. An example was provided in Section 3.4.2. Hallucinations usually happen when a model trained on a broad purview of knowledge is forced to focus on a highly specialized knowledge domain. For example, asking a highly technical question involving the technical specifications of fuel rod cladding will undoubtedly produce a response that is not useful unless the model was specifically trained on nuclear domain knowledge. Hallucination is an area of open research.

Proposed Solution 1: Prompt Engineering

A common approach to enhance the accuracy and relevance of model dialogue output and to reduce hallucination relates to engineering the prompt. Detailed and specific prompts can guide the model toward generating relevant responses. This relevancy can be achieved by iterating over different prompt formats to steer the model away from hallucination. It is also possible to implement constraints in the form of rules to limit the types of responses the model can generate. This can be automated by applying validation rules to a random set of AI-generated prompts to check the relevance and accuracy of the responses before they are presented. These methods can be used directly with autoregressive models or RAG models that were discussed in Section 2.3.3.

7.5.3 Limited Insight

If the LLM model is not tuned to capture specialized knowledge domain, then it may be necessary to provide it insight in the dialogue. However, there is a limit to how much text can be fed to a model at a time. This means that it may be necessary to redact some context from the input prompt, which may result in the LLM providing a response that does not completely satisfy the expectations of the user.

Proposed Solution 1: Using RAG

RAG methods (discussed in Section 2.3.3.5) are a common tool for addressing the input size limitation of LLM models without having to retrain the model with new insight. They operate by incorporating a retrieval system that captures relevant information from a set of documents that represent the needed insight. This approach allows an LLM model to access a broader range of documents and knowledge dynamically.

7.6 Estimation

The main challenge identified in this research regarding estimation is related to missing context or mixed context. The size of text (i.e., too short or too long) used for regression directly correlates with

these context issues used for regression. This finding spurred research into developing methods to achieve optimal context.

7.6.1 Missing Context or Mixed Context

One frequent challenge associated with regression is that text data are too short, which makes it difficult for models to make accurate predictions because the text is ambiguous or lacks the necessary background information. Another challenge is that longer texts are too general or combine confusing contexts, which also makes it difficult for models to make accurate predictions. Both issues result in inconsistent estimates.

Developed Solution 1: Optimizing Text Length

This approach relies on two steps. The first is to determine the regression performance as a function of text length. The second is to expand short text to an optimal length (e.g., see Wang, Yang, and Wei 2023; Nogueira et al. 2019) using methods such as autoregressive models or by trimming long texts to make multiple data points that are regressed separately.

Demonstration of Solution 1: Issue Completion Time

The same experiment used in Section 3.5.2 is used in this example. The aim was to estimate the time it would take to resolve a condition—from the time it originated until it was completed. In this experiment, the goal is to expand on the shortest documents to improve a downstream task. Initially, fields other than the CR core text in the database were used. The model included—"condition," "action taken," "recommendations," "operational notes," "issue notes," "screen notes," "priority," "CAQ," and "criticality"—since they provided more context for the regression model. This resulted in a significant increase in R^2 from 0.33 to 0.52 (Figure 12).

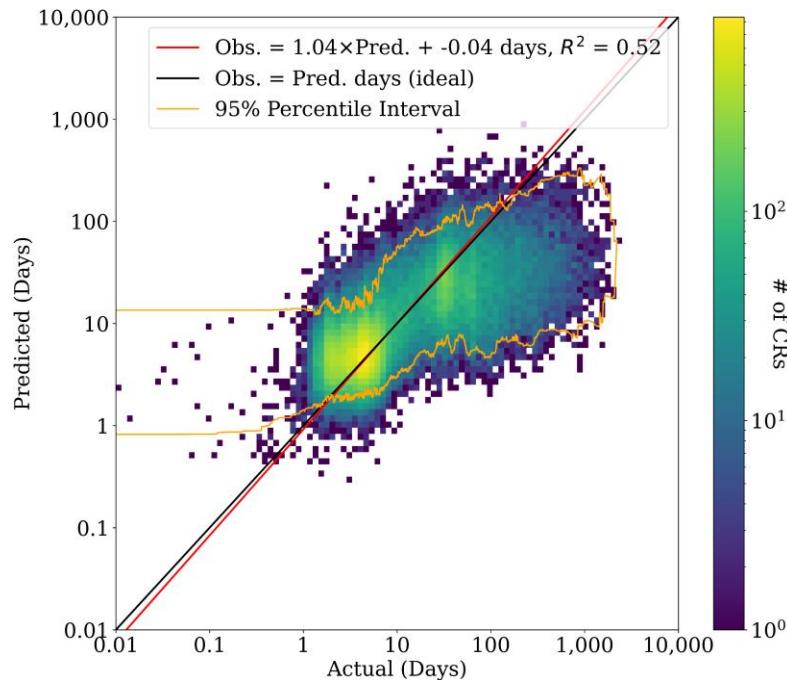


Figure 12. Actual vs. predicted completion time in days after including additional fields.

Figure 13(a) shows that the typical size of the considered CRs. A sliding window of 1,000 CRs was used, which moved across CRs of varying lengths to generate time predictions and R^2 values for each

window. Figure 13(b) illustrates that R^2 results are influenced by the number of tokens; CRs with 100 tokens have an R^2 of around 30%. Yet CRs with 365 tokens (average) have an R^2 of around 58%. The average CR length has nearly double the performance of the shorter CRs. Therefore, it is possible to elaborate on the shortest CRs to increase the predictability of their time prediction. On the other hand, Figure 13(b) shows that longer CRs have lower performance. The 95th percentile is 812 tokens, and CRs above this percentile have the lowest R^2 .

The elaboration of CRs was performed using Phi-3 mini-LLM. The following setup was used:

Prompt:

[CONDITION REPORT] Given the previous condition report, please elaborate.”

Response:

“[ELABORATION.]”

After each CR was elaborated, TF-IDF and ridge regression were used (in a similar manner to Section 3.5.2) to predict the number of days from the origination of a CR until its completion. Five-fold cross validation was performed too. The results are presented in Table 40. The table provides data on the impact of different token thresholds for elaboration on CRs, along with the computational time required. CRs with fewer than 100 tokens each was elaborated to have 100 tokens from an average of 85 tokens. Then the CR token size was incremented by 50 tokens at a time to 250 tokens, then 365 tokens, and the R^2 was given. The table addresses two key questions: First, how does elaboration affect only the CRs that were elaborated? Second, how does elaboration impact overall performance when comparing elaborated and longer CRs to those that were not elaborated?

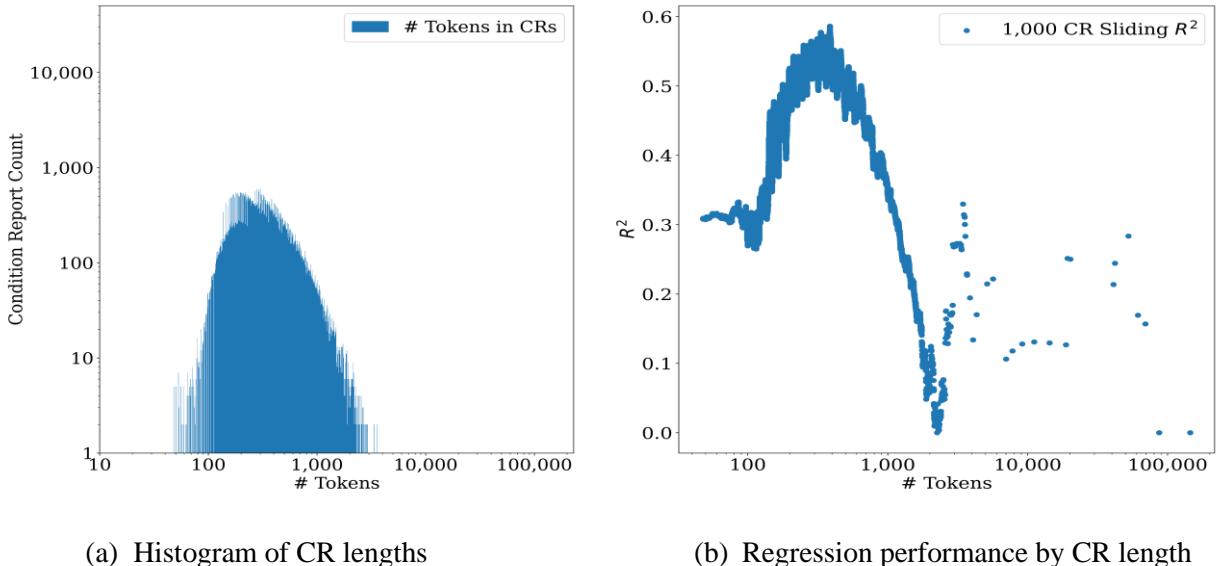


Figure 13. Histogram of CR lengths and regression performance using a sliding window.

In all cases except the 250-token threshold, elaboration led to an increase in R^2 values for the elaborated CRs. In all cases, elaboration resulted in better overall model fitness. However, as the threshold increased, more CRs fell below it, and the elaboration process became computationally expensive. The computational demand appears to grow exponentially for small increases in the elaboration threshold. It is estimated that it would take weeks to calculate the most interesting elaboration threshold of 365 tokens. However, CRs with 365 tokens could result in the highest R^2 . Therefore, a subset of 511 CRs was used as a sample to test this hypothesis. The 511 CRs with fewer than 100 tokens were

used as the baseline (i.e., the same CRs were elaborated in all cases). The original R^2 for those CRs before any elaboration was 32.7% (see Table 41).

Table 40. Impact of text elaboration on model performance and computational time.

Threshold for Elaboration (Tokens)	Number of CRs That Were Elaborated	CRs below Threshold		R^2 (%)	Computational Time Required (Hours)
		R^2 Before Elaboration (%)	R^2 After Elaboration (%)		
Default	0	NA	NA	51.65	0
100	511	32.7	34.6	51.67	0.1
150	6,768	34.3	37.0	51.97	2
200	19,247	40.8	41.8	52.44	6
250	31,626	45.38	45.26	52.59	25
365	NA	NA	NA	NA	Est. Weeks

Table 41. Impact of text elaboration on model performance for 511 CRs with fewer than 100 tokens each.

Threshold for Elaboration (Tokens)	R^2 Before Elaboration (%)	R^2 After Elaboration (%)
100	32.7	34.6
150	32.7	35.4
200	32.7	31.1
250	32.7	28.1
300	32.7	27.2
365	32.7	25.5

Table 41 shows that the elaborated CRs performed worse than the original text except for the CRs with 100 and 150 tokens. The results are also plotted in Figure 14. Unlike Table 40, the optimum R^2 for CRs is no longer about 365 tokens. This indicates that the quality or context of the CRs with 365 human-generated tokens is higher than that of those generated by the LLM. Often, human context is a must-have. For example, consider a scenario in which a WO is recorded with text such as “fix leaking on MSIV valve #3...,” and a duration of 10 days is required to complete it. In a different year, a nearly identical text might require a duration of 15 days for completion. This discrepancy in completion times is a challenge for the model, which cannot be expected to learn a consistent rule for predicting the duration, and the machine making up context would not resolve this discrepancy. The underlying data is inconsistent, which affects the performance of the model. The resulting fitness function after optimizing the size of the CRs is shown in Figure 15, and it demonstrates minimal improvement.

In summary, it is reasonable to conclude that LLMs possess the necessary background information to improve the quality of CRs at NPPs for the purpose of time estimation. However, there appears to be a limit to how much elaboration is useful. This will require more investigation in future research.

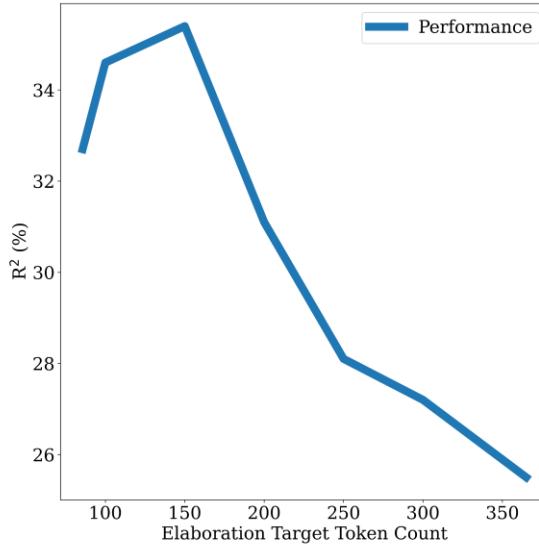


Figure 14. Model performance as a function of token count for 511 CRs with fewer than 100 tokens each.

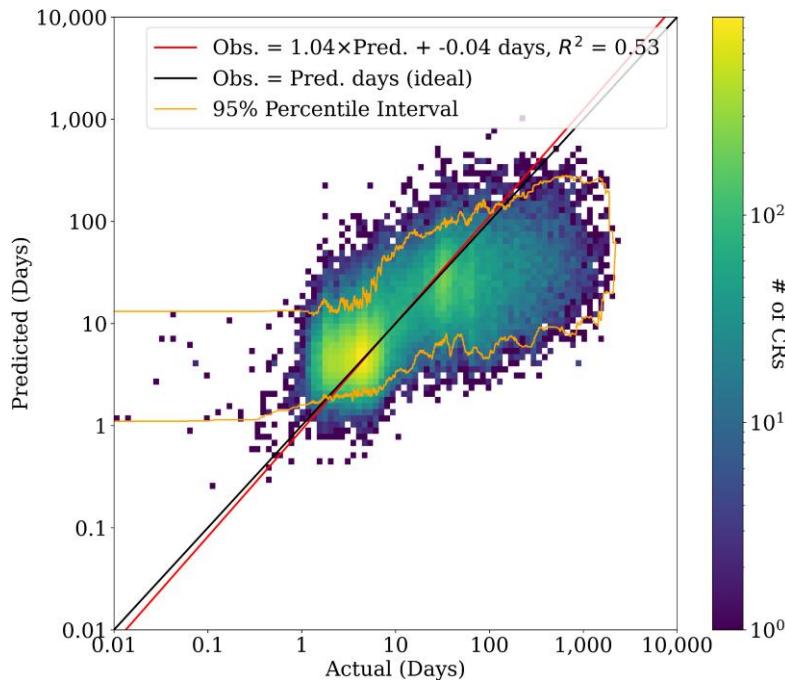


Figure 15. Actual vs. predicted completion time in days after CR size elaboration.

7.7 Generation

Generation often suffers because models lack a clear understanding of what needs to be produced. Similar to humans, these models require substantial information to generate relevant content. This challenge has driven research into the development of methods for creating optimal context to improve the quality and relevance of generated text. Solutions to this challenge exist and are discussed herein.

7.7.1 Irrelevant Text

Using AI to generate new text can sometimes result in irrelevant text due to the AI's limitations regarding understanding context and what the user needs to be generated. This is because the models capture generic knowledge and are biased toward generating text based on what they were trained on. Additionally, these models prioritize generating coherent and human-like text, which is sometimes the focus of the text generation.

Proposed Solution 1: Fine-Tuning the Model for Generation

A model that can generate entirely new content requires a specialized fine-tuning to ensure it can adequately capture the context and relevant facts needed to address the nuances of a highly specialized field, such as NPP operations. It is essential that the model be tuned with many examples of texts from the relevant field to train it to generate text in the desired format. This may necessitate an arduous data-acquisition phase in which thousands of relevant texts are cataloged and appropriate preprocessing is conducted to prepare the texts to be processed through the LLM.

Proposed Solution 2: Providing Example Data of Generated Text

It is also possible for an AI model to generate content based on existing documents retrieved and processed through the RAG method. In this case, the model is provided with sufficient context to generate adequate summaries of documents that resemble the desired new text. Many small excerpts extracted from texts that are relevant to the user's need can be used. The excerpts are presented to the model either individually or in aggregate form. The model must then decide how the excerpts fit together and what sense can be made from them.

7.8 Guidance

Like generation, guidance methods also suffer because models lack a clear understanding of how a model needs to guide the use and how much context is needed. A solution to this challenge is discussed herein.

7.8.1 Confusion

Determining the precise amount of context a machine needs to generate appropriate guidance is challenging. The optimal balance between the amount of information and contextual relevance is unclear. Using excessively long inputs without introducing new context can lead to confusion, making it difficult to gauge the right level of detail required for accurate and meaningful responses. Also, a machine might produce a seemingly counterintuitive guidance that does not help the human.

Proposed Solution 1: Fine-Tuning the Model for Guidance

Providing guidance to a model to ensure a desired response can be helpful for guidance. However, this process must be designed to provide short and valuable text to humans. This will require human-factor evaluation to identify what a human would perceive well, then meticulous fine-tuning so the machine can accurately capture the necessary context and relevant facts. The model would be fine-tuned using a comprehensive dataset of human-developed texts from a specific domain.

7.9 Insertion

Insertion methods are less challenging than other AI function methods because few words are generated to assist the user when typing text. It usually works well for generic text, but when domain-specific text is being inserted by the user, it might not be able to predict relevant words. A solution to this challenge is discussed herein.

7.9.1 Generic Words

As demonstrated in Section 3.8.2, relying on an LLM model for text insertion can be problematic when more than one or two words are needed. Domain-specific knowledge is crucial when predictions depend on a very focused context. Providing only a few words to introduce a highly technical piece and expecting the model to generate relevant text is unlikely to work well. This issue is especially pronounced if the model has not been fine-tuned for the specific field, as its predictions may lack contextual accuracy. For example, if a machine is asked to complete the text “an anomaly has been detected in the functioning of the backup...,” the machine can predict “power supply system” to follow since this is the general words that follow when in a nuclear reactor, but “emergency generator system” would be more suitable in this context.

Proposed Solution 1: Localized Predictors

While it might be useful to refine a model to better predict inserted text, it is possible to train a dedicated autoregressive model for a specific application using historical examples of inserted text. This reduces the possibility of it recommending generic words. The model would only resort to a generic model if its confidence in the predicted word is low.

7.10 Review

Reviewing text is one of the most challenging tasks in text analysis because it requires a machine to thoroughly understand and interpret procedural or requirement documents to evaluate text. Those references are often complex and contain hidden or implied context.

7.10.1 Complex Structure and Confounding Context

Procedural documents often contain complex guidelines and checks that vary and are frequently presented in an unstructured format. The AI must capture and interpret these guidelines, then compare the text against them to determine whether it aligns with the reference. This process involves not only parsing the procedural document but also understanding the context and implications of the requirements. Research in this area is ongoing, with continued work needed to improve and demonstrate effective solutions.

7.11 Structuring

The majority of nuclear data could be classified as unstructured to some extent, given the industry’s reliance on paper-based text (procedures, standards, manual, etc.) for decades. As with review function, structuring unstructured text can be challenging due to complex structuring and confounding context.

7.11.1 Complex Structure and Dispersed Low-Quality Data

Unstructured text lacks a predefined format or organization and often contains information in various formats and artifacts, including diverse language styles, tables, images, and more. Extracting meaningful information from such text requires sophisticated methods to interpret context, handle ambiguity, and identify key entities and relationships. Additionally, unstructured data often necessitates extensive cleaning and preprocessing to remove noise, correct errors, and standardize the text. Research in this area is ongoing, with continued efforts needed to enhance and demonstrate effective solutions.

7.12 Summarization

The key challenge AI models face with text summarization is the quality and breadth of the provided documents they must follow. As with review and structuring functions, the reference text can be challenging due to complex structuring and confounding context.

7.12.1 Misleading the Model

One primary issue in summarization is that providing irrelevant references or references that contain multiple key context areas might confuse the model and mislead it, causing it to summarize irrelevant information. The resulting summary might include unnecessary details or omit crucial points. Excessive or poorly chosen reference material also impacts the coherence and conciseness of the summary and can result in a summary that includes information outside its scope. Achieving seamless integration while maintaining contextual accuracy and coherence is a significant hurdle. Research in this area is ongoing, with continued efforts needed to enhance and demonstrate effective solutions.

7.13 Visualization

Visualizing extensive and complex text data is challenging because humans generally prefer multi-dimensional data presented in 2D or 3D formats that reveal underlying structures. Achieving such visualization often involves making several assumptions.

7.13.1 Data Volume and High Dimensionality

The volume and complexity of data can make it difficult to create visualizations that are both informative and comprehensible. Converting large datasets into graphical formats often risks oversimplifying or distorting the underlying patterns and structures. Additionally, it is not always possible to ensure that visualizations effectively facilitate the exploration and interpretation of the factors driving decisions. Research in this area is ongoing, with continued efforts needed to enhance and demonstrate effective solutions.

7.14 Broader Challenges

This section discusses other challenges that are not specific to a certain AI function. These challenges involved two or more functions and are discussed without a specific AI function in mind.

7.14.1 Obsolescence

New state-of-the-art models are developed nearly every couple of months, and experiments performed with the latest models quickly become obsolete and the models become unavailable. This degrades the ability to make like-for-like comparisons for scientific research. Additionally, if a model is implemented for a task and then that model becomes outdated or unavailable, new models will have to be used. In some cases, the new models offer an improvement; in other cases, the older model may have worked better for a specific task.

7.14.2 Scalability

Each utility has tribal knowledge; for example, a utility may use a certain phrase to describe a specific situation, such as an outage, whereas another will have a different way to describe the same situation. Also, it has been observed that utilities will produce data that are relatively easier to predict for one period, then for another period the data become more difficult to predict. AI models could work well for a specific dataset, but not scale when using other datasets.

7.14.3 Availability and Cost of Computational Power

Nuclear utilities often lack computational power. However, many of the methods discussed in this report—especially the generative AI ones—require large LLMs that are state of the art in performance and often near the limit of what is possible computationally. Therefore, the expense involved with developing this computational infrastructure and staff and keeping up with the latest hardware upgrades can be beyond what is reasonable for a utility. Using cloud-computing resources is possible; however, it might be challenging to use these resources when sensitive information is being used, and they come at a cost, especially when the cloud is used to host large fine-tuned models.

7.14.4 Memory Constraints

The most granular and powerful technique for embedding text is to produce embeddings for each token in every document. Depending on the number of documents this can result in a very large vector database of perhaps millions of entries. It is also possible to partition the document into logical subsections, such as pages perhaps, and then encode the individual pages into text embeddings. The text embeddings for each page are then stored in a database for analysis. The individual pages can then be extracted and used. This approach has the drawbacks of potentially greater memory storage requirements for the vector database, increased processing time for each of the page embeddings, as well as missing across-page context.

8. CONCLUSIONS

This effort successfully demonstrated how AI methods and tools, organized into specific AI functions, can be used to automate decisions and actions in the work management process. The analysis included an overview of the AI methods and the tools that are essential for text preparation. It detailed foundational steps such as tokenization, vectorization, and embedding; followed by supervised learning techniques, such as regression; and unsupervised learning methods, including clustering and topic modeling. Semisupervised learning, combining elements of both supervised and unsupervised methods, was also discussed.

The AI methods and tools were then used in key AI functions relevant for nuclear work management. This included supervised and unsupervised classification, which label operational data to make explicit decisions; comparison for identifying similarities between text and its context; and dialogue systems for real-time human interaction with AI. Estimation predicts continuous outcomes based on historical patterns and was also discussed in this analysis. Other functions covered include generation, which creates new content after AI learns to innovate like a human; guidance for relevant text recommendations; insertion for assisting the human in adding new information; review for unstructured decision-making presented by documents such as procedures or list of requirements; structuring for data organization and extraction; summarization for the quick review of long text; and visualization for making complex data accessible. Together, these functions could enhance operational efficiency, improve decision-making, and ensure compliance.

This research evaluated the readiness of those functions. Supervised classification methods, including embedding models and linear/nonlinear approaches, are ready but could benefit from improved context awareness. Unsupervised classification methods are also ready but need better domain-specific knowledge integration and prompt engineering. Comparison tools are effective for short texts but require further development for use with large documents. Dialogue models are ready for simple interactions but need more research for complex tasks and to reduce hallucinations. Estimation methods need enhancements for handling short text or missing context. Generation models, particularly GANs, are advancing but need to innovate beyond current training datasets. Guidance methods are promising with supervised approaches, but unsupervised methods require more research. Insertion techniques are effective with autoregressive models but need validation for specific knowledge applications. Review capabilities need improvement in understanding complex instructions. Structuring of unstructured data is still in early stages and requires more research into various data formats. Summarization using RAG is effective, but autoregressive models and encoder-decoders need further demonstration for specific tasks. Visualization tools are proven but need further research to refine data transformation techniques.

The AI functions were applied to various work management steps to assess their potential to replace human actions and the resulting cost savings. Human time estimates were assigned to each decision and action to identify the potential for AI to reduce the time and effort required of staff and the resulting cost savings. Quantifying the potential cost savings associated with each AI function offers a clear framework for evaluating the business value and return on investment of AI initiatives. The results showed that there could be potential cost savings across different phases of the work management process. In total, around

\$14.5M per NPP unit per year of cost savings can be accomplished using AI in the work management process. Planning was found to benefit the most from AI, with automation resulting in over 99% cost reduction, representing \$7.7M per NPP unit per year. Several assumptions were made to achieve these numbers. Time estimates and sequence of steps can vary by utility and plant. The potential need for repetition of steps was considered within the existing time estimates. Outlier scenarios were excluded to maintain practical relevance, and not all steps were mapped in detail once the cost of the step diminished. This is to ensure the mapping focus remains on costly decisions and actions. Tangential tasks and outage work were considered outside the scope of this research. Estimates were often rounded to the nearest minute, and any tool used to aid but not fully automate the process was assumed to fully automate the process, resulting in overestimated cost savings. Both simple and complex automation solutions were considered, but simple solutions only were used when possible. Complex solutions were used if simple solutions do not exist.

The top three cost-saving functions relate to unstructured data and decision-making. While review and structuring are primarily used for comprehending data or procedures, generation is mostly associated with taking actions. Unsupervised classification results in lower cost savings because the decisions it applied to a human can typically be made quickly, often taking only a minute. This is not the case with supervised classification, where many decisions depend on explicit or implied labels. Estimation yields low cost savings because the work management step rarely involves predicting values such as time. Guidance and insertion fall at the low end of cost savings, because even if AI issued, humans could still be performing most of the work. It is important to note that the findings of this report are for the current state of AI and the current data structures in utilities. Future technological advancements or plant upgrades and data reorganization could alter the methods used and the potential cost savings.

Various challenges were identified as methods and tools were applied to achieve AI functions. Supervised classification faces issues with sparse, inconsistent, or imbalanced labels, while unsupervised classification struggles with iteration time for new labels and prompt sensitivity. Both types of classification are challenged by short text and input quality. Comparison methods encounter difficulties with text size, embedding model suitability, and identifying similarities and differences. Dialogue systems are affected by prompt sensitivity, hallucination, and limited insight. Estimation methods struggle with missing or mixed context. Generation is challenged by irrelevant text, and guidance and insertion are challenged by confusion and generic words. Review and structuring face issues with complex structures and low-quality data. Summarization risks misleading the model. Broader challenges that were discussed in this research include obsolescence, scalability, availability and cost of computational power, and memory constraints. Several solutions for those challenges were also discussed, with example demonstrations provided for many of them.

Overall, this research underscores the value of optimizing work management processes with AI. Streamlining these processes can reduce operational costs and redirect resources to critical areas. The potential for faster work completion and improved safety measures would contribute to a more efficient and cost-effective operation in the long term.

9. REFERENCES

- Abdi, H., and L. J. Williams. 2010. “Principal Component Analysis.” *Wiley Interdisciplinary Reviews: Computational Statistics* 2 (4): 433–59.
- Abdin, M., et al., 2024. “Phi-3 technical report: A highly capable language model locally on your phone.” arXiv preprint arXiv:2404.14219. <https://arxiv.org/abs/2404.14219>
- Ahmed, M., R. Seraj, and S.M.S. Islam, 2020. “The k-means algorithm: A comprehensive survey and performance evaluation.” *Electronics*, 9(8), p.1295.
- Almazrouei, E., et al. 2023. The falcon series of open language models. arXiv preprint arXiv:2311.16867.
- Al Rashdan, A., J. Browning, and C. Ritter. 2019. “Data Integration Aggregated Model and Ontology for Nuclear Deployment (DIAMOND): Preliminary Model and Ontology.” INL/EXT-19-55610, Idaho National Laboratory.
- Al Rashdan, A., J. Oxstrand, and V. Agarwal. 2016. “Automated Work Package: Conceptual Design and Data Architecture.” INL/EXT-16-38809, Idaho National Laboratory. https://lwrs.inl.gov/Advanced%20IIC%20System%20Technologies/Automated_Work_Package_Conceptual_Design_and_Data_Architecture.pdf.
- Al Rashdan, A., R. Shafer, T. Quinn, M. Bailey, R. Jarrett, and K. Giraud. 2023a. “Regulatory Considerations Regarding the Use of Computer Vision Machine Learning in Safety-Related or Risk-Significant Applications in Nuclear Power Plants.” INL/RPT 23-74920, Idaho National Laboratory.
- Al Rashdan, A., B. Wilcken, N. Mapes, and K. Giraud. 2023b. “Demonstration of an Inspection Data Portal and Data Integration Methods for Nuclear Power Plants.” INL/RPT-23-74611, Idaho National Laboratory.
- Al Rashdan, A., et al. 2020. “Use Cases of DIAMOND for Data-Enabled Automation in Nuclear Power Plants.” INL/LTD-20-59683, Idaho National Laboratory.
- Al Rashdan, A., et al. 2021. “Industry Scaling of Condition Report Data Mining Methods for Automated Compliance.” INL/LTD-21-64306, Idaho National Laboratory.
- Azzopardi, L., M. Girolami, and M. Crowe. 2005. “Probabilistic Hyperspace Analogue to Language,” Paper presented at SIGIR 2005: Proceedings of the 28th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. Association for Computing Machinery: 575–76.
- Church, K. W. 2017. “Word2Vec.” *Natural Language Engineering* 23 (1): 155–62.
- Cover, T., and Hart, P. 1967. “Nearest Neighbor Pattern Classification.” *IEEE Transactions on Information Theory* 13 (1): 21–27.
- Dorogush, A., V. Ershov, and V. Gulin. 2018. “CatBoost: gradient boosting with categorical features support.” arXiv preprint arXiv:1810.11363 <https://catboost.ai/>.
- (2018). Gallagher, R. J., K. Reing, D. Kale, and G. Ver Steeg. 2017. “Anchored Correlation Explanation: Topic Modeling with Minimal Domain Knowledge.” *Transactions of the Association for Computational Linguistics* 5: 529–42.
- Géron, A. 2022. *Hands-on Machine Learning with Scikit-Learn, Keras, and TensorFlow*. 3rd ed. Sebastopol, CA: O'Reilly Media, Inc.

- Gong, H., Y. Shen, D. Yu, J. Chen, and D. Yu. 2020. “Recurrent Chunking Mechanisms for Long-Text Machine Reading Comprehension,” Paper presented at the 58th Annual Meeting of the Association for Computational Linguistics. arXiv preprint arXiv:2005.08056.
- Grefenstette, G. 1999. “Tokenization.” In *Syntactic Wordclass Tagging*, edited by H. van Halteren, 117–33, Text, Speech and Language Technology Series. Dordrecht: Springer.
https://doi.org/10.1007/978-94-015-9273-4_9.
- Grootendorst, M. 2022. “BERTopic: Neural Topic Modeling with a Class-Based TF-IDF Procedure.” *arXiv*: 2203.05794. arXiv:2203.05794v1.
- Hastie, T. 2020. “Ridge Regularization: An Essential Concept in Data Science.” *Technometrics* 62 (4): 426–33.
- Ji, Z., et al. 2023. “Survey of hallucination in natural language generation.” *ACM Computing Surveys*, 55(12), pp.1-38.
- Jiang, A.Q., et al. 2023. “Mistral 7B.” arXiv preprint arXiv:2310.06825. <https://arxiv.org/abs/2310.06825>
- Kabra, M., A. Robie, and K. Branson. 2015. “Understanding Classifier Errors by Examining Influential Neighbors.” In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 3917–25.
- Kasuya, E., 2019. “On the Use of r and r Squared in Correlation and Regression.” *Ecological Research* 34 (1): 235–36.
- Kingma, D. P., and M. Welling. 2013. “Auto-Encoding Variational Bayes.” arXiv preprint arXiv:1312.6114. <https://doi.org/10.48550/arXiv.1312.6114>.
- Lawrence, S. 2023. “Tools and Methods for Optimization of Nuclear Plant Outages.” INL/RPT-23-74439-Rev000, Idaho National Laboratory.
- Lewis, M., et al. 2019. “Bart: Denoising Sequence-to-Sequence Pre-training For Natural Language Generation, Translation, and Comprehension.” arXiv preprint arXiv:1910.13461.
<http://dx.doi.org/10.18653/v1/2020.acl-main.703>.
- Lewis, P., et al. 2020. “Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks.” *Advances in Neural Information Processing Systems* 33: 9459–74.
- Li, J., et al. 2016. “Deep reinforcement learning for dialogue generation”. arXiv preprint arXiv:1606.01541. <https://arxiv.org/abs/1606.01541>.
- Liu, D., and G. Liu. 2019. “A transformer-based variational autoencoder for sentence generation.” 2019 International Joint Conference on Neural Networks (IJCNN). IEEE.
- Liu, T., K. Wang, L. Sha, B. Chang, and Z. Sui. 2018. “Table-to-Text Generation by Structure-Aware seq2seq Learning,” Paper presented at the 32nd AAAI Conference on Artificial Intelligence. AAAI Conference Proceedings 32 (1): 4881–88.
- McDonald, G. C. 2009. “Ridge Regression.” *Wiley Interdisciplinary Reviews: Computational Statistics* 1 (1): 93–100.
- McInnes, L. n.d. “UMAP.” Last accessed August 2022. <https://github.com/lmcinnes/umap>.
- Neculoiu, P., M. Versteegh, and M. Rotaru. 2016. “Learning Text Similarity with Siamese Recurrent Networks.” *Proceedings of the 1st Workshop on Representation Learning for NLP*: 148–57.
- NEI. 2018. “Improving the Effectiveness of Issue Resolution to Enhance Safety and Efficiency.” NEI-16-07, Rev 0, Nuclear Energy Institute.

- NEI. 2023. “Regulatory Assessment Performance Indicator Guideline.” NEI 99-02, Revision 8, Nuclear Energy Institute. <https://www.nrc.gov/docs/ML2329/ML23290A147.pdf>.
- Ng, A. Y. 2004. “Feature Selection, L 1 vs. L 2 Regularization, and Rotational Invariance.” Proceedings of the Twenty-First International Conference on Machine Learning.
- Nogueira, R., et al. 2019. “Document Expansion by Query Prediction.” arXiv:1904.08375. <https://arxiv.org/pdf/1904.08375>.
- Novoselov, S., et al. 2018. “Triplet Loss Based Cosine Similarity Metric Learning for Text-Independent Speaker Recognition.” *Interspeech*: 2242–46.
- Python. 2024. Python Language Reference. Python Software Foundation, Release 3.10.6. <https://docs.python.org/3/reference/index.html>.
- Rahutomo, F., T. Kitasuka, and M. Aritsugi. 2012. “Semantic Cosine Similarity,” Paper presented at the 7th International Student Conference on Advanced Science and Technology ICAST.
- Rai, K. K., et al. 2019. “Question Independent Automated Code Analysis and Grading Using Bag of Words and Machine Learning.” Paper presented at the 2019 International Conference on Computing, Power and Communication Technologies (GUCON). IEEE.
- Ramos, J. 2003. “Using tf-idf to Determine Word Relevance in Document Queries.” Proceedings of the First Instructional Conference on Machine Learning 242 (1): 29–48.
- Rong, X. 2014. “Word2vec Parameter Learning Explained.” arXiv:1411.2738.
- St. Germain, S. W. 2014. “NRC Reactor OE Data.” Proceedings of Probabilistic Safety.
- Schneider, K. M. 2005. “Weighted Average Pointwise Mutual Information for Feature Selection in Text Categorization,” Paper presented at the 9th European Conference on Principles and Practice of Knowledge Discovery in Databases. Knowledge Discovery in Databases: PKDD 2005: 252–63.
- Shen, R., et al. 2016. “Predicting Usefulness of Yelp Reviews with Localized Linear Regression Models.” In 2016 7th IEEE International Conference on Software Engineering and Service Science (ICSESS). IEEE.
- Tong, Z., and H. Zhang. 2016. “A Text Mining Research Based on LDA Topic Modelling.” The 6th International Conference on Computer Science, Engineering, and Information Technology 6 (6): 201–10.
- Vincent, P., H. Larochelle, Y. Bengio, and P. A. Manzagol. 2008. “Extracting and Composing Robust Features with Denoising Autoencoders.” Proceedings of the 25th International Conference on Machine Learning: 1096—1103.
- Wang, L., N. Yang, and F. Wei. 2023. “Query2doc: Query Expansion with Large Language Models.” arXiv:2303.07678. <https://doi.org/10.48550/arXiv.2303.07678>.
- Wang, W., Wei, F., Dong, L., Bao, H., Yang, N. and Zhou, M. 2020. Minilm: “Deep self-attention distillation for task-agnostic compression of pre-trained transformers.” Advances in Neural Information Processing Systems, 33, pp.5776-5788.
- Weng, W-H., et al. 2017. “Medical Subdomain Classification of Clinical Notes Using a Machine Learning-Based Natural Language Processing Approach.” *BMC Medical Informatics and Decision Making* 17: 1–13.
- Wang, Y., Q., Sun, and S. He. 2023. “M3E: Moka Massive Mixed Embedding Model [Computer software].” Hugging Face. <https://huggingface.co/moka-ai/m3e-base>

- White, J., et al. 2023. “A Prompt Pattern Catalog to Enhance Prompt Engineering with Chatgpt.” arXiv:2302.11382. <http://dx.doi.org/10.48550/arXiv.2302.11382>.
- Zheng, L., 2024. “Judging llm-as-a-judge with mt-bench and chatbot arena.” *Advances in Neural Information Processing Systems*, 36.
- Zhang, Y., et al., 2017. “Adversarial Feature Matching for Text Generation.” ICML ’17: Proceedings of the 34th International Conference on Machine Learning 70: 4006–15.
- Zhang, Y., R. Jin, and Z. H. Zhou. 2010. “Understanding Bag-of-Words Model: A Statistical Framework.” *International Journal of Machine Learning and Cybernetics* 1: 43–52.
- Ziwei, J., et al. 2023. “Survey of Hallucination in Natural Language Generation.” *ACM Computing Surveys* 55 (12): 1–38. <https://doi.org/10.1145/3571730>.

Page intentionally left blank

Appendix A

Mapping of Screening

Page intentionally left blank

Appendix A

Mapping of Screening

This appendix applies the approach of Section 4.1 to screening steps according to the flow shown in Figure 16. The following sections discuss each of those steps in detail.

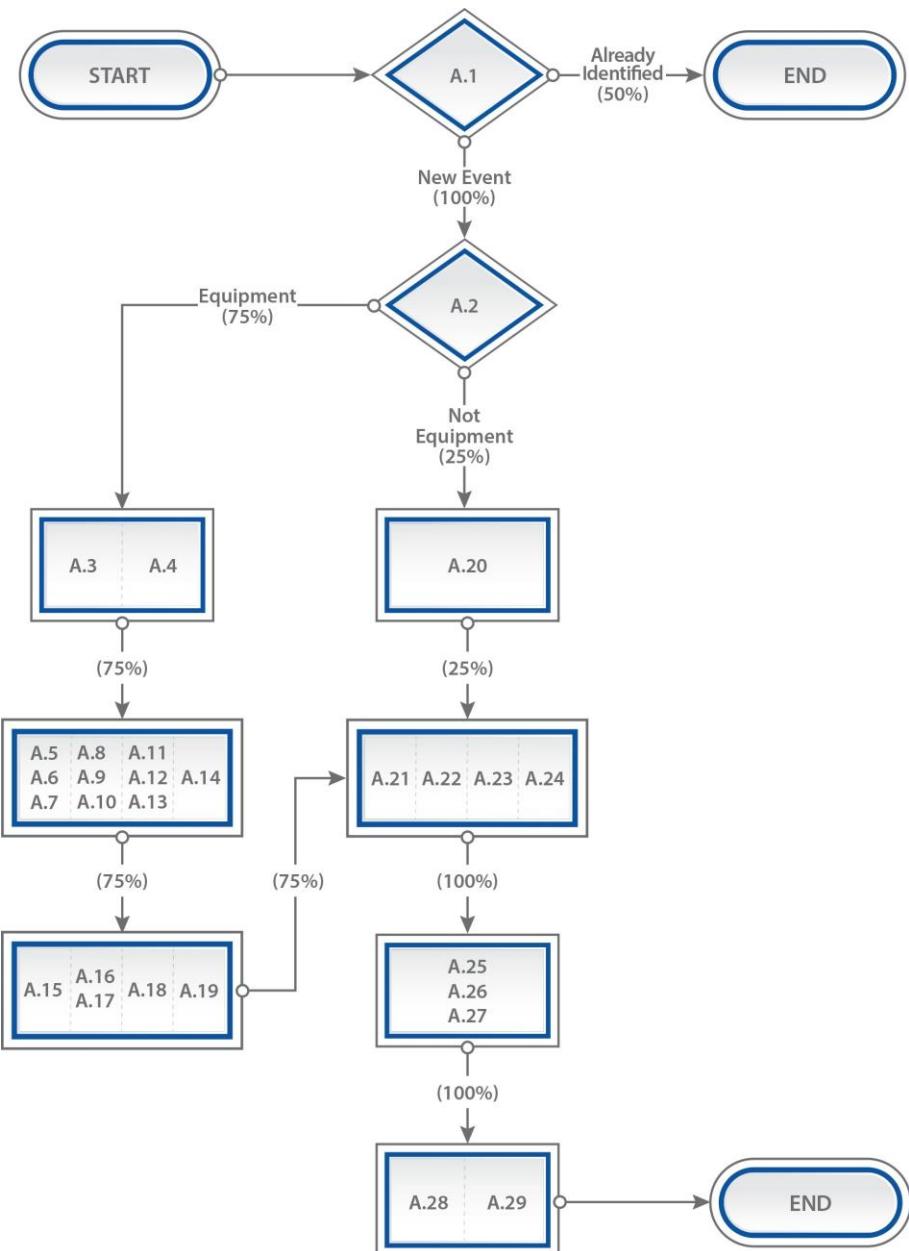


Figure 16. Flowchart of the screening steps that are discussed in Appendix A.

A-1. Issue Identification and Reporting

If the condition has been reported before, then it should not be logged again as it is being handled already. This step aims to determine if this event is already in the system by means of checking for any tags on the equipment or searching through the condition reporting system.

Current Decisions and Conditional Actions

Input: Identified and reported condition.

Was this condition identified before?

Decision	Effort Description	Discipline	Time in Minutes	Frequency Per CR
Process	Check if the condition was previously identified. e.g., there is a tag on the equipment or is in the condition reporting system	Any	3	150%
If yes	Add comments to the existing condition if needed. (End of Process)	Any	2	50%
If no	Write a CR.	Any	5	100%
Total (minutes FTE per CR)				10.5

Automated Decisions and Conditional Actions

Input: Identified and reported condition in text and historical CRs text.

Was this condition identified before?

Decision	AI Description			
	Training/Reference Input	Use Input	Functions	Output
Process	Historical CRs	Identified and reported condition text	Comparison	Yes/No (and previous CR)
If yes	Previous CR	Identified and reported condition text	Generation	Modified CR text
If no	Historical CRs	Identified and reported condition text	Guidance	New CR text
	Historical CRs	Identified and reported condition text	Insertion	New CR text

A-2. Equipment Relevance

A key aspect of events relates to determining if the issue being analyzed relates to equipment. This is necessary as later steps use this information in making decisions.

Current Decisions and Conditional Actions

Input: Current CR.

Is the reported condition related to a piece of equipment?

Decision	Effort Description	Discipline	Time in Minutes	Frequency per CR
Process	Check if the text mentions or refers to any equipment.	Any	0.5	100%
If yes	Insert the equipment ID.	Any	0	75%
If no	None.	NA	0	25%
Total (minutes FTE per CR)			0.5	

Automated Decisions and Conditional Actions

Input: Current CR and engineering master equipment list (MEL).

Is the reported condition related to a piece of equipment?

Decision	AI Description			
	Training/Reference Input	Use Input	Functions	Output
Process	Historical CRs and equipment labels	Current CR	Classification (supervised)	Yes/No
	Equipment words or general-knowledge-enabled LLM	Current CR	Classification (unsupervised)	Yes/No (and Equipment ID)
	MEL equipment description	Current CR	Comparison	Yes/No (and Equipment ID)
If yes	NA			
If no	NA			

A-3. Criticality

In this step, it is desired to determine if the CR relates to equipment that is critical (safety or non-safety) that would impact plant performance and reliability.

Current Decisions and Conditional Actions

Input: Current CR and MEL.

Is this a piece of critical equipment?

Decision	Effort Description	Discipline	Time in Minutes	Frequency per CR
Process	Check if the equipment is identified as critical in the MEL and if the nature of the issue could affect the performance or reliability of the equipment.	Operations	2	75%
If yes	Supervisor notifies shift manager.	Operations	0	5%
If no	None.	NA	0	70%
Total (minutes FTE per CR)			1.5	

Automated Decisions and Conditional Actions

Input: Current and historical CRs, and MEL.

Is this a piece of critical equipment?

Decision	AI Description			
	Training/Reference Input	Use Input	Functions	Output
Process	Historical CRs and criticality labels	Current CR	Classification (supervised)	Yes/No (and criticality label)
	MEL criticality labels	Current CR	Comparison	Yes/No (and criticality label)
If yes	NA			
If no	NA			

A-4. Operability and Impact to Plant

Operations evaluates whether the issue impacts the reliability of equipment that is critical to the operations of the plant. This is different from the criticality analysis, which aims to determine emergent issues that require immediate plant attention. The operability analysis evaluates the longer-term impact to operations of the plant. It is also important to determine if this impacts regulatory compliance by affecting safety-related equipment and technical-specifications requirements.

Current Decisions and Conditional Actions

Input: Current and historical CRs and technical specifications.

Is there an impact to the plant?

Decision	Effort Description	Discipline	Time in Minutes	Frequency per CR
Process	Determine if the issue in the CR affects technical specifications applicability by examining issue for compliance against the technical specifications.	Operations	4*	75%
If yes	Add notes and technical specifications action statement or perform plant configuration Notify regulatory body if needed.	Operations	Out of scope	70%
If no	None.	NA	0	5%
Total (minutes FTE per CR)			3	

* With most CRs, it is a trivial effort to determine whether the work affects technical specification. Some (about 10%), however, require more research (around 40 minutes). This averages to 4 minutes per CR.

Automated Decisions and Conditional Actions

Input: Current and historical CRs and technical specifications.

Is there an impact to the plant?

Decision	AI Description			
	Training/Reference Input	Use Input	Functions	Output
Process	Historical CRs and criticality designation	Current CR	Classification (supervised)	Yes/No
	Technical specifications	Current CR	Review	Yes/No
If yes	NA			
If no	NA			

A-5. Maintenance Validation

Except for operations review, the preceding steps are performed by the individual reporting the condition (i.e., any plant staff). It is next desired for maintenance to validate what was logged by the staff for thoroughness and accuracy. This is performed by the CAP coordinator (CapCo) for the maintenance organization. Validation is performed of the information in each CR.

Current Decisions and Conditional Actions

Input: Current and historical CRs.

Was the CR related to equipment or originated by maintenance characterized correctly?

Decision	Effort Description	Discipline	Time in Minutes	Frequency per CR
Process	Verify the CR entry is comprehensive and accurate	Maintenance	4*	75%
If yes	None.	NA	0	72%
If no	Add notes to clarify and additional insights or recommendations needed to help the screening process.	Maintenance	5	3%
Total (minutes FTE per CR)			3.15	
* In addition to the equipment issues, the maintenance CapCo reviews all events (including those not related to equipment) recorded by maintenance.				

Automated Decisions and Conditional Actions

Input: Current and historical CRs.

Was the CR related to equipment or originated by maintenance characterized correctly?

Decision	AI Description			
	Training/Reference Input	Use Input	Functions	Output
Process	Historical CRs and labels	Current CR	Comparison	Similar CRs for human verification*
	Historical CRs and labels	Current CR	Summarization	Summary of CRs for human verification*
If yes	NA			
If no	Historical CRs	User notes	Guidance	Notes and additional insight
	Historical CRs	User notes	Insertion	Notes and additional insight
* AI can support this activity by finding similar CRs for review, but human review is still needed.				

A-6. Maintenance Compensatory Actions

In some cases, immediate actions are necessary to prevent a condition from worsening. These actions can be as simple as placing a bucket to catch a water leak to prevent an industrial safety incident or securing a panel door to keep dust away from electrical components. This is also performed by the CapCo for maintenance.

Current Decisions and Conditional Actions

Input: Current CR.

Are any immediate or compensatory additional actions needed for all equipment and maintenance-originated CRs?

Decision	Effort Description	Discipline	Time in Minutes	Frequency per CR
Process	Determine if any mitigation is needed for issues that could worsen in time (e.g., leakage).	Maintenance	1	75%
If yes	Take compensatory action.	Maintenance	Out of scope	6%
If no	None.	NA	0	69%
Total (minutes FTE per CR)				0.75

Automated Decisions and Conditional Actions

Input: Current CR.

Are any immediate or compensatory additional actions needed for all equipment and maintenance-originated CRs?

Decision	AI Description			
	Training/Reference Input	Use Input	Functions	Output
Process	NA			
If yes	NA			
If no	NA			

A-7. Maintenance Significance

Every plant has a procedure containing a set of criteria used to determine the significance an event should be assigned. The significance determination is needed to ensure that adequate actions are taken for each CR especially those important to the safe and reliable operation of the plant. The maintenance CapCo assigns this significance based on the criteria.

Current Unconditional Actions

Input: Current CR and CAP procedure.

Assign significance to the equipment related CR or CR that were generated by maintenance.

Action	Effort Description	Discipline	Time in Minutes	Frequency per CR
Process	Compare the condition against the significance criteria in the CAP procedure and assign the appropriate significance.	Maintenance	1	75%
Total (minutes FTE per CR)			0.75	

Automated Unconditional Actions

Input: Current and historical CRs and CAP procedure.

Assign significance to the equipment related CR or CR that were generated by maintenance.

Action	AI Description			
	Training/Reference Input	Use Input	Functions	Output
Process	Historical CRs and significance labels	Current CR	Classification (supervised)	Significance
	CAP procedure	Current CR	Review	Significance

A-8. Operations Validation

Similar to maintenance, it is important for operations to validate the information reported in CRs, especially when it pertains to equipment. As Operations reviews any critical CRs when they are generated, this step involves repeating the maintenance process but with less in-depth investigation.

Current Decisions and Conditional Actions

Input: Current and historical CRs and technical specifications.

Was the CR related to equipment or originated by operations characterized correctly?

Decision	Effort Description	Discipline	Time in Minutes	Frequency per CR
Process	Verify the CR entry is comprehensive and accurate.	Operations	1*	75%
If yes	None.	NA	0	73%
If no	Add notes to clarify and additional insights or recommendations needed to help the screening process.	Operations	5	2%
Total (minutes FTE per CR)			0.85	
* With most CRs, it is a trivial effort to determine whether the CR is characterized correctly. Some (about 30%), however, require more research (around 3 minutes). This averages to 1 minute per CR.				

Automated Decisions and Conditional Actions

Input: Current and historical CRs and technical specifications.

Was the CR related to equipment or originated by operations characterized correctly?

Decision	AI Description			
	Training/Reference Input	Use Input	Functions	Output
Process	Historical CRs and labels	Current CR	Comparison	Most similar CRs for human verification*
	Historical CRs and labels	Current CR	Summarization	Summary of CRs for human verification*
	Technical specifications	Current CR	Review	Any technical specifications issue that results from the CR*
If yes	NA			
If no	Historical CRs	User notes	Guidance	Notes and additional insight
	Historical CRs	User notes	Insertion	Notes and additional insight
* AI can support this activity by finding similar CRs for review, but human review is still needed.				

A-9. Operations Compensatory Actions

As with maintenance, operations evaluates whether any immediate or compensatory actions need to be taken to prevent the condition from worsening. The CapCo for operations performs this evaluation.

Current Decisions and Conditional Actions

Input: Current CR.

Are any immediate or compensatory additional actions needed for all equipment and operations-originated CRs?

Decision	Effort Description	Discipline	Time in Minutes	Frequency per CR
Process	Determine if any mitigation is needed for issues that could worsen in time.	Operations	1	75%
If yes	Consult with operations procedures to identify alternative configurations and take any other necessary action.	Operations	Out of scope	7%
If no	None.	NA	0	68%
Total (minutes FTE per CR)				0.75

Automated Decisions and Conditional Actions

Input: Current CR.

Are any immediate or compensatory additional actions needed for all equipment and operations-originated CRs?

Decision	AI Description			
	Training/Reference Input	Use Input	Functions	Output
Process	NA			
If yes	NA			
If no	NA			

A-10. Operations Significance

Operations utilizes the CAP procedures to determine the significance of the identified issue. The significance determination is to ensure adequate actions are taken to address issues affecting the safety or reliability of the plant. The operations CapCo assigns this significance.

Current Unconditional Actions

Input: Current CR and CAP procedure.

Assign significance to the equipment related CR or CR that was generated by operations.

Action	Effort Description	Discipline	Time in Minutes	Frequency per CR
Process	Using the procedure, compare the condition against the significance criteria (to be provided) and assign the appropriate significance.	Operations	1	75%
Total (minutes FTE per CR)				0.75

Automated Unconditional Actions

Input: Current CR and CAP procedure.

Assign significance to the equipment related CR or CR that was generated by operations.

Action	AI Description			
	Training/Reference Input	Use Input	Functions	Output
Process	Historical CRs and significance labels	Current CR	Classification (supervised)	Significance
	CAP procedure	Current CR	Review	Significance

A-11. Engineering Validation

Similar to maintenance and operations, it is desired for engineering to validate what was reported in a CR if the condition relates to equipment. This step involves repeating the maintenance and operations validation process but with less in-depth investigation.

Current Decisions and Conditional Actions

Input: Current CR and historical WOs.

Was the equipment related CR or CR originated by engineering characterized correctly?

Decision	Effort Description	Discipline	Time in Minutes	Frequency per CR
Process	Verify the CR entry is comprehensive and accurate.	Engineering	1*	75%
If yes	None.	NA	0	74%
If no	Add notes to clarify and additional insights or recommendations needed to help the screening process.	Engineering	5	1%
Total (minutes FTE per CR)			0.80	
* With most CRs, it is a trivial effort to determine whether the CR is characterized correctly. Some (about 10%), however, require more research (around 10 minutes). This averages to 1 minute per CR.				

Automated Decisions and Conditional Actions

Input: Current CR and historical WOs.

Was the equipment related CR or CR originated by engineering characterized correctly?

Decision	AI Description			
	Training/Reference Input	Use Input	Functions	Output
Process	Historical CRs and labels	Current CR	Comparison	Similar CRs for human verification*
	Historical WOs	Current CR	Comparison	Similar WOs for human verification*
	Historical CRs and labels	Current CR	Summarization	Summary of CRs for human verification*
	Historical WOs	Current CR	Summarization	Summary of WOs for human verification*
If yes	NA			
If no	Historical CRs	User notes	Guidance	Notes and additional insight
	Historical CRs	User notes	Insertion	Notes and additional insight
* AI can support this activity by finding similar CRs to review but human review is still needed.				

A-12. Engineering Compensatory Actions

Like operations and maintenance compensatory actions, immediate engineering actions might also be necessary to develop a solution to prevent a condition from worsening. The CapCo for engineering performs this evaluation.

Current Decisions and Conditional Actions

Input: Current CR.

Are any immediate or compensatory additional actions needed for all equipment and engineering-originated CRs?

Decision	Effort Description	Discipline	Time in Minutes	Frequency per CR
Process	Determine if any mitigation is needed for issues that could worsen in time.	Engineering	1	75%
If yes	Take compensatory action.	Engineering	Out of scope	1%
If no	None.	NA	0	74%
Total (minutes FTE per CR)			0.75	

Automated Decisions and Conditional Actions

Input: Current CR.

Are any immediate or compensatory additional actions needed for all equipment and engineering - originated CRs?

Decision	AI Description			
	Training/Reference Input	Use Input	Functions	Output
Process	NA			
If yes	NA			
If no	NA			

A-13. Engineering Significance

Like maintenance and operations, engineering evaluates the significance of the event. The significance determination is needed because a system is important to the operations or generation aspect of the plant. The engineering CapCo assigns this significance.

Current Unconditional Actions

Input: Current CR and CAP procedure.

Assign significance to the equipment-related CR or CR that was generated by engineering.

Action	Effort Description	Discipline	Time in Minutes	Frequency per CR
Process	Using the procedure, compare the condition against the significance criteria (to be provided) and assign the appropriate significance.	Engineering	1	75%
Total (minutes FTE per CR)			0.75	

Automated Unconditional Actions

Input: Current CR and CAP procedure.

Assign significance to the equipment-related CR or CR that was generated by engineering.

Action	AI Description			
	Training/Reference Input	Use Input	Functions	Output
Process	Historical CRs and significance labels	Current CR	Classification (supervised)	Significance
	CAP procedure	Current CR	Review	Significance

A-14. All Others Validation

In addition to maintenance, operations, and engineering, all other organizations will review the CRs that were generated by their organization or related to their discipline. The organization's CapCo will perform this review.

Current Decisions and Conditional Actions

Input: Current and historical CRs.

Was the CR related to the organization or originated by the organization characterized correctly?

Decision	Effort Description	Discipline	Time in Minutes	Frequency per CR
Process	Verify the CR entry is comprehensive and accurate.	All relevant groups	1*	75%
If yes	None.	NA	0	74%
If no	Add notes to clarify and additional insights or recommendations needed to help the screening process.	All relevant groups	5	1%
Total (minutes FTE per CR)			0.80	

* With most CRs, it is a trivial effort to determine whether the CR is characterized correctly. Some (about 10%), however, require more research (around 10 minutes) and is performed by multiple organizations. This averages to 1 minute per CR.

Automated Decisions and Conditional Actions

Input: Current and historical CRs.

Was the CR related to the organization or originated by the organization characterized correctly?

Decision	AI Description			
	Training/Reference Input	Use Input	Functions	Output
Process	Historical CRs and labels	Current CR	Comparison	Similar CRs for human verification*
	Historical CRs and labels	Current CR	Summarization	Summary of CRs for human verification*
If yes	NA			
If no	Historical CRs	User notes	Guidance	Notes and additional insight
	Historical CRs	User notes	Insertion	Notes and additional insight

* AI can support this activity by finding similar CRs for review, but human review is still needed.

A-15. Rework

At this stage maintenance needs to know if this issue is reoccurring (i.e., is a rework). This is different from the previous step in which maintenance was focused on determining if the issue was previously reported.

Current Decisions and Conditional Actions

Input: Current and historical CRs and historical WOs.

Was this worked on before?

Decision	Effort Description	Discipline	Time in Minutes	Frequency per CR
Process	Find similar historical CR and WO records to the CR for the past 2 years.	Maintenance	3	75%
If yes	Perform rework evaluation.	Maintenance	Out of scope	1%
If no	None.	NA	0	74%
Total (minutes FTE per CR)			2.25	

Automated Decisions and Conditional Actions

Input: Current and historical CRs and historical WOs.

Was this worked on before?

Decision	AI Description			
	Training/Reference Input	Use Input	Functions	Output
Process	Historical CRs	Current CR	Comparison	Similar CRs for human verification*
	Historical WOs	Current CR	Comparison	Similar CRs for human verification*
If yes	NA			
If no	NA			

* AI can support this activity by finding similar CRs for review, but human review is still needed.

A-16. Resources for Significant Work

A rough estimate of the necessary resources is needed to determine if this will require significant work and to help determine if the work should be done in-house in the following steps

Current Decisions and Conditional Actions

Input: Current CR.

Is this a significant work?

Decision	Effort Description	Discipline	Time in Minutes	Frequency per CR
Process	Predict the nature of work needed and use it to roughly estimate the resources that will be needed.	Maintenance	3*	75%
If yes	Decide if it can be performed in-house.	Maintenance	0**	3%
If no	None.	NA	0	72%
Total (minutes FTE per CR)		2.25		
* With most CRs, it is a trivial effort to determine whether the work is significant. Some (about 10%), however, require more research (around 30 minutes). This averages to 3 minutes per CR. ** Performed in next step				

Automated Decisions and Conditional Actions

Input: Current and historical CRs and historical WPs.

Is this a significant work?

Decision	AI Description			
	Training/Reference Input	Use Input	Functions	Output
Process	Historical CR assigned WPs and allocated resources	Current CR	Comparison and structuring	Resources available from the WP
If yes	NA			
If no	NA			

A-17. Project/In-house for Significant Work

Based on the resources needed, it is essential to determine early in the process if the job will be performed in-house or if a contractor will be needed, and whether the projects group needs to take over the responsibility of the work.

Current Decisions and Conditional Actions

Input: Current CR.

Is this an in-house job?

Decision	Effort Description	Discipline	Time in Minutes	Frequency per CR
Process	Compare the needed resources for the CR to the available resources and determine if the component is supplier controlled.	Maintenance	1*	75%
If yes	None.	NA	0	74%
If no	Add additional notes on why this is not an in-house job.	Maintenance	1	1%
Total (minutes FTE per CR)				0.76
* With most CRs, it is a trivial effort to determine whether the work can be performed in-house. Some (about 5%), however, require more research (around 20 minutes). This averages to 1 minute per CR.				

Automated Decisions and Conditional Actions

Input: Current and historical CRs and historical WP.

Is this an in-house job?

Decision	AI Description			
	Training/Reference Input	Use Input	Functions	Output
Process	Historical WPs and in-house decision	Current CR	Comparison and structuring	Yes/No
If yes	NA			
If no	Historical CRs	User notes	Guidance	Justification
	Historical CRs	User notes	Insertion	Justification

A-18. Modification/Field Change

Whether the work is conducted in-house or is contracted, it is desired to determine if the work relates to a modification or change, which requires the extensive involvement of engineering and projects.

Current Decisions and Conditional Actions

Input: Current CR, manuals, historical WPs, drawings, and ECs.

Is a modification needed?

Decision	Effort Description	Discipline	Time in Minutes	Frequency per CR
Process	Determine if the issue can be resolved by the type of work that is possible or an equipment modification is needed.	Engineering	1*	75%
If yes	Recommend the level of change needed and initiate the EC request.	Engineering	1	3%
If no	None.	NA	0	72%
Total (minutes FTE per CR)			0.78	

* With most CRs, it is a trivial effort to determine whether modification is needed. Some (about 5%), however, require more research (around 20 minutes). This averages to 1 minute per CR.

Automated Decisions and Conditional Actions

Input: Current and historical CRs, historical WPs, manuals, drawings, and ECs.

Is a modification needed?

Decision	AI Description			
	Training/Reference Input	Use Input	Functions	Output
Process	Historical CRs assigned WP, and modification label	Current CR	Comparison and Structuring	Yes/No
	Historical CRs assigned WP and ECs presence	Current CR	Comparison	Yes/No
	Manuals and drawings	Current CR	Review	Summary of modifications needed*
If yes	Historical CRs and assigned WP's modification label	Current CR	Comparison and classification (supervised)	Modification level
If no	NA			

* AI can support this activity, but human review is still needed.

A-19. Safety Significance

Safety-significant events have been considered as part of criticality and severity decisions, but in this step, they are explicitly flagged as they are tracked by multiple organizations, including the plant risk management staff, licensing group, and performance improvement organization.

Current Decisions and Conditional Actions

Input: Current CR and safety equipment list.

Is this a safety significant event?

Decision	Effort Description	Discipline	Time in Minutes	Frequency per CR
Process	Determine if the issue relates to any safety equipment and if the type of issue impacts the safety function of the equipment.	PI	2	75%
If yes	Flag the CR as CAQ.	PI	0	15%
If no	None.	NA	0	60%
Total (minutes FTE per CR)			1.5	

Automated Decisions and Conditional Actions

Input: Current and historical CRs and safety equipment list.

Is this a safety significant event?

Decision	AI Description			
	Training/Reference Input	Use Input	Functions	Output
Process	Historical CRs and equipment safety labels	Current CR	Classification (supervised)	Yes/No
	Safety equipment list	Current CR and equipment ID	Comparison	Yes/No
If yes	NA			
If no	NA			

A-20. Administrative Impact

Aside from issues directly affecting plant performance, there are other issues that the CRs need to be screened for. In this step, the reported issue is screened for any other potential administrative issue.

Current Decisions and Conditional Actions

Input: Current CR and human resources (HR), security, and information technology (IT) procedures.

Does the condition have any administrative impact?

Decision	Effort Description	Discipline	Time in Minutes	Frequency per CR
Process	Determine if the issue relates to HR, security, or IT procedures and consult relevant procedures if needed.	Administration	5	25%
If yes	Assign to organization.	Administration	0	3%
If no	Add notes.	Administration	0	22%
Total (minutes FTE per CR)			1.25	

Automated Decisions and Conditional Actions

Input: Current and historical CRs and HR, security, and IT procedures.

Does the condition have any administrative impact?

Decision	AI Description			
	Training/Reference Input	Use Input	Functions	Output
Process	Historical CRs and action or department labels	Current CR	Classification (Supervised)	Yes/No (and specific impact)
	HR, security, IT words or general-knowledge-enabled LLM	Current CR	Classification (unsupervised)	Yes/No
	HR, security, and IT procedures	Current CR	Review	Yes/No (and specific impact)
If yes	NA			
If no	NA			

A-21. Radiological Impact

If the identified issue could result in an unplanned radiation release or unplanned dose or could impact the plant's ALARA goals, then the RP group needs to evaluate which actions are necessary to mitigate any radiological impact.

Current Decisions and Conditional Actions

Input: Current CR and radiological survey documentation.

Is there an ALARA impact (increase dose rate) or contamination?

Decision	Effort Description	Discipline	Time in Minutes	Frequency per CR
Process	Determine if the condition relates to a high rad area, involve any release of radiological materials (evident by radiological survey documentation) and contamination, or an increase of dose. Check if the condition impacts the radiological survey data.	RP and chemistry*	2	100%
If yes	RP and chemistry conduct surveys or decontamination.	RP and chemistry	Out of scope	5%
If no	None.	NA	0	95%
Total (minutes FTE per CR)			2	

* The reason both organizations are listed here is because they often have one shared CapCo.

Automated Decisions and Conditional Actions

Input: Current CR and radiological survey documentation.

Is there an ALARA impact (increase dose rate) or contamination?

Decision	AI Description			
	Training/Reference Input	Use Input	Functions	Output
Process	Radiological survey words or general-knowledge-enabled LLM	Current CR	Classification (unsupervised)	Yes/No
	Radiological survey documentation	Current CR	Review	Yes/No (and explanation)
If yes	NA			
If no	NA			

A-22. Chemistry Impact

If the condition could result in an issue such as an unplanned chemical release which could affect plant staff or the environment, then the plant chemistry department needs to evaluate the issue to ensure regulations are not violated and to take action to mitigate the issue. Sampling might be needed to determine the extent of the condition.

Current Decisions and Conditional Actions

Input: Current CR, regulations, data sheets, and sampling results.

Did the chemistry of the plant equipment or environment change?

Decision	Effort Description	Discipline	Time in Minutes	Frequency per CR
Process	Determine if the condition relates to any release of chemical materials (evident using sampling), or if it violates regulatory requirements.	RP and chemistry	5	100%
If yes	RP and chemistry conduct sampling and cleanup, or update data sheets.	RP and chemistry	Out of scope	3%
If no	None.	NA	0	97%
Total (minutes FTE per CR)			5	

Automated Decisions and Conditional Actions

Input: Current CR and chemistry regulations, manuals, datasheets, and sampling results.

Did the chemistry of the plant equipment or environment change?

Decision	AI Description			
	Training/Reference Input	Use Input	Functions	Output
Process	Chemistry words or general-knowledge-enabled LLM	Current CR	Classification (unsupervised)	Yes/No (relates to chemistry) *
	Chemistry regulations, manuals, datasheets, and sampling results	Current CR	Review	Yes/No (and explanation)
If yes	NA			
If no	NA			

* AI can support this activity, but human review is still needed.

A-23. Severity

Severity relates to the consequences of the event to the plant; it is different than criticality, which is focused on the importance of the event. Severity is a key metric that PI is interested in tracking to ensure plant performance is sustained at satisfactory levels.

Current Decisions and Conditional Actions

Input: Current CR and work management procedure.

Is this a severe event?

Decision	Effort Description	Discipline	Time in Minutes	Frequency per CR
Process	Determine if the issue relates to a high or moderate severity based on the work management procedure. Severe events relate to safety-related equipment, events that could impact ALARA, or events that indicate a broader problem in the plant.	PI	3	100%
If yes	Add notes to justify the ranking.	PI	1	10%
If no	None.	NA	0	90%
Total (minutes FTE per CR)		3.1		

Automated Decisions and Conditional Actions

Input: Current and historical CRs and work management procedure.

Is this a severe event?

Decision	AI Description			
	Training/Reference Input	Use Input	Functions	Output
Process	Historical CRs and severity labels	Current CR	Classification (supervised)	Severity level
	Work management procedure	Current CR	Review	Severity level
If yes	Historical CRs	User notes	Guidance	Justification
	Historical CRs	User notes	Insertion	Justification
If no	NA			

A-24. Actions Needed

Each CR is evaluated to determine whether further investigation or evaluation is necessary, or whether additional corrective actions need to be taken depending on the severity of the issue. For example, the condition may require a HP evaluation or improvement or a root cause evaluation.

Current Decisions and Conditional Actions

Input: Current CR and CAP procedure.

Is an action needed? What type?

Decision	Effort Description	Discipline	Time in Minutes	Frequency per CR
Process	Determine if the event relates to the scope of one of the action categories in the work management procedure. Assign that action and the appropriate department.	PI	2*	100%
If yes	Assign the action to a department.	PI	0	5%
If no	None.	NA	0	95%
Total (minutes FTE per CR)			2	
* With most CRs, it is a trivial effort to determine if an action is needed. Some (about 10%), however, require more research (around 20 minutes). This averages to 2 minutes per CR.				

Automated Decisions and Conditional Actions

Input: Current CR and CAP procedure.

Is an action needed? What type?

Decision	AI Description			
	Training/Reference Input	Use Input	Functions	Output
Process	Historical CRs and action or department labels	Current CR	Classification (supervised)	Yes/No (and action type)
	CAP procedure	Current CR	Review	Yes/No (and action type explanation)
If yes	NA			
If no	NA			

A-25. Prescreening Meeting

After all relevant groups have had a chance to provide input, they all meet to evaluate the decisions made collectively. This step does not apply to all utilities as some do not perform a prescreening meeting.

Current Unconditional Actions

Input: Current CR.

Provide any feedback and validate the individual feedback.

Action	Effort Description	Discipline	Time in Minutes	Frequency per CR
Process	Discuss in meeting all decisions made	All relevant groups	6*	100%
Total (minutes FTE per CR)				6

* Assuming 5 staff meet for 30 minutes and 25 CRs per day, this would result in 6 minutes per CR.

Automated Unconditional Actions

Input: Current CR.

Provide any feedback and validate the individual feedback.

Action	AI Description			
	Training/Reference Input	Use Input	Functions	Output
Process	NA			

A-26. Screening Meeting Preparation

Some staff might have actions to take to prepare for the screening meeting. Those actions could include reviewing the decisions made or providing more information. The PI staff is responsible for tracking all actions and making sure they are performed.

Current Decisions and Conditional Actions

Input: Current CR.

Has all the feedback from the prescreening meeting been incorporated into the screening package?

Decision	Effort Description	Discipline	Time in Minutes	Frequency per CR
Process	Incorporate prescreening meeting updates.	PI	2	100%
If yes	No further actions.	PI	0	67%
If no	Make updates to the screening meeting package.	PI	15	33%
Total (minutes FTE per CR)				6.95

Automated Decisions and Conditional Actions

Input: Current CR.

Has all the feedback from the prescreening meeting been incorporated into the screening package?

Decision	AI Description			
	Training/Reference Input	Use Input	Functions	Output
Process	NA			
If yes	NA			
If no	NA			

A-27. Screening Meeting

After all relevant groups have had a chance to update the input, they all meet to evaluate the decisions made collectively to prepare the WR, if applicable.

Current Unconditional Actions

Input: Current CR.

Provide any feedback and validate the individual feedback.

Action	Effort Description	Discipline	Time in Minutes	Frequency per CR
Process	Discuss in meeting the decisions made	All relevant groups	6*	100%
Total (minutes FTE per CR)		6		

* Assuming 5 staff meet for 30 minutes and 25 CRs per day, this would result in 6 minutes per CR.

Automated Unconditional Actions

Input: Current CR.

Provide any feedback and validate the individual feedback.

Action	AI Description			
	Training/Reference Input	Use Input	Functions	Output
Process	NA			

A-28. Condition Report Update and Work Request Notification

After the screening meeting, the PI staff ensures that all necessary actions in the screening process are completed and a WR is initiated if needed.

Current Decisions and Conditional Actions

Input: Current CR.

Have all the CR decisions been updated in the CR system and the actions created?

Decision	Effort Description	Discipline	Time in Minutes	Frequency per CR
Process	Ensure all needed fields of the CR have been addressed and incorporate any prescreening meeting recommendations into the CR and WR. Evaluate the screening findings and determine if a WR is needed.	PI	2	100%
If yes	Create the WR and populate it with all critical information from the CR.	PI	0	45%
If no	Contact departments for actions and/or updates.	PI	15	55%
Total (minutes FTE per CR)			10.25	

Automated Decisions and Conditional Actions

Input: Current CR and CAP procedure.

Have all CR decisions been updated in the CR system and actions created?

Decision	AI Description			
	Training/Reference Input	Use Input	Functions	Output
Process	CAP procedure	Current CR	Review	Yes/No (and explanation)
If yes	NA			
If no	NA			

A-29. Maintenance Work Order Creation

Using the WR, create a WO to provide detailed instructions and allocate resources for the completion of the identified work.

Current Decisions and Conditional Actions

Input: Current CR and WR (notifications).

Does the WR require a WO?

Decision	Effort Description	Discipline	Time in Minutes	Frequency per CR
Process	Determine if the WR and associated condition require maintenance-related work.	Maintenance	1	45%
If yes	Initiate a WO in the WO system.	Maintenance	1	18%
If no	None.	NA	0	27%
Total (minutes FTE per CR)			0.63	

Automated Decisions and Conditional Actions

Input: Current and historical WRs (notifications) and historical WOs.

Does the WR require a WO?

Decision	AI Description			
	Training/Reference Input	Use Input	Functions	Output
Process	Historical WRs and WOs assignment	Current WR	Classification (supervised)	Yes/No
	Historical WOs	Current WR	Comparison	Yes/No
If yes	NA			
If no	NA			

Appendix B

Mapping of Scoping

Page intentionally left blank

Appendix B

Mapping of Scoping

This appendix applies the approach of Section 4.1 to scoping steps according to the flow shown in Figure 17. The following sections discuss each of those steps in detail.

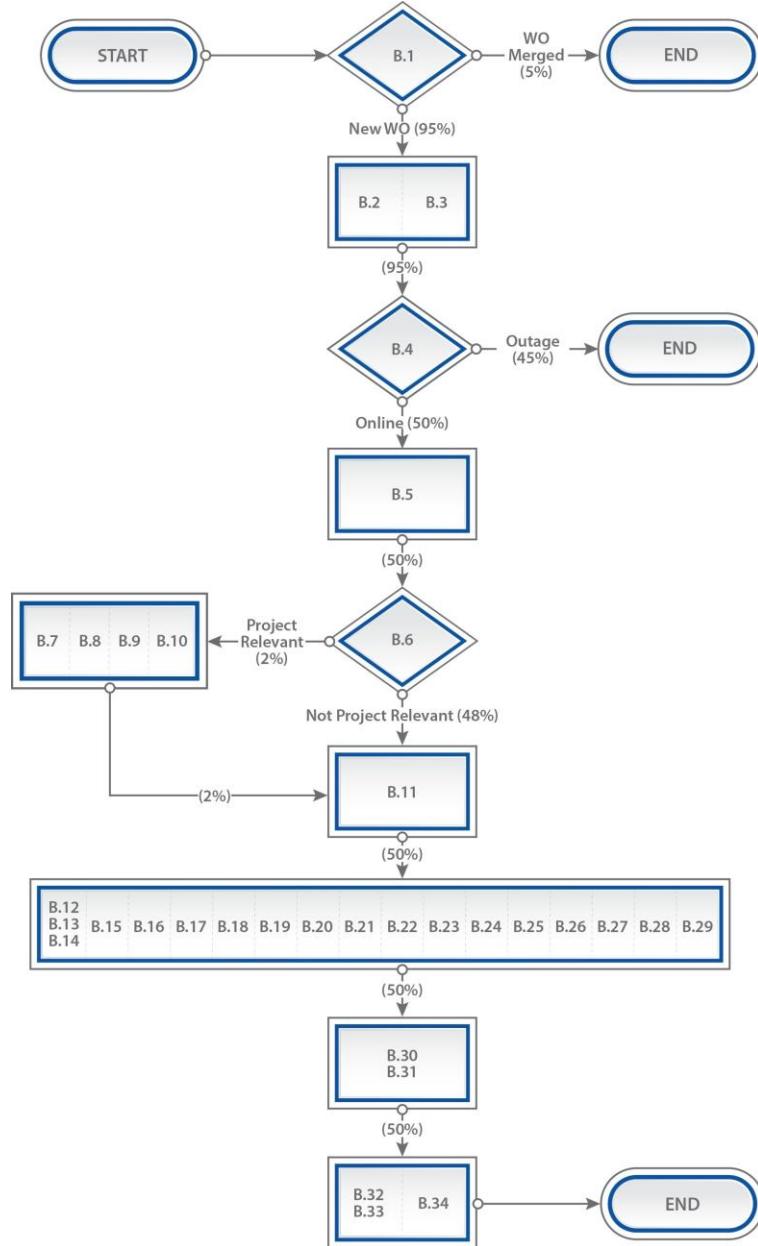


Figure 17. Flowchart of the scoping steps that are discussed in Appendix B.

B-1. Merger of Work Orders for Same Issue

Whether the WO was initiated by maintenance, projects, or PM, it is desired to first determine if this new WO should be merged with an existing one (i.e., to combine multiples WOs in one WO) or if a new one should be created. This is done because the same equipment or system should be worked on at the same time to reduce equipment unavailability.

Current Decisions and Conditional Actions

Input: Current WO and other existing WOs.

Are there multiple WOs that can be merged into one WO?

Decision	Effort Description	Discipline	Time in Minutes	Frequency per WO
Process	Determine if the new WO should be merged with an existing WO for the same system/component by verifying whether the defects described within the new WO will be addressed during the execution of the work for the existing WO.	Maintenance senior reactor operator (SRO)*	5	100%
If yes	Merge new WO into an existing WO.	Maintenance SRO	5	5%
If no	Keep the new WO.	Maintenance SRO	0	95%
Total (minutes FTE per WO)				5.25
* A Maintenance SRO is a maintenance staff member who is qualified and licensed to operate the plant, enabling the individual to scope work with insights from both maintenance and operations perspectives.				

Automated Decisions and Conditional Actions

Input: Current WO and other existing WOs.

Are there multiple WOs that can be merged into one WO?

Decision	AI Description			
	Training/Reference Input	Use Input	Functions	Output
Process	Existing WOs	Current WO	Comparison	Yes/No (and similar WO)
If yes	Similar WO	Current WO	Generation	Modified existing WO
If no	NA			

B-2. Work Order Priority

Priority is used to indicate urgency. This step validates the WR-provided priority by determining if it impacts technical specifications, ALARA, or other characteristics that would impact priority. This priority is assigned based on work management procedures, which list the criteria that are used to assign priority.

Current Unconditional Actions

Input: Current and historical WOs, and the work management procedure.

What priority level should be assigned to the WO?

Action	Effort Description	Discipline	Time in Minutes	Frequency per WO
Process	Assign a priority level to the WO based on the current work scope, the priority assigned to similar previous WOs, and work management procedure criteria.	Maintenance SRO	3	95%
Total (minutes FTE per WO)			2.85	

Automated Unconditional Actions

Input: Current and historical WOs, and the work management procedure.

What priority level should be assigned to the WO?

Action	AI Description			
	Training/Reference Input	Use Input	Functions	Output
Process	Historical WOs and priority labels	Current WO	Classification (supervised)	Priority
	Work management procedure	Current WO	Review	Priority

B-3. Work Order Criticality

Criticality is an indicator of importance (not urgency like priority). It is usually assigned based on the equipment being worked on. Equipment that relates to critical functions in the plant, specifically safety-related equipment listed in the technical specifications, is usually assigned high criticality. There are also other circumstances for which equipment criticality might be increased, and these are usually detailed in the work management procedure. This label is usually listed in the MEL.

Current unconditional Actions

Input: Current and historical WOs, MEL, and the work management procedure.

What criticality level should be assigned to the WO?

Action	Effort Description	Discipline	Time in Minutes	Frequency per WO
Process	Use the criticality of the equipment being worked on from the MEL, the criticality assigned to similar previous WOs, and current work scope and work management procedure criteria to assign criticality.	Maintenance SRO	3	95%
Total (minutes FTE per WO)			2.85	

Automated unconditional Actions

Input: Current and historical WOs, MEL, and the work management procedure.

What criticality level should be assigned to the WO?

Action	AI Description			
	Training/Reference Input	Use Input	Functions	Output
Process	Historical WOs and criticality labels	Current WO	Classification (supervised)	Criticality
	MEL criticality labels	Current WO	Comparison	Criticality
	Work management procedure	Current WO	Review	Priority

B-4. Outage or Online Work

Whether or not work is performed as outage work depends on the nature of issue, the impacted equipment, and the associated risk. If the work is to be performed during an outage, it follows a different schedule than online work and is considered out of scope in this effort. Thus, at this stage it is desired to determine whether the work fits in the outage or online (daily) schedule.

Current Decisions and Conditional Actions

Input: Current and historical WOs and technical specifications.

Is this outage work?

Decision	Effort Description	Discipline	Time in Minutes	Frequency per WO
Process	Review the technical specifications and look for any requirements indicating the equipment must be worked during an outage. Search for similar WOs for the same or similar equipment to see if the work was done in an outage.	Maintenance SRO	2*	95%
If yes	Select an outage to scope the work into by determining the minimum lead time for work and deciding which is the closest outage in which to perform it. (End of process)	Maintenance SRO	Out of scope	45%
If no	Select a workweek to scope the work into.	Maintenance SRO	5	50%
Total (minutes FTE per WO)		4.4		

* With most WOs, it is a trivial effort to determine whether the work is outage work or online work. Some (about 10%), however, require more research (around 20 minutes). This averages to 2 minutes per WO.

Automated Decisions and Conditional Actions

Input: Current and historical WOs, technical specifications, and historical outage dates.

Is this outage work?

Decision	AI Description			
	Training/Reference Input	Use Input	Functions	Output
Process	Historical WOs and outage labels (using outage dates)	Current WO	Classification (supervised)	Yes/No
	Technical specifications	Current WO	Review	Yes/No (and explanation)
If yes	NA			
If no	Historical WOs and time to execution	Current WO	Estimation	Lead time needed

B-5. Preliminary Work Schedule

At this point, work is typically given a provisional placeholder in the schedule (a specific workweek for online work). Then the rest of the scoping process is completed, and the work may either continue to stay in that spot in the schedule or be moved to a more suitable spot at any time during the process. The goal is that by the end of the scoping process, all relevant factors will have been considered and the WO is positioned in a suitable spot in the schedule based on all factors.

Current Unconditional Actions

Input: Current WO.

Add WO to a preliminary position in the schedule.

Action	Effort Description	Discipline	Time in Minutes	Frequency per WO
Process	Estimate needed lead time and glance at the work schedule. Input the WO in the scheduling software.	Cycle planning*	5	50%
Total (minutes FTE per WO)				2.5

* Cycle planning is part of the work management organization.

Automated Unconditional Actions

Input: Current and historical WOs.

Add WO to a preliminary position in the schedule.

Action	AI Description			
	Training/Reference Input	Use Input	Functions	Output
Process	Historical WOs time to execution	Current WO	Estimation	Usual time to execute the work

B-6. Project Relevance

If the WO was initiated by the project team specifically to complete project work, then obviously the WO is related to that specific project. However, there may be other WOs that relate to a project work even if they were not generated by the project team, and they should be performed as part of the overall project. For example, if there is a project to replace a large valve with a new style of valve, and an issue with the valve controller has been reported separately, then perhaps the scope of work for the controller can be absorbed into the overall project.

Current Decisions and Conditional Actions

Input: Current WO and active projects list.

Should this WO be included with planned project work?

Decision	Effort Description	Discipline	Time in Minutes	Frequency per WO
Process	Search the project database and determine if this WO should be included in the work scope for an existing project based on the planned scope for the WO overlapping with planned project work.	Cycle planning	2*	50%
If yes	Mark this as a relevant project effort and proceed with project-scoping activities.	Cycle planning	0	2%
If no	Skip project steps.	Cycle planning	0	48%
Total (minutes FTE per WO)			1	
* With most WOs, it is a trivial effort to determine whether the work is project work. Some (about 10%), however, require more research (around 20 minutes). This averages to 2 minutes per WO.				

Automated Decisions and Conditional Actions

Input: Current WO and active projects list.

Should this WO be included with planned project work?

Decision	AI Description			
	Training/Reference Input	Use Input	Functions	Output
Process	Active projects list	Current WO	Comparison	Yes/No*
If yes	NA			
If no	NA			
* AI can support this activity by providing the relevant projects, but human confirmation is still needed.				

B-7. Project Sequence

Some projects comprise many other smaller projects that are scoped separately. If the WO relates to project work, it should be scheduled in accordance with the master project plan. Work management ensures that the WO tied to other projects is completed in accordance with the master project plan.

Current Unconditional Actions

Input: Current WO, active projects list, and master project plan.

Assign project sequence.

Action	Effort Description	Discipline	Time in Minutes	Frequency per WO
Process	Compare the WO scope to the project work scoped into an execution week or weeks and move the WO to a new week if needed. Verify the WO scope is on a system that is selected for work during the selected week. Determine whether the WO is being performed in the intended sequence in accordance with the master project plan. (i.e., logic should tie this work to the other related projects).	Cycle planning	30	2%
Total (minutes FTE per WO)			0.6	

Automated Unconditional Actions

Input: Current WO, active projects list, and master project plan.

Assign project sequence.

Action	AI Description			
	Training/Reference Input	Use Input	Functions	Output
Process	Active projects list and scheduled week or weeks for each in the master project plan	Current WO	Comparison	Relevant scope week
	Active projects list and master project plan	Current WO	Review	Sequence validation

B-8. Project Week(s) Readiness

It is necessary to determine if the project work is approved by projects management (which occurs during individual discussions among project team members or during one or more project review meetings) and is feasible in the preliminary scheduled week, given constraints associated with existing WOs.

Current Decisions and Conditional Actions

Input: Current and existing WOs.

Are all projects scheduled for the preliminary scheduled week ready for implementation?

Decision	Effort Description	Discipline	Time in Minutes	Frequency per WO
Process	Check status of all the relevant projects for the week and consider any constraints.	Cycle planning	60	2%
If yes	Move WO to the project execution week	Cycle planning	3	0.2%
If no	None.	NA	0	1.8%
Total (minutes FTE per WO)			1.21	

Automated Decisions and Conditional Actions

Input: Current and existing WOs.

Are all projects scheduled for the preliminary scheduled week ready for implementation?

Decision	AI Description			
	Training/Reference Input	Use Input	Functions	Output
Process	Existing relevant project WOs and ready labels	Current WO	Comparison	Yes/No
If yes	NA			
If no	NA			

B-9. Project Engineering Evaluation Review

To successfully plan the WPs necessary to implement a project, engineering products such as drawing revisions, updated parameters, and set-point revisions must be performed, reviewed, and approved so they can be handed over to the maintenance work planners. This step reviews each of those engineering products to ensure that they will be ready to support WP planning.

Current Decisions and Conditional Actions

Input: Current WO and EC package.

Is the applicable EC on track for completion to meet the project schedule?

Decision	Effort Description	Discipline	Time in Minutes	Frequency per WO
Process	Verify that all ECs listed in the EC package and referenced in the WO are in the complete status and that all drawings have been released for work.	Engineering	30*	2%
If yes	None.	NA	0	1.9%
If no	Identify additional resources necessary for completion of EC documentation or rescope the work into a different execution week.	Engineering	Out of scope	0.1%
Total (minutes FTE per WO)		0.6		
* With most project WOs, it is a reasonable effort (around 15 minutes) to determine whether the EC is ready. Some (about 1%), however, require more research (around 45 minutes). This averages to 30 minutes per project WO.				

Automated Decisions and Conditional Actions

Input: Current WOs and EC package, and engineering procedures.

Is the applicable EC on track for completion to meet the project schedule?

Decision	AI Description			
	Training/Reference Input	Use Input	Functions	Output
Process	General-knowledge-enabled LLM	Current EC package	Classification (unsupervised)	Yes/No
	Engineering procedure	Current EC package	Review	Yes/No (and reason why it is not ready)
If yes	NA			
If no	NA			

B-10. Project Need for Contractor Services

Some projects are very involved and require many resources to implement. If the work cannot be performed within the allotted timeframe by the plant maintenance organization, then contracted services may be enlisted to ensure that the project is completed on time. This step evaluates the need for contracted services based on the work scope and available plant resources.

Current Decisions and Conditional Actions

Input: Current WO, contractor list, and staff resources dataset.

Will some or all the project work need to be contracted out to an external vendor or vendors?

Decision	Effort Description	Discipline	Time in Minutes	Frequency per WO
Process	Determine if contracted resources (from the contractor list) will be necessary to execute the project by comparing work scope from WO, WO history for similar work, in-house resources, and contractor resources.	Projects	10	2%
If yes	Prepare procurement requisition and purchase order for needed services.	Procurement	Out of scope	0.7%
If no	None.	NA	0	1.33%
Total (minutes FTE per WO)				0.2

Automated Decisions and Conditional Actions

Input: Current and historical WOs and EC package, and staff resources dataset.

Will some or all the project work need to be contracted out to an external vendor or vendors?

Decision	AI Description			
	Training/Reference Input	Use Input	Functions	Output
Process	Historical WOs and EC package contract decisions	Current WO and EC package	Structuring and classification (supervised)	Yes/No
	Historical WOs and EC packages resources	Current WO and EC package	Structuring and classification (supervised)	Resources needed*
If yes	NA			
If no	NA			
* AI can support this activity by providing the resources needed, but human review is needed after consulting with scheduling.				

B-11. Work Order Activities Creation

Each WO has initially one default activity. Additional WO activities may need to be created to ensure that related work steps, indirectly associated with the main scope (e.g., scaffolding), are completed. This step creates as many WO activities as necessary.

Current Decisions and Conditional Actions

Input: Current and historical WOs.

Do additional WO activities need to be created to complete the work scope?

Decision	Effort Description	Discipline	Time in Minutes	Frequency per WO
Process	Evaluate the work scope and previous similar WOs to determine if additional WO activities need to be created to ensure that the work scope can be properly scoped, planned, scheduled, and executed.	Maintenance	30	50%
If yes	Create additional WO activities.	Maintenance	8*	35%
If no	None.	NA	0	15%
Total (minutes FTE per WO)			17.8	
* With most project WOs, it is a reasonable effort (around 10 minutes) to create activities. Project WOs (about 2%), however, require more research (around 240 minutes). This averages to 8 minutes per project WO.				

Automated Decisions and Conditional Actions

Input: Current and historical WOs.

Do additional WO activities need to be created to complete the work scope?

Decision	AI Description			
	Training/Reference Input	Use Input	Functions	Output
Process	Historical WOs and activities	Current WO	Classification (supervised)	Yes/No
If yes	Historical WOs and activities	Current WO	Generation	Activities for new WOs
	Historical WOs	New activities	Insertion	Detailed activities
	Historical WOs	New activities	Guidance	Detailed activities
If no	NA			

B-12. Resources Estimation

At this stage of scoping, it is desired to determine the needed work resources for the work scope defined earlier. The goal is to make an initial estimate, providing a basis for early decision-making in the scoping process.

Unconditional Actions

Input: Current and historical WOs.

Prepare a high-level overview of resources needed to complete the work.

Action	Effort Description	Discipline	Time in Minutes	Frequency per WO
Process	Prepare an overview of resource estimation to complete the work. Review similar previous projects for reference. Additionally, prepare an overview of work hour estimation to complete the work.	Maintenance	6	50%
Total (minutes FTE per WO)			3	

Automated Unconditional Actions

Input: Current and historical WOs, and qualification database.

Prepare a high-level overview of resources needed to complete the work.

Action	AI Description			
	Training/Reference Input	Use Input	Functions	Output
Process	Historical WOs and resources allocated	Current WO	Classification (supervised)	Resources needed for similar work
	Qualification database and general-knowledge-enabled LLM	Current WO	Classification (unsupervised)	Needed qualifications
	Historical WOs and resources allocated	Current WO	Estimation	Hours needed

B-13. Resources Availability

All work that is scheduled for a particular workweek must be tabulated to ensure that adequate resources are available to perform all scheduled work. This step ensures that all available resources (plant and contracted) are available to perform the work scheduled within the execution week. Assigning specific staff is handled during scheduling, when resources are matched to the work scope to ensure full utilization.

Current Decisions and Conditional Actions

Input: Current WO, maintenance staffing plan, and qualification database.

Are sufficient in-house and/or contractor staff available during the preliminary scheduled week to perform the work?

Decision	Effort Description	Discipline	Time in Minutes	Frequency per WO
Process	Determine if maintenance or contracted resources are available to support the work scope for the equipment to be worked on. Search the training and qualification databases to determine if workers have the current qualifications to support the work scope to be performed.	Maintenance	12*	50%
If yes	None.	NA	0	45%
If no	Verify options for rescheduling the work to an alternative time or acquiring additional resources to perform the work in the selected timeframe.	Maintenance	30	5%
Total (minutes FTE per WO)				7.5
* With most WOs, it is a trivial effort to determine whether the work will involve significant or unusual resources. Some (about 20%), however, require more research (around 60 minutes). This averages to 12 minutes per WO.				

Automated Decisions and Conditional Actions

Input: Current WO, maintenance staffing plan, and qualification database.

Are sufficient in-house and/or contractor staff available during the scheduled week to perform the work?

Decision	AI Description			
	Training/Reference Input	Use Input	Functions	Output
Process	Maintenance staffing plan, and qualification database	Current WO	Review	Yes/No (and staff availability)
If yes	NA			
If no	NA			

B-14. Plant/Fleet/Corporate Schedule

When a plant is part of a fleet or there are other proximal plants that share resources (e.g., personnel, tools) for a specific type of work, the availability of these resources must be considered to ensure that scheduling milestones are met. Some items in the plant/fleet/corporate schedule (e.g., outage at other unit, inspection) could impact whether resources are available to perform this work.

Current Decisions and Conditional Actions

Input: Current WO, resources available at other plant/fleet units, and schedule for other plant/fleet units.

Does this work involve resources that could be needed from another unit in the plant or fleet?

Decision	Effort Description	Discipline	Time in Minutes	Frequency per WO
Process	Determine if the resources needed are not in-unit and need to be resourced from other plants.	Cycle planning	2*	50%
If yes	Evaluate options to align with other plant/fleet resources availability, including rescheduling the work, or move the work to the outage schedule.	Cycle planning	30	3%
If No	None.	NA	0	47%
Total (minutes FTE per WO)			1.9	

* With most WOs, it is a trivial effort to determine whether the work will involve fleet or corporate resources. Some (about 5%), however, require more research (around 40 minutes). This averages to 2 minutes per WO.

Automated Decisions and Conditional Actions

Input: Current WO, resources available at other plant/fleet units, and schedule for other plant/fleet units.

Does this work involve resources that could be needed from another unit in the plant or fleet?

Decision	AI Description			
	Training/Reference Input	Use Input	Functions	Output
Process	Historical WOs out-of-plant used resources	Current WO	Structuring and classification (supervised)	Yes/No
If yes	Resources available at other plant/fleet units, and schedule for other plant/fleet units	Current WO (resources estimate)	Review	Staff availability
If no	NA			

B-15. Technical Specifications

Technical specifications need to be considered when scoping work. This ensures that inoperability time of associated equipment is minimized, and that other related equipment is available to provide required redundancy for probabilistic risk assessment (PRA) considerations.

Current Decisions and Conditional Actions

Input: Current WO and technical specifications.

Will the scheduled work have an impact on equipment governed by technical specifications?

Decision	Effort Description	Discipline	Time in Minutes	Frequency per WO
Process	Determine whether the equipment affected by the work is safety-related equipment, and whether the scope of work affects technical specifications compliance.	Operations	6*	50%
If yes	Risk analysis will be needed to complete scoping of the work.	Risk management	Out of scope	5%
If no	None.	NA	0	45%
Total (minutes FTE per WO)			3	
* With most WOs, it is a trivial effort to determine whether the work will affect technical specifications. Some (about 10%), however, require more research (around 60 minutes). This averages to 6 minutes per WO.				

Automated Decisions and Conditional Actions

Input: Current and historical WOs, MEL, and technical specifications.

Will the scheduled work have an impact on equipment governed by technical specifications?

Decision	AI Description			
	Training/Reference Input	Use Input	Functions	Output
Process	Historical WOs and technical specifications labels*	Current WO	Classification (supervised)	Yes/No
	MEL technical specifications labels	Current WO	Comparison	Yes/No
	Technical specifications	Current WO	Review	Yes/No
If yes	NA			
If no	NA			
* If the technical specification label is available.				

B-16. Unit Down Power

In some cases, a unit down power will be necessary to ensure that the work is performed safely, and to minimize the risk of a unit trip, or the potential for an unplanned impact on reactivity. Other considerations that affect whether a derate is necessary include radiological and unfavorable work environment, depending on the area in which the work is being performed.

Current Decisions and Conditional Actions

Input: Current WO, operations procedures, MEL, and technical specifications.

Is a down power necessary to execute the work during the execution week?

Decision	Effort Description	Discipline	Time in Minutes	Frequency per WO
Process	Determine if the work scope will force a derate due to technical specifications compliance, and if the work is intrusive and would disable the equipment Compare the WO equipment to the equipment of MEL that requires a derate.	Operations	6*	50%
If yes	Initiate plans for performing a unit derate during the execution week. Coordinate with the utility grid operator.	Operations	Out of scope	1%
If no	None.	NA	0	49%
Total (minutes FTE per WO)			3	

* With most WOs, it is a trivial effort to determine whether the work requires a down power. Some (about 10%), however, require more research (around 60 minutes). This averages to 6 minutes per WO.

Automated Decisions and Conditional Actions

Input: Current and historical WOs, operations procedures, plant process computer, MEL, and technical specifications.

Is a down power necessary to execute the work during the execution week?

Decision	AI Description			
	Training/Reference Input	Use Input	Functions	Output
Process	Historical WOs and downrate labels (using plant computer)	Current WO	Classification (supervised)	Yes/No
	MEL down power label	Current WO	Comparison	Yes/No
	Operations procedures and technical specifications	Current WO	Review	Yes/No
If yes	NA			
If no	NA			

B-17. Chemistry Impact

There are various systems and components that can impact reactors or other important system chemistry parameters when they are unavailable. The sustained unavailability of plant-chemistry-related equipment can degrade a reactor vessel and associated piping. Reducing the unavailability of this equipment is the desired goal when working on any component or system that can affect the availability of this equipment. The impact of the work on chemistry needs to be evaluated.

Current Decisions and Conditional Actions

Input: Current WO and plant chemistry parameters and procedures.

Will the scheduled work impact plant chemistry?

Decision	Effort Description	Discipline	Time in Minutes	Frequency per WO
Process	Determine if the equipment within the WO is chemistry-related equipment and assess the impact of the work on plant chemistry.	Chemistry	3*	50%
If yes	Identify necessary mitigation strategies and associated resources (evaluations, chemicals, vendors, equipment, other resources).	Chemistry	Out of scope	5%
If no	None.	NA	0	45%
Total (minutes FTE per WO)			1.5	
* With most WOs, it is a trivial effort to determine whether the work impacts the plant chemistry. Some (about 10%), however, require more research (around 60 minutes). This averages to 3 minutes per WO.				

Automated Decisions and Conditional Actions

Input: Current WO, MEL, and plant chemistry parameters and procedures.

Will scheduled work impact plant chemistry?

Decision	AI Description			
	Training/Reference Input	Use Input	Functions	Output
Process	Chemistry words and general-knowledge-enabled LLM	Current WO	Classification (unsupervised)	Yes/No
	MEL chemistry labels	Current WO	Comparison	Yes/No
	Chemistry parameters and procedures	Current WO	Review	Yes/No
If yes	NA			
If no	NA			

B-18. Reactivity Effect

Various plant systems and components can either directly or indirectly impact nuclear reactivity when they are worked on, or when they are made unavailable to allow for work on other components within a functional location. If reactivity will be affected, compensatory work may also be performed to ensure that all reactivity impact is planned and quantified as expected.

Current Decisions and Conditional Actions

Input: Current WO, MEL, and system drawings.

Will this work, possibly in conjunction with other scheduled work, result in unacceptable reactivity effects?

Decision	Effort Description	Discipline	Time in Minutes	Frequency per WO
Process	Compare the WO equipment to system drawings to determine if the equipment has been identified as reactivity-related equipment. Consult the MEL to identify reactivity-related equipment and leverage operations staff experience for any other equipment that could indirectly affect reactivity.	Operations	6*	50%
If yes	Evaluate options for rescheduling the work or moving the work to the outage schedule.	Cycle planning	0**	2%
If No	None.	NA	0	48%
Total (minutes FTE per WO)				3
* With most WOs, it is a trivial effort to determine whether the work will impact reactivity. Some (about 10%), however, require more research (around 60 minutes). This averages to 6 minutes per WO. ** This action involves repeating some steps outlined in this appendix. The time is set to zero because this time has already been accounted for in the step time estimates, as specified in the report assumptions.				

Automated Decisions and Conditional Actions

Input: Current WO, MEL, operations procedures, and system drawings.

Will this work, possibly in conjunction with other scheduled work, result in unacceptable reactivity effects?

Decision	AI Description			
	Training/Reference Input	Use Input	Functions	Output
Process	Reactivity words	Current WO	Classification (unsupervised)	Yes/No
	MEL reactivity labels	Current WO	Comparison	Yes/No
	Operations procedures and system drawings	Current WO	Review	Yes/No
If yes	NA			
If no	NA			

B-19. Reactor Coolant System Effect

Maintaining reactor coolant parameters such as level and pressure ensures that required and adequate reactor cooling margins are maintained throughout the timeframe during which the work on the equipment or related equipment is being executed. Maintaining those margins is necessary to ensure that the reactor coolant can adequately perform its function.

Current Decisions and Conditional Actions

Input: Current WO, reactor coolant system equipment list, lessons learned in operating experience (OE) database.

Will this work, possibly in conjunction with other scheduled work, result in unacceptable effects on reactor coolant system parameters?

Decision	Effort Description	Discipline	Time in Minutes	Frequency per WO
Process	Determine if the WO affects reactor coolant system equipment by comparing the equipment listed in the WO with the reactor coolant system equipment list and its associated systems. Additionally, search external OE to assess whether the work on this equipment could impact reactor coolant system equipment.	Operations	3*	50%
If yes	Evaluate options for rescheduling the work or moving the work to the outage schedule.	Cycle planning	0**	2%
If no	None.	NA	0	48%
Total (minutes FTE per WO)			1.5	
* With most WOs, it is a trivial effort to determine whether the work will impact reactor coolant systems. Some (about 10%), however, require more research (around 30 minutes). This averages to 3 minutes per WO. ** This action involves repeating some steps outlined in this appendix. The time is set to zero because this time has already been accounted for in the step time estimates, as specified in the report assumptions.				

Automated Decisions and Conditional Actions

Input: Current WO, reactor coolant system equipment list, operation procedure, and lessons learned in OE database.

Will this work, possibly in conjunction with other scheduled work, result in unacceptable effects on reactor coolant system parameters?

Decision	AI Description			
	Training/Reference Input	Use Input	Functions	Output
Process	Reactor coolant words	Current WO	Classification (unsupervised)	Yes/No
	Reactor coolant system equipment list	Current WO	Comparison	Yes/No
	Operations procedures and OE	Current WO	Review	Yes/No
If yes	NA			
If no	NA			

B-20. Redundant Structures, Systems, and Components Effect

A robust plant design ensures reactor safety and reliability with redundant equipment that must remain unaffected during work execution to maintain its functional availability. If redundancy is impacted, the work can be rescheduled, or compensatory measures can be implemented to ensure plant safety and reliability. Additionally, a further review of technical specifications is conducted to understand any limitations on the availability or operability of redundant equipment.

Current Decisions and Conditional Actions

Input: Current WO, system drawings, and work schedule.

Will this work affect redundancy of a component, train, or system?

Decision	Effort Description	Discipline	Time in Minutes	Frequency per WO
Process	Determine if there is any redundant equipment by comparing the WO with system drawings. Verify if a redundant component, train, or system has work scheduled during the same timeframe.	Operations	2*	50%
If yes	Evaluate options for rescheduling the work or moving the work to the outage schedule.	Cycle planning	0**	5%
If no	None.	NA	0	45%
Total (minutes FTE per WO)			1	

* With most WOs, it is a trivial effort to determine whether the work will impact redundant SSCs. Some (about 10%), however, require more research (around 20 minutes). This averages to 2 minutes per WO.

** This action involves repeating some steps outlined in this appendix. The time is set to zero because this time has already been accounted for in the step time estimates, as specified in the report assumptions.

Automated Decisions and Conditional Actions

Input: Current WO, MEL, system drawings, and work schedule.

Will this work affect redundancy of a component, train, or system?

Decision	AI Description			
	Training/Reference Input	Use Input	Functions	Output
Process	MEL redundant systems labels	Current WO	Comparison	Yes/No
	System drawings	Current WO	Structuring	Yes/No
If yes	NA			
If no	NA			

B-21. Plant Transient

Engineered safety features (ESFs) are design elements intended to prevent accidental releases of reactor fission products or to protect the public in the event of such releases from the reactor coolant system. ESFs can include systems like emergency core cooling. They should not be activated unintentionally during normal plant operations. Unplanned activation of these systems or other plant transients may occur during maintenance work, potentially eroding safety margins and reducing plant reliability.

Current Decisions and Conditional Actions

Input: Current WO and operational procedures.

Could this work, possibly in conjunction with other scheduled work, potentially result in a plant transient or the actuation of an ESF?

Decision	Effort Description	Discipline	Time in Minutes	Frequency per WO
Process	Compare the work scope against operations procedures to identify if a plant transient or ESF actuation could result from performing the work.	Operations	2*	50%
If yes	Determine if the operations procedures identify alternative configurations for completing the work or for mitigating risk or move to outage.	Operations	45	2%
If no	None.	NA	0	48%
Total (minutes FTE per WO)			1.9	

* With most WOs, it is a trivial effort to determine whether the work will result in a transient. Some (about 10%), however, require more research (around 20 minutes). This averages to 2 minutes per WO.

Automated Decisions and Conditional Actions

Input: Current WO, MEL, and operations procedures.

Could this work, possibly in conjunction with other scheduled work, potentially result in a plant transient or the actuation of an ESF?

Decision	AI Description			
	Training/Reference Input	Use Input	Functions	Output
Process	Historical WOs and plant transient (from process computer)	Current WO	Classification (supervised)	Yes/No
	MEL transient labels	Current WO	Comparison	Yes/No
	Operations procedures	Current WO	Review	Yes/No
If yes	Operations procedures	Current WO	Summarization	Alternative configuration
If no	NA			

B-22. Surveillance Schedule

To maintain technical specification compliance as well as to ensure that condition-based monitoring is performed for plant reliability, surveillances are performed within set frequencies. If the surveillances are not performed within the required window, regulatory requirements may not be met, or important information that is necessary to prevent unplanned equipment failures may not be obtained. The WO needs to be evaluated against the surveillance schedule to ensure surveillances are performed as required.

Current Decisions and Conditional Actions

Input: Current WO and operational surveillance database.

Will this work interfere with any scheduled surveillances that need to be performed during the same time?

Decision	Effort Description	Discipline	Time in Minutes	Frequency per WO
Process	Compare the WO scope to operational surveillance databases to determine if any scheduled surveillances will be impacted or delayed because of the work to be performed.	Operations	6*	50%
If yes	Evaluate options for rescheduling the work.	Cycle planning	0**	5%
If no	None.	NA	0	45%
Total (minutes FTE per WO)			3	

* With most WOs, it is a trivial effort to determine whether the work will impact the surveillance schedule. Some (about 10%), however, require more research (around 60 minutes). This averages to 6 minutes per WO.

** This action involves repeating some steps outlined in this appendix. The time is set to zero because this time has already been accounted for in the step time estimates, as specified in the report assumptions.

Automated Decisions and Conditional Actions

Input: Current WO, MEL, and operational surveillance database.

Will this work interfere with any scheduled surveillances that need to be performed during the same time?

Decision	AI Description			
	Training/Reference Input	Use Input	Functions	Output
Process	Operational surveillance equipment and planned time, and MEL surveillance labels	Current WO	Comparison	Yes/No
If yes	NA			
If no	NA			

B-23. Seasonal Impacts

Weather extremes can impact work that is to be performed and can reduce required equipment availability times as well as result in damage to equipment if associated heating and cooling are not available as designed. Also, the work may result in unacceptable risk during a sensitive season of the year. Seasonal impacts need to be considered when scoping work on equipment that can be impacted by the weather.

Current Decisions and Conditional Actions

Input: Current WO, historical WPs, and operations procedures.

Is there the potential for weather to negatively impact the work for when it is scheduled?

Decision	Effort Description	Discipline	Time in Minutes	Frequency per WO
Process	Search operations procedures and historical WPs to determine if the subject equipment or any downstream equipment can be affected by extreme weather conditions.	Operations	6*	50%
If yes	Evaluate options for rescheduling the work.	Cycle planning	0**	1%
If no	None.	NA	0	49%
Total (minutes FTE per WO)			3	
* With most WOs, it is a trivial effort to determine whether the work will be impacted by seasonal factors. Some (about 10%), however, require more research (around 60 minutes). This averages to 6 minutes per WO. ** This action involves repeating some steps outlined in this appendix. The time is set to zero because this time has already been accounted for in the step time estimates, as specified in the report assumptions.				

Automated Decisions and Conditional Actions

Input: Current WO, historical WPs, and operations procedures.

Is there the potential for weather to negatively impact the work for when it is scheduled?

Decision	AI Description			
	Training/Reference Input	Use Input	Functions	Output
Process	Historical WPs for the WO equipment and references to weather-related issues	Current WO	Comparison and structuring	
	Operations procedures	Current WO	Review	Yes/No
If yes	NA			
If no	NA			

B-24. Grid Instability

Although most utilities have electricity contracts that provide a certain level of predictability with the electrical grid, factors outside the plant's control, such as unexpected low-voltage conditions, can impact the availability of certain equipment. The potential for unexpected grid stability issues and factors affecting the compensatory equipment supporting the plant must be considered. Additionally, any work that may pose an unacceptable risk to the grid during the scheduled time should also be evaluated.

Current Decisions and Conditional Actions

Input: Current WO, electrical diagrams, and utility grid policy/procedures.

Will this work result in unacceptable risk to the grid or to the plant because of potential grid instability?

Decision	Effort Description	Discipline	Time in Minutes	Frequency per WO
Process	Review electrical diagrams and utility grid policies/procedures to determine if the equipment is related to electrical systems that are critical to or dependent on the grid, and evaluate the available backup power equipment options	Operations	2*	50%
If yes	Evaluate options for rescheduling the work or moving the work to the outage schedule.	Cycle planning	0**	1%
If no	None.	NA	0	49%
Total (minutes FTE per WO)			1	

* With most WOs, it is a trivial effort to determine whether the work will impact the grid. Some (about 3%), however, require more research (around 60 minutes). This averages to 2 minutes per WO.

** This action involves repeating some steps outlined in this appendix. The time is set to zero because this time has already been accounted for in the step time estimates, as specified in the report assumptions.

Automated Decisions and Conditional Actions

Input: Current WO, MEL, electrical diagrams, OE, and utility grid policy/procedures.

Will this work result in unacceptable risk to the grid or to the plant because of potential grid instability?

Decision	AI Description			
	Training/Reference Input	Use Input	Functions	Output
Process	Grid electrical equipment words and general-knowledge-enabled LLM	Current WO	Classification (unsupervised)	Yes/No
	MEL grid electrical equipment or OE	Current WO	Comparison	Yes/No
	Electrical diagrams and utility grid policy/procedures	Current WO	Review	Yes/No
If yes	NA			
If no	NA			

B-25. Containment Integrity/Heat Removal

The potential impact of the work on containment integrity and the ability to remove heat generated by plant operations or during shutdown conditions needs to be evaluated. Any potential negative impacts on any critical nuclear safety boundaries such as containment (and its structural and leak integrity) or heat removal systems must be considered when scoping the work to ensure that these impacts are minimized or that compensatory actions are taken.

Current Decisions and Conditional Actions

Input: Current WO, MEL, system drawings, and operations procedures.

Does this work have the potential to impact the containment integrity or the ability to remove heat from the reactor?

Decision	Effort Description	Discipline	Time in Minutes	Frequency per WO
Process	Determine if the WO equipment is designated as part of the containment boundary or heat removal in the MEL. Use system drawings and operating procedures to verify the typical requirements and constraints for conducting the work.	Operations	6*	50%
If yes	Evaluate options for rescheduling the work or moving the work to the outage schedule.	Cycle planning	0**	1%
If no	None.	NA	0	49%
Total (minutes FTE per WO)			3	
<p>* With most WOs, it is a trivial effort to determine whether the work will impact containment integrity/heat removal. Some (about 10%), however, require more research (around 60 minutes). This averages to 6 minutes per WO.</p> <p>** This action involves repeating some steps outlined in this appendix. The time is set to zero because this time has already been accounted for in the step time estimates, as specified in the report assumptions.</p>				

Automated Decisions and Conditional Actions

Input: Current WO, MEL, system drawings, and operations procedures.

Does this work have the potential to impact the containment integrity or the ability to remove heat from the reactor?

Decision	AI Description			
	Training/Reference Input	Use Input	Functions	Output
Process	Containment or reactor coolant words	Current WO	Classification (unsupervised)	Yes/No
	MEL containment and heat removal labels	Current WO	Comparison	Yes/No
	System drawings and operations procedures	Current WO	Review	Yes/No
If yes	NA			
If no	NA			

B-26. Cycle Timing

During unit startup after an outage, or during unit coast down when a plant is long into its operational cycle, the characteristics of the plant are different from those during normal plant operations. Cyclical impacts can change the operational characteristics of certain plant equipment, making them different than during normal operations. These differences must be considered when performing work within these timeframes.

Current Decisions and Conditional Actions

Input: Current WO, operations procedures, and OE database.

Will this work have an adverse impact on personnel or equipment based on when it is scheduled in the cycle?

Decision	Effort Description	Discipline	Time in Minutes	Frequency per WO
Process	Compare the work scope to the operations procedures and OE database to determine any cyclical impacts.	Operations	6*	50%
If yes	Evaluate options for rescheduling the work.	Cycle planning	0**	1%
If no	None.	NA	0	49%
Total (minutes FTE per WO)			3	

* With most WOs, it is a trivial effort to determine whether the work will be impacted by cycle timing. Some (about 10%), however, require more research (around 60 minutes). This averages to 6 minutes per WO.

** This action involves repeating some steps outlined in this appendix. The time is set to zero because this time has already been accounted for in the step time estimates, as specified in the report assumptions.

Automated Current Decisions and Actions

Input: Current WO, operations procedures, and OE database.

Will this work have an adverse impact on personnel or equipment based on when it is scheduled in the cycle?

Decision	AI Description			
	Training/Reference Input	Use Input	Functions	Output
Process	Operations procedures and OE database	Current WO	Review	Yes/No
If yes	NA			
If no	NA			

B-27. Plant Security

In some cases, work can impact the plant security plan by causing equipment necessary for meeting security plan requirements to be unavailable or of diminished capacity. Security requirements need to be considered when planning work for equipment that can affect the security plan.

Current Decisions and Conditional Actions

Input: Current WO and security equipment configuration documentation.

Will this work adversely affect the site protection strategy?

Decision	Effort Description	Discipline	Time in Minutes	Frequency per WO
Process	Determine if the work to be performed can impact security equipment that is necessary to protect the plant and consult with site security.	Security	1*	50%
If yes	Evaluate options for mitigating risk.	Security	Out of scope	3%
If no	None.	NA	0	47%
Total (minutes FTE per WO)			0.5	

* With most WOs, it is a trivial effort to determine whether the work will impact security. Some (about 10%), however, require more research (around 10 minutes). This averages to 1 minute per WO.

Automated Decisions and Conditional Actions

Input: Current WO, MEL, and security equipment configuration documentation.

Will this work adversely affect the site protection strategy?

Decision	AI Description			
	Training/Reference Input	Use Input	Functions	Output
Process	Security words and general-knowledge-enabled LLM	Current WO	Classification (Unsupervised)	Yes/No
	MEL security equipment labels	Current WO	Comparison	Yes/No
	Security equipment configuration documentation	Current WO	Review	Yes/No
If yes	NA			
If no	NA			

B-28. Maintenance Rule

Maintenance-rule-related equipment has associated regulatory time clocks tracking the system or component unavailability. When maintenance-rule-related equipment unavailability exceeds the set unavailability limit, additional regulatory measures are taken to preserve the remaining availability time of this equipment. It is therefore important to consider that maintenance rule time clocks for maintenance-rule-related equipment.

Current Decisions and Conditional Actions

Input: Current WO and maintenance rule database.

Will the work result in unacceptable levels of equipment availability?

Decision	Effort Description	Discipline	Time in Minutes	Frequency per WO
Process	Compare the work equipment to the maintenance rule database, and if it is included, determine whether the work will impact equipment availability.	Engineering	3*	50%
If yes	Evaluate options for rescheduling the work to an alternative time or means for minimizing unavailability time.	Cycle planning	30	1%
If no	None.	NA	0	49%
Total (minutes FTE per WO)		1.8		
* With most WOs, it is a trivial effort to determine whether the work will impact maintenance rule functions. Some (about 10%), however, require more research (around 30 minutes). This averages to 3 minutes per WO.				

Automated Decisions and Conditional Actions

Input: Current and historical WO and maintenance rule database.

Will the work result in unacceptable levels of equipment availability?

Decision	AI Description			
	Training/Reference Input	Use Input	Functions	Output
Process	Maintenance rule database	Current WO	Comparison	Yes/No
	Historical WOs and equipment loss time	Current WO	Estimation	Unavailability estimate
If yes	NA			
If no	NA			

B-29. Risk Screening

If the work performed could impact plant risk, it is input into a software program that aggregates overall risk to make sure that approved risk levels are not exceeded. The aggregate PRA risk is calculated for all work scheduled within the execution week to ensure the risk is acceptable.

Current Decisions and Conditional Actions

Input: Current WO and PRA software model.

Does the work scope introduce risk to the plant?

Decision	Effort Description	Discipline	Time in Minutes	Frequency per WO
Process	Update the PRA model software with the equipment condition to determine if there is any impact change to the PRA.	Engineering	Out of scope	50%
If yes	Consider compensatory actions that can be taken to reduce the aggregate risk to the plant's safety.	Engineering	240	1%
If no	None.	NA	0	49%
Total (minutes FTE per WO)			2.4	

Automated Decisions and Conditional Actions

Input: Current WO and PRA software model.

Does the work scope introduce risk to the plant?

Decision	AI Description			
	Training/Reference Input	Use Input	Functions	Output
Process	NA			
If yes	NA*			
If no	NA			

* Risk optimization can benefit from using AI, but this is outside the scope of the current effort.

B-30. Work Scope Clarity and Conciseness

The scope of the work needs to be clear and concise. A final check ensures that the appropriate resources, evaluations, and timeframes are well understood and that the work is accurately scoped within the allotted week. Additionally, the accuracy of the scope of work is evaluated to ensure it is useful for work planning.

Current Decisions and Conditional Actions

Input: Current WO.

Is the WO scope clear and concise?

Decision	Effort Description	Discipline	Time in Minutes	Frequency per WO
Process	Review the work scope to ensure that all decisions are made correctly and documented appropriately, and that the WO is written in an acceptable manner.	Maintenance SRO	3	50%
If yes	Add additional notes if needed.	NA	1	40%
If no	Clarify work scope and consider rescoping if the work scope changes are significantly different than when WO was created.	Maintenance SRO	30	10%
Total (minutes FTE per WO)		4.9		

Automated Decisions and Conditional Actions

Input: Current WO and work management procedure.

Is the WO scope clear and concise?

Decision	AI Description			
	Training/Reference Input	Use Input	Functions	Output
Process	Work management procedure	Current WO	Review	Yes/No
If yes	Historical WOs	Current WO	Guidance	Notes
	Historical WOs	Current WO	Insertion	Notes
If no	Work management procedure	Current WO	Generation	Issues and discrepancies

B-31. Integrated Risk Screening for Execution Week

Aside from PRA there are other types of safety risks that must be considered for an execution week. Each task itself may not pose a decreased safety margin, but in aggregate these risks could result in eroded safety margins such that risk-mitigating measures need to be taken. This step evaluates the overall integrated risk to the plant for the respective execution week.

Current Unconditional Actions

Input: Current and historical WOs, work management database, OE, and PRA results.

Consider risk associated with the work and develop mitigation strategies to be employed with the work.

Action	Effort Description	Discipline	Time in Minutes	Frequency per WO
Process	Aggregate the risk for all work scheduled for the execution week, compare the risk against historical WOs data and OE, and identify any impacts that will reduce any type of safety, radiological, or environmental margins not previously evaluated in any other step of the scoping process.	Cycle planning	15	50%
Total (minutes FTE per WO)			7.5	

Automated Unconditional Actions

Input: Current and historical WOs, work management database, OE, and PRA results.

Consider risk associated with the work and develop mitigation strategies to be employed with the work.

Action	AI Description			
	Training/Reference Input	Use Input	Functions	Output
Process	NA*			

* This process is too risky and complicated to automate.

B-32. Priority Assignment Validation

When WOs are created from condition screening, projects, or other sources, they are given an immediate priority to help with the scoping of the work. After all the research is done, especially for the risk evaluations, the priority of the work can either be escalated or downgraded. This step is performed to validate the initial priority assigned to the work and to ensure that work management requirements for priority are followed correctly.

Current Decisions and Conditional Actions

Input: Current and historical WOs and work management procedures.

Is the assigned priority accurate?

Decision	Effort Description	Discipline	Time in Minutes	Frequency per WO
Process	Compare the current WO information to previous work and the work management procedure to ensure that the proper WO priority is assigned to the subject work.	Cycle planning	0.5	50%
If yes	None.	NA	0	47%
If no	Modify the priority and provide justification.	Cycle planning	0	3%
Total (minutes FTE per WO)		0.25		

Automated Decisions and Conditional Actions

Input: Current and historical WOs and work management procedures.

Is the assigned priority accurate?

Decision	AI Description			
	Training/Reference Input	Use Input	Functions	Output
Process	Historical WOs and priority labels	Current WO	Classification (supervised)	Yes/No
	Work management procedures	Current WO	Review	Yes/No
If yes	NA			
If no	Work management procedures	Current WO	Generation	Justification

B-33. Priority Work Rank

Once the priorities are assigned to each WO scoped into the work execution week, the work is ranked to ensure that the highest-priority WOs are completed within their scheduled requirements. Each WO is ranked against the rest of the work scheduled for the week to determine which tasks can be delayed, if necessary, and which must be completed as scheduled. As a result, it is important to avoid scheduling too many high-priority WOs in the same week.

Current Unconditional Actions

Input: Current WO and planned WOs for the same duration.

Rank the current WO according to priority in comparison to other work.

Action	Effort Description	Discipline	Time in Minutes	Frequency per WO
Process	Compare the priority of the WO to the priorities of other planned WOs, consider overlap of needed resources, and consult the work management procedure for requirements on work sequencing.	Cycle planning	3	50%
Total (minutes FTE per WO)			1.5	

Automated Unconditional Actions

Input: Current and historical WOs and planned WOs for the same duration.

Rank the current WO according to priority in comparison to other work.

Action	AI Description			
	Training/Reference Input	Use Input	Functions	Output
Process	Planned WOs and assigned resources	Current WO	Comparison	Similar resources
	Historical WOs completion times	Current and planned WOs	Estimation	Time to execute the work
	Planned WOs resources and execution times	Current WO	Review	Optimal ranking of WOs

B-34. Preventive Maintenance Grace Periods

PM WOs are assigned a priority along with all WOs scheduled to be performed within the workweek. Some PM is regulatorily required, with grace periods being strictly adhered to when the WO priority is assigned. If a PM WO is to be moved to a different week, it is important that the PM has minimal impact on regulatory margins or plant reliability. If the grace period needs to be exceeded, then a PM change is needed.

Current Decisions and Conditional Actions

Input: Current WO and planned PM WO for the same duration.

Is a PM change needed for a PM WO that is scheduled in the same week?

Decision	Effort Description	Discipline	Time in Minutes	Frequency per WO
Process	Consider all priorities of work assigned to the week. Determine if there is a risk of the PM WO going beyond the grace period.	Cycle planning	3	50%
If yes	Create a PM change request.	Maintenance	Out of scope	1%
If no	None.	NA	0	49%
Total (minutes FTE per WO)		1.5		

Automated Decisions and Conditional Actions

Input: Current WO and planned PM WO for the same duration.

Is a PM change needed for the PM WO that is scheduled in the same week?

Decision	AI Description			
	Training/Reference Input	Use Input	Functions	Output
Process	NA*			
If yes	NA			
If no	NA			

* This is a simple lookup from a database table. No AI automation is needed.

Page intentionally left blank

Appendix C

Mapping of Planning

Page intentionally left blank

Appendix C

Mapping of Planning

This appendix applies the approach of Section 4.1 to planning steps according to the flow shown in Figure 18. The following sections discuss each of those steps in detail.

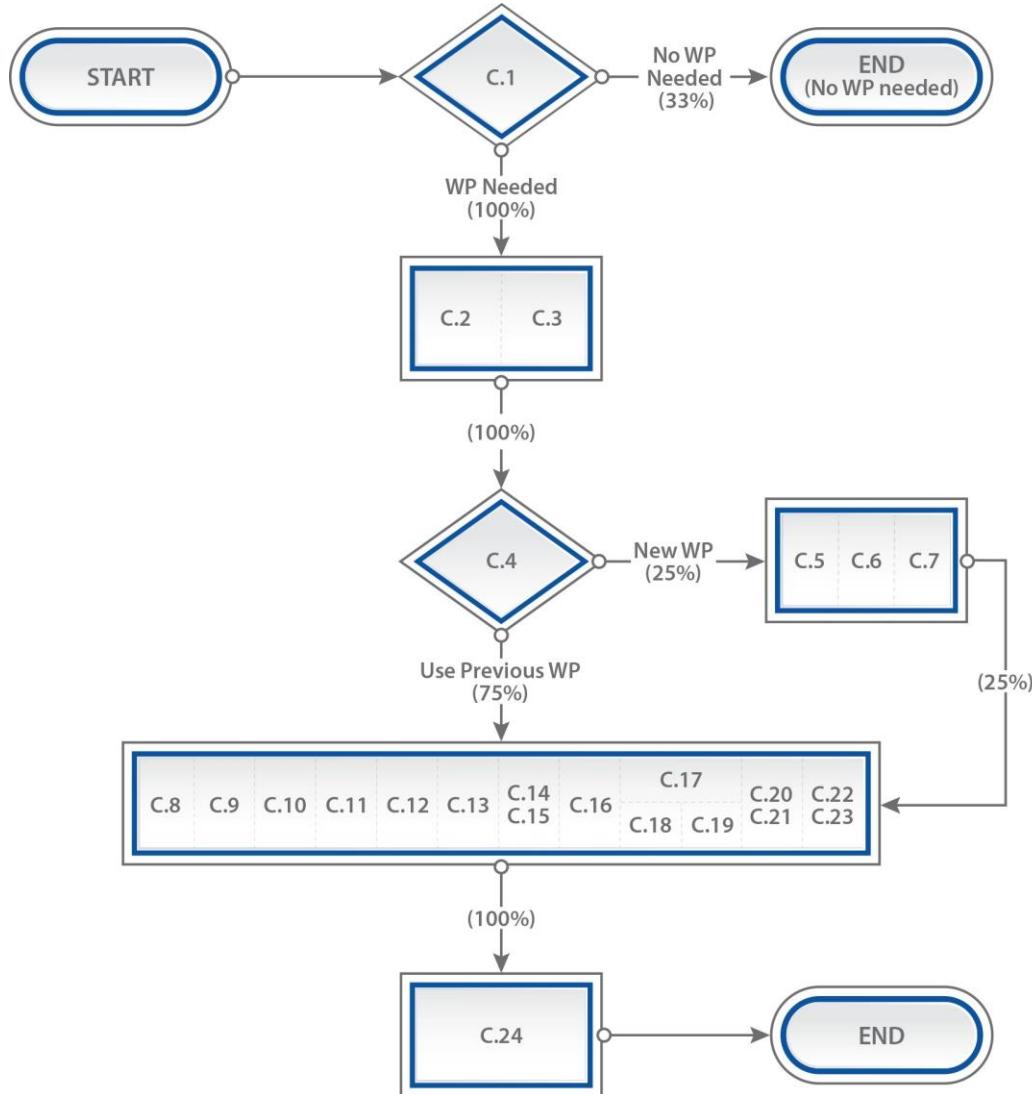


Figure 18. Flowchart of the planning steps that are discussed in Appendix C.

C-1. Need for Work Package

The work scope for most WOs at an NPP are complex enough that instructions and various permits are required to complete the work. All these instructions and information related to the WO scope are known collectively as a WP. This step determines if the work scope is complex enough to require a WP.

Current Decisions and Conditional Actions

Input: Current WO.

Does the WO need a WP for guidance?

Decision	Effort Description	Discipline	Time in Minutes	Frequency Per WP
Process	Determine if the work scope is complex (requires detailed work instructions, multiple work steps, system isolation, or permits for such as hot work, confined space, or FME, etc.).	Work planning*	3	133%
If yes	Flag that a WP is needed.	Work planning	0	100%
If no	Create minor or tool pouch maintenance instructions and pass it to fix-it-now team. (End of process)	Fix-it-now planning	30	33%
Total (minutes FTE per WP)		14		
* Work planning is part of the maintenance organization.				

Automated Decisions and Conditional Actions

Input: Current and historical WOs and historical WO to WP assignment table.

Does the WO need a WP for guidance?

Decision	AI Description			
	Training/Reference Input	Use Input	Functions	Output
Process	Historical WOs and WPs assignment	Current WO	Classification (supervised)	Yes/No
If yes	NA			
If no	Historical WOs instructions.	Current WO	Generation	Few instructions to conduct the work

C-2. Previous Similar Work Packages

If a WP is required to complete the work scope, typically a previously completed WP for a similar work scope can be used as a template for the new work scope WP. This template can be updated and inform the WP planning process. If no previous WPs are similar enough to serve as the template for the new WP, an entirely new WP must be created with new content, though it may still benefit from content from previous WPs that have a somewhat similar work scope (but not similar enough to serve as a template).

Current Unconditional Actions

Input: Current WO and historical WPs

Identify previously completed WPs that are similar to the WO scope.

Action	Effort Description	Discipline	Time in Minutes	Frequency per WP
Process	Search for relevant previous WPs with a work scope similar to the current WP. Save all identified WPs for reference during the planning process.	Work planning	5	100%
Total (minutes FTE per WP)			5	

Automated Unconditional Actions

Input: Current and historical WOs and historical WPs.

Identify previously completed WPs that are similar to the WO scope.

Action	AI Description			
	Training/Reference Input	Use Input	Functions	Output
Process	Historical WOs and WPs assignment	Current WO	Comparison	Similar WPs

C-3. Long Lead-Time Parts

An important early step during the planning process is to determine if there are any parts needed for the work that are not readily available and will take a relatively long time (months) to procure and receive on site. Such parts need to be identified early to allow adequate time for procurement. Thus, this step is pursued as early as possible in the planning process. Additional long-lead parts may be identified as the planning process continues and additional information is gained, but initial efforts to procure long-lead parts are pursued early on with the best available information.

Current Decisions and Conditional Actions

Input: Current WO, similar WPs, and inventory dataset.

Are any parts or materials that have a long lead time needed for the work to be performed?

Decision	Effort Description	Discipline	Time in Minutes	Frequency per WP
Process	Develop an initial parts list for the work scope focusing on major items that may have a longer lead time. Consider what was used in the similar WPs. Confirm if any part has a long lead time in the procurement system or request a lead time from the vendor.	Work planning	15	100%
If yes	Reserve parts.	Work planning	10	20%
If no	None.	NA	0	80%
Total (minutes FTE per WP)			17	

Automated Decisions and Conditional Actions

Input: Current WO, similar WPs, and inventory dataset.

Are any parts or materials that have a long lead time needed for the work to be performed?

Decision	AI Description			
	Training/Reference Input	Use Input	Functions	Output
Process	Historical WOs WPs assignment or similar WPs materials or parts list and lead time (from the inventory dataset)	Current WO	Structuring and estimation	Materials or parts lead time
If yes	NA			
If no	NA			

C-4. Existing or New Work Package Use

After similar WPs have been identified as potential candidates to serve as templates for updating the new WP (see previous step), the work scope must be examined in a broader context to determine if the selected similar WPs are suitable templates based on all factors. For example, if it has been many years since the most recent similar WP was created, there may have been enough changes to procedures and policies on work instructions and permits that it would be better to start with a new WP rather than use this particular WP as a template. Or perhaps there have been changes to equipment or systems that mean the previous WP no longer adequately describes the current work scope.

Current Decisions and Conditional Actions

Input: Current WO and most similar WP.

Has the work scope significantly changed since the last time the similar WP was used?

Decision	Effort Description	Discipline	Time in Minutes	Frequency per WP
Process	Compare the similar WP to the current WO scope and look for significant differences or inconsistencies.	Work planning	5	100%
If yes	The work scope has significantly changed. Begin creating a new WP from a blank template.	Work planning	5	25%
If no	Use the scope and instructions from the previous WP as a starting point for the new WP.	Work planning	5	75%
Total (minutes FTE per WP)			10	

Automated Decisions and Conditional Actions

Input: Current WO, similar WP, and historical WPs.

Has the work scope significantly changed since the last time the similar WP was used?

Decision	AI Description			
	Training/Reference Input	Use Input	Functions	Output
Process	Similar WP	Current WO	Summarization	WO meets WP scope or similarities and differences
If yes	Similar WP	Current WO	Generation	Template WP
If no	Historical WPs	Current WO	Generation	Generic Template WP
	Historical WPs	Current WO	Guidance	General instructions
	Historical WPs	Current WO	Insertion	General instructions

C-5. Similar Work Packages Review for Information

If no previously completed WP is similar enough to directly serve as a template to be updated for the new WP, there may still be previously completed WPs that are close enough to provide useful content and information for developing the new WP. These WPs may not be for the exact same equipment or may have slightly different work scope, and such differences need to be considered when developing work instructions and new WP content.

Current Decisions and Conditional Actions

Input: Current and similar WPs.

Are there any similar previously completed WPs information that can be used to help create the new WP?

Decision	Effort Description	Discipline	Time in Minutes	Frequency per WP
Process	Search in similar WPs with a focus on content, that will be helpful in creating the new WP.	Work planning	15	25%
If yes	Save content from similar WP(s) to a new WP template.	Work planning	15	13%
If no	None.	Work planning	0	12%
Total (minutes FTE per WP)				5.7

Automated Decisions and Conditional Actions

Input: Current WP and similar WPs.

Are there any similar previously completed WPs information that can be used to help create the new WP?

Decision	AI Description			
	Training/Reference Input	Use Input	Functions	Output
Process	Similar WPs	Current WP	Summarization	Similarities
If yes	Similar WPs	Current WP	Generation	Similar instructions
	Similar WPs	Current WP	Structuring	Useful information
If no	NA			

C-6. Needed Technical Information

For a new WP (created from a blank template instead of from a previously completed WP used as a template), a certain level of understanding of the equipment involved is necessary to complete the work instructions for the desired work scope. Technical documentation may be in various repositories and formats at the power plant, and it may be necessary to compile these documents for easy access. In addition to finding and collecting equipment technical information, a work planner will review the information and take notes on aspects of the identified data that will be useful for the given work scope.

Current Unconditional Actions

Input: Current and historical WPs, ECs and drawings, and manufacturer/vendor technical manuals.
Compile applicable technical information that is available.

Action	Effort Description	Discipline	Time in Minutes	Frequency per WP
Process	Search out and compile applicable technical information that is available for the equipment and become familiar with how the equipment works and what needs to be done in the work scope to ensure future equipment performance. Consult with engineering personnel as needed and make notes of any particular aspects of the equipment to keep in mind for planning the work scope.	Work planning	60	25%
Total (minutes FTE per WP)				15

Automated Unconditional Actions

Input: Current and historical WPs, ECs and drawings, and manufacturer/vendor technical manuals.
Compile applicable technical information that is available.

Action	AI Description			
	Training/Reference Input	Use Input	Functions	Output
Process	Historical WPs technical information	Current WP	Comparison and structuring	Critical technical information
	ECs and drawings, and manufacturer/vendor technical manuals	Current WP	Summarization	Critical technical information

C-7. Work Instructions

If there is not an existing ready-to-use template for work instructions that can be directly applied to the new work scope of a new WP, some level of writing will likely be required to produce the work instructions for the new scope. In some cases, generating work instructions may be as simple as referencing existing procedures, whereas in other cases it may require drafting new work steps.

Current Unconditional Actions

Input: Current and historical WPs.

Draft work instructions.

Action	Effort Description	Discipline	Time in Minutes	Frequency per WP
Process	Draft work instructions for the work scope, referencing existing procedures as needed.	Work planning	60	25%
Total (minutes FTE per WP)			15	

Automated Unconditional Actions

Input: Current and historical WPs and procedures.

Draft work instructions.

Action	AI Description			
	Training/Reference Input	Use Input	Functions	Output
Process	Historical WPs procedures	Current WP	Generation	New instructions
	Historical WPs procedures	Current WP	Guidance	New instructions
	Historical WPs procedures	Current WP	Insertion	New instructions

C-8. Incorporate Operating Experience

Each WP should incorporate the latest industry lessons learned, industry OE, and improvements initiated by the CAP. This involves including the most applicable OE reports in the WP for use in pre-job briefs and incorporating any applicable lessons learned for all applicable OE into the work instructions. For WPs being updated, the previous similar WP must be updated with any recent information in this area for the new WP. In the case of developing a new WP, there may be some relevant OE from similar WPs that can be incorporated into the new WP, or new OE may be found that can be included.

Current Decisions and Conditional Actions

Input: Current and historical WP, OE database, and historical CRs.

Are there CRs or OE evaluations that are pertinent to this WP?

Decision	Effort Description	Discipline	Time in Minutes	Frequency per WP
Process	Search for relevant CRs or OE lessons learned and evaluations for inclusion in the WP.	Work planning	10	100%
If yes	Update the WP and its pre-job brief with new info.	Work planning	10	50%
If no	None.	NA	0	50%
Total (minutes FTE per WP)		15		

Automated Decisions and Conditional Actions

Input: Current and historical WP, OE database, and historical CRs.

Are there CRs or OE evaluations that are pertinent to this WP?

Decision	AI Description			
	Training/Reference Input	Use Input	Functions	Output
Process	Historical WPs found issues	Current WP	Comparison and structuring	Relevant issues
	Historical CRs	Current WP	Summarization	Relevant CRs*
	OE database	Current WP	Summarization	Relevant OE*
If yes	Relevant issues, CR, or OE	Current WP	Generation	Modified WP with issues, CRs or OE insight
If no	NA			

* A human will determine if the CR or OE are relevant.

C-9. Worker Feedback from Previous Similar Work

Feedback from workers from the most recent time similar work was previously performed may be available. The WP must be updated to incorporate any feedback from workers. Feedback from workers may be captured in completed WPs, in the CAP database, or in some other lessons learned database. Reviewing feedback for new WPs is not the same as for previously completed WPs, as the previous worker feedback is not for the exact same work scope. However, worker feedback can still be incorporated from previous similar work scopes.

Current Decisions and Conditional Actions

Input: Current and historical WP, lessons learned database, and historical CRs.

Is there any worker feedback from similar work that was performed previously?

Decision	Effort Description	Discipline	Time in Minutes	Frequency per WP
Process	Find previous similar WPs and relevant CRs and check for any relevant worker feedback that should be incorporated into the current work scope.	Work planning	5	100%
If yes	Update the WP based on worker feedback.	Work planning	15	20%
If no	No change to the WP.	Work planning	0	80%
Total (minutes FTE per WP)			8	

Automated Decisions and Conditional Actions

Input: Current and historical WP, lessons learned database, and historical CRs.

Is there any worker feedback from similar work that was performed previously?

Decision	AI Description			
	Training/Reference Input	Use Input	Functions	Output
Process	Historical WPs and feedback	Current WP	Comparison and structuring	Relevant feedback
	Historical CRs	Current WP	Summarization	Relevant CRs*
	OE database	Current WP	Summarization	Relevant OE*
If yes	Relevant feedback, CR, or OE	Current WP	Generation	Suggestions based on worker feedback
If no	NA			

* A human will determine if the CR or OE are relevant.

C-10. Human Performance Tools

The work scope should be reviewed to determine if there are work steps that should have specific HP tools incorporated into the work instructions to mitigate risk. Those HP tools could include self-check, STAR (stop, think, act, review), peer check, and three-way communication should be incorporated into the work instructions.

Current Decisions and Conditional Actions

Input: Current WP and HP procedure.

Do specific HP tools need to be used by workers due to potential risk situations?

Decision	Effort Description	Discipline	Time in Minutes	Frequency per WP
Process	Evaluate the HP risk for work steps and determine if there are any high-risk steps for which HP tools—as per HP procedure—should be incorporated into the instructions.	Work planning	2*	100%
If yes	Update WP to incorporate HP tools.	Work planning	15	20%
If no	None.	NA	0	80%
Total (minutes FTE per WP)		5		
* With most WPs, it is a trivial effort to determine whether the work will need HP tools. Some (about 10%), however, require more research (around 20 minutes). This averages to 2 minutes per WP.				

Automated Decisions and Conditional Actions

Input: Current WP, historical WPs, and HP procedure.

Do specific HP tools need to be used by workers due to potential risk situations?

Decision	AI Description			
	Training/Reference Input	Use Input	Functions	Output
Process	Historical WPs and HP tools assignment	Current WP	Structuring and classification (supervised)	Yes/No (and HP tools)
	HP procedure	Current WP	Review	Yes/No
If yes	Historical WPs HP tools and HP procedure	Current WP	Generation	HP tools
If no	NA			

C-11. Clearance Order Request

If system isolation is necessary to complete the work, a CO is required. System or process drawings, such as piping and instrumentation diagrams or electrical one-line diagrams, can help identify sources of operational hazards, along with CO or hazardous energy administrative procedures. Typically, a CO for a similar scope of work will have been completed previously and can be used as a reference to request a new CO. This process will determine if a CO is needed for the current work scope.

Current Decisions and Conditional Actions

Input: Current and historical WPs, system drawings, and administrative procedures for hazardous energy or COs.

Does a CO request need to be initiated for new WP?

Decision	Effort Description	Discipline	Time in Minutes	Frequency per WP
Process	Compare the work scope to the hazardous energy or CO administrative procedures to determine whether system isolation is required to perform the work safely. Use system/process drawings to identify sources of hazardous energy. Consider previously completed similar WPs to make the determination.	Work planning	3	100%
If yes	Create a new CO request.	Work planning	15	33%
If no	None.	NA	0	67%
Total (minutes FTE per WP)			8	

Automated Decisions and Conditional Actions

Input: Current and historical WPs, system drawings, and administrative procedures for hazardous energy or COs.

Does a CO request need to be initiated for the new WP?

Decision	AI Description			
	Training/Reference Input	Use Input	Functions	Output
Process	Historical WPs and COs assignment	Current WP	Structuring and classification (supervised)	Yes/No
	System drawings and administrative procedures for hazardous energy or COs	Current WP	Review	Yes/No
If yes	Historical WPs COs	Current WP	Generation	New CO
If no	NA			

C-12. Engineering and Operations Documentation

The WP needs to be checked to see if the equipment to be worked on has an applicable engineering configurations or operational requirements (technical specifications) that need to be considered in the work instructions. For WPs created from a previously completed WP used as a template, this review focuses on changes that occurred since the previous WP was completed. For new WPs, this review ensures that the draft work instructions are not missing any information based on how the equipment is described in engineering and technical specifications.

Current Decisions and Conditional Actions

Input: Current WP, ECs, and technical specifications.

Have there been any changes to the engineering specifications or technical specifications regarding the equipment in the WP?

Decision	Effort Description	Discipline	Time in Minutes	Frequency per WP
Process	Search for ECs or technical specifications changes related to the equipment referenced in the WP and compare them to those in the WP.	Work planning	15	100%
If yes	Update the WP with engineering/technical specifications information.	Work planning	30	5%
If no	None.	NA	0	95%
Total (minutes FTE per WP)			16.5	

Automated Decisions and Conditional Actions

Input: Current WP, ECs, and technical specifications.

Have there been any changes to the engineering specifications or technical specifications regarding the equipment in the WP?

Decision	AI Description			
	Training/Reference Input	Use Input	Functions	Output
Process	ECs and technical specifications (multiple versions)	Current WP	Summarization	Relevant EC and technical specifications and version changes
If yes	ECs and technical specifications	Current WP	Generation	Suggested text based on changes from EC and technical specifications
If no	NA			

C-13. Administrative or Technical Procedures

Procedures may need to be referenced for certain administrative processes (such as establishing a confined space) and technical processes (such as performing the calibration of instrumentation) that need to occur as part of performing the work. Current revisions of the procedures must be included in the WP, and the content of those procedures must be incorporated correctly.

Current Decisions and Conditional Actions

Input: Current WP and administrative or technical procedures.

Does the WP need to reference administrative or technical procedures?

Decision	Effort Description	Discipline	Time in Minutes	Frequency per WP
Process	Examine the work scope to determine if there are technical and administrative processes that require a reference to a procedure.	Work planning	2*	100%
If yes	Add applicable content from existing procedures by reference or by including applicable content in the work instructions. Compare the procedure versions included in the WP to the latest procedures.	Work planning	15	67%
If no	None.	Work planning	0	33%
Total (minutes FTE per WP)				12.05

* With most WPs, it is a trivial effort to determine whether the work will reference administrative or technical procedures. Some (about 10%), however, require more research (around 20 minutes). This averages to 2 minutes per WP.

Automated Decisions and Conditional Actions

Input: Current and historical WPs and administrative or technical procedures.

Does the WP need to reference administrative or technical procedures?

Decision	AI Description			
	Training/Reference Input	Use Input	Functions	Output
Process	Historical WPs and administrative or technical procedures assignment	Current WP	Structuring and classification (supervised)	Yes/No
	Administrative or technical procedures	Current WP	Review	Yes/No
If yes	Historical WPs and administrative or technical procedures	Current WP	Generation	Added procedures
If no	NA			

C-14. Fire Protection and Industrial Safety Controls

When creating a WP, particular attention must be given to ensure that all currently applicable fire protection and industrial safety controls are included in the WP. This step typically involves interfacing with fire protection engineers and industrial safety professionals to ensure that all currently applicable controls are included in the WP, such as limits on combustible loading for a given area or policies on the establishment of confined spaces.

Current Decisions and Conditional Actions

Input: Current WP, and procedures and forms related to fire protection and industrial safety.

Will the work scope impact fire protection or industrial safety policies or practices in the area where work will be performed?

Decision	Effort Description	Discipline	Time in Minutes	Frequency per WP	
Process	Evaluate the location and scope of work according to procedure and forms related to fire protection and industrial safety.	Work planning	1*	100%	
If yes	Update the WP instructions to include fire protection or industrial safety controls.	Work planning	20	5%	
If no	No change to the WP.	Work planning	0	95%	
Total (minutes FTE per WP)		2			
* With most WPs, it is a trivial effort to determine whether the work will impact fire protection or industrial safety. Some (about 10%), however, require more research (around 10 minutes). This averages to 1 minute per WP.					

Automated Decisions and Conditional Actions

Input: Current and historical WPs, MEL, and procedures and forms related to fire protection and industrial safety.

Will the work scope impact fire protection or industrial safety policies or practices in the area where work will be performed?

Decision	AI Description			
	Training/Reference Input	Use Input	Functions	Output
Process	Fire protection or industrial safety words and general-knowledge-enabled LLM	Current WP	Classification (unsupervised)	Yes/No
	MEL fire protection labels	Current WP	Comparison	Yes/No
	Procedures and forms related to fire protection and industrial safety	Current WP	Review	Yes/No
	Historical WPs and fire protection or industrial safety controls assignment	Current WP	Structuring and classification (supervised)	Yes/No
If yes	Procedures and forms related to fire protection and industrial safety	Current WP	Generation	Modified WP with controls
If no	NA			

C-15. Fire Protection and Industrial Safety Permits

Many WPs require one or more permits be generated to authorize the work to be performed. While the previous step incorporates the applicable fire protection and industrial safety controls into the work instructions themselves, this step completes the necessary permits to prepare to authorize the work. Some permits may be totally prepared and approved at this time, while others can be initially prepared but can only be approved closer to the time of work implementation.

Current Decisions and Conditional Actions

Input: Current WP and maintenance procedures.

Do any hot work, fire protection, or combustible loading permits need to be initiated?

Decision	Effort Description	Discipline	Time in Minutes	Frequency per WP
Process	Verify whether the work scope requires a permit as per the requirements in maintenance procedures.	Work planning	2*	5%
If yes	Initiate relevant permit.	Work planning	30	4%
If no	None.	Work planning	0	1%
Total (minutes FTE per WP)				1.3
* With most WPs, it is a trivial effort to determine whether the work will need hot work, fire protection, or combustible loading permits. Some (about 1%), however, require more research (around 20 minutes). This averages to 2 minutes per WP.				

Automated Decisions and Conditional Actions

Input: Current and historical WPs and maintenance procedures.

Do any hot work, fire protection, or combustible loading permits need to be initiated?

Decision	AI Description			
	Training/Reference Input	Use Input	Functions	Output
Process	Historical WPs fire protection permit assignment	Current WP	Structuring and classification (supervised)	Yes/No
	Maintenance procedures	Current WP	Review	Yes/No
If yes	Historical WPs fire protection permits	Current WP	Generation	Permit forms
If no	NA			

C-16. Work Package Walkdown

An important step during WP planning, when possible, is to perform a walkdown of the area where work will be performed to evaluate actual conditions in the area and determine if draft work instructions are adequate based on the work environment and proposed scope of work. Sometimes walkdowns are not possible during planning due to inaccessibility of the work area. Frequently, maintenance work groups and support groups are invited to participate in walkdowns to provide input.

Current Decisions and Conditional Actions

Input: Current WP and plant conditions.

Is a walkdowns needed and feasible?

Decision	Effort Description	Discipline	Time in Minutes	Frequency per WP
Process	Determine if work is complex or critical and require a walkdown. Check if the equipment is accessible given the plant status.	Work planning	5	100%
If yes	Perform walkdown and update the WP to reflect the walkdown findings.	Work planning	15*	20%*
If no	No change to the WP.	Work planning	0	80%
Total (minutes FTE per WP)		8		
* Since walkdowns are not automatable, this time and frequency is only for the WP update.				

Automated Decisions and Conditional Actions

Input: Current and historical WPs and plant conditions.

Is a walkdowns needed and feasible?

Decision	AI Description			
	Training/Reference Input	Use Input	Functions	Output
Process	Historical WPs notes from walkdown	Current WP	Structuring and classification (supervised)	Yes/No
If yes	Current notes from walkdown	Current WP	Generation	Updates to WP
If no	NA			

C-17. Radiological Controls

If the work will be performed in a radiological area, the planner is responsible to ensure applicable radiological controls are incorporated into the WP. Usually this means referencing a radiological work permit (RWP). If a more than negligible amount of radiation dose is expected to be received by workers performing the work, the dose must be accounted for in overall dose estimates and some degree of planning must take place to minimize the dose.

Current Decisions and Conditional Actions

Input: Current WP, RWP, and radiological survey databases.

Will the work be performed in a radiological protected area (RPA)?

Decision	Effort Description	Discipline	Time in Minutes	Frequency per WP
Process	Determine, based on equipment location and current radiological conditions (from the most recently completed radiological surveys), if work will be performed in RPA.	Work planning	1	100%
If yes	Include relevant radiological controls in the WP.	Work planning	30	25%
If no	None.	NA	0	75%
Total (minutes FTE per WP)			8.5	

Automated Decisions and Conditional Actions

Input: Current WP, RWP, and radiological survey databases.

Will the work be performed in an RPA?

Decision	AI Description			
	Training/Reference Input	Use Input	Functions	Output
Process	Referenced locations in the plant in the radiological work permit and radiological survey databases	Current WP	Comparison	Yes/No
	Historical WPs ALARA plans assignment	Current WP	Structuring and classification (supervised)	Yes/No
If yes	Historical WPs radiological controls	Current WP	Generation	Radiological controls
If no	NA			

C-18. ALARA Plan

If the work will be performed in a radiological area and the potential for radiation dose received by workers exceeds a defined threshold, the planner is responsible to initiate an ALARA plan. An ALARA plan includes an assessment of radiation risks and strategies to minimize dose to workers. Most of the work of creating the ALARA plan falls to RP personnel.

Current Decisions and Conditional Actions

Input: Current and historical WPs, RP procedures, dose-tracking and radiological survey databases.

Will the work require an ALARA plan?

Decision	Effort Description	Discipline	Time in Minutes	Frequency per WP
Process	Use the dose-tracking and radiological survey database Determine if the work scope is in an area that will result in radiation exposure and study applicable RP procedures to identify if the work requires an ALARA plan. Review historical work in the area to determine if an ALARA plan is needed.	RP	3	25% *
If yes	Create an ALARA plan.	RP	Out of scope	5%
If no	None.	NA	0	20%
Total (minutes FTE per WP)			0.75	

* Frequencies are different based on plant design (boiling water reactor vs. pressurized water reactor).

Automated Decisions and Conditional Actions

Input: Current and historical WPs, RP procedures, dose-tracking and radiological survey databases.

Will the work require an ALARA plan?

Decision	AI Description			
	Training/Reference Input	Use Input	Functions	Output
Process	Referenced locations in the plant in the dose-tracking and radiological survey databases	Current WP	Comparison	Yes/No
	Historical WPs ALARA plan assignment	Current WP	Structuring and classification (supervised)	Yes/No
If yes	NA			
If no	NA			

C-19. Radiological Work Impact on Workers

When work is to be performed in an RPA, the work planner must work with RP personnel to determine the potential hazards to workers who will be performing the work. Impacts to consider include dose rate changes during the work, contamination levels changing during the work, and whether dose rates may challenge worker dose limits. If there are potential significant impacts to workers, work instructions and/or the ALARA plan must be updated accordingly with appropriate warnings and controls.

Current Decisions and Conditional Actions

Input: Current WPs, dose-tracking and historical radiological survey databases.

Could the work scope result in an increased radiological impact to workers?

Decision	Effort Description	Discipline	Time in Minutes	Frequency per WP
Process	Evaluate the history of dose rates and contamination in the work location focusing on historical sudden changes.	RP	10	25%
If yes	Adjust work instructions and/or ALARA plan to include appropriate warnings and controls.	RP	20	1%
If no	None.	NA	0	24%
Total (minutes FTE per WP)		2.7		

Automated Decisions and Conditional Actions

Input: Current WPs, dose-tracking and historical radiological survey databases.

Could the work scope result in an increased radiological impact to workers?

Decision	AI Description			
	Training/Reference Input	Use Input	Functions	Output
Process	Historical WPs and dose rate and contamination level at the execution time (from the dose-tracking and historical radiological survey databases)	Current WP	Estimation	Dose rates and contamination levels changes in the work location
If yes	Historical WPs ALARA plan and current ALARA plan	Current WP	Generation	Update to ALARA plan
If no	NA			

C-20. Worker Training and Qualification

Some work tasks are considered “skill of the craft” tasks and do not require specific training or qualification to perform. Tasks beyond skill of the craft typically require some form of training or qualification. These may be general (for example, fall protection/ladder safety training to access elevated areas where equipment may be located) or technical (for example, training and qualification required to use a torque wrench) in nature. When work instructions are created for the WP, the training and qualifications needed by workers to perform the work must be determined and specified.

Current Decisions and Conditional Actions

Input: Current WP, training and qualification list, and training policies/procedures.

Does the work scope require specific training or qualifications?

Decision	Effort Description	Discipline	Time in Minutes	Frequency per WP
Process	Compare the work scope to training policies/procedures to determine if any work tasks require training and/or worker qualifications.	Work planning	3*	100%
If yes	Ensure work tasks are covered by existing training and qualifications and document the training and qualifications needed to perform the work in the WP.	Work planning	15	25%
If no	None.	NA	0	75%
Total (minutes FTE per WP)			6.75	

* With most WPs, it is a trivial effort to determine whether the work will require training or qualifications. Some (about 10%), however, require more research (around 30 minutes). This averages to 3 minutes per WP.

Automated Decisions and Conditional Actions

Input: Current and historical WPs, training and qualification list, and training policies/procedures.

Does the work scope require specific training or qualifications?

Decision	AI Description			
	Training/Reference Input	Use Input	Functions	Output
Process	Training and qualification words and general-knowledge-enabled LLM	Current WP	Classification (unsupervised)	Yes/No
	Historical WPs and training or qualifications assignment	Current WP	Structuring and classification (supervised)	Yes/No
	Training policies/procedures	Current WP	Review	Yes/No
If yes	Historical WPs and training or qualifications	Current WP	Generation	Training and qualification requirements
If no	NA			

C-21. Resources Assignment to Work Package

After determining the needed training and qualifications, the staff's skill set is compared to the needed skill set and tasks are assigned accordingly. At this stage, the crew is assigned to the WP, not the specific staff as those are assigned by the supervisor during execution.

Current Unconditional Actions

Input: Current WP, training and qualification list, and staff resources dataset.

Assign crew to the WP.

Action	Effort Description	Discipline	Time in Minutes	Frequency per WP
Process	Compare the required training and qualifications for the work to those of the staff, and assign a crew if they meet the requirements for the scope	Work planning	5	100%
Total (minutes FTE per WP)				5

Automated Unconditional Actions

Input: Current WP, training and qualification list, and staff resources dataset.

Assign crew to the WP.

Action	AI Description			
	Training/Reference Input	Use Input	Functions	Output
Process	Staff training and qualification (from the staff resources dataset)	Current WP	Comparison	Crew that meets the requirement

C-22. Equipment Qualification Requirements

Certain equipment in an NPP must be purchased and controlled as either environmental qualification (EQ) equipment or seismically qualified equipment. As part of the work execution work planners must verify whether EQ or seismically qualified parts are needed as part of the work scope and verify that such parts are available and controlled as qualified parts.

Current Decisions and Conditional Actions

Input: Current WP, MEL, and inventory dataset.

Are any of the parts or components needed for the work scope classified as EQ or seismic?

Decision	Effort Description	Discipline	Time in Minutes	Frequency per WP
Process	Check the work scope against equipment or components that require EQ or seismic qualifications.	Work planning	5	100%
If yes	Notate work as EQ or seismic in the WP and ensure that replacement parts are correctly qualified.	Work planning	30	5%
If no	None.	NA	0	95%
Total (minutes FTE per WP)				6.5

Automated Decisions and Conditional Actions

Input: Current and historical WPs, MEL, inventory dataset, and EQ and seismic qualification requirements.

Are any of the parts or components needed for the work scope classified as EQ or seismic?

Decision	AI Description			
	Training/Reference Input	Use Input	Functions	Output
Process	Historical WPs and EQ or seismic notation	Current WP	Structuring and classification (supervised)	Yes/No
	MEL EQ or seismic labels or inventory dataset	Current WP	Comparison	Yes/No
If yes	EQ and seismic qualification requirements	inventory dataset	Review	Yes/No
If no	NA			

C-23. Parts and Materials

A work planner must evaluate what parts and materials will be needed as part of the work scope and whether those parts already exist in the plant inventory database. Previous similar WOs may be helpful in determining what parts and materials may be needed for the work scope and how those are defined in the inventory database. The planner puts in a parts and materials request so that procurement personnel can work to obtain the needed parts if necessary and have all parts and materials ready for the work before it is performed.

Current Unconditional Actions

Input: Current and historical WPs, and inventory database.

Reserve required parts and materials.

Action	Effort Description	Discipline	Time in Minutes	Frequency per WP
Process	Check the parts and/or materials needed as part of the work scope and reserve required parts and materials by initiating a procurement request.	Work planning	20	100%
Total (minutes FTE per WP)				20

Automated Unconditional Actions

Input: Current and historical WPs, and inventory database.

Reserve required parts and materials.

Action	AI Description				
	Training/Reference Input	Use Input	Functions	Output	
Process	Historical WPs and parts and materials assignment	Current WP	Structuring and classification (supervised)	Needed parts and materials	

C-24. Work Package Review and Approval

After the WP planning steps are completed and the WP is prepared, applicable representatives or supervisors from impacted work groups review and approve the WP.

Current Decisions and Conditional Actions

Input: Current WP, and applicable subject-matter-specific policies/procedures.

Is the WP ready to be performed as prepared?

Decision	Effort Description	Discipline	Time in Minutes	Frequency per WP
Process	Check the completion of the WP content and determine if all processes and procedures have been correctly incorporated and the WP is ready to be worked.	All relevant groups	20*	100%
If yes	Mark the WP as reviewed and/or approved.	All relevant groups	0	90%
If no	Discuss any needed changes with the work planner until changes are agreed upon, then mark the WP as reviewed and/or approved.	All relevant groups	60**	10%
Total (minutes FTE per WP)		26		
* Assuming 4 staff spend 5 minutes per WP, this would result in 20 minutes per WP. ** Assuming 2 staff spend 30 minutes per WP, this would result in 60 minutes per WP.				

Automated Decisions and Conditional Actions

Input: Current WP, and applicable subject-matter-specific policies/procedures.

Is the WP ready to be performed as prepared?

Decision	AI Description			
	Training/Reference Input	Use Input	Functions	Output
Process	Relevant policies/procedures	Current WP	Review	Yes/No
If yes	NA			
If no	Relevant policies/procedures	Current WP	Review	Needed change

Page intentionally left blank

Appendix D

Mapping of Scheduling

Page intentionally left blank

Appendix D

Mapping of Scheduling

This appendix applies the approach of Section 4.1 to scheduling steps according to the flow shown in Figure 19. The following sections discuss each of those steps in detail.

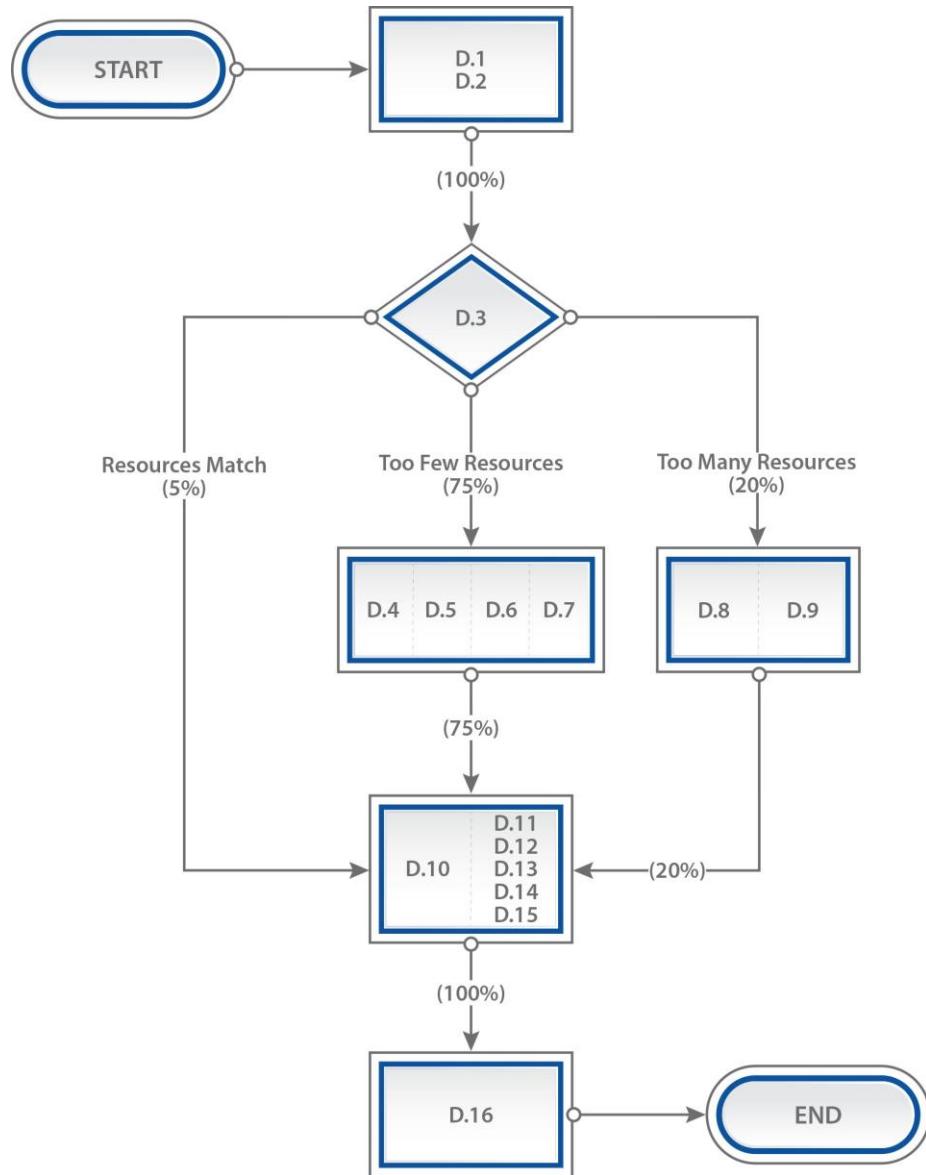


Figure 19. Flowchart of the scheduling steps that are discussed in Appendix D.

D-1. Schedule Tasks Creation

The work scope of a WO described in comprehensive procedures in a WP is performed in STs. Each WO can have one or more STs depending on the work scope. In this step, the STs associated with the WOs are created in the schedule, given the resources available within the workweek. The tasks are not yet sequenced in this step; this is performed in a later step.

Current Unconditional Actions

Input: Planned WOs.

Create the ST for the planned WO in the workweek.

Action	Effort Description	Discipline	Time in Minutes	Frequency per ST
Process	Create an ST for each WO activity, using the ST template. Move all relevant schedule assignments and constraints into the ST from the WO. Add additional STs if needed to accommodate any scope that is not included in the WP but is needed to execute the work.	Scheduling*	5	100%
Total (minutes FTE per ST)		5		
* Work scheduling is part of the work management organization.				

Automated Unconditional Actions

Input: Planned WOs.

Create the ST for the planned WO in the workweek.

Action	AI Description			
	Training/Reference Input	Use Input	Functions	Output
Process	Historical WOs and their STs	Planned WO	Generation	STs

D-2. Resources Assignment to All Schedule Tasks

Before scheduling begins, it is necessary to ensure that needed resources are assigned to all STs that require them. Resource estimation and crew assignment were performed earlier in the process, so this step involves verifying that the crew has the necessary resources to perform the work, considering the entire week's schedule.

Current Decisions and Conditional Actions

Input: Current ST.

Are all needed resources assigned to the ST?

Decision	Effort Description	Discipline	Time in Minutes	Frequency per ST
Process	Evaluate the planned ST resources and assigned crew to ensure that crew resources are properly allocated to the ST.	Scheduling	1	100%
If yes	None.	NA	0	99%
If no	Notify ST owners to assign/allocate resources.	Scheduling	1	1%
Total (minutes FTE per ST)			1.01	

Automated Decisions and Conditional Actions

Input: Current ST.

Are all needed resources assigned to the ST?

Decision	AI Description			
	Training/Reference Input	Use Input	Functions	Output
Process	NA*			
If yes	NA			
If no	NA			

* This can be automated by a database query to ensure that all tasks have a resource assigned.

D-3. Deficiency or Surplus of Resources for Execution Week

Aggregating all STs, the total available resources—by work discipline—are matched to the aggregate work scope. Then schedulers must determine if there are enough resources to complete the scoped work, or if there are too many or too few resources available. This helps them ensure that resources are fully utilized to accomplish as much work as possible. A mismatch typically occurs due to an unexpected increase in work or insufficient work scoped into the execution week to utilize the available resources in specific departments. This latter usually happens when work is removed from the execution week, or when resource predictions change after the work has been scoped into the week.

Current Decisions and Conditional Actions

Input: Current STs, staff resources dataset, and week schedule.

Do available resources match required resources for some time slots in the week?

Decision	Effort Description	Discipline	Time in Minutes	Frequency per ST
Process	Aggregate the resources needed by discipline, and then subtract the available resources from the needed resources by the STs for the scheduled work and quantify the difference in resources by each department.	Scheduling	3	100%
Match	Skip schedule resource optimization.	Scheduling	0	5%
Too few	Proceed to systematically determine how to obtain the required resources or to rescope some of the work out of the execution week schedule.*	Cycle planning	0**	75%
Too many	Proceed to systematically determine the best approach to utilizing the available resources.*	Cycle planning	0**	20%
Total (minutes FTE per ST)			3	
* In some cases, it may take a combination of methods to optimize resources.				
** Performed in next step.				

Automated Decisions and Conditional Actions

Input: Current STs, staff resources dataset, and week schedule.

Do available resources match required resources for some time slots in the week?

Decision	AI Description			
	Training/Reference Input	Use Input	Functions	Output
Process	NA*			
If yes	NA			
If no	NA			
*	This is a database query to ensure that the number of resources available to work in each week matches the number required for all STs for each execution week.			

D-4. Overtime Use

Initial resource availability estimates are based on a 40-hour workweek, considering absences and other issues affecting resources. Work groups can consider offering overtime to the staff as a measure to provide additional resources to compensate for any shortage of available resources.

Current Decisions and Conditional Actions

Input: Current STs and staff resources dataset.

Can overtime be used to fill the resource gap?

Decision	Effort Description	Discipline	Time in Minutes	Frequency per ST
Process	Calculate the amount of overtime above the base available resource amount that would be necessary by department to meet resource shortfalls and consult with the work execution staff.	All relevant groups	1	75%
If yes	Initiate overtime request.	All relevant groups	5	50%
If no	None.	NA	0	25%
Total (minutes FTE per ST)		3.25		

Automated Decisions and Conditional Actions

Input: Current STs and staff resources dataset.

Can overtime be used to fill the resource gap?

Decision	AI Description			
	Training/Reference Input	Use Input	Functions	Output
Process	NA*			
If yes	Staff resources dataset	Current STs	Generation	Overtime request
If yes	NA			
If no	NA			

* This involves a simple calculation.

D-5. Contractors Use

When all available plant resources, including overtime, are not enough to complete work scope estimates, and plant leadership does not want to remove the scope from the execution week, then using contracted workers to perform some of the work scope is an option for augmenting plant staff. In this step, a determination is made whether contracted resources should be used to perform some of the work scope.

Current Decisions and Conditional Actions

Input: Current and historical STs, Current and historical WOs, staff resources dataset, and historical service contracts.

Can contractors be used to fill the resource gap?

Decision	Effort Description	Discipline	Time in Minutes	Frequency per ST
Process	Consider the estimated shortfall numbers for plant resources. Review work scope history and determine if contractors could be used to complete the scope.	Maintenance	2*	75%
If yes	Initiate a service request for contractors to augment plant resources.	Maintenance	15	3%
If no	None.	NA	0	72%
Total (minutes FTE per ST)			1.95	

* With most STs, it is a trivial effort to determine whether the work can be supported by a contractor. Some (about 7%), however, require more research (around 30 minutes). This averages to 2 minutes per ST.

Automated Decisions and Conditional Actions

Input: Current and historical STs, Current and historical WOs, staff resources dataset, and historical service contracts.

Can contractors be used to fill the resource gap?

Decision	AI Description			
	Training/Reference Input	Use Input	Functions	Output
Process	Historical WOs and their assigned contracts in historical STs	Current WO	Structuring and classification (supervised)	Yes/No
	Historical service contracts	Current WO	Review	Yes/No
If yes	Historical service contracts	Current ST and WO	Generation	Contract service request
If no	NA			

D-6. Resources Overload

An option to help close differences between available resources and estimated work scope resources is to overload the work schedule to more than 100% of its resources. This is often performed when leadership is optimistic that the estimated work scope can be performed more efficiently than current work scope estimates and does not want to remove any work scope from the execution week.

Current Decisions and Conditional Actions

Input: Current and historical STs, Current and historical WOs, historical staff resources dataset, and week schedule

Can resource loading of the schedule be greater than 100 percent?

Decision	Effort Description	Discipline	Time in Minutes	Frequency per ST
Process	Evaluate if the similar work is usually accomplished with less time.	Scheduling	3*	75%
If yes	Allocate more work than there are available resources in the week schedule.	Scheduling	10	3%
If no	None.	NA	0	72%
Total (minutes FTE per ST)			2.55	
* With most STs, it is a trivial effort to determine whether the work can be accomplished with less time. Some (about 20%), however, require more research (around 15 minutes). This averages to 3 minutes per ST.				

Automated Decisions and Conditional Actions

Input: Current and historical STs, Current and historical WOs, historical staff resources dataset, and week schedule.

Can resource loading of the schedule be greater than 100 percent?

Decision	AI Description			
	Training/Reference Input	Use Input	Functions	Output
Process	Historical STs	Current ST	Estimation	Revised resources needed
	Historical WOs and their used resources in historical STs or historical staff resources dataset	Current WO	Structuring and estimation	Revised resources needed
If yes	Revised resources needed	Week schedule	Review	Modified schedule
If no	NA			

D-7. Work Removal from Execution Week Schedule

When it is not possible to perform all the execution week work scope due to a lack of available resources, and there is work that can be moved to a different execution week without incurring much risk to the plant, then work management will remove some of the lower-priority work scope. This step determines if some of the work scope can be removed from the subject execution week.

Current Decisions and Conditional Actions

Input: Current STs and WOs and week schedule.

Can work scope be removed from the execution week schedule?

Decision	Effort Description	Discipline	Time in Minutes	Frequency per ST
Process	Identify lower-priority work scope that can be moved from the execution week to another week.	Cycle planning	1	75%
If yes	Move the lower-priority work scoped into the execution week schedule to another week.	Cycle planning	0*	25%
If no	None	NA	0	50%
Total (minutes FTE per ST)			0.75	

* This action involves repeating some steps outlined in the appendices. The time is set to zero because this time has already been accounted for in the step time estimates, as specified in the report assumptions.

Automated Decisions and Conditional Actions

Input: Current STs and WOs and week schedule.

Can work scope be removed from the execution week schedule?

Decision	AI Description			
	Training/Reference Input	Use Input	Functions	Output
Process	NA*			
If yes	NA			
If no	NA			

* Low-priority work is already labeled.

D-8. Work Addition to Execution Week Schedule

If there is a surplus of resources for the execution week, it is important to utilize these resources to complete any pending work. A typical plant often has work items that can take advantage of such opportunities, in addition to any emergent or urgent work. Once work management determines which types of surplus resources are available, work management decides what type of work can be added to the execution week schedule.

Current Decisions and Conditional Actions

Input: Existing WOs, staff resources dataset, and week schedule.

Are there other WOs that can be added to fully utilize execution week resources?

Decision	Effort Description	Discipline	Time in Minutes	Frequency per ST
Process	Evaluate unscoped WOs scope and needed resources and compare them to available resources for opportunities to add work to the execution week schedule.	Cycle planning	2	20%
If yes	Perform scoping for the WOs to be added.	Cycle planning	0*	10%
If no	None.	NA	0	10%
Total (minutes FTE per ST)			0.4	

* This action involves repeating some steps outlined in the appendices. The time is set to zero because this time has already been accounted for in the step time estimates, as specified in the report assumptions.

Automated Decisions and Conditional Actions

Input: Existing and historical WOs, staff resources dataset, qualification database and week schedule.

Are there other WOs that can be added to fully utilize execution week resources?

Decision	AI Description			
	Training/Reference Input	Use Input	Functions	Output
Process	Historical WOs and resources allocated	Existing WOs	Classification (supervised)	Resources needed for work
	Qualification database and general-knowledge-enabled LLM	Existing WOs	Classification (unsupervised)	Needed qualifications
	Historical WOs and resources allocated	Existing WOs	Estimation	Hours needed
	Available staff resources and week schedule	Existing WOs*	Review	WOs to consider for this week
If yes	NA			
If no	NA			

* If resources for the unscoped WO are unknown, this option is performed after the previous three options have been completed to estimate the needed resources.

D-9. Lower-Priority Work Addition to Execution Week Schedule

There are many lower-priority WOs waiting to be scoped into a workweek. Higher-priority WOs have procedural deadlines, but lower-priority work is added when it can be, depending on the available resources and other scoping factors. In this step, work management evaluates the lower-priority WOs and determines if any of those can be scoped into the work execution week.

Current Decisions and Conditional Actions

Input: Existing WOs, staff resources dataset, and week schedule.

Can other lower-priority WOs be added to the schedule to fill resource gaps?

Decision	Effort Description	Discipline	Time in Minutes	Frequency per ST
Process	Identify all lower-priority WOs and evaluate their scope and needed resources to compare them to available resources for opportunities to add work to the execution week schedule.	Scheduling	2	20%
If yes	Perform scoping on the WOs to be added.	Maintenance SRO	0*	10%
If no	None.	NA	0	10%
Total (minutes FTE per ST)				0.4
* This action involves repeating some steps outlined in the appendices. The time is set to zero because this time has already been accounted for in the step time estimates, as specified in the report assumptions.				

Automated Decisions and Conditional Actions

Input: Existing and historical WOs, staff resources dataset, qualification database and week schedule.

Can other lower-priority WOs be added to the schedule to fill resource gaps?

Decision	AI Description			
	Training/Reference Input	Use Input	Functions	Output
Process	Historical WOs and resources allocated	Existing WOs	Classification (supervised)	Resources needed for work
	Qualification database and general-knowledge-enabled LLM	Existing WOs	Classification (unsupervised)	Needed qualifications
	Historical WOs and resources allocated	Existing WOs	Estimation	Hours needed
	Available staff resources and week schedule	Existing WOs*	Review	WOs to consider for this week
If yes	NA			
If no	NA			
* If resources for the low-priority WO are unknown, this option is performed after the previous three options have been completed to estimate the needed resources.				

D-10. Impact of New Items Screened on Existing Execution Week Scope

CRs and WRs are screened daily as issues are identified. Depending on the severity of the issue and the impact on the plant, a priority is added to the new WO and it is allocated to the workweek that best ensures the most favorable outcome for the plant. Sometimes if the new issue is a high priority, then lower-priority previously allocated work must be rescoped into a different execution week.

Current Decisions and Conditional Actions

Input: Current CRs and WOs, and week schedule.

Will any new issues impact the planned work scope of the execution week?

Decision	Effort Description	Discipline	Time in Minutes	Frequency per ST
Process	Compare the priority of any new screened issues or work requested and unscheduled against the WOs allocated to the execution week and determine the impact on the execution week schedule.	Cycle planning	1*	100%
If yes	Allocate the new work to the execution week schedule.	Cycle planning	0**	1%
If no	None.	NA	0	99%
Total (minutes FTE per ST)			1	

* With most STs, it is a trivial effort to determine whether it will be impacted. Some (about 3%), however, require more research (around 30 minutes). This averages to 1 minute per WO.

** This action involves repeating some steps outlined in the appendices. The time is set to zero because this time has already been accounted for in the step time estimates, as specified in the report assumptions.

Automated Decisions and Conditional Actions

Input: Current CRs and WOs, and week schedule.

Will any new issues impact the planned work scope of the execution week?

Decision	AI Description			
	Training/Reference Input	Use Input	Functions	Output
Process	Existing CRs or WOs and priority label	Current (new) CR or WO	Comparison	Similar scope of work*
If yes	NA			
If no	NA			

* Identifies similar work to the already planned work and the priorities assigned to them.

D-11. Logical Order of Schedule Tasks

Once all STs have been created, there is a logical progression of work tasks that must be performed to make sure each work group can perform its tasks at the correctly scheduled time most efficiently. Work management must ensure that the work is performed in a logical order that supports the most efficient work performance. In this step work management reviews the initial sequence of the schedule for the work execution week and places tasks in the most logical order.

Current Unconditional Actions

Input: Current STs and week schedule.

Schedule tasks in the most efficient order to facilitate the best possible execution week outcome.

Action	Effort Description	Discipline	Time in Minutes	Frequency per ST
Process	Evaluate resource availability in the week and work scope sequence requirements. Schedule tasks in the most logical and efficient order. Adjust as necessary. Determine if there are gaps in workflow and add tasks as necessary to ensure schedule continuity.	Scheduling	5	100%
Total (minutes FTE per ST)				5

Automated unconditional Actions

Input: Current and existing STs and week schedule.

Schedule tasks in the most efficient order to facilitate the best possible execution week outcome.

Action	AI Description			
	Training/Reference Input	Use Input	Functions	Output
Process	Existing STs constraints and requirements	Current ST	Review*	ST placed in schedule

* Optimization methods can be used to allocate the tasks in the schedule, but AI review can also be used to perform this.

D-12. Execution Week Schedule Freeze

The work management “freezes” the week schedule so that no more changes can be made to the schedule. This allows all managers to review the upcoming schedule and provide comments during the certification meeting.

Current Unconditional Actions

Input: Current STs and week schedule.

Freeze work execution week schedule.

Action	Effort Description	Discipline	Time in Minutes	Frequency per ST
Process	Lock execution week schedule and add comments for the work execution.	Scheduling	1	100%
Total (minutes FTE per ST)				1

Automated Unconditional Actions

Input: Current and historical STs and week schedule.

Freeze work execution week schedule.

Action	AI Description			
	Training/Reference Input	Use Input	Functions	Output
Process	Historical STs	Current ST	Guidance	Comments
	Historical STs	Current ST	Insertion	Comments

D-13. Execution Week Certification Meeting

The week schedule certification meeting is the last chance for each department to challenge the upcoming week schedule. Based on the comments and concerns gathered during the certification meeting, some changes may be made to the upcoming week schedule.

Current Decisions and Conditional Actions

Input: Current STs and week schedule.

Are there any changes to the work execution week schedule?

Decision	Effort Description	Discipline	Time in Minutes	Frequency per ST
Process	Gather all comments and concerns and identify if any changes need to be made to the workweek schedule.	All relevant groups	3*	100%
If yes	Recommend changes and notify concerned manager that changes are needed.	All relevant groups	5	1%
If No	No changes	All relevant groups	0	99%
Total (minutes FTE per ST)			3.05	
* Assuming 10 staff meet for 120 minutes and 400 STs per week, this would result in 3 minutes per ST.				

Current Decisions and Conditional Actions

Input: Current STs and week schedule.

Are there any changes to the work execution week schedule?

Decision	AI Description			
	Training/Reference Input	Use Input	Functions	Output
Process	NA			
If yes	NA			
If no	NA			

D-14. Changes from Certification Meeting

Depending on the magnitude of the change, actions will be taken in this step to address the concern by the relevant work group. In extreme cases, the cycle planner may remove or rescope work if the concern cannot be resolved.

Current Unconditional Actions

Input: Current STs and WOs and week schedule.

Make changes to address certification week issues.

Action	Effort Description	Discipline	Time in Minutes	Frequency per ST
Process	Address the comments received during the certification meeting. Make changes as necessary. Request work management to reschedule if needed.	All relevant groups	10*	1%
Total (minutes FTE per ST)				0.1
* Any significant changes identified during the certification meeting can result in the work being rescoped and rescheduled. This involves repeating some steps outlined in the appendices. The time is not added because it has already been accounted for in the step time estimates, as specified in the report assumptions.				

Automated Unconditional Actions

Input: Current STs and WOs and week schedule.

Make changes to address certification week issues.

Action	AI Description			
	Training/Reference Input	Use Input	Functions	Output
Process	NA			
* Those activities can vary significantly so automating them might be challenging.				

D-15. Turnover of Execution Week from Cycle Planner to Workweek Manager

In this step, the work management cycle planner turns the subject execution week schedule over to the workweek manager. This turnover includes any discussion of concerns and comments about potential issues that may arise during the execution week.

Current Unconditional Actions

Input: Current STs and WOs and week schedule.

Perform execution week turnover meeting.

Action	Effort Description	Discipline	Time in Minutes	Frequency per ST
Process	Hold a meeting, discuss STs, resolve any minor issues.	Cycle planning and workweek management*	0.5**	100%
Total (minutes FTE per ST)			0.5	

* The workweek management is part of the work management organization.
** Assuming 2 staff meet for 100 minutes and 400 STs per week, this would result in 0.5 minute per ST.

Automated Unconditional Actions

Input: Current STs and WOs and week schedule.

Perform execution week turnover meeting.

Action	AI Description			
	Training/Reference Input	Use Input	Functions	Output
Process	NA			

D-16. Daily Work Schedule Issuance

To keep the schedule current with work execution conditions, work management issues a daily schedule for all work groups. The schedule is updated with the progress of the day and reissued (electronically) before each day shift.

Current Unconditional Actions

Input: Daily schedule.

Issue work management daily schedule.

Action	Effort Description	Discipline	Time in Minutes	Frequency per ST
Process	Confirm the schedule is up to date, perform any updates, and uploads the schedule to the intranet.	Scheduling	0.5	100%
Total (minutes FTE per ST)			0.5	

Automated Unconditional Actions

Input: Daily schedule.

Issue work management daily schedule.

Action	AI Description			
	Training/Reference Input	Use Input	Functions	Output
Process	Daily schedule	Completed work schedule (updated by the staff during execution)	Summarization	Summary of work schedule changes and daily schedule

Page intentionally left blank

Appendix E

Mapping of Execution

Page intentionally left blank

Appendix E

Mapping of Execution

This appendix applies the approach of Section 4.1 to work execution steps according to the flow shown in Figure 20. The following sections discuss each of those steps in detail.

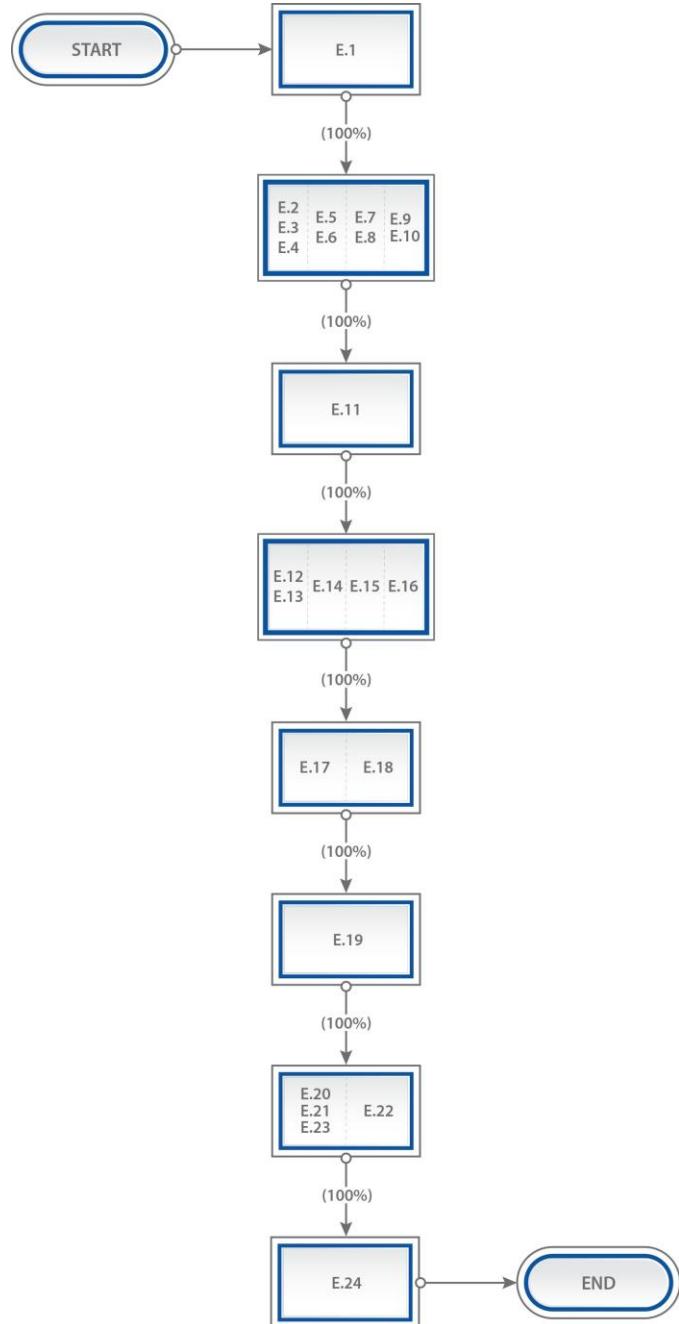


Figure 20. Flowchart of the work execution steps that are discussed in Appendix E.

E-1. Daily Work Schedule Review

Each work group supervisor reviews the daily schedule within their own groups, and then there is a meeting each morning of execution week before the day shift in which collegial discussions occur between the work groups on issues that need resolution that day.

Current Unconditional Actions

Input: Current WP and Week schedule.

Review daily schedule.

Action	Effort Description	Discipline	Time in Minutes	Frequency per WP
Process	Check the daily schedule for applicable tasks for the department to ensure task readiness.	All relevant groups	10	100%
Total (minutes FTE per WP)			10	

Current Unconditional Actions

Input: Current WP and Week schedule.

Review daily schedule.

Action	AI Description			
	Training/Reference Input	Use Input	Functions	Output
Process	NA*			

* This cannot be automated since supervisors need to familiarize themselves with the scheduled work.

E-2. RP Work Associated with the Work Package

Before maintenance work can be performed, RP needs to check if the work will be done in a radiological area to perform any necessary work such as dosimetry, review RWP, log into RPA.

Current Decisions and Conditional Actions

Input: Current WP, daily schedule, and dose-tracking database.

Do RP predecessor tasks need to be performed?

Decision	Effort Description	Discipline	Time in Minutes	Frequency per WP
Process	Check the WP to determine if there are any RP predecessor tasks that must be completed ahead of the maintenance work.	RP	0.5	100%
If yes	Get dosimetry, review RWP, and log into RPA if needed.	RP	15	25%
If no	None.	NA	0	75%
Total (minutes FTE per WP)			4.25	

Automated Decisions and Conditional Actions

Input: Current WP, daily schedule, radiological control procedure and dose-tracking database.

Do RP predecessor tasks need to be performed?

Decision	AI Description			
	Training/Reference Input	Use Input	Functions	Output
Process	Radiological control words or general-knowledge-enabled LLM	Current WP	Classification (unsupervised)	Yes/No
	Radiological control procedure	Current WP	Review	Yes/No
If yes	Dose-tracking database and RWP	Current WP	Generation	Radiological impact
If no	NA			

E-3. RP Work Execution Week Survey

Depending on the scope of the work, a new survey might be necessary if there are outdated surveys or concerns about high doses or contamination. In this step, the scope of each activity is assessed to decide if a new survey is required before the work proceeds.

Current Decisions and Conditional Actions

Input: Current WP, daily schedule, and RP survey maps.

Does the work evolution need a new survey?

Decision	Effort Description	Discipline	Time in Minutes	Frequency per WP
Process	Determine if the work location poses a risk of high-dose exposure or contamination. Check whether recent surveys have been completed or if new surveys are needed before executing the work.	RP	5*	25%
If yes	Perform survey.	RP	60	3%
If no	None.	NA	0	22%
Total (minutes FTE per WP)			3.05	
* These frequencies are different based on the type of plant—boiling water reactor vs. pressurized water reactor).				

Automated Decisions and Conditional Actions

Input: Current and historical WPs, daily schedule, RP survey dataset, and radiological control procedure.

Does the work evolution need a survey?

Decision	AI Description			
	Training/Reference Input	Use Input	Functions	Output
Process	Radiological survey words or general-knowledge-enabled LLM	Current WP	Classification (unsupervised)	Yes/No
	Historical WPs and survey assignment	Current WP	Structuring and classification (supervised)	Yes/No
	RP survey dataset and radiological control procedure	Current WP	Review	Yes/No
If yes	NA			
If no	NA			

E-4. RP Work Schedule Updates

RP is required to update the work schedule each shift to ensure that schedule compliance will be achieved if problems occur during execution of work, or to allow for the completion of other related work activities.

Current Unconditional Actions

Input: Current WP and ST and daily schedule.

Provide ST status updates.

Action	Effort Description	Discipline	Time in Minutes	Frequency per WP
Process	Provide updates to the ST as the work starts, progresses, and completes. Those updates can be just marking steps completed or provided details description of what was accomplished.	RP	2	25%
Total (minutes FTE per WP)			0.5	

Current Unconditional Actions

Input: Current and historical WPs and Current ST, daily schedule.

Provide ST status updates.

Action	AI Description			
	Training/Reference Input	Use Input	Functions	Output
Process	Historical WPs	Current WP and ST	Guidance	Notes and updates
	Historical WPs	Current WP and ST	Insertion	Notes and updates

E-5. Operations CO Associated with the Work Package

Operations reviews the work scope for each WO in the execution day and determines if the work will need a tag-out to protect the maintenance workers during the execution of the work.

Current Decisions and Conditional Actions

Input: Current WP and daily schedule.

Does the WP have a CO?

Decision	Effort Description	Discipline	Time in Minutes	Frequency per WP
Process	Identify if a CO is associated with the WP before work execution.	Operations	1	100%
If yes	Perform tagout, log into RPA dosimetry information, if applicable, and/or configure the plant as needed.	Operations	Out of scope	20%
If no	None.	NA	0	80%
Total (minutes FTE per WP)			1	

Automated Decisions and Conditional Actions

Input: Current WP, historical COs, and daily schedule.

Does the WP have a CO?

Decision	AI Description			
	Training/Reference Input	Use Input	Functions	Output
Process	CO words or general-knowledge-enabled LLM	Current WP	Classification (unsupervised)	Yes/No
	Historical COs	Current WP	Review	Yes/No
If yes	NA			
If no	NA			

E-6. Operations Work Schedule Update

Operations is required to update the work schedule each shift to ensure that schedule compliance will be achieved if problems occur during execution of work, or to allow for the completion of other related work activities.

Current Unconditional Actions

Input: Current WP and ST and daily schedule.

Provide ST status updates.

Action	Effort Description	Discipline	Time in Minutes	Frequency per WP
Process	Provide updates to the ST as the work starts, progresses, and completes. Those updates can be just marking steps completed or providing a detailed description of what was accomplished.	Operations	2	20%
Total (minutes FTE per WP)			0.4	

Current Unconditional Actions

Input: Current and historical WPs and Current ST, daily schedule.

Provide ST status updates.

Action	AI Description			
	Training/Reference Input	Use Input	Functions	Output
Process	Historical WPs	Current WP and ST	Guidance	Notes and updates
	Historical WPs	Current WP and ST	Insertion	Notes and updates

E-7. Engineering Determination of Work

Engineering determines whether they need to assist with work or perform duties (inspections, surveillances, and potential troubleshooting) in the plant in conjunction with work being performed by other work groups.

Current Decisions and Conditional Actions

Input: Current WP and daily schedule.

Does engineering have work to be performed?

Decision	Effort Description	Discipline	Time in Minutes	Frequency per WP
Process	Determine if the WP involves any inspections, surveillances, and/or potential troubleshooting.	Engineering	1	100%
If yes	Prepare to perform inspections, surveillances, and/or plan troubleshooting as applicable.	Engineering	30	10%
If no	None.	NA	0	90%
Total (minutes FTE per WP)			4	

Automated Decisions and Conditional Actions

Input: Current WP and daily schedule.

Does engineering have work to be performed?

Decision	AI Description			
	Training/Reference Input	Use Input	Functions	Output
Process	Inspections, surveillances, and troubleshooting keywords or general-knowledge-enabled LLM	Current WP	Classification (Unsupervised)	Yes/No
If yes	Inspections, surveillances, and troubleshooting documents	Current WP	Summarization	Specific engineering scope
If yes	NA			
If no	NA			

E-8. Engineering Work Schedule Update

Engineering is required to update the work schedule each shift to ensure that schedule compliance will be achieved if problems occur during execution of work, or to allow for the completion of other related work activities.

Current Unconditional Actions

Input: Current WP and ST and daily schedule.

Provide ST status updates.

Action	Effort Description	Discipline	Time in Minutes	Frequency per WP
Process	Provide updates to the ST as the work starts, progresses, and completes. Those updates can be just marking steps completed or providing a detailed description of what was accomplished.	Engineering	2	10%
Total (minutes FTE per WP)			0.2	

Current Unconditional Actions

Input: Current and historical WPs and Current ST, daily schedule.

Provide ST status updates.

Action	AI Description			
	Training/Reference Input	Use Input	Functions	Output
Process	Historical WPs	Current WP and ST	Guidance	Notes and updates
	Historical WPs	Current WP and ST	Insertion	Notes and updates

E-9. Other Work Groups Predecessor Tasks Prior to Maintenance

Before work can be performed, many tasks need to be completed by other organizations, like security, etc. In this step, the working groups determine if they need to perform any work tasks prior to the actual execution work.

Current Decisions and Conditional Actions

Input: Current WP and daily schedule.

Do predecessor tasks need to be performed?

Decision	Effort Description	Discipline	Time in Minutes	Frequency per WP
Process	Check the WP to determine if there are any predecessor tasks that must be completed ahead of the maintenance work.	All relevant groups	1	100%
If yes	Prepare and execute the task.	All relevant groups	30	10%
If no	None.	NA	0	90%
Total (minutes FTE per WP)			4	

Automated Decisions and Conditional Actions

Input: Current WP and daily schedule.

Do predecessor tasks need to be performed?

Decision	AI Description			
	Training/Reference Input	Use Input	Functions	Output
Process	Keywords or general-knowledge-enabled LLM	Current WP	Classification (Unsupervised)	Yes/No
If yes	NA*			
If no	NA			

* This depends on conditions outside the work management (i.e., plant conditions) and can't be automated.

E-10. Other Groups Work Schedule Update

Other groups are required to update the work schedule each shift to ensure that schedule compliance will be achieved if problems occur during execution of work, or to allow for the completion of other related work activities.

Current Unconditional Actions

Input: Current WP and ST and daily schedule.

Provide ST status updates.

Action	Effort Description	Discipline	Time in Minutes	Frequency per WP
Process	Provide updates to the ST as the work starts, progresses, and completes. Those updates can be just marking steps completed or providing a detailed description of what was accomplished.	All relevant groups	2	10%
Total (minutes FTE per WP)			0.2	

Current Unconditional Actions

Input: Current and historical WPs and Current ST, daily schedule.

Provide ST status updates.

Action	AI Description			
	Training/Reference Input	Use Input	Functions	Output
Process	Historical WPs	Current WP and ST	Guidance	Notes and updates
	Historical WPs	Current WP and ST	Insertion	Notes and updates

E-11. Maintenance Determination of Work Predecessor Tasks

In this step, maintenance determines if all predecessor tasks indicated in the WP are completed and maintenance is clear to start.

Current Decisions and Conditional Actions

Input: Current WP and ST, daily schedule.

Are all predecessor tasks to the main work scope completed?

Decision	Effort Description	Discipline	Time in Minutes	Frequency per WP
Process	Check the WP to verify that listed predecessor tasks are performed.	Maintenance	3	100%
If yes	Proceed.	Maintenance	0	90%
If no	Wait or contact responsible department.	Maintenance	0	10%
Total (minutes FTE per WP)			3	

Automated Decisions and Conditional Actions

Input: Current WP and ST, daily schedule.

Are all predecessor tasks to the main work scope completed?

Decision	AI Description			
	Training/Reference Input	Use Input	Functions	Output
Process	WP predecessor task documents and STs	Current WP	Summarization	Summary of all completed and pending tasks
If yes	NA			
If no	NA			

E-12. Pick-up of Materials and Parts from Warehouse

In most cases when maintenance is performed, materials and parts are needed to facilitate the maintenance. In this step, materials and parts are issued to maintenance by the warehouse personnel.

Current Decisions and Conditional Actions

Input: Current WP and daily schedule.

Do any materials and parts need to be picked up from warehouse?

Decision	Effort Description	Discipline	Time in Minutes	Frequency per WP
Process	Check if any materials and parts are needed and are reserved and ready in the warehouse.	Maintenance	2	100%
If yes	Pick up materials and parts from warehouse.	Maintenance	15	75%
If no	None.	NA	0	25%
Total (minutes FTE per WP)			13.25	

Automated Decisions and Conditional Actions

Input: Current WP and daily schedule.

Do any materials and parts need to be picked up from warehouse?

Decision	AI Description			
	Training/Reference Input	Use Input	Functions	Output
Process	NA*			
If yes	NA			
If no	NA			

* This a data query as the WO and WP will be associated with the parts.

E-13. Temporary Materials and Parts Storage

If materials and parts are necessary to complete a work activity, then maintenance is required by procedure to set up a temporary storage area. This is done for personnel safety and to protect the materials and parts from work being performed in proximity to the subject work activity.

Current Decisions and Conditional Actions

Input: Current WP and daily schedule.

Does a temporary materials and parts storage area need to be established?

Decision	Effort Description	Discipline	Time in Minutes	Frequency per WP
Process	Check the materials and parts list to identify if the number of materials and parts requires that a temporary storage area be established.	Maintenance	2	75%
If yes	Set up storage area.	Maintenance	30	15%
If no	No storage area needed.	Maintenance	0	60%
Total (minutes FTE per WP)			6	

Automated Decisions and Conditional Actions

Input: Current WP and daily schedule.

Does a temporary materials and parts storage area need to be established?

Decision	AI Description			
	Training/Reference Input	Use Input	Functions	Output
Process	NA*			
If yes	NA			
If no	NA			

* This step relies on the number of parts from the previous step since the need for storage is not labeled.

E-14. Check-out of Tools for Work

In this step, maintenance personnel check out tools from the tool crib as needed. This process is a key component of budgetary control and asset management. By tracking the use of tools, the organization can optimize tool usage, prevent loss or misuse, and ensure that tools are returned in good condition.

Current Decisions and Conditional Actions

Input: Current WP and daily schedule.

Do any tools need to be checked out for work?

Decision	Effort Description	Discipline	Time in Minutes	Frequency per WP
Process	Check the WP to identify tools needed to be checked out for work execution.	Maintenance	2	100%
If yes	Check out tools from tool crib.	Maintenance	15	75%
If no	No tools needed to be checked out.	Maintenance	0	25%
Total (minutes FTE per WP)				13.25

Automated Decisions and Conditional Actions

Input: Current WP, tools list, and daily schedule.

Do any tools need to be checked out for work?

Decision	AI Description			
	Training/Reference Input	Use Input	Functions	Output
Process	Tools keywords or general-knowledge-enabled LLM	Current WP	Classification (unsupervised)	Yes/No
	Tools list	Current WP	Comparison	Needed tools
If yes	NA			
If no	NA			

E-15. Check-out of Calibrated Tools and Equipment

Some safety and reliability-related work will need to be performed with precision tools and equipment. Those tools usually exist in a different storage area. In this step, calibrated tools and equipment are checked out for the work in the plant.

Current Decisions and Conditional Actions

Input: Current WP and daily schedule.

Are any calibrated tools or equipment needed to perform work?

Decision	Effort Description	Discipline	Time in Minutes	Frequency per WP
Process	Check the WP to identify if calibrated tools or equipment need to be checked out for work.	Maintenance	2	100%
If yes	Check out calibrated tools and equipment.	Maintenance	30	25%
If no	None.	NA	0	75%
Total (minutes FTE per WP)			9.5	

Automated Decisions and Conditional Actions

Input: Current WP, calibrated tools and equipment list, and daily schedule.

Are any tools or calibrated equipment needed to perform work?

Decision	AI Description			
	Training/Reference Input	Use Input	Functions	Output
Process	Calibrated tools and equipment keywords or general-knowledge-enabled LLM	Current WP	Classification (unsupervised)	Yes/No
	Calibrated tools and equipment list	Current WP	Comparison	Needed calibrated tools and equipment
If yes	NA			
If no	NA			

E-16. Chemicals and Lubricants

All chemicals utilized in the plant are controlled due to regulations, to protect plant chemistry, and for personnel safety. In this step, maintenance obtains chemicals and lubricants as well as the documentation authorizing the use of the chemicals in the plant.

Current Decisions and Conditional Actions

Input: Current WP and daily schedule.

Are any chemicals or lubricants needed for work?

Decision	Effort Description	Discipline	Time in Minutes	Frequency per WP
Process	Check the WP to identify if any chemical and lubricants needed to perform the work.	Maintenance	2	100%
If yes	Gather chemicals and lubricants.	Maintenance	30	10%
If no	No chemicals or lubricants needed.	Maintenance	0	90%
Total (minutes FTE per WP)			5	

Automated Decisions and Conditional Actions

Input: Current WP, chemicals and lubricants list, and daily schedule.

Are any chemicals or lubricants needed for work?

Decision	AI Description			
	Training/Reference Input	Use Input	Functions	Output
Process	Chemicals and lubricants keywords or general-knowledge-enabled LLM	Current WP	Classification (Unsupervised)	Yes/No
	Chemicals and lubricants list	Current WP	Comparison	Needed chemicals and lubricants
If yes	NA			
If no	NA			

E-17. Maintenance Check-in with Operations

Each maintenance crew that is scheduled to perform work in the plant will go to the work execution center and discuss the work activity scope with operations. This is done to make sure that workers are aware of hazards while they are working, and that operations are aware of work being performed in the plant.

Current Decisions and Conditional Actions

Input: Current WP and daily schedule.

Is the plant in a safe condition and in the proper configuration for maintenance to begin work?

Decision	Effort Description	Discipline	Time in Minutes	Frequency per WP
Process	Discuss the work activities with operations and determine if they can be performed based on current plant conditions.	Maintenance	10*	100%
If yes	Verify tagout, and log into RPA dosimetry information if applicable.	Maintenance	Out of scope	95%
If no	Notify workweek manager for resolution.	Maintenance	1	5%
Total (minutes FTE per WP)		10.05		

* Assuming 2 staff discuss for 5 minutes, this would result in 10 minutes per WP.

Automated Decisions and Conditional Actions

Input: Current WP and daily schedule.

Is the plant in a safe condition and in the proper configuration for maintenance to begin work?

Decision	AI Description			
	Training/Reference Input	Use Input	Functions	Output
Process	NA			
If yes	NA			
If no	NA			

E-18. Maintenance FME Zone Actions

When there is a possibility that foreign material can be introduced into process systems during work execution, then FME zone actions are necessary to prevent foreign material intrusion.

Current Decisions and Conditional Actions

Input: Current WP and daily schedule.

Are there FME zones associated with work to be performed?

Decision	Effort Description	Discipline	Time in Minutes	Frequency per WP
Process	Check the WP to identify whether FME zone actions will need to be performed to prevent foreign material intrusion during the work activity.	Maintenance	1*	100%
If yes	Set up FME zones.	Maintenance	30	5%
If no	None.	NA	0	95%
Total (minutes FTE per WP)			2.5	

* With most WPs, it is a trivial effort to determine whether the work will need FME zone. Some (about 10%), however, require more research (around 10 minutes). This averages to 1 minute per WP.

Automated Decisions and Conditional Actions

Input: Current WP and daily schedule.

Are there FME zones associated with work to be performed?

Decision	AI Description			
	Training/Reference Input	Use Input	Functions	Output
Process	FME zone keywords or general-knowledge-enabled LLM	Current WP	Classification (unsupervised)	Yes/No
	FME zone locations	Current WP	Review	Yes/No
If yes	NA			
If no	NA			

E-19. Work Groups Perform Work

In this step, the subject work group performs the work in the plant based on the WP associated with the work activity.

Current Unconditional Actions

Input: Current WP and daily schedule.

Perform work.

Action	Effort Description	Discipline	Time in Minutes	Frequency per WP
Process	Perform field activity.	All relevant groups	Out of scope	100%
Total (minutes FTE per WP)		0		

* This is mainly a physical activity with variable duration and is outside the scope of this work.

Automated Unconditional Actions

Input: Current WP and daily schedule.

Perform work.

Action	AI Description			
	Training/Reference Input	Use Input	Functions	Output
Process	NA			

E-20. Work Groups Update of Work Package Notes/Status

During the performance of the WP, all groups will need to log their actions and plant conditions encountered while performing work. This is done to ensure procedural requirements have been maintained as well as help with condition-based monitoring.

Current Unconditional Actions

Input: Current WP.

Provide ST status updates.

Action	Effort Description	Discipline	Time in Minutes	Frequency per WP
Process	Update the WP based on the work execution and document any interactions.	Maintenance	2	100%
Total (minutes FTE per WP)				2

Automated Unconditional Actions

Input: Current and historical WPs.

Provide ST status updates.

Action	AI Description			
	Training/Reference Input	Use Input	Functions	Output
Process	Historical WPs	Current WP	Guidance	Notes and updates
	Historical WPs	Current WP	Insertion	Notes and updates

E-21. Maintenance Supervisor Approval of Work

Once the work is complete in the field, the supervisor will ensure that the full scope of work has been completed and that the work that was performed has been adequately documented. Once verified, the supervisor will approve the completion of the WP.

Current Decisions and Conditional Actions

Input: Current WP.

Is the work completed and documented?

Decision	Effort Description	Discipline	Time in Minutes	Frequency per WP
Process	Review the WP and discusses with staff to determine if work is performed and documented adequately.	Maintenance	5	100%
If yes	Supervisor updates WP status.	Maintenance	5	95%
If no	Rework or improve documentation.	Maintenance	60	5%
Total (minutes FTE per WP)				12.75

Automated Decisions and Conditional Actions

Input: Current and historical WPs and work management procedure.

Is the work completed and documented?

Decision	AI Description			
	Training/Reference Input	Use Input	Functions	Output
Process	Work management procedure	Current WP	Review	Yes/No
If yes	Historical WPs	Current WP	Guidance	Notes and updates
	Historical WPs	Current WP	Insertion	Notes and updates
If no	Work management procedure	Current WP	Generation	Notes on improvement or recommendation for rework

E-22. Maintenance Work Schedule Updates

Maintenance is required to update the work schedule each shift to ensure that schedule compliance will be achieved if problems occur during execution of work, or to allow for the completion of other related work activities.

Current Unconditional Actions

Input: Current WP and ST and daily schedule.

Provide ST status updates.

Action	Effort Description	Discipline	Time in Minutes	Frequency per WP
Process	Provide updates to the ST as the work starts, progresses, and completes. Those updates can be just marking steps completed or providing a detailed description of what was accomplished.	Maintenance	2	100%
Total (minutes FTE per WP)				2

Current Unconditional Actions

Input: Current and historical WPs and Current ST, daily schedule.

Provide ST status updates.

Action	AI Description			
	Training/Reference Input	Use Input	Functions	Output
Process	Historical WPs	Current WP and ST	Guidance	Notes and updates
	Historical WPs	Current WP and ST	Insertion	Notes and updates

E-23. Work Package Closeout

After work is complete, all working departments are required to complete the WP documentation. This includes updating the as found and as left equipment condition. This task is mainly for maintenance to complete, but may also involve engineering, operations, RP, etc.

Current Unconditional Actions

Input: Current WO and WP.

Close out WP.

Action	Effort Description	Discipline	Time in Minutes	Frequency per WP
Process	Update as-left condition and work completion notes and mark WP as complete.	All relevant groups	15	100%
Total (minutes FTE per WP)			15	

Automated Unconditional Actions

Input: Current and historical WPs.

Close out WP.

Action	AI Description			
	Training/Reference Input	Use Input	Functions	Output
Process	Historical WPs	Current WP	Guidance	Notes and updates
	Historical WPs	Current WP	Insertion	Notes and updates

E-24. Work Package Completion Check

After WP is closed, a review is conducted by records/document management personnel to ensure all signatures are there and all work has been completed.

Current Decisions and Conditional Actions

Input: Current WP and work management procedure.

Is WP completed properly?

Decision	Effort Description	Discipline	Time in Minutes	Frequency per WP
Process	Check WPs for completeness including all signatures.	Document control	10	100%
If yes	None.	NA	0	95%
If no	Add notes and send back to responsible entity for missing work.	Document control	30	5%
Total (minutes FTE per WP)				11.5

Automated Decisions and Conditional Actions

Input: Current and historical WPs and work management procedure.

Is WP completed properly?

Decision	AI Description			
	Training/Reference Input	Use Input	Functions	Output
Process	Work management procedure	Current WP	Review	Yes/No
If yes	NA			
If no	Work management procedure	Current WP	Generation	Notes for missing parts of the WP

Appendix F

Mapping of Post-Execution

Page intentionally left blank

Appendix F

Mapping of Post-Execution

This appendix applies the approach of Section 4.1 to work post-execution steps according to the flow shown in Figure 21. The following sections discuss each of those steps in detail.

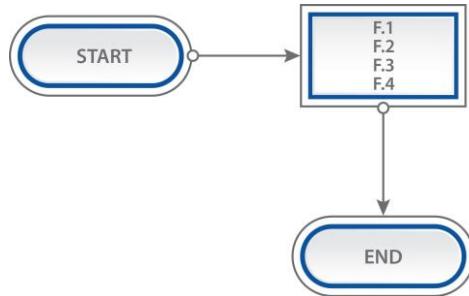


Figure 21. Flowchart of the work post-execution steps that are discussed in Appendix F.

F-1. Work Management Evaluation of Week Performance Information

Once execution week is complete, the work management group will gather execution week information to ensure that they can continuously improve the work management process. This information will be fed into performance indicators and will be used during the post-execution week critique.

Current Unconditional Actions

Input: Completed WPs and week schedule.

Gather information about the week and generate a report.

Action	Effort Description	Discipline	Time in Minutes	Frequency per WP
Process	Run reports to provide results on any outstanding tasks that need to be performed and document any insights on work performance.	Workweek manager	2	100%
Total (minutes FTE per WP)			2	

Automated Unconditional Actions

Input: Completed WPs, week schedule, and historical performance reports,

Gather information about the week and generate a report.

Action	AI Description			
	Training/Reference Input	Use Input	Functions	Output
Process	Week schedule and historical performance reports	Completed WP	Generation	Week performance report
	Week schedule and historical performance reports	Completed WP	Summarization	Summary of key week performance

F-2. Relevant Groups Analysis of Execution Week Performance Information

Information that is gathered by work management is provided to each department's supervisor for analysis. Any performance shortfalls are analyzed in preparation for the post-execution week critique meeting.

Current Unconditional Actions

Input: Completed WPs, week schedule, week performance report, and work group documents.

Analyze weekly report.

Action	Effort Description	Discipline	Time in Minutes	Frequency per WP
Process	Review performance reports and identify any department shortfalls.	All relevant groups	1	100%
Total (minutes FTE per WP)			1	

Automated Unconditional Actions

Input: Completed WPs, week schedule, week performance report, and work group documents.

Analyze weekly report.

Action	AI Description			
	Training/Reference Input	Use Input	Functions	Output
Process	Week schedule, week performance report and work group documents	Completed WP	Review	Performance shortfalls

F-3. Post-Execution Week Performance Critique Meeting

All relevant groups representatives meet to discuss the past week's execution of work, compare it to key performance indicators, and analyze any issues that came up during the execution week.

Current Unconditional Actions

Input: Completed WPs, week schedule, work performance report, and work group documents.

Discuss performance in a meeting.

Action	Effort Description	Discipline	Time in Minutes	Frequency per WP
Process	Participate in a meeting and provide feedback on week schedule to include in future work.	All relevant groups	0.75	100%
Total (minutes FTE per WP)				0.75

* Assuming 5 staff meet for 30 minutes and 200 WPs per week, this would result in less than 0.75 minutes per WP.

Automated Unconditional Actions

Input: Completed WPs, week schedule, work performance report, and work group documents.

Discuss performance in a meeting.

Action	AI Description			
	Training/Reference Input	Use Input	Functions	Output
Process	NA			

F-4. Performance Shortfalls Documentation

If any of the issues that came up during the post-execution week critique meet the threshold for a CR, then a CR will be generated for more detailed analysis of the issue.

Current Decisions and Conditional Actions

Input: Completed WPs, week schedule, work performance report, critique meeting notes, and work group documents.

Are there any issues to report?

Decision	Effort Description	Discipline	Time in Minutes	Frequency per WP
Process	Determine if the WP issues or performance shortfalls discussed in the meeting include an issue that requires comprehensive review and evaluation.	All relevant groups	1	100%
If yes	Report issues or shortfalls in a CR.	All relevant groups	1	5%
If no	None.	NA	0	95%
Total (minutes FTE per WP)			1.05	

Automated Unconditional Actions

Input: Completed WPs, week schedule, work performance report, critique meeting notes, and work group documents.

Are there any issues to report?

Decision	AI Description			
	Training/Reference Input	Use Input	Functions	Output
Process	Work management procedure for CRs criteria	Completed WP and critique meeting notes	Review	Yes/No
If yes	Historical CRs	Critique meeting notes	Generation	New CR
If no	NA			