Development of Probabilistic Risk Assessment Methodology Using Artificial Intelligence Technology
4. Automatic Fault Detection Method for Building Reliability Database-Analysis and Classification Method for Incidents with Complex Causal Networks

Hiroshi UJITA [a*], Tatsuya MORIMOTO [a], Satoshi FUTAGAMI [b], Kenichi KURISAKA [b], Hidemasa YAMANO [b]
[a] AdvanceSoft Corporation, Chiyoda-ku, Tokyo, Japan
[b] Japan Atomic Energy Agency, Narita-cho, Ibaraki, Japan

Abstract: For the construction of reliability database in Japan, the Central Research Institute of Electric Power Industry and the Japan Nuclear Safety Institute have been developing the NUCIA database for light water reactors, and the Japan Atomic Energy Agency has been developing the CORDS database for sodium-cooled fast reactors. Recently generative artificial intelligence (AI) technologies improvement is remarkable, it is believed that the solutions can be resolved by using generative AI as chat GPT, with the greatest expectation being the possibility of discovering new insights that may not be discernible even by the individual expert. By creating an event scenario network diagram, it becomes possible to structure the content described in NUCIA in a unified format (which can be considered as an advanced summarization process), while a rule-based classification method based on past cases is appropriate for CORDS. Therefore the determination can be addressed through text classification processing. The text was converted into vectors considering the context, and the distance between vectors were calculated to determine text similarity.

Keywords: Reliability Database, Text Mining, Data Mining, Generative AI, Rule-based Approach.

1. INTRODUCTION

For the construction of reliability database in Japan, the Central Research Institute of Electric Power Industry and the Japan Nuclear Safety Institute have been developing the NUCIA database for light water reactors [2], and the Japan Atomic Energy Agency has been developing the CORDS database for sodium-cooled fast reactors [1]. These databases allow us referencing individual cases and support analysis through database searches such as keyword retrieval and statistical functions. These have also been used for creating failure rate databases.

However, due to the vast amount of data, significant human resources have been expended on analysis. Additionally, the involvement of many analysts has led to potential variability and bias in the analyses (referred to as primary analysis, targeted in this study for FY2022). Furthermore, the sheer volume of data makes it challenging to overview the entirety and extract the temporal and spatial characteristics of events, which is considered the most significant issue (referred to as secondary analysis, targeted in this study for FY2023-2024). Therefore, it is believed that utilizing artificial intelligence (AI) techniques can solve these issues as described below. The greatest expectation lies in the potential to discover new insights through AI analysis that human analysts may not find.

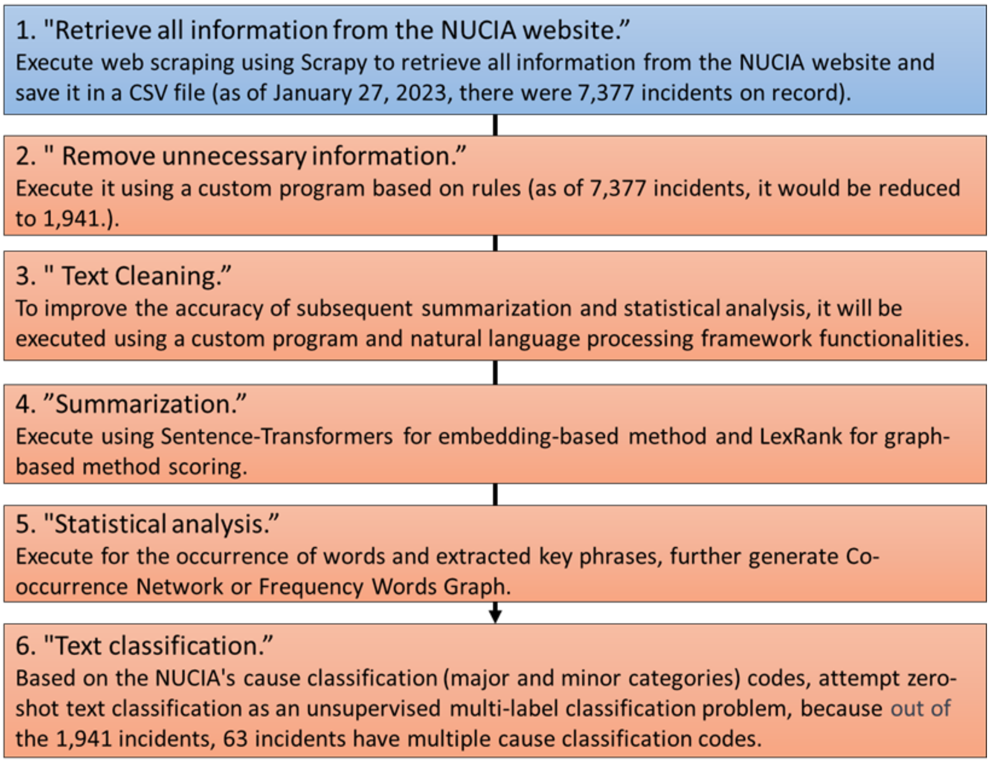
- Issue 1 (Primary Analysis): Improving the Efficiency of Reliability Database Construction
  - Promoting speed and labor-saving (big data processing)
  - Promoting accuracy and uniformity (eliminating individual differences among analysts)
- Issue 2 (Secondary Analysis): Enhancing Analytical Capabilities for Extracting Temporal and Spatial Characteristics
  - Improving uniformity (promoting comprehensive analysis using text mining and data mining techniques)
  - Acquiring new insights (extracting common factors and failure characteristics such as temporal and spatial features through automatic big data processing)

In order to calculate failure rates in a failure rate database, it is crucial to accurately understand the causes indicated in the failure information. The NUCIA database basically adopts a method of identifying one cause for each case. However, in practice, there are cases where multiple causes exist within a single case or where there are chains of causality. Extracting these comprehensively is essential for creating an accurate database. By utilizing AI methods, it is expected that the appropriate causes can be identified from complex causal relationships, thereby contributing to the accurate calculation of failure rates.

Recently generative AI technologies improvement is remarkable, it is believed that the following solutions can be resolved by using generative AI as chat GPT, with the greatest expectation being the possibility of discovering new insights that may not be discernible even by the individual expert.

2. OVERVIEW OF THE DEVELOPED AI TOOL IN FY2023

In FY2022, as information necessary for constructing the reliability database, the methodology of an AI tool was developed and prototyped to extract and database failure occurrences (systems and equipment) and causes from NUCIA using AI technologies [3]. The overall flow is shown in Figure 1.



1. "Retrieve all information from the NUCIA website."
Execute web scraping using Scrapy to retrieve all information from the NUCIA website and save it in a CSV file (as of January 27, 2023, there were 7,377 incidents on record).

2. " Remove unnecessary information."
Execute it using a custom program based on rules (as of 7,377 incidents, it would be reduced to 1,941.).

3. " Text Cleaning."
To improve the accuracy of subsequent summarization and statistical analysis, it will be executed using a custom program and natural language processing framework functionalities.

4. ”Summarization."
Execute using Sentence-Transformers for embedding-based method and LexRank for graph-based method scoring.

5. "Statistical analysis."
Execute for the occurrence of words and extracted key phrases, further generate Co-occurrence Network or Frequency Words Graph.

6. "Text classification."
Based on the NUCIA's cause classification (major and minor categories) codes, attempt zero-shot text classification as an unsupervised multi-label classification problem, because out of the 1,941 incidents, 63 incidents have multiple cause classification codes.

Figure 1 Development Flow of AI tool for NUCIA database (2022 Fiscal Year).

We have prototyped using so-called "traditional" natural language processing techniques. However, the release of ChatGPT, a chat-based generative AI using OpenAI's large language model GPT-3.5, in November 2022, and the subsequent release of an even larger and more precise model, GPT-4, by OpenAI in March 2023, drastically changed the landscape. It became possible to easily and accurately handle natural language-related processing that had previously been considered difficult. Therefore, in FY2023, new trial system of ChatGPT was developed using the process flow building on the element technologies prototyped in FY2022.
The key point is the creation of an event scenario network diagram (structured data) from the data described in NUCIA (unstructured data in the form of report texts) and its use in analysis. To explain the reason for creating the event scenario network diagram, we outline the characteristics of the trouble information reports in NUCIA and the corresponding countermeasures (as examined in FY2022) below.

- Multiple causes
  Identifying a cause (cause classification) is necessary for failure determination. Often, existing cause classifications select only one cause for a single trouble report, but is this truly accurate? Moreover, are the existing classification items themselves appropriate? Therefore as for the countermeasure, multi-label classification based on appropriate classification items is required.
- Event scenario network (Complex causal network)

Is the information necessary for identifying causes truly comprehensive? Can missing information be interpolated? Therefore as for the countermeasure, constructing an event scenario from root cause to event occurrence and preventive measures, and visualizing it as a network diagram, can potentially confirm missing information and provide highly explanatory cause identification. Additionally, it should facilitate the analysis of event and time differences.

By creating an event scenario network diagram, it becomes possible to structure the content described in NUCIA in a unified format (which can be considered an advanced summarization process) and to clarify the event scenarios. In this process, "direct causes" corresponding to PRA failure modes, as well as "root causes" such as human error, organizational factors, or safety culture, can be extracted, enabling true cause classification. Furthermore, by understanding the commonality of "direct causes" and "root causes" among events as a graph network, this method is expected to be effective for the common characteristics analysis targeted in FY2024. The key points are discussed in Chapter 3.

3. NEW APPROACH TO ANALYZE COMPLEX CAUSAL NETWORKS

3.1. Extraction of Key Phrases and Relationships for Specified Items

By using the multimodal large language model GPT-4 (gpt-4-0125-preview released in January 2024) developed by OpenAI, we created JSON format data (structured data) necessary to create event scenario network diagrams from the data recorded in NUCIA (unstructured report texts). Generally, the larger the language model, the higher the response accuracy, and the newer the model, the more up-to-date information and better tuning it has. Thus, we adopted gpt-4-0125-preview, which was the latest at the time of this trial. Another option was GPT-3.5 (gpt-3.5-turbo-0125) by OpenAI, which is cost-effective. However, considering the number of parameters (though not disclosed, it is significantly fewer than GPT-4), it was deemed less accurate and therefore not adopted. Additionally, Claude3 by Anthropic, which has reported superior performance to GPT-4, was released in March 2024 and hence was unavailable at the time of this trial (the same applied to Meta's Llama3). Other models were excluded from the outset due to their lower response accuracy based on the number of parameters, ease of implementation, required memory, and processing speed when using free models locally.

Based on the analysis of the content recorded in NUCIA by our experts, we examined the ontology (information representation specification) shown in Figure 2 and set it as the input prompt for GPT-4. Specifically, we defined roles, instructions, conditions, and each item (response, observation, direct cause, intermediate cause, root cause, preventive measures), and organized key phrases and their causal relationships for each item in the form of nodes and edges as the input prompt. This allowed the content recorded for each event in NUCIA to be organized according to the same specification.

Furthermore, by using OpenAI's Function Calling feature, we limited GPT-4's responses to the JSON format for network diagram creation. The Function Calling feature automatically generates input for functions in pre-defined JSON format based on the input prompt content and automatically selects the functions to be used. This ensured that the JSON format for network diagram creation was reliably generated.

3.2. Knowledge Base Creation

The content recorded in NUCIA, the entire JSON format for network diagram creation, and the extracted "direct causes" and "root causes" were stored in an sqlite3 database. SQLite is a serverless relational database management system that manages the database in a single file through a C language library. The sqlite3 module, included in Python's standard library, provides access to and manipulation of the SQLite database, functioning as a binding to the C language implementation of the SQLite engine. The configuration of the created database includes 36 items.

In this manner, while we created the knowledge base using a relational database in 2023, we plan to consider the use of a graph database in 2024, as we aim to analyze the data as a network graph, as mentioned above.

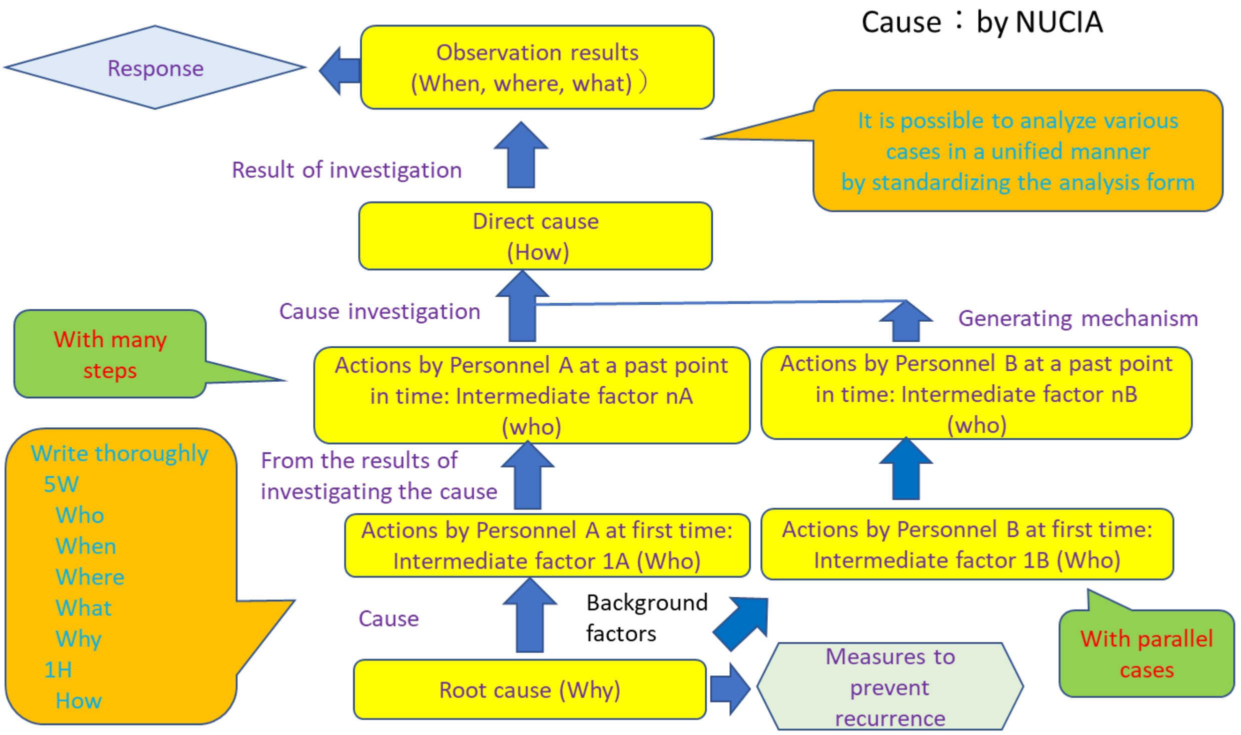

Cause：by NUCIA
Response
Observation results (When, where, what)）
It is possible to analyze various cases in a unified manner by standardizing the analysis form
Result of investigation
Direct cause (How)
Cause investigation
Generating mechanism
With many steps
Actions by Personnel A at a past point in time: Intermediate factor nA (who)
Actions by Personnel B at a past point in time: Intermediate factor nB (who)
Write thoroughly
5W
Who
When
Where
What
Why
1H
How
From the results of investigating the cause
Actions by Personnel A at first time: Intermediate factor 1A (Who)
Actions by Personnel B at first time: Intermediate factor 1B (Who)
Cause
Background factors
Measures to prevent recurrence
With parallel cases
Root cause (Why)

Figure 2 Creation of standard event scenario network diagram.

3.3. Visualization of Event Scenario Network Diagram

We visualized the JSON format for network diagram creation using Graphviz, an open-source graph drawing tool suitable for networks, flowcharts, and hierarchies. By using Graphviz, each item (response, observation, direct cause, intermediate cause, root cause, preventive measures) can be represented as clusters, making it easy to visualize keywords (nodes) for each item and their connections (edges) in a format that allows for easy determination of whether they conform to the ontology. Therefore, we adopted this tool for the current project.

3.4. Various Analyses

To visualize the differences in "direct causes" (assuming PRA failure modes) and "root causes" (mainly assuming human error and organizational factors) of each event extracted by creating event scenario network diagrams, before and after the Fukushima Daiichi Nuclear Power Plant accident, as well as differences by power company and plant, the following analyses were attempted:

- Word Occurrence Frequency by Category (Company Name, Power Plant)
  - ➢ Purpose: To clarify characteristics specific to particular utilities or power plants and identify issues unique to organizations or facilities.
  - ➢ Feature: The frequency of occurrence of words belonging to each category.
  - ➢ Applications: Identification of problems and risks within companies or power plants. Proposal of customized safety measures for each organization or facility. Understanding differences in safety culture between organizations and facilities.
- Yearly Word Occurrence Frequency
  - ➢ Purpose: To track changing language patterns over time and analyze long-term trends.
  - ➢ Feature: Yearly word occurrence frequency.
  - ➢ Applications: Capturing the impact of changes in safety standards and technology over time. Evaluating the effectiveness of long-term safety measures. Identifying significant events or changes that occurred in specific years.
- Chi-Square Analysis

  - ➢ Purpose: To evaluate whether differences in word occurrence frequency between categories (power plant name, company name) are statistically significant, verifying whether specific words are disproportionately present in certain categories.
  - ➢ Feature: Word occurrence frequency in each category.
  - ➢ Applications: Identifying common risks or issues within specific categories through the discovery of significant words. Visualizing statistically significant relationships in a heat map to understand patterns and aid in developing preventive measures and safety strategies.
- PCA (Principal Component Analysis)
  - ➢ Purpose: To capture the main structure within the dataset and reveal correlations by linearly transforming high-dimensional data into a new lower-dimensional feature space.
  - ➢ Feature: Word occurrence frequency in each category.
  - ➢ Applications: Simplifying the analysis of trends and patterns in accident data by reducing the dimensionality of the data. Efficiently identifying issues in power plants and operations by extracting principal components from multi-dimensional data.
- Differences in Normalized Word Occurrence Frequency Before and After Fukushima
  - ➢ Purpose: To quantitatively capture changes before and after the Fukushima accident and analyze the impact of the accident.
  - ➢ Feature: Normalized word occurrence frequency before and after the Fukushima accident.
  - ➢ Applications: Quantitative evaluation of changes due to the impact of the accident. Verification of content shifts associated with changes in safety measures and regulations. Monitoring the effectiveness of policy revisions.

4. RESULTS EVALUATION

The following describes the evaluation conducted by our experts on the results produced by the AI tool prototyped in 2023.

4.1. Evaluation of Event Scenario Network Diagrams

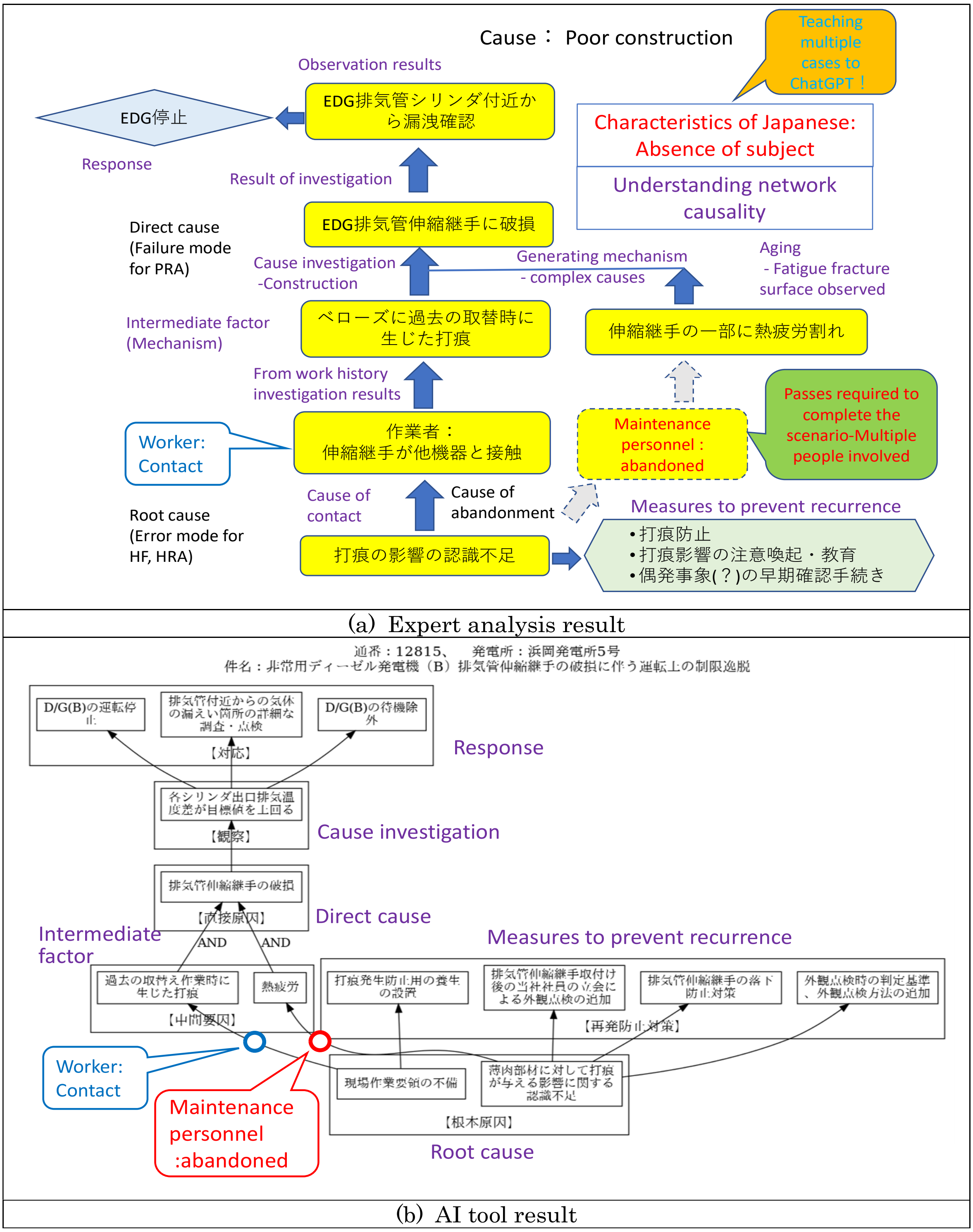
The evaluation focused on how well the event scenario network diagrams created by the AI tool could replicate the five event analysis results from NUCIA performed by our experts. For five characteristic events analyzed: (a) results from our expert analysis, and (b) results from the AI tool, one scenario was taken as an example shown in Figure 3 as standardization of event scenario analysis (assignment for 2024), giving a framework for analysis in advance for learning ChatGPT.

- The event scenarios produced by the AI tool generated more detailed networks compared to our experts' analysis. However, the causes are not classified based on understanding the meaning of the sentence. It is necessary to develop a learning process using the created event networks as training data for graph neural networks or similar techniques to classify and predict causes. This will be addressed as a challenge for the common cause analysis method in 2024.
- Learning from the network to understanding causal relationships requires the ability to infer subjects, which necessitates a process for learning through examples. This involves considering subject, verb, and object sets to understand causal relationships.

4.2. Evaluation of Various Analysis Results

Various analyses (such as identifying multiple causes and analyzing differences in events and time) indeed yielded results, but interpreting these insights remains challenging.

- Differences between power companies and plants were clarified. While Tokyo Electric Power Company, which has many plants, experiences many issues, the Kashiwazaki-Kariwa plant tends to have fewer troubles.
- Differences in trouble causes before and after the Fukushima accident were clarified (e.g., increase in construction defects, decrease in maintenance deficiencies). It was also observed that the number of trouble cases decreased after the Fukushima accident.


Cause： Poor construction
Teaching multiple cases to ChatGPT！
Observation results
EDG停止
EDG排気管シリンダ付近から漏洩確認
Response
Result of investigation
Characteristics of Japanese: Absence of subject
Understanding network causality
Direct cause (Failure mode for PRA)
EDG排気管伸縮継手に破損
Cause investigation -Construction
Generating mechanism - complex causes
Aging - Fatigue fracture surface observed
Intermediate factor (Mechanism)
ベローズに過去の取替時に生じた打痕
伸縮継手の一部に熱疲労割れ
From work history investigation results
Worker: Contact
作業者：伸縮継手が他機器と接触
Maintenance personnel : abandoned
Passes required to complete the scenario-Multiple people involved
Cause of contact
Cause of abandonment
Measures to prevent recurrence
Root cause (Error mode for HF, HRA)
打痕の影響の認識不足
•打痕防止
•打痕影響の注意喚起・教育
•偶発事象(？)の早期確認手続き
通番：12815、　発電所：浜岡発電所5号
件名：非常用ディーゼル発電機（B）排気管伸縮継手の破損に伴う運転上の制限逸脱
D/G(B)の運転停止
排気管付近からの気体の漏えい箇所の詳細な調査・点検
D/G(B)の待機除外
【対応】
Response
各シリンダ出口排気温度差が目標値を上回る
【観察】
Cause investigation
排気管伸縮継手の破損
【直接原因】
Direct cause
Intermediate factor
AND
AND
Measures to prevent recurrence
過去の取替え作業時に生じた打痕
熱疲労
【中間要因】
打痕発生防止用の養生の設置
排気管伸縮継手取付け後の当社社員の立会による外観点検の追加
排気管伸縮継手の落下防止対策
外観点検時の判定基準、外観点検方法の追加
【再発防止対策】
Worker: Contact
Maintenance personnel :abandoned
現場作業要領の不備
薄肉部材に対して打痕が与える影響に関する認識不足
【根本原因】
Root cause

(a) Expert analysis result

(b) AI tool result

Figure 3 Event scenario: Trouble NUCIA-12815, EDG-B exhaust pipe expansion joint damage.

5. CONSTRUCTION OF ‘CORDS’ ANALYSIS METHOD

The following challenges for the CORDS database are included in the study.

- A rule-based classification method based on past cases is appropriate.
- The determination can be addressed through text classification processing. The text was converted into vectors considering the context, and the distance between vectors were calculated to determine text similarity. The process came to a successful conclusion.

For calculating failure rates, we organized the existing event data recorded in the CORDS fast breeder reactor equipment reliability database, which includes 443 events (185 from Joyo and 258 from Monju). Compared to NUCIA, the number of analyzed cases, the scope of analysis, and the analysts are limited, leading to a uniform granularity of each case as shown in Figure 4 [3]. Therefore, we judged that a rule-based approach is more appropriate than using generative AI for this context and decided to adopt a different preprocessing method for failure analysis. CORDS, in addition to raw field data, includes expert analysis data as training data, enabling supervised learning.

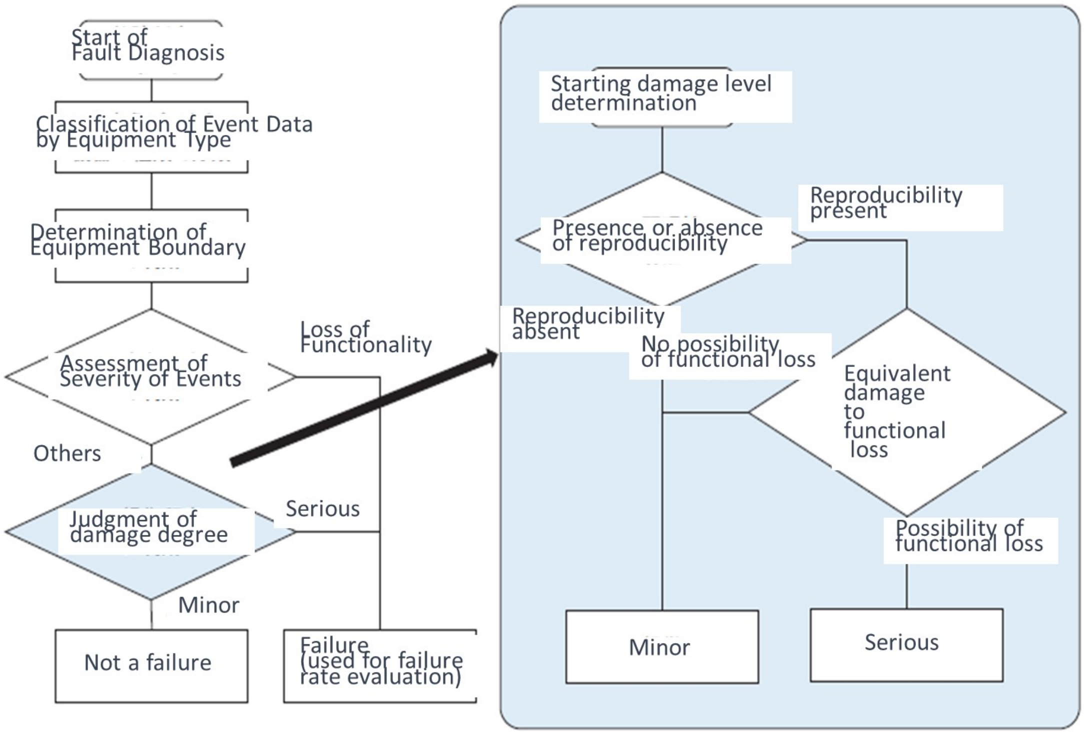

Start of Fault Diagnosis
Classification of Event Data by Equipment Type
Determination of Equipment Boundary
Assessment of Severity of Events
Loss of Functionality
Others
Judgment of damage degree
Serious
Minor
Not a failure
Failure (used for failure rate evaluation)
Starting damage level determination
Presence or absence of reproducibility
Reproducibility present
Reproducibility absent
No possibility of functional loss
Equivalent damage to functional loss
Possibility of functional loss
Minor
Serious

Figure 4 The expert-driven fault diagnosis flow.

Using failure determination data (FBR reliability) from JAEA's experts as training data, we considered the following method for linking the names of failed equipment and failure modes (failure causes) by text mining from event data (repair reports):

- Identification of equipment scope based on the equipment boundary definition table (binary classification).
- Determination of failure presence using rule-based failure judgment criteria (two-step failure determination):
- Determining functional loss based on the severity of the event (binary classification).
- Determining the magnitude based on the extent of damage (urgency, reproducibility, severity) (binary classification).
- Determination of equipment failure modes based on the equipment-failure mode table (multi-class classification).

Upon inputting event data, the system will output the following judgment results for the event:

- Whether it is within the equipment boundary.
- If within the equipment boundary, whether the event results in functional loss.
- If not resulting in functional loss, whether the extent of damage is significant.
- If judged as functional loss or significant damage, which failure mode it corresponds to.

The overall image of the processing flow for the tool prototyped for automatic failure determination in CORDS is illustrated as follows; First, the text is converted into vectors considering the context. Then, by calculating the distance between vectors or cosine similarity, the similarity of the text can be quantified. The tool also includes a GUI interface.

6. CONCLUSION

For the construction of reliability database in Japan, the Central Research Institute of Electric Power Industry and the Japan Nuclear Safety Institute have been developing the NUCIA database for light water reactors, and the Japan Atomic Energy Agency has been developing the CORDS database for sodium-cooled fast reactors. These databases allow us referencing individual cases and support analysis through database searches and have also been used for creating failure rate databases.

Recently generative artificial intelligence (AI) technologies improvement is remarkable, it is believed that the following solutions can be resolved by using generative AI as chat GPT, with the greatest expectation being the possibility of discovering new insights that may not be discernible even by the individual expert.

The following challenges for the NUCIA database were identified and completed successfully.

- The cases are scenarios where various causal relationships are intertwined, so we have created a framework with predefined multiple paths to describe multiple players and causes and steps to describe various causal relationship. On top of this framework, we have structured the events as a network, allowing for their description.
- As a characteristic of Japanese language, there are cases where the subject is not explicitly stated, so it is necessary to infer and add the subjects or players. As a response in cases of uncertainty, a feature is set to request necessary information at the origin.
- Since there are cases where the analyses are insufficient, a mechanism is established to always describe the 5W1H, and if there is no clear description, it is completed by inference. In cases where it is not possible, a feature is set to request necessary information at the origin.
- The above three challenges are problems in the analysis of a single event, but there are also challenges that span multiple cases. For unresolved challenges that occur in multiple plants, a function is needed to collect similar cases and identify common causes through time-series analysis.

The following challenges for the CORDS database were identified and completed successfully.

- A rule-based classification method based on past cases is appropriate.
- The determination can be addressed through text classification processing. The text was converted into vectors considering the context, and the distance between vectors were calculated to determine text similarity. The process came to a successful conclusion.

Acknowledgements

This work and ‘Automatic Fault Tree Creation’ study [4] were supported by MEXT Innovative Nuclear Research and Development Program Grant Number JPMXD0222682583.

References

[1] http://www.nucia.jp/

[2] K. KURISAKA, "Development of Fast Reactor Equipment Reliability Database," Japan Atomic Energy Research Institute, JNC Technical Report No. 98, pp. 18-31, June 1996.

[3] H. UJITA, T. MORIMOTO, S. FUTAGAMI, H. YAMANO and K. KURISAKA, Development of Probabilistic Risk Assessment Methodology Using Artificial Intelligence Technology, 2. Automatic Fault Detection Method for Building Reliability Database, in Probabilistic Safety Assessment and Management (PSAM) Topical, 2023, Virtual.

[4] S. FUTAGAMI, H. YAMANO and K. KURISAKA, H. UJITA, Development of Probabilistic Risk Assessment Methodology Using Artificial Intelligence Technology, 3. Automatic Fault Tree Creation, in PSAM17&ASRAM2024, October 7-11, 2024, Sendai.