



# ICONE31

31<sup>st</sup> International Conference on  
Nuclear Engineering

---

CONFERENCE AUG. 4–8, 2024  
HILTON PRAGUE  
PRAGUE, CZECH REPUBLIC



# A LLM-Based Approach For Automatic Generation Of Fault Trees In Nuclear Power Plants

Xingyu Xiao

xxxy23@mails.Tsinghua.edu.cn

15630451413

Institute of Nuclear and New Energy Technology, Tsinghua University, Beijing 100084, China



# Contents



**INTRODUCTION**



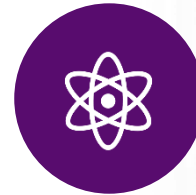
**RESULTS AND DISCUSSION**



**CONCLUSION**



**MATERIALS AND METHODS**



**APPLICATION**



**REFERENCES**

## Fault Tree Analysis (FTA)

It is systematic and structured method used to identify and analyze potential failure modes and their causes within complex systems. It involves constructing a fault tree model that graphically represents various failure events and their interrelationships, aiding in the identification and assessment of system reliability and safety. Below are the key steps and fundamental concepts of FTA:

### Key Steps

#### 1. Define the Top Event:

- The first step is to clearly define the analysis objective, which is the undesired system failure or event to be avoided or explained. This is known as the top event.

#### 2. Identify Basic Events:

- Determine all potential basic events that could lead to the occurrence of the top event. These basic events are the most fundamental failure units within the system and typically cannot be further decomposed.

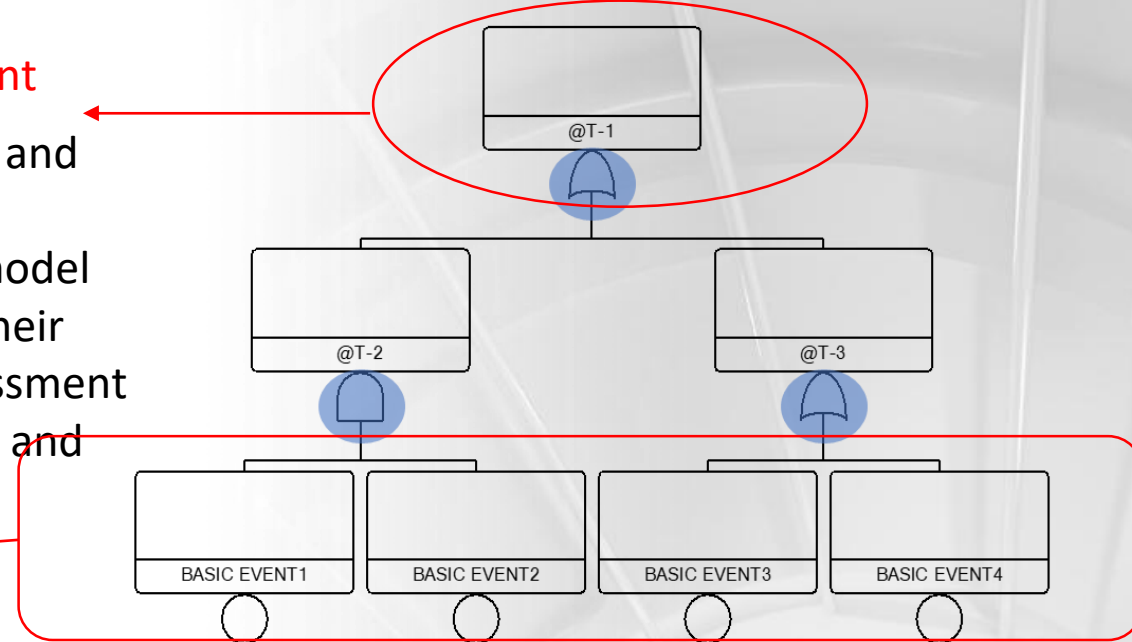
#### 3. Construct the Fault Tree Structure:

- Connect the basic events to the top event using logical gates (such as AND gates and OR gates) to represent the relationships between events. An **AND** gate indicates that the output event occurs only if all input events occur simultaneously, while an **OR** gate indicates that the output event occurs if any one of the input events occurs.

#### 4. Qualitative Analysis and Quantitative Analysis

Top Event

Basic event







## » Fault Tree Analysis

- ✓ a complicated and time-consuming task: the necessity of interdisciplinary and intricate knowledge

computer aided methods	flowchart-based methods: require a lot of manual input and time <b>more than 90%</b>
	based on other data types: mainly relying on a data-driven approach to achieve the cause of failure, failure mode mining (time-series data , text ) <b>less and immature: lack logical relationships</b>

## » Artificial Intelligence

- ✓ large language models: These models have demonstrated a remarkable capability to comprehend and generate intricate textual data, opening doors to myriads of applications in both academia and industry.



GPT-doctor

Wang et al., 2023

Their framework offers step-by-step principles and guidance for real-world problems.



DISC-LawLLM

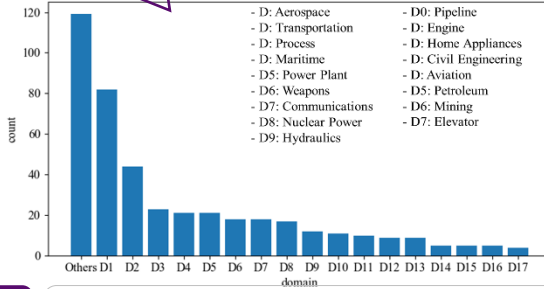
Yue et al., 2023

They adopt legal syllogism prompting strategies to construct supervised fine-tuning datasets in the Chinese Judicial domain.



## Dataset Description

### Supervised Fine-Tuning Data



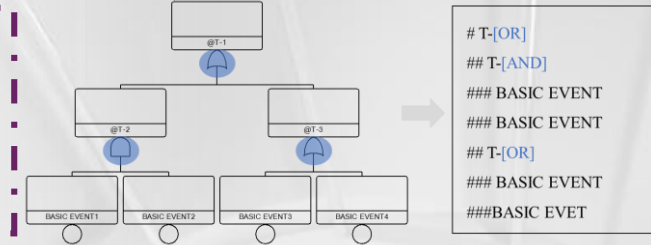
### User Message

description of the system or device  
top event  
failure modes of the system or device

### System Message

Few-shots  
Thinking Development  
Role Playing

### Textual Structure



## Methodology For Automated Fault Tree Generation



### 01. Base Model

Since the ideal performance exhibiting in both English and Chinese, we choose baichuan 2-13B-Chat as our base model for automated fault tree generation.



### 02. Full-parameter fine-tuning

Full-parameter fine-tuning is by far the most effective method of parameter fine-tuning



### 03. Parameter Settings

The global batch size of 64, the learning rate of 1e-5, 50 epochs training stage, maximum source length of 2048 tokens, and maximum target length of 1024 tokens. The training process is carried out on 8\*A100 GPUs and the training cost is further reduced with the help of deepspeed.

## » Evaluation Metrics

- ✓ we unravel the “black box” and examine the sources of model performance improvement from the perspectives of horizontal conversation pattern alignment and vertical fault tree knowledge evolution.

**Conversation Pattern Alignment.**

**Cosine Similarity Calculation.**

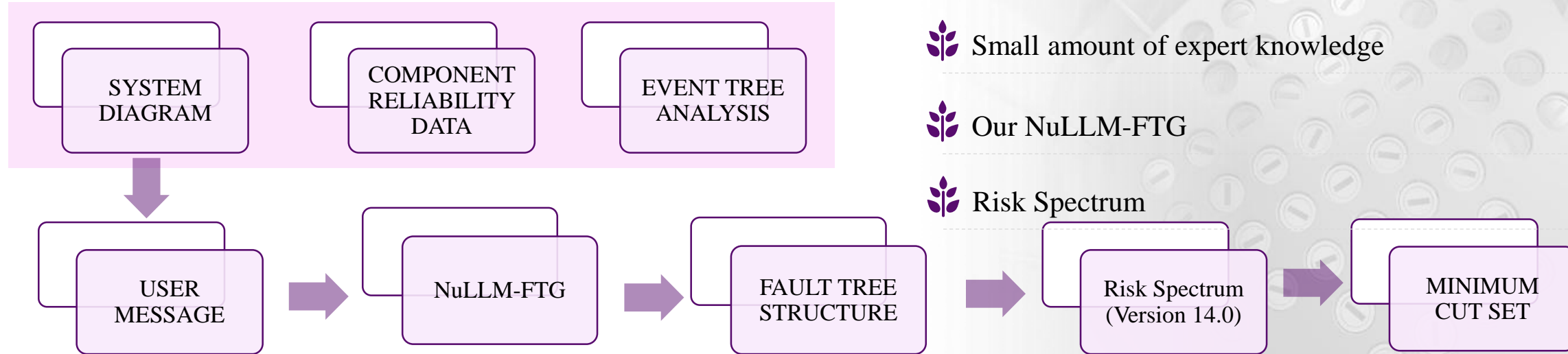
$$\text{similarity} = \frac{Z_1 Z_2}{\|Z_1\| \|Z_2\|}$$

**Qualitative Evaluation.**

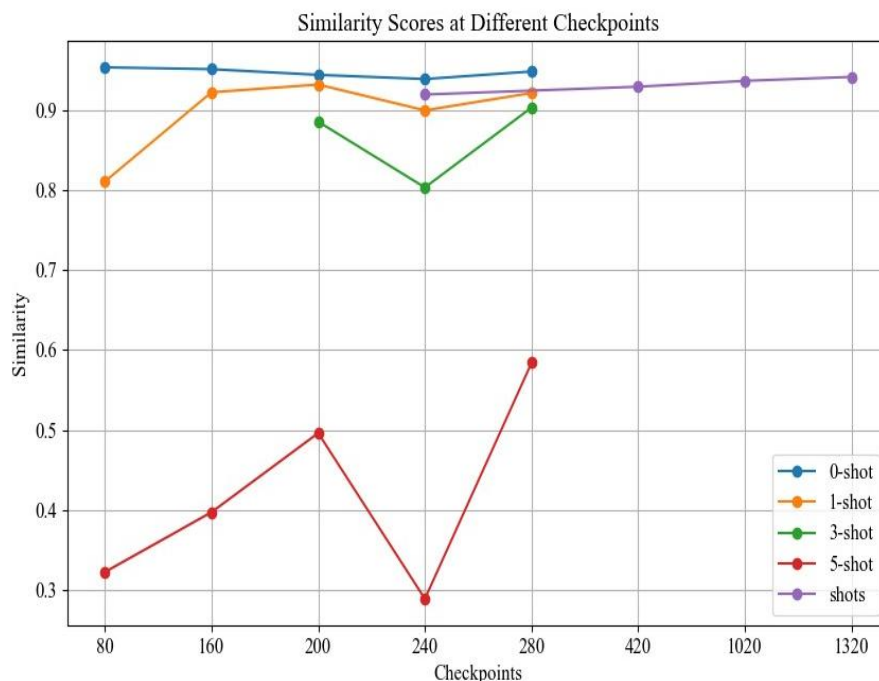
professionalism, completeness, and analysis, we employ two distinct evaluative

## » Application

- ✓ Case Study: To demonstrate the applicability of our NuLLM-FTG, we present a simple case study as an illustration.



# 1. Experimental Results



It can be observed that the 5-shots scenario shows a marked dip and recovery in performance, suggesting a momentary learning difficulty or anomaly at the 240-checkpoint.

The 0-shot scenario remains consistently high, indicating robust performance even without prior data exposure.

## Comparative Analysis Of Assessment Metrics Across Different Language Models.

Model	Strategies	Cosine Similarity	Conversations on Pattern
Baseline Model	0-shot	0.4691	N
	1-shot	0.2433	N
	3-shot	0.4229	N
	5-shot	0.5705	N
GPT-3.5	0-shot	0.6920	N
	1-shot	0.6124	N
	3-shot	0.6219	N
	5-shot	0.6255	N
GPT-4	0-shot	0.7920	N
	1-shot	0.8091	N
	3-shot	0.8373	Y
	5-shot	0.8368	Y
NuLLM-FTG	0-shot	0.9390	Y
	1-shot	0.9321	Y
	3-shot	0.9409	Y
	5-shot	0.9359	Y

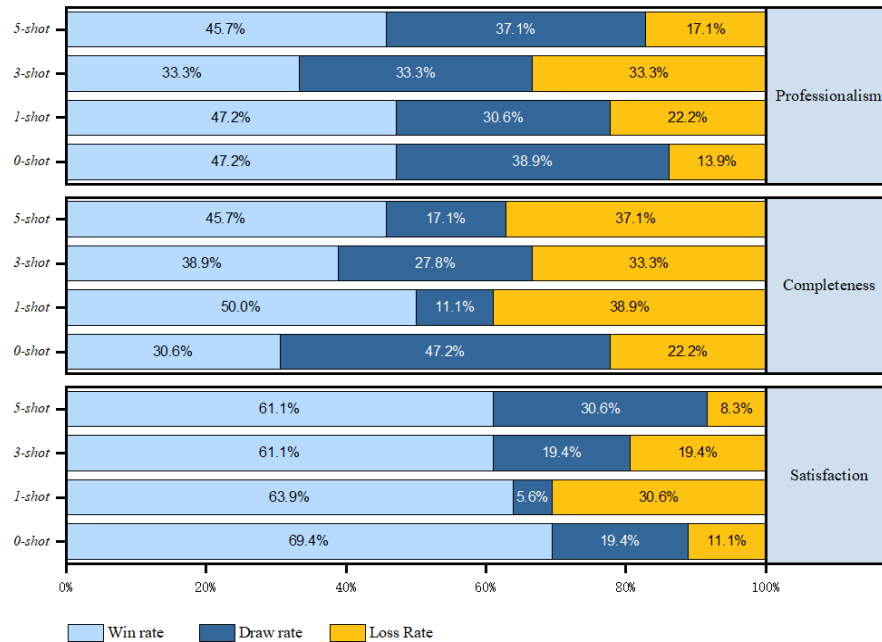
→ Our NuLLM-FTG model demonstrates the highest cosine similarity scores across all strategies.



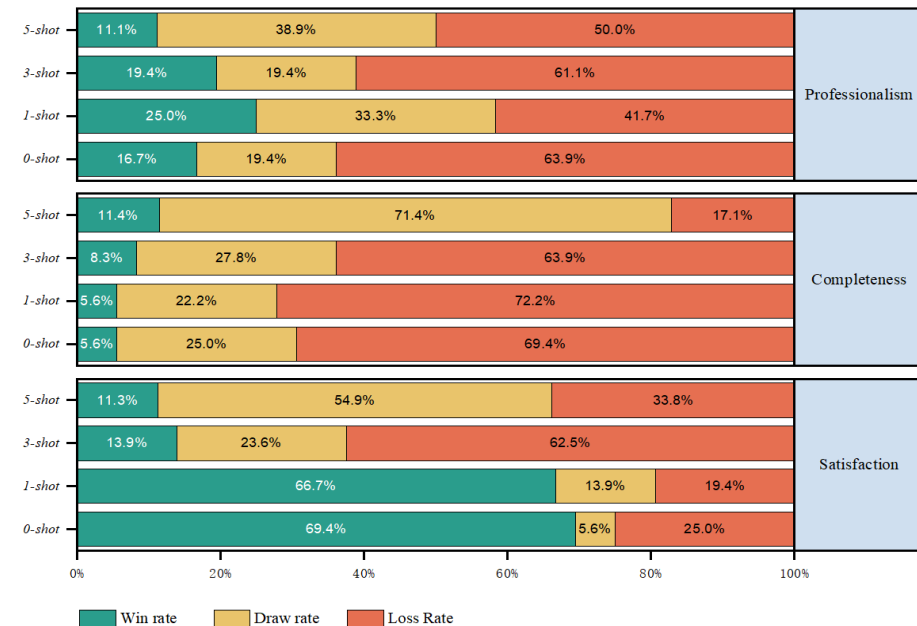
## 2. Qualitative Evaluation



single-blind evaluation against GPT-4



double-blind evaluation against GPT-4

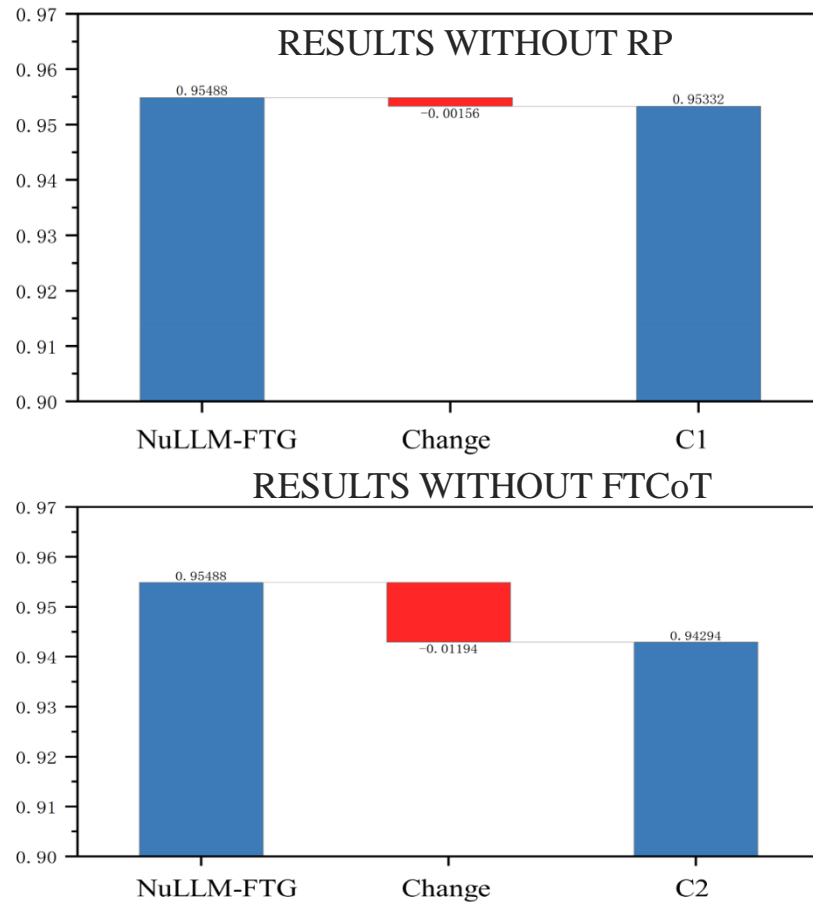


- There is a **threshold** after which additional context does not yield better quality and even is detrimental.
- While **GPT-4 generates** responses that are more **complete**, our strengths lie in the close resemblance to the provided **standard** answers. Consequently, when experts are presented with a standard answer for context in the single-blind setup, **our responses exhibit a higher win rate**.



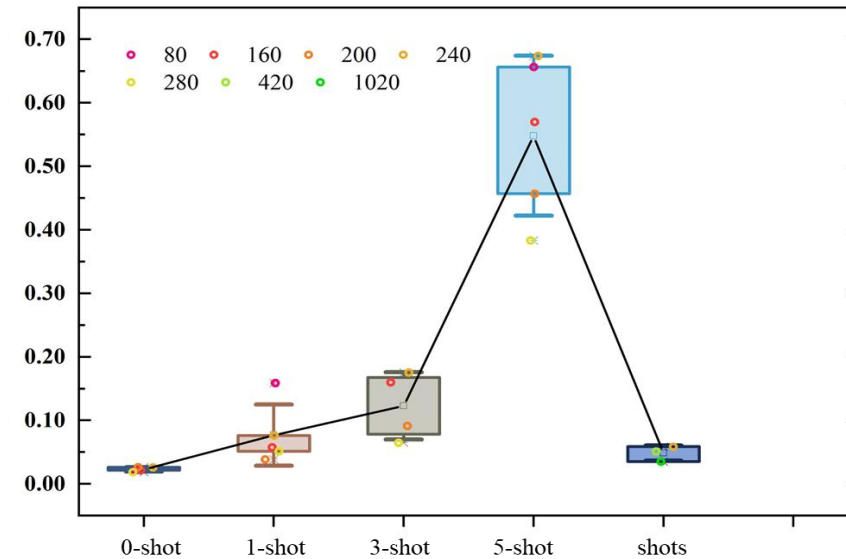
### 3. Ablation Experiment and Language Corpus Impact Assessment

- ✓ The omission of **FTCoT** resulted in a more pronounced decrease in the textual similarity metric (down to 0.94294) compared to the **exclusion of RP** (where the metric decreased to 0.95332).



### Language Corpus Impact Assessment

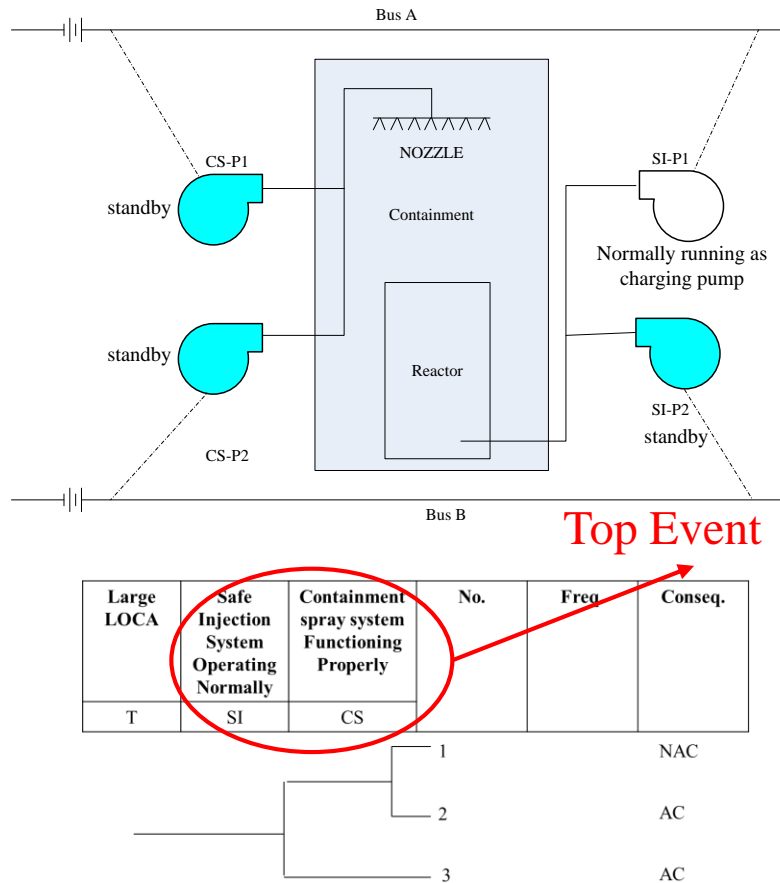
difference



The results with English Corpus shows better.

# APPLICATION

✓ System structure-> Event Tree -> Top Event -> User Input->NuLLM-FTG -> Output -> Risk Spectrum ->Minimum



## NuLLM-FTG

### Input

There are two pumps, labeled "SI-P1" and "SI-P2". SI-P1 operates normally under no special conditions, while SI-P2 serves as a standby pump, with both functioning as safety injection pumps under safety injection conditions. If Bus A fails, then pump SI-P1 fails; if Bus B fails, then pump SI-P2 fails. The operation of either pump is sufficient. The potential failures of the Safety Injection System (SI system), consisting of the SI-P1 and SI-P2 pumps, are: (1) failure to start and (2) failure during operation. Additionally, the failure of the safety injection system only occurs if both pumps fail. Please exclude common-cause failures of the pumps and failures of Bus A/B in your consideration. The fault tree should be constructed with "Safety Injection System Failure" as the top event.

### Output

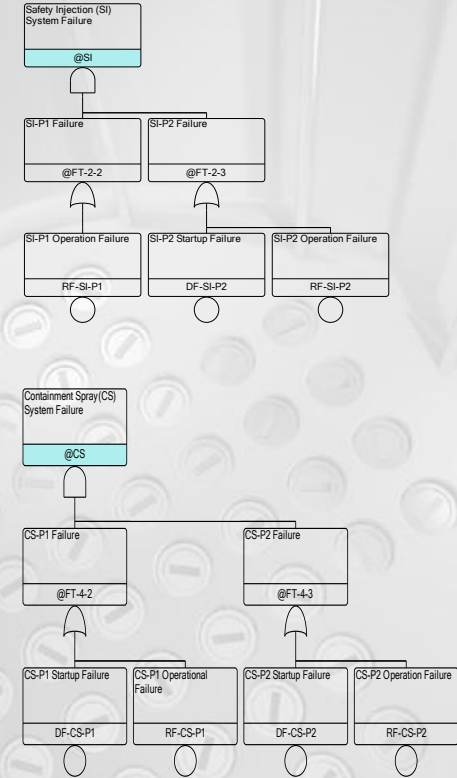
Hello! Thank you for your inquiry. The top event in the system is 'Safety Injection System Failure'. The constructed fault tree structure is as follows:  
 # Safety Injection System Failure [AND]  
 ## SI-P1 Failure [OR]  
 ### SI-P1 Operation Failure  
 ## SI-P2 Failure [OR]  
 ### SI-P2 Startup Failure  
 ### SI-P2 Operation Failure

### Input

There are two pumps, identified as "CS-P1" and "CS-P2". Both CS-P1 and CS-P2 are in standby mode prior to a Loss of Coolant Accident (LOCA) and are required to start upon the occurrence of such an incident. Should Bus A fail, pump CS-P1 becomes non-functional, and similarly, a failure in Bus B results in the malfunction of pump CS-P2. The system is operational as long as one pump remains functional. Potential failures for the CS-P1 and CS-P2 pumps include start-up failure and operational failure. Furthermore, a failure in the container spray system occurs only when both pumps are non-functional. Please exclude the consideration of common-cause failures of the pumps and failures of Bus A/B in this analysis. The fault tree is requested to have the "container spray system failure" as its top event. Please construct the fault tree accordingly.

### Output

Hello! Thank you for your inquiry. The top event of the system in question is the "Container Spray System," and the structure of the fault tree constructed is as follows:  
 # Container Spray System [AND]  
 ## CS-P1 Failure [OR]  
 ### CS-P1 Start-up Failure  
 ### CS-P1 Operational Failure  
 ## CS-P2 Failure [OR]  
 ### CS-P2 Startup Failure  
 ### CS-P2 Operational Failure



Output

✓ Our model is highly capable of constructing fault trees.

## APPLICATION

- ✓ System structure-> Event Tree -> Top Event -> User Input->NuLLM-FTG -> Output -> Risk Spectrum -> Minimum cut set

MINIMUM CUT SET

No	Probability	%	Event1	Event2	Event3
1	1.44E-12	28.70	IE-LOCA	RF-CS-P1	RF-CS-P2
2	1.20E-12	23.92	IE-LOCA	DF-CS-P1	RF-CS-P2
3	1.20E-12	23.92	IE-LOCA	DF-CS-P2	RF-CS-P1
4	1.00E-12	19.93	IE-LOCA	DF-CS-P1	DF-CS-P2
5	1.20E-13	02.39	IE-LOCA	DF-SI-P2	RF-SI-P1
6	5.76E-14	01.15	IE-LOCA	RF-SI-P1	RF-SI-P2







## » Conlucison

- ✓ we unravel the “black box” and examine the sources of model performance improvement.
- ✓ we introduce the NuLLM-FTG and evaluation metrics for fault tree generation.
- ✓ we introduce a textual data structure for fault tree.
- ✓ English corpus, role playing, chain of thought contribute to enhancing model performance.
- ✓ effectiveness and robustness: we conduct a practical case study in conjunction with Risk Spectrum.



# Reference

- [1] Lazaroova-Molnar, S., Niloofar, P., & Barta, G. K, Automating Reliability Analysis: Data-driven Learning and Analysis of Multistate Fault Trees, Proceedings of the 30th European Safety and Reliability Conference and 15th Probabilistic Safety Assessment and Management Conference (2020) 1805–1812.
- [2] Loshchilov, I., & Hutter, F., Decoupled weight decay regularization, arXiv preprint (2017) 1711.05101.
- [3] Lv, K., Yang, Y., Liu, T., Gao, Q., Guo, Q., & Qiu, X., Full Parameter Fine-tuning for Large Language Models with Limited Resources, arXiv preprint (2023) 2306.09782.
- [4] Majdara, A., & Wakabayashi, T., Component-based modeling of systems for automated fault tree generation, Reliability Engineering & System Safety 94(6) (2009) 1076–1086.
- [5] OpenAI, R., Gpt-4 technical report., View in Article, (2023) 2303.08774.
- [6] Rasley, J., Rajbhandari, S., Ruise, O., & He, Y., Deepspeed: System optimizations enable training deep learning models with over 100 billion parameters, Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (2020) 3505–3506.
- [7] Shao, Y., Li, L., Dai, J., & Qiu, X., Character-LLM: A Trainable Agent for Role-Playing, arXiv preprint (2023) 2310.10158.
- [8] Thangaratinam, S., & Redman, C. W., The delphi technique, The Obstetrician & Gynaecologist 7(2) (2005) 120–125.
- [9] Wang, W., Zhao, Z., & Sun, T., GPT-doctor: Customizing Large Language Models for Medical Consultation, arXiv preprint (2023).
- [10] Watson, H., Bell telephone labs, Launch Control Safety Study, Bell Telephone Laboratories (1961).
- [11] Yang, A., Xiao, B., Wang, B., Zhang, B., Bian, C., Yin, C., Lv, C., Pan, D., Wang, D., Yan, D., Yang, F., Deng, F., Wang, F., Liu, F., Ai, G., Dong, G., Zhao, H., Xu, H., Sun, H., Wu, Z., Baichuan 2: Open Large-scale Language Models, arXiv preprint (2023) 2309.10305
- [12] Yue, S., Chen, W., Wang, S., Li, B., Shen, C., Liu, S., Zhou, Y., Xiao, Y., Yun, S., Huang, X., & Wei, Z., DISC-LawLLM: Fine-tuning Large Language Models for Intelligent Legal Services, arXiv preprint (2023) 2309.11325.



# Thanks for watching!

Xingyu Xiao:xxy23@mails.Tsinghua.edu.cn  
15630451413