FaultExplainer: Leveraging Large Language Models for Interpretable Fault Detection and Diagnosis

Can Li
Assistant Professor
Davidson School of Chemical Engineering

Recent Advances

- Huge success in the past decade.

- Mainly due to the advance in hardware and the abundance of data

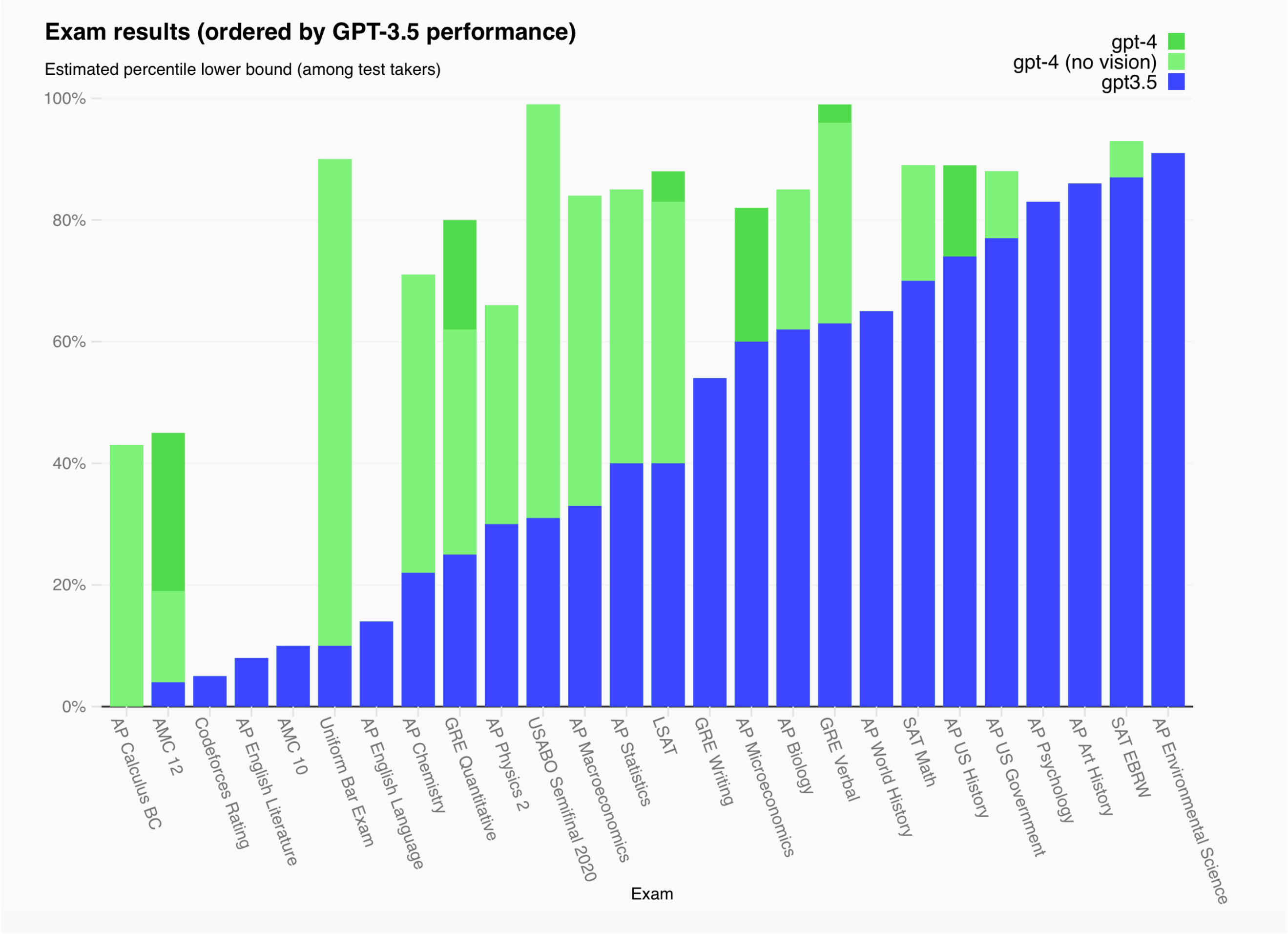
Cray 2 supercomputer (1985)
1.9GFlops

40.8GFlops

Credit: Prof. Venkat Venkatasubramanian

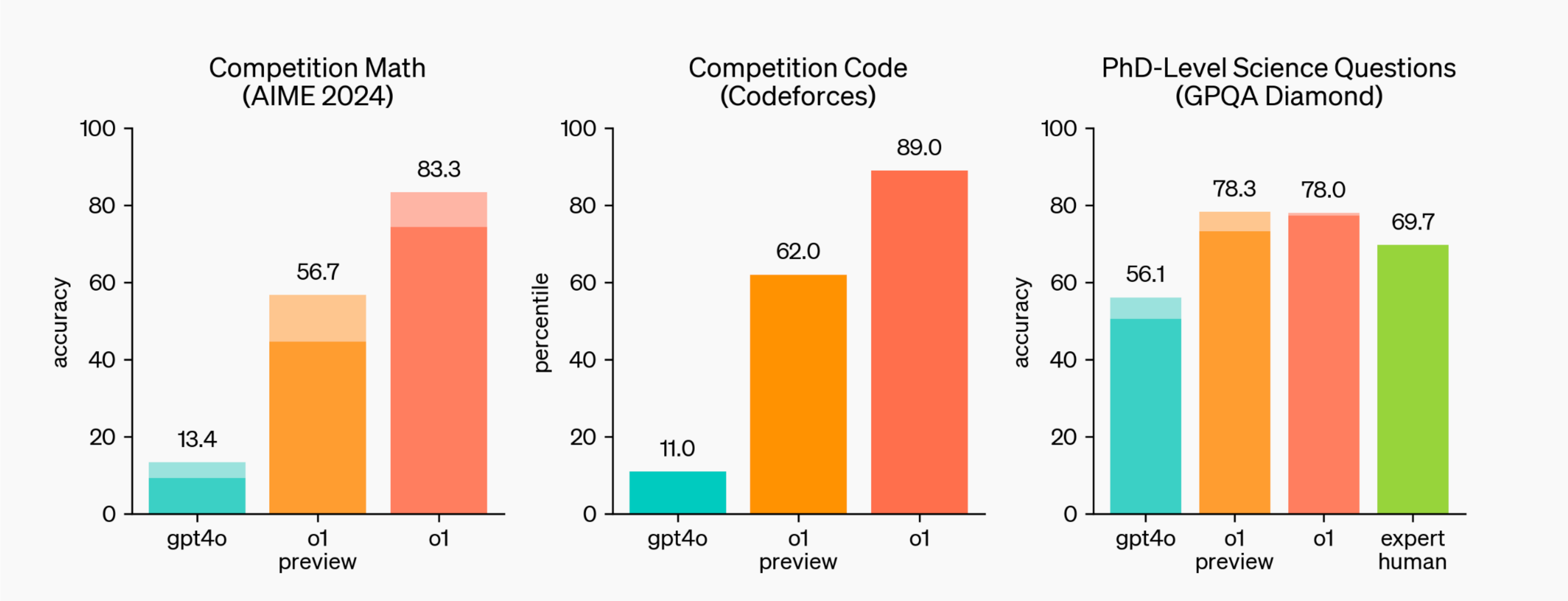
What can LLMs do

- GPT-4, released on March 2023. Excel at qualitative subjects: History, biology, bar exam, and verbal. Struggles with math, coding, and reasoning

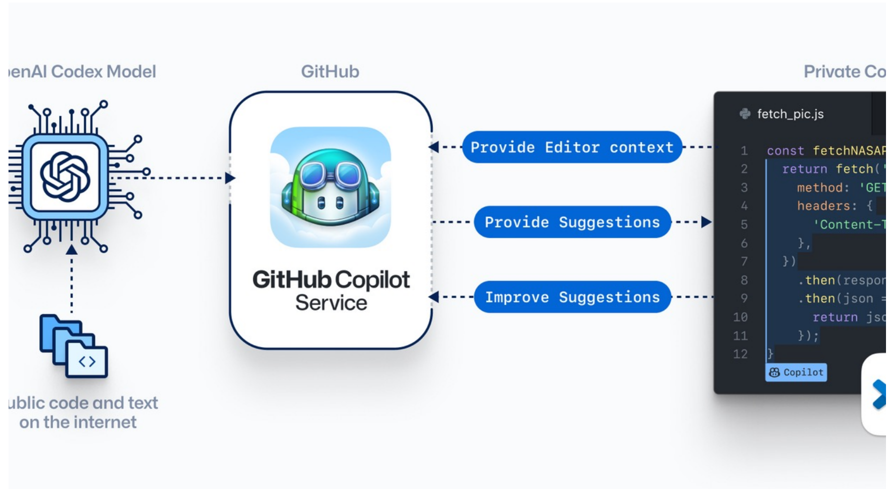


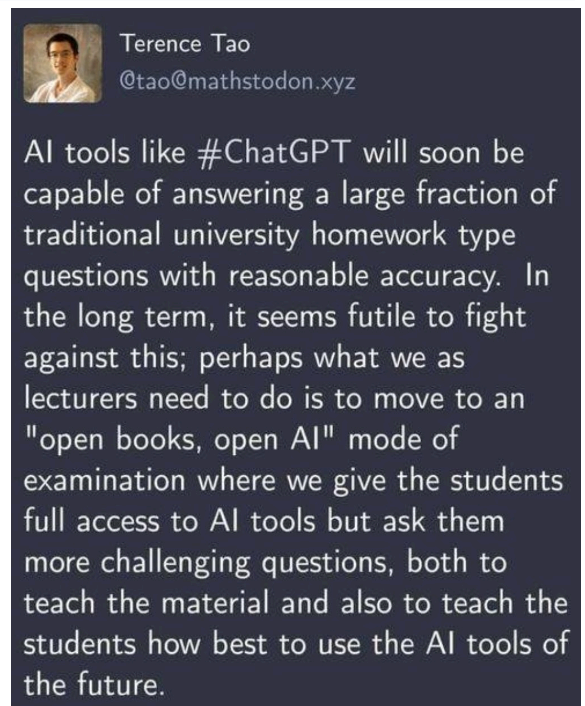
What's new lately?

- OpenAI o1 model: released on September 2024. Significant improvement in math, science, and reasoning.
- Similar to how a human may think for a long time before responding to a difficult question, o1 uses a **chain of thought** when attempting to solve a problem



What can LLMs do




Terence Tao
@tao@mathstodon.xyz

AI tools like #ChatGPT will soon be capable of answering a large fraction of traditional university homework type questions with reasonable accuracy. In the long term, it seems futile to fight against this; perhaps what we as lecturers need to do is to move to an "open books, open AI" mode of examination where we give the students full access to AI tools but ask them more challenging questions, both to teach the material and also to teach the students how best to use the AI tools of the future.

Perhaps, you have used ChatGPT for

- Correcting grammar
- Answering the homework problems
- Drafting emails and letters
- Writing code (GitHub copilot)

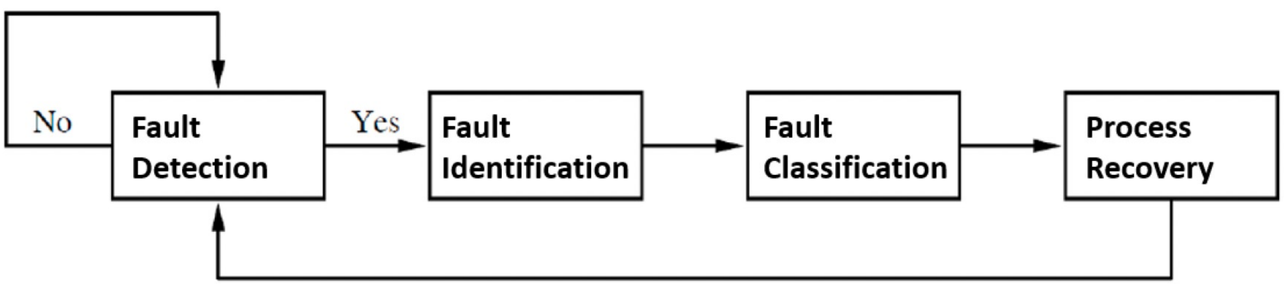
Applications of LLM in Research

- Coscientist (Boiko et al., 2023): an AI copilot for chemical research, autonomously designing, planning, and conducting complex experiments.
- LLaVA-Med (Li et al., 2024), BiomedGPT (Zhang et al., 2024), Med-Gemini (Lin et al., 2024): AI Assistants that help doctors in biomedical imaging by analyzing medical images and generating diagnostic reports.
- PILOT (Schweidtmann 2024): LLM for process engineering.
- scChat (Lu et al. 2024) AI copilot for interpreting RNA sequencing data and providing suggestions for experimental design.
- OptiChat (Chen et al. 2024) A chatbot for explaining optimization models to nonexperts.
- FaultExplainer (this talk): an interactive user interface for monitoring and interpreting the fault occurring in TEP.

Challenge: LLM tends to hallucinate.

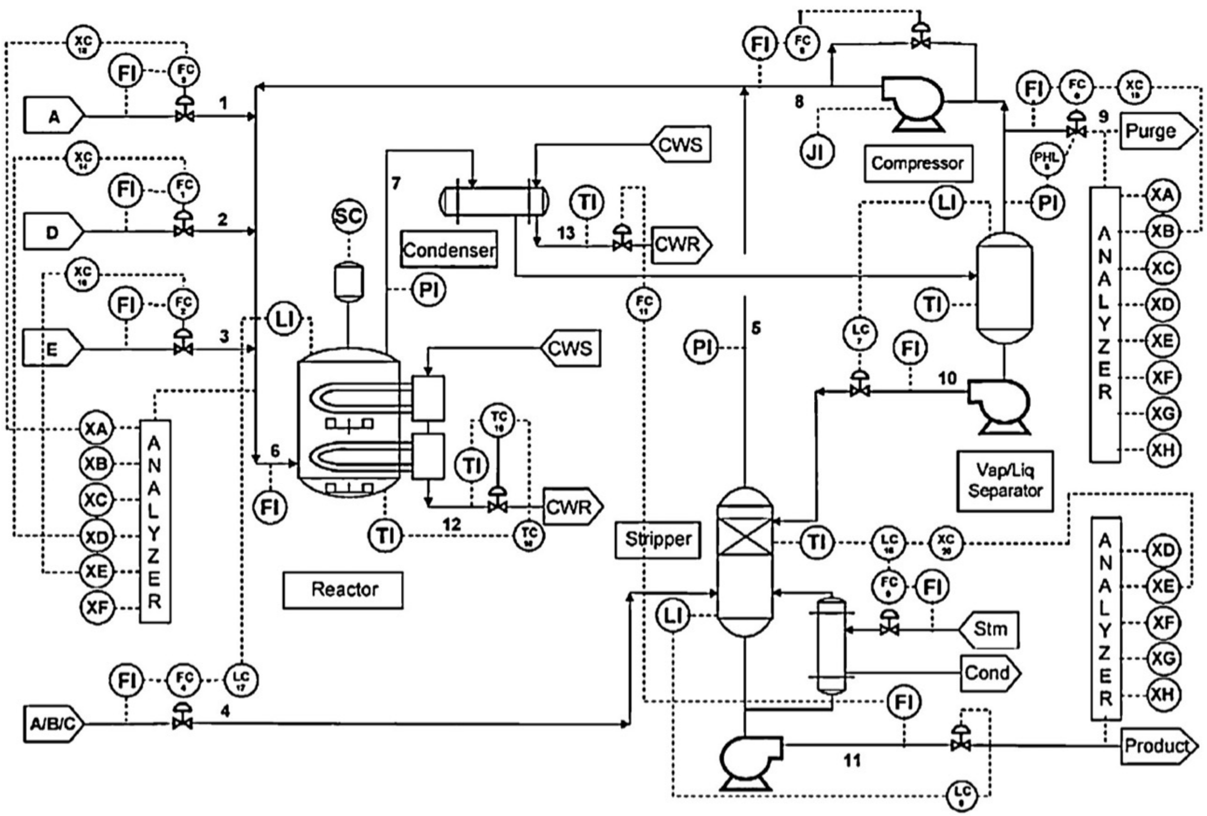
Combine LLM with Machine Learning Models for Process Monitoring

- Fault **Detection**: Detect if a fault has occurred
- Fault **Identification**: Identify the variables most relevant to the fault
- Fault **Diagnosis** (or Classification): Diagnose the root cause of the fault



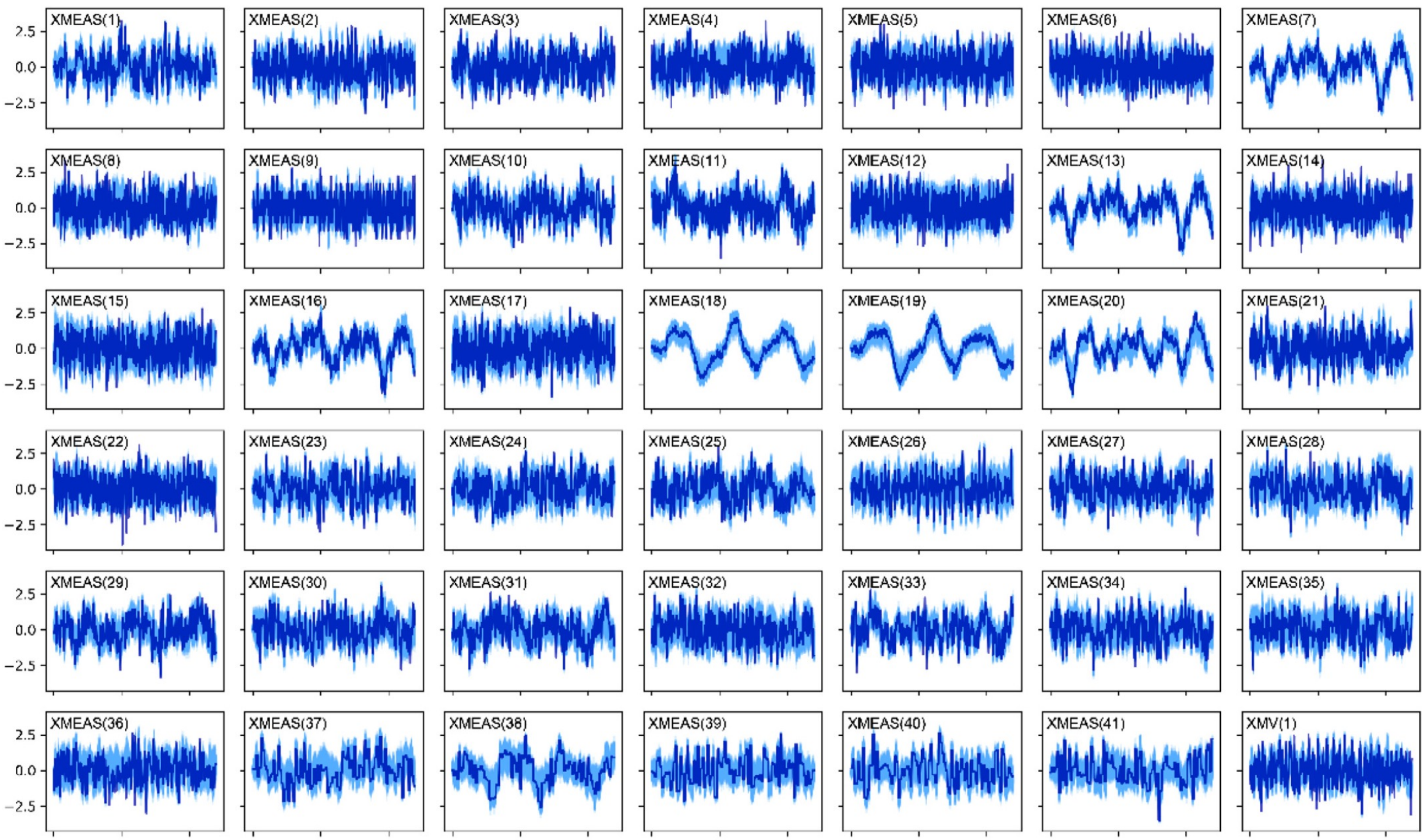
Tennessee Eastman Process

- TEP is an open-source simulator written in Fortran that resembles a real chemical process by Eastman
- Time series data can be collected from over 40 sensors that measure the state variables.
- Task: From measured state variables, perform fault detection using ML/AI



Examples of State Variables with Sensor Data

- Examples include feed flow rates, temperatures, pressures



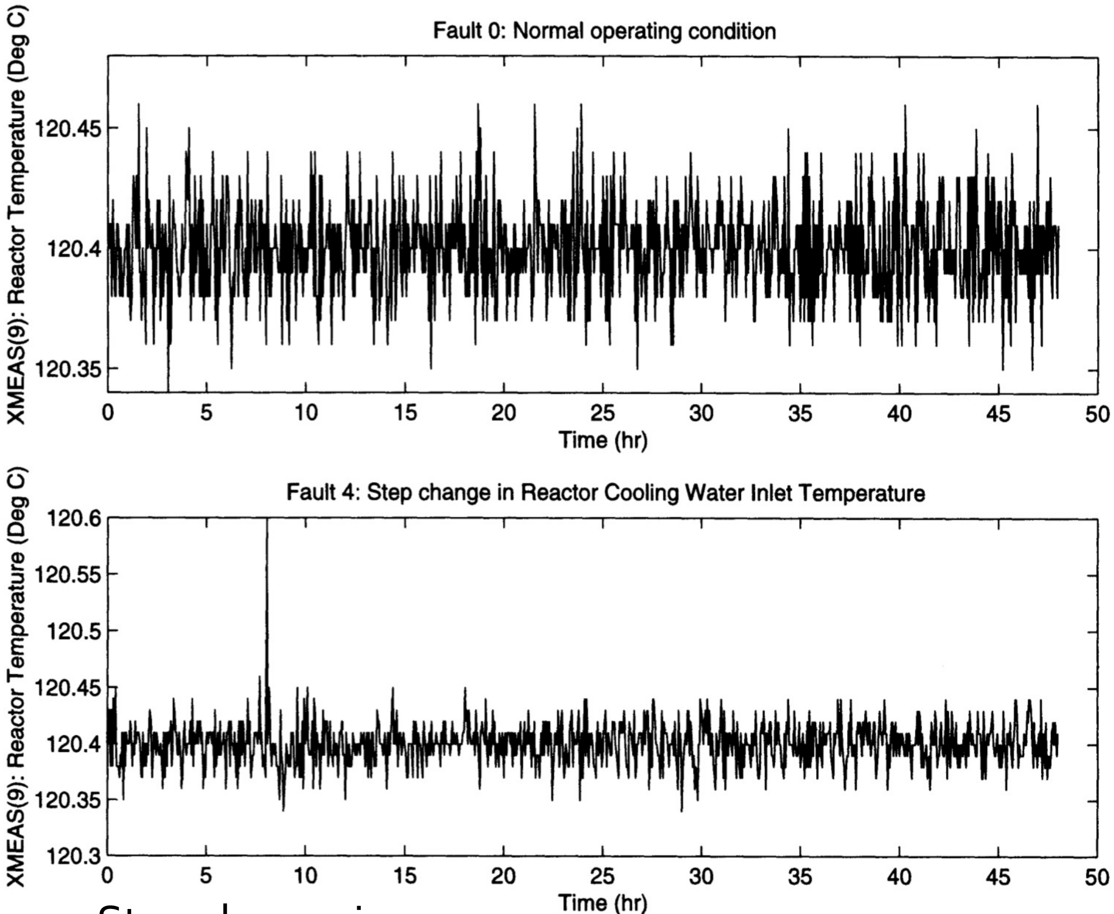
List of Potential Faults

- The following "faults" are created synthetically by the simulator
- These faults will cause the measured state variables to change from their normal operating conditions which further cause safety hazards.

Variable	Description	Type
IDV(1)	A/C Feed Ratio, B Composition Constant (Stream 4)	Step
IDV(2)	B Composition, A/C Ratio Constant (Stream 4)	Step
IDV(3)	D Feed Temperature (Stream 2)	Step
IDV(4)	Reactor Cooling Water Inlet Temperature	Step
IDV(5)	Condenser Cooling Water Inlet Temperature	Step
IDV(6)	A Feed Loss (Stream 1)	Step
IDV(7)	C Header Pressure Loss - Reduced Availability (Stream 4)	Step
IDV(8)	A, B, C Feed Composition (Stream 4)	Random Variation
IDV(9)	D Feed Temperature (Stream 2)	Random Variation
IDV(10)	C Feed Temperature (Stream 4)	Random Variation
IDV(11)	Reactor Cooling Water Inlet Temperature	Random Variation
IDV(12)	Condenser Cooling Water Inlet Temperature	Random Variation
IDV(13)	Reaction Kinetics	Slow Drift
IDV(14)	Reactor Cooling Water Valve	Sticking
IDV(15)	Condenser Cooling Water Valve	Sticking
IDV(16)	Unknown	
IDV(17)	Unknown	
IDV(18)	Unknown	
IDV(19)	Unknown	
IDV(20)	Unknown	
IDV(21)	The valve for Stream 4 was fixed at the steady state position	Constant Position

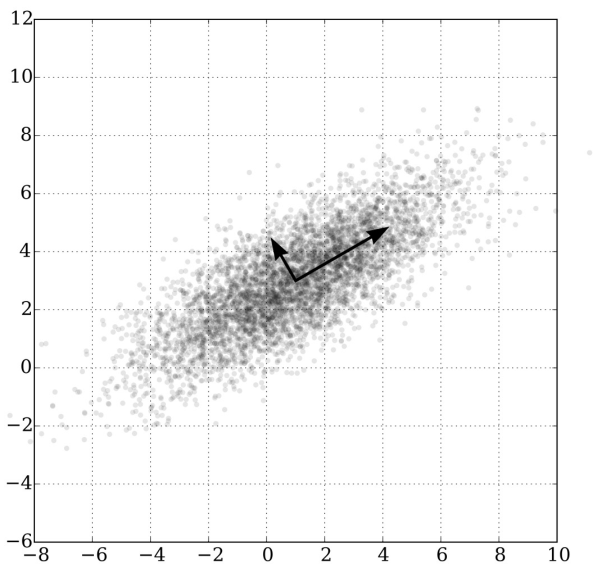
Step change in reactor cooling water temperature

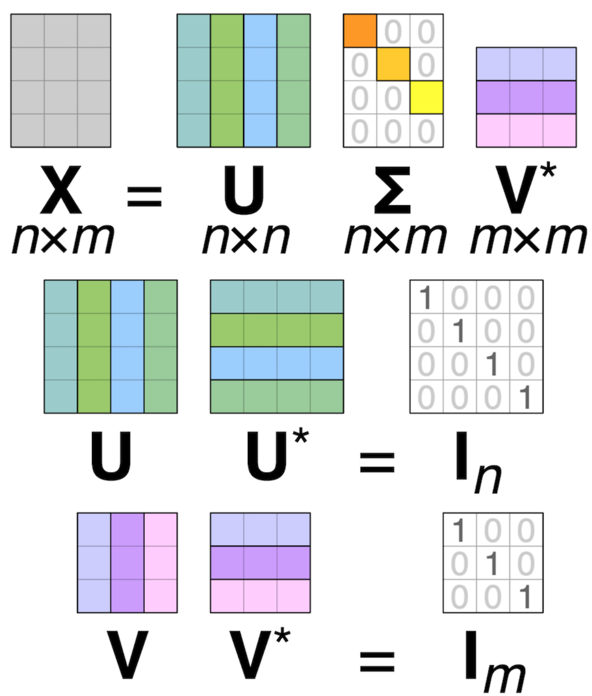
- This fault could cause runaway reaction. The controller will increase the cooling water flowrate to bring the temperature down



Step change in reactor temperature

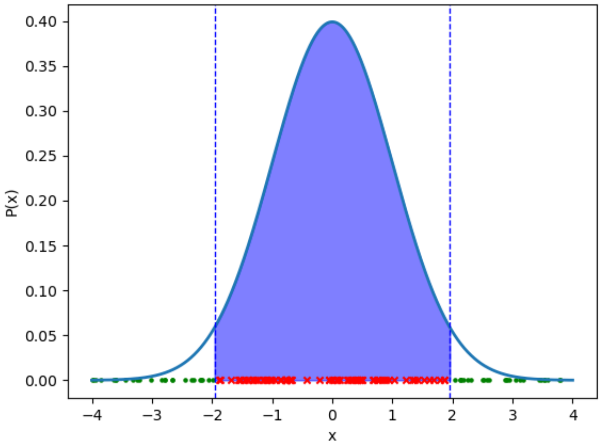
- **Principal component analysis:** identify the principal components where the data have the largest variance. The non-principal components are “noise”.
- Approach: singular value decomposition

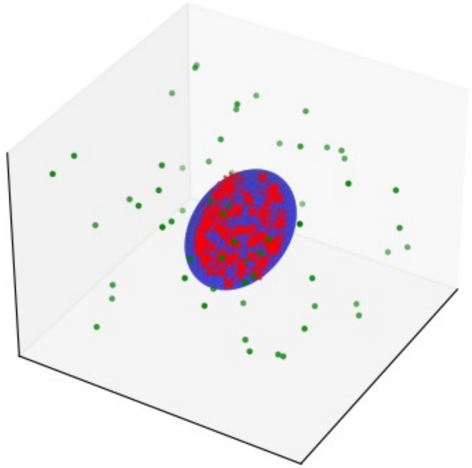




PCA

- The region within the thresholds represents the Normal Operating Condition (NOC) under random noise.
- The region outside of the thresholds represents the systematic variation from NOC.
- Test statistics measure the variation in the reduced space and indicate any fault if the thresholds are violated.

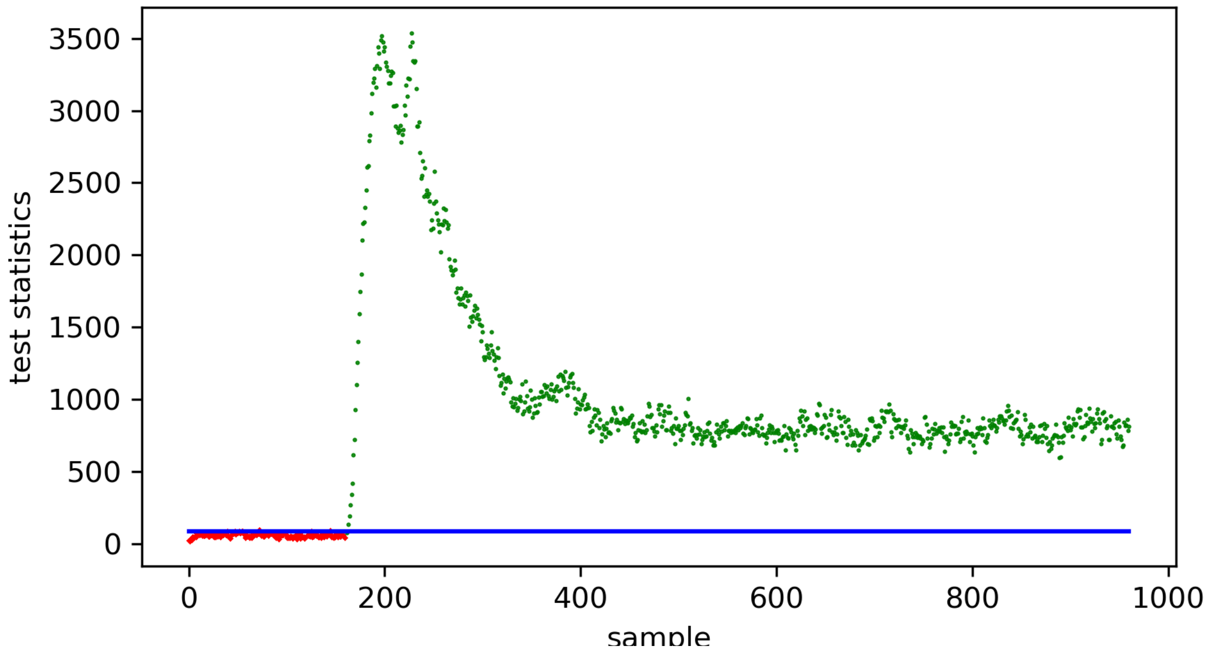




Anomaly Detection Using PCA

- T^2 statistics: describes how far the data is from normal operating condition

$$T^2 = \mathbf{x}'\mathbf{S}^{-1}\mathbf{x}$$



Limitation: Lack of interpretation.

Variable Contribution

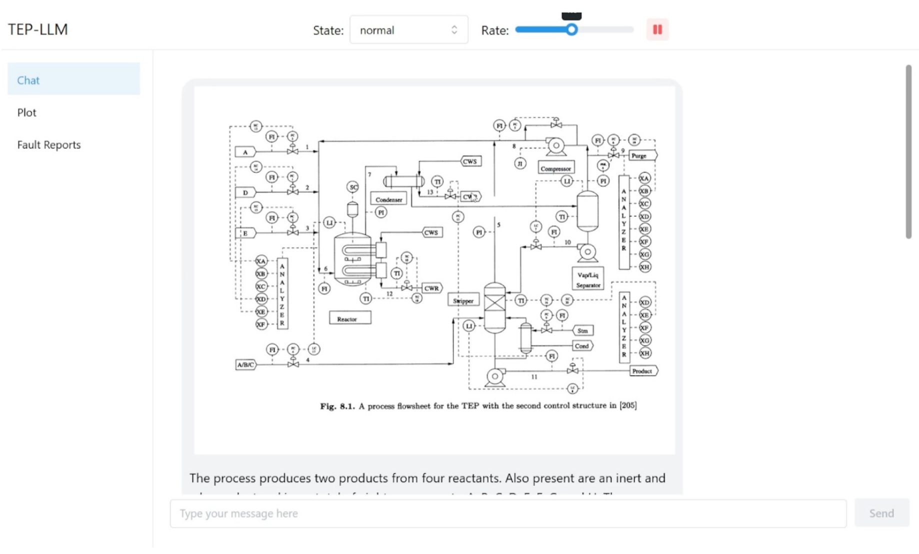
- Motivation: identify the process variables that contribute the most to the fault.

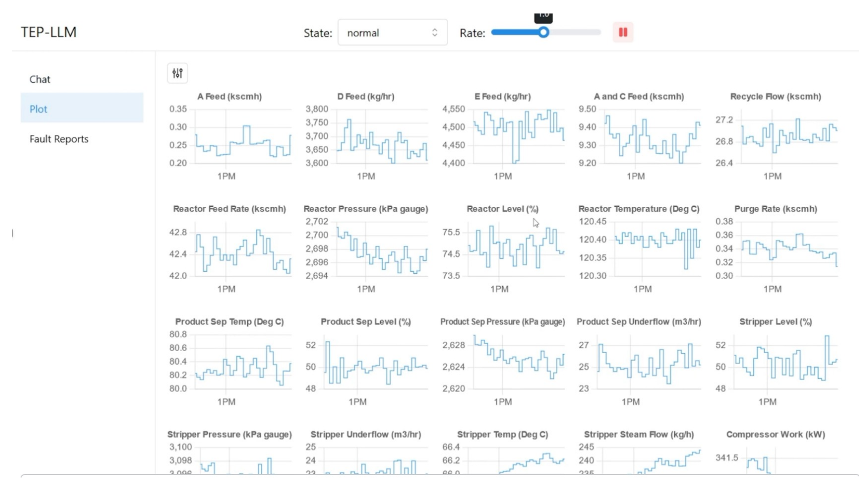
$$C_i = \sum_{j=1}^{k} \left(\frac{T_j^2}{\lambda_j} \right) p_{ij}^2$$

Variable	Description	Units
XMEAS(1)	A Feed (Stream 1)	kscmh
XMEAS(2)	D Feed (Stream 2)	kg/hr
XMEAS(3)	E Feed (Stream 3)	kg/hr
XMEAS(4)	Total Feed (Stream 4)	kscmh
XMEAS(5)	Recycle Flow (Stream 8)	kscmh
XMEAS(6)	Reactor Feed Rate (Stream 6)	kscmh
XMEAS(7)	Reactor Pressure	kPa gauge
XMEAS(8)	Reactor Level	%
XMEAS(9)	Reactor Temperature	Deg C
XMEAS(10)	Purge Rate (Stream 9)	kscmh
XMEAS(11)	Product Sep Temp	Deg C
XMEAS(12)	Product Sep Level	%
XMEAS(13)	Prod Sep Pressure	kPa gauge
XMEAS(14)	Prod Sep Underflow (Stream 10)	m^3/hr
XMEAS(15)	Stripper Level	%
XMEAS(16)	Stripper Pressure	kPa gauge
XMEAS(17)	Stripper Underflow (Stream 11)	m^3/hr
XMEAS(18)	Stripper Temperature	Deg C
XMEAS(19)	Stripper Steam Flow	kg/hr
XMEAS(20)	Compressor Work	kW
XMEAS(21)	Reactor Cooling Water Outlet Temp	Deg C
XMEAS(22)	Separator Cooling Water Outlet Temp	Deg C

Overview of FaultExplainer

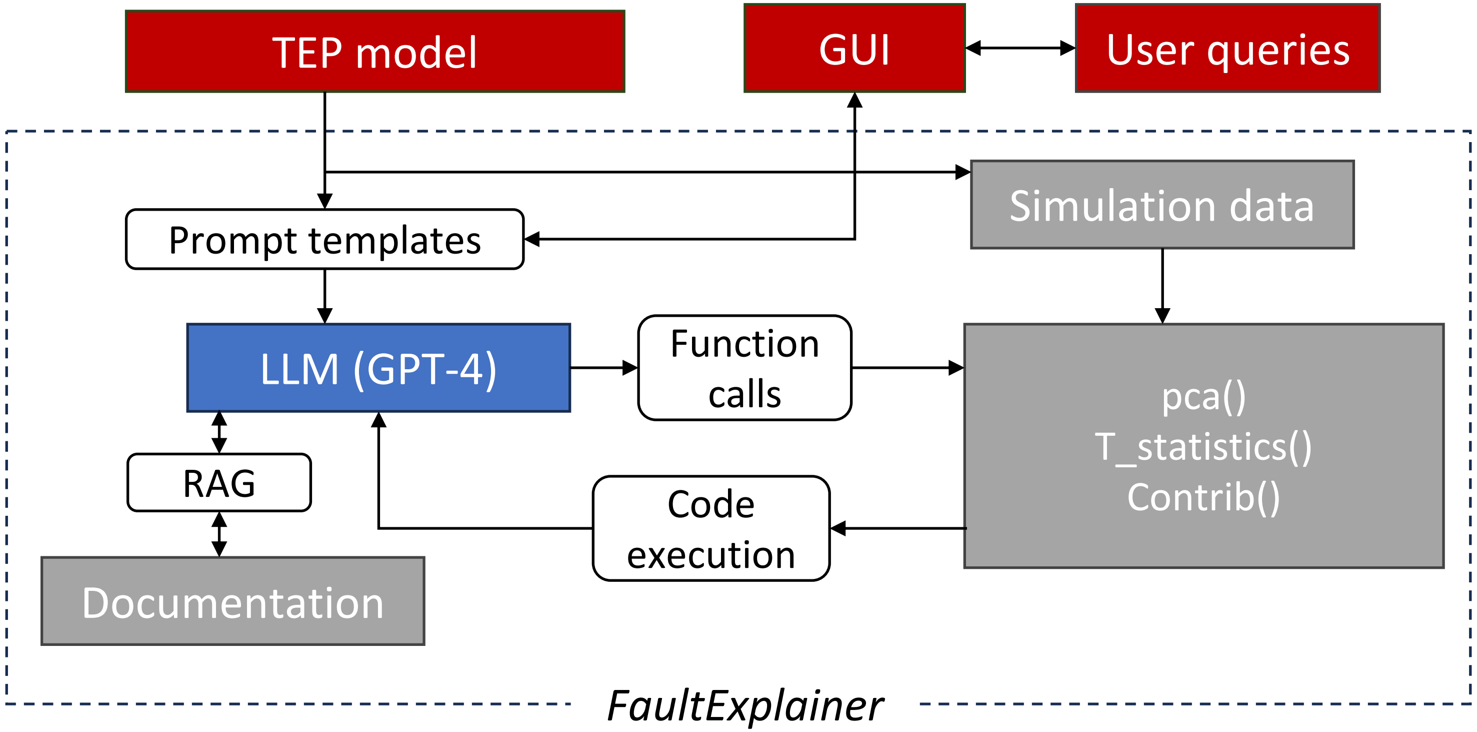
- FaultExplainer has a GUI to monitor all the process variables
- Fault detection and identification is conducted by PCA.
- LLM generates a fault report to explain the potential causes of the fault to process operators.
- FaultExplainer can also answer general queries.





What's inside the box?

- **LLM:** the brain (coordinate user queries and all the tasks).
- **Prompt templates:** instructions to the LLM with demonstrations
- **Function calls:** conduct quantitative tasks such as PCA.
- **Retrieval-Augmented Generation (RAG):** provide external information to alleviate hallucination, e.g., process knowledge of TEP.



LLM for Process Monitoring and Fault Detection

TEP-LLM demo

localhost:5173

120%

TEP-LLM

State: normal

Rate: 1.0

Chat

Plot

Fault Reports

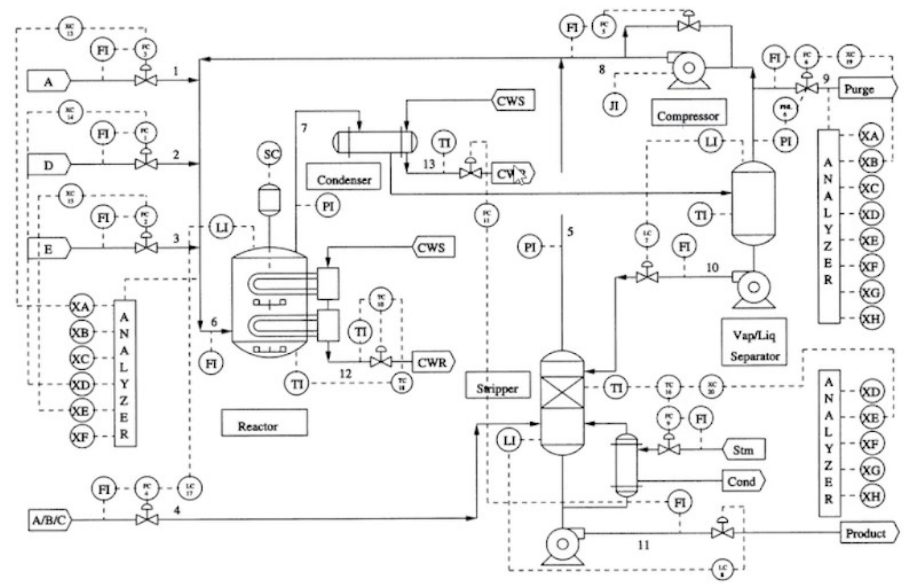


Fig. 8.1. A process flowsheet for the TEP with the second control structure in [205]

The process produces two products from four reactants. Also present are an inert and

Type your message here

Send

Qualitative Analysis of GPT-4o and o1 Models

- Both models are prompted to diagnose the step change in the feed composition of inert B based on the top six changes in the sensor measurement.

- **Root Cause 1: Increased Reactor Pressure or Flow Rate:** An increase in reactor pressure or flow rate could lead to higher purge rates and changes in the composition of the purge stream. The control system might respond by increasing the purge valve opening to manage the pressure, leading to the observed increase in purge rate and component B in the purge. The increased flow rate could also result in more component B being carried over to the reactor, explaining the increase in component B to the reactor. The increase in A and E feed loads could be a response to maintain reaction rates and product output under the new conditions. This root cause can explain all six feature deviations.

GPT-4o

O1 is able to provide a more logical and quantitative reasoning of the fault propagation, including how the control system reacted to the fault.

Root Cause 1: Increased Ingress of Inert B into the System
An unexpected increase in the influx of inert Component B into the process could explain the observed feature deviations. This may result from contamination in one of the feed streams, particularly Stream 4, which includes a mixture of A, B, and C. If the concentration of B in this stream increases, more inert B enters the reactor feed, leading to an elevated mole percentage of Component B to the Reactor (Feature 4), as observed (+4.17%). The accumulation of inert B in the process causes higher concentrations of B in the recycle loop. To prevent excessive buildup of inert gas, the control system responds by increasing the Purge Valve opening (Feature 3, +20.86%) to remove the excess inert. This action increases the Purge Rate (Feature 2, +20.96%) and results in higher concentrations of Component B in the Purge stream (Feature 1, +5.63%).
The increased presence of inert B in the reactor feed dilutes the concentrations of reactive species, potentially reducing the overall reaction rates. The control system compensates for this by increasing the feed rates of reactants A and E. This is evidenced by the significant increase in A Feed Load (Feature 6, +30.74%) and the smaller increase in E Feed Load (Feature 5, +1.67%). Since A participates in all reactions and is more critical, a larger increase in its feed rate is justified. Therefore, this root cause explains all six observed feature deviations.
Features explained: 6/6.

O1

Quantitative Results

- We prompt the LLMs to give top 3 possible causes of the fault based on the top 6 contributing features
- PCA can identify 11/15 of the faults
- The success rate of identifying the root cause of the fault

GPT-4o: 6/11 o1: 9/11

Variable	Description	Type
IDV(1)	A/C Feed Ratio, B Composition Constant (Stream 4)	Step
IDV(2)	B Composition, A/C Ratio Constant (Stream 4)	Step
IDV(3)	D Feed Temperature (Stream 2)	Step
IDV(4)	Reactor Cooling Water Inlet Temperature	Step
IDV(5)	Condenser Cooling Water Inlet Temperature	Step
IDV(6)	A Feed Loss (Stream 1)	Step
IDV(7)	C Header Pressure Loss - Reduced Availability (Stream 4)	Step
IDV(8)	A, B, C Feed Composition (Stream 4)	Random Variation
IDV(9)	D Feed Temperature (Stream 2)	Random Variation
IDV(10)	C Feed Temperature (Stream 4)	Random Variation
IDV(11)	Reactor Cooling Water Inlet Temperature	Random Variation
IDV(12)	Condenser Cooling Water Inlet Temperature	Random Variation
IDV(13)	Reaction Kinetics	Slow Drift
IDV(14)	Reactor Cooling Water Valve	Sticking
IDV(15)	Condenser Cooling Water Valve	Sticking
IDV(16)	Unknown	
IDV(17)	Unknown	
IDV(18)	Unknown	
IDV(19)	Unknown	
IDV(20)	Unknown	
IDV(21)	The valve for Stream 4 was fixed at the steady state position	Constant Position

Conclusions and Future Work

- Combining LLMs like GPT and o1 with mathematical models can get both explainability and rigor.
- With sufficient training data, we can fine-tune a LLM for tasks such as PHA, HAZOP, explaining controller behavior.