

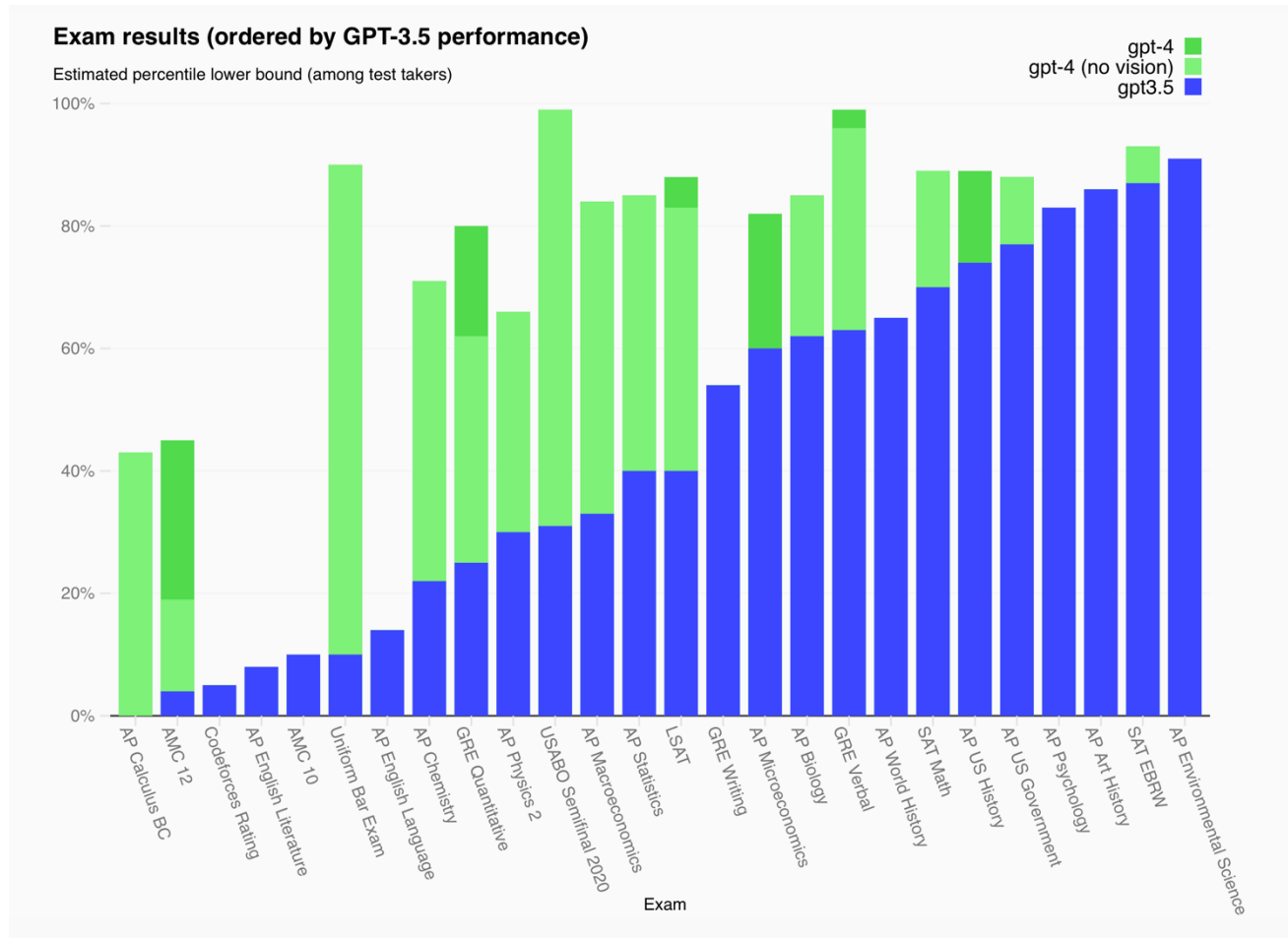
FaultExplainer: Leveraging Large Language Models for Interpretable Fault Detection and Diagnosis

Can Li
Assistant Professor
Davidson School of Chemical Engineering

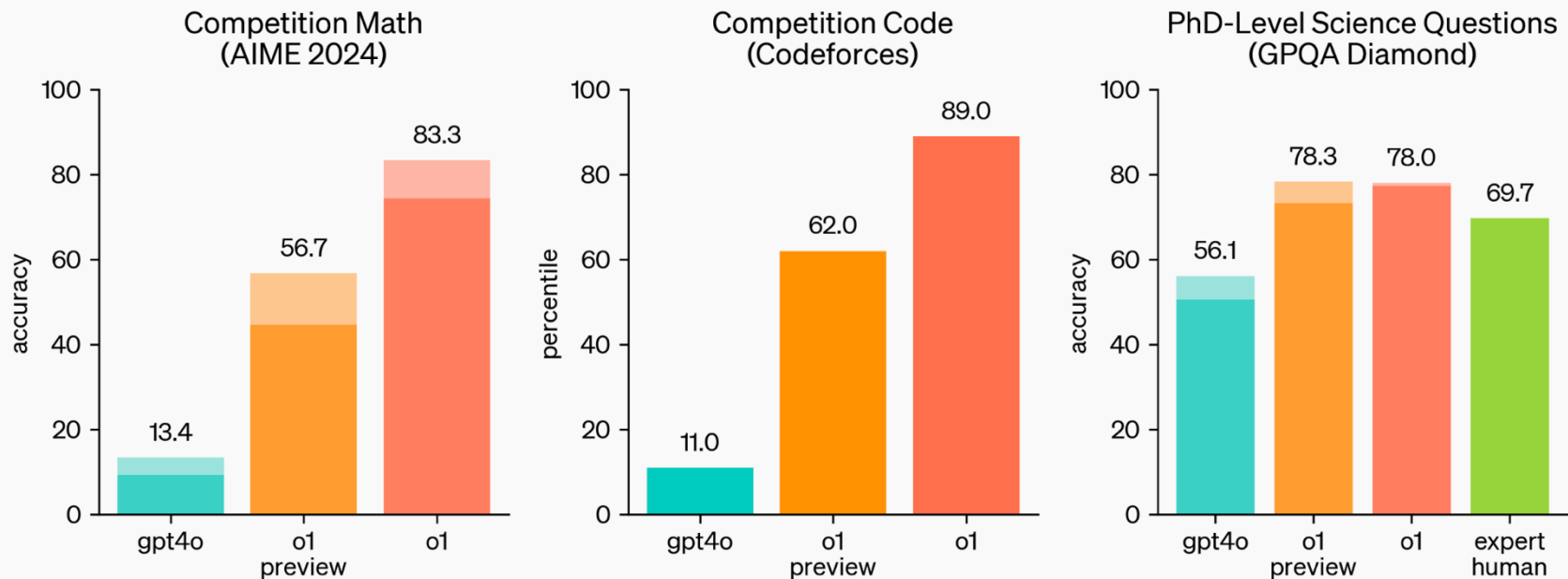
What can LLMs do



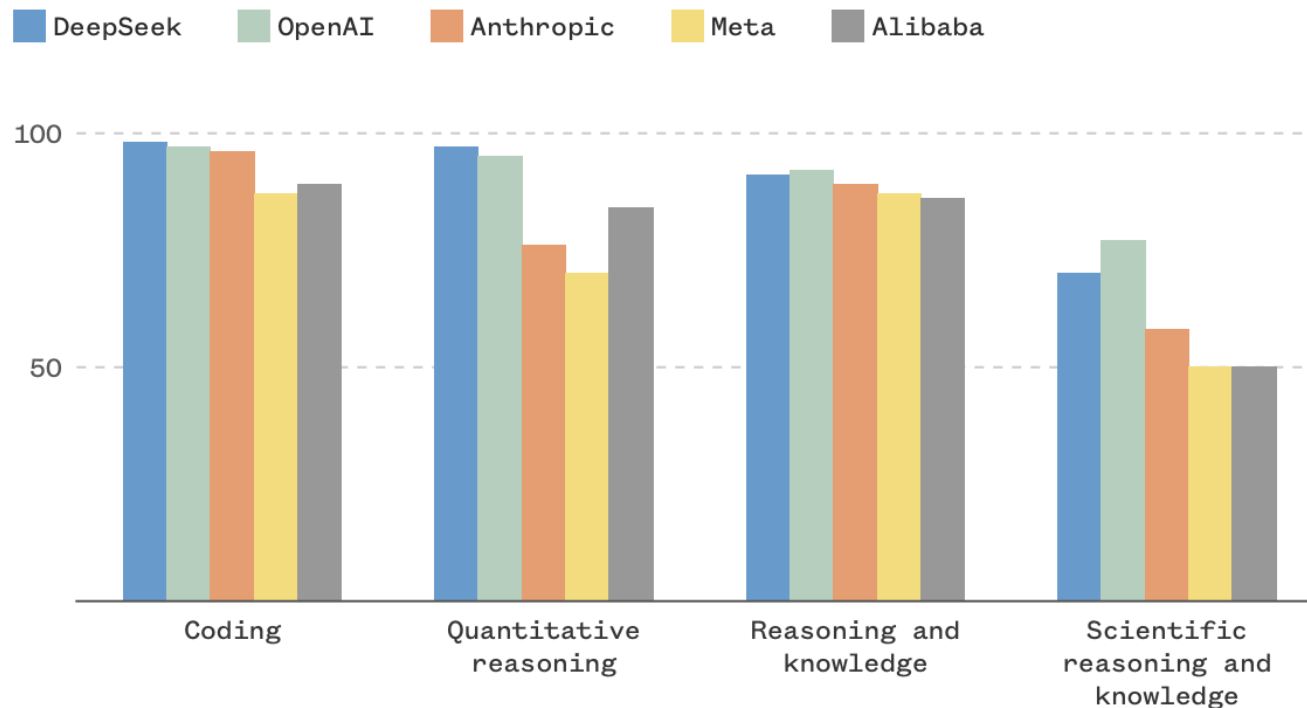
- GPT-4, released on March 2023. Excel at qualitative subjects: History, biology, bar exam, and verbal. Struggles with math, coding, and reasoning



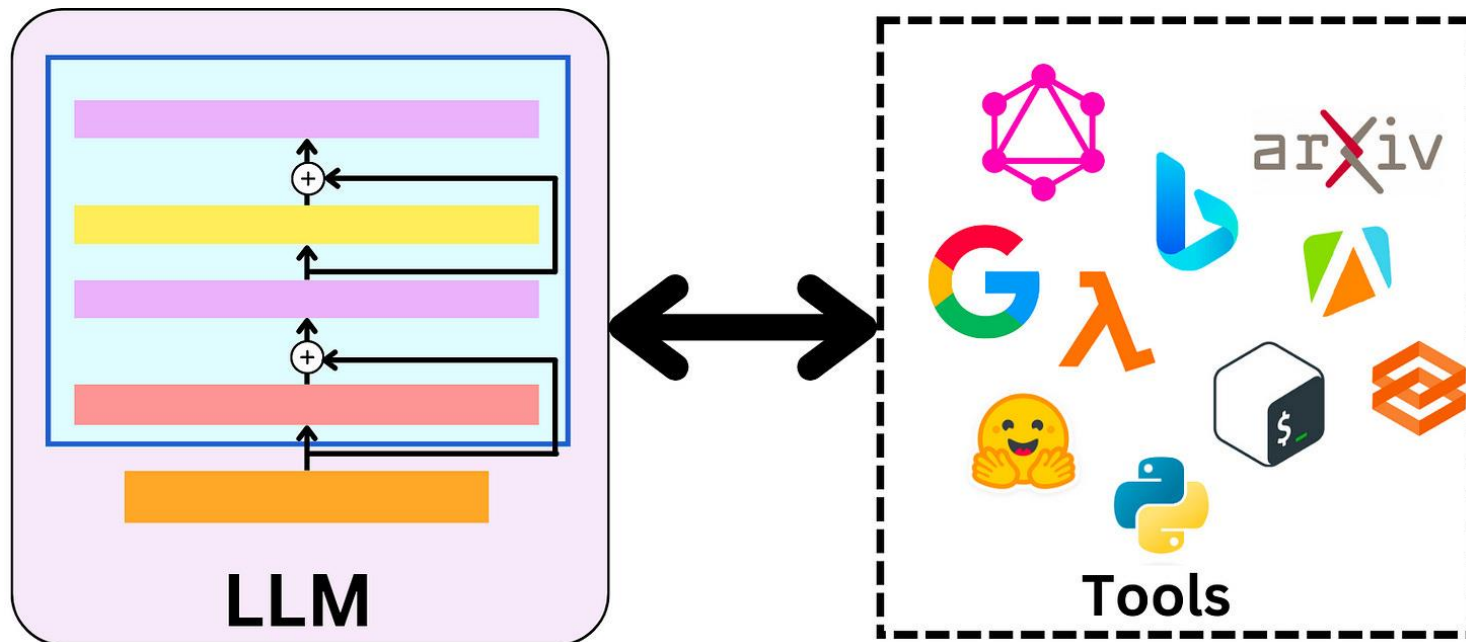
- OpenAI o1 model: released on September 2024. Significant improvement in math, science, and reasoning.
- Similar to how a human may think for a long time before responding to a difficult question, o1 uses a **chain of thought** when attempting to solve a problem



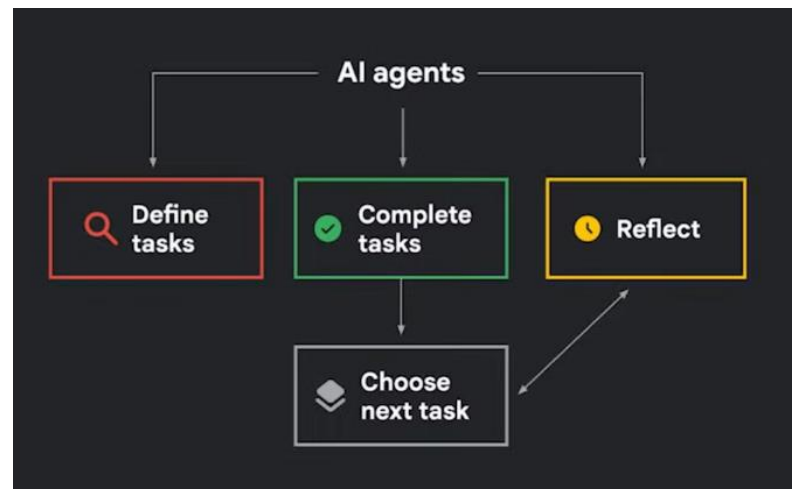
- DeepSeek-R1 released on January 20, 2025 : \$5.6 million, it delivers performance comparable to models like OpenAI's GPT-4, which reportedly cost over \$100 million to develop.
- DeepSeek-R1 utilizes a Mixture-of-Experts (MoE) architecture, activating only relevant parts of the model for each task, thereby enhancing efficiency.



- LLMs can be augmented with external tools to extend their capabilities
- Integration with tools mitigates hallucination and improves performance on complex, real-world tasks.
- Web search (Google, Bing), scientific databases (arXiv), code execution (Python shell), APIs (GraphQL), and libraries (Hugging Face, LangChain).



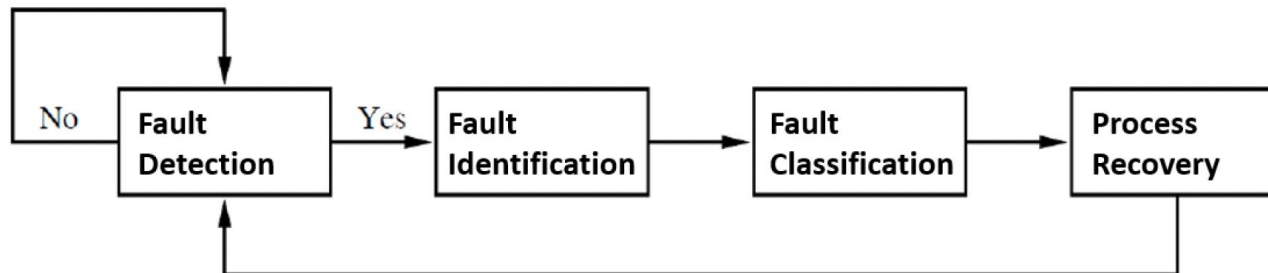
- Unlike traditional workflows, AI agents can define, complete, and reflect on tasks independently.
- Agents not only perform tasks but also evaluate outcomes and select the next steps without human input.
- Key capabilities of AI agents:
 - Define Tasks: Identify what needs to be done.
 - Complete Tasks: Execute tasks using tools and knowledge.
 - Reflect: Evaluate outcomes to improve performance.
 - Choose Next Task: Dynamically adapt goals and strategies.



- Cosientists
 - An AI copilot for chemical research (Boiko et al., 2023), autonomously designing, planning, and conducting complex experiments
 - Google's AI co-scientist (Gottweis et al. 2025): automated research hypothesis generation
- LLaVA-Med, BiomedGPT, Med-Gemini
 - AI Assistants that help doctors in biomedical imaging by analyzing medical images and generating diagnostic reports
- scChat
 - AI Assistants to help stem cell researchers analyze RNA sequencing data, generate research hypothesis and experimental design.
- FaultExplainer
 - Interpret faults identified in a chemical plant by a PCA algorithm

Combine LLM with Machine Learning Models for Process Monitoring

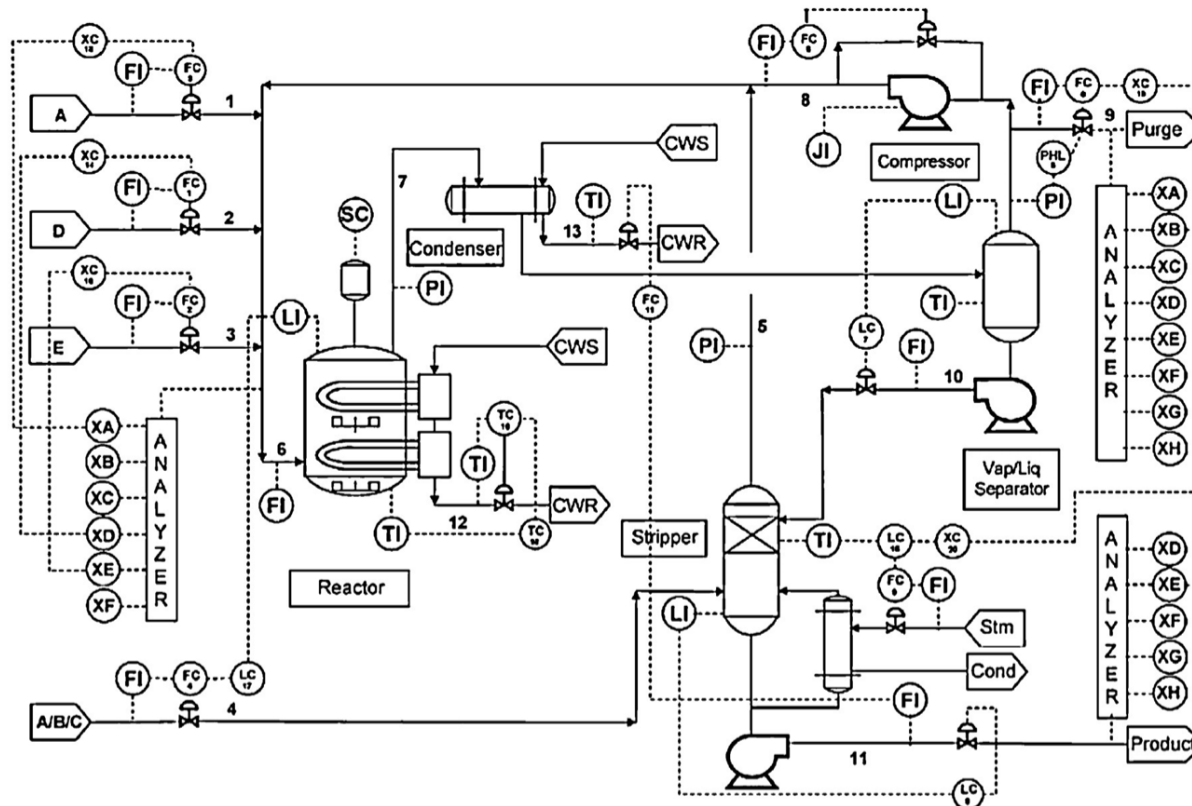
- Fault **Detection**: Detect if a fault has occurred
- Fault **Identification**: Identify the variables most relevant to the fault
- Fault **Diagnosis** (or Classification): Diagnose the root cause of the fault



Tennessee Eastman Process



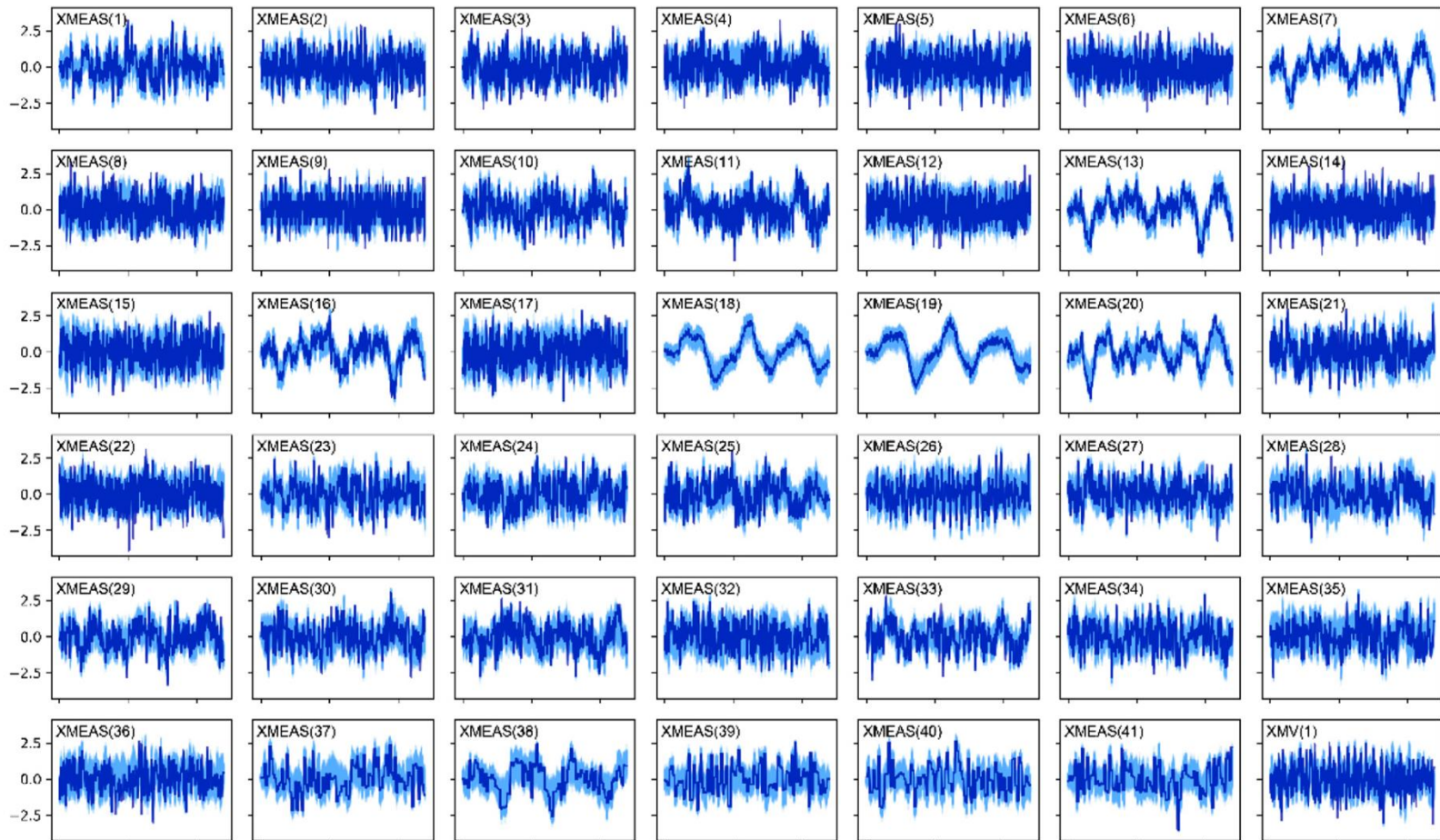
- TEP is an open-source simulator written in Fortran that resembles a real chemical process by Eastman
- Time series data can be collected from over 40 sensors that measure the state variables.
- Task: From measured state variables, perform fault detection using ML/AI



Examples of State Variables with Sensor Data



- Examples include feed flow rates, temperatures, pressures



List of Potential Faults



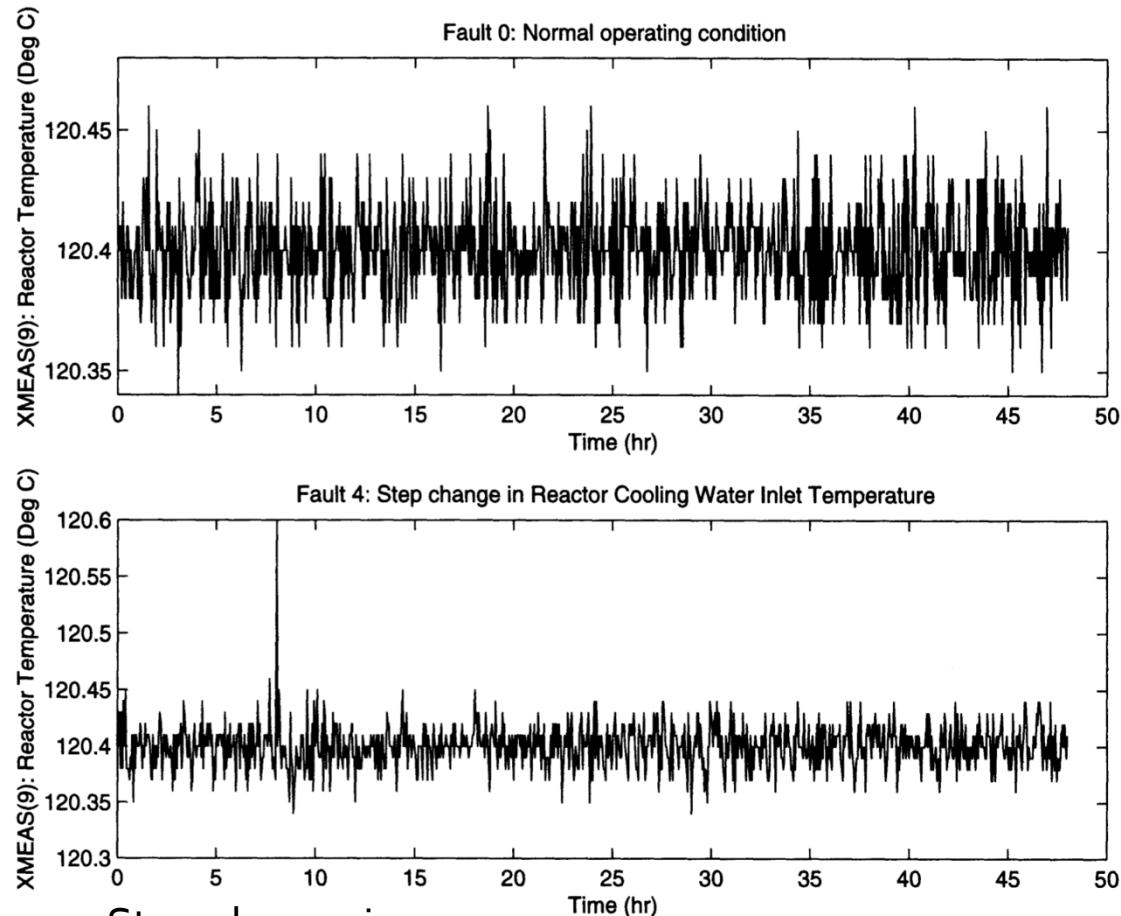
- The following “faults” are created synthetically by the simulator
- These faults will cause the measured state variables to change from their normal operating conditions which further cause safety hazards.

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

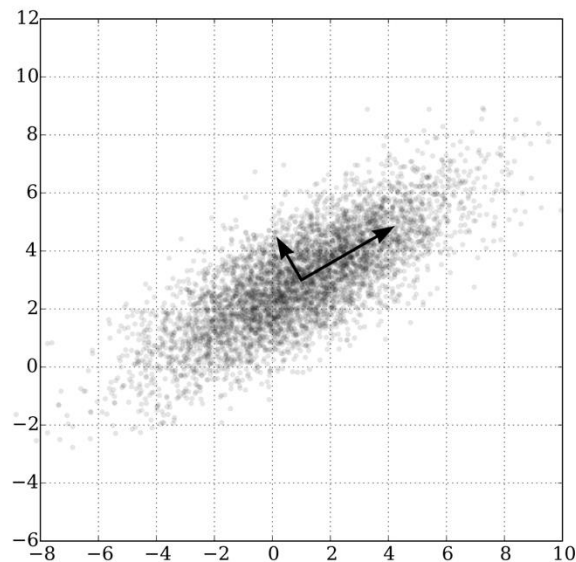


- This fault could cause runaway reaction. The controller will increase the cooling water flowrate to bring the temperature down



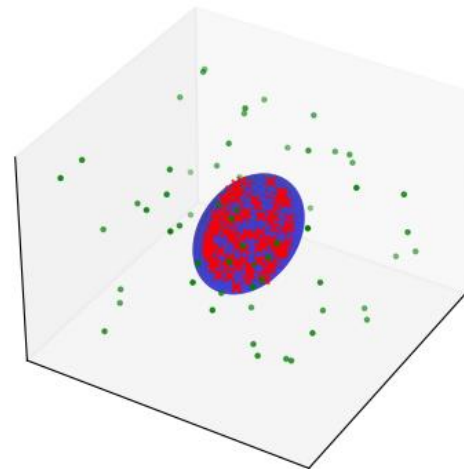
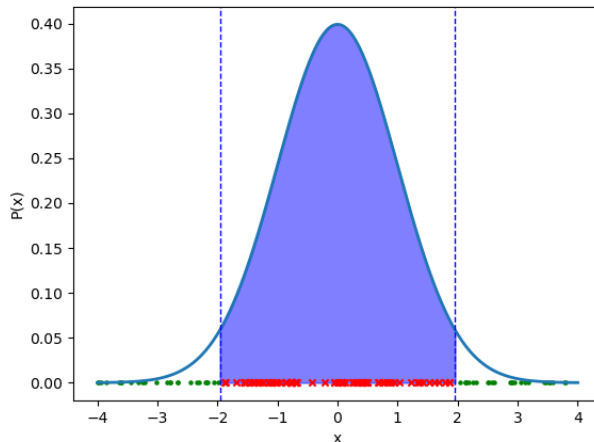
Step change in
reactor temperature

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



$$\begin{array}{cccc}
 \begin{array}{|c|c|c|} \hline \square & \square & \square \\ \hline \square & \square & \square \\ \hline \square & \square & \square \\ \hline \end{array} &
 \begin{array}{|c|c|c|c|} \hline \square & \square & \square & \square \\ \hline \square & \square & \square & \square \\ \hline \square & \square & \square & \square \\ \hline \end{array} &
 \begin{array}{|c|c|c|} \hline \square & \square & \square \\ \hline \square & \square & \square \\ \hline \square & \square & \square \\ \hline \end{array} &
 \begin{array}{|c|c|c|} \hline \square & \square & \square \\ \hline \square & \square & \square \\ \hline \square & \square & \square \\ \hline \end{array} \\
 \mathbf{X} = \mathbf{U} \mathbf{\Sigma} \mathbf{V}^* \\
 n \times m & n \times n & n \times m & m \times m \\
 \\
 \begin{array}{|c|c|c|c|} \hline \square & \square & \square & \square \\ \hline \square & \square & \square & \square \\ \hline \square & \square & \square & \square \\ \hline \end{array} &
 \begin{array}{|c|c|c|c|} \hline \square & \square & \square & \square \\ \hline \square & \square & \square & \square \\ \hline \square & \square & \square & \square \\ \hline \end{array} &
 \begin{array}{|c|c|c|c|} \hline \square & \square & \square & \square \\ \hline \square & \square & \square & \square \\ \hline \square & \square & \square & \square \\ \hline \end{array} & = \mathbf{I}_n \\
 \mathbf{U} & \mathbf{U}^* & & \\
 \\
 \begin{array}{|c|c|c|} \hline \square & \square & \square \\ \hline \square & \square & \square \\ \hline \square & \square & \square \\ \hline \end{array} &
 \begin{array}{|c|c|c|} \hline \square & \square & \square \\ \hline \square & \square & \square \\ \hline \square & \square & \square \\ \hline \end{array} & = \mathbf{I}_m \\
 \mathbf{V} & \mathbf{V}^* &
 \end{array}$$

- 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.

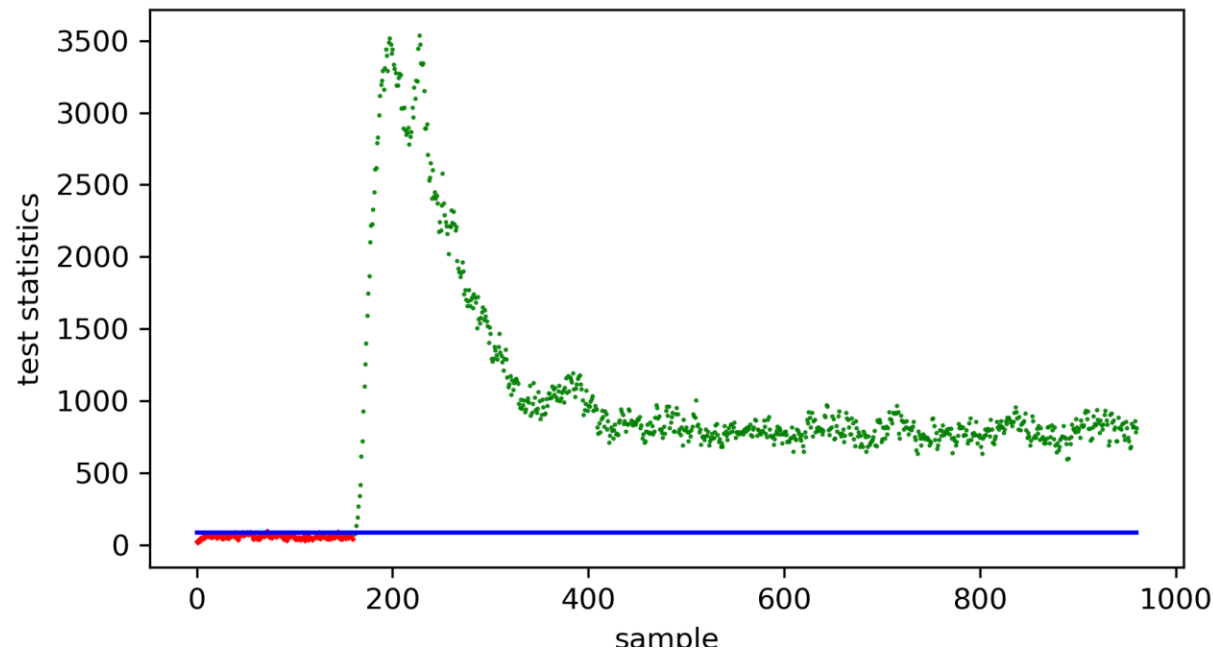


Anomaly Detection Using PCA



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

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



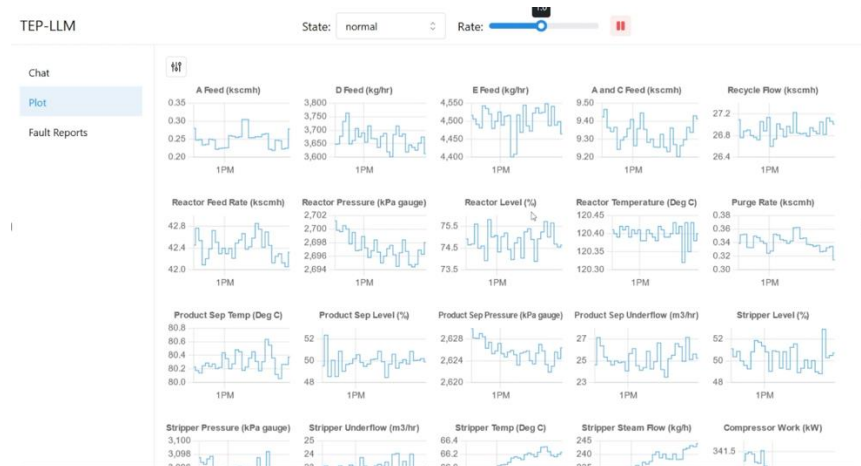
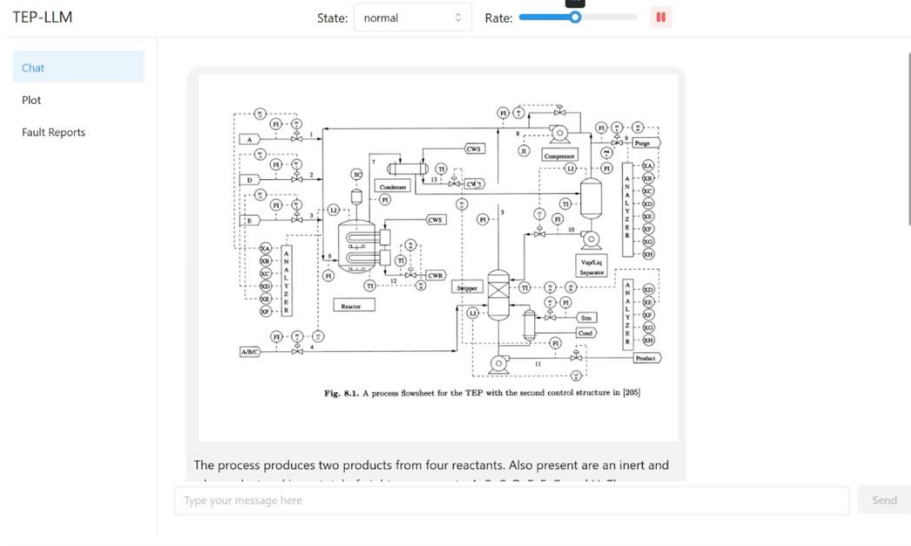
Limitation: Lack of interpretation.

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

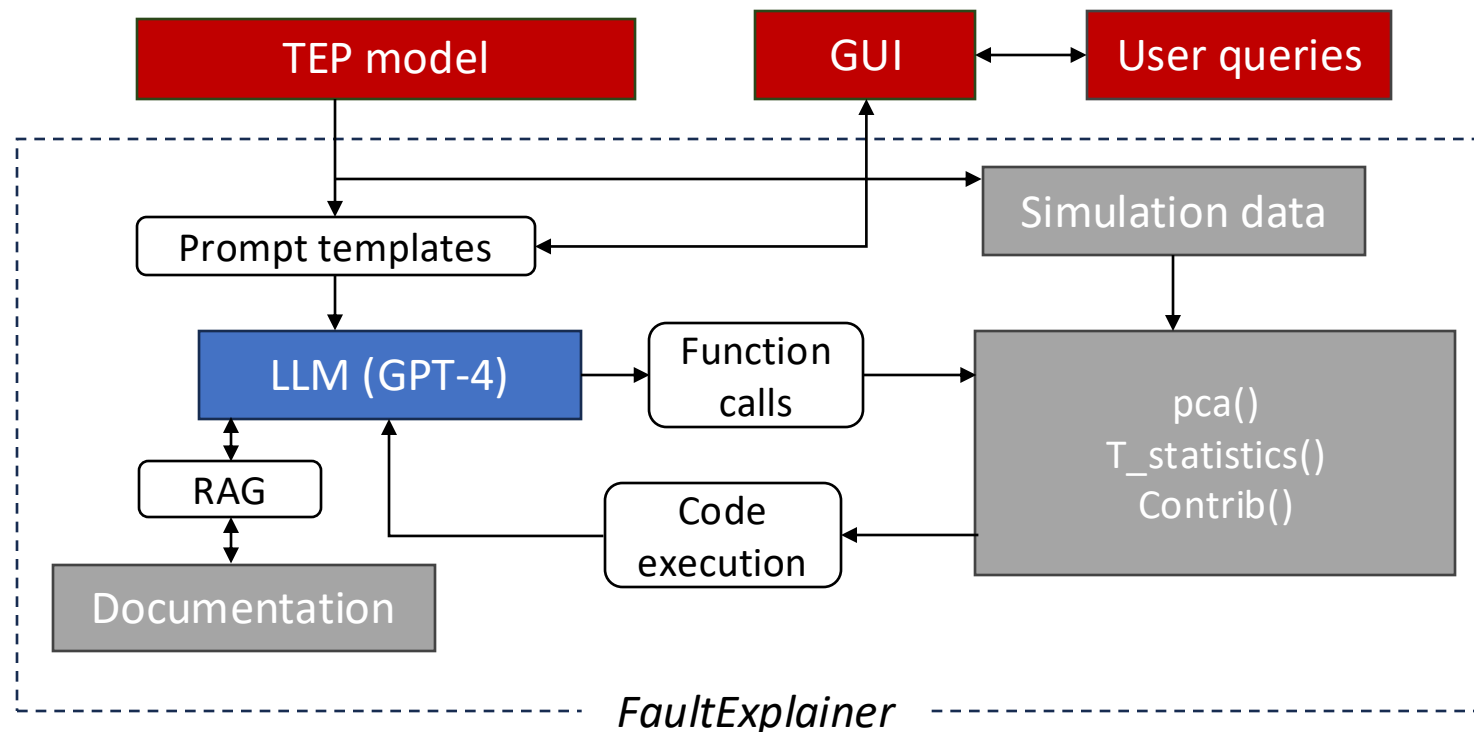
$$C_i = \sum_{j=1}^k \left(\frac{T_j^2}{\lambda_j} \right) p_{ij}^2$$

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

- 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.



- **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.



FaultExplainer Demo



FaultExplainer

Fault: Normal Operation



Monitoring

Assistant

Fault History



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 a byproduct making a total of eight components: A, B, C, D, E, F, G, and H. The reactions are:

$A(g) + C(g) + D(g) \rightarrow G(liq)$: Product 1,

$A(g) + C(g) + E(g) \rightarrow H(liq)$: Product 2,

$A(g) + E(g) \rightarrow F(liq)$: Byproduct,

$3D(g) \rightarrow 2F(liq)$: Byproduct.

All the reactions are irreversible and exothermic. The reaction rates are a function of temperature through an Arrhenius expression. The reaction to produce G has a higher activation energy resulting in more sensitivity to temperature. Also, the reactions are approximately first-order with respect to the reactant concentrations.

The process has five major unit operations: the reactor, the product condenser, a vapor-liquid separator, a recycle compressor and a product stripper. Figure showing a diagram of the process is attached.

The gaseous reactants are fed to the reactor where they react to form liquid products. The gas phase reactions are catalyzed by a nonvolatile catalyst dissolved in the liquid phase. The reactor has an internal cooling bundle for removing the heat of reaction. The products leave the reactor as vapors along with the unreacted feeds. The catalyst remains in the reactor. The reactor product stream passes through a cooler for condensing the products and from there to a vapor-liquid separator. Noncondensed components recycle back through a centrifugal compressor to the reactor feed. Condensed components move to a product stripping column to remove remaining reactants by stripping with feed stream number 4. Products G and H exit the stripper base and are separated in a downstream refining section which is not included in this problem. The inert and byproduct are primarily purged from the system as a vapor from the vapor-liquid separator.

Ask something here ...



- 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.

- 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

- When the root cause is **C Feed Temperature Random Variation**
- O1-preview gives a plausible explanation
- However, the top six features given to the LLM are directly related to the root causes.

Top six feature changes

The **Stripper Temperature** increased by 2.45%, the **A Feed Load** increased by 43.90%, and the **A Feed Rate** increased by 43.63%. Meanwhile, the **E Feed Rate** decreased by 2.38%. The **Stripper Steam Valve** position increased by 14.59%, and the **Stripper Steam Flow** increased by 9.93%.

o1-preview explanation

The most plausible root cause is **IDV(13): Reaction Kinetics & Slow Drift**.

A gradual deterioration of reaction kinetics, possibly from catalyst deactivation or fouling, would reduce the conversion rates of reactants to products. To maintain the desired production levels, the control system responds by increasing the **A feed load** and **A Feed**, supplying more reactant A to the reactor. This explains the significant increases in these features. The slight decrease in **E Feed** may result from an accumulation of unreacted E in the reactor due to slower reaction rates, prompting the control system to reduce its feed rate. The reduced reaction efficiency leads to higher concentrations of unreacted reactants in the reactor effluent. Consequently, the control system increases the **Stripper steam valve** opening and the **Stripper Steam Flow** to enhance the removal of residual reactants from the product stream. This accounts for the observed increases in these features. The increased steam flow raises the **Stripper Temperature**, aligning with the observed temperature rise. All six of the top features can be explained by **IDV(13)**.

- **PCA Limitations:** Current reliance on PCA may miss subtle or indirect variable-root cause relationships.
- **LLM Limitations:** Explanations can be overly generic or hallucinated, especially when input features lack direct relevance.
- **Better Feature Selection:**
 - Integrate domain-informed feature engineering or causal inference to improve fault detection.
- **Hybrid Model Architecture:**
 - Combine explainable AI with black-box deep learning for richer intermediate outputs to guide LLMs.
- **Domain-Specific Language Models:**
 - Fine-tune LLMs using process data and expert knowledge to enhance explanation quality.
- **Human Feedback Loop:**
 - Apply RLHF to align model outputs with user needs (e.g., operators and R&D scientists).
- **Evaluation Metrics:**
 - Develop trust-based metrics for operator validation of explanations.

➤ Papers and preprints:

- Khan, A., Nahar, R., Chen, H., Flores, G. E. C., & Li, C. (2025). FaultExplainer: Leveraging large language models for interpretable fault detection and diagnosis. Computers & Chemical Engineering, 109152.

➤ Code

- <https://github.com/li-group/FaultExplainer>

Thanks for the financial support and advice from P2SAC members.

