Applications of AI in the Chemical Industry: A Tutorial

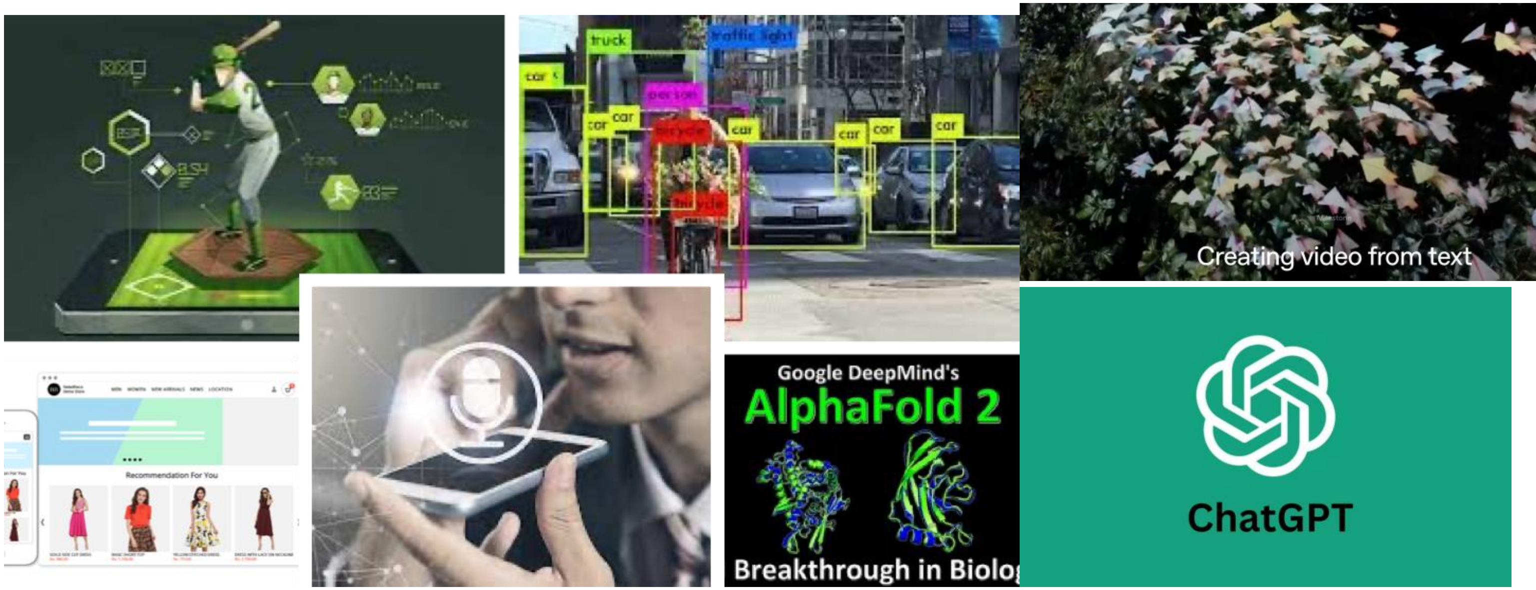
Can Li
Davidson School of Chemical Engineering

Outline

- Background and trends
- Introduction to the basics of modern AI/machine learning
- Existing applications of AI in the chemical industry
- Limitations of existing approaches and future prospects

Recent Advances

- Huge success in the past decade.



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

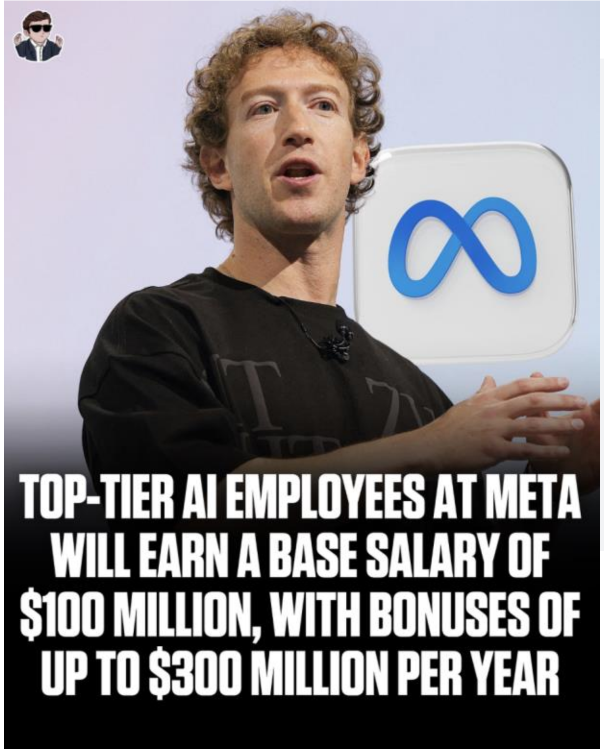
Cray 2 supercomputer (1985)
1.9GFlops

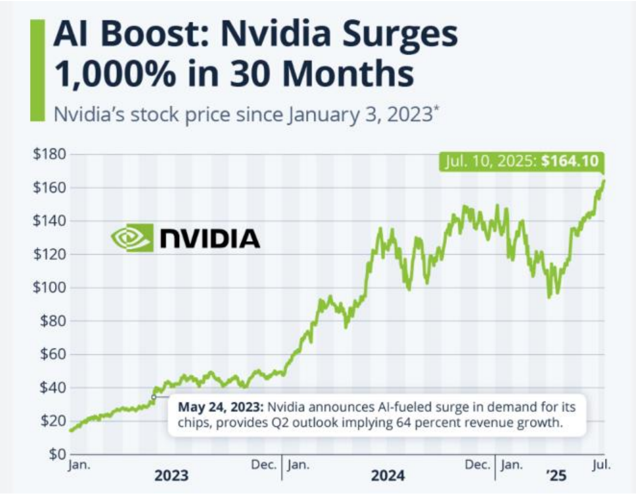
40.8GFlops

Credit: Prof. Venkat Venkatasubramanian

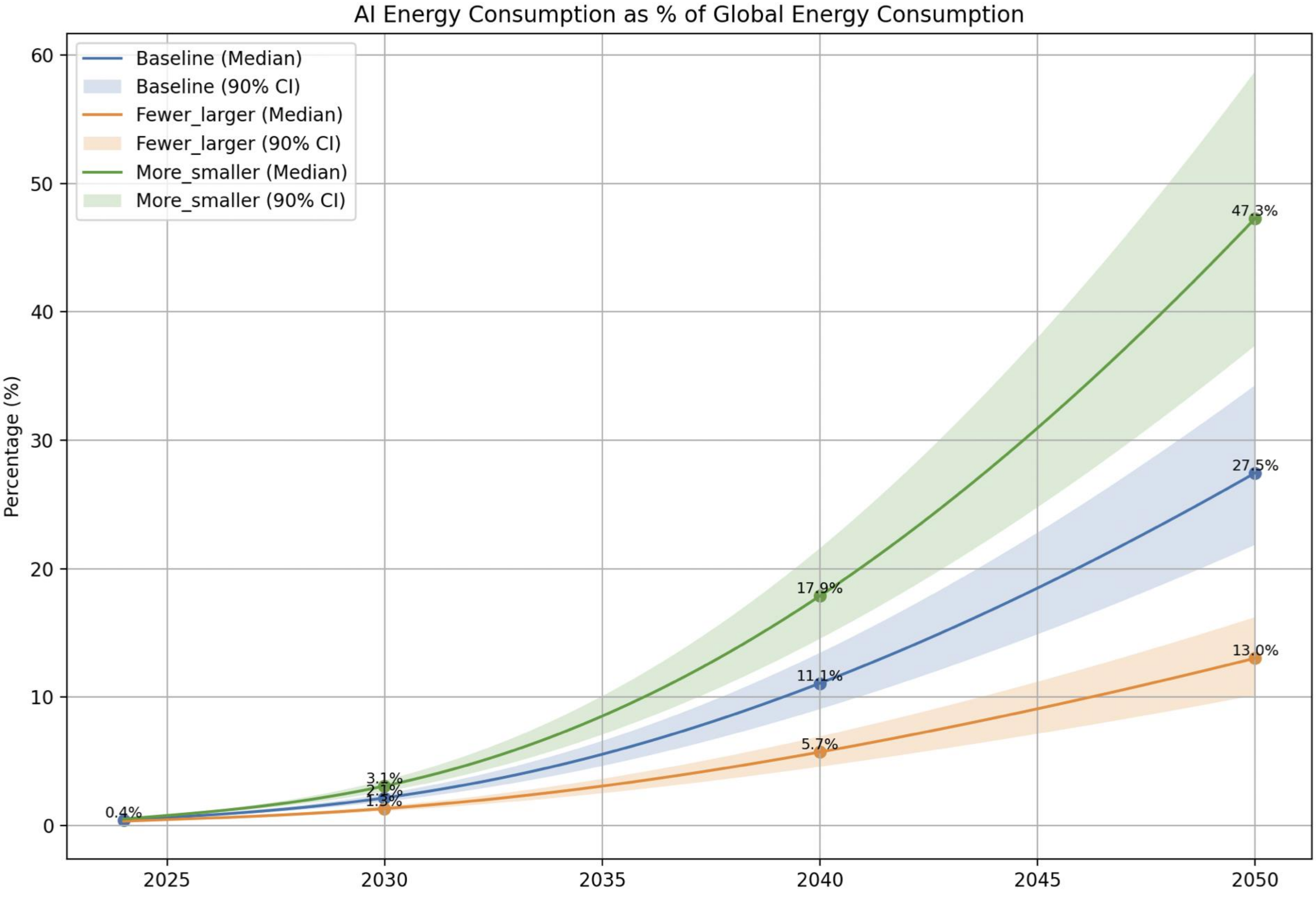
AI: Gold Rush

- Billion-dollar GPU clusters running under hurricane-proof canvas not steel, in just 3 weeks.
- 400 MW on-site power substations, $2-3B worth of compute under each tent
- 20,000+ GPUs per site





AI: Energy Consumption

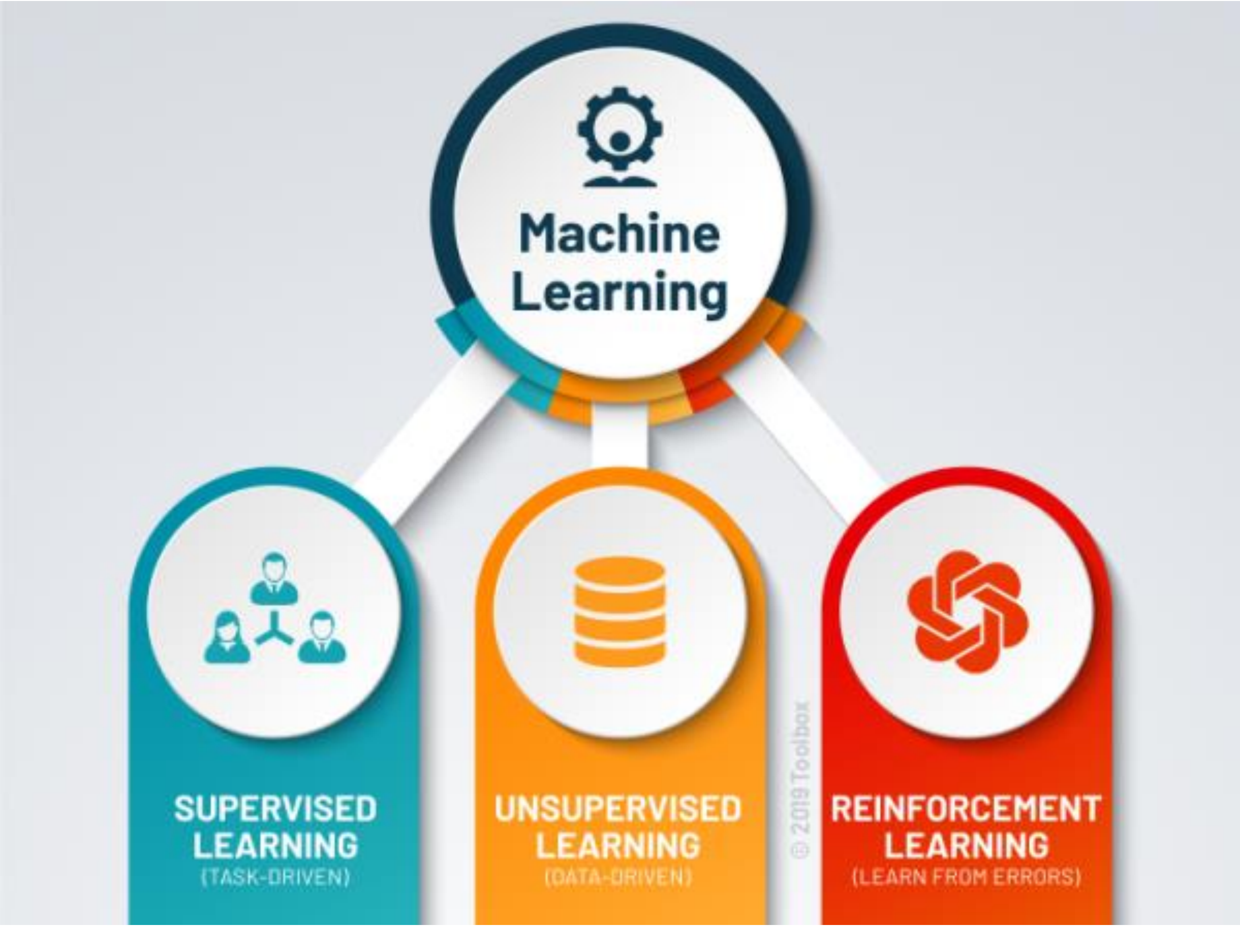


Source: Türkay et al (2025)

Machine Learning
Introduction

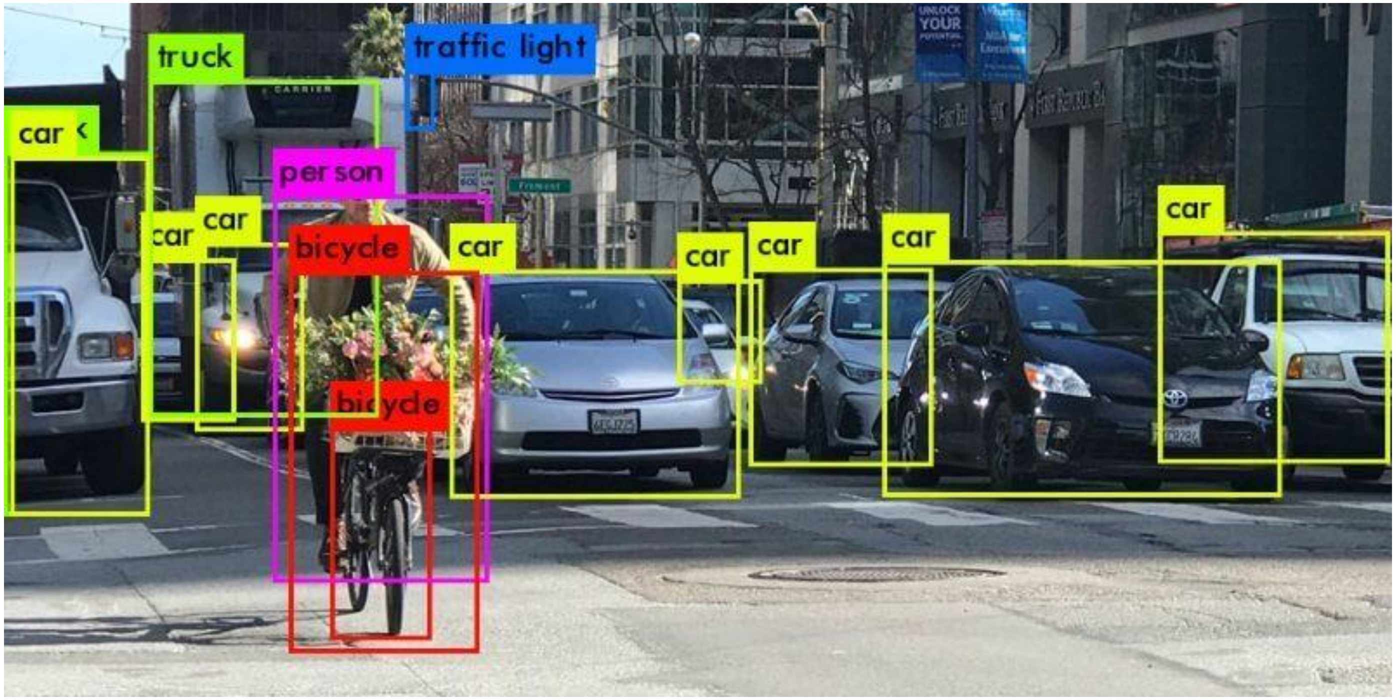
Types of Machine Learning

- Machine learning uses data to train a model and then use that model for prediction or decision-making.



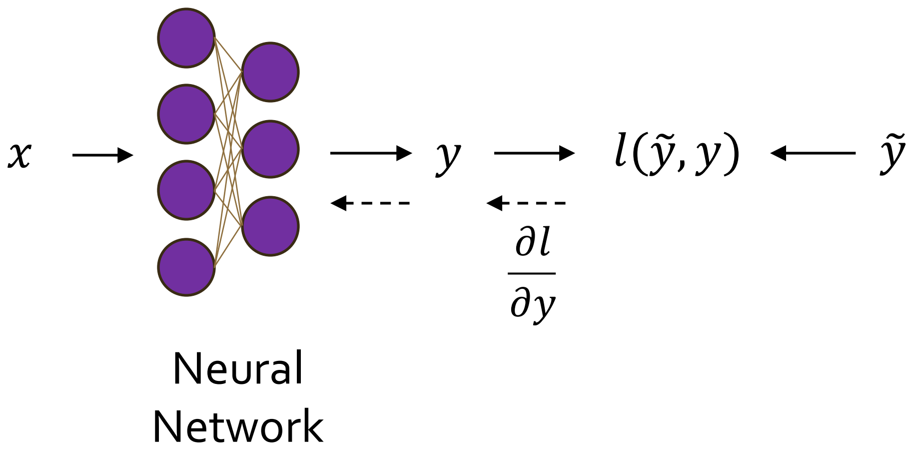
Supervised Learning

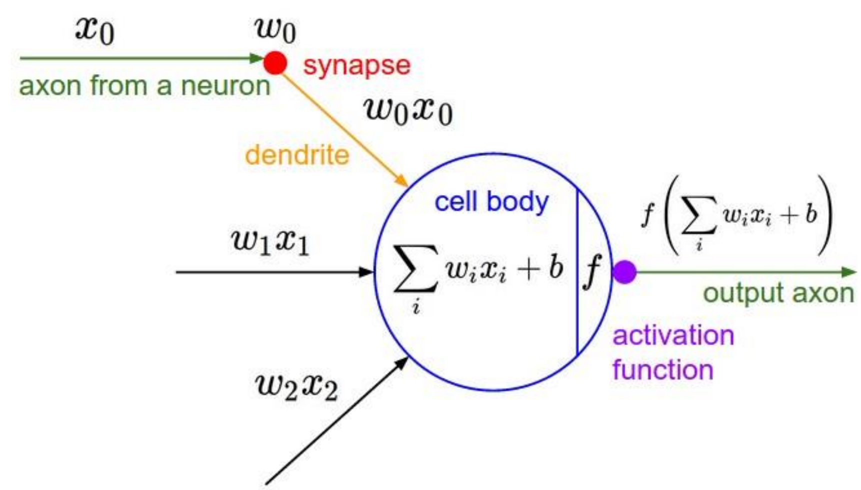
- Task: Given X predict y
- Given i.i.d. data $X_1, X_2, \dots, X_n$ and labels provided by an expert $y_1, y_2, \dots, y_n$
- Train a classifier f such that $f(X_i) \approx y_i$ (This is a mathematical optimization problem). Examples include neural networks, support vector machines.
- Test the model you have trained and use it in practice



Neural Networks

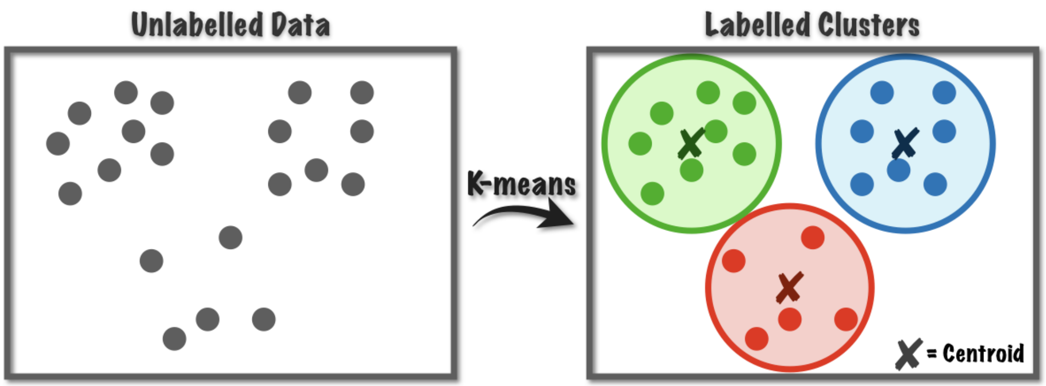
- **Neural Network (NN) Architecture**
  - Input x, Label $\tilde{y}$
  - Neural network can represent any arbitrary nonlinear mapping between x and y
  - Loss function: evaluate how far the prediction is from the expert label.

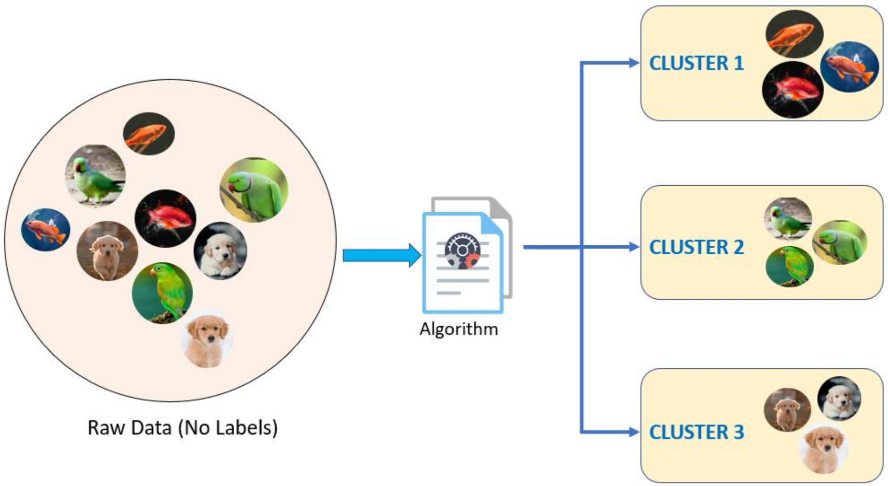




Unsupervised learning

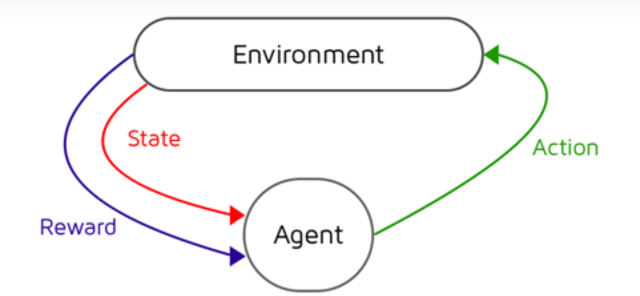
- Expert labels are unavailable. Usually used for identifying patterns from data
- Hidden Markov models, k-means, hierarchical clustering, and Gaussian mixture models

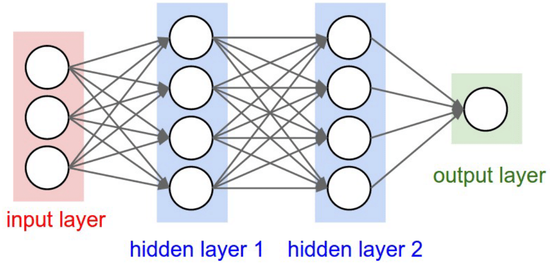


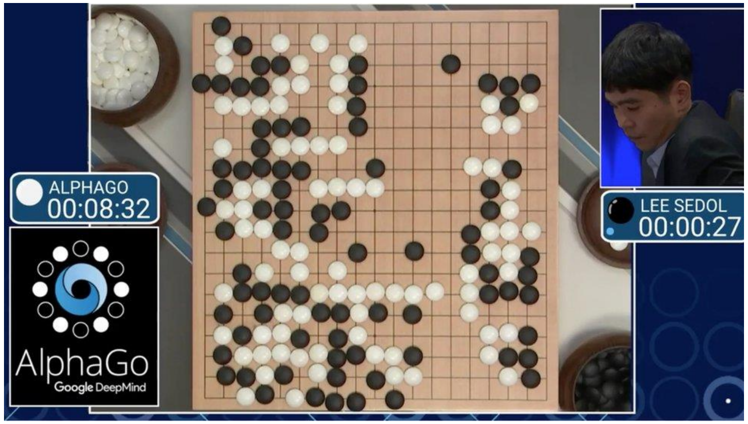


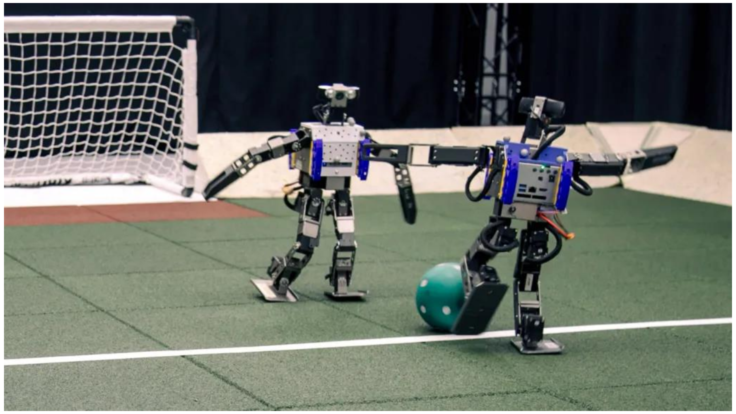
Reinforcement Learning

- Use AI to make decisions (take actions). Learning by doing.
- Exploration v.s. Exploitation. ϵ-greedy policy.



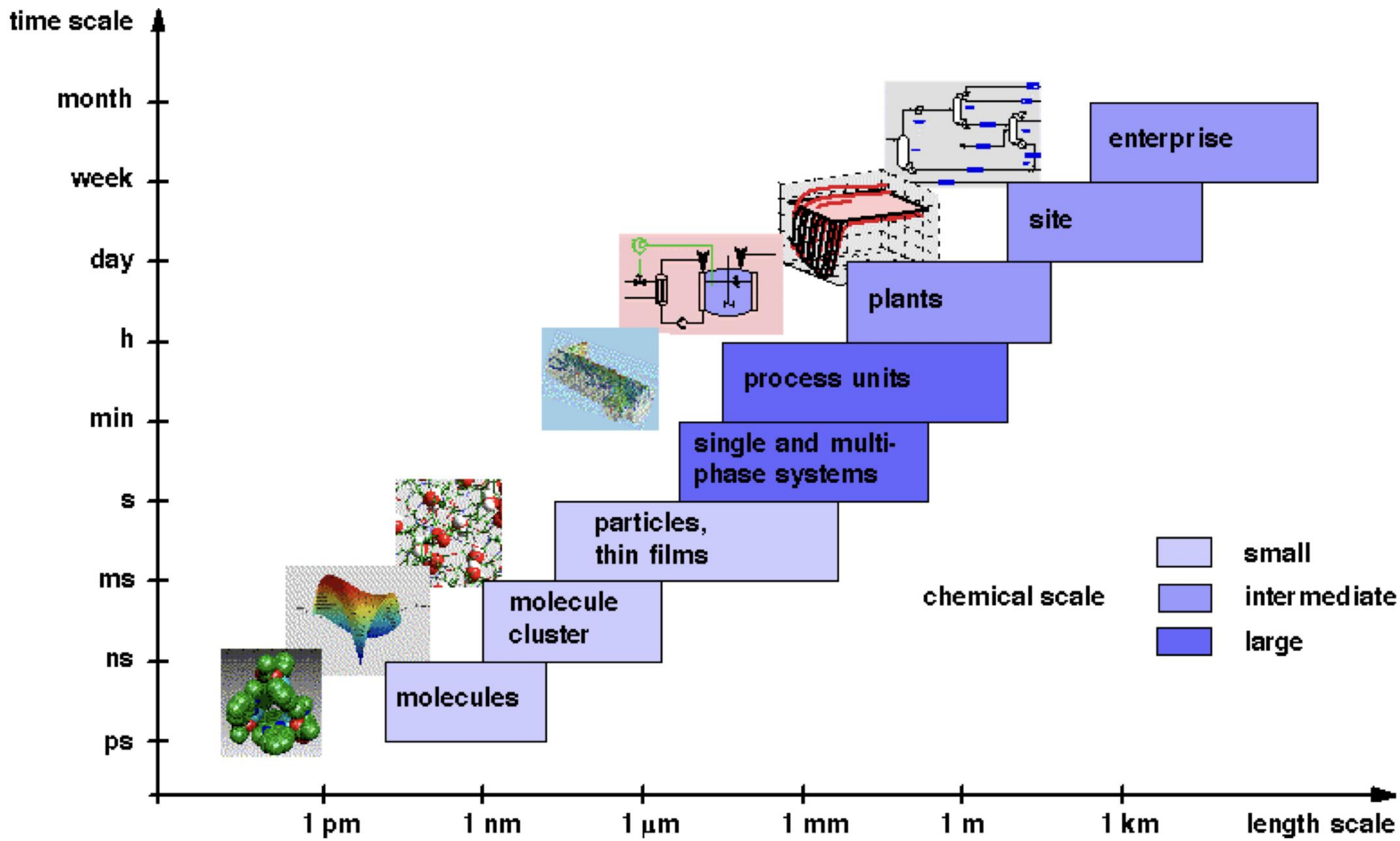






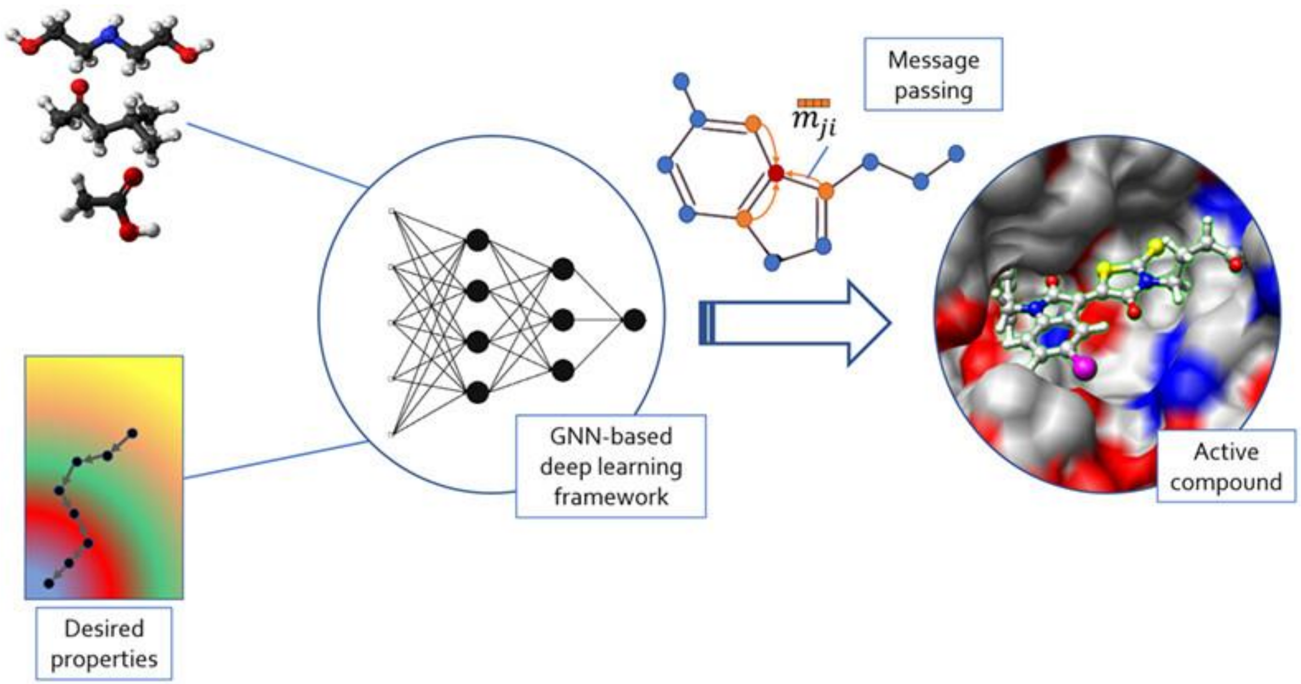
Existing Applications of AI in the Chemical Industry

Multi-Scale Applications



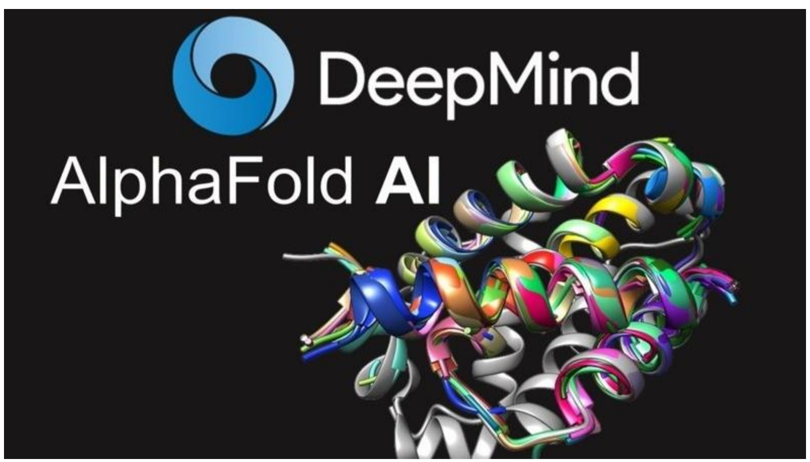
Predict Molecular Property

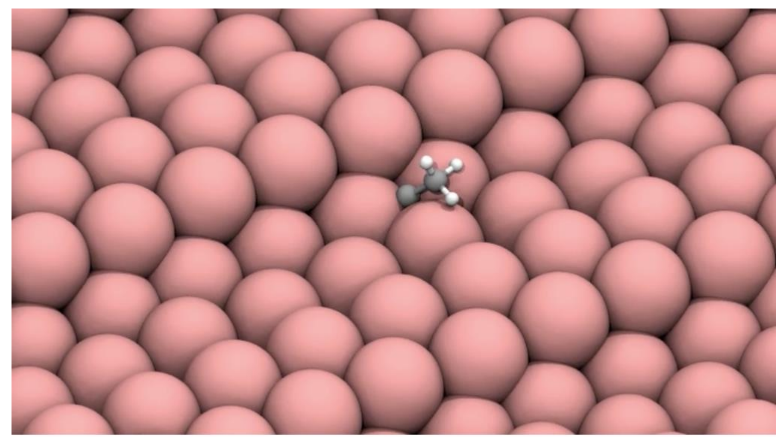
- Predict key molecular properties (e.g., solubility, toxicity, activity) from structure
- Accelerate materials and drug discovery through data-driven modeling
- Reduce need for costly experiments or simulations
- Enable inverse design of molecules with desired properties



Source: Abate et al. 2023

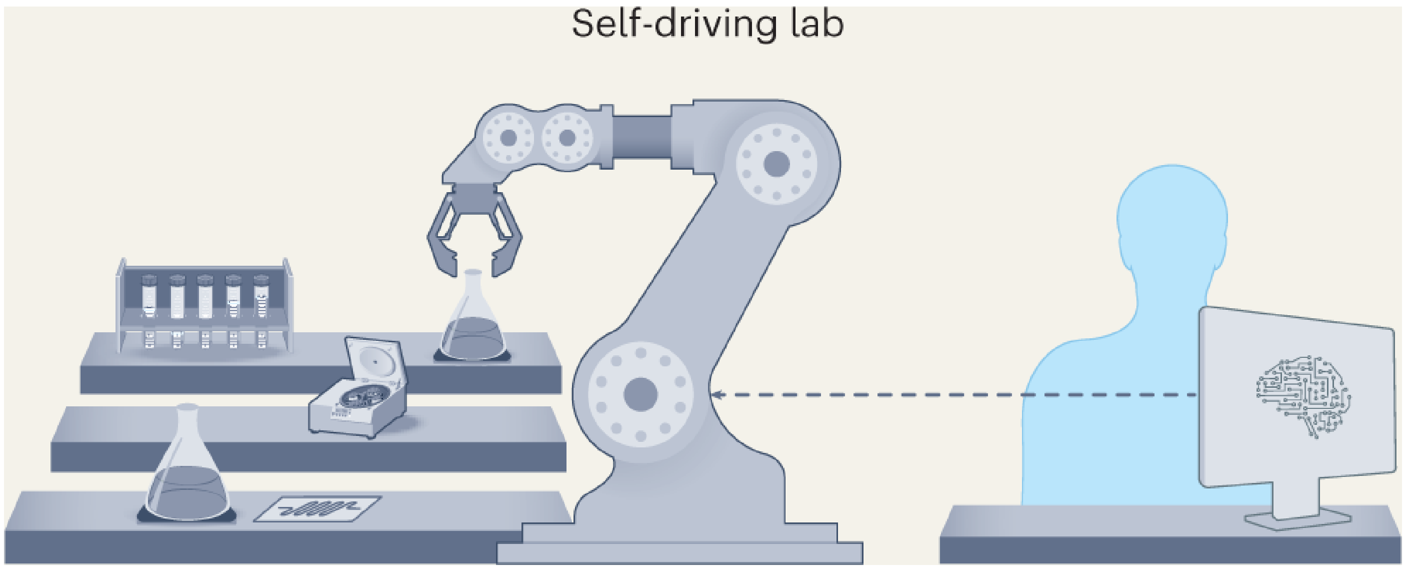
- AlphaFold: Nobel Prize in Chemistry (2024). Predict 3D protein structure from the sequence of amino acids. Replace expensive X-ray experiments
- Meta AI's Open Catalyst project: use AI for catalyst discovery. Replace DFT simulation and experiments.





Self-Driving Lab

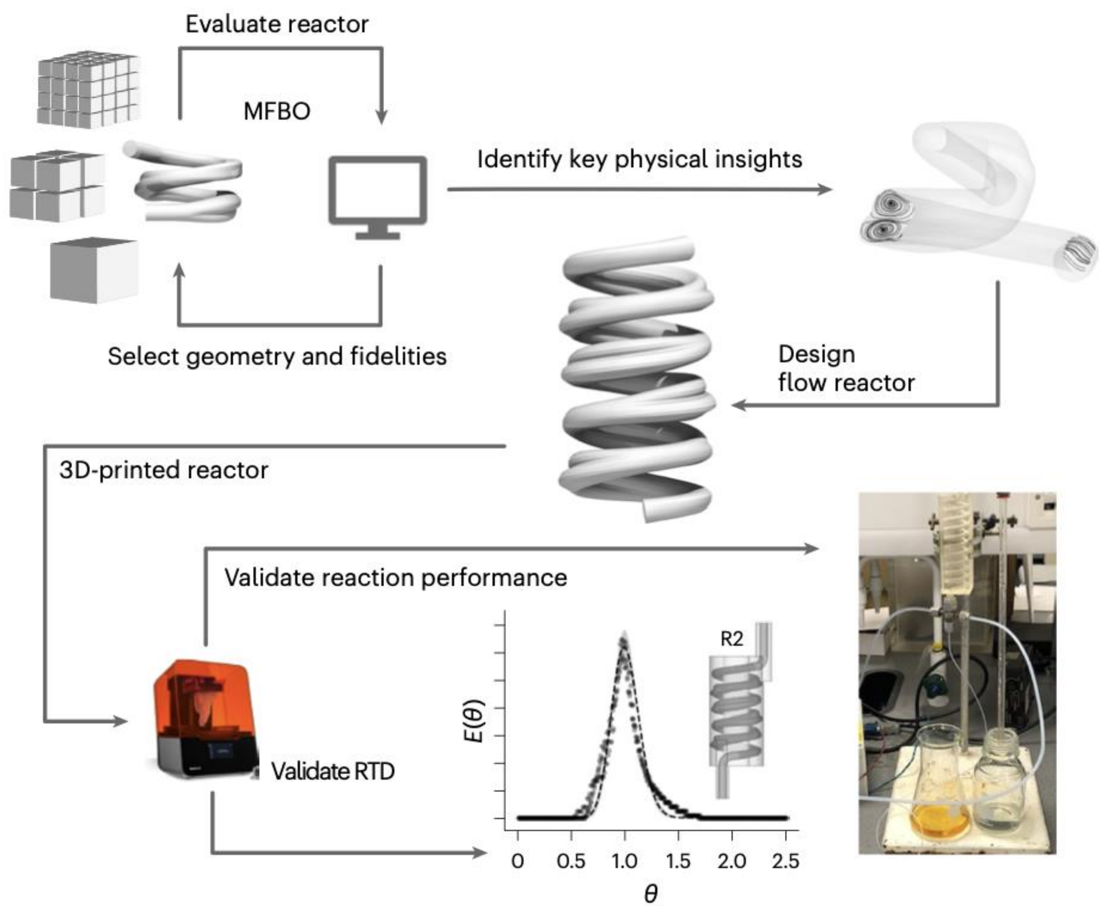
- Automate experimentation with AI-driven design–make–test–learn cycles
- Integrate robotics, machine learning, and real-time analytics for closed-loop discovery
- Optimize reaction conditions or material synthesis autonomously
- Accelerate discovery of catalysts, polymers, and battery materials
- Reduce human error and experimental cost while improving reproducibility



Source: Abolhasani & Kumacheva (2023)

Reactor Design

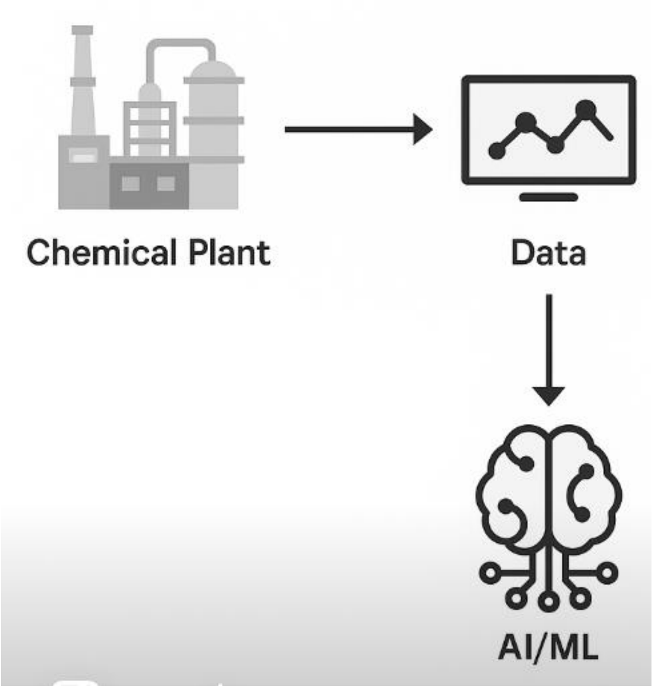
- Use Multi-Fidelity Bayesian Optimization (MFBO) to design efficient flow reactors
- Optimize reactor geometry by combining simulations and experimental data
- 3D-print and validate optimized designs through RTD and performance tests
- Iteratively refine models to uncover key physical insights and improve design

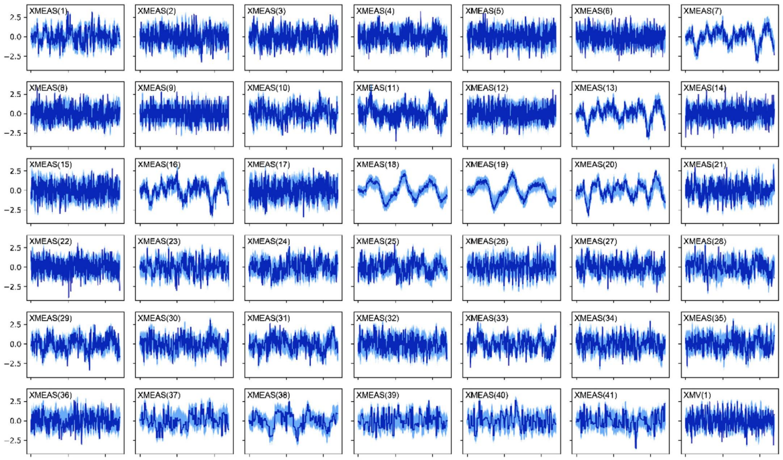


Source: Savage et al. 2024

Fault Detection and Diagnosis

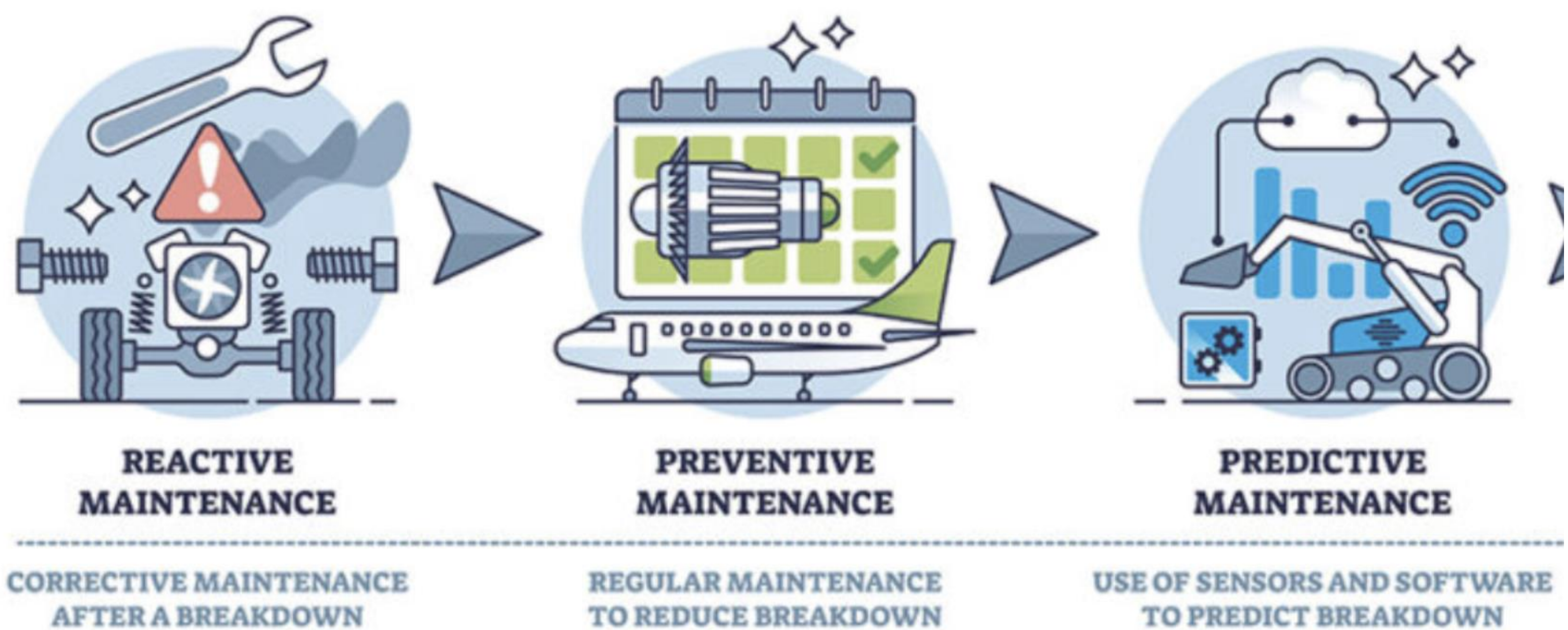
- Detect abnormal process behavior using sensor and operational data
- Diagnose root causes with machine learning and hybrid (physics + data) models
- Enable early intervention to prevent safety incidents and equipment damage





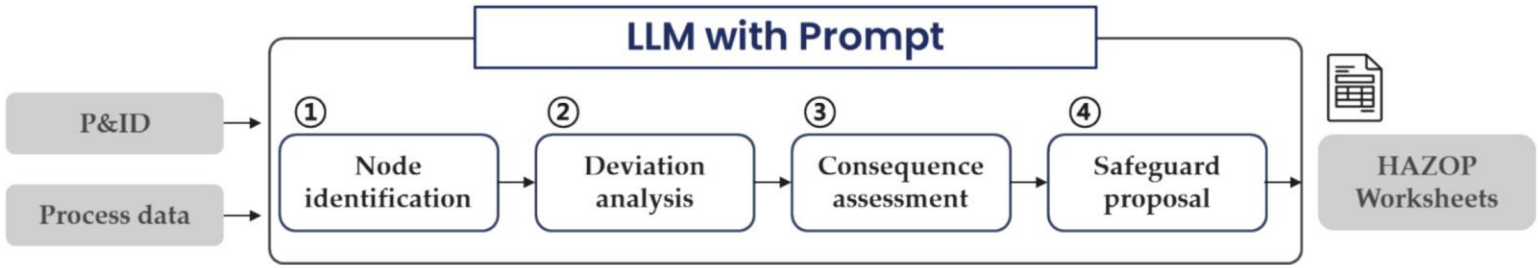
Predictive Maintenance

- **Reactive Maintenance:** Fix equipment only after a failure occurs — leads to unplanned downtime.
- **Preventive Maintenance:** Perform scheduled maintenance to avoid breakdowns — based on time or usage.
- **Predictive Maintenance:** Use sensors and AI to forecast failures before they happen — minimizes cost and disruption.



AI Assisted HAZOP Analysis

- Large Language Models (LLMs) can assist in automating the HAZOP (Hazard and Operability) study process.
- With proper prompts, LLMs can process P&IDs and process data to identify nodes, deviations, consequences, and safeguards.

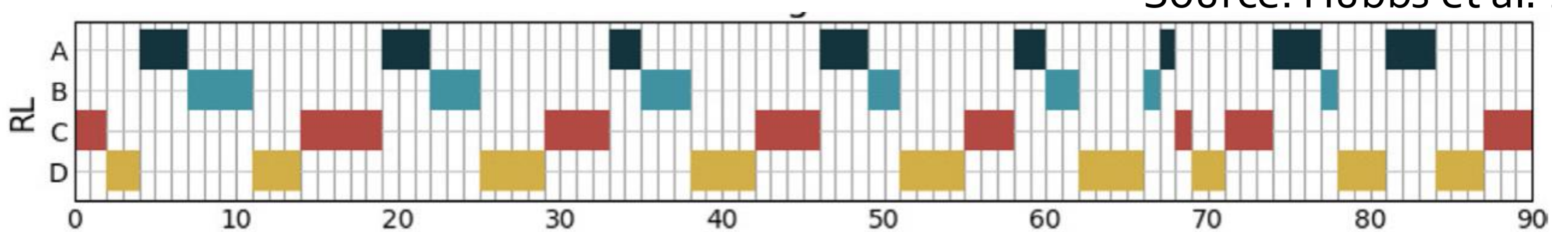


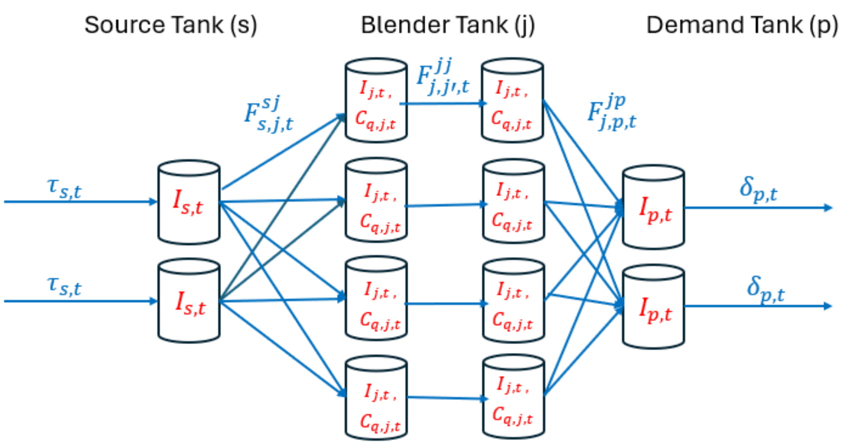
Lee et al. Can large language models automate the HAZOP process without human intervention?
See presentations by Ray Mentzer and students.

Scheduling of Chemical Production

- AI can automatically make scheduling decisions—such as which product to produce and for how long—in real time.
- It learns through trial and error, gradually improving its decisions even under changing or uncertain conditions, like fluctuating demand.

Source: Hubbs et al. 2020



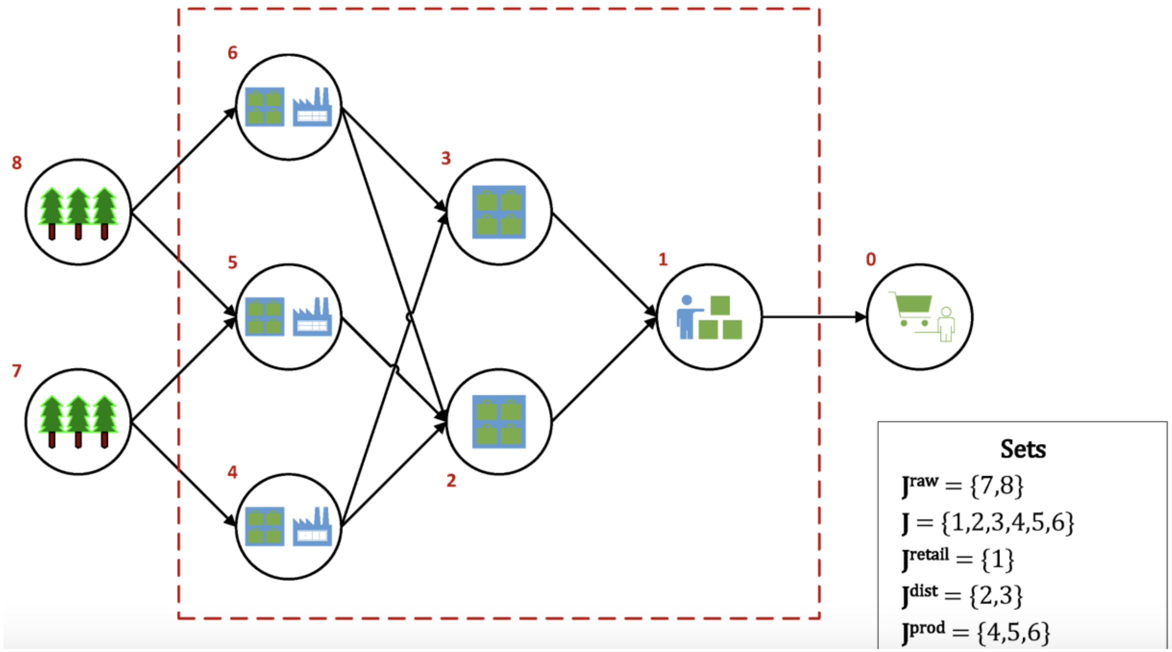


Our previous use reinforcement learning to schedule a reactor owned by Dow.

Ongoing work with ExxonMobil using AI for oil blending

Supply Chain Inventory Management

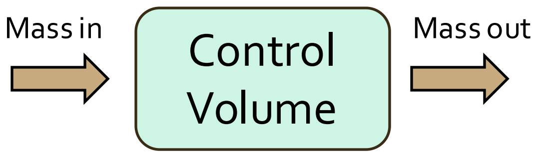
- Inventory management (IM) involves deciding how much to order, when to order, and from whom to keep products flowing efficiently.
- Key challenges include uncertain demand, variable lead times, and limited production or storage capacity.
- AI, especially reinforcement learning, can learn from experience to make smarter inventory decisions over time.
- AI can adapt to changes in demand or supply chain disruptions and find balanced policies that trade off profit, service level, and inventory cost.

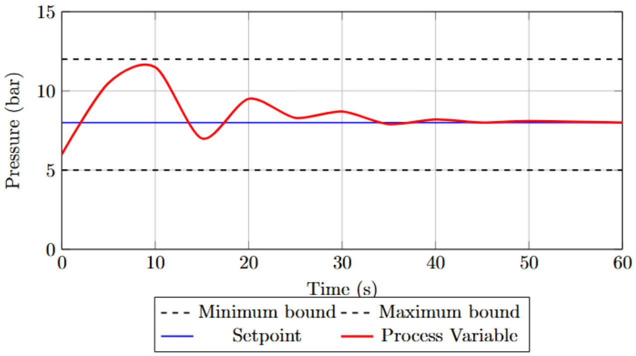


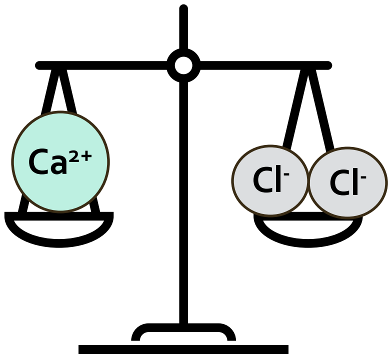
Limitations and Future Directions

Limitations of Existing AI Methods

- **Data efficiency**
  - Works well in a large data regime.
  - Scientific data is scarce.
  - Improve performance in a low data regime.
- **Hard constraints**
  - No hard rules in language or image texts.
  - Hard constraints in engineering domains, e.g.: conservation laws, operational bounds.

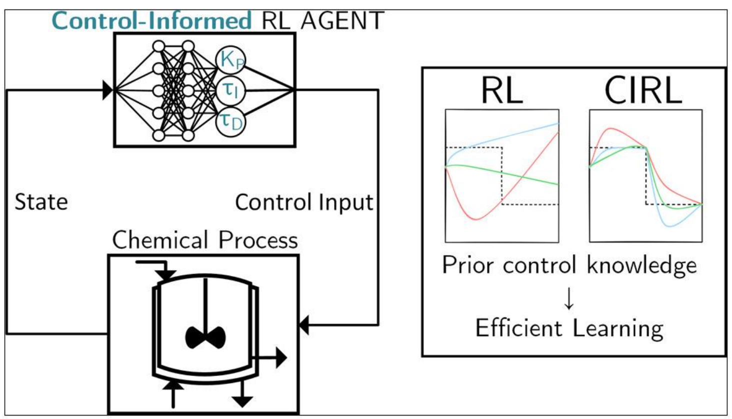


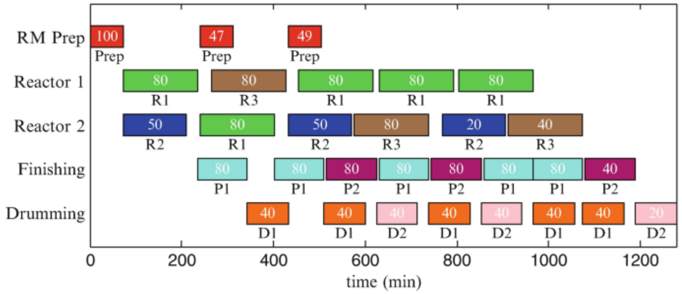




Limitations of Existing AI Methods

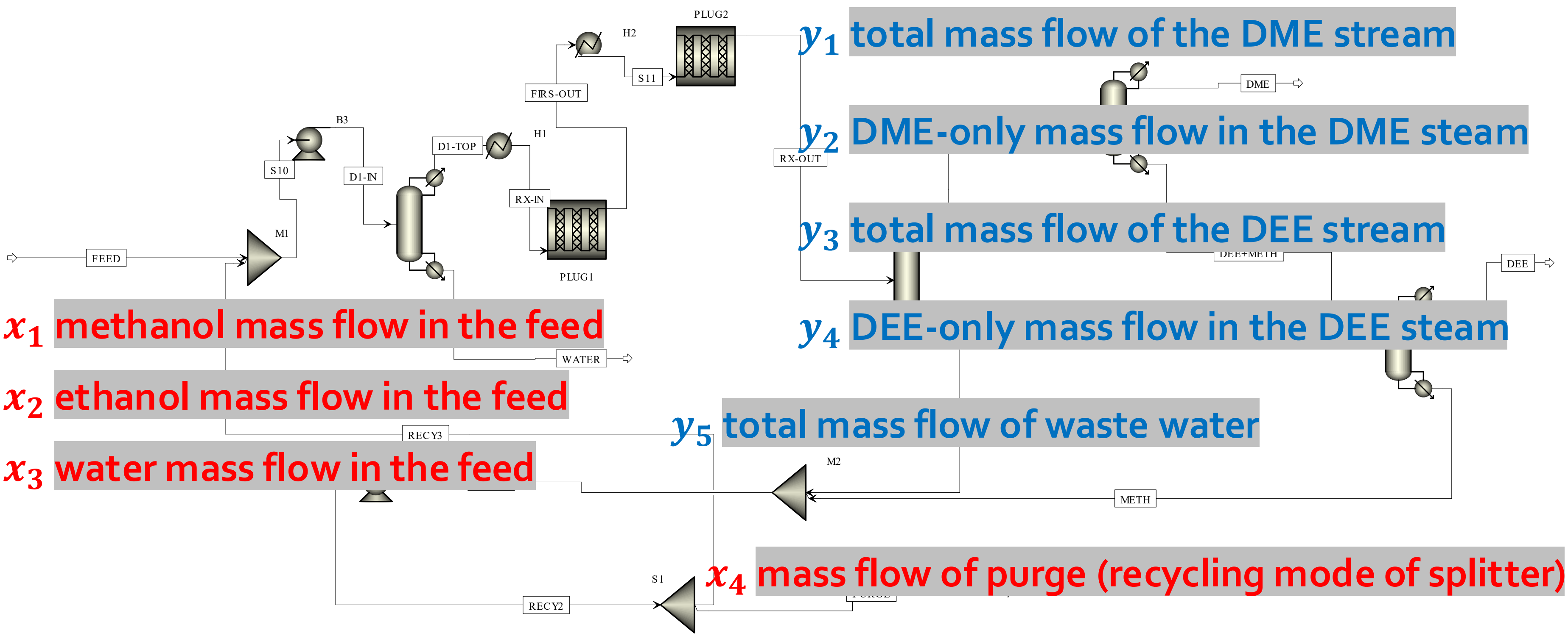
- **Interpretability**
  - Difficult to interpret. Lack of trust.
  - Why the model is predicting this?
- **Privacy**
  - Chemical engineering data are typically sensitive and confidential.
  - Equipment health status, drug properties, plant operating schedules





Satisfy Physical Constraints and Data Efficiency

Satisfy Mass Energy Blanace

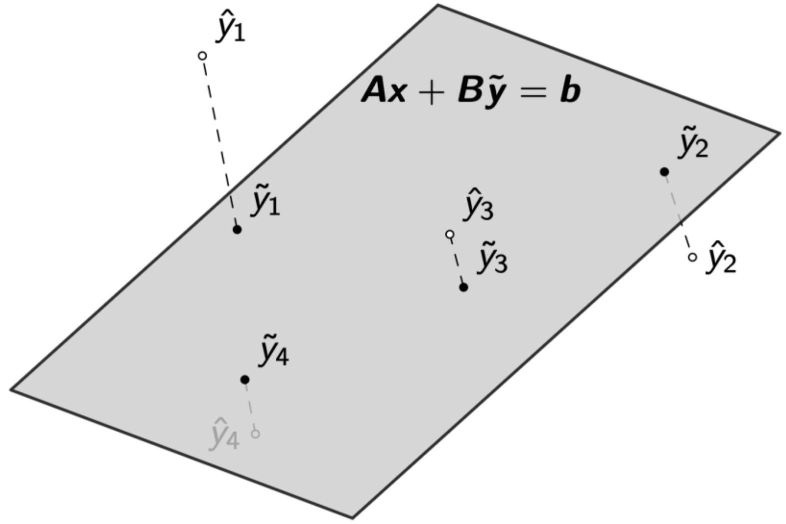


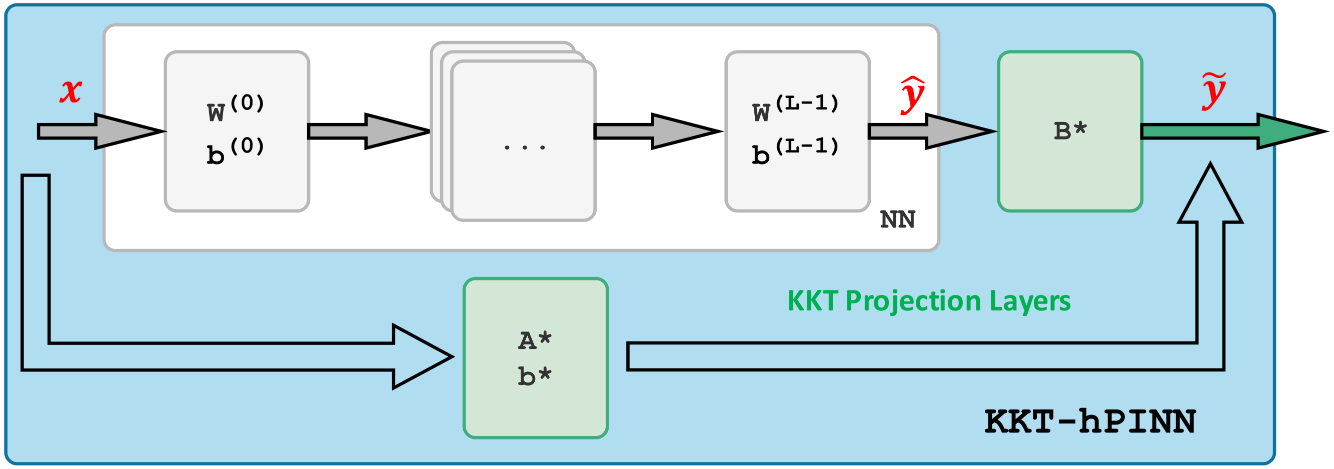
a chemical plant that uses methanol, ethanol, and water to manufacture dimethyl ether (DME) and diethyl ether (DEE)

- Surrogate model that predicts the outflow, given the inflow and recycling model
- Hard linear constraint embedded into KKT-hPINN: mass balance (flow-in = flow-out)

Orthogonal Projection

Geometric sense: **Orthogonal projection** of the old prediction, $\hat{\boldsymbol{y}}$, onto a hyperplane where $\mathbf{A}\mathbf{x} + \mathbf{B}\tilde{\boldsymbol{y}} = \boldsymbol{b}$ (linear equality constraints of interest are satisfied)



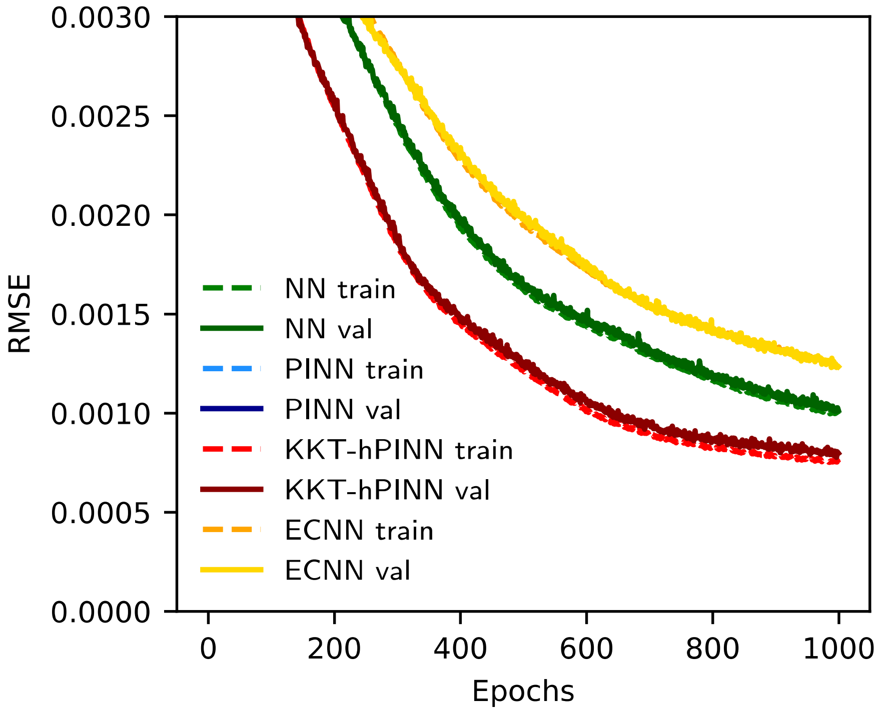


Model Performance: Learning Curves

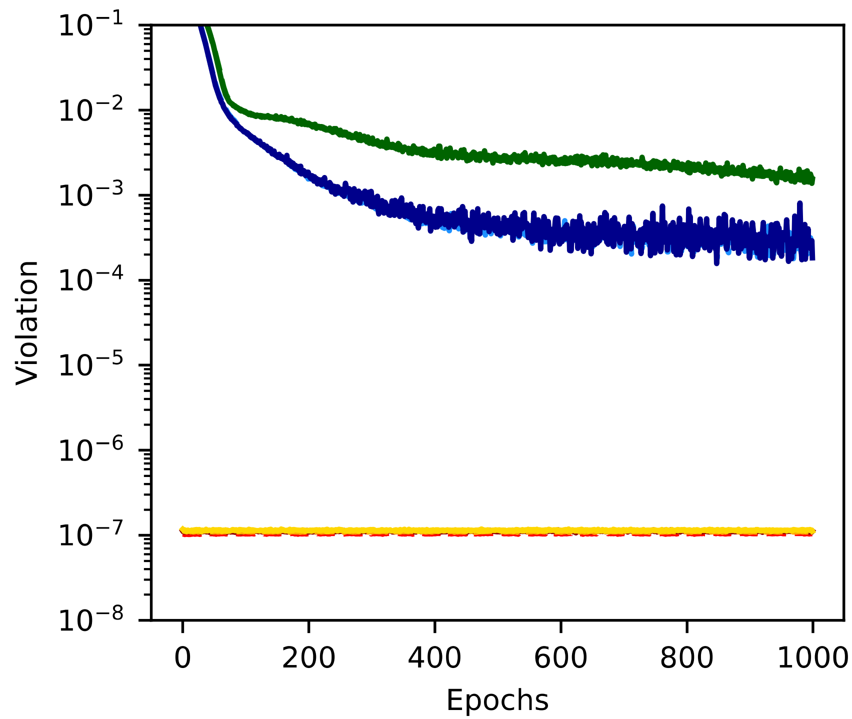
- ✓ **Higher accuracy:** outperforms neural network with a significant margin
- ✓ **Generalizability:** no overfitting and can be generalized well to unseen dataset
- ✓ **Inviolable hard constraints:** strictly satisfies the mass balance constraint

RMSE: root mean square error
between prediction and ground truth

Model performance on validation dataset

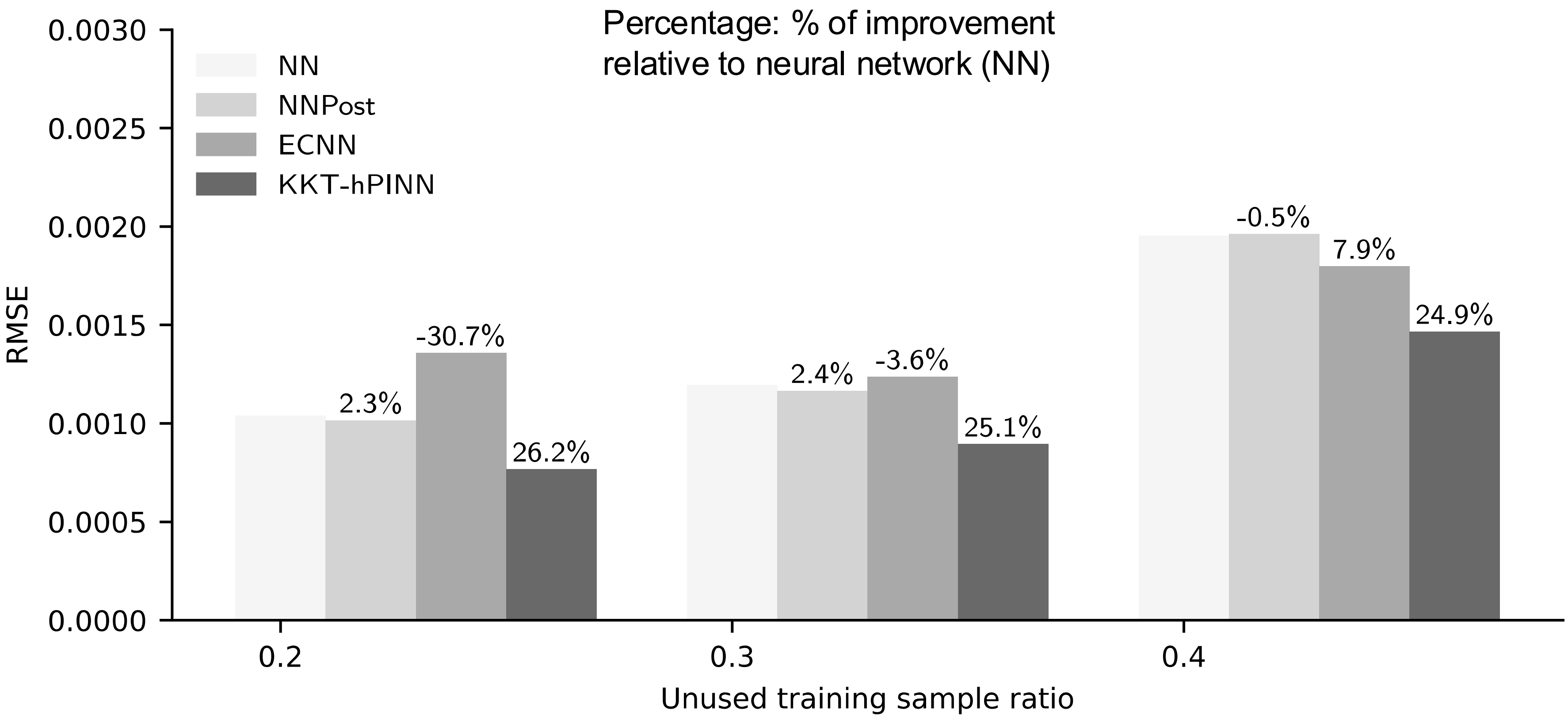


extent of mass balance violation
$|\text{inflow} - \text{outflow}|$



Model Performance vs. Training Samples

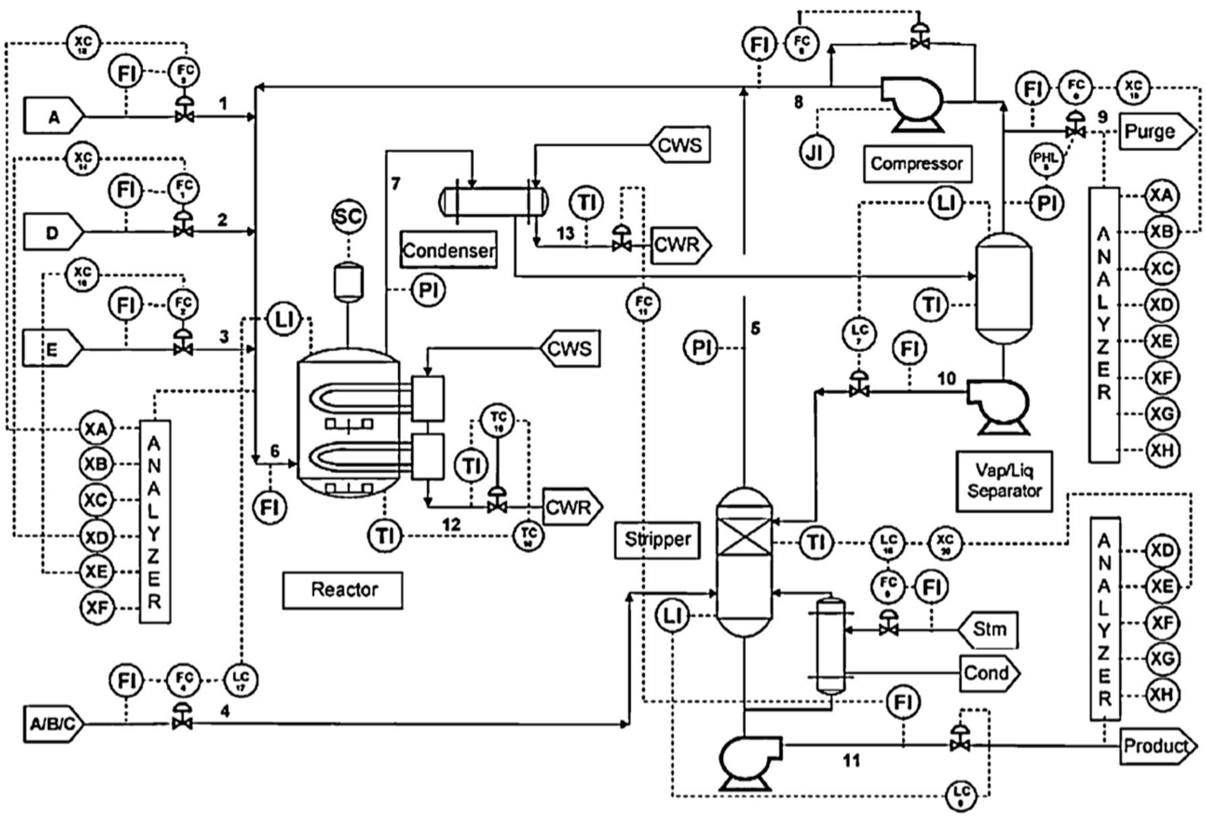
✓ **Minimal data requirement:** relatively good performance with less data



Improve Interpretability

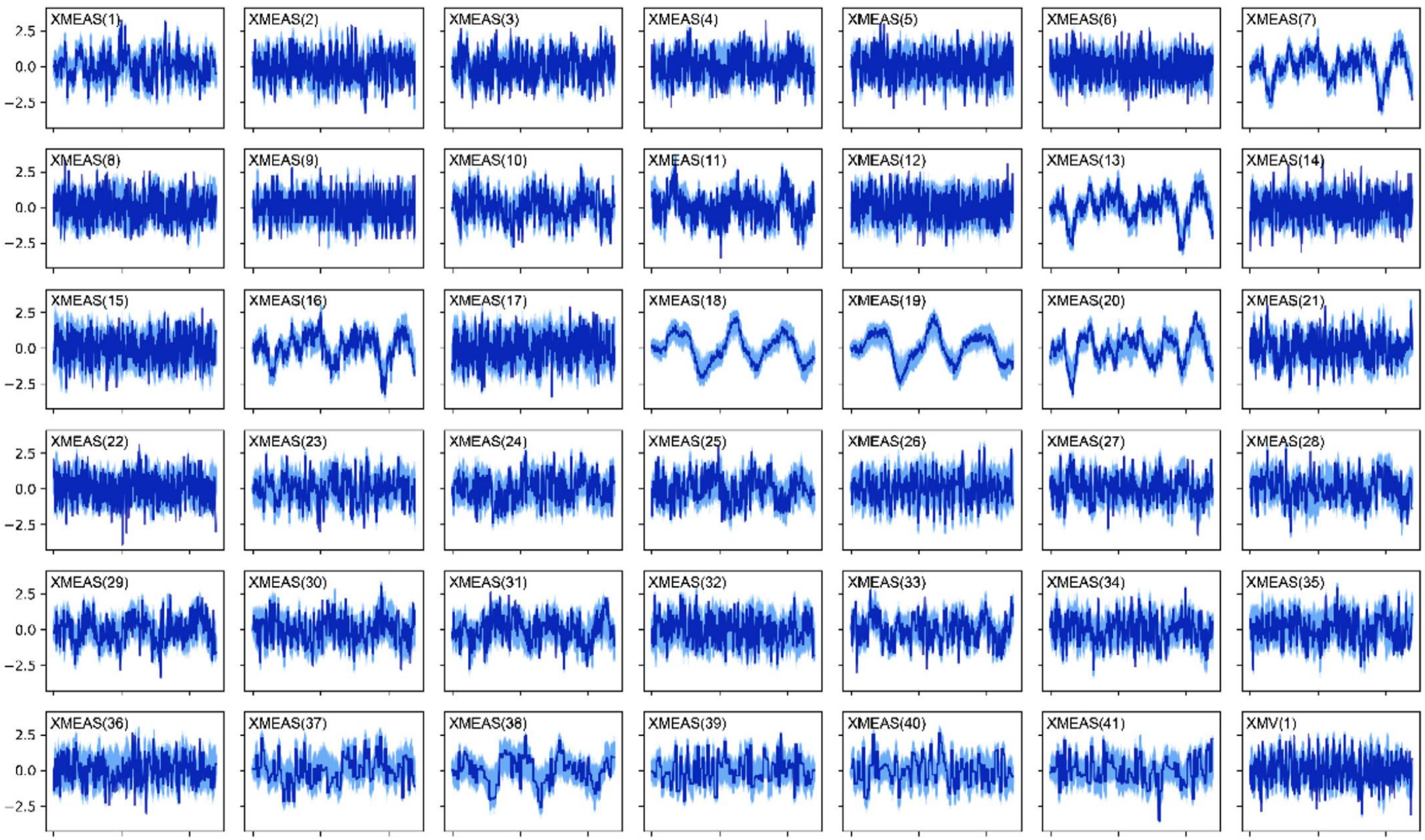
Tennessee Eastman Process

- 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

FaultExplainer Demo

FaultExplainer

Fault: Normal Operation

- Monitoring
- Assistant
- Fault History

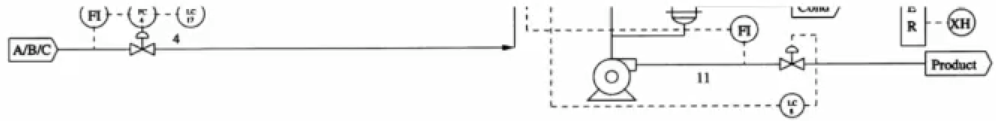


Fig. 8.1. A process flowsheet for the TEP with the second control structure in [205]

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) - G(liq): Product 1,

A(g) + C(g) + E(g) - H(liq): Product 2,

A(g) + E(g) - F(liq): Byproduct,

3D(g) - 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 ...

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

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

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

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

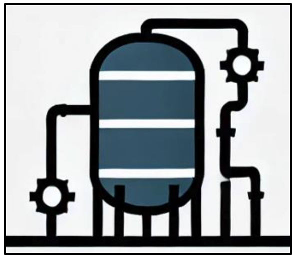
  GPT-4o: 6/11 o1: 9/11

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

Protect Privacy

Forecasting Maintenance

- Predictive maintenance of equipment



Air separation unit

Compressor in the plant

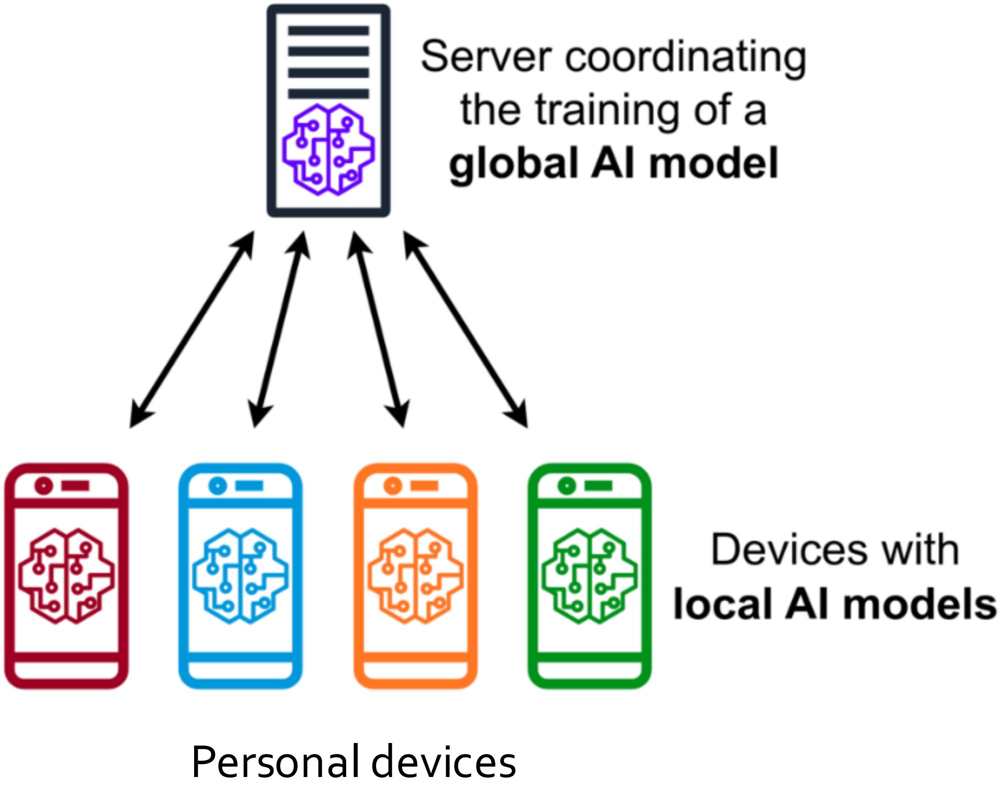
- Observed disruption in the plant!
- Plant aims to forecast such disruption using ML
- ***Challenge***: Insufficient data
- **Solution**: Federated Learning

Sensor

S. No.	Historical Data Points	State of Compressor
1	....	Healthy
2	....	Healthy
3	....	About to get Faulty
4	....	Faulty
	....	....

Federated Learning

- **Federated Learning**: Extensively used by Big-Tech companies to trainn prediction models, such as 'next letter that you will type'



Differential Privacy

- **Differential Privacy**: Technique to provide guarantees that private data won't be sacrificed
- Routinely used by Big-Tech companies and for high stakes government projects

APPLE

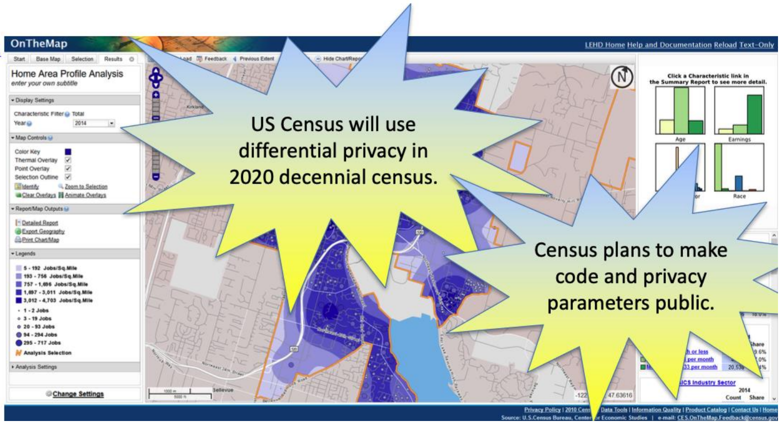
ANDY GREENBERG SECURITY 06.13.16 07:02 PM

APPLE'S 'DIFFERENTIAL PRIVACY' IS ABOUT COLLECTING YOUR DATA—BUT NOT *YOUR* DATA

*

U.S. CENSUS BUREAU

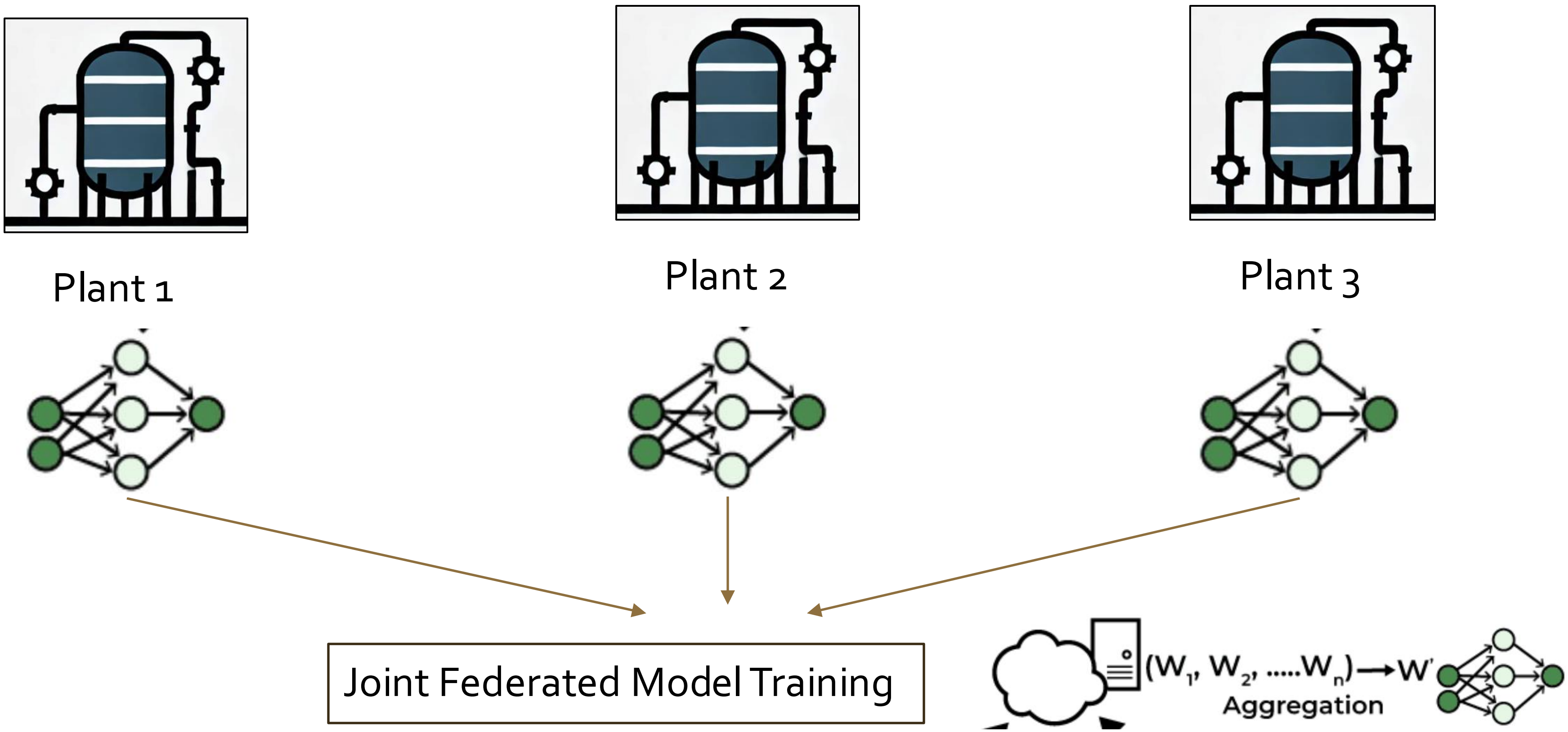
http://onthemap.ces.census.gov



2008 AD

Federated Forecasting Model

- Federated predictive maintenance of equipment



- Issues in joint training: Privacy of data points
- Solution: ***Differential privacy*** based → <u>Differentially private federated learning</u>

Acknowledgement

PhD students

Asha Ramanujam
(5^{th} year)

Hao Chen
(4^{th} year)

Sai Kompalli
(3^{rd} year)

Akshdeep Singh
(3^{rd} year)

Hsuan-Han Chiu
(2^{nd} year)

Shraman Pal
(2^{nd} year)

Piyush Sharma
(2^{nd} year)

MS students

Ashley Varghese

Quinn Huckaba

Sponsors

P2SAC Purdue Process Safety & Assurance Center
Air Liquide
ExxonMobil
ORACLE
amazon
U.S. National Science Foundation
American Cancer Society
NIH National Institutes of Health
Office of Naval Research Science & Technology
ACS Chemistry for Life

References

- Flores, G. E. C., Chen, H., & Li, C. Enforcing Hard Linear Constraints in Deep Learning Models with Decision Rules. In The Thirty-ninth Annual Conference on Neural Information Processing Systems. 2025
- Li, C. (2025). Breaking data silos in drug discovery with federated learning. Nature Chemical Engineering, 2(5), 288-289.
- 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.
- Chen, H., Flores, G. E. C., & Li, C. (2024). Physics-informed neural networks with hard linear equality constraints. Computers & Chemical Engineering, 189, 108764.

Quiz

If I want to train a machine learning model to play the game of go, or robot to play soccer. Which of the following machine learning method should I use

A. Supervised learning
B. Unsupervised learning
C. Reinforcement Learning

Correct answer: C