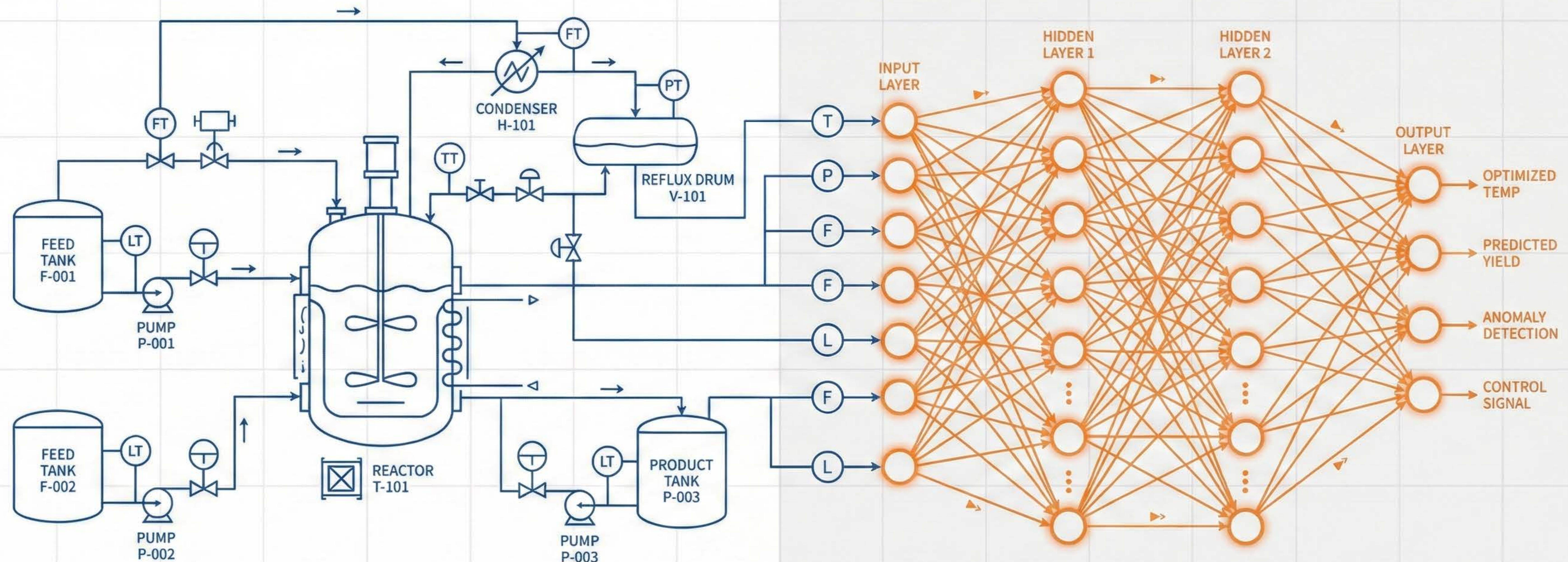


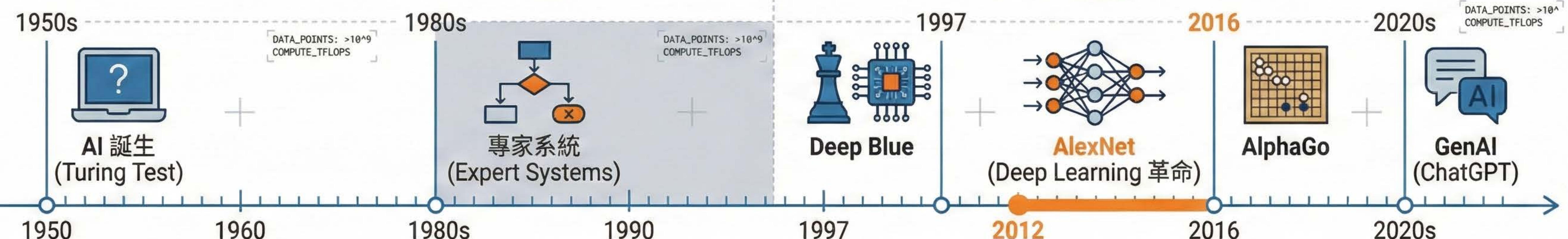
# AI 在化工上的應用：從原理到實踐

## Unit 01 課程概論與導讀

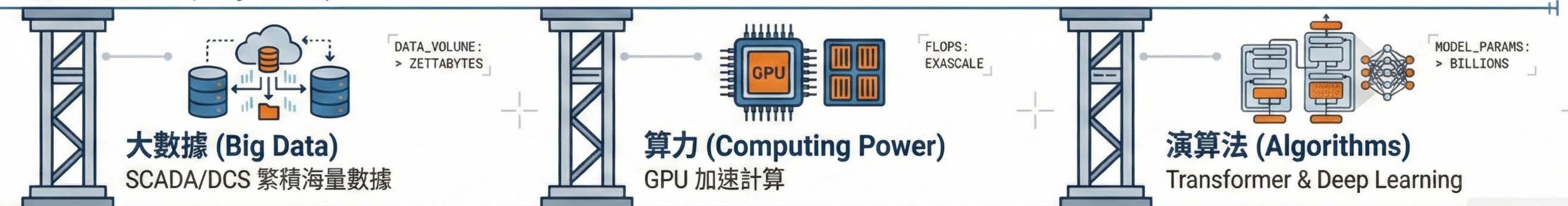


課程目標：建立人工智慧與機器學習基礎，解決化工  
領域實際問題

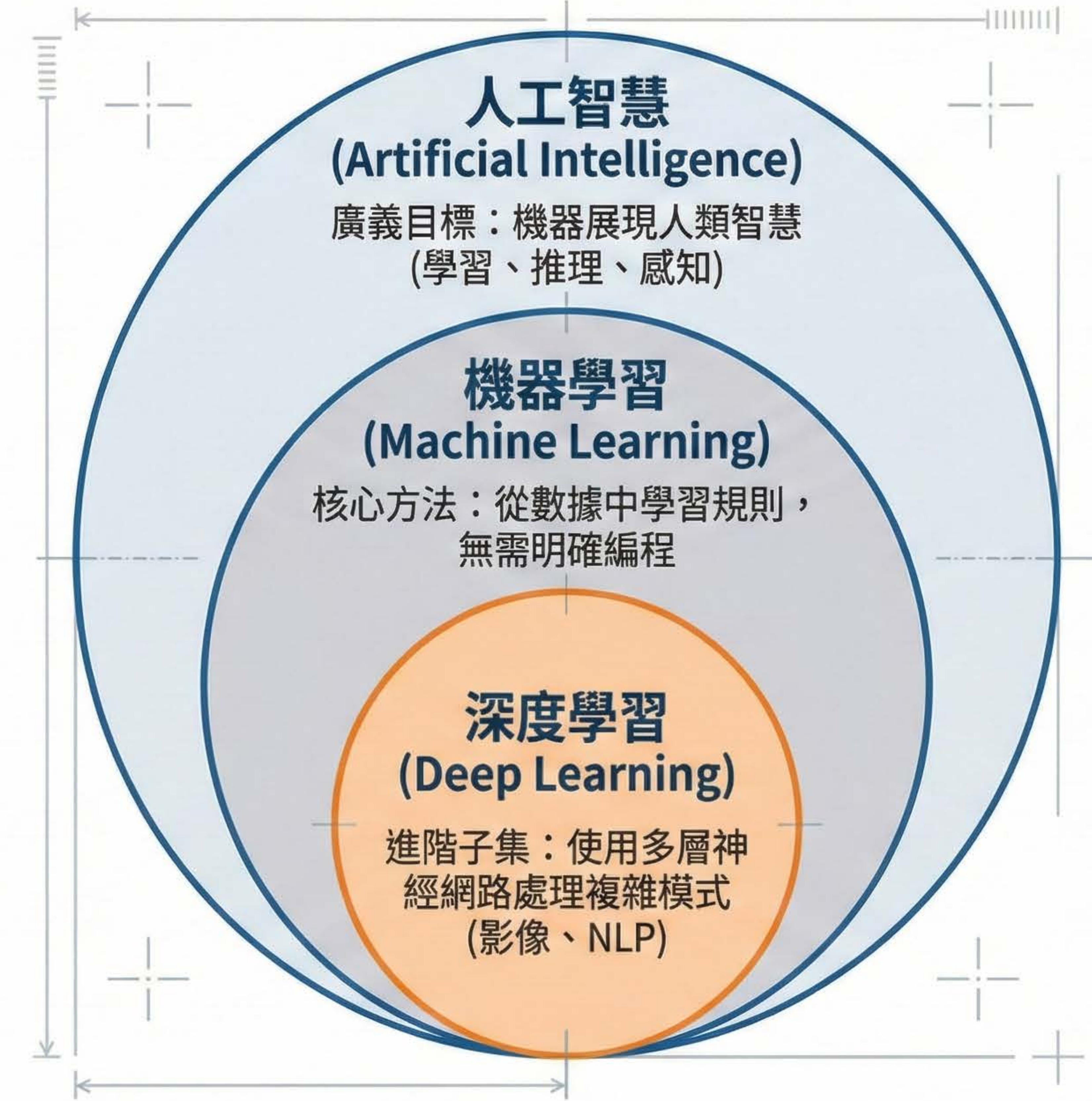
# 化工工程師的演進：從經驗法則到數據驅動



為何是現在？(Why Now?)



# 解密 AI 術語：理解層次結構



## 關鍵概念 (Key Concept)

**AI** 是大傘，**ML** 是方法，**DL** 是工具。

在此課程中，我們主要聚焦於 **Machine Learning (ML)** 技術在化工數據上的應用。

# 機器學習地圖：如何選擇正確的工具？

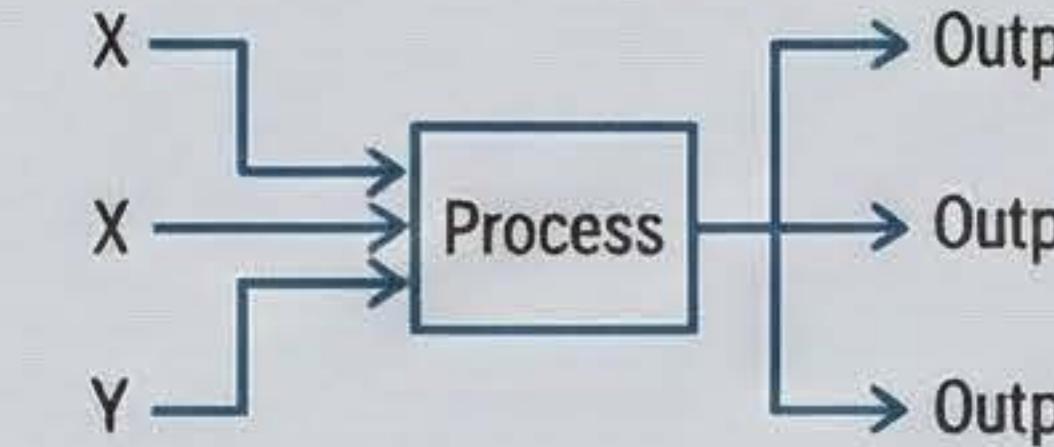
## 監督式學習 (Supervised Learning)



**數據 (Data) :**  
有標籤數據 (Input X + Output y)

**目標 (Goal) :**  
預測 (Prediction)

**化工案例 (ChemE Examples) :**  
品質預測 (Quality Prediction)、  
故障分類 (Fault Classification)



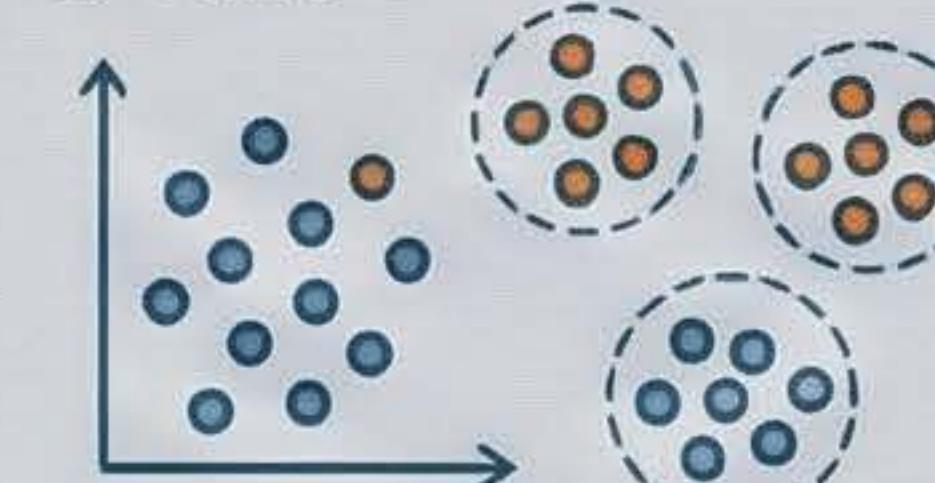
## 非監督式學習 (Unsupervised Learning)



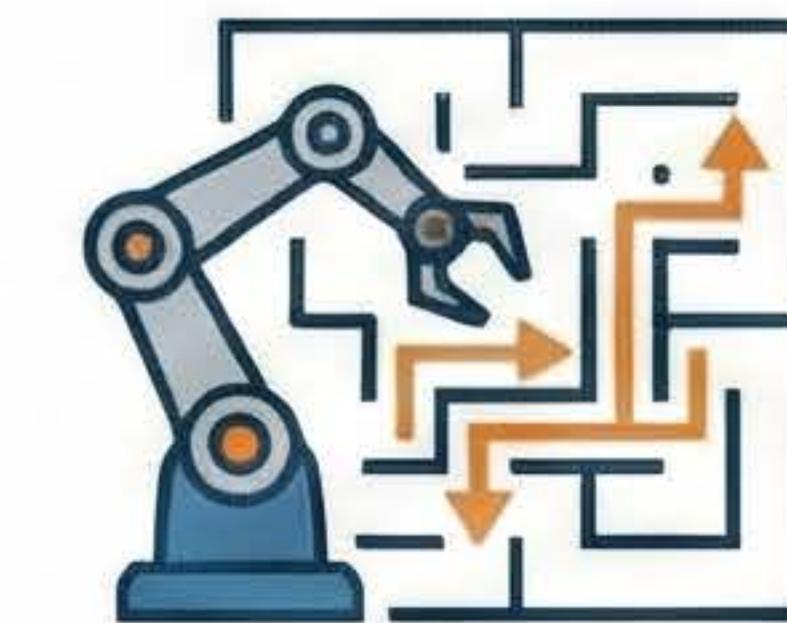
**數據 (Data) :**  
無標籤數據 (Input X only)

**目標 (Goal) :**  
發現模式 (Discovery)

**化工案例 (ChemE Examples) :**  
異常檢測 (Anomaly Detection)、  
操作模式識別



## 強化學習 (Reinforcement Learning)



**數據 (Data) :**  
環境互動 (Action/Reward)

**目標 (Goal) :**  
決策優化 (Optimization)

**化工案例 (ChemE Examples) :**  
製程控制策略 (Process Control)



不同的化工問題需要對應不同的演算法類型。

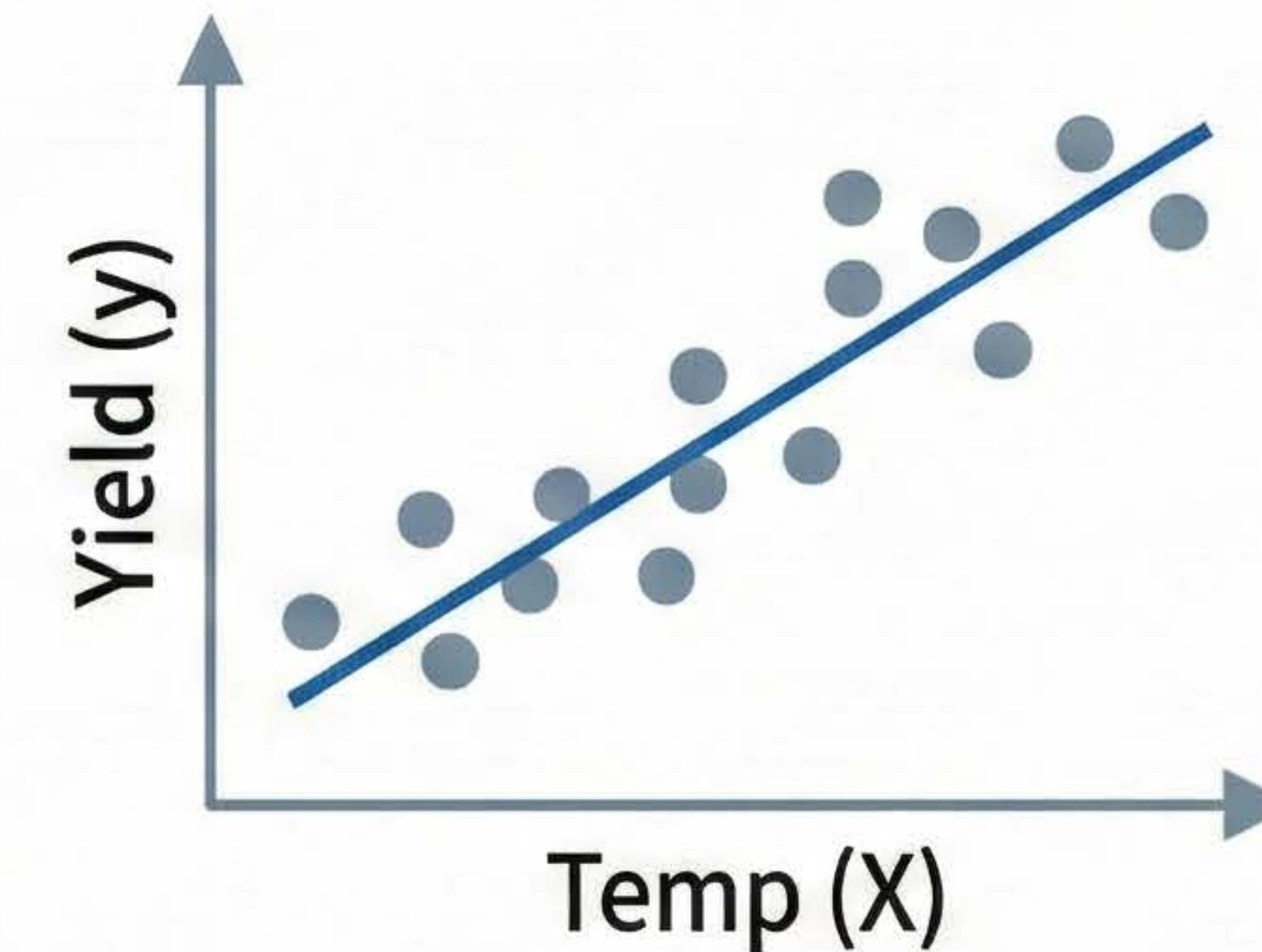
AI

# 監督式學習：預測與分類

核心任務：學習函數  $f(X) \rightarrow y$

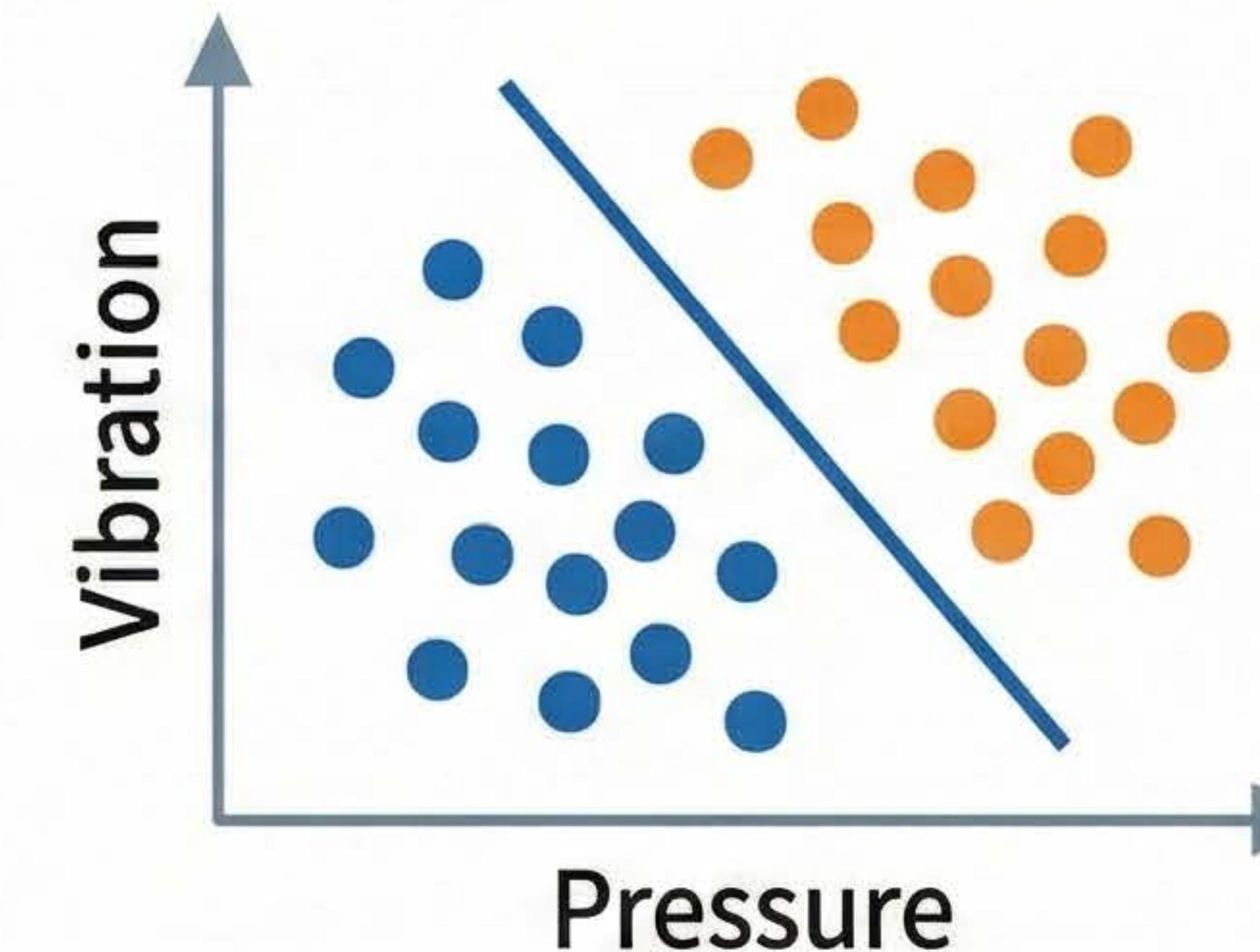
## 回歸 (Regression)

預測連續數值 (Predict Continuous Values)



## 分類 (Classification)

預測離散類別 (Predict Categories)



### 化工應用：軟感測器 (Soft Sensors)

利用容易量測的變數 (如溫度、壓力) 預測難以量測的指標 (如黏度、純度)。

算法概數：

Linear Regression, Random Forest, SVR

### 化工應用：故障診斷 (Fault Diagnosis)

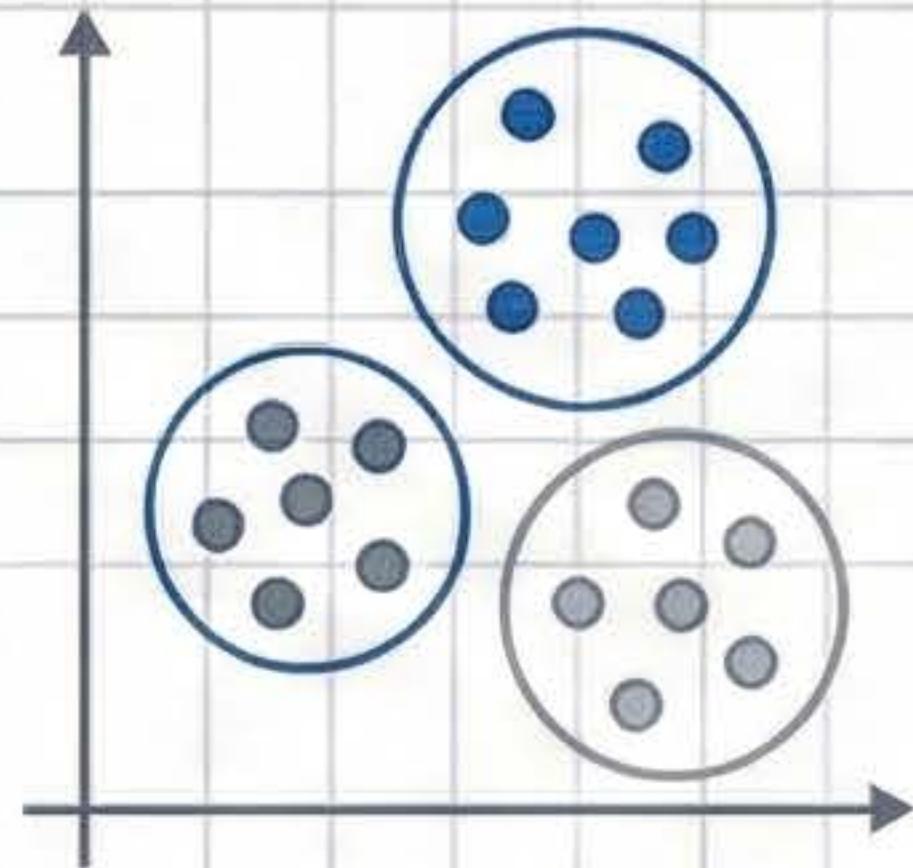
判斷設備狀態：[正常] vs [警告] vs [故障]。

算法概數：

Logistic Regression, SVM, XGBoost

# 非監督式學習：從數據中發現隱藏模式

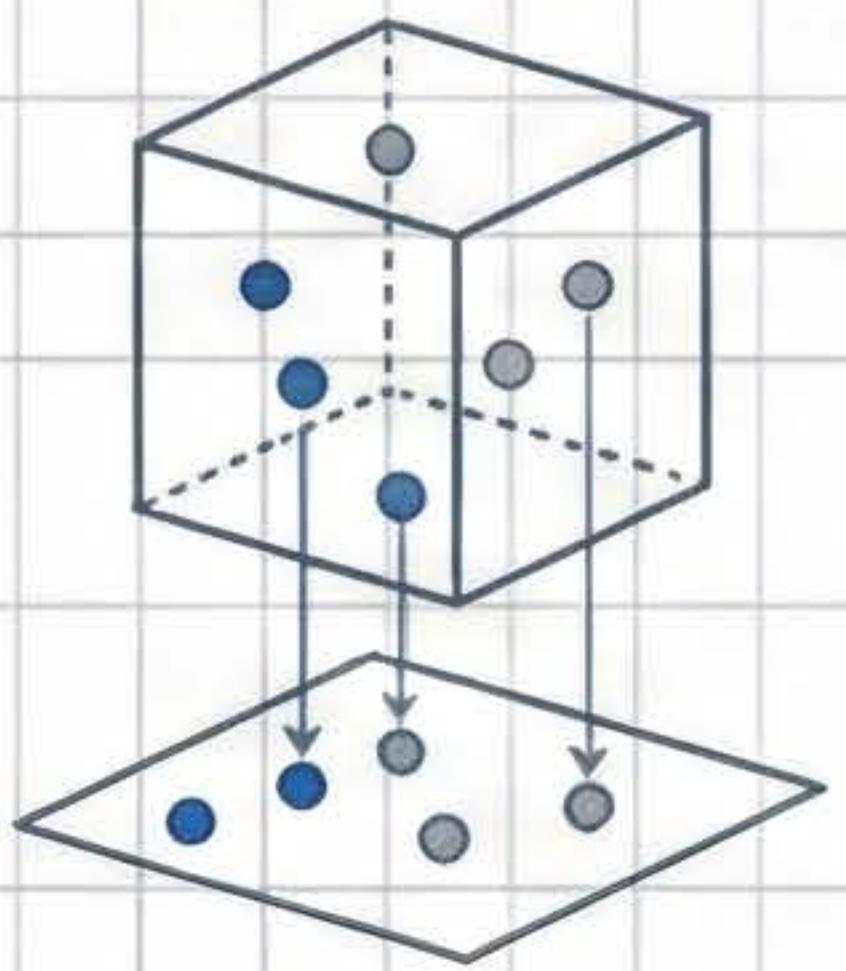
在沒有標準答案的情況下尋找結構 (Input X only)



## 分群 (Clustering)

將相似的數據點分組。

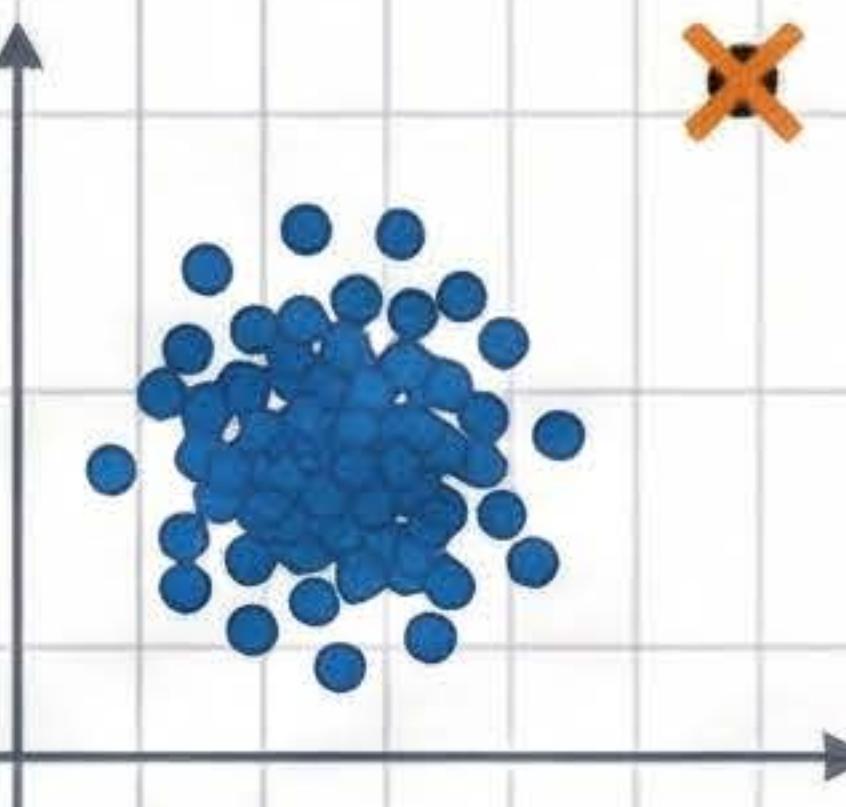
化工案例：**操作模式識別** - 找出反應器的不同穩定狀態 (Steady States)。



## 降維 (Dimensionality Reduction)

壓縮資訊，保留關鍵特徵。

化工案例：**製程視覺化** - 將成百上千個感測器數據降維至 2D 圖表以便監控 (PCA)。



## 異常檢測 (Anomaly Detection)

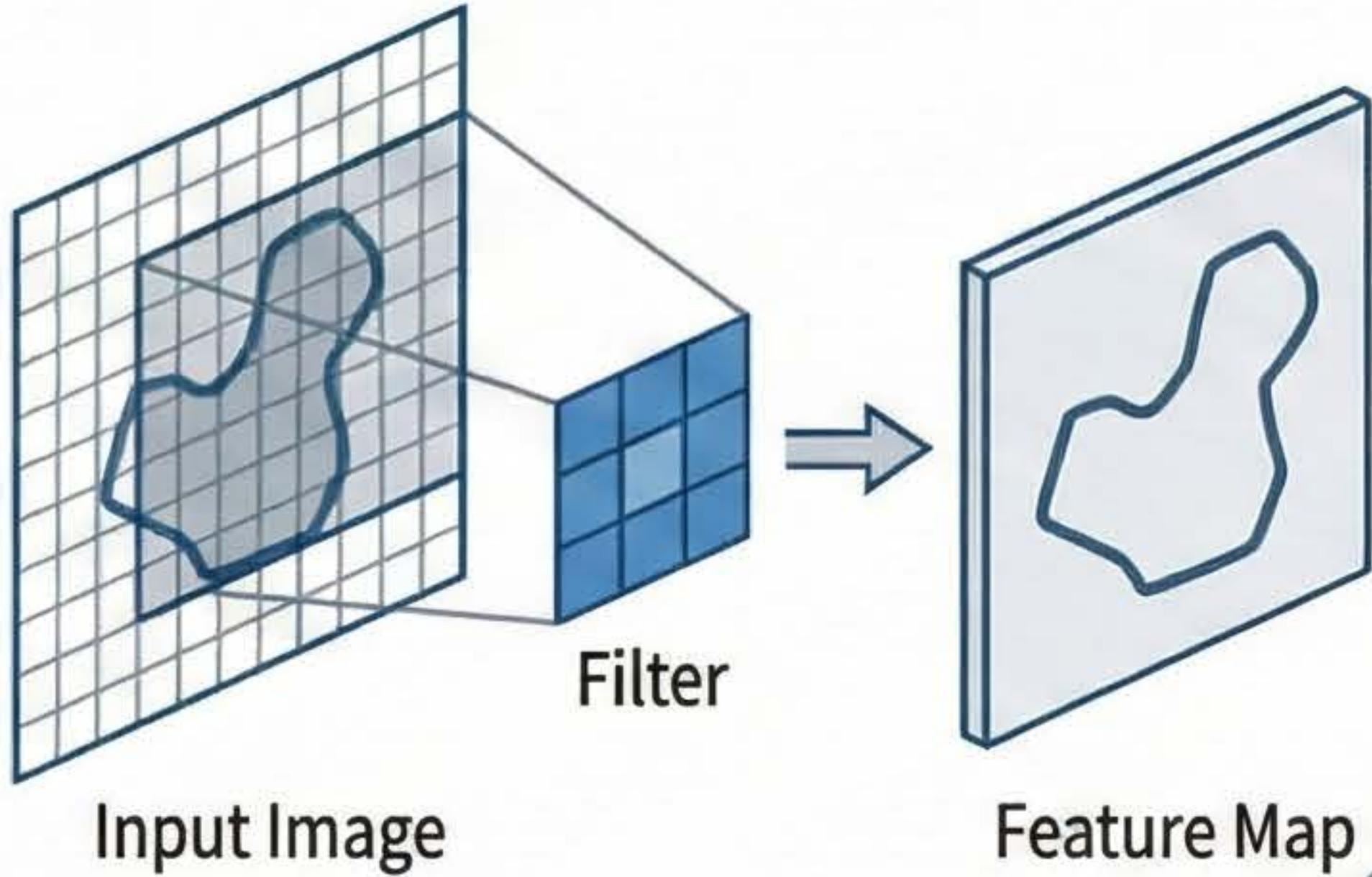
識別不屬於正常分佈的孤立點。

化工案例：**早期預警系統** - 在安全界限被突破前偵測到製程偏移。

# 進階前沿：深度學習與生成式 AI

處理非結構化數據 (影像、序列、分子結構) 的強大工具。

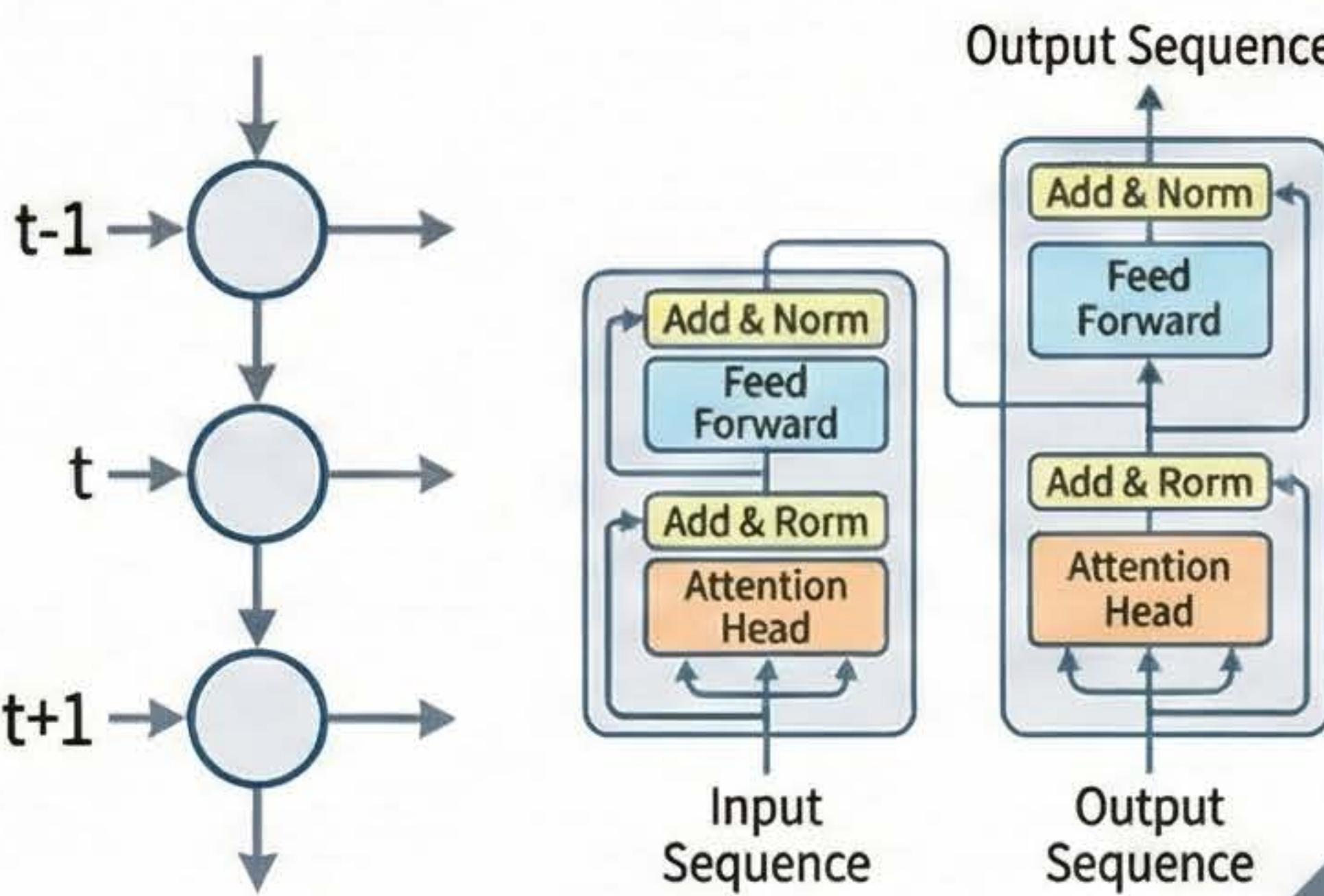
## CNN (卷積神經網路)



**數據類型：**影像數據 (Image)

**應用：**晶體形態分析 - 從顯微鏡影像中自動識別晶體大小與形狀。

## RNN / Transformer



**數據類型：**序列數據

(Time-series / Sequence)

**應用：**動態建模 - 預測反應器隨時間的動態變化或是SMILES分子字串。

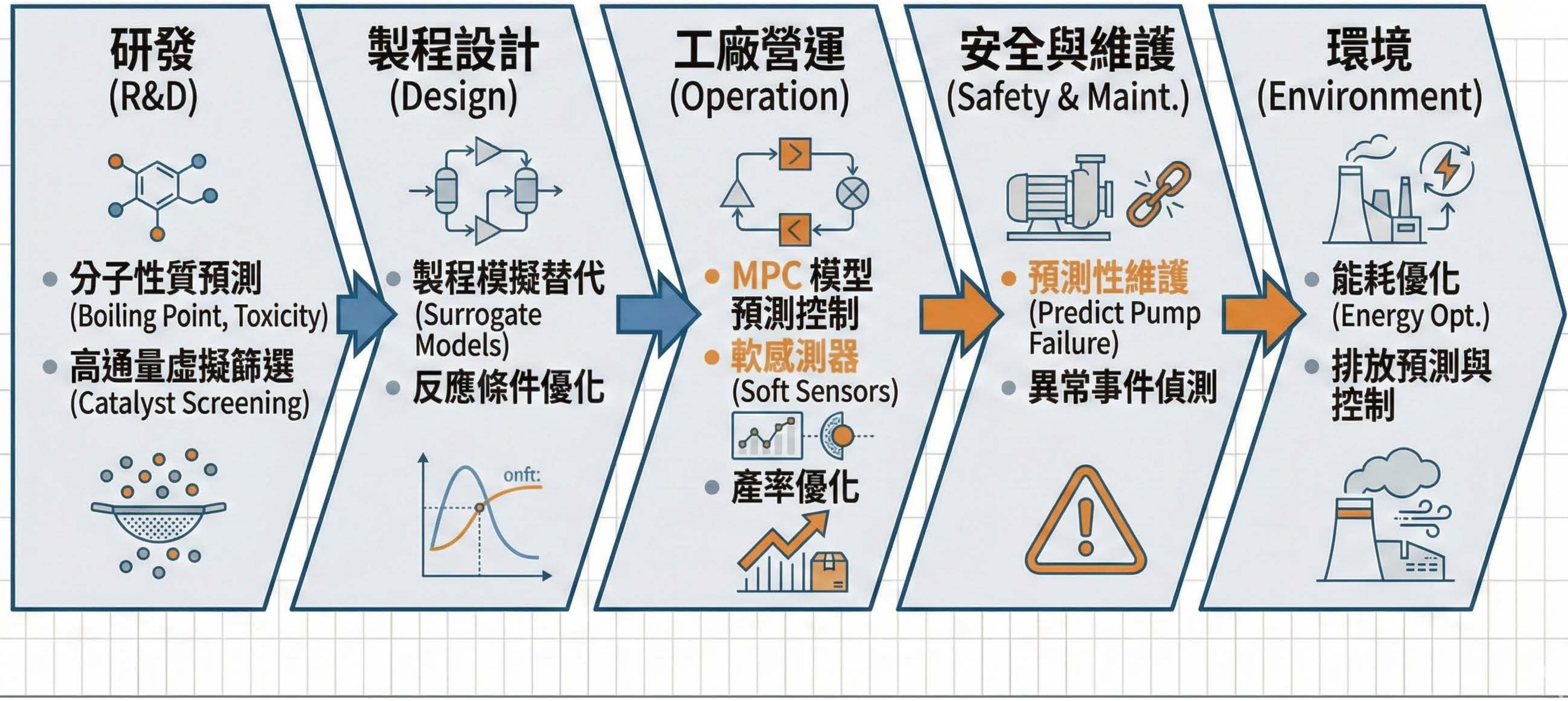
## Generative AI (生成式 AI)



**數據類型：**創造新數據

**應用：**逆向設計 (Inverse Design) - 根據目標性質 (如高溶解度) 自動生成新的分子結構或配方。

# AI 在化工全生命週期的價值

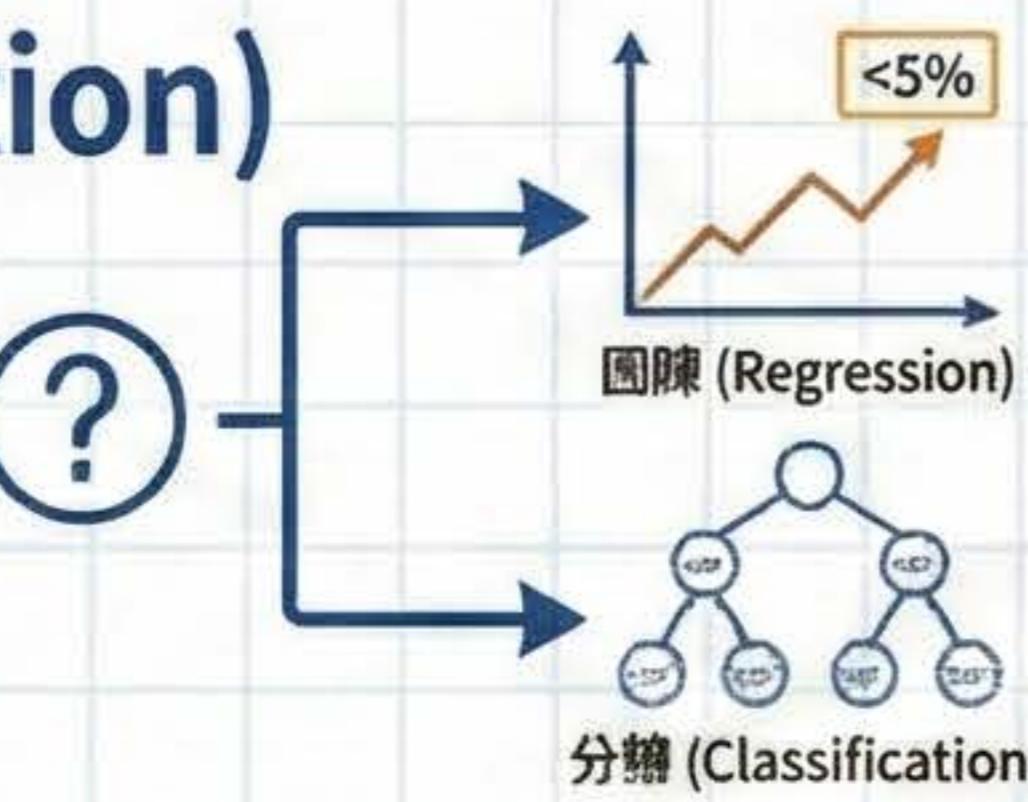


# 機器學習工作流程 Part 1：數據準備

## ⚠ 原則：Garbage In, Garbage Out (垃圾進，垃圾出)

### 1. 問題定義 (Problem Definition)

定義成功指標。這是回歸問題還是分類問題？(例如：預測分子量誤差  $< 5\%$ )



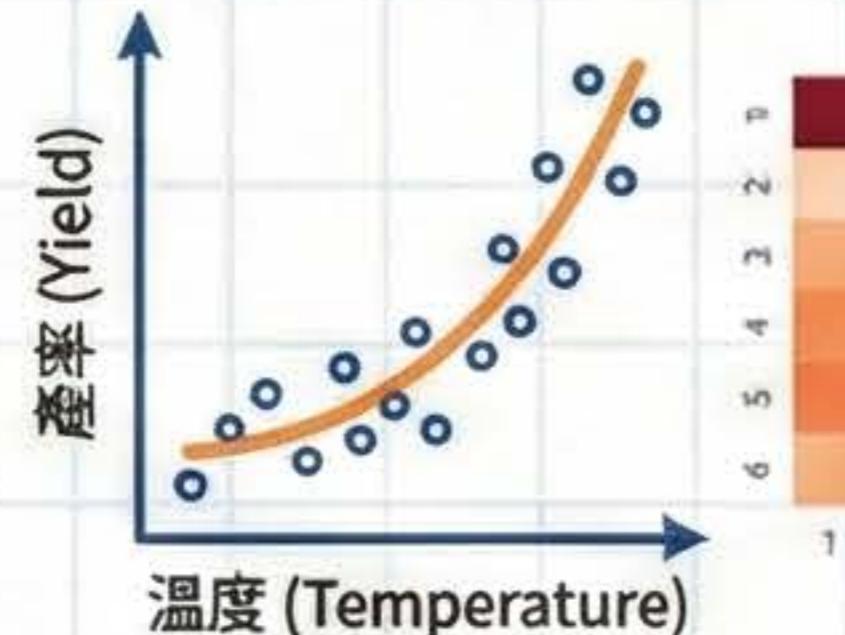
### 2. 數據收集 (Data Collection)

來源：SCADA, LIMS, Sensors.  
檢查：代表性、完整性。



### 3. 探索性數據分析 (EDA)

視覺化分佈與相關性。發現非線性關係 (例如：溫度 vs 產率)。



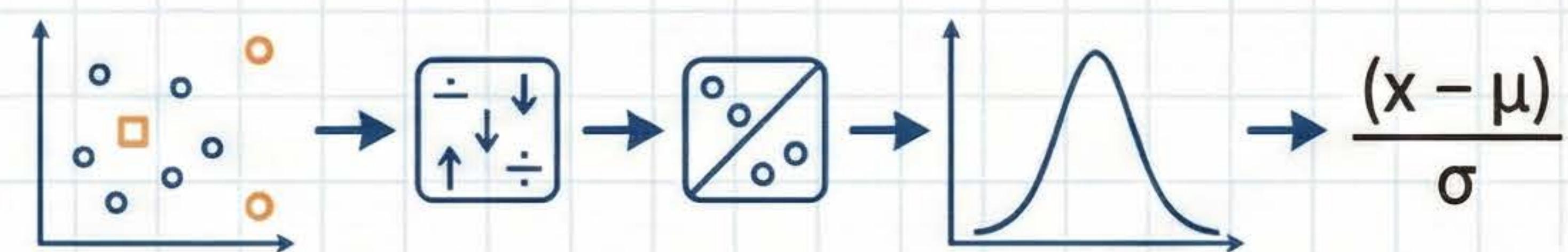
### 2. 數據收集 (Data Collection)

來源：SCADA, LIMS, Sensors  
檢查：代表性、完整性。



### 4. 前處理 (Pre-processing)

- 缺失值填補 (Imputation)
- 離群值處理 (Outlier Removal)
- 標準化 (Standardization):  $(x - \mu) / \sigma$



# 關鍵交集：特徵工程 (Feature Engineering)

這是化工專業知識發揮最大價值的環節。

## 原始數據 (Raw Data)

Sensor\_T1  
Flow\_Rate  
Density  
Pipe\_Diameter



Thermodynamics, Kinetics,  
Fluid Dynamics

## 領域知識轉換 (Domain Transformation)

## 工程特徵 (Engineered Features)

Reynolds Number (Re)  
(Dimensionless)

Reaction Rate (k)  
 $(= A * \exp(-E/RT))$

Enthalpy (H)

## 更佳的模型 (Better Model)

更高的準確度  
(Accuracy)

+

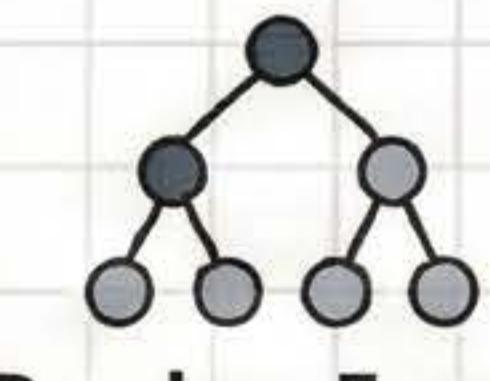
更好的可解釋性  
(Interpretability)

與其直接輸入原始感測器讀數，不如輸入具備物理意義的參數。

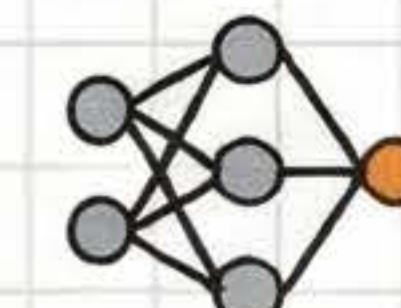
# 機器學習工作流程 Part 2：建模與部署

## 6. 模型選擇 (Model Selection)

依樣本數量與可解釋性需求選擇  
(例如：Random Forest vs LSTM)。



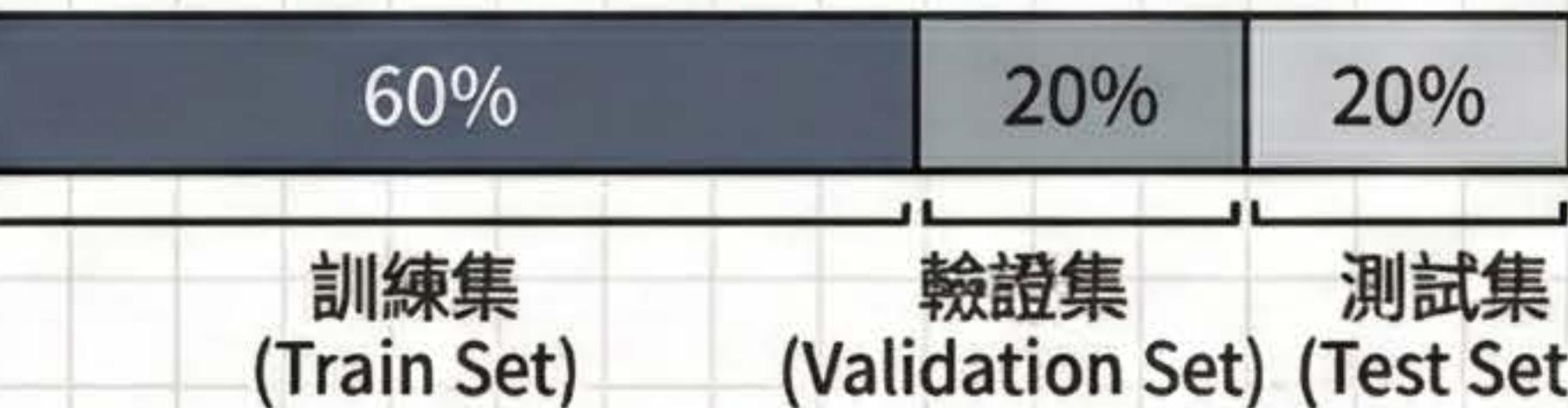
Random Forest  
(可解釋性高)



LSTM  
(適合序列數據)

數據量 / 可解釋性 (Data / Interpretability)

## 7. 訓練 (Training)

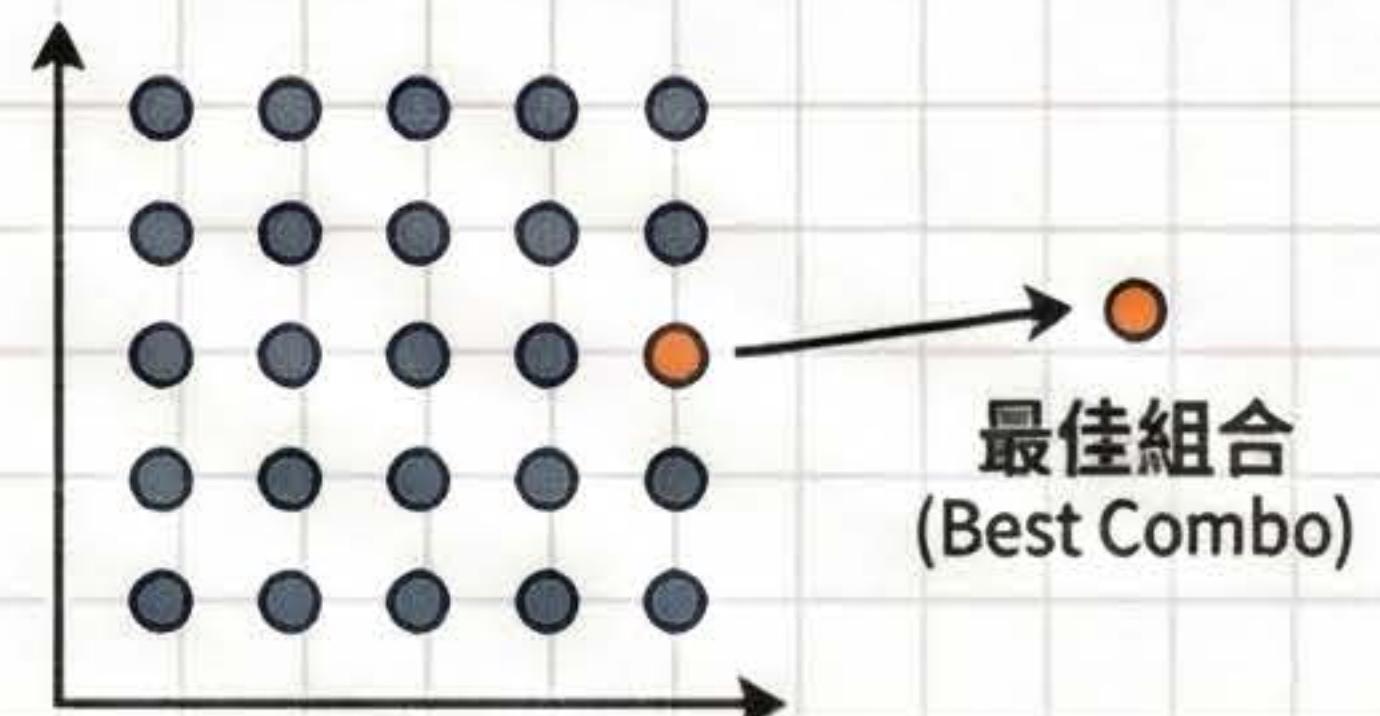


⚠ 絕對不可使用測試集進行訓練！

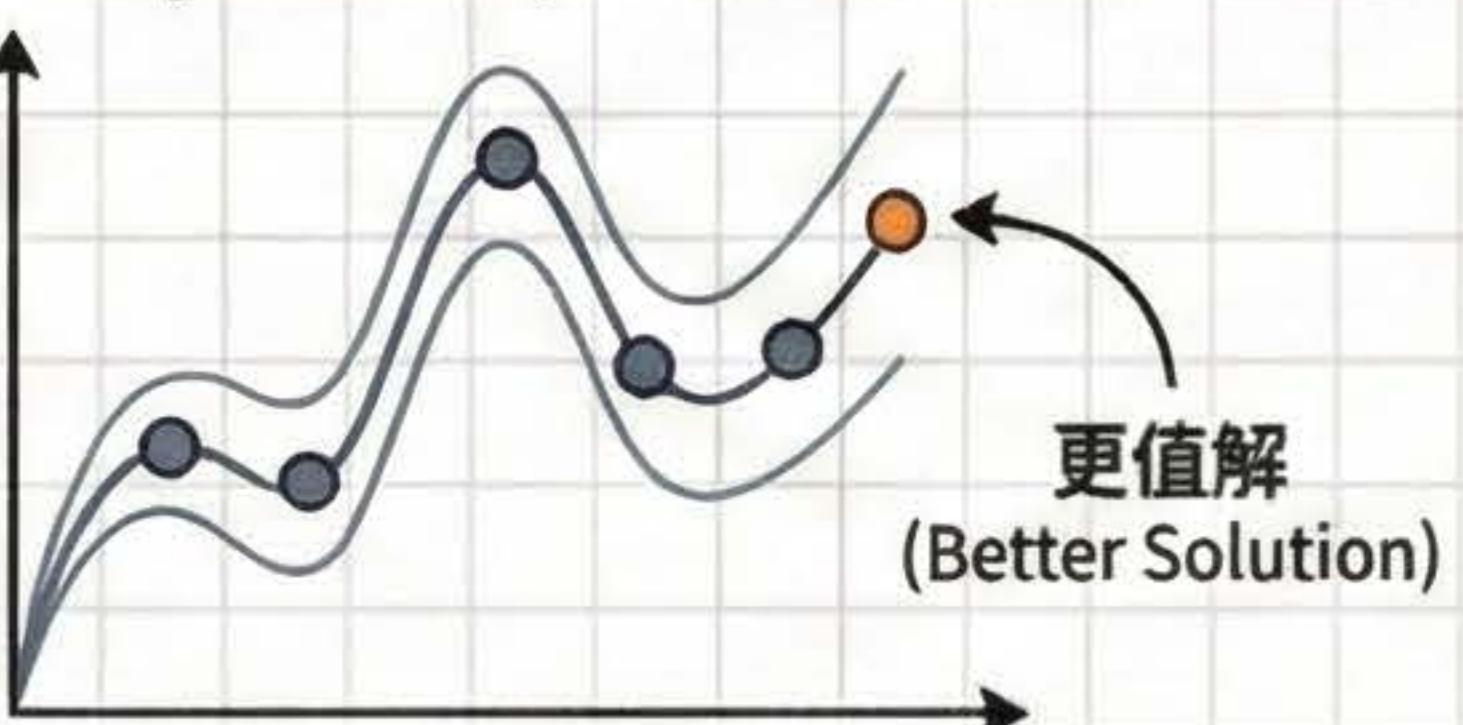
## 9. 優化 (Optimization)

超參數調整 (Hyperparameter Tuning)

Grid Search



Bayesian Optimization

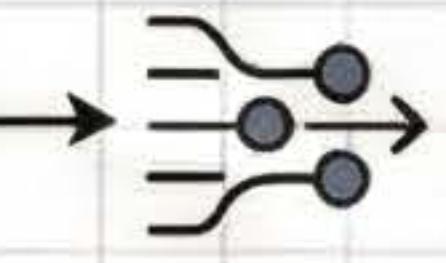


## 10. 部署與監控 (Deployment & Monitoring)

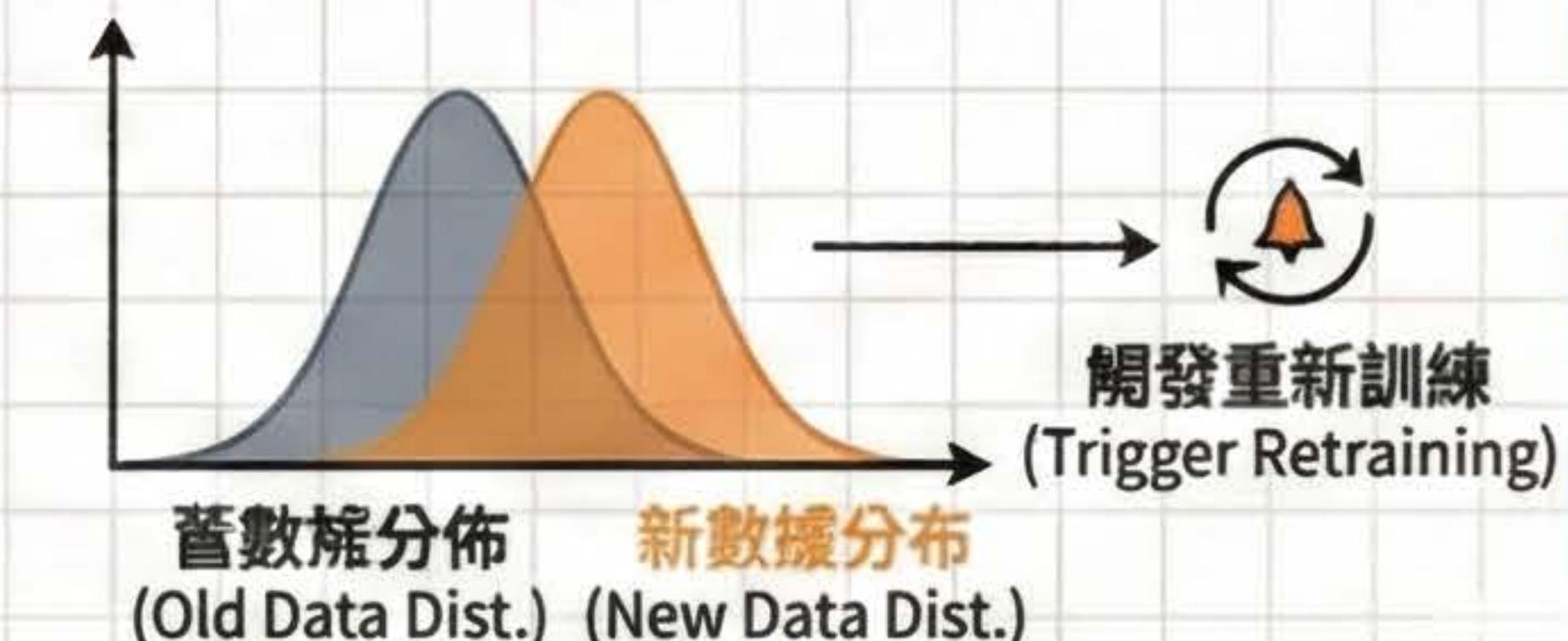
監測 Concept Drift (概念漂移)：當設備老化或原料改變導致模型失效時，需重新訓練。



部署模型  
(Deployed Model)



Live 數據



# 現實世界的挑戰 (Challenges in the Real World)

化工廠不是 Kaggle 競賽，物理世界有其限制。

**小數據 (Small Data) & 不平衡數據**  
實驗成本高昂，異常數據極其稀少。

→ 解決方案 (Solution):

- 數據擴增 (Data Augmentation)
- 遷移學習 (Transfer Learning)
- 合成數據 (Synthetic Data)

**物理限制 (Physics Constraints)**  
純數據模型可能會預測出負的濃度或違反質量守恒。

→ 解決方案 (Solution):

混合建模 (Hybrid Modeling): 結合第一原理模型與機器學習。

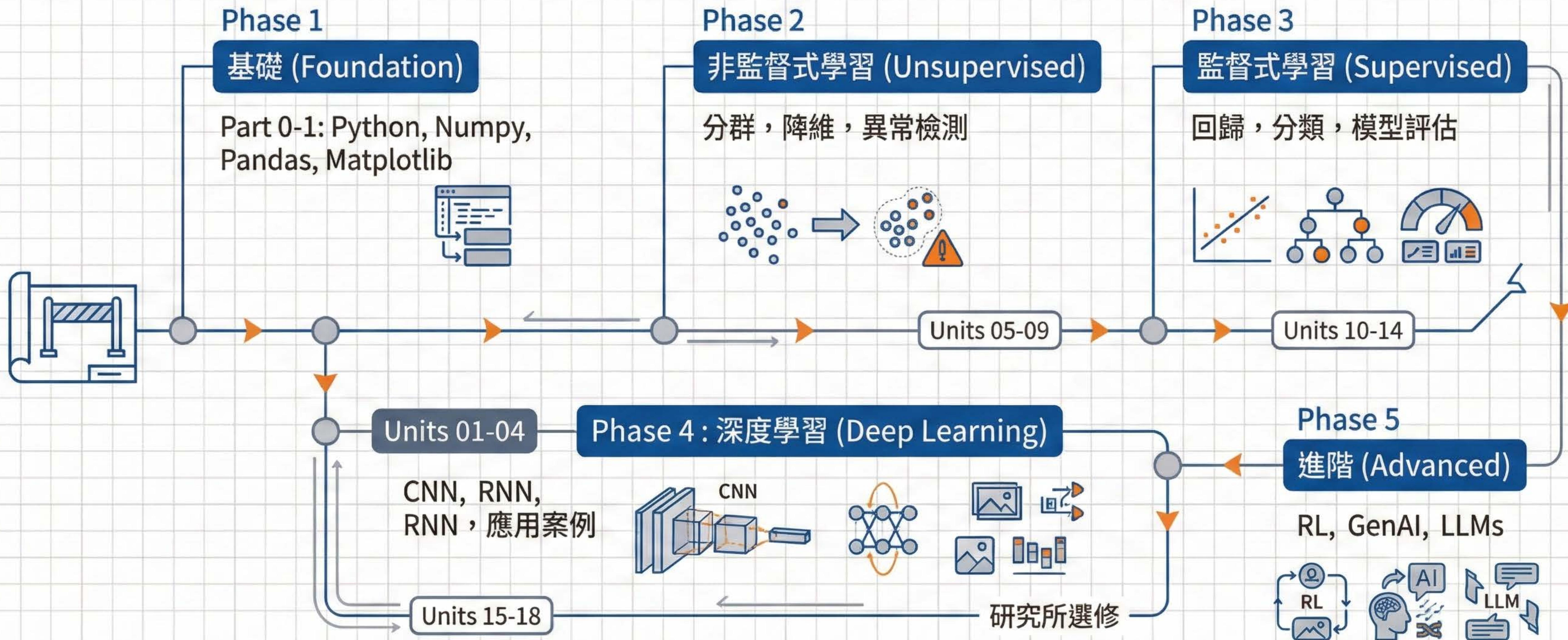
**可解釋性 (Interpretability)**  
操作員需要知道『為什麼』模型建議關閉閥門 (信任與安全問題)。

→ 解決方案 (Solution):

- SHAP Values
- Feature Importance Analysis

Feature A  
Feature B  
Feature C

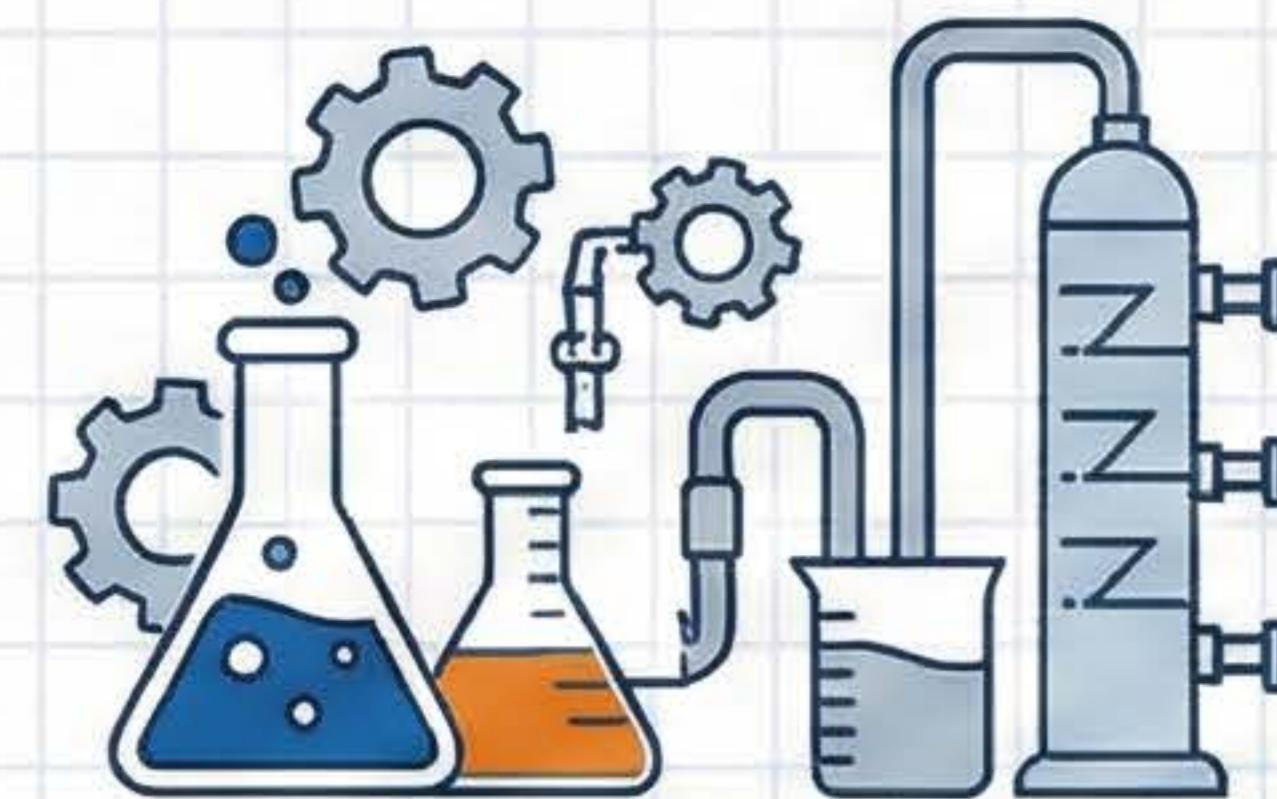
# 課程地圖與學習路徑



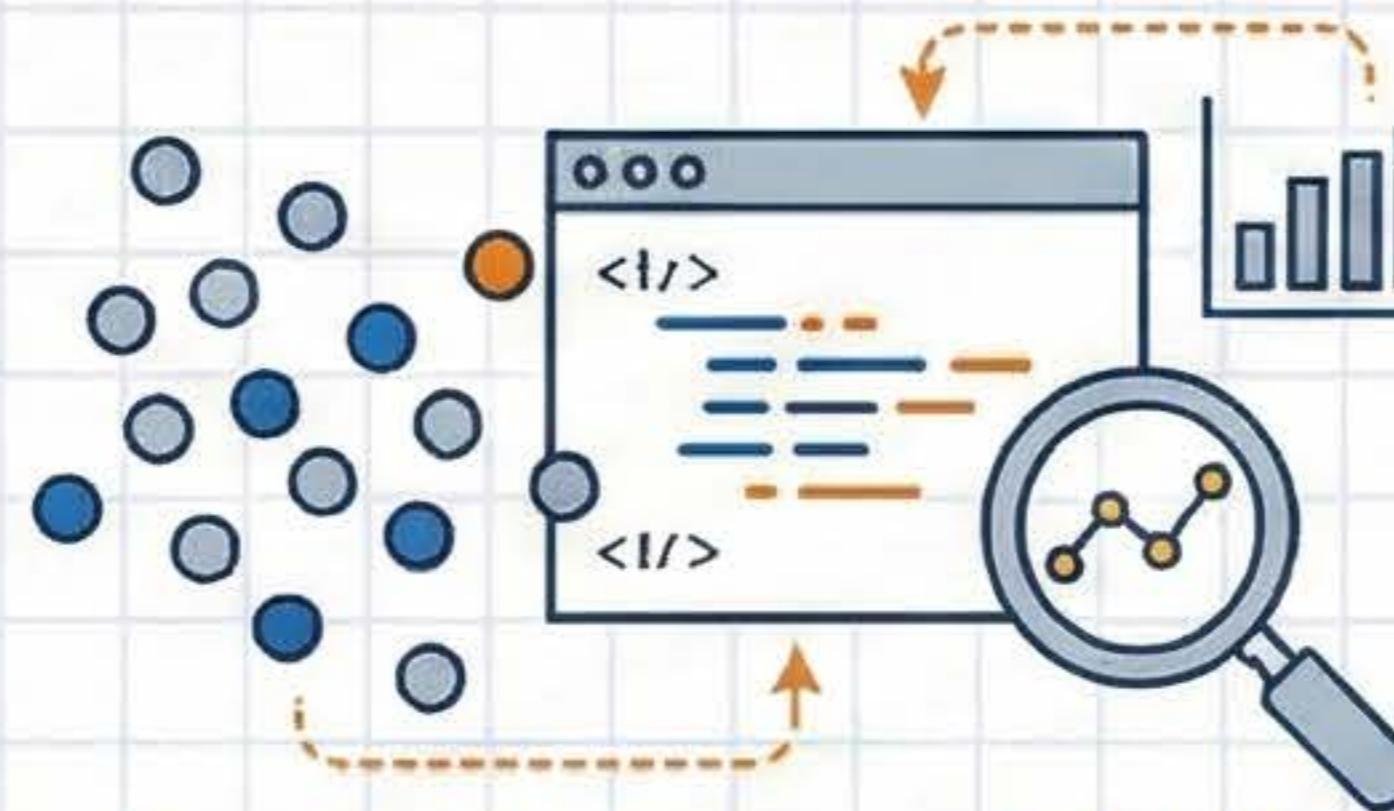
■ 建議初學者從基礎 Python 開始；有經驗者可直接進入 Part 2。

# 如何在 AI 時代取得成功

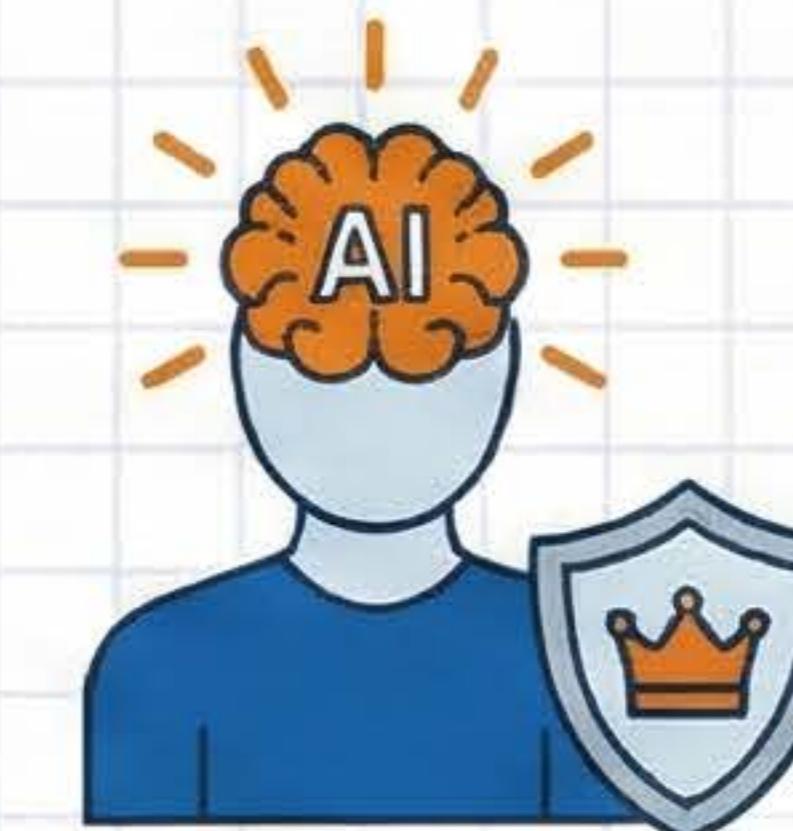
學習建議與心態



Domain Knowledge  
(化工專業)



Data Science Skills  
(數據技能)



不可替代的人才  
(Unbeatable)

不要懼怕程式碼

```
import pandas as pd
data = pd.read_csv("file.csv")
# notes on
for i in some(seo()):
    ar = data.ast('1.1')
    return pxs.amd(1.1)
...

```

Python 是易學的工具。專注於邏輯，而非死背語法。

從簡單開始

線性回歸  
(Linear Regression)      隨機森林  
(Random Forest)      深度神經網路  
(Deep Neural Networks)

先精通線性回歸與隨機森林，再挑戰深度神經網路。

動手實踐 (Hands-On)

理論是不夠的。你需要親自寫 code  
並分析真實數據。

推薦資源：Kaggle, Coursera, Stack Overflow, Scikit-learn Documentation.

# 結語：您是未來的定義者

AI 不會取代化工工程師；  
但懂得使用 AI 的化工工程師將取代不懂的人。



## 影響力 (Impact)



加速創新 (Innovation) →



提升安全 (Safety) →



優化永續性 (Sustainability) →

## 下一步 (Next Step) :

前往 **Unit 02 - Python 程式語言基礎**

開始您的 AI 學習之旅！

