

# AI 的未來與機器學習的基石

## 1. AI 的未來能力

我認為20年後AI能做到、且對人類有重大意義的事是：針對特定患者的特定疾病狀態，從零開始設計出全新的藥物分子，並在該患者的人體模型中完整模擬其療效及分析藥物機理。

想像一位被診斷出患有胰臟癌的病患。目前的標準療法對他無效。在20年後的未來，醫療流程如下：

- 數據採集：醫生會採集該病患的腫瘤樣本進行全基因組以及蛋白質定序。
- 建立人體模型：AI接收這些數據，建立一個該病患的數位人體模型。這個模型不僅是靜態的資料庫，而是一個動態的、可計算的系統，模擬了從分子交互到器官功能的完整生理機制。
- 設計藥物：科學家向AI下達指令：「設計一種或多種化合物，要求能精準抑制此病患癌細胞中的A通路與B通路，同時對其肝臟和心臟的毒性需低於閾值，且分子結構在化學上具備可合成性。」
- 模擬與分析：AI並非盲目搜索，而是在其內部的化學與生物學知識庫中，利用生成模型（Generative Models）創造出數百萬個候選分子。接著再將這些分子投放到病患的人體模型中進行模擬。幾小時後，AI提交一份報告，內容包含：

最佳方案：推薦由兩種新分子（Molecule-X 和 Molecule-Y）組成的雞尾酒療法。

機理推演：報告會附上詳盡的分析與數據，解釋為何這些分子能夠產生療效，並解釋藥物機理。

## 2. 涉及的機器學習類型

- 生成模型 (Generative Models)：從零開始創造全新的分子結構。AI需要學習現有數百萬種化合物的結構規則，然後再進行生成。

資料來源：大型化學資料庫，如 ZINC、PubChem、ChEMBL，其中包含數億種已知分子的結構。

目標訊號：在此階段屬於自監督學習 (Self-Supervised Learning)。模型學習的目標是重建或去噪輸入的分子，從而掌握構成有效分子的底層化學規則與結構模式。

- 監督式學習 (Supervised Learning)：在生成海量候選分子後，系統需要快速、準確地預測每個分子的性質，以篩選掉劣質選項。

3. 強化學習 (Reinforcement Learning)：整個藥物設計並非一次性的生成。AI需要不斷地「生成-評估-修正」，以逐步優化分子，使其滿足所有約束條件（高療效、低毒性、可合成等）。強化學習是適合解決此類問題的框架之一。
- 

### 3. 簡化模型問題

我們的目標為針對單一、剛性蛋白質靶點，利用強化學習生成具備高親和力與新穎性的小分子。

- 單一靶點：代表特定致病通路。
- 分子生成：對應從零開始設計藥物。
- 結合親和力：作為療效的量化指標。
- 化學新穎性：驗證模型的創造能力而非記憶。

**可測試性：**

可透過以下三部分驗證模型的成功：

1. 計算驗證 (In Silico)：使用分子對接軟體 (如 AutoDock) 評估生成分子的對接分數。
2. 專家評估：由化學家評估高分分子的結構合理性與合成潛力。
3. 實驗室驗證 (In Vitro)：合成最佳候選分子，並在生物實驗中測試其活性 (如 IC50)。

LLM is used to aid writing and structurizing