

## 一、未來的 AI

功能：自動產生並篩選藥物候選的

它能從公開化學分子資料與疾病相關文獻裡，針對某特定疾病，設計並合成具成藥性、可製造性且低毒性的候選藥物，再自動打分、留下最有希望的一小批，並說明這個分子被留下的原因。

重要性：能提高新藥的開發速度，提升醫療水平

因現在藥物早期探索成本高、失敗率高，很多候選最後卡在不可合成或毒性。如果能在電腦端就先排掉這些，整體研發流程會省很多時間和錢，這個 AI 也能在材料與農用化學方面使用。

## 二、會用到的機器學習類型

需要「自監督 + 監督式 + 強化學習 + 規則」

自監督：用來在大分子庫上學到分子語法與結構表徵（掩碼還原、圖對比）。

監督式：學 ADMET、理化性質、毒性警訊、合成可及性，這些都有公開或可計算的標籤。

強化學習：作用是「往高分區域繼續找」，因為化學空間是開的，必須要有互動式回饋（每產一個分子，打分器就回一個分數）。

規則：SELFIES 合法性、黑名單、基本逆合成可行是安全邊界，負責把不合規的分子去除掉。

## 三、第一步的簡化模型問題

目標：做出成藥性高、合成可行性高、毒性低

操作方法：在通過語法（基本化學約束）與合規檢查的合法分子中搜尋，並最大化一個綜合分數。
$$\text{綜合分數} = (\text{權重一} \times \text{成藥性指標}) + (\text{權重二} \times \text{合成可行性指標}) - (\text{權重三} \times \text{毒性指標})$$
，三個權重皆為正。

這能代表最終能力，因為它把「像藥、做得出來、不要明顯毒性」三條最硬門檻先穩住，未來只需在綜合分數的計算公式上加上「靶點結合或作用機制的代理分數」，就能變成完整系統。

可測試性：

（1）生成品質：看合法率、去重率、新穎率。同時檢查平均成藥性、平均合成可行性、毒性警訊比例是否達標。

（2）多目標表現：在相同生成配額下，觀察「前一百分位、前五百分位」的綜合分數是否隨迭代穩定上升。

（3）穩定與可解釋：不同隨機種子結果方差是否可控、是否維持多樣性。

（4）合規：黑名單與法規檢查應為零違規，並保留所有決策記錄。

所需工具：

生成端：可用圖神經網路或擴散式生成模型。

評分端：用多任務性質預測模型，提供成藥性、合成可行性與毒性分數。

搜尋端：使用基於高斯過程的多目標貝葉斯最佳化或加權標量化的多起點搜尋。