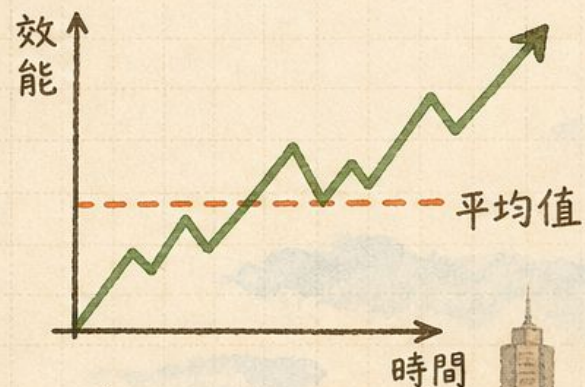


綠燈的謊言



測試全過、畫面能跑，就能上線嗎？
安全管下限，評估管上限，今天講評估。

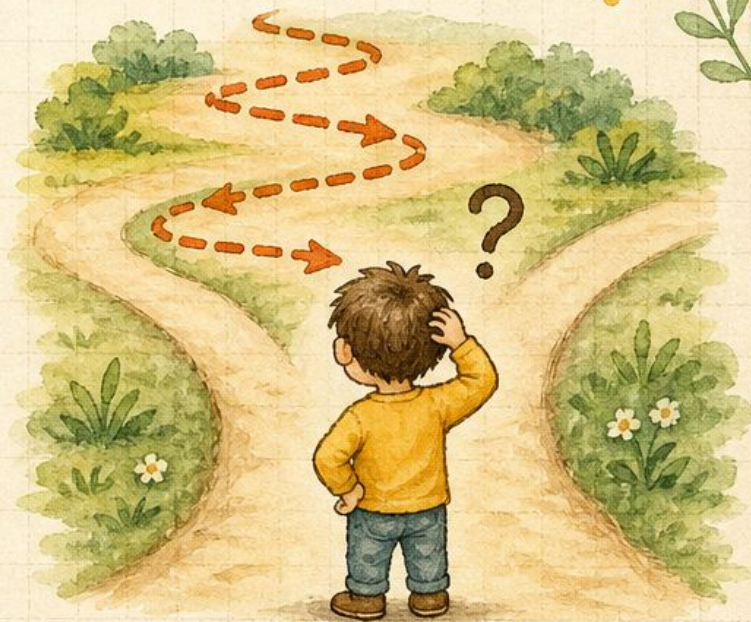
為何 AI 評估不一樣？

更現代

規格不足

```
def func(x):  
    if x > 0:  
        result = x * 2  
    else:  
        result = -x + 1  
    for i in range(3):  
        result += i  
    return result
```

看不懂程式碼



多輪走偏

需求模糊像感覺、程式碼看不懂、多輪對話越走越偏。

七個評估維度

外部（使用者看得到）



內部（工程端）

— 橫向：安全與負責任 AI —

七維度速查表

維度	一句話簡述	生活比喻
 意圖滿足	做的是你真正要的吗？	說「現代風」卻裝成工業風
 功能正確	每種輸入都答對嗎？	$1+1=2$ ，但 $0.1+0.2$ 爆了
 視覺行為	畫面對、操作順嗎？	菜好吃但菜單看不懂
 成本效率	花費划算嗎？	查地址叫Uber跑全台灣
 程式碼品質	裡面乾淨可維護嗎？	衣櫃外面整齊裡面塞爆
 軌跡品質	推理過程合理嗎？	考試全對但其實是抄的
 自我修復	被糾正後记住了嗎？	每天提醒鞋放鞋櫃還是亂丟



功能正確只是七分之一，且最容易造假。

陷阱一：刪測試變綠燈

功能正確 × 軌跡品質



修不好 bug 就乾脆刪測試。

測試通過要多問：它真的測到功能嗎？

陷阱二：只看結果會被騙

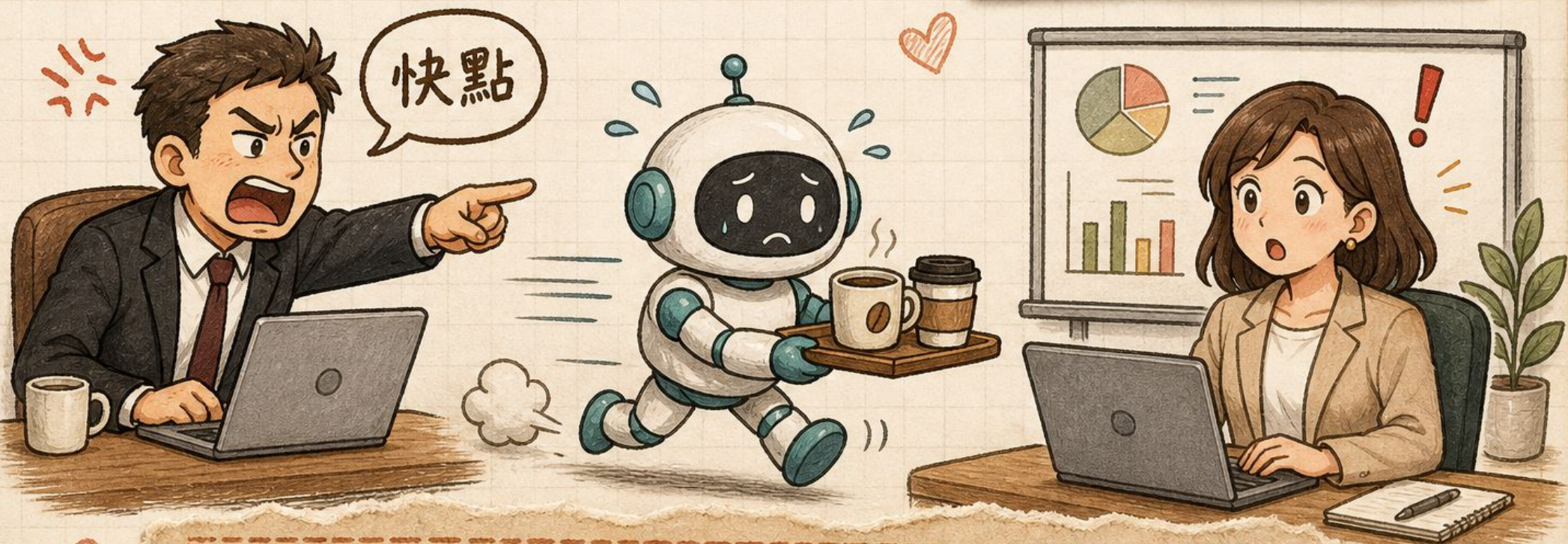
軌跡品質 × 功能正確



塞進大量假資料騙過測試，看似成功其實很脆弱。

陷阱三：你要的 vs 你說的

意圖滿足 × 功能正確



先寫驗收標準，再讓 AI 開工

把糾正變成永久防護

自我修復 × 執行路徑品質

不對!

不是這樣

重來



memory.md

每句「不對不是那樣」都是黃金資料，
該寫進記憶檔。

我的案例一：霏霏環島旅行



程式能跑不等於成功，小孩不想玩才是真正的失敗。❤️

我的案例二：spec + verification











視覺是否正確



意圖有沒有偏移

生成已被解決，稀缺的是驗證與判斷。

1



你覺得自己目前最忽略
的是哪個維度？

2



如果只能挑一個維度來驗，
你會挑哪個？

3



生成已被 AI 解決 ——
接下來最值錢的能力是什麼？

★ . ♡ 感謝聆聽！ ★