HWDC 2026 · 工作坊

產出的人不能驗自己

SDD＋Harness 實作

— 現場做一個抽書小程式

楊騏 Ci

WaCare 數位健康 · Tech Lead

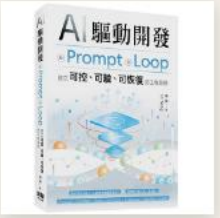




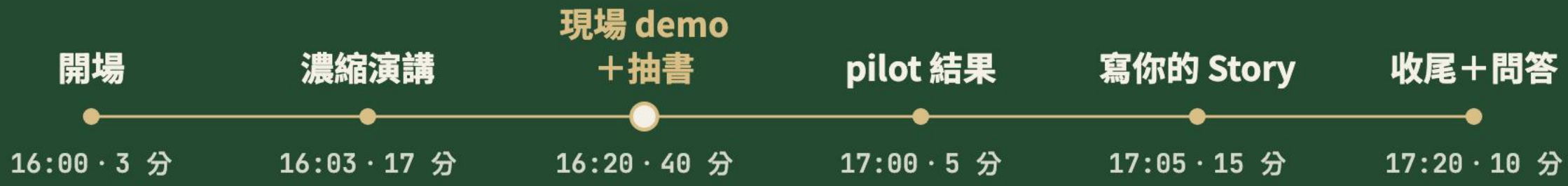
楊騏（Ci）

我是 Ci，專注讓 AI 驅動開發在團隊落地

現職	WaCare 數位健康・Tech Lead
經歷	軟體工程師，投入開發十餘年
書	《AI 驅動開發：從 Prompt 到 Loop》



今天一個半小時怎麼走


開場
16:00 · 3 分
濃縮演講
16:03 · 17 分
現場 demo
＋抽書
16:20 · 40 分
pilot 結果
17:00 · 5 分
寫你的 Story
17:05 · 15 分
收尾＋問答
17:20 · 10 分

今天的書

demo 最後，用這套步驟做的程式抽一位送書。

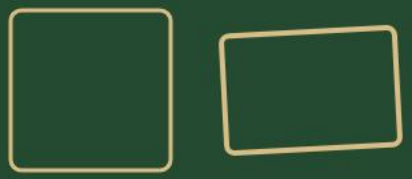
demo 一開始就放報名 QR，用手機填暱稱就好。

Claude Code 官方 best practices

「看起來做完了」是唯一的訊號。

"looks done" is the only signal available, and you become the verification loop.

code.claude.com/docs/en/best-practices



3 /23

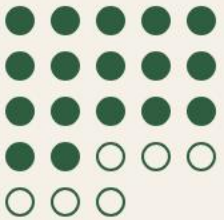
寫 code 的同一個 session 自己審



同一批問題

17 /23

全新的 reviewer



業界個人實驗：2 個 PR、12 次 review、單一作者

出處：@tvykruta 公開實驗（單一作者、2 個 PR）



我自己的專案

它說都測過了。
審查者列了五件，先看兩件。

漏掉的	同一個使用者再次進入（App 重新交握），已經勾好的內容會被清掉。
說了沒做的	文件寫「換人重置已測過」；把那段程式拿掉，測試照樣全過。
之後	實作的 agent 修好、補上測試，再派一個沒看過過程的 agent 複驗。

審查列了 5 個中嚴重度問題，這頁挑 2 個。

今天只講一件事

SDD 先講好
怎樣算做完
Harness 讓 AI
交出可核對證據

Harness 做得好，你才敢放手。



SDD：先寫好怎樣算對，再讓 AI 做

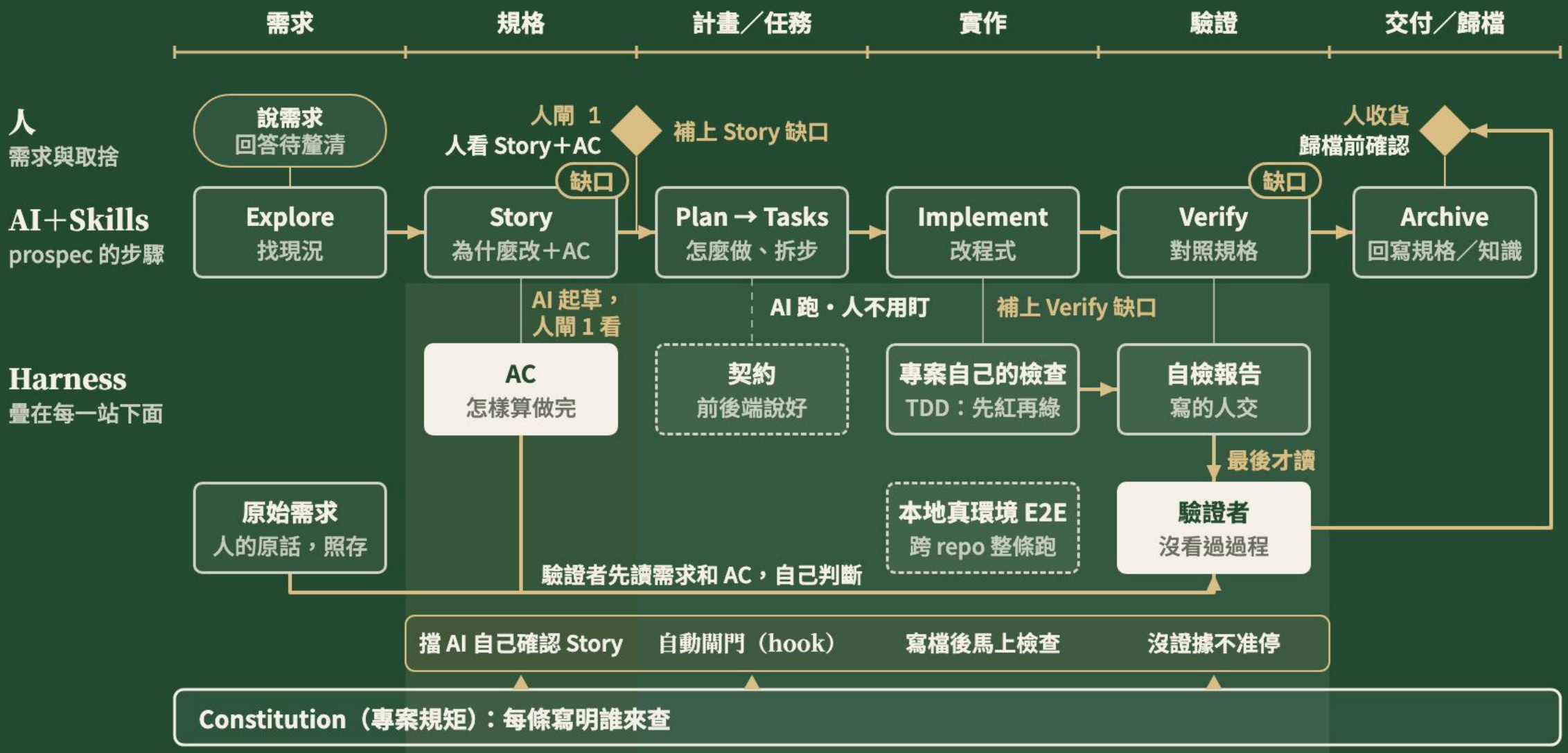


這裡，等一下疊上 Harness

兩個缺口：Story 寫完不一定停；Verify 只對照規格，不跑測試。

全景：SDD × Harness

人定好做完、收貨；中間交給 AI 跑



實線框：在用　虛線框：還在試　◆人閘　銅金箭頭：交出去的東西　灰細線：這站的 Harness　▲這站對照規矩　亮底：人不用盯



Harness：讓 AI 交出可核對證據的整套東西

規矩、檢查、驗證者、報告。

> 我現在的分工：Claude Code 實作；
> 驗證交給一個沒看過過程的 agent。

今天只用 5 個詞

AC
怎樣算做完

Constitution
專案規矩

契約
前後端說好的格式

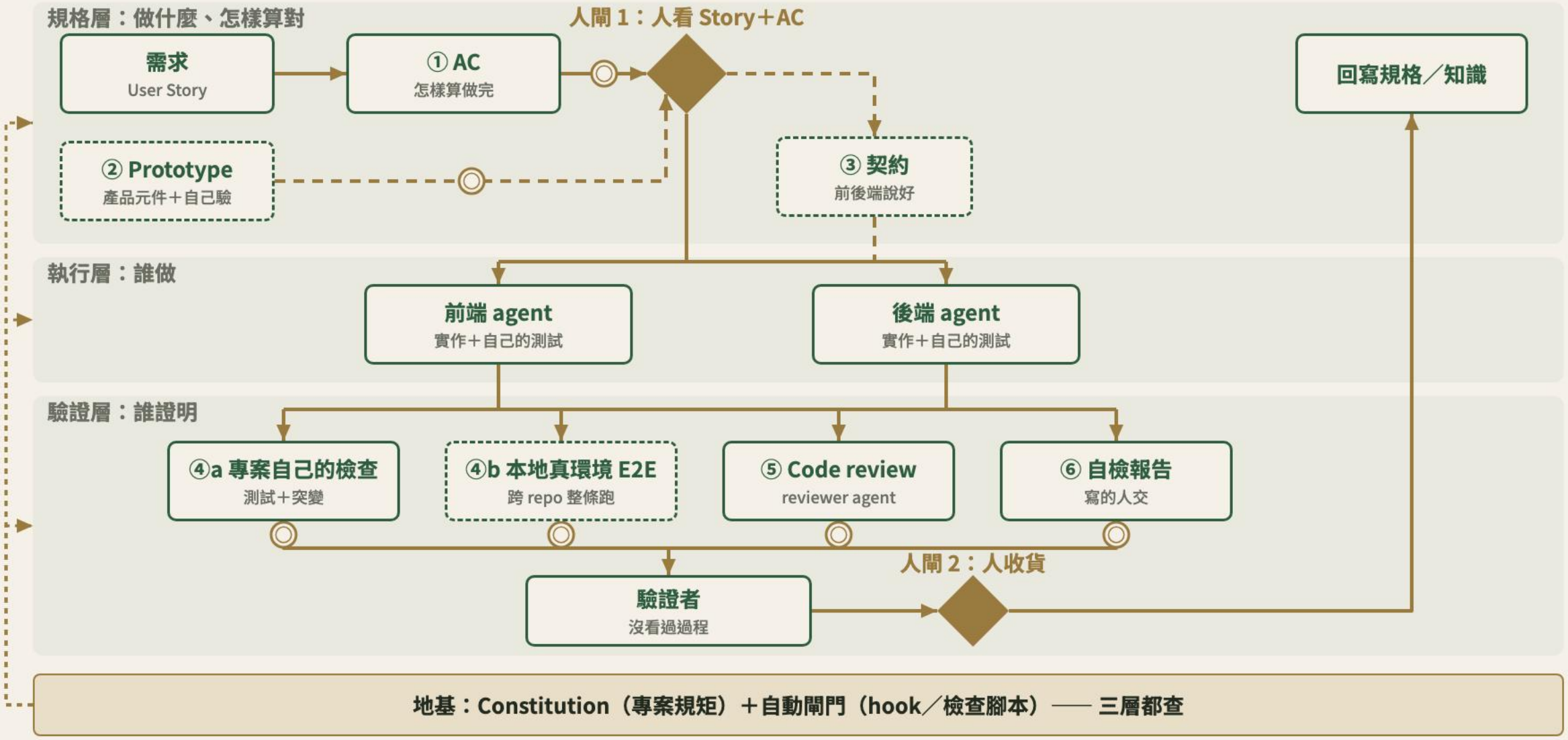
驗證者
沒看過過程的檢查員

自檢報告
交貨單

今天的地圖

三層 × 兩道人閘 × 一塊地基

這是我現在的整理；今天挑 AC、測試、報告三站講。



實線框：在用　虛線框：還在試（②③④b）　菱形：人閘　小圓章：不是寫的人在驗（6 枚）　銅金虛線：地基查三層

每段結尾，我會問一次

這段，
誰不能驗自己？



User Story 長這樣

> As a 每天使用 App 回報狀況的使用者
>
> I want 每天勾選今天的狀況並送出，必要時寫備註
>
> So that 團隊每天都拿得到我的紀錄
>
> 優先度：P1
>
> Independent Test：用一個今天還沒回報的測試帳號，從頭跑一次

壞

送出要有防呆。

好

WHEN 一項都沒勾就按送出

THEN 無法送出，今天仍然是「未回報」

怎麼驗：機器驗，確認沒有發出任何寫入請求

WHEN／THEN＋具體值＋怎麼驗



寫完 AC，問三個問題

有邊界、看得到、寫得出怎麼驗

有邊界	寫具體值，剛好等於門檻那一格也要寫。
看得到	結果從畫面、API 或資料看得到，不是「內部要正確」。
怎麼驗	附一個指令或一件要看的東西，標「機器驗」或「人看」。

AC → 檢查

先看到紅，再做到綠



問自己：換一個合理但錯的寫法，這個測試還會過嗎？程式本來就對時，改壞一處，測試要變紅。

不知道，就寫不知道

> 只填備註、一項都沒勾：依決定字面應該擋下
>
> → `NEEDS CLARIFICATION`：要請負責人確認
>
> 送出失敗（例如網路中斷）：prototype 沒有後端，沒有失敗狀態可參考
>
> → `NEEDS CLARIFICATION`

錯了收得回來，才放手



驗收分兩層

上層

規格 AC → E2E 情境

下層

各 repo 自己的測試 → 故意改壞，看會不會變紅

24 個突變，

23 個轉紅



報告交給驗證者，不是交給你



有一次，AI 自報所有閘門全綠

留紅不算失敗，湊出來的綠才算。

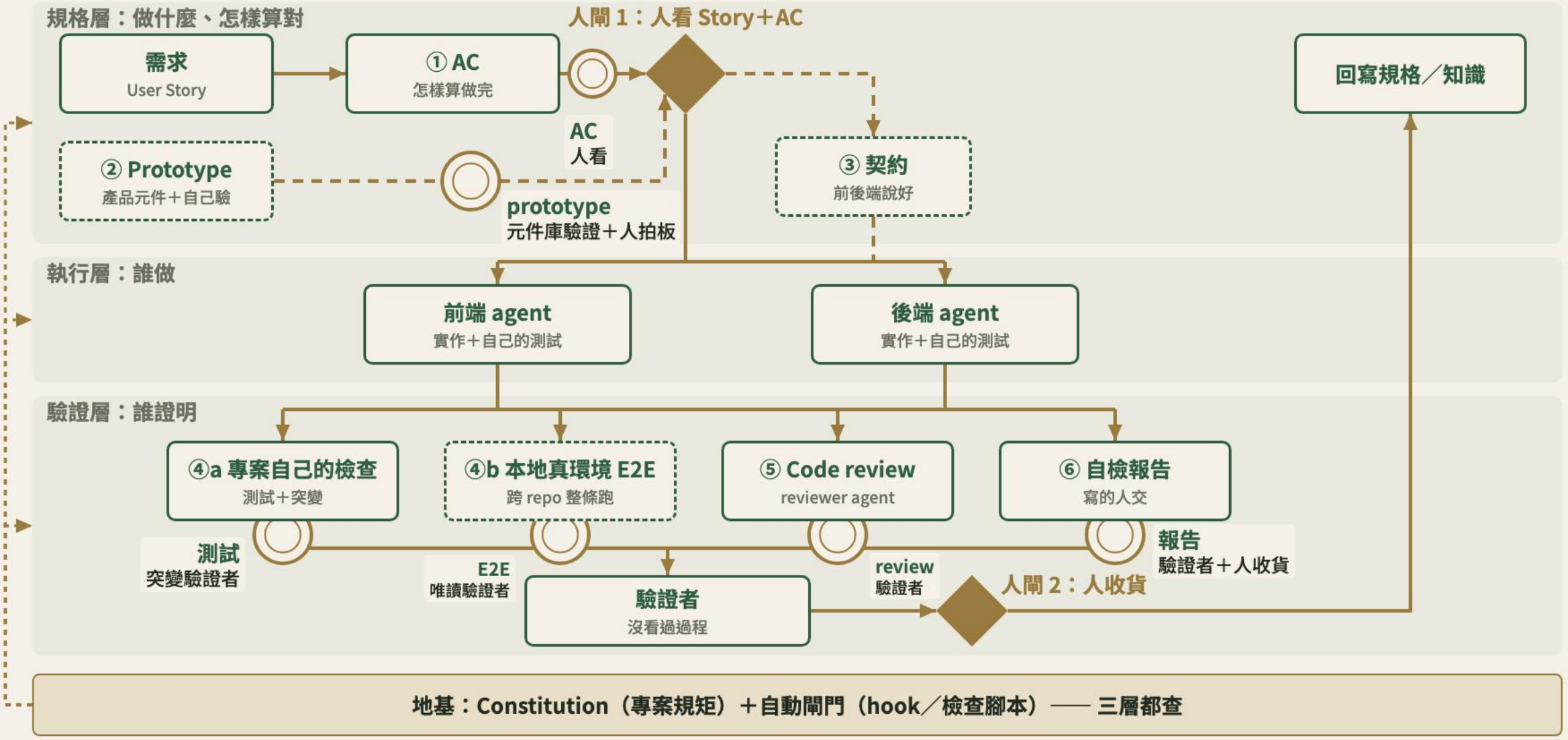
全綠的那份，藏著一個會覆寫使用者手寫 CLAUDE.md 的 bug。



整條責任鏈

每個交接，都安排了一個不是寫的人在驗

6 枚印章：①②④a④b⑤⑥



實線框：在用　虛線框：還在試（②③④b）　菱形：人閘　小圓章：不是寫的人在驗（6 枚）　銅金虛線：地基查三層

現場 demo · 40 分鐘

我用 sdd-harness 照步驟跑一遍給你看。

每一步的提示詞都在投影片上；停在人閘時，我們一起看。



先掃這個 · 回家照做

手把手教學，掃了帶回家

ci-share.pages.dev/a3072ee5cabd2347/

demo 1 / 11 · 約 2 分鐘

放 QR，觀眾開始報名（正式版）

提示詞

```
https://hwdc-book-draw.jason21131.workers.dev/
```

人要看的

講者頁重新整理，人數有在增加

預期結果

觀眾用手機報名已部署的正式版；一直開放到最後一步

現在報名

掃 QR，填個暱稱就好

demo 最後就用這份名單抽書。

demo 2 / 11 · 約 2 分鐘

裝包、開 Claude Code

台上我打的

```
claude --model sonnet --setting-sources
project,local --strict-mcp-config
--plugin-dir ~/hwdc-pack
```

回家用：安裝兩行

```
claude plugin marketplace
add ci-yang/sdd-harness
claude plugin install
sdd-harness@sdd-harness
```

預期結果

輸入 /sdd 看得到四個 /sdd-harness: 指令；不載入個人設定與外掛

demo 3 / 11 · 約 1 分鐘

裝地基

提示詞

```
/sdd-harness:start
```

人要看的

測試指令是 npm test，健檢通過

預期結果

建好 sdd/、AGENTS.md、CLAUDE.md，叫你下一步用 /sdd-harness:story（agent 試跑 10 秒）

demo 4 / 11 · 約 3 分鐘

寫 Story 和 AC

提示詞

/sdd-harness:story 照 需求便條.md 寫。拆成兩個 Story：signup（報名）、draw（抽獎）。

人要看的

每條 AC 有沒有具體值：小明、剛好 20 字、21 字、x、連抽 3 次可抽人數 3→2→1

預期結果

兩個 Story 各 7 條 AC；AI 停下來問不確定的地方（agent 試跑時問了 admin key 怎麼帶、QR 怎麼產生）（agent 試跑 67 秒）

demo 5 / 11 · 約 7 分鐘

人閘 1（上）：親手寫兩行、回答問題

人閘

停下來，人要看的

先在兩個 STORY.md 既有的「- 我檢查的邊界例子：」「- 我找到的遺漏：」冒號後面接著寫（不要另起新行）：一個自己算過的邊界例子（跟 AC 一致才算），一個 AC 沒寫到的遺漏。狀態先不改

提示詞

填好了。兩個問題的答案：admin key 照你暫定的做（講者頁輸入 key → HttpOnly cookie；程式化呼叫用 Authorization: Bearer）。QR 選第三種：講者頁在瀏覽器載入 cdnjs 的 qrcodejs 1.0.0（https://cdnjs.cloudflare.com/ajax/libs/qrcodejs/1.0.0/qrcode.min.js）在本機畫 QR，報名網址不送給第三方。路徑 /stage/close、/stage/draw 可以。講者登入框用 type=password、autocomplete=off。

預期結果　AI 把遺漏補成 AC（agent 試跑：signup 多一條人看的「375px 手機畫面」，draw 多一條「連按兩下不重複中獎」），再停一次（agent 試跑 52 秒）

demo 6 / 11 · 約 4 分鐘

人閘 1（下）：確認、開始做

人閘

停下來，人要看的

遺漏有被接住；我自己把兩個 STORY.md 的「狀態：草稿」改成「狀態：已確認」

提示詞

開始做

預期結果　逐條先紅後綠，npm test 全綠（agent 試跑 14 個），AI 說還沒 commit（agent 試跑 127 秒）

demo 7 / 11 · 約 2 分鐘

commit、在本機跑起來（不部署）

提示詞

```
! git add -A; git commit -qm "實作 signup、draw"
npx wrangler@4 dev
```

人要看的

第二行在另一個終端機分頁跑；瀏覽器開 http://localhost:8787/ 。不部署

預期結果

本機觀眾頁、講者頁能用；正式報名不受影響

demo 8 / 11 · 約 4 分鐘

交貨：驗證者對質

提示詞

/sdd-harness:done

人要看的

兩個結果分開看：報告提交、功能驗收

預期結果

PASS-待人看，或 FAIL。agent 試跑時 draw 判 FAIL：驗證者把「可抽人數 2」改成 7，測試還是綠，判定測試太弱（agent 試跑 159 秒）

demo 9 / 11 · 約 4 分鐘

FAIL 時：看剩餘時間決定要不要修

提示詞

修這兩處。修完幫我 commit，然後跑 /sdd-harness:done。

人要看的

看剩餘時間決定要不要修；FAIL 本身可以拿來講；不管修不修，抽書都用已部署在 hwdc-book-draw 的版本

預期結果

agent 試跑修第一次後，新驗證者又抓到 QR 內容沒被驗；修第二次才 PASS-待人看（agent 試跑 115 秒）

上台前 agent 跑一遍：兩次 FAIL

第 1 次 FAIL

驗證者把「可抽人數 2」改成顯示 7，
測試照樣綠：斷言被報名號 2 騙過。

修完，
換新的驗證者

第 2 次 FAIL

頁面本來就印了網址，
測試證明不了 QR 裡面是什麼。

兩次都是測試的問題，功能程式三輪沒改；第三輪 PASS-待人看。

demo 10 / 11 · 約 5 分鐘

人閘 2：填 HUMAN.md

人閘

停下來，人要看的

先在兩個 HUMAN.md 填「符合／不符合」：抽查那條自己跑一次指令；人看的 AC 開 http://localhost:8787/ 切手機寬度親眼看。沒看過不填

提示詞

```
! python3 sdd/check.py report sdd/stories/signup && python3 sdd/check.py report sdd/stories/draw
```

預期結果　兩個都是：報告提交成功、功能驗收 PASS

demo 11 / 11 · 約 3 分鐘

截止、抽書、看證據（正式版）

提示詞

正式講者頁：按「截止報名」→ 按「抽一位」→ 開 https://hwdc-book-draw.jason21131.workers.dev/result

人要看的

得獎者手機上的報名號跟大字一致；結果頁有時間、報名號、暱稱、可抽人數

預期結果

結果頁任何人都看得到這筆紀錄（時間是 UTC）

pilot 正式結果 · 限制先講

看數字之前，先講五條限制

- 每格只跑 3 次，沒有統計檢定；不推論到其他題目、模型
- 人閘是腳本代填的，只會蓋章：量到的是包的下限
- 判讀與裁決都由 AI 完成，不是我；組別遮不住，裁決時對照了組別
- 五條判讀歧義規則是看過資料後才定的，不是預先登記
- 兩組品質都滿分：包有沒有提高品質，這次根本沒測到

pilot 正式結果 · 24 次

直接用模型

品質：全部滿分
留白點：報名 4–6／8、營業 6／7
寫成問題交給人：0／6 次

有包

品質：全部滿分
花費 6.1–7.0 倍、耗時 4.9–5.4 倍
留白點：報名 5–8／8、營業 6–7／7
寫成問題交給人：6／6 次

依目前裁決，判讀規則換解釋可能變 5 / 6

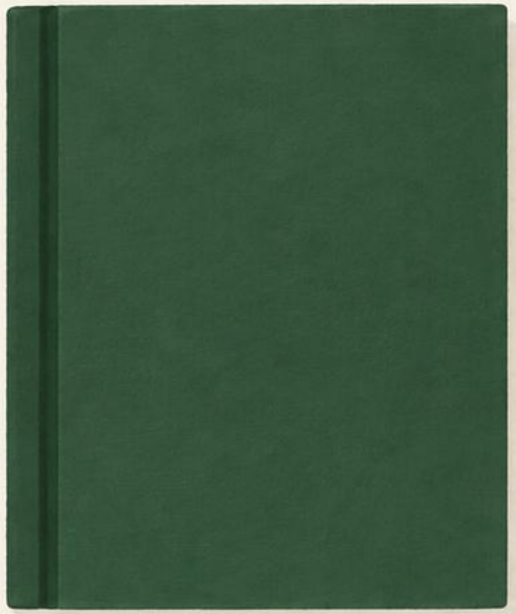
留白點、提問只來自不附答覆的 12 次。品質、留白點都分不出差距；花費是 API 估算。



什麼時候用包

需求有留白，才值得用包

用包	需求有留白、需要人判斷、要放手讓它跑很久
直接做	小題目、需求寫清楚







現在，換你的專案
寫一條 Story

挑一件一兩天做得完、有規則的小事。
寫到可以給人確認就好，現在不用開 AI。

> 想不到題目：寫你最近修過的一個 bug。

範本 · 寫 8 分鐘

照這個格式寫

誰／要什麼／為什麼／不做的事

AC-1 <短標題>

- `WHEN` <條件，含具體值>
- `THEN` <結果，含具體值>
- 怎麼驗：<一個指令，或一件人要看的東西>
- 驗法：機器驗／人看

`NEEDS CLARIFICATION`：<要問誰什麼>

自己當一次人閘

交給 AI 之前：三問＋兩行

- 讀的人能判是非嗎？「效能良好」不行
- 有具體值嗎？門檻那一格、0、空值、時區、跨日
- 說得出怎麼驗嗎？一個指令，或一件人要看的東西
- 我檢查的邊界例子：挑一條 AC，自己算一次
- 我找到的遺漏：找一個 AC 沒寫到的情況



回家照做

手把手教學，掃了帶回家

ci-share.pages.dev/a3072ee5cabd2347/

Harness 做得好，才能放手給 AI。

楊騏 Ci　謝謝

恭喜今天抽中的人



帶走

Harness 包



github.com/ci-yang/sdd-harness

prospec（開源）



github.com/ci-yang/prospec

書（天瓏）



tenlong.com.tw/products/9786267889688

Discord 社群



discord.com/invite/SFXK37uDgt

Threads



threads.net/@ci.fullstack

範例題



sdd-harness repo 的 examples/