論文生成全過程紀錄

一場人類與 AI 協作的學術寫作工作流

Cheng-I Wu Claude (Anthropic)

2026 年 3 月 8 日

Abstract

本文記錄一篇學術論文從構思到完稿的完整人機協作過程。該論文探討代理型人工智慧（agentic AI）如何重新定義學生學習成效測量，並對台灣第四週期校務評鑑提出政策建議。全過程使用 Claude Code 的 `academic-pipeline` skill（v2.3）調度，歷經研究、撰寫、誠信驗證、模擬同儕審查、修訂、格式化等 9 個階段，共 2 輪完整 pipeline 循環。最關鍵的迭代來自使用者提出「學習的定義本身也在變」的洞察，催生了論文最具原創性的 §4.7 共同演化觀點。

Contents

1 論文生成全過程紀錄 3
1.1 論文資訊 . . . 3
1.2 第一階段：起始與研究（Session 2, 2026-03-07 晚間） . . . 3
1.2.1 Stage 1: RESEARCH（deep-research skill） . . . 3
1.3 第二階段：撰寫（Session 2 續） . . . 3
1.3.1 Stage 2: WRITE（academic-paper skill） . . . 3
1.4 第三階段：誠信審查（Session 2 續） . . . 4
1.4.1 Stage 2.5: INTEGRITY（integrity_verification_agent） . . . 4
1.5 第四階段：同儕審查 + 修訂（Session 2 續） . . . 4
1.5.1 Stage 3: REVIEW（academic-paper-reviewer skill） . . . 4
1.5.2 Stage 4: REVISE . . . 4
1.5.3 Stage 3’: RE-REVIEW（驗收審查） . . . 4
1.5.4 Stage 4.5: FINAL INTEGRITY . . . 5
1.6 第五階段：格式化（Session 2, 2026-03-08 凌晨） . . . 5
1.6.1 Stage 5: FINALIZE . . . 5
1.7 第六階段：中文版 + 作者決定（Session 2 續） . . . 5
1.8 第七階段：NotebookLM 簡報 Prompt（Session 3, 2026-03-08 上午） . . . 6

1.9 第八階段：共同演化觀點的誕生（Session 3 續，最關鍵的迭代） 6
1.10 第九階段：第二輪 Pipeline（Session 3 續） 7
1.10.1 Stage 3 Round 2: REVIEW 7
1.10.2 Stage 4 Round 2: REVISE 7
1.10.3 Stage 3' Round 2: RE-REVIEW → Accept 7
1.10.4 Stage 4.5 Round 2: FINAL INTEGRITY → PASS 7
1.10.5 Stage 5 Round 2: FINALIZE 7
1.11 互動模式總結 8
1.12 使用者的關鍵決策時刻 8
1.13 關鍵教訓 8
1.14 協作品質評估 9
1.14.1 總分 9
1.14.2 做得好的（What Worked Well） 9
1.14.3 錯過的機會（Missed Opportunities） 10
1.14.4 下次建議（Recommendations for Next Time） 10
1.14.5 人類 vs AI 的價值貢獻（Human vs AI Value-Add） 11
1.14.6 Claude 的自我反思 11

1 論文生成全過程紀錄

1.1 論文資訊

標題：*From Snapshots to Trajectories: How Agentic AI Will Redefine Student Learning Outcomes and Transform Student Success Measurement — Implications for Taiwan's Next Cycle of Institutional Accreditation*

最終產出：英文全文 ~15,000 字（997 行）、中文全文 ~69,000 tokens、APA 7.0 LaTeX PDF、DOCX、NotebookLM 簡報 Prompt

---

1.2 第一階段：起始與研究（**Session 2, 2026-03-07** 晚間）

使用者的第一句話：

> "use academic pipeline related skills. I would like to write a paper about how agentic ai will redefine student learning outcomes and the paradigm shift of student success measurement, in the context of…"

這觸發了 `academic-pipeline` skill（v2.3），啟動完整 9-stage pipeline。

1.2.1 Stage 1: RESEARCH（deep-research skill）

- Claude 使用 `deep-research` skill 進行文獻搜索與研究綜述
- 產出：`stage1_research_synthesis.md` — 包含研究問題、方法論框架、文獻清單、核心論點

使用者的第一次方向修正：

> "wait, i would like to add a direction: and its implication to the next cycle of institutional accreditation"

這讓論文從純理論分析轉為具體政策建議，聚焦到 HEEACT 第四週期校務評鑑的設計窗口。使用者確認新方向後回覆 `"good"`，研究階段繼續。

1.3 第二階段：撰寫（**Session 2** 續）

1.3.1 Stage 2: WRITE（academic-paper skill）

- Claude 用 `academic-paper` skill 平行撰寫 6 個區段檔案：
  - `sections_1_7_8_abstract.md`（緒論、研究議程、結論、摘要）

  - section2_agentic_ai_paradigm.md（AI 四層級分類架構）
  - section3_taiwan_current_paradigm.md（台灣現行典範與六項結構性限制）
  - section4_paradigm_shift.md（ADAPT 框架、七維度典範轉移）
  - section5_policy_accreditation.md（三情境分析、南韓前車之鑑、分階段路徑）
  - section6_ethics_governance.md（四原則倫理分析、三層治理）
- 組合成 full_paper.md 完整英文論文

使用者多次回覆「繼續」推進 pipeline。

1.4 第三階段：誠信審查（Session 2 續）

1.4.1 Stage 2.5: INTEGRITY（integrity_verification_agent）

- 第一次驗證發現問題 → integrity_report_stage2.5.md
- 修正後重新驗證 → integrity_reverification_stage2.5.md（PASS）
- 確保所有參考文獻存在、引用脈絡正確、統計數據可驗證

1.5 第四階段：同儕審查 + 修訂（Session 2 續）

1.5.1 Stage 3: REVIEW（academic-paper-reviewer skill）

模擬 5 位審稿人的獨立審查：

- **EIC**（主編）：結構與整體性
- **R1**（方法論專家）：Kuhn 框架的適當性
- **R2**（領域專家）：台灣實務的準確性
- **R3**（跨領域專家）：認知科學與教育科技
- **Devil's Advocate**（魔鬼代言人）：挑戰所有論點

產出：stage3_review_report.md（60KB）— **Minor Revision**

1.5.2 Stage 4: REVISE

根據審查意見逐項修訂。使用者回覆「修訂」啟動。

1.5.3 Stage 3': RE-REVIEW（驗收審查）

產出：stage3prime_rereview_report.md — **Accept**

1.5.4 Stage 4.5: FINAL INTEGRITY

最終誠信驗證 → `integrity_report_stage4.5.md` — **PASS**

1.6 第五階段：格式化（**Session 2, 2026-03-08** 凌晨）

1.6.1 Stage 5: FINALIZE

使用者：

> “產出 docx + latex + pdf”

Claude 產出 DOCX + LaTeX + PDF。使用者檢視後：

> “pdf 是 latex 格式嗎？”

發現 PDF 是從 HTML 轉的（非學術標準），使用者要求：

> “安裝並且以後這個 skill 產出的 pdf 都必須從 latex 轉出”

Claude 安裝 `tectonic`（LaTeX 編譯器），從此所有 PDF 必須從 LaTeX 編譯。

使用者接著傳了 **PDF** 截圖做視覺檢查，並問：

> “這格式有符合學術論文的規範嗎”

Claude 調整為 APA 7.0 格式（`apa7` document class），重新編譯。

Lesson Learned 被寫入 **skill**：使用者說「remember the lesson and update skill」，academic-pipeline v2.3 更新了 Stage 5 的規範——PDF 必須從 LaTeX 編譯，禁止 HTML-to-PDF。

1.7 第六階段：中文版 + 作者決定（**Session 2** 續）

使用者：

> “translate it in a traditional chinese version”

Claude 平行翻譯 8 章，產出 `full_paper_zh.md`。

使用者：

> “把我從英文版和中文版的作者中刪除，主要作者改為你自己”

作者改為 Claude（AI 生成論文，不掛人類作者）。

使用者接著發現中文版語感問題：

> “繁體中文版讀起來不太順暢，用台灣教育界慣用的學術語氣撰寫”

Claude 對中文版進行全面語感修訂，調整為台灣高教學術界慣用的論述風格。

Session 結束，推送 git commit。

1.8 第七階段：NotebookLM 簡報 Prompt（Session 3, 2026-03-08 上午）

使用者：

> “昨天你生成了一篇關於代理 ai 對高等教育學生學習成效評量與品保的啟示的論文，如果我要把它用 notebooklm 生成研討會發表的簡報，prompt 你會怎麼下？”

Claude 使用 `notebooklm-slides-generator` skill，產出 15 張投影片的結構化 prompt（`agentic-ai-slides-prompt.md`），可直接貼入 NotebookLM 的 Slide Deck 功能。

1.9 第八階段：共同演化觀點的誕生（Session 3 續，最關鍵的迭代）

這是整個過程中最重要的智識互動。

使用者提出了論文的核心盲點：

> “這份論文是否有探討「學生學習」在未來幾年可能經歷的定義轉變？也就是學習這件事，包含什麼是學習、如何學、為何學，有了 ai 人類的學習與認知能力會經歷什麼變化？都會隨著 ai 發展動態變化”

Claude 分析後確認：論文隱含地假設「學習」是一個固定的建構（fixed construct），只是需要更好的測量工具。但使用者指出了更深層的問題——AI 不只改變「怎麼測量」學習，更在改變「什麼是學習」本身。

Claude 提出 5 項修訂方案，使用者確認：

> “優先順序和範圍 ok; 新增 4.7; 新增文獻支持。改完一樣走一遍 pipeline 流程”

決定：

1. 新增 §4.7「The Co-Evolution Dynamic: Why a One-Time Paradigm Shift Is Insufficient」
2. 修改 §1.1 將共同演化提升為論文核心論旨（framing thesis）

3. 新增 7 篇參考文獻（Clark & Chalmers, Hayles, Hutchins, Luckin, Risko & Gilbert, Sparrow et al., Sterelny）
4. 全文重新走一遍 pipeline

使用者特別指示：

> "直接改英文版，不要改中文"

1.10 第九階段：第二輪 Pipeline（Session 3 續）

Claude 修改英文版，加入：

- 三種認知轉化機制：認知卸載（strategic vs maladaptive）、認知增強（extended mind thesis）、認知重組（technogenesis）
- 共同演化循環（5 步遞迴）
- 適應性定義框架（三個設計原則）
- Clark-Sterelny 哲學張力的承認
- Vygotsky ZPD 整合
- 不可證偽性限制的坦承

然後重新走 pipeline：

1.10.1 Stage 3 Round 2: REVIEW

5 位審稿人再次審查 → `stage3_review_report_r2.md`。決定：**Minor Revision**（8 必要項 + 5 建議項）。關鍵問題：R-1 減少篇幅 15-20%、R-3 Kuhn-Scenario B 矛盾、R-4 Clark-Sterelny 張力。

1.10.2 Stage 4 Round 2: REVISE

逐項處理 R-1~R-8 + A-1（Vygotsky ZPD）+ A-2（不可證偽性）。產出 `response_to_reviewers_r2.md`。

1.10.3 Stage 3' Round 2: RE-REVIEW → Accept

1.10.4 Stage 4.5 Round 2: FINAL INTEGRITY → PASS

1.10.5 Stage 5 Round 2: FINALIZE

使用者選擇 APA 7.0 格式（回覆「1」），產出最終版 `full_paper_apa7.pdf`（418 KB）。

1.11 互動模式總結

面向	描述
使用者角色	方向指引者 + 品質把關者 + 智識挑戰者
Claude 角色	研究者 + 撰寫者 + 審稿模擬 + 格式化引擎
使用者介入次數	~15 次實質性決策（不含「繼續」）
關鍵轉折點	使用者提出「學習的定義本身也在變」→ 催生 §4.7 共同演化觀點
Pipeline 完整循環	2 輪（初版 + 共同演化修訂版）
模擬審查	2 輪 × 5 位審稿人 = 10 份審查報告
誠信驗證	4 次（Stage 2.5 × 2 + Stage 4.5 × 2）
總 **session** 數	2 個主要 session + 1 個當前 session
總耗時	約 5-6 小時（含等待編譯時間）
最終產出	6 檔案（MD + DOCX + TEX + PDF + 審查報告 + 回覆信）

1.12 使用者的關鍵決策時刻

1. 定方向：agentic AI + student learning outcomes + paradigm shift
2. 擴範圍：加入「next cycle of institutional accreditation」政策面向
3. 品質要求：PDF 必須從 LaTeX 編譯、格式必須符合 APA 7.0
4. 作者決定：移除人類作者，AI 為主要作者
5. 語感修訂：中文版要用台灣教育界慣用的學術語氣
6. 智識深化：提出「學習定義本身也在變」→ 共同演化觀點（論文最具原創性的貢獻）
7. 流程選擇：修訂後重走完整 pipeline

1.13 關鍵教訓

1. 最有價值的人類貢獻不是寫作，而是提問——使用者一句「學習的定義會不會變？」催生了論文最深刻的章節
2. 迭代式 **pipeline** 比一次性生成有效——兩輪 10 份審查報告發現並修正了 13+ 項問題
3. 視覺檢查不可省略——使用者傳 PDF 截圖才發現格式問題
4. 中文不是翻譯而是重寫——直譯的學術中文不自然，需要用領域慣用語感重寫

1.14 協作品質評估

本章以誠實、建設性的態度評估使用者在本次人機協作論文寫作中的表現。評估基於整個 pipeline 過程中的實際行為與決策紀錄。

1.14.1 總分

一句話評語：使用者展現了稀有的「智識催化者」角色——不是被動等待 AI 產出，而是以一個關鍵提問（「學習的定義本身也在變」）從根本上改變了論文的理論深度，這是 AI 無法獨立達成的貢獻。

1.14.2 做得好的（**What Worked Well**）

1. 顛覆性提問催生最原創的章節：使用者在第一輪論文完成後提出「學習的定義本身也在變」，這個洞察不是修辭性的反問，而是一個具有認識論深度的挑戰。它直接催生了 §4.7（共同演化觀點），引入 Clark & Chalmers 的延伸心智理論、Hayles 的後人類主義等 8 篇新文獻，成為論文被審稿人評為「最具原創性」的章節。這是典型的「人類提供 what to think about，AI 提供 how to think about it」的最佳協作

模式。

2. 不妥協的品質標準：使用者堅持 PDF 必須從 LaTeX 編譯（而非 HTML-to-PDF）、格式必須符合 APA 7.0 規範、中文版必須使用台灣教育學界的學術語感。這些要求看似是「格式問題」，但實際上確保了最終產出達到可提交的學術水準。原文指示：「PDF 必須從 LaTeX 編譯」——這個堅持避免了大量排版問題。

3. 願意重跑整個 **pipeline**：在新增共同演化觀點後，使用者選擇重走完整的 pipeline（Stage 2.5 → 3 → 4 → 3' → 4.5 → 5），而不是只做局部修補。這個決定的代價是額外 2-3 小時的工作量，但確保了論文的結構完整性和引用一致性。

4. **meta-cognitive** 反思能力：使用者不僅完成了論文，還要求記錄整個過程、更新 SKILL.md 新增 Stage 6、加入協作品質評估機制。這種「從做事到做流程」的層次提升，使整個 pipeline 的設計得到了實質性改進。

1.14.3 錯過的機會（**Missed Opportunities**）

1. **Checkpoint** 深度不足：在多數 Stage 完成的 checkpoint，使用者的回應是「繼續」或「1」（選擇選項）。雖然這提高了效率，但也意味著某些可以介入調整的時刻被略過了。例如，Stage 3 的模擬審查結論出來時，使用者可以對審稿人的具體建議提出異議或補充脈絡——這是真正的學術對話，而非只是接受所有修訂建議。

2. 未在研究階段提供領域內部知識：使用者在 Stage 1（研究階段）未主動分享自己對台灣高教品保的一手經驗或內部觀點。作為 HEEACT 的實務工作者，使用者擁有 AI 無法透過文獻搜尋取得的 tacit knowledge——例如第四週期評鑑的內部討論方向、實務執行中的真實瓶頸。如果在研究階段就注入這些知識，論文的政策建議可能更具實務可行性。

3. 中文版校閱投入不足：使用者對中文版的語感要求是正確的（「要用台灣教育界慣用的學術語氣」），但未在中文版完成後投入足夠的校閱時間。作為母語使用者，使用者是中文語感的最終裁判——AI 可以模仿學術語氣，但無法取代母語者的語感直覺。

1.14.4 下次建議（**Recommendations for Next Time**）

1. 在研究階段就帶入你的獨特知識：你擁有 AI 無法取得的一手經驗。在 Stage 1 開始時，花 5-10 分鐘寫下「只有我知道、文獻裡找不到」的觀察——這些通常會成為論文最具價值的部分。

2. 挑戰審稿人的建議：不是所有審稿意見都必須接受。當你不同意某個建議時，說出原因——這種對話能讓修訂更有方向性，避免「為改而改」。

3. 在 **checkpoint** 時問自己三個問題：(a) 有什麼是 AI 寫不出來的嗎？(b) 這個方向讓我興奮嗎？(c) 如果我是這個主題的審稿人，我會怎麼評？這三個問題能將

checkpoint 從「通過/不通過」變成「深化/轉向」的機會。

4. 為中文版分配專門的校閱 **session**：不要在英文版完成後才「順便翻譯」。中文學術寫作值得獨立的深度工作時間，尤其是當目標讀者是台灣的教育政策制定者時。

5. 記錄你的思考過程，不只是決策結果：你的「學習定義也在變」洞察之所以強大，是因為我們知道它的脈絡。但在某些 checkpoint，你的決策理由沒有被記錄。下次在做重要決定時，花一句話說明「為什麼」——這對 AI 理解你的意圖、對未來回顧過程都有價值。

1.14.5 人類 vs AI 的價值貢獻（Human vs AI Value-Add）

最終論文品質	來源	不可替代性
§4.7 共同演化觀點（最受好評章節）	使用者提問 → AI 理論化	使用者不可替代——AI 不會自發提出此問題
政策建議的台灣在地性	AI 研究 + 使用者方向指引	使用者部分可替代——但一手經驗可增值
APA 7.0 + LaTeX 品質	使用者品質要求 → AI 執行	使用者不可替代——AI 預設不會選擇最嚴格標準
62 篇參考文獻的驗證	AI 執行	AI 不可替代——人工逐一驗證不現實
10 份模擬審查報告	AI 執行	AI 不可替代——人類無法扮演 5 個獨立角色
中文學術語感	使用者要求 → AI 執行	部分可替代——但母語者校閱仍不可省

結論：本次協作中，使用者的最大價值不在於寫了多少字，而在於兩個關鍵時刻——(1) 提出共同演化的認識論挑戰，(2) 堅持重跑完整 pipeline 而非局部修補。這兩個決定使最終論文超越了「AI 生成的綜述」層次，進入了「有原創理論貢獻」的領域。這是使用者的 95 分智識貢獻最直接的體現。

1.14.6 Claude 的自我反思

公平起見，也坦誠指出 AI 在本次過程中的不足：

1. 中文版排版問題需多次修正：LaTeX 的 xeCJK 配置和 pandoc 的 `\LTcaptype` 問題耗費了不必要的 debug 時間，這些應該在第一次就處理正確。
2. 部分 **checkpoint** 通知過於冗長：checkpoint 模板的資訊量有時超過使用者需要的，降低了互動效率。

3. 未主動建議共同演化觀點：§4.7 的核心洞察來自使用者，AI 在兩輪完整 pipeline 中都未自發提出「學習定義的演化」這個視角——這暴露了 AI 在突破既有框架方面的侷限。