

學生與生成式人工智慧共作《紅樓夢》人物圖像之多模態學習歷程分析——以賈探春圖像生成為例

A Multimodal Learning Process Analysis of Student-Generative AI Co-creation of Character Imagery from Dream of the Red Chamber: A Case Study on the Generation of Jia Tanchun's Image

李瑄^{1*}, 林伊玲², 張玉明³

^{1,2} 金門大學華語文學系

³ 廈門工學院博雅教育與藝術傳媒學院

hsuanli@email.nqu.edu.tw

【摘要】 本研究探討大學生在數位人文課程中，透過生成式人工智慧（Generative Artificial Intelligence, GAI）共作《紅樓夢》人物圖像的學習歷程。研究以學生與 ChatGPT 共構「賈探春雪中著大紅猩猩氈」形象為例，分析其多模態提示語操作、文化補綴與圖像修正策略。研究採質性分析法，蒐集學生互動語料與後設反思，歸納其提示語策略發展與角色理解歷程。結果顯示，學生能從初始模糊描述逐步精緻化語言策略，主動補充文化資訊、修正生成誤圖，最終生成具文學詮釋深度的圖像作品。此共作歷程有助於學生發展提示素養與文化理解能力，亦對教師設計跨域課程與生成式 AI 教學具實務啟示。

【關鍵字】 生成式人工智慧；多模態學習；提示素養；ChatGPT；紅樓夢

Abstract: This study explores how students co-construct a literary character image from Dream of the Red Chamber with Generative AI. Focusing on the creation of Jia Tanchun dressed in a crimson Qing dynasty winter cloak (xingxing felt) in the snow, it analyzes students' multimodal prompts and revision strategies. Findings highlight students' development of prompt literacy and cultural interpretation through iterative AI interaction.

Keywords: Generative AI, Multimodal Learning, Prompt Literacy, ChatGPT, Dream of the Red Chamber

1. 前言

生成式人工智慧（Generative Artificial Intelligence, GAI，以下簡稱生成式 AI）與多模態學習（Multimodal Learning）近年在語文學習中快速發展，有助學生深化文化理解與語言素養（Liu et al., 2025）。模態（mode）係指人們用以傳達意義的符號系統與表達形式，包含語言、圖像、物件、比喻等文化性意義資源，多模態素養被視為多元識讀（multiliteracies）的一環，強調學生在數位與文化脈絡中靈活應用各種符號資源進行表達與詮釋（Bezemer & Kress, 2016; Cope & Kalantzis, 2016）。然而，多數研究聚焦於教師設計的單一模態任務，對學生如何透過多輪互動提示 AI 並建構文學角色形象之歷程仍較少探討（Lee & Hannafin, 2016）。

《紅樓夢》可推估其文化氛圍對應清朝初年，書中細膩描寫角色心理、服飾與社會階級，判詞與判曲象徵人物命運，提供 AI 共作圖像的文化詮釋依據（朱蕙君等人，2024）。賈探春

為書中四春之三，聰穎果敢，群芳夜宴上抽到杏花籤寓其朝氣蓬勃，敢為人先的凜然姿態，判詞與判曲反映其精明幹練卻終難掌握命運的悲劇宿命，象徵封建時代中女性的才志與侷限。

本研究奠基於多模態學習觀點，聚焦學生與 ChatGPT 共作「賈探春雪中著大紅猩猩氅」圖像歷程，分析其如何透過多模態提示語修正、文化補綴與視覺詮釋，逐步建構角色形象與文化意義。研究問題如下：

- (1) 學生如何透過多模態提示與 ChatGPT 互動，建構賈探春圖像？
- (2) 學生在共作歷程中展現出哪些提示素養與文化詮釋行為？

2. 研究方法

研究資料來自臺灣某大學華語文學系數位人文課程成果，學生以《紅樓夢》人物為題，結合生成式 AI 進行圖像共作。本文選擇一組具代表性的案例「賈探春雪中著大紅猩猩氅」進行質性分析。研究資料包含該組學生與 ChatGPT 之完整互動紀錄，含文字與圖像。文字約 14,540 字、共 103 輪對話，圖像計 115 幅，其中 32 幅為學生策略性提供。另外輔以任務後回顧式訪談（約 40 分鐘）。語料第 51 至 93 句展現提示語與圖像回應密集互動歷程，為本研究分析核心。共作學生為兩位華文系大四學生（S1、S2）。S1 具空間構圖經驗，自評美術技能 2 分、美感 1 分；S2 曾學習壓克力畫，自評技能 3 分、美感 4 分（皆為五點量表）。兩人皆無電腦繪圖背景，主要透過語言與 AI 互動進行圖像共作與策略調整。

研究採質性分析，參考 Bezemer 與 Kress（2016）提出之多模態學習觀為架構，探討學生如何運用語言、圖像與文化資源引導 AI 協同生成，並展現提示素養與文化詮釋能力。

3. 研究結果與討論

本研究透過質性分析，呈現學生在生成式 AI 共作中的提示語發展與文化理解歷程：

3.1. 學生提示語歷程與多模態操作策略（對應研究問題一）

學生在與 ChatGPT 共作圖像的歷程中，展現出從模糊輸入到精準提示的策略性發展。初始階段，學生僅以「想生成有探春感覺的圖像」作為起點，輸入如《紅樓夢》判詞、判曲文字，以及影視作品的截圖或描述作為視覺風格參考。AI 回應後，學生透過持續多輪互動修正提示語，逐步建立如「大紅猩猩氅」、「觀音兜」、「雪景背景」等關鍵視覺元素，提示語策略逐漸從模糊描述轉為多層次細化。

在語料第 51 至 93 句中，學生展現出連續修正與多模態提示操作的能力。面對 AI 生成誤差（如帽子形狀錯誤、臉部比例不符等），學生使用形容詞強化、語序調整與生活語彙進行語意修正。例如以「像帽 T 把鬆緊帶拉緊一樣」的比喻協助 AI 理解兜帽貼合臉部的設計需求，並反覆釐清「帽子不是兜帽」的語義邊界。

根據訪談內容，學生亦提及在構思前即已有朦朧的視覺畫面，再以語言將圖像分項描述，進行視覺轉語言的表達轉譯（S1）；另一位學生則強調語言與圖像需相互建構，並透過圖像與語言交替描述以更精準傳達想法（S2）。學生也提到有時語言難以完全表達，需透過圖像輔助補充，說明其提示語建構實際涉及多感官模態整合。

綜合而言，學生從最初模糊的語意輸入出發，歷經視覺語言轉譯、語句細化與比喻補強等策略歷程，不僅展現其多模態協同操作的能力，也顯示其提示素養已由感知型描述逐步進展為具文化與視覺邏輯的引導型策略。

3.2. 學生文化詮釋與角色理解發展（對應研究問題二）

在與 AI 互動過程中，學生展現主動文化考據與詮釋意識。除了詢問猩猩氅、盤金繡、觀音兜等清代服飾的歷史特徵，學生亦查閱《紅樓夢》文本細節，辨識角色所代表的文化意涵。

當 AI 無法正確生成特定物件（如觀音兜）時，學生會提供劇照、相關圖片，甚至拍攝生活中可參照的服飾照片補餵給 AI，強化其視覺理解。這些文化補餵行為使生成圖像逐漸貼近清代語境與角色原貌，顯示學生已能靈活轉化文本知識與視覺再現。

訪談中，學生表示其文化知識背景亦影響提示語選擇與識別能力，例如 S1 因家庭信仰熟悉觀音形象，理解「觀音兜」的造型意象；S2 則來自日常觀影經驗判斷兜帽風格。學生亦會在 AI 生成錯誤後，查找文獻或討論進行交叉確認與修正，逐步建立文化正確性的審查意識。

最終生成之探春圖像融合文本描寫、文化符號與學生詮釋：角色身著繡杏花的大紅猩猩氈，頭戴觀音兜、姿態凜然，背景為大雪紛飛，側身前行象徵其遠嫁邊疆、命運難違的情境。學生指出，此圖像回應了判詞與判曲所呈現的宿命，也體現探春精明幹練但無可奈何的氣質。任務後，學生 S1 與 S2 在角色理解自評上分別從 1 與 2 分提升至 3 與 4 分（五點量表），圖像生成歷程促使其更深入查閱文本與服飾資料，將語言理解轉化為視覺詮釋與文化重構。



圖 1

學生與 ChatGPT 共作生成之探春圖像成果：著大紅猩猩氈、頭戴觀音兜，融合清初服飾與原著判詞詮釋，背景雪景經 AI 生成與後製合成，象徵其遠嫁邊疆與命運蕭瑟；判詞文字亦為後續視覺設計時加入。（圖像來源：學生共作成果）

3.3. 模態整合策略與角色轉變（綜合回應研究問題一與二）

語料第 51 至 93 句為學生嘗試生成賈探春頭戴觀音兜的冬季造型，呈現學生提示語策略由初階描述向高階調控的發展歷程。初期（第 51 至 74 句），學生以輸入觀音菩薩圖片為參照，搭配語言描述服飾與臉型特徵（如「兜帽要蓋住耳朵」、「不要露出瀏海」）協助 AI 生成貼近文化背景之造型。中後期（第 75 至 93 句）學生進一步釐清語義邊界，針對帽兜樣式明確劃分詞彙使用（如「帽子不是兜帽」），並出現情緒語句（如第 91 句學生再次餵補一張兜帽圖片標準樣式後說明：「你再出現一次帽子我就生氣」），強化生成風格規範與語義控制，顯示其角色已由使用者轉化為圖像風格的審美制定者與語義引導者。這一角色的轉變，顯示學生不僅具備生成意識，也具備「生成監督」的能力，在多模態提示過程中逐步建構 AI 理解的準則與文化框架。學生亦會在 AI 生成失誤時採取行動修正，包括直接提供兜帽照片、進行語言補述，或在必要時手動使用軟體進行圖像調整。此外，學生亦曾將生成結果交由教師或紅

樓夢專家協助判斷，強化其生成作品的文化正確性與審核意識。表 1 整理語料第 51 至 93 句所展現的模態資源與策略。模態定義與分類依循多模態學習觀點，視語言、圖像、文化物件與身體動作等為建構意義的符號資源（Jewitt et al., 2016）。

整體互動歷程展現學生從初階生成提示，逐步發展出具有文化意涵與語意精準度的高階策略。其操作不僅體現在語言上的調整，更涉及圖像補綴與符號再設計（re-design）行為，呼應 Bhutoria（2022）所指出，學生在 AI 互動中可由被動生成者轉變為主動引導者，並展現視覺、語言與文化之協同設計與審美調控能力。此結果亦與 Liu et al.（2025）所提出的 GAI 素養架構相呼應，其將生成式 AI 學習歷程中的學生能力分為情意、行為、認知與倫理四個面向。本研究呈現學生語言細化、視覺判斷、文化轉譯與術語釐清等行為，具體展現出其在多模態協同生成中逐步發展的認知調整與文化意識，並為教學實務提供可觀察的策略參照。

表 1. 學生 AI 共作歷程中的模態—行為—策略對應整理（語料第 51–93 句）

模態資源	學生行為類型	對應提示策略
語言提示（文字）	逐步描述服飾細節、語序調整、否定詞使用	語意細化、邊界釐清、語意加強
AI 圖像回應	回應錯誤辨識、選擇性接受、要求再生	結果比對、風格修正、偏好調整
圖片補綴（靜態圖）	補充觀音兜樣式、提供兜帽照片	圖像餵補、文化澄清、概念再建構
生活語彙與比喻	使用「帽 T 鬆緊帶」等日常經驗進行視覺轉譯	視覺語言轉化、感知連結
語意規範與命名控管	嚴格區辨「帽子」與「兜帽」詞彙，明確拒絕誤稱	概念規範、語言控制、術語修正

4. 結論

本研究回應兩項研究問題。針對研究問題一，學生透過多模態提示與 ChatGPT 展開圖像共作，歷經從模糊概念到細緻視覺元素建構的過程，展現逐步調整與策略性指令運用的能力。針對研究問題二，學生在互動歷程中展現出文化補充、歷史考據、圖像錯誤修正與物件逐一澄清等提示素養與文化詮釋行為，並透過後設反思回顧自身學習歷程。最終生成之探春圖像呈現高層次文學詮釋，學生亦從角色使用者轉變為 AI 生成邏輯的審核與指導者。教師若能設計兼具文化深度與生成彈性的任務，將有助於學生發展多模態素養與 AI 理解力，呼應近期 GAI 素養研究所強調的提示精煉、文化意識與倫理操作等層面（Liu et al., 2025）。

本研究以單一共作歷程為起點，未來將延伸進行跨角色與跨任務的互動比較分析，發展提示素養量表與教學應用模式，擴展學生與 AI 共作之評估工具與文化詮釋能力的理論建構。

參考文獻

朱蕙君、許佳穎、王俊傑（2024）。生成式 AI 繪圖應用於古詩文學課程對高中生學習表現之影響。收錄於黃國禎、涂芸芳（編），*生成式 AI：融入教育的理論、策略與研究設計*（頁 203-228）。華藝數位。

Bezemer, J., & Kress, G. (2016). *Multimodality, learning and communication: A social semiotic frame*. Routledge.

- Bhutoria, A. (2022). Personalized education and Artificial Intelligence in the United States, China, and India: A systematic review using a Human-In-The-Loop model. *Computers and Education: Artificial Intelligence*, 3, 100068. <https://doi.org/10.1016/j.caeai.2022.100068>
- Cope, B., & Kalantzis, M. (2016). The Things You Do to Know: An Introduction to the Pedagogy of Multiliteracies. In B. Cope, & M. Kalantzis (Eds.), *A Pedagogy of Multiliteracies: Learning by Design* (pp. 1-36). Palgrave Macmillan. https://doi.org/10.1057/9781137539724_1
- Jewitt, C., Bezemer, J., & O'Halloran, K. L. (2016). *Introducing multimodality*. Routledge.
- Lee, E., Hannafin, M.J. (2016). A design framework for enhancing engagement in student-centered learning: own it, learn it, and share it. *Education Tech Research Dev*, 64, 707–734. <https://doi.org/10.1007/s11423-015-9422-5>
- Liu, M., Zhang, L. J., & Zhang, D. (2025). Enhancing student GAI literacy in digital multimodal composing through development and validation of a scale. *Computers in Human Behavior*, 166, 108569. <https://doi.org/10.1016/j.chb.2025.108569>