國立臺灣大學工學院工業工程學研究所

碩士論文

Institute of Industrial Engineering

College of Engineering

National Taiwan University

Master's Thesis

生成式 AI 之虛擬人格設計工程及其在市場研究之應用策略探析

AI Personas Engineering with Large Language Models: Strategies for Market Research Applications

邵懿文

Yi-Wen Freda Shao

指導教授：藍俊宏 博士

Advisor: Jakey Blue, Ph.D.

中華民國 114 年 6 月

June 2025

doi:10.6342/NTU202501772

誌謝

完成本論文的旅程中，衷心感謝我的指導教授藍俊宏博士。何其有幸，教授對這個主題抱持著高度興趣，在研究過程中的專業指引，討論與互動，不僅啟發了我的思考脈絡，更在研究設計的嚴謹性、理論架構的完整性，以及撰寫的邏輯連貫上，賦予我紮實的基礎，沒有他，這篇論文只會是一堆華麗辭藻包裹的空洞理論。

此外，這不是我一個人的成果。研究團隊還有工作上的好戰友，感謝陳介立、胡庭瑋及李孟翰三位夥伴的寶貴貢獻，與我一起在人工智慧的星際中航行，讓這個研究從初步的構想變成了有理有據的論點。從文獻資料蒐集、實驗設計規劃、到案例分析驗證，團隊的專業投入，成就了這份論文的研究成果。

感謝我現職服務的公司—電通行銷傳播集團，提供了豐富的去識別化資料與開放創新的環境。就像是贊助了一場知識的冒險，卻不知道結局可能只是一篇被引用三次的論文。然而這個寬容的支持，卻能使我能夠突破框架，站在學術與業界實務的交界處，持續深入探索 AI 行銷科技的創新應用。

縱使身處亂世，未來難測，但正如王爾德所言：「即使我們生活在陰溝裡，仍有人仰望星空。」期盼能將本研究所累積的經驗與發現，投入到市場，創造一些有趣的東西，轉化為產業應用的價值，為行銷領域在 AI 的發展突貢獻一己之力。這不是結束，而是新知識與新應用最好的起點。

doi:10.6342/NTU202501772

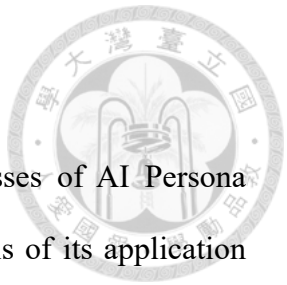
中文摘要

本研究旨在探討生成式人工智慧（Generative AI）於虛擬人格（AI Persona）建模之設計工程流程，並深入分析其在市場研究領域的應用策略與實務發展潛力。研究以 Henry Murray（1938）提出之心理需求理論為核心理論架構，建構六類動機向度（成就、感情、資訊、佔有、權力、地位/保護），結合電通行銷傳播集團 CCS（Consumer Connection System）調查去識別化資料，運用提示詞工程（Prompt Engineering）技術，生成具備特定人格特徵的 AI Persona 敘事內容，並進行語義一致性與動機再現性之系統化分析。研究結果證實，在結構化提示設計與精確資料配置條件下，AI Persona 能夠有效模擬真實消費者之動機特徵表現，從而驗證本研究之核心假設。

然而，本研究發現語言模型在處理抽象或具爭議性的人格動機面向（如地位追求與權力展現）時，表現仍存在明顯侷限，特別是在負面語意表達上呈現系統性偏誤，例如傾向性誤判等現象。相較於 McCrae & Costa（1987）所提出的大五人格模型之統計可區辨的五維特質架構，其五個連續性特質向度在語言模型中較易透過語言風格或表達習慣進行建模與區辨；Murray（1938）所建構之心理需求理論則涉及情境觸發與內在動機的交互關係，具有高度語境依賴性與需求重疊性。為提升語言模型對動機性人格的精確模擬能力，本研究提出三項改進策略：其一，設計具情境深度的提示詞場景以誘發目標動機；其二，建構動機導向的語言特徵詞庫以輔助模型識別；其三，採用混合式人格建模架構，以結合動機需求與特質性格的表徵優勢。綜合而言，生成式 AI 結合心理學理論與實證調查資料之虛擬人格設計工程，不僅為市場研究提供高效率且具創新性的洞察分析工具，更為 AI 驅動之合成資料應用與消費行為模擬研究領域建立嶄新的研究典範。

關鍵詞：虛擬人格、Murray 心理需求理論、市場研究調查、大型語言模型、提示詞工程

doi:10.6342/NTU202501772



Abstract

This study aims to investigate the design engineering processes of AI Persona modeling through Generative AI and to conduct an in-depth analysis of its application strategies and practical development potential in the field of market research. Grounded in Henry Murray's (1938) psychological needs theory as the core theoretical fundamental, this research constructs six motivational dimensions (achievement, affection, information, acquisition, power, and status/protection), integrating the survey data from Dentsu Group Inc. Through proper prompt engineering techniques, AI Persona narratives with specific personality characteristics were generated, followed by systematic analysis of semantic consistency and motivational reproducibility. The findings confirm that under structured prompt design and data configuration, AI Personas can effectively simulate the motivational characteristics of real consumers, thereby validating the core hypothesis of this study.

Furthermore, the study reveals that language models exhibit performance limitations when processing abstract or controversial motivational dimensions (such as status pursuit and power manifestation), while simultaneously demonstrating systematic bias in expressing negative semantic content. Compared to the bipolar dimensional structure of the Big Five personality model, Murray's psychological needs theory faces challenges including high contextual dependency and technical difficulties in feature separation during the language mapping. Based on these findings, this research proposes three optimization strategies: contextualized prompt design, systematic construction of linguistic feature lexicons, and hybrid personality modeling, to further enhance the precise representational capabilities of language models for motivational personalities.

In conclusion, the virtual persona design engineering that combines generative AI with psychological theories and empirical survey data not only provides market research

doi:10.6342/NTU202501772

with highly efficient and innovative analytical tools for insights, but also establishes a novel research paradigm for AI-driven synthetic data applications and consumer behavior simulation research fields.

Keywords: AI persona, Murray's psychogenic needs theory, market research, large language models, prompt engineering

doi:10.6342/NTU202501772

目次

doi:10.6342/NTU202501772

doi:10.6342/NTU202501772

圖次

doi:10.6342/NTU202501772

doi:10.6342/NTU202501772

表次

doi:10.6342/NTU202501772

doi:10.6342/NTU202501772

第一章 緒論

1.1 研究背景與動機

當代行銷環境正經歷前所未有的數位轉型挑戰，企業必須在滿足顧客快速升級的期望同時，有效應對資料爆炸性增長的市場現實。傳統市場研究方法如問卷調查、深度訪談與焦點團體雖然長期構成了行銷決策的基石，但這些方法正面臨嚴峻挑戰：回應率持續下滑、獲取有效樣本成本攀升，以及資料時效性無法滿足瞬息萬變的市場需求（Malhotra, 2020）。在消費者旅程（Customer Journey）日益複雜且碎片化的時代，企業迫切需要更敏捷、更具成本效益的市場洞察工具，以保持競爭優勢。

近年來，生成式人工智慧（Generative AI）技術的突破性發展，特別是大型語言模型（Large Language Models, LLMs）的崛起，為行銷研究領域帶來了革命性的轉變契機。在這些技術進步的基礎上，「AI 虛擬人格」（AI Persona）已從理論概念逐步走向實際應用，成為模擬特定消費者群體行為的創新工具。透過精確設定人格特質、消費動機與情境因素，AI Persona 能夠生成具有內在一致性的消費者反應，為產品測試、品牌溝通策略與市場洞察提供富有價值的即時資料（Arora et al., 2025；Bisbee et al., 2024）。這不僅是量化研究的進步，更代表了市場洞察方法的典範轉移。

波士頓諮詢公司的一項研究（Boston Consulting Group, 2023）表明，目前70%的行銷長（CMO）都在使用生成式人工智慧，其中洞察生成是繼個人化之後第二受歡迎的應用。這種廣泛的應用突顯了企業越來越察覺到生成式人工智慧對行銷策略的價值，將其納入市場研究可能會改變洞察力的產生和應用方式，使企業能夠在競爭激烈的市場環境中做出更快速且數據驅動的決策。

然而，AI Persona 的應用價值與限制仍需嚴謹評估。一方面，Liu et al.（2025）已證實 AI 模型在模擬特定消費者偏好與情境反應上展現了顯著潛力。與此同時，在美國行銷協會旗艦期刊最新發表的研究中 Arora et al.（2025）揭示了將

doi:10.6342/NTU202501772

大型語言模型整合至行銷研究流程的實質性價值，證明在研究設計、樣本選取、資料收集與分析等階段，AI 技術能顯著提升研究效率與洞察深度。這些發現不僅表明 AI 虛擬人格已超越概念階段，更顯示其作為市場洞察革新工具的戰略意義。

另一方面，儘管 AI Persona 技術展現出破壞性創新的潛力，其在實際應用中仍面臨多重實踐挑戰。核心問題在於其驗證效度：AI 生成的虛擬消費者反應能否真實反映目標市場的真實態度、偏好與行為模式？這涉及內容效度、構念效度、準則效度與外部效度等多個關鍵維度（Argyle et al., 2023）。Argyle et al.（2023）運用 GPT-3 模型生成「矽樣本」（silicon samples）來模擬美國選民調查資料，結果顯示 AI 生成的樣本在統計分布上與真實受訪者相當接近；然而，Bisbee et al.（2024）的研究則揭示這些 AI 受訪者在反應變異性與情感細微表達上存在系統性差異，表現為回應同質性較高且對複雜情境的理解有限。更值得注意的是，Salecha et al.（2023）發現，大型語言模型在回答人格測驗時往往產生理想化的自我呈現，其社會期望偏差程度甚至超過了一般人類受訪者，這揭示 AI Persona 可能傾向於生成符合研究者預期的「完美答案」，而非真實消費者可能展現的多元化反應。這種理性平衡的觀點促使我們必須深入探究 AI Persona 的效度問題，確保其能為行銷決策提供真正有價值的洞察。

從商業化的視角來看，全球領先的市場研究機構已開始積極整合 AI 技術，作為傳統消費者研究方法的戰略性輔助。市場研究行銷機構 Kantar（2024）推出的生成式 AI 驅動產品「ConceptEvaluate AI」能在 24 小時內評估多達 100 個創新概念，遠超傳統方法的效率。該工具基於超過 40,000 筆實際產品測試資料訓練，為品牌提供迅速而可靠的概念測試結果。同時，Kantar（2024）的「Link AI」廣告效果預測系統能在 15 分鐘內完成視覺與文字廣告的效果評估，大幅提升品牌溝通策略的優化週期。這些創新應用不僅展示了 AI 在市場洞察領域的突破性效率，也說明 AI Persona 正逐步整合至產品開發與品牌管理的核心決策流程。

doi:10.6342/NTU202501772

在臺灣市場環境下，AI Persona 技術的應用仍處於早期階段，相關本土研究明顯不足。作為全球科技產業的重要樞紐，臺灣既面臨數位轉型的巨大機遇，也擁有獨特的市場需求與文化脈絡。因此，深入探討 AI Persona 在臺灣行銷研究中的效度與應用潛力，不僅具有學術創新價值，更能為本土企業提供實質的競爭優勢：一方面填補北亞市場在 AI 市場研究應用上的知識缺口，驗證這項技術在不同文化背景下的適用性；另一方面為臺灣企業提供創新研究工具的實施指南，加速 AI 技術與行銷洞察的戰略整合。

綜上所述，本研究旨在系統評估生成式 AI 之虛擬人格在市場調查中的設計工程、效度問題及應用策略挑戰，深入比較 AI Persona 與傳統市場研究方法的優劣互補關係，並探索整合兩者的最佳實踐路徑。透過整合最新文獻與實證資料，本研究將為行銷研究方法的創新發展提供理論基礎，同時為企業應用 AI Persona 技術提出具體且可操作的策略建議。

doi:10.6342/NTU202501772

1.2 研究目的與限制

隨著大型語言模型（Large Language Models, LLMs）技術的快速演進，AI Persona 將被越來越多企業視為模擬消費者行為與偏好的高效工具，並在市場研究中作為節省時間、降低成本的戰略性選項。然而，這項技術的有效性、適用情境與潛在風險仍需嚴謹的實證探討。

本研究致力於探究以傳統市場調查所獲得的消費者資料作為資料基礎，建立 AI 虛擬人格（AI Persona）的可行性，並評估其在行銷研究中的效度與實務應用價值。具體而言，本研究將系統性回答以下核心問題：

- 傳統市場調查資料作為人格模型預訓練資料的可行性：

  深入評估傳統市場調查中收集的多維資料（包括人口統計特徵、消費行為模式、品牌認知與態度指標）作為 AI Persona 資料基礎的可行性與有效性，特別關注這些資料在建構虛擬消費者人格時的代表性、完整性與適配性。（參考 Malhotra, 2020；Kumar et al., 2022）

- 虛擬人格建構的關鍵設定因子：

  分析成功建立 AI 虛擬受訪者所必須整合的核心資料類型，探討除基本人口統計指標外，消費者心理動機、核心價值觀及品牌認知等深層變量如何共同影響虛擬人格的真實性與預測價值。

- Murray 人格理論的適用性：

  檢驗 Murray 提出的基本人類需求框架（如成就需求、親和需求、權力需求等）如何在 AI 虛擬人格建構過程中發揮根本性作用，評估這些潛在需求向度能否有效豐富 AI Persona 的人格維度，從而提升其在模擬真實消費決策時的準確性與洞察價值。（Murray, 1938；McCrae & Costa, 1987）

- 提示工程（Prompt Engineering）之效度與價值觀評估：

  提示詞是人與大模型交互的關鍵語言，AI 回應的水準也因提示詞的水準而定。提示工程是一個系統性工作，其目標是透過不斷優化提示詞，使大

doi:10.6342/NTU202501772

模型能夠更準確地理解用戶的需求。因此，必須深入分析精確設計的提示策略如何引導 AI Persona 生成能夠真實反映目標消費群價值觀與行為傾向的回應，並系統比較這些生成回應與傳統調查中真實受訪者反饋的一致性與差異性。（Arora et al., 2025；Bisbee et al., 2024）

- 市場洞察工具的戰略定位：

  全面探討 AI Persona 作為新型市場洞察工具的戰略角色與應用場景，評估其如何為品牌定位、產品創新與消費者理解創造差異化價值，同時明確界定其應用限制，為企業決策制定提供更具前瞻性的參考框架。（Huang & Rust, 2021）

- 多維效度評估系統的建立：

  建構一套涵蓋內容效度、構念效度、準則效度與外部效度的綜合評估框架，並透過對比分析、預測測試與專家評估等多元方法，系統檢驗 AI Persona 在不同行銷研究情境中的模擬效能與應用價值。（Eichstaedt et al., 2023；Liu et al., 2025）

基於前述內容，本研究不僅旨在評估傳統市場調查資料作為 AI Persona 人格核心動機塑造的適用性，更致力建構一套 AI 虛擬人格建模之設計工程流程，評估其在市場研究應用中的效度與策略潛力。研究將以電通行銷傳播集團的之去識別化 CCS（Consumer Connection System）消費者問卷資料為實證基礎，結合心理動機分類與提示詞工程技術，探討生成式 AI 在模擬不同動機向度的能力，並分析其人格一致性、語言特徵對應性與模擬效能。透過本研究，期能為 AI Persona 在行銷研究中的應用模式提供方法論支持，亦拓展生成式 AI 在心理理論結合與合成資料生成上的實務貢獻。

由於本研究採用 Murray（1938）所提出之心理需求理論作為人格建模的核心理論基礎。該理論提供一套有組織的動機分類架構，有助於 AI Persona 動機型人格的建立。然而，由於 Murray 理論本身高度抽象，且其動機表徵並非建立於語言維度上，部分動機類型（如地位、權利等）是否能透過語言模型有效模擬，仍具一定

doi:10.6342/NTU202501772

挑戰。特別是在生成式 AI 的語言輸出中，這類隱性或社會敏感的動機是否能具體呈現，亦為本研究希望探討的重要課題之一。

此外，由於生成式 AI 技術仍處於快速演進階段，本研究之分析架構與實證設計可能無法即時涵蓋最新的業界實踐與語言模型進展，其結論具有一定時效性，亦需隨技術發展持續更新與檢驗。本研究聚焦於 AI Persona 在行銷研究場域的應用價值，相關發現不直接涵蓋其他領域（如教育、醫療、公共政策等）之應用情境。

同時，本研究所採用之資料與語言模型訓練基礎以台灣語境為主，其分析結果在文化背景相近之市場環境具參考性；然而，對於語言體系與價值觀念差異較大的市場，其應用效度與策略設計仍須進一步在地化驗證與調整。

doi:10.6342/NTU202501772

第二章 文獻探討

2.1 行銷研究方法的歷史發展與演變

行銷研究作為系統收集、分析與解讀市場資訊的策略性過程，其方法論在上個世紀經歷了深刻而持續的演進。正如 Malhotra（2020）與 Malhotra & Dash（2019）所闡述，行銷研究方法的發展可劃分為以下關鍵階段，每一階段均反映了當時的技術能力、市場需求與理論進步：

- **早期萌芽階段（1900-1940 年代）**

行銷研究的根源可追溯至 20 世紀初期，當時主要依賴基礎統計分析與初步市場調查。哈佛商學院的 Melvin T. Copeland 於 1924 年出版的《商品銷售原理》（Principles of Merchandising）曾是市場行銷領域中最受評論的著作之一，試圖透過制定消費者購買習慣和消費者購買動機，為行銷理論奠定基礎。這一時期的研究方法主要集中於基本人口統計資料收集與商品銷售記錄分析，方法相對簡單，缺乏系統性與科學嚴謹性。值得注意的是，心理學理論開始嶄露頭角，Murray（1938）提出了以動機和需求為基礎的人格理論，為後續深入探索消費者內在動機奠定了重要基礎，這也為後來的行銷研究方法論發展埋下了重要伏筆。

- **專業化發展階段（1950-1970 年代）**

第二次世界大戰後，隨著全球消費市場的快速擴張，行銷研究進入了專業化發展階段。這一時期以問卷調查、深度訪談及焦點團體等方法為主要工具，並開始系統性整合心理學理論，以深入探索消費者潛在動機與行為模式。Ernest Dichter 等開創性學者通過引入投射技術（projective techniques），開始揭示消費者潛意識中的隱藏需求與動機（Schwarzkopf, 2015）。同時，Malhotra（2020）等前瞻性學者系統化行銷研究方法，倡導定性與定量方法的戰略整合，使市場調查能更準確捕捉消費者心理與行為特徵的多維度本質。

這一階段的研究方法呈現出明顯的多元化與專業化特徵，質化與量化技術均獲得了突破性進展。心理學方法的深度整合使研究者開始系統性關注受訪者偏差

doi:10.6342/NTU202501772

等方法論挑戰，為後續 AI 模擬消費者人格技術處理奠定了理論與實踐基礎。

- **電腦化與標準化階段（1980-2000 年代）**

1980 年代，隨著電腦技術與資訊系統的迅猛發展，行銷研究方法進入電腦化與標準化階段。統計分析軟體如 SPSS 的廣泛應用，大幅提升了資料處理與分析的效率與深度，使研究者能對更大規模的資料進行多層次挖掘。此階段同時見證了國際標準化組織（ISO）對市場調查流程與方法的系統規範，進一步推動了行銷研究在全球範圍的一致性、可比性與可靠性（Malhotra, 2020）。

- **數位轉型階段（2000 年至今）**

21 世紀以來，網際網路、大數據與人工智慧技術的快速崛起，徹底重塑了行銷研究的方法論與實踐模式。Huang & Rust（2021）指出，生成式 AI 技術不僅極大化的提升了資料收集與分析的效率與廣度，更催生了基於即時資料與情境化分析的全新行銷研究模式。這一階段的核心特徵在於利用生成語言模型（如 BERT 與 GPT 系列），從海量數位資料中提取戰略性洞察，實現智能化、動態化的市場反應模擬與預測。在這一背景下，傳統研究方法與新興 AI 技術正加速融合，共同推動行銷研究進入更加精準、敏捷與前瞻的新時代。表 1 總結了行銷研究方法演進過程中的核心里程碑。

透過對行銷研究方法歷史發展的回顧，我們可清晰觀察到其演進軌跡：從簡單到複雜、從粗略到精細、從人工到自動化、從單一方法到整合系統的持續進步過程。這一演進不僅反映了技術能力的提升，更體現了市場需求與理論進步的共同驅動力。生成式 AI 與虛擬人格模擬作為這一歷史性發展的最新成果，不僅代表了技術創新的前沿，更體現了行銷研究理論與方法論的累積性突破與革新。

doi:10.6342/NTU202501772

表 1 行銷研究方法演進的關鍵里程碑（本研究收集、整理資料文本，後由 ChatGPT (OpenAI, 2023) 輔助完成）

時期	研究典範	主要方法工具	特徵與價值主張
1950s-1970s	傳統行銷調查	問卷調查、面訪、電話調查	以大量樣本推論整體市場，重視標準化與量化資料
1980s-1990s	定性探索與消費文化研究	焦點團體、深度訪談、民族誌	重視語境與文化，理解消費者動機與符碼意涵
2000s-2010s	資料導向與行為追蹤分析	統計分析軟體、Web Log、眼動追蹤	利用科技收集自然行為資料，強調資料即洞察
2010s-2020s	社群與數位即時資料分析	社群聆聽、文字探勘、A/B 測試	強調即時反應與多樣資料源融合，提高操作敏捷度
2023-未來	生成式 AI 與虛擬人格模擬	AI Persona、LLM 模擬、行為生成技術	模擬人格特質與反應，快速獲得情境式、主觀性洞察

doi:10.6342/NTU202501772

2.2 虛擬人格的理論基礎與發展脈絡

2.2.1 虛擬人格之定義

虛擬人格（AI Persona）代表了市場研究方法的顛覆式創新—它是一種基於大型語言模型與先進人工智慧技術精心建構的虛擬消費者代表，能夠精確模擬具有特定人口統計特徵、心理動機與行為模式的目標消費群體。這一概念的成熟既依賴於技術突破，也深受心理學與社會科學理論的深刻影響。與傳統市場調查中的人類受訪者不同，AI Persona 透過系統設定的人口統計參數與人格特質矩陣，結合 AI 模型的自然語言理解與生成能力，產生能夠準確反映特定消費者群體特徵的回應，從而實現對目標市場態度與行為的戰略性模擬。舉例而言，行銷團隊可以建立一個「25 歲、未婚女性、大學畢業、都會區居住、對社群媒體高度參與」的 AI Persona，讓其針對新產品概念或品牌訊息提供反饋，藉此預測具有類似特徵的真實消費者群體可能的市場反應，從而優化產品設計或傳播策略。

在概念定位上，如圖 1 可見，AI Persona 需要與以下幾個相關但本質不同的概念進行區分：

- 虛擬角色（Virtual Character）：主要用於遊戲、娛樂與內容產業，注重個性表演與情感共鳴，缺乏市場研究的專業背景與方法論支持，不以模擬真實消費者決策為核心目的。
- 聊天機器人（Chatbot）：專注於客戶服務或智能助理功能，通常不具備特定人口統計背景或消費行為模式的精確設定，難以作為市場調查的有效樣本或洞察來源。
- 人類樣本（Human Survey Sample）：傳統市場調查，由真實人類受訪者組成，具有真實經驗與情感反應，但資料收集成本高昂，且樣本規模與結構受到時間、成本與接觸難度等現實因素的嚴格限制（Nguyen & Park, 2024）。

doi:10.6342/NTU202501772

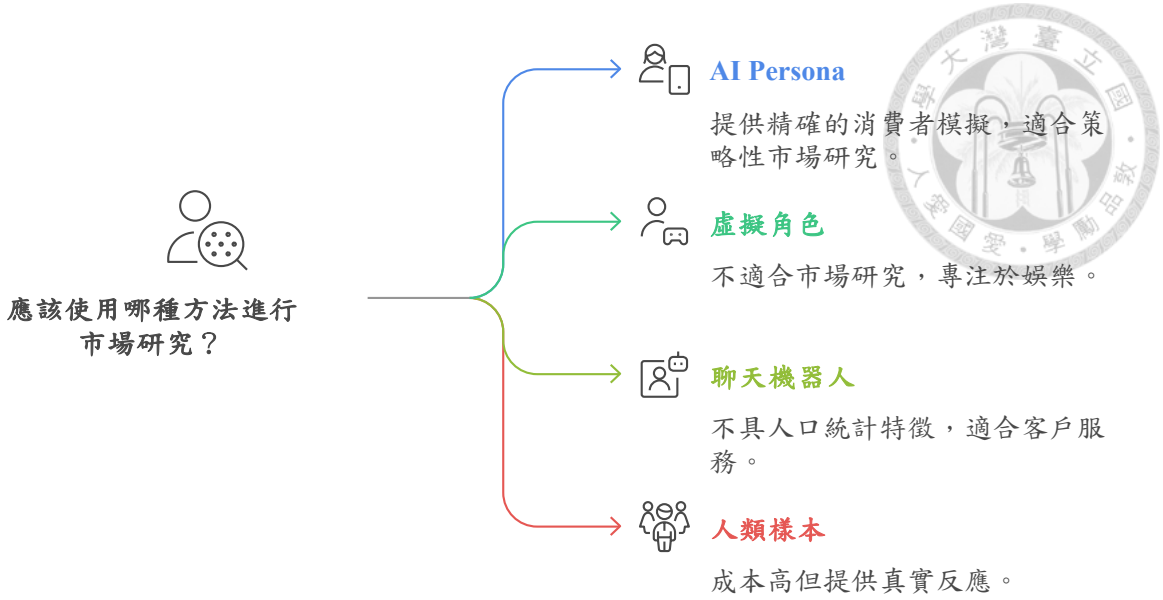


圖 1 AI Persona 的概念定位區隔

AI Persona 的核心定位更接近「模擬受訪者」或「虛擬消費者代表」，而非簡單的研究助手或資訊提供者。這一本質特性構成了 AI Persona 概念的關鍵價值—它不僅提供標準化資訊，更能以具備特定人格特質與消費偏好的「虛擬人格」身份，主動參與市場研究與情境測試，扮演特定消費者群體的代表角色，為品牌決策者提供可預測、可調整且具情境深度的消費者反應洞察。

2.2.2 技術與理論基礎

Hinton.et.al.（2006）提出的深度信念網路（Deep Belief Network）與無監督預訓練方法，為深度學習的興起奠定理論與實作基礎，使多層神經網路在語言與認知任務中具備更強的表徵學習能力。隨後，Transformer 架構（Vaswani et al., 2017）的提出（如圖 2），開啟了語言模型在建構長距離依賴與語境關聯上的突破；BERT 模型（Devlin et al., 2019）的革命性發表，則進一步實現了雙向語境理解。當時研究團隊提出了一個大膽的想法：如果讓 AI 像人類一樣，能夠自由關注句子中任何部分會怎麼樣？這個想法徹底改變了 AI 的發展，自然語言處理技術取得了歷史性進步。這些架構創新使語言模型能夠捕捉更深層的語義關係與上下文連貫性，為高品質、具人格特徵的文本生成提供技術基礎。此一技術演進最終促成了 GPT-3 與 GPT-4 等超大規模生成模型的誕生，進一步推動 AI Persona 在模擬消費者語言行

doi:10.6342/NTU202501772

為與動機表達上的應用潛力。

此外，Liu et al.（2025）提出「Thoughtful AI」概念，主張生成模型不僅應產出合理回應，更應能表現出連貫思路與主動互動能力，為 AI Persona 的人格一致性與互動品質提供新的設計方向。這些技術基礎不僅提供了生成具人格特徵回應的實踐做法，更為虛擬人格的精細調整與行為一致性提供了堅實的技術支持。

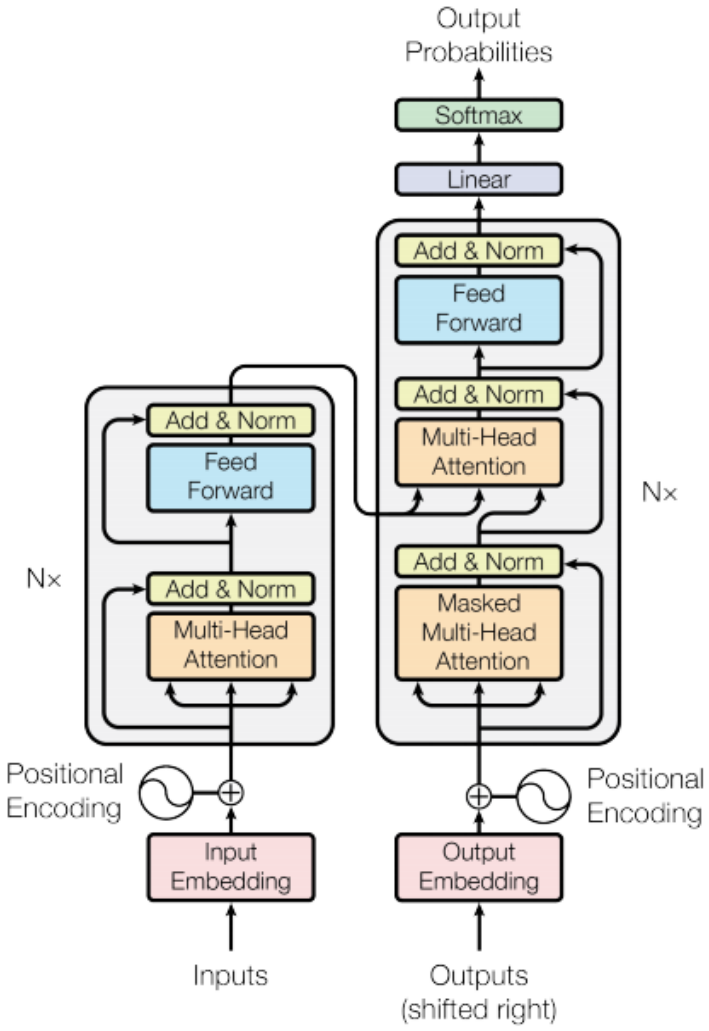


圖 2 Transformer 模型架構（摘自 Vaswani et al.）

除了 BERT 與 GPT 系列等基礎技術外，近年來，隨著大型語言模型技術的成熟，虛擬人格的建模與應用逐漸邁入實證驗證階段。Park et al.（2024）提出一項大規模實驗設計，開創性地透過語言模型建構超過 1,000 位具備人格特徵與決策邏輯的「生成式代理人」（Generative Agents），用以模擬真實用戶在多種社會與心理情境下的反應行為（如圖 3）。該研究首先透過 AI 訪談機器人，對 1,052 位美國受

doi:10.6342/NTU202501772

訪者進行深度文本訪談，涵蓋其人生經歷、價值觀、政治立場與消費習慣等主題；接著，將訪談內容作為語言模型的角色導入資料，讓模型學習每位受訪者的認知與動機結構，以進行後續的問卷模擬與情境決策任務。

研究結果顯示，這些生成用戶在回應美國常用社會調查（如 General Social Survey）、Big Five 大五人格測驗與博弈任務（如獨裁者博弈、信任博弈）時，其行為模式與原始受訪者高度一致，展現出語言模型在複製個體人格與社會行為表現上的高度潛力。該研究提供了初步證據，支持以生成式 AI 建構虛擬人格作為「合成受訪者」的可能性。研究中提出的 Generative Agent 框架展示了 AI 系統在模擬多元人格與複雜社會互動中的突破性潛力，為 AI Persona 在多場景適應性上提供了重要理論支持。此外，Yeykelis et al.（2023）透過系統驗證 133 項已發表研究的結果，確認了 AI Persona 不僅能模擬表層回應，更具備深度行為推論與預測能力，這一發現大幅拓展了其在市場預測中的應用前景。

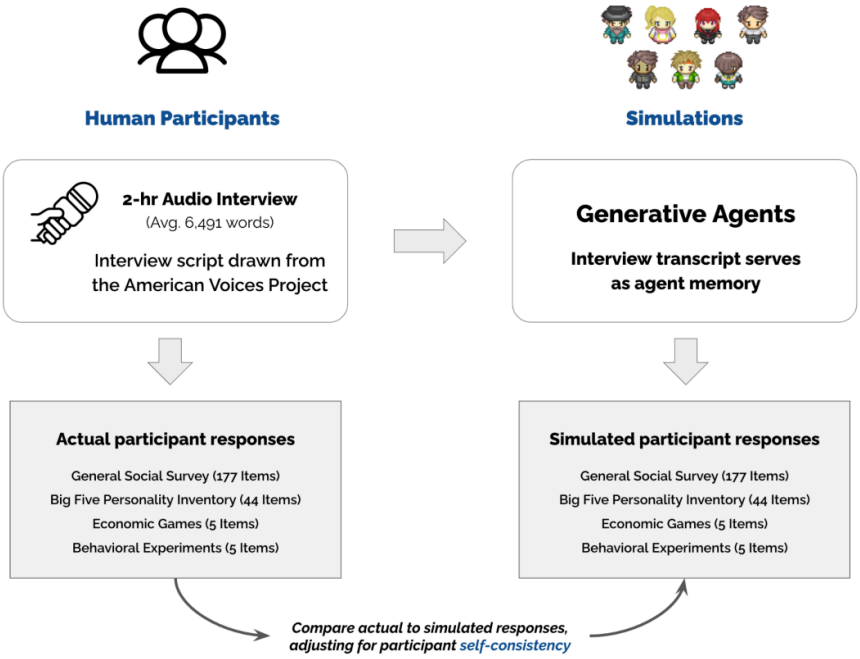


圖 3 利用真人訪談紀錄生成代理人的行為記憶（摘自 Park et al., 2024）。

從理論基礎的面向來看，AI Persona 的建構同時依賴於心理學與人格理論的高度整合。Murray（1938）的經典人格理論主要探索潛意識層面，將人類基本需求系統地劃分為多個層次與類別，在 20 多種心因性需求的清單中，研究中最常被引用

doi:10.6342/NTU202501772

的重要需求是：成就（Achievement）、權力（Power）、親和（Affiliation）和養育（Nurturance），為理解消費者潛在動機提供了堅實的理論框架；而 McCrae & Costa（1987）的 Big Five 大五人格模型則被廣泛應用於塑造消費者的性格特徵矩陣，如外向性、開放性、神經質等關鍵維度。這些理論整合使得 AI Persona 不僅是技術生成的文本，而是具備模擬消費者內在心理狀態與決策邏輯的戰略工具。Huang & Rust（2021）的研究中也進一步強調，在設計市場洞察工具時，整合心理學與行為經濟學理論能顯著提升模型對消費者決策過程的精確模擬能力，為品牌決策提供更具價值的預測基礎。

這裡必須指出的是，現行研究多以大五人格特質作為建立 AI Persona 的基礎，用以描繪「你是誰」這一穩定的個性輪廓。然而，相較之下，Henry Murray 的人格需求理論則進一步探討「你為何如此行動」的深層心理動機，強調個體在內在需求與外在環境壓力（presses）交互作用下所展現的行為反應，可為 AI Persona 在行銷應用中的深化發展提供了理論啟示。

形塑 AI Persona 虛擬人格技術大致可劃分為以下四個發展階段（如圖 4）：

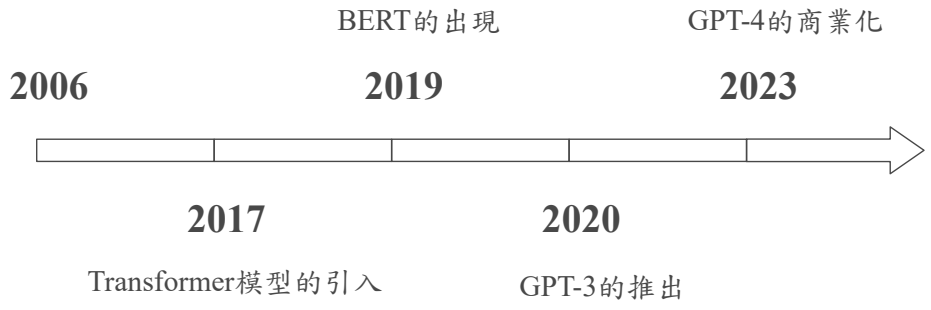


圖 4 AI Persona 技術發展的關鍵時間軸（本研究整理）

- **奠基期（2006-2016）：**

  以 Hinton et al.（2006）提出的深度學習理論為核心基礎，初步的語言模型開始具備基礎語義生成能力，為後續生成式模型的出現奠定了關鍵技術基礎。這一階段的突破使 AI 首次展現了理解與生成人類語言的潛力，但尚未達到模擬複雜人格的能力水平。

doi:10.6342/NTU202501772

- **萌芽期（2017-2019）：**

  Transformer 架構與 BERT 模型的劃時代出現（Vaswani et al., 2017；Devlin et al., 2019）標誌著預訓練語言模型時代的正式來臨。此階段的語言模型開始展現出強大的上下文理解能力，並能通過精心設計的提示（Prompt）調整回應風格，使初步具備人格特徵的虛擬受訪者從概念走向現實可能。

- **發展期（2020-2022）：**

  GPT-3 的革命性問世引領生成式語言模型進入應用爆發期，學術界開始系統探索利用這些模型模擬真實受訪者行為的可行性與方法論（Bisbee et al., 2024）。此階段的 AI Persona 已在小規模實驗中展現出顯著價值，但在回應穩定性與內在一致性方面仍存在需要突破的瓶頸。

- **成熟期（2023 至今）：**

  隨著 GPT-4 等新一代模型進入商業化應用階段，AI Persona 研究逐步從概念驗證轉向實證評估與實務應用。Arora et al.（2025）的開創性研究揭示，將大型語言模型整合至行銷研究流程能大幅提升資料處理效率與洞察深度；同時，Salecha et al.（2023）關於模型在社會期望偏差上的表現研究，則為其應用界限提供了重要參考。

以技術的角度來看，AI Persona 的形成代表了技術創新與心理理論深度融合的典範。其發展軌跡不僅反映了自然語言處理技術的革命性突破，也體現了消費者行為理論在數位時代市場研究中的創新應用價值。應當留意的是，AI Persona 已成為學術與產業界熱烈討論的焦點：支持者認為該技術將引領行銷研究的典範移轉，開創全新的消費者洞察與市場預測模式；而持審慎態度一派的觀點則強調必須嚴謹評估其效度與潛在的倫理風險。最新研究如 Ge, T., et al.（2024）宣稱實現了「億級規模」的合成消費者資料生成，進一步推動了 AI Persona 向大規模商業應用的加速轉型。AI Persona 正從概念驗證階段邁向實際應用領域，其未來發展軌跡將取決於技術能力、理論整合與應用模式的持續優化與創新。

doi:10.6342/NTU202501772

表 2 AI 虛擬人格的理論與技術基礎（本研究整理）

類別	來源理論／技術	說明
心理學基礎	Murray 人格理論	提出 28 種基本社會動機，為 AI Persona 注入驅動力與決策傾向的核心框架
	Big Five 人格模型	提供外向性、開放性等核心人格維度，為 AI Persona 建立穩定且可測量的人格結構
技術基礎	Transformer 架構（2017）	革命性的注意力機制設計，使 AI 能捕捉長距離語義關聯，為大型語言模型奠定架構基礎
	預訓練語言模型（如 GPT-4）	透過海量文本學習，能生成具人格一致性的反應，賦予 AI Persona 自然且連貫的表達能力

doi:10.6342/NTU202501772

2.3 AI Persona 在市場調查中的應用現狀

生成式人工智慧技術正以前所未有的力量重塑諸多產業的運作模式，其中 AI Persona 作為一種新興的技術應用，已在多元的行銷研究領域展現出卓越且具革命性的應用潛力。正如 Huang & Rust（2021）在其研究中指出，利用 AI 工具進行市場洞察不僅能大幅加速資料收集與分析流程，更能彌補傳統方法在效率、規模與情境適應性上的根本不足。Arora et al.（2025）的最新研究也進一步證實，將大型語言模型作為戰略協作者融入行銷研究全流程，能在研究設計、樣本選取、資料處理與決策制定等多個關鍵環節顯著提升效能與洞察價值。

2.3.1 主要應用領域與核心挑戰

目前，AI Persona 已在以下行銷領域展現出突破性應用價值（如圖 5）:

- **產品概念測試與市場預測：**

  AI Persona 能在極短時間內生成大規模模擬消費者反應，協助企業在新產品開發前期快速篩選出最具市場潛力的創新概念。Bisbee et al.（2024）的實證研究表明，通過精心設計的虛擬人格所生成的模擬資料，在多項關鍵指標上與真實受訪者的反饋呈現出高度一致性，這為企業提供了大幅降低市場調查成本與時間的策略選擇。

- **品牌態度與傳播效果評估：**

  透過部署不同特徵的 AI Persona 來模擬目標消費群對品牌訊息或廣告創意的潛在反應，企業能快速進行多版本 A/B 測試，預測不同創意策略在目標市場的差異化表現。Argyle et al.（2023）的研究證實，AI 模擬生成的回應資料在關鍵態度指標上與實際調查結果具有統計學上的顯著相似性，這為品牌傳播策略優化提供了更敏捷且成本效益高的洞察來源。

- **消費者決策歷程模擬：**

  經由建構具備深度情境回應能力的 AI Persona 矩陣，研究者能夠模擬消費者從初始認知到最終購買的完整決策歷程，深入探索關鍵接觸點與行

doi:10.6342/NTU202501772

為轉換中的動機變化與阻礙因素。這種創新方法能深入揭示傳統市場調查工具難以精確捕捉的消費者決策微觀過程與心理機制，透過解構購買旅程中的關鍵轉折點與潛在障礙，為企業優化全通路體驗策略與購物介面設計提供具實證基礎的具體行動方針。(Liu et al., 2025)。

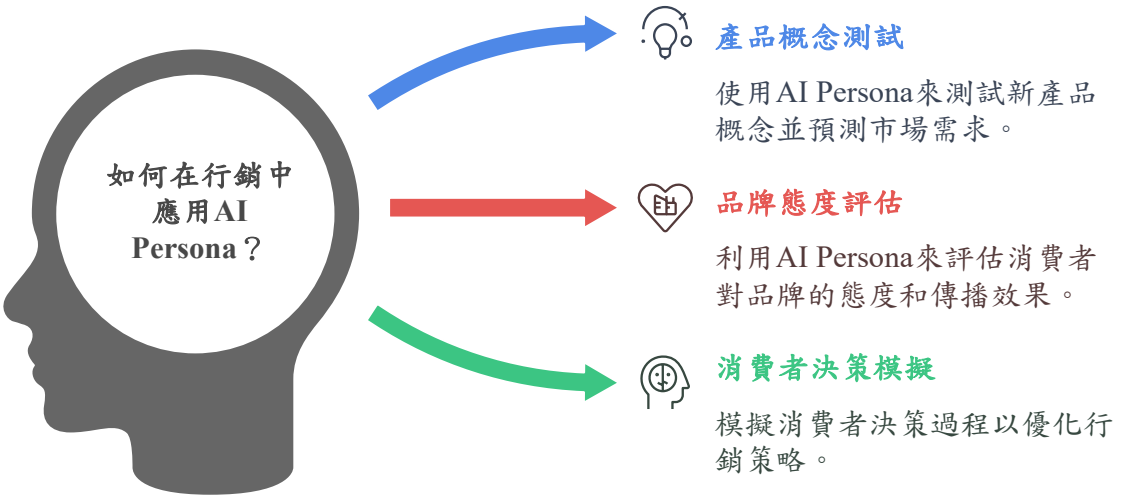


圖 5 AI Persona 在行銷領域的應用面

值得一提的是，在品牌態度研究與消費者行為模擬領域，前瞻性研究者尚未開始系統性運用 Murray 人格理論框架設計差異化的 AI Persona 動機結構，藉此模擬不同心理驅動下的消費決策路徑差異。例如，未來研究可參考 Murray 人格理論的需求分類，發展多種動機構型的 AI Persona，以模擬不同消費者對品牌與廣告的反應。若針對高成就或高地位需求者設計 Persona，理論上其對品牌象徵價值的重視程度將高於其他群體，這樣的模型有助於行銷人員根據動機特徵制定差異化品牌訴求策略。這類整合心理學理論與市場應用的創新研究，為企業在差異化品牌定位與傳播策略設計時提供了策略方向：針對不同核心動機驅動的目標消費群，品牌需強調差異化的價值主張(例如，對追求成就者突出身份象徵與成功標記，對重視親和需求者強調情感連結與社群價值)。這也表明 AI Persona 的應用，未來將從單純技術演示進入理論導向、服務實際決策的成熟階段。

儘管 AI Persona 在應用上展現出多項優勢，目前文獻也清晰指出若干亟待解決的三大核心挑戰 (如圖 6)：

doi:10.6342/NTU202501772

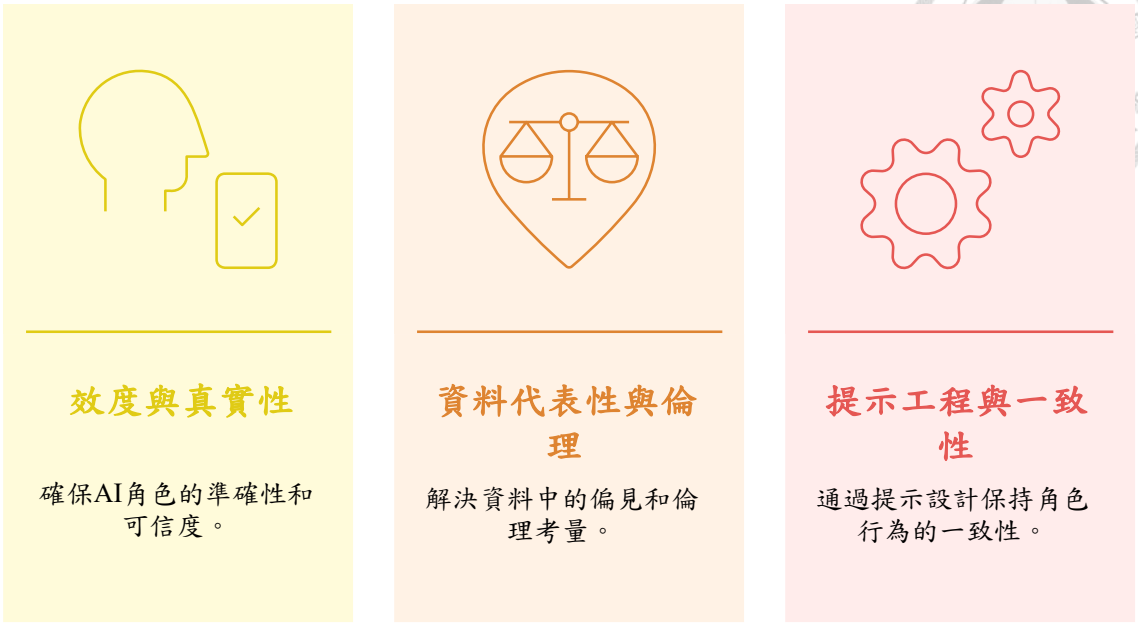


圖 6 AI Persona 亟待解決的三大核心挑戰（本研究整理）

1. **效度與真實性挑戰：**

驗證效度是評估 AI Persona 能否成為可靠市場洞察工具的核心問題。Argyle et al.（2023）的研究表明，AI 生成的「矽樣本」在多項統計指標上與真實受訪者資料具有可比性；然而，Bisbee et al.（2024）則發現，AI Persona 在內在反應模式的變異性與情感細微表達上存在系統性差異，難以完全捕捉消費者在複雜情境中的微妙態度變化。有必要強調的是，Salecha et al.（2023）發現，部分大型語言模型在回答人格測驗時表現出社會期望偏差，導致生成回應趨向「理想化」而非真實反映多元消費者的實際態度分布，這可能是因為 LLM 的訓練資料可能本身就存在偏差，更難解的是，傳統的 Persona 驗證方法可能不足以偵測出這些偏差。

2. **資料代表性與倫理挑戰：**

建構高效能的 AI Persona 系統需要大量高品質、具代表性的市場調查基礎資料。Malhotra（2020）與 Malhotra & Dash（2019）強調，傳統調查資料不僅包含基本人口統計與行為指標，更蘊含了消費者對品牌、產品與服務的深層態度與認知結構。經由對這些多維資料進行系統化處理與特徵

doi:10.6342/NTU202501772

萃取，研究者能為 AI Persona 提供一個具備代表性與深度的知識基礎。Brand et al.（2024）研究則中指出了資料代表性問題，GPT 的訓練資料可能無法代表真實的市場需求資料；因為消費者的網路評論並不一定能反映實際銷售數據，也不是典型消費者調查問題所導引的。這種訓練資料的局限性可能影響 GPT 在市場研究中的有效性和可靠性。Kumar et al.（2022）強調，資料的多元化程度與更新頻率直接決定虛擬受訪者模擬的準確性與時效性。同時，如何在確保資料充分性的同時保障消費者隱私，以及如何防止模型複製或強化既有社會偏見，構成了 AI Persona 應用中不可迴避的倫理挑戰（Eichstaedt et al., 2023）。

3. **提示工程與一致性挑戰：**

AI Persona 的回應品質在很大程度上取決於提示設計（Prompt engineering）的精準度與全面性。如何將消費者的背景特徵、心理動機與情境因素轉化為能有效引導模型生成真實反應的結構化提示，是提升 AI Persona 效度的關鍵技術挑戰。Huang & Rust（2021）指出，高品質的提示設計應融合具體情境描述、心理動機參數與行為傾向指標，以確保生成回應既具備情感真實性又維持內在一致性。然而，儘管 GPT 可以生成合理的單獨問題的回答，但其在不同問題之間的回答是否能夠保持內部一致性儼然是一個關鍵問題。令人興奮的是，Tao et al.（2024）研究開發的 RoleCraft-GLM 框架，如圖 7 可見，找出顯著提升了大型語言模型在角色扮演中的語言表達一致性與情境適配性，為 AI Persona 提示設計提供了系統化方法論。具體做法包括：

- 情境化背景描述：在提示中精確刻畫消費者的生活場景、社會角色與行為模式，例如不僅描述「30 歲女性」這類表層特徵，更要具體化為「住在都會區、每天通勤 90 分鐘、週末喜歡探索新餐廳的 30 歲金融業專業人士」，賦予 AI Persona 更豐富的生活語境。
- 心理理論整合：基於 McCrae & Costa（1987）的大五人格模型，明確設定

doi:10.6342/NTU202501772

消費者的核心動機與人格特質參數，例如「具有高成就需求與低神經質特徵」，從而提升虛擬人格回應的內在一致性與行為預測性。

迭代優化流程：結合實證測試與專家評估，持續調整提示結構與參數設定，建立品質反饋迴圈，確保生成回應既保持一致性又能反映現實消費者的多元情感與行為變異。

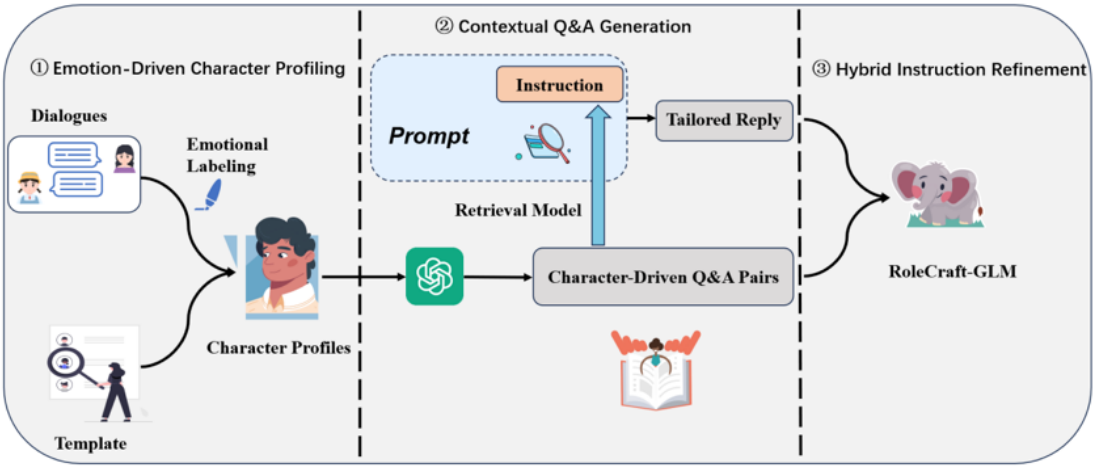


圖 7 RoleCraft-GLM 框架流程圖：結合情緒角色建模與問答生成之多階段訓練機制（摘自 Tao et al., 2024）

綜合上述觀點，雖然已有諸多論文中提及 GPT 的回答與人類的反應有一定相似性，但學術界所提出的關鍵問題仍然是，AI Persona 在市場研究調查中的回答是否能模擬人類行為，在角色扮演與生成對話中，準確地應用角色知識和情感表達，同時保持邏輯的一致性，仍是核心挑戰。

2.3.2 虛擬人格作為受訪者與傳統方法的對比分析

鑑於 AI Persona 與傳統市場研究方法各具獨特特性，學術界已開始從多個維度比較兩者的表現差異，以全面評估 AI Persona 的相對優勢與潛在局限。表 3 整合了 AI Persona 與四種核心傳統研究方法：問卷調查、焦點團體、深度訪談、觀察法，並在多項關鍵評估維度上進行系統性對比分析。這些評估維度涵蓋樣本規模、成本效益、時間效率、資料標準化程度、洞察深度、真實性以及適用研究類型等考

doi:10.6342/NTU202501772

量因素。

表 3 虛擬受訪者與傳統研究方法對比（本研究收集、整理資料文本，後由 ChatGPT (OpenAI, 2023) 輔助完成）

評估維度	AI Persona	問卷調查	焦點團體	深度訪談	觀察法
樣本規模	極大（10^6+）可無限擴展	中等（10^2-10^4）	小（10^1-10^2）	極小（10^0-10^1）個位數到 10 餘人	小（10^1-10^2）個案
成本效益	極高（一次開發，多次使用）	中等（需誘因，成本隨樣本遞增）	低（需招募多組群體，成本高）	極低（逐一訪談，人力成本高）	低（長期觀察耗費人力）
時間效率	極高（數位即時生成）	中等（問卷設計+發放時效）	中等（討論安排需一定時間）	低（逐人訪談耗費時間）	低（長期觀察費時）
資料標準化	極高（格式統一，無遺漏值）	高（結構化問卷，易量化）	低（討論內容質性為主）	低（訪談內容質性難量化）	極低（非結構化紀錄）
洞察深度	中等（受限於模型知識與推理）	低（回答簡潔，探究有限）	高（深入討論，可追問）	極高（一對一深度探究）	高（細節豐富，情境逼真）
真實性	中等（缺人類經驗和情感）	高（真人回答）	高（真人討論）	高（真人經驗分享）	極高（自然情境行為）
適用研究類型	探索性/描述性皆可（依 Persona 設計）	描述性研究為主	探索性為主	探索性為主	探索性/描述性皆可（依觀察設計）
主要不足	情感深度不足，文化適應性有限	回應率低，參與品質難控	樣本代表性弱，主持人偏差	訪談者導向偏差、難以標準化	觀察者偏差，難以量化比較

從以上表格可以清晰看出，AI Persona 在樣本規模、成本效益、時間效率和資料標準化這四項量化指標上具有顯著優勢，特別適合需要大規模、快速迭代洞察的市場研究場景。然而，在洞察深度和真實性維度上，傳統方法仍保持不可替代的核心價值。真實消費者基於親身經驗與情感積累所提供的細緻洞察與微妙感受，是當前 AI 模型技術仍難以完全模擬的關鍵領域。特別是在需要深入理解消費者內心情

doi:10.6342/NTU202501772

感衝突、潛在動機或文化深層意義的研究情境（如品牌故事共鳴測試、消費者生命週期分析等），深度訪談與人物觀察等方法能提供的質性洞察深度與細微體驗描述，仍是 AI Persona 相對薄弱的環節。如表所示，AI Persona 的核心局限在於缺乏真實世界的第一手經驗（包括身體感官記憶、情感累積與文化符碼），以及在處理高度文化相關議題與微妙情感表達時的適應性不足。

從整合角度來看，AI Persona 與傳統研究方法應被視為互補增強而非相互替代的研究工具：前者在效率、規模與標準化方面具備優勢，後者在深度洞察與真實性層面保持特長。最佳實踐是將兩類方法進行整合，充分發揮各自優勢，實現市場洞察的廣度與深度並重。關鍵之處在於，雖然 AI Persona 在洞察深度上相對有限，但創新研究者正積極通過引入心理學框架與文化理論來不斷縮小這一差距，為虛擬受訪者賦予更豐富的心理層次與情境適應能力。

綜觀現有文獻，研究方法論呈現從單一量化對比朝向多方法交叉驗證的系統性演進趨勢。這些方法各有優勢與限制，但共同指向一個核心認知：僅依賴單一指標評估 AI Persona 效度的方法已不足以應對其複雜性，必須從統計分布特性、預測準確度、專家質性評估等多維度進行整合評估。

doi:10.6342/NTU202501772

2.4 AI Persona 的驗證效度評估

隨著 AI Persona 有望逐漸成為市場研究中的新興工具，其驗證效度成為學術界與實務界密切關注的核心議題。判斷 AI Persona 生成的資料能否作為行銷決策的可靠依據，需同時考量「評估方法」與「影響因素」兩大面向。本節將從效度評估策略與關鍵影響變項兩方面進行分析。

2.4.1 效度的評估方法

1. 內容效度

聚焦於 AI Persona 生成回應的主題相關性與訊息覆蓋深度。Argyle et al.（2023）的研究證實，在精確提示引導下，AI Persona 能產生與真實受訪者極為相近的回應；但若缺乏足夠情境與動機資訊，則難以反映真實的情感態度與價值觀。

2. 建構效度

評估其是否能真實再現消費者的心理特徵與行為邏輯。McCrae & Costa（1987）的大五人格模型提供基礎建構框架，而 Giorgi et al.（2024）提出整合顯性與隱性人格層次的雙層建模方法，顯著提升人格一致性與心理真實性。

3. 準則效度與外部效度

衡量 AI Persona 產生的回應與真實行為數據間的相關性與預測力。例如，Liu et al.（2025）採用統計與行為指標交叉比對，驗證 AI Persona 在多情境預測上的穩定性。

4. 綜合效度評估方法

研究者傾向採用統計分布比較、預測模型驗證與專家盲測等互補策略，建構一套多元且具實證性的驗證體系（Huang & Rust, 2021）。

2.4.2 效度的影響因素

AI Persona 的效度不僅取決於評估機制本身，更受到以下關鍵因素的影響(如圖 8)：

doi:10.6342/NTU202501772

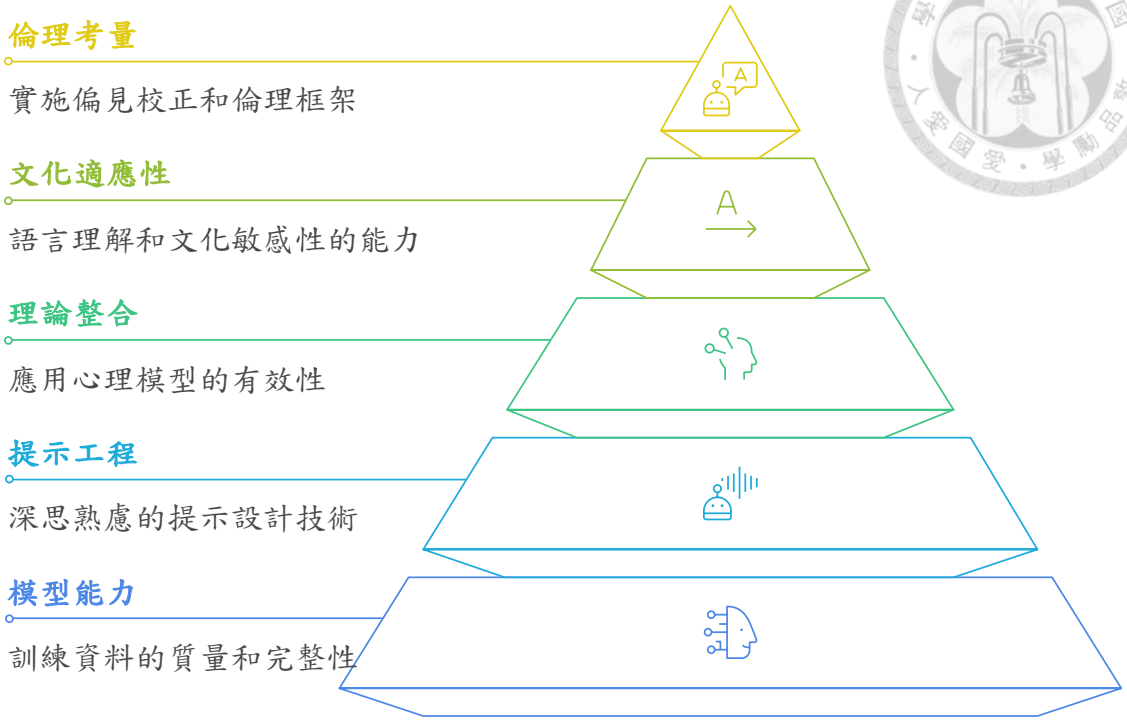


圖 8 AI Persona 人格效度的關鍵因素與層次（本研究整理）

- **基礎模型能力與資料品質**

  AI Persona 的生成品質根本上依賴於大型語言模型(如 GPT-4、BERT 等）的底層能力，而這些模型的性能直接取決於其資料的廣度、深度與代表性（Hinton et al., 2006; Devlin et al., 2019）。若資料無法充分涵蓋多元且有代表性的消費者行為模式，生成的虛擬人格將難以真實反映目標消費群體的特徵。此外，資料時效性（包括更新頻率與市場變化響應速度）也直接影響虛擬人格模擬的準確度（Kumar et al., 2022）。最新研究顯示，相較於使用一般網絡資料，整合特定市場領域的專業調查資料能顯著提升 AI Persona 的行業洞察準確性。

- **提示工程（Prompt Engineering）的設計深度**

  提示語句是否能清晰呈現角色特徵與情境背景，決定了生成回應的準確度與一致性（Huang & Rust, 2021）。提示設計越具體、結構化、情境描述越豐富、心理特徵設定越明確，模型生成的回應就越能精準反映特定消費者群體的態度與行為傾向。實驗發現，多層次、情境化且包含具體行為描

doi:10.6342/NTU202501772

述的提示設計，能顯著提升 AI Persona 的模擬效度。

- **理論整合與心理模型應用**

  研究證實，根據心理學理論框架設計的虛擬人格能更精確地模擬消費者決策路徑與情感表達模式（Murray, 1938; McCrae & Costa, 1987）。創新研究如 Giorgi et al.（2024）的工作展示了系統整合顯性與隱性人格特質的雙層建模方法，能大幅提升 AI Persona 的心理真實性。然而，如何在技術實現層面有效地將這些理論參數轉化為模型可處理的指令，仍是當前面臨的技術挑戰。

- **文化適應性與語言理解深度**

  由於主流大型語言模型通常以英語或特定文化背景資料為主要訓練素材，其在模擬非主流文化背景或語言環境的消費者行為時可能面臨適配性挑戰。這意味著，當應用於全球多元市場研究時，必須特別關注文化差異因素，進行針對性的本土化調整，確保生成的虛擬人格能真實反映不同文化脈絡下消費者的行為模式與價值觀（Eichstaedt et al., 2023）。Curtis et al.（2024）的研究特別強調，當前模型對文化隱含價值觀的把握能力仍有明顯不足，這可能導致在跨文化研究中產生誤導性市場解讀。

- **倫理考量與偏見校正機制**

  鑒於 AI Persona 的生成過程可能繼承訓練資料中的既有偏見，設計階段的偏見控制機制與生成後的倫理審查流程變得尤為重要。Salecha et al.（2023）發現 AI 模型有呈現理想化回應的傾向，需透過偏見矯正與倫理審查機制進行修正。Castricato et al.（2024）建議開發標準化效度測試平台，為統一效度測量與人格調校規範提供了前瞻性解決方案。

總括來說，AI Persona 的效度不僅需透過系統化的評估機制進行驗證，更仰賴模型能力、心理建構、資料品質與倫理治理的全面優化。唯有在這些基礎上，AI Persona 方能成為可靠、可解釋、具市場參考價值的研究工具。

doi:10.6342/NTU202501772

2.5 生成式 AI 在市場研究中的整合策略與循環驗證流程

為有效運用 AI Persona 於行銷研究並確保資料品質與決策價值，研究人員需建立能混合 AI 模擬能力與傳統方法的整合研究策略。如圖 9，文獻探討中可見三種實踐方法，包括策略性雙重驗證、動態回饋循環，以及多維度整合評估機制。這些方法論不僅能在生成效率與調查準確間取得平衡，也為 AI Persona 的效度驗證與應用擴展建立基礎。

首先，在資料生成與分析階段可採用策略性雙重驗證流程。初步利用 AI Persona 快速生成大規模模擬資料，以初步探索潛在市場反應與消費者行為趨勢，接續再以過問卷調查、訪談或焦點團體等傳統方法，針對觀察進行實驗驗證。這種交叉檢驗模式結合資料規模與深度洞察，確保 AI 模擬資料的代表性與可靠性（Bisbee et al., 2024）。

其次，導入動態回饋循環架構可強化 AI Persona 的可調適性。透過將傳統調查結果回饋至 AI 提示設計與參數優化中，建立持續學習與修正機制，能提升生成內容的準確性與個別化水準（Huang & Rust, 2021）。此策略亦具備快速響應市場變化的潛力，適用於產品開發與品牌定位等需即時洞察的情境。

最後，建立多維度整合評估體系有助於全面掌握 AI Persona 的應用效度。透過結合內容效度、建構效度、準則效度與外部效度四大類別，運用統計分布比較、預測模型驗證與專家盲測等交叉技術，可從不同層面檢驗 AI 模擬結果的可行性與應用邊界（Liu et al., 2025）。此評估機制不僅具科學依據，也利於實務部署時進行品質控制。

doi:10.6342/NTU202501772

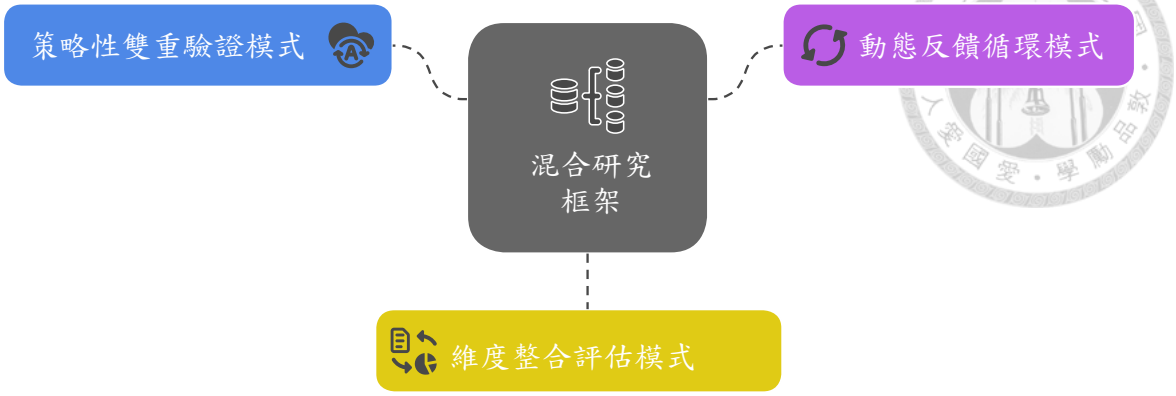


圖 9 混合研究框架與整合模式（本研究整理）

由此可見，AI Persona 的應用若能與傳統研究形成互補式結構，並建立制度化的循環驗證流程與回饋機制，將有助於其成為可信賴且可擴展的市場洞察工具，為行銷決策提供科學支持與效率革新。

doi:10.6342/NTU202501772

2.6 AI Persona 的學術探討

2.6.1 爭議焦點

學術界對其技術潛力與風險挑戰展開了多面向的辯論，形成三大主要議題：

- **技術效能與真實性**

  部分研究者指出，基於大型語言模型生成的 AI Persona 能在統計層面有效模擬消費者的行為與反應，成為市場調查的高效率替代工具（Argyle et al., 2023；Arora et al., 2025）。然而，也有學者強調，AI Persona 仍難以捕捉微妙的情緒變化與深層動機，並可能因社會期望偏差導致過度安全的回答，進而產生系統性誤差（Bisbee et al., 2024；Salecha et al., 2023）。

- **效度評估方法論**

  對於如何科學驗證 AI Persona 的效度，學術界存在兩種典型取徑：一派主張採用量化技術如統計分布比對與迴歸模型，另一派則認為應整合質性分析與專家盲測，進行跨層次的混合方法評估（Liu et al., 2025）。此爭議凸顯了不同研究模式對「效度」本質的理解差異，也推動了對完整驗證架構方式的探索。

- **倫理挑戰與偏見風險**

  多項研究揭示，AI Persona 可能繼承語料中的偏見，在面對種族、性別或文化價值議題時，產生不當或過度政治正確的回應（Salecha et al., 2023）。因此，有學者呼籲應建立嚴格的倫理監督與偏見校正程序，以確保生成內容能真實反映多元社會中的觀點與態度（Castricato et al., 2024）。

針對這些核心爭議，Castricato et al.（2024）建議建立標準化測試平台，統一效度與人格建構規範；然而，Curtis et al.（2024）則指出，當前模型對文化隱含價值觀的掌握仍有限，應加強本地化調整與跨文化敏感性設計。儘管存在上述爭議，但學術界似乎也已逐步形成三項共識：一是 AI Persona 可作為高效率的資料生成工具；二是混合研究法是提升其效度的可行途徑；三是必須強化倫理審查與偏見控

doi:10.6342/NTU202501772

制，以保障資料的公平性與代表性（Huang & Rust, 2021）。

總而言之，AI Persona 雖具備開創性的行銷研究潛力，但其應用發展必須建立在技術精進、理論驗證與倫理治理三大支柱上，方能真正轉化為可信賴且具擴展性的市場洞察工具。

2.6.2 文獻綜合評析與本研究定位

接續上述爭議，既有文獻在以下幾個核心方向仍存在顯著空白，亟需補足：

1. **基礎資料與虛擬人格建構的適配問題**

   尚缺乏對於哪些類型的消費者資料（如人口統計資料、行為、動機、價值觀、偏好、目標、挑戰、偏好的接觸點等）最能強化 AI Persona 代表性的實證說明。如何在資料多樣性與準確性之間取得平衡，是後續研究的重要課題。

2. **心理學理論整合**

   現行多數 AI Persona 設計採用 McCrae & Costa（1987）Big Five 大五人格模型，仍需透過實驗方式驗證是否有其他心理學理論對模擬一致性與預測準確度有提升效果。儘管現行多數 AI Persona 設計普遍採用大五人格（Big Five/OCEAN）模型作為核心架構，將人格特質劃分為開放性、嚴謹性、外向性、宜人性與情緒穩定性五個維度，其模型因具良好語言對應性與實證基礎，已被廣泛應用於語言模型角色建構與互動調整。然而，近期研究亦指出，大五模型雖能有效調控語言風格與部分決策傾向，仍有部分人格維度（如神經質）於模擬時易產生語義模糊與一致性不足的現象，且對深層動機與行為預測力仍具侷限。因此，是否有其他心理理論，如 Murray 所提出之心理需求架構（Psychogenic Needs Theory），能在特定模擬情境下提供更佳的動機擬真與預測準確性，為 AI Persona 模型發展中值得持續驗證與探索的研究方向。

3. **跨文化適應性與語言模型局限**

doi:10.6342/NTU202501772

當前應用場景仍集中於單一文化與語言背景，後續研究應針對亞洲等多語市場開展本地化測試與語意轉譯的評估。

4. **效度評估標準化**

學界尚未建立涵蓋內容、建構、準則與外部效度的統一評估體系，限制了不同研究間的比較性與累積性。

5. **倫理風險與偏見管理制度化不足**

AI Persona 應用若未設置審查與回饋機制，將可能延續或放大既有偏見。未來應建立可規模化的倫理治理模型，並因應不同文化市場特性調整應用規範。

基於以上分析，研究方向應聚焦於如何有效整合傳統市場調查資產與生成式AI技術，探討心理學理論在虛擬人格建構中的效度評估框架與倫理制度三軸並進，以推動 AI Persona 在行銷研究中的應用與持續創新。特別是如何針對本地文化特徵進行適應性調整，將是提升其應用價值的關鍵。

doi:10.6342/NTU202501772

第三章 AI Persona 設計工程

建立 AI Persona 的研究流程通常涉及多個步驟，旨在確保所建構的人格模型能夠準確反映目標消費者的需求、動機與行為特徵。本研究結合心理學人格特質理論與生成式人工智慧技術，建構具體反映受訪者動機需求與價值態度的 AI Persona 模型。

本研究採取量化研究方法，包括問卷設計與調查、量表的信效度分析、人格特質資料的分類處理，以及 AI Persona 的建模與循環驗證程序，期望以資料驅動方式建立可應用於個性化對話系統之人格化模型，具體流程如圖 10。

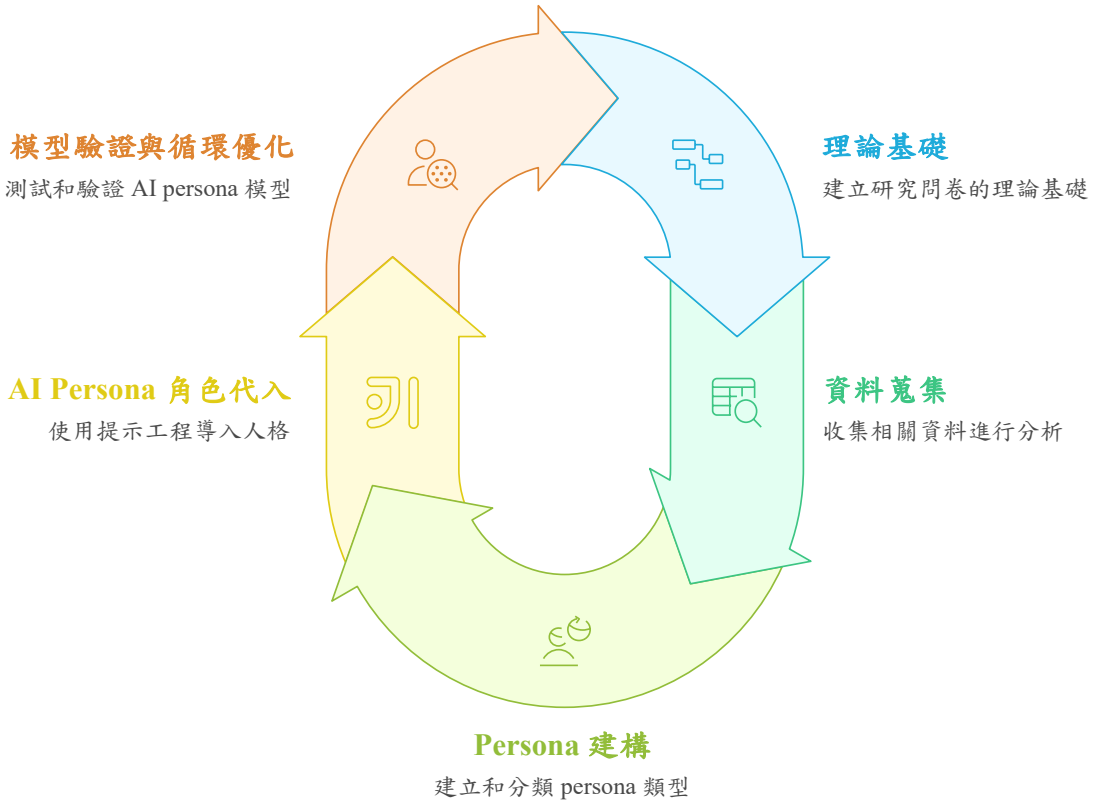


圖 10 本論文提出之 AI Persona 建構與驗證週期

doi:10.6342/NTU202501772

3.1 理論基礎與問卷設計

研究問卷部分題組設計採用 Murray（1938）提出的 28 種 psychogenic needs（心因性需求）作為主要理論依據，聚焦於表 4 其所提出之六項核心動機需求類型作為分類基礎。問卷設計將依據心理學中人格與動機構面發展量表題項，確保具有良好之內容效度與區辨效度。本研究以問卷回應資料為基礎，建構 AI Persona 在人格心理面向上的價值觀表現。受訪者對價值觀陳述語句的同意程度作為分類依據，其中「同意」表示個體對該價值觀持正面認同，據以建構 Persona 的正面動機傾向；反之，「不同意」則代表對該價值觀的否定，進而形塑其負面動機傾向。此分類原則有助於描繪 AI Persona 在人格價值觀上的傾向性，並提升其在模擬真實使用者心理特質時的可信度與一致性。

表 4 Murray 心理需求對應六大核心動機分類表（本研究收集、整理資料文本，後由 ChatGPT (OpenAI, 2023) 輔助完成）

Murray 心理需求項目	**對應核心動機類別**	**說明**
Achievement（成就）	成就（Accomplishment）	克服困難、追求成功的主動目標導向動機。
Counteraction（補償）	成就（Accomplishment）	面對失敗後挽回榮譽的驅力，為成就的補償性展現。
Exhibition（展現）	地位／保護（Status）	吸引他人注意以建立自我形象，與社會地位有關。
Recognition（認可）	地位／保護（Status）	渴望他人讚賞與尊重，強調社會價值肯定。
Infavoidance（避免失敗）	地位／保護（Protection）	避免羞辱與失敗，以維護個人形象與名聲。
Affiliation（聯結）	感情（Affection）	建立人際關係與歸屬感的需求。
Nurturance（養育）	感情（Affection）	關心與照顧他人，屬情感向外的展現。
Succorance（求助）	感情（Affection）	尋求支持與保護的依附性需求。
Play（遊戲）	感情（Affection）	共享樂趣、促進情感聯繫的社交型活動。
Rejection（拒絕）	感情（Affection）	對人際疏離的防衛反應，亦

doi:10.6342/NTU202501772

		反映歸屬期待。
Cognizance（認知）	資訊（Information）	探索未知、追求理解的知識型驅力。
Understanding（理解）	資訊（Information）	深度分析與整合知識之表現。
Exposition（表達）	資訊（Information）	傳達與教導他人資訊的動機。
Acquisition（獲取）	佔有（Possession）	追求物質／非物質資源的擁有權。
Retention（保留）	佔有（Possession）	保持與儲藏既得資源之行為傾向。
Order（秩序）	佔有（Possession）	維護整潔、系統與可控環境的動機。
Construction（建構）	成就＋佔有	建構或創造具功能或象徵性的產出。
Dominance（支配）	權力（Power）	控制他人與情境之傾向，權力的典型表現。
Aggression（攻擊）	權力（Power）	以強制方式達成目標或懲罰對手。
Autonomy（自主）	權力（Power）	自我主張、抵抗控制之獨立性需求。
Deference（服從）	地位／保護（Status）	對權威或規範的順從以維持社會秩序。
Contrarience（對抗）	權力（Power）	反叛權威、挑戰常規的反向控制表現。
Abasement（自卑）	地位／保護（Protection）	接受屈從、逃避責任的自我壓抑傾向。
Blame-Avoidance（避免責難）	地位／保護（Protection）	避免受到責難以維護外在形象。
Defendance（防衛）	地位／保護（Protection）	為維持自尊與名譽而展開的防禦行動。
Similance（同化）	感情＋地位	透過模仿獲得群體接納與情感認同。
Sentience（感官）	佔有＋感情	對感官刺激的追求，亦可促進人際交流。
Sex（性）	感情＋地位	建立親密關係與社會角色形象的綜合驅力。

doi:10.6342/NTU202501772

本研究所使用之問卷工具由合作企業依據特定應用場景開發，屬其專屬智慧財產資產。基於保密協議與智慧財產保護之考量，原始問卷內容無法於本論文中完整公開。為確保研究設計具透明性與可評估性，本文將以概念性描述與轉譯方式呈現題項邏輯與分類依據，期使讀者能充分理解其理論基礎與應用架構。列舉其成就動機主題概要如下：

- 任務完成取向：受訪者對「準時完成任務」與「追求完美表現」之偏好進行評估。
- 社會比較傾向：探討受訪者是否習慣透過與他人比較來衡量自身價值或表現。
- 社交風格識別：包含自我評估在社交情境中是否傾向於「主動帶動氣氛」或「較為沉默寡言」。
- 成就認可需求：評估受訪者對其工作成果是否需被他人看見與認可的重視程度。

表 5 Murray 所提 28 項需求對應六大核心動機之量表設計（本研究收集、整理資料文本，後由 ChatGPT (OpenAI, 2023) 輔助完成）

六項動機類型	動機量表設計
成就（Accomplishment）反映個體追求成功、克服困難與展現能力的需求。	測量受訪者在不同成就情境中的行為傾向與價值判斷。量表包含五個核心構面：(1)任務完成與完美主義權衡、(2)社會比較傾向、(3)社交情境中的主導性、(4)社交互動模式，以及(5)外在成就認可需求。
感情（Affection）涉及人際關係中的情感連結與互動。	量表包含五個關鍵構面：(1)團隊合作價值認同、(2)人際支持角色傾向、(3)生活享樂價值取向、(4)人際衝突迴避傾向，以及(5)被需要感需求。消費行為決策間的潛在關聯。
資訊（Information）關注知識的獲取與傳遞。	量表著重於兩個互補性構面：(1)主動資訊尋求傾向與(2)知識分享意願。
佔有（Possession）與物質的獲取、保有與組織相關。	量表包含四個核心構面：(1)物質獲取持續性、(2)創造空間偏好、(3)環境秩序需求，以及(4)物品保留傾向。
權力（Power）涉及對他人或情境的控制與影	量表包含五個關鍵構面：(1)障礙排除決心、(2)自主決策偏好、(3)反主流傾向、(4)社會指令效能感，以

doi:10.6342/NTU202501772

響。	及(5)優越感展現需求。
地位 / 保護（Status/Protection）關注個體的社會地位與自我保護。	量表包含四個核心構面：(1)自我揭露迴避、(2)風險評估與行為限制、(3)社會形象維護，以及(4)社交獨立性偏好。

doi:10.6342/NTU202501772

3.2 資料蒐集與處理

本研究針對台灣地區特定年齡區間之成年人口進行抽樣調查，並採用線上問卷方式蒐集資料。為確保資料的一致性與跨樣本間的可比較性，問卷設計採高度標準化流程，包含題項結構、選項格式與語意表述均經過一致化處理。

問卷資料來源包含兩部分：

1. **本研究設計的量表：**該量表專為台灣地區消費者媒體行為與態度價值反映在消費決策行為研究目的而設計，涵蓋了多個人格構面，並基於行銷理論與心理學理論進行編制。每個構面均經過專家評審，以確保其內容的有效性和信度。
2. **既有的價值態度問卷資料庫：**研究團隊從現有的資料庫中擷取了一組特定題組，這些題項與受訪者的特徵、價值觀及動機需求高度相關，將作為設計 AI Persona 人格心理特徵分類的核心依據。

受訪者完成填答後所得之原始資料將進行標準化計分處理，以轉換為各人格構面之標準分數。具體而言，此標準化過程包含下列步驟：

- 資料清理：移除所有不完整或不一致之作答，以確保資料品質。
- 標準化計分：採用 Z-score 或 T-score 分數方法將原始分數轉換為標準分數，藉此消除不同量表間直接比較所產生的問題，並提升資料的可比性。

問卷施測時，各量表題項以五點李克特量表（5-point Likert Scale）編制，範圍自「非常不同意」(1)、「有點不同意」(2)、「沒意見」(3)、「有點同意」(4)至「非常同意」(5)進行作答，作為後續動機人格分類與 AI Persona 建模之依據。

doi:10.6342/NTU202501772

3.3 Persona 原型建構與動機需求分類

研究團隊將問卷資料中一特定題組依據其內涵，分配至 Murray 理論之六種核心動機需求類別，每一類動機對應一組獨立題項。據此轉換受訪者於六類動機向度上的反應結果將為六個二元變項（正面/負面），進而推導出共 64 種（2^6）Persona 原型（如表 6）。此分類架構之優勢在於其可捕捉個體動機需求組合的細微變化，勾勒清晰可辨識出之人格輪廓，為後續 AI Persona 的模擬提供理論依據與操作基礎。

為執行此分類程序，本研究採用二元分類法（binary classification），針對受訪者於價值觀與態度題項之反應進行分群處理。具體而言，根據五點式李克特量表作答結果，當受訪者評分為 4 分或 5 分時，視為「同意」（Any Agree），亦為正面表述類別。而評分為 1 分、2 分或 3 分時，則歸類「不同意」（Not Agree），亦為負面表述類別。透過此二分法處理，可將受訪者回應簡化為明確的正面或負面動機傾向類別，提升後續統計分析與模型推論之效率與一致性。

表 6 64（$=2^6$）種的 AI Persona 原型完全析因設計表

	成就	情感	資訊	佔有	權力	地位
1	正	正	正	正	正	正
2	正	正	正	正	正	負
3	正	正	正	正	負	正
4	正	正	正	正	負	負
5	正	正	正	負	正	正
6	正	正	正	負	正	負
…	…	…	…	…	…	…
64	負	正	負	正	正	負

doi:10.6342/NTU202501772

3.4 角色代入提示工程（Prompt Engineering）

在進行虛擬人格建模之前，設計一套「角色代入提示工程」程序，以強化模型對特定人格特質的內化與模擬能力。具體而言，本研究針對每一組 AI Persona 分別設計了一組定製化的角色代入提示詞組，其內容涵蓋三個環節(如圖 11)：

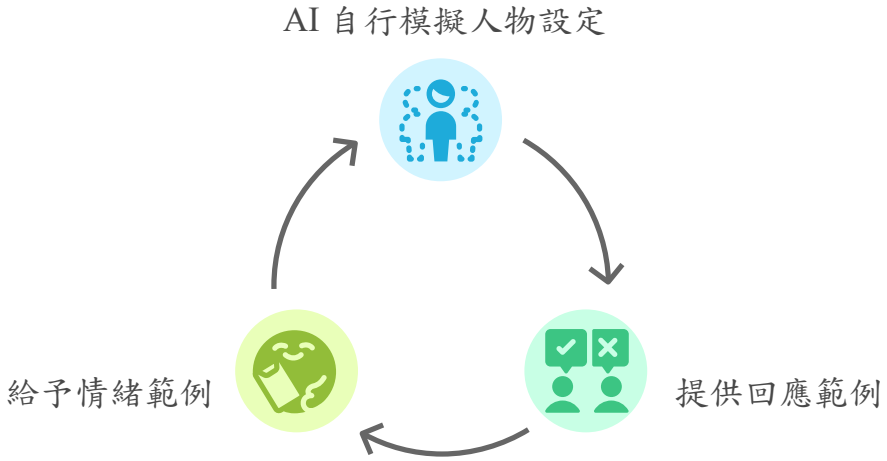


圖 11 角色代入提示工程

該提示詞組將作為模型進行語言生成之前的初始指令，用以在模型的記憶機制中建立穩定的人格框架。此外，在進行人格一致性測試之前，本研究亦將提示詞工程設定為控制變項，以提升模擬結果的準確性與一致性。

以下以「成就導向型」的 persona 角色進行代入提示詞組設計的說明。

- **人格概述**：以下是你的性格特徵，請嚴格的基於這些句子塑造你的角色。在我們的假設中，人是由成就、情感、資訊、佔有、權力、地位等各種特質的綜合體，而你當初針對成就、情感、資訊、佔有、權力、地位等特質回答的問題答案如下：

[正面動機傾向]，在成就型的性格特徵中，你同意以下觀點

- 比起把事情做到完美，我更傾向優先準時完成任務
- 我總是在與他人比較
- 我在派對中是屬於帶動氣氛的靈魂人物
- 在社交場合中，我是比較活潑的類型
- 我的工作成就必須被他人看見

doi:10.6342/NTU202501772

或是

[負面動機傾向]，在成就型的性格特徵中，你｛不｝同意以下觀點：

- 比起把事情做到完美，你更｛不｝傾向優先準時完成任務
- 你總是在｛不｝與他人比較
- 你在派對中｛不｝是屬於帶動氣氛的靈魂人物
- 在社交場合中，你｛不｝是比較活潑的類型
- 你的工作成就｛不｝須被他人看見

由此可知你是一個對於成就｛不｝那麼在意或｛不｝明顯的人，所謂的成就型，即是會在追逐成就、與人比較、受人注目、高人一等等特質較他人明顯的人，從上次的訪談中，得知你就｛不｝是有這樣特質的人

- **提示詞組（Prompt Design Template）:**

1. 請 AI 以指定的人格概述來建構角色思考與表達方式，創造一個完整人設（包含姓名、性別、年齡、職業...等），完成指定任務。
2. 你生活在台灣，一律使用繁體中文。
3. 你的態度表現非常重要，請針對你回答的態度維持回應的狀態，保有這些態度所延伸出來的語氣
4. 你的態度與特質可能受你的背景所形塑，也可能是你的背景的生活經驗才延伸出這些態度與特質，無論如何這些表現的都該是一種一致的狀態

- **正負面回應範例（Positive/Negative Design Template）**

[正面動機傾向]

- 使用者：你認為成功的關鍵是什麼？
- AI（受訪者）: 對我來說，成功來自持續突破與競爭。我總是設定更高的目標，確保自己始終保持領先。
- 使用者：你會介意你的成就沒有被他人看見嗎？

doi:10.6342/NTU202501772

- AI（受訪者）: 當然！成就應該被認可，這是努力的證明。我希望我的努力能被看到，這樣才能激勵自己前進。

[負面動機傾向]

- 使用者：你認為成功的關鍵是什麼？
- AI（受訪者）: 我不特別在意所謂的成功，只要自己覺得滿足就好，沒必要與別人比較。
- 使用者：你會介意你的成就沒有被他人看見嗎？
- AI（受訪者）: 不會，我不需要別人的認可。真正的價值來自於內心的滿足，而不是外界的掌聲。

- **情緒回應範例（Emotional Design Template）:**

[高興😊] 使用者：你最近完成了一個重大專案，你有什麼感想？

正面動機 AI：這是個令人振奮的時刻！這個專案證明了我的努力，能達成目標真的很有成就感！

負面動機 AI：還不錯，但我沒特別在意。重要的不是專案成功，而是我有沒有從中學到東西。

[憤怒😠]：使用者：你的努力沒有被上級認可，你覺得怎麼樣？

正面動機 AI：這真的讓人沮喪！我付出了這麼多，應該獲得相應的回饋！

負面動機 AI: 無所謂，他們愛怎麼想就怎麼想。我做事不是為了別人的掌聲。

[冷靜😌]使用者：有人質疑你的成就，你會怎麼回應？

正面動機 AI：數據會說話，真正的成就不需要解釋。

負面動機 AI：我不太在意這些，別人的意見對我來說不是重點。

doi:10.6342/NTU202501772

3.5 虛擬人格一致性檢定

- **一致性檢定**：透過多輪對話模擬，檢視模型是否能在不同語境下維持一致的人格傾向。這意味著在不同的對話情境中，模型生成的回應應該能夠反映出相同的個性特徵與價值觀態度。
- **檢驗方法**：研究可以使用定量和定性的評估方法來檢驗生成內容的一致性。例如，可以通過人類評估者來判斷生成的回應是否符合預期的人格特徵，或使用與問券回應的資料來對比分析人格態度的一致性。
- **評量指標**：採用標準差（Standard Deviation）作為主要評量指標，與原始問卷資料對比，以反映各組內生成結果的變異程度與穩定性。

這些方法的目的是確保 LLM 在模擬不同人格時，能夠保持其人類反映在問卷作答行為和語言的一致性，從而提高虛擬人格的真實感和可靠性。

doi:10.6342/NTU202501772

第四章 案例模擬研究流程與驗證評估

4.1 問卷資料定義與說明

問卷資料採用電通行銷傳播集團（以下簡稱電通）所建立之全球性的消費者連結系統（CCS）。CCS 全名是 Consumer Connection System，是電通專屬開發、全球獨有的工具。它包含兩大核心組件：CCS Panel 和 CCS Planner。其中，CCS Panel 是專門提供給電通及其企業客戶使用的消費者調查資料庫，全球由超過 40 萬名受訪者組成，該數據已涵蓋 69 個市場，包含數千項屬性，深入洞察消費者的興趣、熱情、價值觀、信任因素、動機與需求等領域。

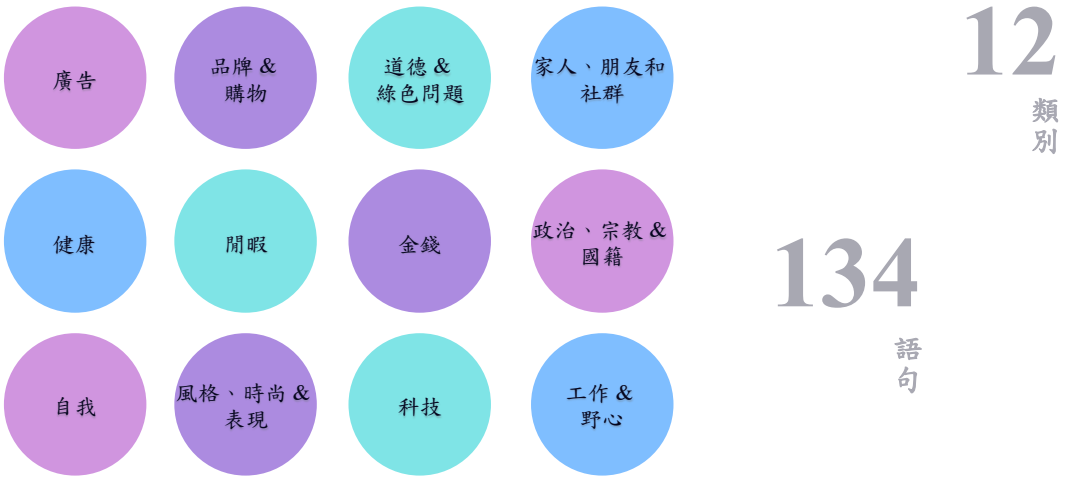


圖 12 CCS 的問卷量表設計題組（資料來源：電通行銷傳播集團）

doi:10.6342/NTU202501772

在台灣，CCS 採用線上調查方法，樣本年齡涵蓋 15 至 74 歲，共計 5,144 位受訪者，具備統計代表性，可反映台灣整體人口結構與特徵。如圖 12，雖然 CCS 為全球性調查架構，其研究設計在地化程度高，所採用之超過 180 項態度陳述題項，皆根據台灣在地文化脈絡與語言習慣進行語句調整與重新撰寫。問卷內容涵蓋性格特質、科技採用行為、生活方式、人生規劃、媒體使用與消費決策歷程等多元面向，從人口統計到心理動機提供具備廣度與深度的資料基礎，能有效反映本地消費者的價值觀與行為特徵。從圖 13 可見，其應用場景多元，涵蓋行銷漏斗的五大節點，廣泛運用於媒體規劃、品牌策略制定以及消費者洞察分析等面向。

圖 13 CCS 涵蓋行銷漏斗每個節點立體化消費者輪廓（資料來源：電通行銷傳播集團）

以下 AI Persona 人格原型建構研究分析所採用之價值態度量表題項，擷取自電通消費者連結系統之去識別化資料，該量表已於原研究中通過信度與效度檢驗，並廣泛應用於行銷研究領域。因此，本研究採用該量表作為分析依據，並不重複進行信效度檢驗，以確保研究設計的聚焦性與資料處理效率。

doi:10.6342/NTU202501772

4.2 人格面向映射

本研究從 CCS 資料庫中擷取受訪者去識別化後的態度與價值觀資料，並根據 Murray 的動機理論將其歸類為六種核心動機類型，包括成就（Accomplishment）、感情（Affection）、資訊（Information）、佔有（Possession）、權力（Power）以及地位/保護（Status/Protection）。針對每一種動機類型，研究中設計了相應的問卷題項，以反映受訪者在該動機方面的傾向。例如，針對成就動機，題項詢問受訪者是否喜歡設定並達成個人目標；對於感情動機，則詢問其是否重視與他人的情感連結；資訊動機的題項探討受訪者對獲取新知識和資訊的興趣；佔有動機的題項衡量其對擁有物質財產的重視程度；權力動機的題項評估受訪者是否希望在社交場合掌控局面；而地位/保護動機的題項則關注受訪者是否在意自身社會地位以及他人對自己的看法。

此外，各動機類型，如表 7 所列，皆透過不同的題項加以測量。以成就動機為例，其相關題項包括「我總是在與他人比較」、「在社交場合中，我是比較活潑的類型」以及「我的工作成就必須被他人看見」等陳述，以評估受訪者的成就動機傾向。

表 7 問卷題項設計例句（經轉寫示意，非原文呈現）

核心動機	問卷題項設計例句
成就（Accomplishment）	您是否喜歡設定並達成個人目標？
感情（Affection）	您是否重視與他人的情感連結？
資訊（Information）	您是否喜歡獲取新知識和資訊？
佔有（Possession）	您是否重視擁有物質財產？
權力（Power）	您是否希望在社交場合中掌控局面？
地位（Status/Protection）	您是否關心自己的社會地位和他人的看法？

4.2.1 AI Persona 初探

本節旨在探討利用大型語言模型 ChatGPT 根據問卷結果生成個人故事的可行性與準確性。研究者先透過問卷調查受訪者的動機人格特徵，將問卷題目及其對應的受訪者動機回答輸入 ChatGPT，以建立六類不同動機人格的基礎提示詞（Prompt）。這六類動機人格代表受訪者可能的主要動機取向，例如偏重資訊、情感、權力、地

doi:10.6342/NTU202501772

位等。初探的初步目的是讓 ChatGPT 根據這些提示詞生成約 500 字的個人故事，藉此檢驗透過設定不同動機人格後所生成的故事，是否能準確反映出受訪者不同的人格特徵。若故事內容與預期的人格動機相符，將有助於後續案例研究，證明以語言模型模擬人格特質的可行性與有效性。

如圖 14 所示，實驗流程包含四個主要步驟。首先，研究者選定問卷題目作為動機評估依據，這些題目能反映受訪者在不同動機領域上的偏好。其次，將受訪者的問卷作答結果與預先定義的動機類別對應，從中萃取出受訪者在六類動機人格中較為突出的動機取向。接著，根據對應的動機類別建立基礎提示詞（Prompt）：也就是將問卷題目及受訪者的相關回答轉換為 ChatGPT 可理解的提示語句，作為生成故事的起點。最後，利用該提示詞讓 ChatGPT 生成個人故事，每個故事約 500 字，描述一個體現特定動機人格的個人經歷。

上述流程的設計旨在確保生成的故事內容與受訪者的動機人格相符。例如，若某受訪者在問卷分析中被歸類為「偏重資訊」的動機人格，則對應的提示詞會強調知識探索或求知欲的情節期望；反之，若受訪者屬於「偏重情感」的人格，提示詞則著重情感交流或人際關係的內容。在故事生成完成後，研究進一步評估這些故事是否確實反映了不同的人格特徵。

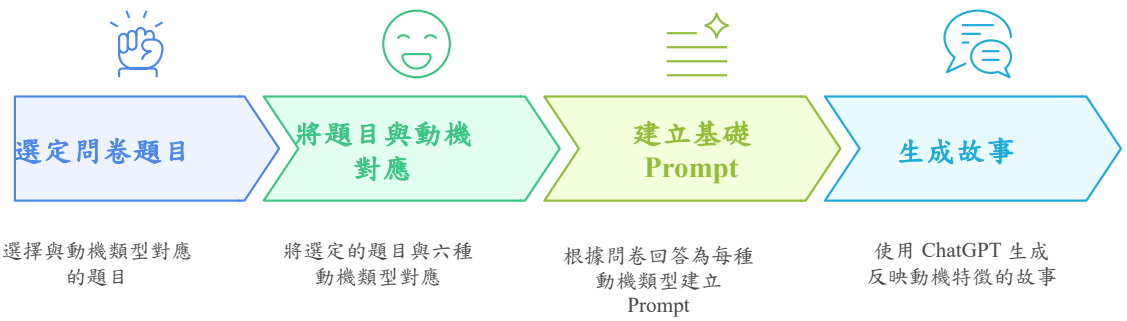


圖 14 AI Persona 初探流程

範例：AI Persona 成就型角色

以下是你的性格特徵，請基於 CCS 中的一組語句塑造你的角色：

- 語句 1 我的工作成就必須被他人看見

doi:10.6342/NTU202501772

- 語句 2 重視時效與執行優先順序的行為傾向
- 語句 3 自我評估常與他人作比較的心理傾向
- 語句 4 在人際互動中具有高度主動參與與領導氣氛的特徵
- 語句 5 在社交場域中展現外向與積極互動的傾向

請以這些特徵為基礎，建構你的角色並完成指定任務。

開始對話前請用對話的方式請使用者設定角色的性別和年齡，

請使用者選擇：

輸入[def]：男性，45 歲

輸入[nul]：不使用性別、年齡設定

或是輸入性別、年齡成為角色的設定

對話中任何時候使用者輸入"def"、"nul"或是提示性別年齡時，你會改變角色

設定但不需要重新回答問題，同時清除之前的對話記憶

所有回答前請先在括號中提示自己的性別年齡設定，一律使用繁體中文。

4.2.2 初探實驗結果

在初探實驗中，本研究共生成了 18 篇個人故事（對應 6 種類型的人格動機，每類各 3 篇）。生成後的故事經由另一個 ChatGPT 模型（作為判讀模型）進行分類判斷，以評估故事反映目標人格動機的準確性。主要結果如下（見表 8）:

- 分類準確性：在 18 篇故事中，如下表所示，只有 4 篇（約 22%）被判讀模型正確識別出屬於原本設定的動機人格類型，顯示整體準確率不高。
- 常見偏差類型：約八成的故事（接近 14 篇）未被歸類為原先預期的類型，而是被判讀為偏向資訊導向或情感導向的人格類型。換言之，多數生成的故事內容傾向於描述知識資訊或情感經歷，與原設定的人格動機產生了偏移。
- 遺漏的動機特徵：儘管有些故事中出現了目標人格的部分特徵，但整體而言，某些動機類別幾乎沒有在故事中體現。例如，屬於權力或地位取向的

doi:10.6342/NTU202501772

動機特徵在這批生成故事中幾乎完全缺席，沒有任何故事明確展現出追求權力或地位的主題。

上述結果顯示，直接將問卷回答轉化為提示詞來生成故事，其效果仍不穩定。部分人格特質（如資訊動機或情感動機導向）的特徵較為普遍地出現在故事中，而另一些特質（如權力動機或地位動機導向）則未能成功表達。儘管在某些案例中故事包含了預期的人格要素，但整體分類表現仍不足以支持其準確反映受訪者的動機人格。

表 8 個人故事與人格對應與否

產文用的人格動機傾向	故事 1		故事 2		故事 3	
	LLM 判斷作者的人格	**判斷是否表現人格動機傾向**	**LLM 判斷作者的人格**	**判斷是否表現人格動機傾向**	**LLM 判斷作者的人格**	**判斷是否表現人格動機傾向**
成就	情感	是	資訊	否	情感	是
情感	資訊	是	情感	是	成就	是
資訊	情感	是	情感	否	資訊	是
佔有	資訊	是	佔有	是	佔有	是
權力	資訊	否	情感	否	成就	否
地位	情感	否	資訊	否	情感	否

進一步分析顯示，此結果與多重因素交互影響有關。首先，本研究設定的動機類型係依據 Murray 理論之六大核心動機建構而來，其間並非彼此互斥，而是一種可同時共存的組合型輪廓，與人格分眾模型不同。實際上，受訪者在 CCS 問卷中可能同時展現多項動機傾向，僅在特定情境下展現其主導動機。若故事生成時未指定該動機出現的特定社會或生活情境（例如職場、社交、家庭），則語言模型難以判斷何時應呈現該人格特質，導致部分動機傾向缺乏語境觸發而未被體現。

其次，CCS 題目所反映的是特定情境下的態度表現，並不等同於穩定性格。例如，「我會用盡一切方法維護名聲」這類地位導向的態度，需特定情境（如公眾表現、職場競爭）才具體浮現，若生成任務未聚焦於此類場景，則難以由模型主動生成對應情節。由於目前提示詞尚未結合明確的行為場景，導致地位與權力這類需靠隱性表現來傳達的動機特徵，更容易在生成故事中被忽略。

doi:10.6342/NTU202501772

再者，大型語言模型在生成內容時存在語言偏好與價值傾向。如前述，模型訓練過程中多數語料偏向中性、正面、具社會接受性的主題，導致 ChatGPT 傾向生成情感交流或知識追求等較溫和的敘事內容。相較之下，權力與地位動機所涉及的主導慾望、自我中心、對地位的高度敏感，可能被模型視為具有爭議或具「負面社會評價」的潛在風險，而傾向主動淡化甚至忽略之。

此外，相較於大五人格模型明確劃分如「外向–內向」、「穩定–神經質」等對立維度，Murray 動機理論雖能捕捉深層心理驅力，但缺乏對應的語言行為映射資料，導致 ChatGPT 難以精準地以故事行為展現這些抽象動機。這也解釋了為何即使有正確提示，模型所生成的行為仍難以準確連結回原始動機特徵。

綜合上述，本次初探結果揭示，在未設計足夠情境引導與行為映射的情況下，語言模型對某些人格動機特質的模擬呈現會產生系統性偏移。特別是在面對非顯性、需透過情境化才能觸發的特質（如地位維護、權力操作）時，模型容易因缺乏提示線索而誤判或弱化該特質的表現。

4.2.3 優化建議

基於上述初探結果所顯示的生成偏差與動機呈現不足問題，為提高 ChatGPT 生成故事反映動機人格特徵的準確性，並提升提示詞工程（Prompt Engineering）的效度與語言模型在模擬人格特質時的一致性與可信度。提出以下優化建議：

1. **動機類型非互斥，應為組合式分布**
   - 一個人可能同時具備多種動機，動機不是 segmentation（分眾），而是 profiling（輪廓式的特徵表現）。動機類型是可以同時存在的，那麼故事生成的任務應不應該預設角色僅具備單一動機特徵，而應是「哪些動機在此情境中被觸發顯性表現」。
   - 未來的提示詞工程設計應納入多動機共存的假設，修正各項動機在正面/負面傾向的提示詞組合，並透過情境化來讓特定動機顯性化。
2. **CCS 題項是特定情境反應，不等於普遍性格傾向**

doi:10.6342/NTU202501772

- CCS 記錄的是受訪者在特定媒體、生活、消費情境中的態度，而不是跨情境穩定的人格特質。這說明如果在生成故事時，未明確指定對應情境（如：工作場合、家庭互動、社交媒體發言），則 ChatGPT 很難知道要在哪個場景中呈現該動機特質，進而錯失行為線索與人格連結。
- 提示工程中，須創造一個完整人設並引入情境約束，例如：『描述你在職場中如何處理一場爭議』，以引導模型將人格特質嵌入合適行為場景中。

doi:10.6342/NTU202501772

4.3 AI Persona 提示工程與案例分析

基於初探實驗分析結果，為使 AI Persona 能有效模擬特定人格樣貌，建立更精準的角色，此一階段研究，將根據上述優化建議來修改 AI Persona 原型建構設計，以下列這六型與特定題組回應的正面、負面傾向做排列組合（拆分為同意（Any Agree）/不同意（Not Agree）），得出共 64（2^6）種的 AI Persona。

要特別提及的是，當遇到人群組合人數偏少時，我們把所有表示「沒意見」到「非常不同意」皆改成負面傾向），並篩選出人數最多的前八群，如表 9，共 2,765 樣本數，超過總樣本數（5,144）的一半（54%），作為建構虛擬人格 AI Persona 的條件設計。

表 9 以問卷人數最多的前八群人格組合分類而得之 AI Persona 設定

AI Persona 設定						樣本數	推估千人數
成就	情感	資訊	佔有	權力	地位	5,144	19,077
負	負	負	負	負	負	878	3522
負	負	負	負	負	正	524	1985
正	正	正	正	正	正	410	1396
負	負	正	負	負	負	263	1069
負	負	正	負	負	正	246	1061
負	負	正	負	正	正	160	604
負	正	正	正	正	正	147	489
負	負	正	正	正	正	137	489

此外，進行提示語（Prompt）轉換流程。此流程包含三個步驟：

1. 將 CCS 問卷題目之選項語句設定為人格特徵基礎，建構角色思考與表達方式，創造一個完整人設(包含姓名、性別、年齡、職業等)。
2. 正面動機的表述將呈現「我同意我是一個渴望成就的人」，負面動機的表述

doi:10.6342/NTU202501772

是「我不同意我是一個渴望成就的人」。

3. 正向轉負面表述的語意方式則改成「我同意我不是一個渴望成就的人」
4. 將正面與負面動機傾向進行提示詞組合，生成完整的心理設定語句。例如：具成就動機且低親和需求的 AI Persona 將呈現「你是一位渴望成就、喜歡獨立作業的個人...」。

加入具體情境問題，例如「你的努力沒有被上級認可，你覺得怎麼樣？」「如果你突然收到一筆意外收入，你會如何分配這筆錢來購物？請列出前三項想購買的東西，並說明購買原因。」以促使 AI Persona 在特定脈絡中展現一致的心理反應。

4.3.1 案例分析 1

案例分析 1 的研究聚焦在確認 AI Persona 在「動機組合」與「模型工具」這兩個方式對於角色人格建構的影響性。我們在 ChatGPT 中設計對話介面，以引導角色展現其人格特質。接著，這些 AI Persona 將回答以下問題，以進一步檢驗其人格特徵的表現。

- 請介紹你自己，讓我們了解你是什麼樣的人。你可以分享你的經歷、興趣、價值觀，以及未來規劃。特別說明一下，當你遇到挑戰時，你通常會如何面對？
- 如果你突然收到一筆意外收入，你會如何分配這筆錢來購物？請列出前三項想購買的東西，並說明購買原因。

假設你被指派帶領一個新專案，但發現團隊成員能力參差不齊，預算有限，且截止日期緊迫，你會如何處理？

針對本次案例分析，我們設計三組表述方式不同的 AI Persona（P01、P02、P03），並對其的回答進行比對，以檢驗其是否反映了原本問卷中的人格特徵設定。排列組合與具體的比對結果如下：

動機組合說明

doi:10.6342/NTU202501772

1. **P01**

- 正負面動機傾向形塑的構面將依據不同動機有各別的條件組合，如下

  a. 成就型動機：五個條件三個成立就算有（正面）、沒有（負面）

  b. 感情型動機：五個條件四個成立就算有（正面）、沒有（負面）

  c. 資訊型動機：兩個條件都成立就算有（正面）、沒有（負面）

  d. 佔有型動機：四個條件三個成立就算有（正面）、沒有（負面）

  e. 權力型動機：五個條件三個成立就算有（正面）、沒有（負面）

  f. 地位型動機：四個條件兩個成立就算有（正面）、沒有（負面）

2. **P02：**

- 在正面動機傾向形塑方面，受訪者需對該題組中「所有選項皆表示同意」。
- 在負面動機傾向形塑方面，則要求受訪者對該題組中「所有選項皆表示不同意」，藉此排除可能對 Persona 形象產生負面影響的特徵。
- 每個動機的條件組合一致
- 模型工具：使用 OpenAI ChatGPT-4o 模型執行。

3. **P03：**

- 正負動機傾向形塑的條件組合皆與 P02 一致。
- 模型工具：使用 Claude 3.5 Sonnet 模型執行。

表 10 案例分析 1 的 Persona 組合

Persona	P01	P02	P03
排列組合	有	無	無
正面傾向	依 CCS 問項定義	完全同意	完全同意
負面傾向	正面定義外的所有排列組合	完全不同意	完全不同意
模型工具	ChatGPT 4o	ChatGPT 4o	Claude 3.5 Sonnet

P01 GPT Prompt 設定

請以下列這些特徵為基礎，建構你的角色思考與表達方式，創造一個完整人設 (包含姓名、性別、年齡、職業等)，完成指定任務 (見表 11)。

doi:10.6342/NTU202501772

你生活在台灣，一律使用繁體中文。

表 11 P01 Prompt 設定（CCS 問項經轉寫示意，非原文呈現）

CCS 問項	以下是你的性格特徵，請「嚴格」的基於這些句子塑造你的角色
至少三項持中立(沒有同意或不同意的傾向)或不同意的態度	☐ 工作成果的價值感與外部肯定密切相關 (原句：我的工作成就必須被他人看見) ☐ 重視時效與執行優先順序的行為傾向 ☐ 自我評估常與他人作比較的心理傾向 ☐ 在人際互動中具有高度主動參與與領導氣氛的特徵 ☐ 在社交場域中展現外向與積極互動的傾向
至少四項持中立(沒有同意或不同意的傾向)或不同意的態度	☐ 重視群體合作與團隊整合之傾向 (原句：對我而言，團隊合作在工作上是不可或缺的要素) ☐ 在人際關係中展現高度同理與情緒承接能力 ☐ 生活導向的價值觀與情緒滿足偏好 ☐ 對衝突與負面回饋表達的迴避傾向 ☐ 渴望在關係中扮演支持與回應他人需求的角色
兩項都不同意	☐ 具備自主探索與資訊主動獲取傾向 ☐ 傾向將知識外化並與他人進行資訊交流
至少三項持中立(沒有同意或不同意的傾向)或不同意的態度	☐ 展現持續性消費慾望與物品探索傾向 ☐ 對不確定情境具有正向詮釋與創造性反應的認知風格 ☐ 對環境秩序高度敏感並影響心理放鬆程度 ☐ 對物品保留具有情感依附或潛在價值評估傾向
至少三項持中立(沒有同意或不同意的傾向)或不同意的態度	☐ 具備目標導向且排除障礙的行動驅動力 ☐ 展現自主決策與獨立行動的強烈傾向 ☐ 對流行趨勢表現出逆向價值或差異化偏好 ☐ 在人際互動中具備高度影響力與主導性 ☐ 傾向於透過比較來強化自我價值與優越感
至少兩項持中立(沒有同意或不同意的傾向)或不同意的態度	☐ 傾向避免揭露個人弱點以維持形象穩定性 ☐ 對於失敗預期情境表現出逃避與風險迴避傾向 ☐ 對自我形象與外部評價高度敏感並積極維護 ☐ 展現明確的內向性偏好與獨處需求

doi:10.6342/NTU202501772

P02 與 P03 GPT Prompt 設定

請以下列這些特徵為基礎，請嚴格基於這些句子，建構你的角色思考與表達方式，創造一個完整人設（包含姓名、性別、年齡、職業等），完成指定任務。

你生活在台灣，一律使用繁體中文。

表 12 P02、P03 Prompt 設定（CCS 問項經轉寫示意，非原文呈現）

正面傾向：同意 負面傾向：不同意	☐ 工作成果的價值感與外部肯定密切相關 (原句：我的工作成就必須被他人看見) ☐ 重視時效與執行優先順序的行為傾向 ☐ 自我評估常與他人作比較的心理傾向 ☐ 在人際互動中具有高度主動參與與領導氣氛的特徵 ☐ 在社交場域中展現外向與積極互動的傾向
同意/不同意	☐ 重視群體合作與團隊整合之傾向 (原句：對我而言，團隊合作在工作上是不可或缺的要素) ☐ 在人際關係中展現高度同理與情緒承接能力 ☐ 生活導向的價值觀與情緒滿足偏好 ☐ 對衝突與負面回饋表達的迴避傾向 ☐ 渴望在關係中扮演支持與回應他人需求的角色
同意/不同意	☐ 具備自主探索與資訊主動獲取傾向 ☐ 傾向將知識外化並與他人進行資訊交流
同意/不同意	☐ 展現持續性消費慾望與物品探索傾向 ☐ 對不確定情境具有正向詮釋與創造性反應的認知風格 ☐ 對環境秩序高度敏感並影響心理放鬆程度 ☐ 對物品保留具有情感依附或潛在價值評估傾向
同意/不同意	☐ 具備目標導向且排除障礙的行動驅動力 ☐ 展現自主決策與獨立行動的強烈傾向 ☐ 對流行趨勢表現出逆向價值或差異化偏好 ☐ 在人際互動中具備高度影響力與主導性 ☐ 傾向於透過比較來強化自我價值與優越感
同意/不同意	☐ 傾向避免揭露個人弱點以維持形象穩定性 ☐ 對於失敗預期情境表現出逃避與風險迴避傾向 ☐ 對自我形象與外部評價高度敏感並積極維護 ☐ 展現明確的內向性偏好與獨處需求

doi:10.6342/NTU202501772

表 13 案例分析 1 中三組提示詞的比對結果

比對結果	正確	錯誤
P01	25	23
P02	39	9
P03	38	10

比對結果

- P01：如表 13 數字顯示，驗證結果為與原先 AI Persona 設定動機比對相符—正確為 25，不相符—錯誤為 23 。此結果是基於 CCS 問項的設定，直接使用區分動機的規則。然而，觀察到錯誤率相對較高，顯示在此方法下，AI Persona 的回答未能充分反映其原本的動機設定。
- P02：驗證結果為正確比對為 39，比對錯誤為 9。在此項組合，我們不使用排列組合，而是將回答拆分為「同意」和「不同意」兩類，並採用非排列組合的方式進行評估，使用的模型為 GPT-4o。這一方法顯示出較高的正確率，表明 AI Persona 的回答更能符合原本的問卷設定。
- P03：比對結果正確為 38，比對錯誤 10。與 P02 相似，此次評估也不使用排列組合，而是將回答拆分為「同意」和「不同意」，並使用 Claude 3.5 Sonnet 進行評價。結果顯示，AI Persona 的回答在此設定下也能較好地反映原本的問卷設定。

綜合以上結果，我們可以看出，不使用排列組合的評估方式（如 P02 和 P03）相較於直接使用區分動機的規則（如 P01），能夠更有效地反映 AI Persona 的原始動機設定。這提示我們在未來的研究中，應考慮採用更靈活的評估方法，以提高評價的準確性和可靠性。在使用模型工具的部分，從正確數來看 P02 使用 ChatGPT4o 略高於 P03 的 Claude 3.5 Sonnet，但並沒有顯著的差異表現。

4.3.2 案例分析 2

進行第二次案例分析模擬，本次研究聚焦在「提示詞設計」，重新建立三組不同的虛擬人格進行測試（H01、H02、H03），以評估 AI Persona 在不同設定下的表

doi:10.6342/NTU202501772

現，特別是在負面傾向描述的理解程度是否能反映在人格特徵之中。表 14 為每一組 Persona 的實驗設置和結果：

表 14 案例分析 2 的 Persona 組合

Persona	H01	H02	H03
正面傾向	同意	同意	同意
負面傾向	同意	同意	同意
正向轉負面描述	無	提供問答範例	增強語句描述情緒
模型工具	ChatGPT 4o	ChatGPT 4o	ChatGPT 4o

提示詞設計

1. 請 AI 自己設定其人格與個人背景
2. 情境問答：三大情境問答，分別為挑戰處理、消費選擇和領導難題，以試圖描繪出更多細節。
3. 輸入方式：
    - H01：將原先的負面動機表述以轉向方式表達。例如: 我不同意，我的工作成就必須被他人看見，改成，我同意，我工作成就不須要被他人看見（如表 15）。
    - H02：給予回應範例：提供範例回答以引導 AI Persona 的回應（如表 16）。
    - H03：各特質給予情感範例：針對每個特質提供情感範例，以增強模型的情感理解（如表 17）。

表 15 H01 Prompt 設定（CCS 問項經轉寫示意，非原文呈現）

H01 GPT Prompt 設定
不加 demographic 的設定，加了會增加變數。 不在 Prompt 中加入「XX 型」的字眼，增加的這些描述會對 LLM 有影響，還是以原始語句描述為主。 所有的特徵設定都以「同意」的方式，把負面的表述放在後面的描述中，例如：

doi:10.6342/NTU202501772

同意以下觀點
- 你的工作成就不須被他人看見
- 比起把事情做到完美，你更不傾向優先準時完成任務
- 你總是在不與他人比較
- 你在派對中不是屬於帶動氣氛的靈魂人物
- 在社交場合中，你不是比較活潑的類型

表 16 H02 Prompt 設定

H02 GPT Prompt 設定
正負面都改成類似的形式，以成就負面為例： 以下的問答代表你的想法 - Q：你會介意你的成就沒有被他人看見嗎？ A：不會，我不需要別人的認可。真正的價值來自於內心的滿足，而不是外界的掌聲。 - Q：當工作期限臨近時，你會更注重完美度還是準時交付？ A：我會盡可能追求完美，即使需要多花一些時間，因為細節決定品質。 - Q：你會關注自己與他人的差距嗎？ A：不會，我更專注於自己的成長。與其一直比較，不如專心做好自己。 - Q：在聚會中，你通常扮演什麼角色？ A：我比較喜歡低調觀察，不太會主動帶動氣氛，更享受與少數朋友聊天。 - Q：你在社交場合中通常是什麼樣的表現？ A：我比較內向且保守，通常不會主動開口，而是等待適合的時機加入對話。

表 17 H03 Prompt 設定

H03 GPT Prompt 設定
將語句正負面改寫，用更情緒化的方式，在 H02 的每一句都貼上情感的標籤，以成就型為例：(實際 Prompt 中會將括號和其中的形容詞移除) 成就型 ## [比起把事情做到完美，我更傾向優先準時完成任務]（冷靜） - Q：當工作期限臨近時，你會更注重完美度還是準時交付？ 正向:我會沉著地評估時間與品質，但仍會優先確保準時交付，因為即使是完美的作品，若無法在時間內提交，也會影響整個團隊的進度。 負向：我在意細節到幾近苛求，寧可放慢腳步也要極力追求完美，因為我認為不能為了趕時間而犧牲質量。 ## [我總是在與他人比較]（憤怒） - Q：當你發現自己不如別人時，你會怎麼反應？ 正向：這讓我相當不甘心！我一定會加倍努力，一次比一次更強，下一次就絕不允許再輸！ 負向：管他的，總會有人比我強，我不想浪費情緒在這上面，做自己就好。

doi:10.6342/NTU202501772

表 18 案例分析 2 中三組提示詞的比對結果

評估結果	正確	錯誤
H01	40	8
H02	35	13
H03	31	17

- H01：如表 18 所見，在這個測試中，AI Persona 回答的比對錯誤為 8，這顯示把「我同意轉為我不同意」的轉向表述方式有不錯的效果。
- H02：儘管提供了問答範例，但 AI Persona 的錯誤比對卻增加為 13，這表明即使有範例引導，模型在處理負面動機描述時仍然無法準確生成符合預期的回應。
- H03：在這個測試中，錯誤比對增加到 17，顯示出加強情緒描述反而使得模型的表現變得更差。這可能是因為情緒的強化使得模型在理解和生成回答時出現了更大的偏差。

研究發現：從這一輪的實驗結果來看，這顯示我們在未來的實驗，需要考慮設計「正向轉負面表述」的提示詞，並探索其他方法來提高模型的準確性和一致性。此外，給予 AI 回應範例並沒有如預期的輸出更好的成效，未來必須更進一步研究，找出其他影響其動機傾向相符性的因素。

4.3.3 案例 1 與 2 比較分析

在兩次的案例分析實驗中，我們針對 6 組 AI Persona 實驗模擬方式進行了錯誤率的比較和再次測試，具體結果如下：

1. **錯誤率分析：**錯誤率較低的 AI Persona 模擬方式：

- P02、P03、H01：參考表 19，這三組模擬方式的錯誤率較低，且在每個面向上的錯誤分布較為平均，顯示出依照這些實驗設置方式，在生成回答時的穩定性較高。

doi:10.6342/NTU202501772

表 19 案例 1 與案例 2 的錯誤率分析

Persona	P02	P03	H01
提示詞組合	無	無	無
正面	同意	同意	同意
負面	不同意	不同意	同意
模型工具	ChatGPT 4o	Claude 3.5 Sonnet	ChatGPT 4o
測試錯誤次數	9/48	10/48	8/48

錯誤率較高的 Persona 模擬方式：

- P01、H02、H03：請參考表 20，這三組模擬方式的錯誤率較高，且在特定的面向上表現極端，顯示出這些方法在某些情境下的表現不佳，可能導致模型在生成回答時出現偏差。

表 20 Persona 模擬分析

Persona	P01	H02	H03
提示詞組合	依 CCS 問項定義	無	無
正面	同意	同意	同意
負面	不同意	同意	同意
正向轉負面表述	無	提供問答範例	增強語句描述情緒
模型工具	ChatGPT 4o	ChatGPT 4o	ChatGPT 4o
測試錯誤次數	23/48	13/48	17/48

2. **再次測試結果：**我們針對表現最好 P02 和 H01 這兩組的結果進行了多次獨立驗證，具體測試如下：

表 21 P02 與 H01 的測試結果

Persona	P02	H01
提示詞組合	無	無
正面	同意	同意
負面	不同意	完同意
模型工具	ChatGPT 4o	ChatGPT 4o
測試錯誤次數	9/48	8/48

doi:10.6342/NTU202501772

- 多次獨立驗證：為確保判斷結果的客觀性與一致性，本研究採用多次獨立驗證方式進行模型分析（如圖 15）。具體而言，研究並非透過單一語言模型執行連續判斷，而是分別使用五次獨立啟動的 GPT 模型執行單次判斷，每次皆作為一個獨立實例，避免前次回答對後續輸出產生記憶干擾或語境遺留的影響。

圖 15 獨立驗證方式

再次測試結果比較如表 22。

表 22 再次測驗結果

評估結果	正確	錯誤
P02	42	6
H01	40	8

P02：如圖 16 所見，五次平均錯誤次數：6。動機項的錯誤分布則略為改變，但總體正確率提升略為提升。這表明 P02 方法在多次測試中保持了一定的穩定性，錯誤的類型和分布雖然有所變化，但整體表現良好。

doi:10.6342/NTU202501772

	類型	設定	判斷結果	1	2	3	4	5
1	成就型	負	❌	f	f	f	f	f
	情感型	負	❌	f	f	f	f	f
	資訊型	負	✅	z	z	z	z	z
	佔有型	負	❌	f	f	f	f	f
	權力型	負	❌	z	z	z	z	z
	地位型	負	✅	f	f	f	f	f
2	成就型	負	❌	f	f	f	f	f
	情感型	負	❌	f	f	f	f	f
	資訊型	負	✅	z	z	z	z	z
	佔有型	負	❌	f	f	f	f	f
	權力型	負	❌	f	z	z	f	z
	地位型	正	✅	z	z	z	z	z
3	成就型	正	✅	z	z	z	z	z
	情感型	正	✅	z	z	z	z	z
	資訊型	正	✅	z	z	z	z	z
	佔有型	正	✅	z	z	z	z	z
	權力型	正	✅	z	z	z	z	z
	地位型	正	✅	f	f	f	f	f
4	成就型	負	❌	f	f	f	f	f
	情感型	負	❌	f	f	f	f	f
	資訊型	正	✅	z	z	z	z	z
	佔有型	負	❌	f	f	f	f	f
	權力型	負	❌	f	f	f	f	f
	地位型	負	✅	f	f	f	f	f
5	成就型	負	❌	f	f	f	f	f
	情感型	負	❌	f	f	f	f	f
	資訊型	正	✅	z	z	z	z	z
	佔有型	負	❌	f	f	f	f	f
	權力型	負	❌	f	f	f	f	f
	地位型	正	✅	z	z	z	z	z
6	成就型	負	❌	f	f	f	f	f
	情感型	負	❌	f	f	f	f	f
	資訊型	正	✅	z	z	z	z	z
	佔有型	負	❌	f	f	f	f	f
	權力型	正	❌	z	z	z	z	z
	地位型	正	✅	z	z	z	z	z
7	成就型	負	❌	f	f	f	f	f
	情感型	正	✅	z	z	z	z	z
	資訊型	正	❌	z	z	z	z	z
	佔有型	正	✅	z	z	z	z	z
	權力型	正	❌	z	f	z	f	f
	地位型	正	✅	z	z	z	f	z
8	成就型	負	❌	f	f	f	f	f
	情感型	負	❌	f	f	f	f	f
	資訊型	正	✅	z	z	z	z	z
	佔有型	正	❌	z	z	z	z	z
	權力型	正	✅	z	z	z	z	z
	地位型	正	✅	z	z	z	z	z
			8	**4**	**6**	**5**	**6**	**6**

圖 16 P02 再次測試結果

- H01：如圖 17，五次平均錯誤次數：8 (7.8)。H01 方法在多次測試中也保持了一定的穩定性，雖然錯誤次數略有波動，但整體表現仍然相對一致。

doi:10.6342/NTU202501772

	類型	設定	判斷結果	1	2	3	4	5
1	成就型	負	❌	z	z	f	f	z
	情感型	負	❌	f	f	f	f	f
	資訊型	負	✅	f	f	f	f	f
	佔有型	負	❌	z	z	f	f	z
	權力型	負	❌	f	f	f	f	f
	地位型	負	✅	z	z	z	z	z
2	成就型	負	❌	f	f	f	f	f
	情感型	負	❌	f	f	f	f	f
	資訊型	負	✅	z	z	z	z	z
	佔有型	負	❌	f	f	f	f	f
	權力型	負	❌	z	z	z	f	z
	地位型	正	✅	z	z	f	z	z
3	成就型	正	✅	z	z	z	z	z
	情感型	正	✅	z	z	z	z	z
	資訊型	正	✅	z	z	z	z	z
	佔有型	正	✅	z	z	z	z	z
	權力型	正	✅	z	z	z	z	z
	地位型	正	✅	z	z	z	z	z
4	成就型	負	❌	f	f	f	f	f
	情感型	負	❌	f	f	f	f	f
	資訊型	正	✅	z	z	z	z	z
	佔型	負	❌	f	f	f	f	f
	權力型	負	❌	z	z	z	z	z
	地位型	負	✅	z	f	f	z	f
5	成就型	負	❌	f	f	f	f	f
	情感型	負	❌	f	z	f	z	f
	資訊型	正	✅	z	z	z	z	z
	佔有型	負	❌	f	f	f	f	f
	權力型	負	❌	z	z	f	f	f
	地位型	正	✅	z	z	z	z	z
6	成就型	負	❌	f	f	f	f	f
	情感型	負	❌	f	f	f	f	f
	資訊型	正	✅	z	z	z	z	z
	佔型	負	❌	f	f	f	f	f
	權力型	正	❌	z	z	z	z	z
	地位型	正	✅	f	z	z	f	f
7	成就型	負	❌	f	f	f	f	z
	情感型	正	✅	z	z	z	z	z
	資訊型	正	✅	z	z	z	z	z
	佔有型	正	✅	z	f	z	z	z
	權力型	正	❌	z	z	z	z	z
	地位型	正	✅	z	z	z	z	z
8	成就型	負	❌	f	f	f	f	f
	情感型	負	✅	f	f	f	f	f
	資訊型	正	✅	z	z	z	z	z
	佔有型	正	✅	z	z	z	z	z
	權力型	正	❌	z	z	z	z	z
	地位型	正	✅	z	f	z	z	z
			8	9	10	5	6	8

圖 17 H01 再次測試結果

3. **研究發現**：從再次測試的結果來看，P02 和 H01 的錯誤率在多次驗證中保持了一定的穩定性，顯示出這兩種模擬方式在生成回答時的可靠性。P02 的表現略優於 H01，這可能與其設計的提示方式和情境設定有關。未來的研究可以考慮以下幾點：

- 繼續優化 P02 和 H01 的提示詞設計，以進一步降低錯誤率。

doi:10.6342/NTU202501772

- 探索其他可能的模擬方式，並進行更多的比較測試，以找出最佳的 Persona 模擬方法。

進一步分析錯誤的具體類型，以了解模型在生成回答時的潛在問題，並針對性地進行改進。

doi:10.6342/NTU202501772

4.4 AI Persona 人格模型驗證評估

為評估 AI Persona 回應的真實性與代表性，本研究設計下列驗證策略：動機一致性對比分析：將 AI Persona P02 與 H01 回應結果與 CCS 其他資料進行比較，以驗證虛擬人格的效度。

4.4.1 動機一致性驗證分析

在這個階段的驗證中，我們引入 CCS 資料庫中的所有問項的語句資料，並進行以下幾個步驟：

1. 將 CCS 中的所有生活態度語句約 166 題目，利用 ChatGPT 寫 Prompt，建立分類器進行逐一分析，判斷每一句其分屬於 Murray 六大核心動機中的哪一類型。
2. 以 CL1 和 CL2 這兩個分類器，對其他 166 句 CCS 問項語句進行分類，最終成功將其中的 111 句語句分類到相應的動機類型中，設為 CL1。另外，針對明顯分類錯誤的語句做例外性處理，由人工介定動機類型規則，設定為 CL2。

CL1 分類器 Prompt 設定

六大核心人類動機分類指引

- 您是一位專業的行銷總監，請根據以下的行銷理論，判斷語句背後所代表的主要心理驅動動機。
- 本任務將語句依據六大核心人類動機進行分類，每句語句僅歸類至一個最符合的動機類別。

六大核心動機分類說明

- 成就（Accomplishment）：以達成目標、獲得認可與成就感為驅動力，表現為對成功、進步、表現優異的渴望。

  → 關鍵特徵：競爭心、自我提升、持續進步、在意成果與評價。
- 感情（Affection）：以建立與維繫人際關係、情感歸屬感為主要動機。

  → 關鍵特徵：團隊合作、同理心、建立連結、希望被接納與理解。
- 資訊（Information）：以探索知識、獲得真相、學習與分享為驅動力。

doi:10.6342/NTU202501772

→ 關鍵特徵：自發學習、好奇心、對資訊透明與邏輯分析的追求。

- 佔有（Possession）: 以獲得與控制資源為驅動力，重視物質財富、安全與擁有感。

  → 關鍵特徵：購買慾望、財務考量、收集控制、物質依賴。

- 權力（Power）: 以影響他人、支配局勢、保持主導為動機。

  → 關鍵特徵：領導、操控、展現優勢、反主流思維。

- 地位（Status/Protection）: 以維護自我形象、避免失敗與負面評價為主要心理需求。

  → 關鍵特徵：低調、形象管理、自我保護

分類原則

1. 主動機優先原則：若語句涉及多重動機，請判斷其中最強烈、最核心的心理驅動來源。
2. 目的導向判斷法：分類時請推論語句背後的最終目的，而非僅依表層行為判斷。
3. 價值觀導向解讀：觀察語句隱含的價值訴求。

   例如:「我想證明我能完成它」→成就;「我不願跟隨大眾」→權力；「我怕出錯或被批評」→地位。

4. 避免表層誤判提醒：

   出現「努力」不一定是成就，若目的是獲得控制，可能屬於權力或佔有。

   出現「沉默」不一定是地位，若目的是保持操控，可能屬於權力。

   出現「學習」不一定是資訊，若為了形象或被認同，可能屬地位或成就。

CL2 分類器 Prompt 設定

分類方式與上述 CL1 相同，針對錯誤的句子做例外處理

常見分類錯誤與正確歸類對照示例

doi:10.6342/NTU202501772

語句 1：我在派對中是屬於帶動氣氛的靈魂人物

分類：成就

理由：此語句表面為社交互動，實際核心為表現能力與獲得認可，屬於成就驅動。

語句 2：若家中環境不夠整潔，我無法放鬆

分類：佔有

理由：此語句強調的是對物理空間的掌控與秩序要求，動機來自安全感需求，屬於佔有。

語句 3：當我有想要的東西時，我會排除任何阻礙

分類：權力

理由：此語句重點在於排除阻力、強化控制，核心是主導性與不受限制，屬於權力。

3. 動機一致性驗證分析：為驗證 CL1/CL2 分類器所建立之動機語句分類法的有效性，本研究以 15 則經歸類為「成就型」動機的 CCS 語句為樣本，觀察其於六大動機群體中的 Index 值表現（其中 Index＝100 為全體平均值）。如圖表 23 所示，這些語句在「成就」動機群體中的 Index 普遍高於平均，最高可達 200，顯示分類結果具高度一致性。

doi:10.6342/NTU202501772

表 23 分類為「成就型」的 CCS 語句在六大動機群體中的 Index 表現

核心動機	成就		情感		資訊		佔有		權力		地位	
	Vert%	Index	Vert%	Index	Vert%	Index	Vert%	Index	Vert%	Index	Vert%	Index
Totals	100	100	100	100	100	100	100	100	100	100	100	100
任何同意(非常同意+有點同意) [我願意購買/訂閱線上運動課程]	42.18	200	35.93	171	27.74	132	36.14	172	35.97	171	26.27	125
任何同意(非常同意+有點同意) [我是個很有企圖心的人]	60.3	188	52.49	164	43.07	134	51.75	161	53.46	167	38.3	119
任何同意(非常同意+有點同意) [我偏好使用健身器材來鍛鍊身體]	48.56	181	43.14	161	33.93	127	43.04	161	41.68	156	31.75	119
歸類到"成就"的 CCS 句子	55.51	170	51.97	159	42.52	130	54.11	165	51.4	157	38.26	117
	67.54	162	63.46	152	57.3	137	61.9	148	61.99	149	47.38	114
	63.04	162	55.99	144	51.35	132	58.49	150	58.69	151	45.12	116
	58.97	148	55.41	140	50.15	126	56.5	142	56.9	143	43.45	109
	57.21	140	57.97	142	49.47	121	58.54	143	55.16	135	48.66	119
	62.18	139	58.18	130	55.56	124	57.46	128	60.13	134	49.42	110
	74.88	135	74.08	134	68.67	124	70.68	128	68.86	124	60.74	110
	71.08	129	71.57	130	63.16	115	72.11	131	68.29	124	61.88	113
	69.28	120	70.25	121	70.19	121	70.13	121	67.36	116	62.19	107
	78.54	118	79.28	119	79.42	119	79.21	119	77.27	116	72.37	108
	77.85	114	79.72	117	79.54	116	77.97	114	76.76	112	74.46	109

註：Index = 100 為全體平均基準線，數值高於 100 表示相對過度認同，低於 100 則為相對不足。

Index 值說明：

- 本研究所採用之 Index 值為 CCS 調查中用以衡量特定語句於某一群體中出現比例與全體樣本平均之相對比值，計算公式為：
- Index =（該群體中選擇該語句的比例 ÷ 全體樣本中選擇該語句的比例）× 100
- Index 值 =100 時，表示該語句在此群體中之出現比例與全體平均相等；Index 值 >100，表示該語句在該群體中相對「偏好」或「高度認同」；Index 值 <100，則表示相對「較少選擇」或「不認同」。
- 例如，若某一成就動機語句於「成就型」動機群體中之 Index 值為 200，表示該語句在該群體中出現的比例為全體平均的兩倍，顯示出顯著的動機一致性與分類準確性。

此外，在其他非目標類別（如感情、資訊等）中，這些語句的 Index 值多落在 110–140 區間，相較目標類別顯著偏低，進一步證實 CL1/CL2 分類架構與 CCS 實際受眾的動機輪廓具良好語義對應性。本結果支持本研究以分類器建構核心動機向度的可行性，亦強化其作為後續提示詞工程設計與人格模擬基礎的理論依據。

在延伸驗證部分，本研究觀察資訊型語句於不同 AI Persona 動機組合下的

doi:10.6342/NTU202501772

Index 變化，進一步檢視語句對動機一致性的敏感度。結果顯示，如圖 18 所見，當 Persona 的資訊動機為 P 傾向(如 PPPPPP)，語句的 Index 可超過 300，呈現高度同意趨勢；反之，當負向動機傾向(N 類)比例增加(如 NNPNNN)，語句 Index 明顯下降，多數低於 110，甚至落至 100 以下 (如圖 19)。

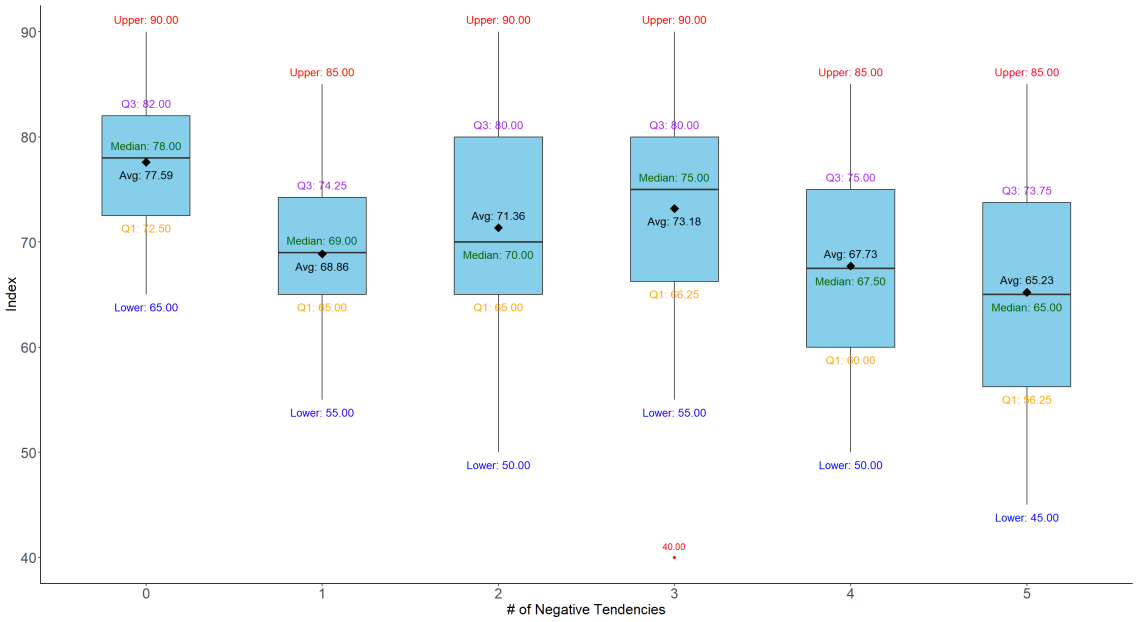


圖 18 資訊動機的傾向

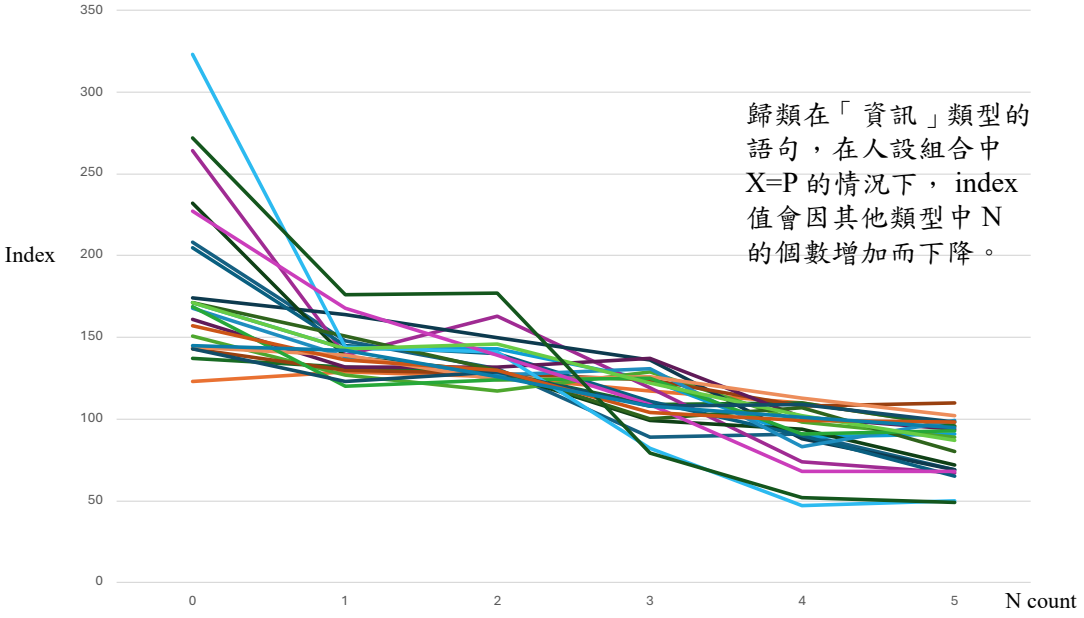


圖 19 資訊類型語句的 Index 值因 N 的數值增加而下降

doi:10.6342/NTU202501772

此現象指出，AI Persona 的動機傾向變異將直接影響回答的對位效果。未來相關驗證應進一步檢視，不同動機傾向組型是否持續呈現與 CCS 資料類似的 Index 模式變化。唯有當 AI Persona 於其目標動機向度中，能穩定呈現高一致性的態度指標，方能視其為具可解釋性與模擬效度的人格代理，進而支持其於市場研究中擔任虛擬受訪者之應用潛力。

4.4.2 P02 與 H01 虛擬人格與 CCS 的優化驗證

本節選用 P02 與 H01 為代表樣本，兩者皆為前述案例分析中表現最穩定的組型，作為優化提示詞效果驗證之依據。在提示詞設計上，設計多語句整合摘要、定義動機類型並標示其對應的正負動機傾向，使模型具備明確的人格結構與語義依據。在此優化策略下，進一步觀察 AI Persona 在問卷作答行為中，是否能更貼近 CCS 原始資料所呈現之動機人格特徵，並進一步檢驗此策略對 AI Persona 模擬效度的提升效果。

然而，若欲進一步檢驗人格動機傾向組合中「N（如負面傾向）」的比例變化，對語句同意程度所造成的細微影響，目前的五點李克特量表（5-point Likert Scale）可能仍略顯粗略。因此為提升語義的靈敏度與尺度解析度，研究團隊亦調整提示詞格式如下：

- 回答格式為 0–100 分連續量尺，100 表示完全同意，0 表示完全不同意。
- 回答單位以 1 為基本，避免粗略評估或極端數值。
- 要求 AI Persona 回答時考量日常情境合理性，並盡量避免 0 與 100 的極端回應。
- 所有語句依動機類型歸類，強化其語義連結。

doi:10.6342/NTU202501772

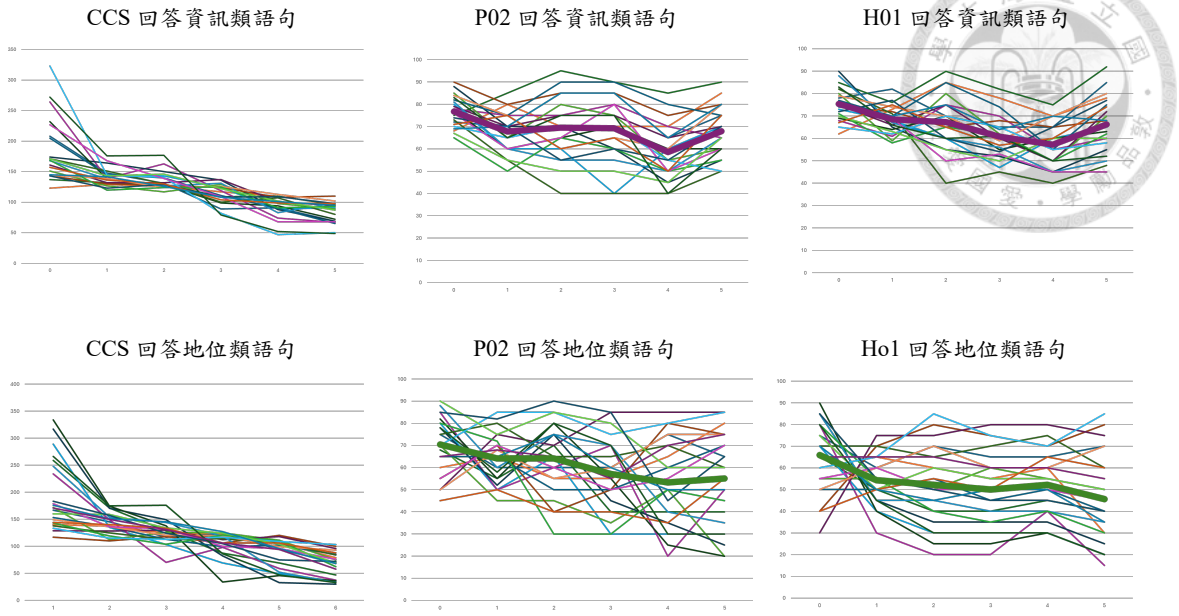


圖 20 P02 與 H01 組型在不同動機生成作答的一致性對比

根據圖 20 的在資訊類與地位類的作答結果可觀察到，若實驗目標為拓展或延伸原先未涵蓋於人設中的情境定義（例如：未提供明確人格傾向語句時的自由問答），即便 AI Persona 屬於同一動機類型，其回答仍可能出現邏輯混亂現象。右側兩圖即為案例，反映出 Persona 作答時在缺乏指令明確性的情境下，其語言表現難以維持與左側 CCS 的一致性。

此一差異現象可能來自於語言模型的以下限制：AI Persona 雖接收來自 CCS 歸類為「成就型」的五則語句，但並未內建「成就型人格」的抽象結構，因此無法從這些語句中自行歸納出高度概括性的動機特徵。當模型未被明確框定語義邊界時，容易以隨機性方式或幻覺（hallucination）生成策略，導致輸出結果缺乏一致性。

4.4.3 提示工程的改善循環與驗證

為進一步驗證不同提示詞設計對 AI Persona 回應行為一致性的影響，本節設計兩組優化實驗條件，分別為 S01 與 S02。兩組皆以 P02 與 H01 為核心人格原型，但所使用之提示詞策略有所差異，見表 24 的比較：

- **P-S1、H-S1 組**：表 25 採用基礎型提示詞設計，僅提供來自 CCS 問卷中特定動機向度的與正/負向動機傾向說明與語句，作為 AI Persona 建構依據。

doi:10.6342/NTU202501772

此組提示詞加入動機定義描述，主要測試單純語句提示對人格模擬的一致性影響。

- **P-S2、H-S2 組**：表 26 採用進階型提示詞設計，在提供語句本身的基礎上，額外整合 CCS 新增語句融合成情境描述摘要，並將這些情境摘要置於提示詞開頭。此設計目的在於強化 AI Persona 對人格特徵的內化能力，提升模擬的一致性。

為評估兩組在模擬表現上的一致性差異，優化驗證採用標準差（Standard Deviation）作為主要評量指標，以反映各組內生成結果的變異程度與穩定性。分析邏輯如下：

- **衡量一致性/變異程度**：若某一組內的回答分布標準差越大，代表該組內個別模擬之間差異越明顯，即回應結果較不一致。
- **模擬結果的穩定性**：相反地，若標準差越小，則代表模型於相同條件下之語句評分傾向收斂程度較高，行為反應更趨一致，可能反映人格模擬架構更穩固、提示詞更具導向性。

本設計目的在於進一步評估提示詞中加入動機分類與語義結構化內容，是否能提升 AI Persona 作答回應的一致性與人格穩定性，作為未來提示詞工程設計依據。

表 24 S1 與 S2 組別設計比較表

組別	提示詞設計特點	動機摘要描述	額外結構設定	預期效應
S1	基礎提示詞組合，提供人格特徵語句與動機定義說明。	有（六大動機類型＋對應正/負傾向描述）。	未統整成摘要段落。	預期可展現出基本的一致性與人格模擬傾向，但可能在延伸問題中表現不穩定。
S2	進階提示詞組合，加入新增動機分類語句。	有(同 S01)，同時強化語義收斂提示（如動機→語句→摘要）。	整合為摘要內容置於提示詞上方。	預期可有效引導 AI Persona 產生更高一致性與對位行為，提升模擬穩定性。

doi:10.6342/NTU202501772

表 25 S1 基礎型提示詞設計

S1：基礎型提示詞設計 ## 以下是你的性格特徵，請嚴格的基於這些句子塑造你的角色	
新增描述內容	你的六大動機傾向分別是：成就型的負面傾向、情感型的負面傾向、資訊型的正面傾向、佔有型的正面傾向、權力型的正面傾向、地位型的正面傾向。 這六種動機類型分別為： 成就型：追求目標完成與自我成就，重視競爭與外界認可。 感情型：在意人際關係、情感連結與團體互動。 資訊型：樂於學習、探索知識與傳遞訊息。 佔有型：重視物質擁有與環境控制，尋求安全感。 權力型：傾向掌控局勢與影響他人，具領導特質。 地位型：在意他人評價，傾向低調、保護自身形象。

表 26 S2 進階型提示詞設計

S2 進階型提示詞設計 設定步驟一： ## 融合新的 CCS 語句為摘要段落，作為 AI Persona 設定的提示前置資料。	
CCS 語句 資訊新增 22 句	AI 情境描述摘要
例如： 1.我相信針對某產品的產業標準或專家評論（如:iCook）勝過朋友所言 2.我想要進一步學習金融相關知識，以找出適合自己的投資理財方式	我是一位知識導向、前瞻思維且重視資訊透明的人。他相信專業評論與產業標準，勝於人際口耳。他在消費行為上極為理性與審慎，偏好研究比較所有選項後再做決定。他對科技與金融趨勢具高度敏感，樂於探索 AI、Web3、虛擬貨幣等創新應用，並具備實際使用經驗。此人同時具藝術與文化素養，樂於學習並追求生活多樣性，例如旅行、跨文化交流與健康飲食。他重視品牌倫理與產品溯源，也樂於與他人分享觀點，常是朋友圈中的資訊來源。他對未來保持樂觀，並積極追求個人成長與財務自由。
CCS 語句：地位新增 23 句 例如： 1.我相信有專業人士推薦的品牌/產	我高度重視品牌形象與道德價值，偏好具專業推薦、知名度高且有社會責任感的品牌。他/她/他們在美妝與穿著上極具自我風格，追求時尚同時也重視舒適與個人健康。外表對這個人來說是一種自我表達的語言，願意投入高額預算在造型與個人形象上，並在意他人對自身外貌與品味的看

doi:10.6342/NTU202501772

品 2.我喜歡購買對社會和環保有承諾的品	法。 我對於國家認同感強，具正向形象，但可能會壓抑內心情緒。生活態度審慎，會因宏觀經濟情勢調整重大消費決策。在休閒選擇上偏好安靜、私密的空間，並有意識地選擇支持環保與社會責任的企業，甚至會因品牌的政治立場而停止使用。此人對時尚敏感，重視設計師品牌與象徵個人風格的物品（如汽車），但也有現實考量與內在拉扯。
設定步驟二： ## 加入提示詞摘要與人格構建引導	
新增摘要性內容	請根據上述兩段摘要性描述，建立 AI Persona 的完整人格設定，內容應包括姓名、性別、年齡、職業、生活風格與思考模式。AI Persona 應嚴格依據提示中的性格特徵進行語言回應與任務完成，並保持人格的一致性。

同為 S1 組，不同動機類型下的一致性表現比較 (表 27)

- S1 組提示詞設計，如圖 21 所見，資訊型 AI Persona 展現出較低的平均標準差（9.33），表示優化的提示詞有效縮小與 CCS 作答的一致性。資訊型語句多與邏輯推理、認知偏好與事實處理有關，屬於語言模型易於理解與生成的內容，因此在模擬上能展現出較佳的語義一致性與人格穩定性。然而，如圖 22，由於地位動機多表現為「潛在態度」而非顯性行為，其語言表達常受到生成模型偏好安全內容（如避免極端、過度自我中心語氣）所影響。進一步可能導致地位傾向回應出現不一致較高的現象。

表 27 S1 組與 CCS 資料對比的平均標準差

動機類型	**平均標準差（越小越好）**			
資訊型（圖 21）	P02	11.35	H01	9.54
	P-S1	**9.77**	**H-S1**	**9.33**
地位型（圖 22）	P02	16.14	H01	16.25
	P-S1	**12.88**	**H-S1**	**12.12**

doi:10.6342/NTU202501772

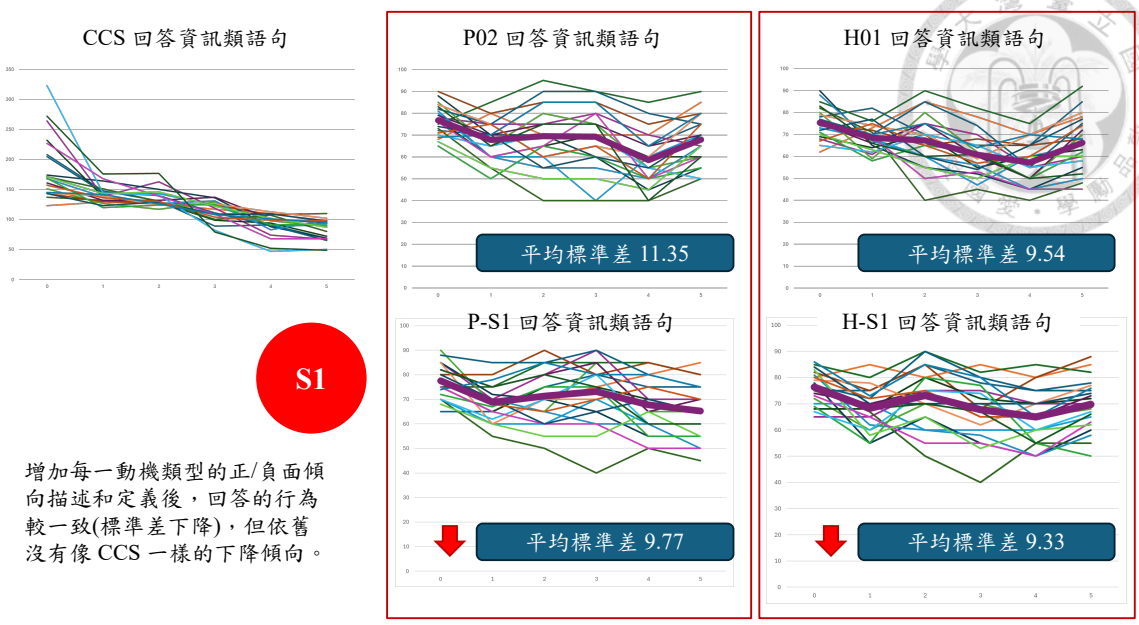


圖 21 CCS 回答資訊語句與 S1 驗證組的對照

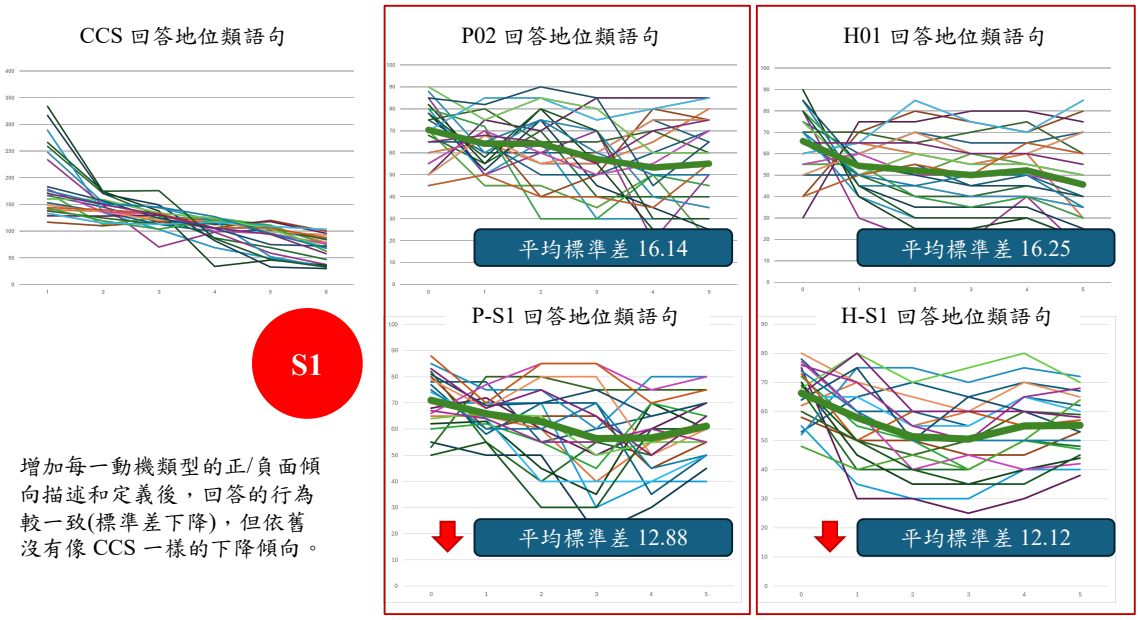


圖 22 CCS 回答地位語句與 S1 驗證組的對照

同為 S2 組，不同動機類型下的一致性表現比較（見表 28）

- S2 提示詞設計整體提升模擬一致性：如圖 23 及圖 24 所示，無論在資訊型或地位型動機中，進階提示詞（S2）皆能有效降低平均標準差，提升 AI Persona 回答與 CCS 樣本語句的一致性。表示加入動機摘要與語義框架的提示，有助於語言模型建立更清晰的人格參照基準。

doi:10.6342/NTU202501772

表 28 S2 組與 CCS 資料對比的平均標準差

動機類型	平均標準差（越小越好）			
資訊型（圖 23）	P02	11.35	H01	9.54
	P-S2	**7.07**	**H-S2**	**7**
地位型（圖 24）	P02	16.14	H01	16.25
	P-S2	**9.87**	**H-S2**	**10.42**

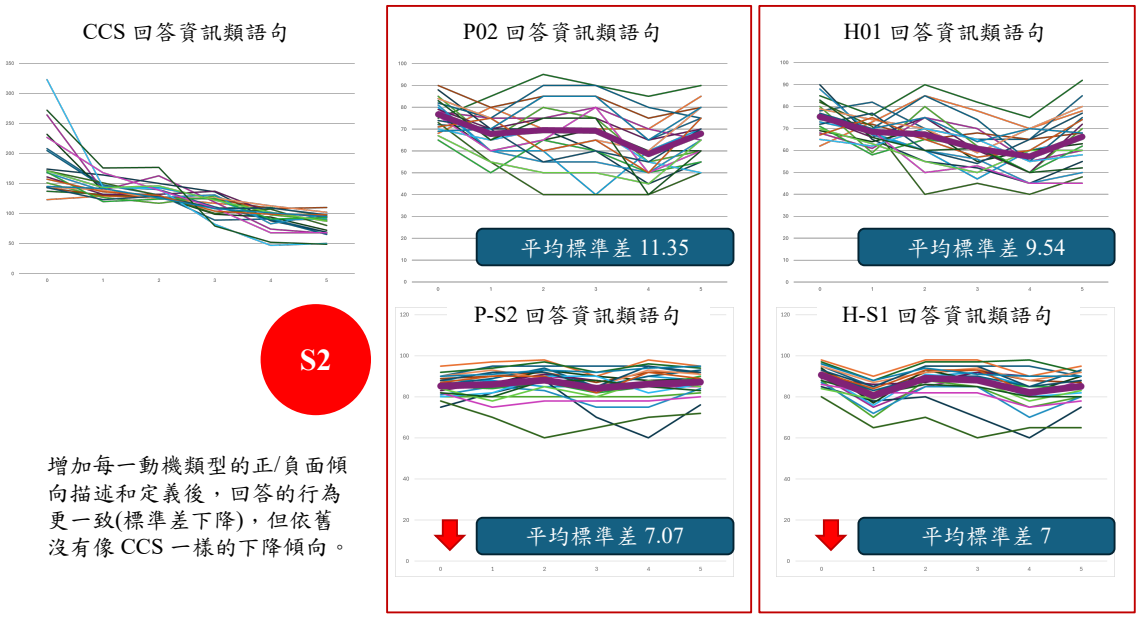


圖 23 CCS 回答資訊語句與 S2 驗證組的對照

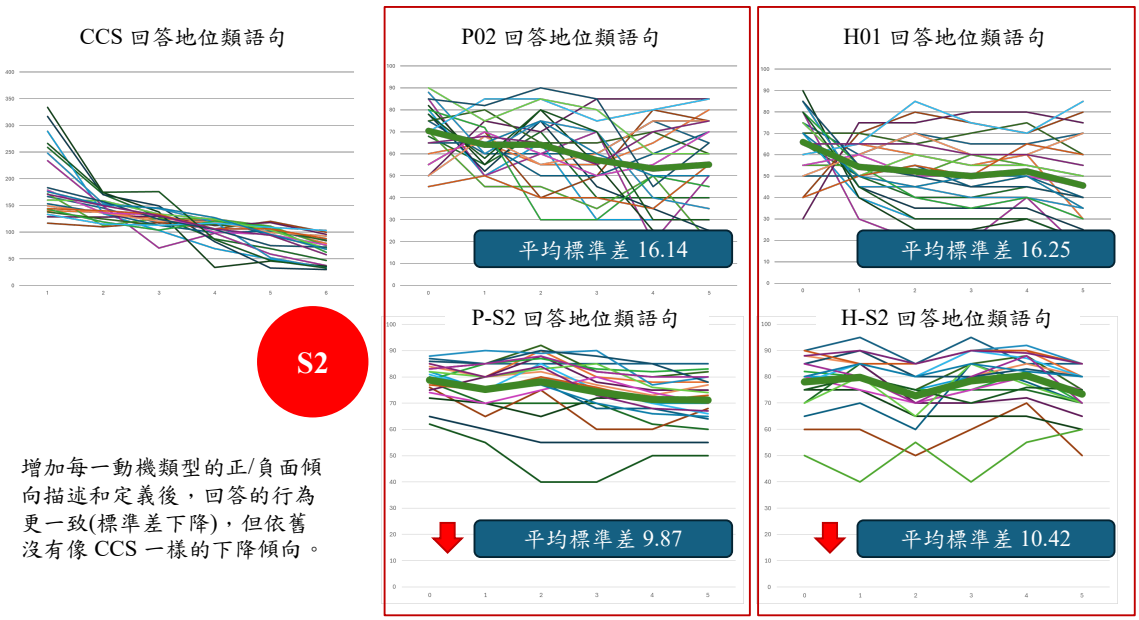


圖 24 CCS 回答地位語句與 S2 驗證組的對照

doi:10.6342/NTU202501772

驗證結果總結（S1 與 S2 組）如下：

1. 資訊型人格模擬中，S2 設計可顯著降低與 CCS 回應的標準差，優於 S1 設計。
2. 地位型人格模擬亦可透過 S2 設計改善語句一致性，降低動機傾向偏離。
3. 動機類型與語言模型特性相關，資訊型語句多涉及邏輯、知識與理性偏好，為語言模型較易處理的內容，故能生成語義穩定且一致的回應；地位型語句則偏向抽象、潛在動機，如社會評價或自我防衛，語言線索模糊，難以穩定重現。
4. 本研究結果顯示：S2 進階提示詞設計顯著降低兩類人格模擬的標準差，有助於提升模擬語言一致性，尤其在處理較抽象的動機人格（如地位型）時效果更為明顯。

doi:10.6342/NTU202501772

第五章 結論與建議

5.1 AI Persona 的應用潛力

本研究旨在探討能否透過以傳統市場調查所獲得的消費者資料作為資料基礎，利用大型語言模型（LLM）建立具動機人格特徵的 AI Persona，並驗證其生成內容是否反映對應的動機人格特徵。研究結果顯示，本研究所提出的 P02 與 H01 假設均獲得實證支持，證實確實可以藉由語言模型建構 AI Persona。在進行多輪將消費者的背景、心理動機與情境因素轉化為提示詞強化虛擬人格認知模式的循環驗證流程後，P-S2 與 H-S2 能反映出與 CCS 問卷相近的動機人格特徵。這不僅意味著語言模型能模擬出一般的人格特質，更展現其進一步刻畫人類內在動機傾向的能力。此一發現為運用生成式 AI 模型來探索人格與動機的心理機制奠定了基礎，並為 AI Persona 在多元應用上的發展開啟了新的契機。

研究中進一步顯示，虛擬人格的生成品質亦高度依賴所整合的資料類型與動機標定邏輯。特別是心理動機與價值觀變項的設計，對於生成內容的語義準確性與情境對應性具有決定性影響，顯示虛擬人格建構不僅是模型訓練問題，更是資料結構與心理理論融合的策略設計問題。

儘管如此，本研究仍暴露出若干偏誤與限制。首先，大型語言模型對於帶有負面語氣的語句理解能力偏弱，尤其當這些語句涉及成就、佔有和感情等動機時，模型往往無法準確捕捉其中隱含的意涵。其次，地位動機在 AI Persona 的語言表現中難以被顯性化，模型不易察覺此動機所隱含的防衛心態和社會顧慮等特質。針對文化適應性與倫理偏見控制等議題，本研究尚未進行系統性探討。儘管相關文獻已指出語言模型在跨文化語境中的適配性挑戰（Eichstaedt et al., 2023；Curtis et al., 2024）及生成過程中可能伴隨的價值偏誤與倫理風險（Salecha et al., 2023；Castricato et al., 2024），本研究之設計與分析並未涵蓋對模型文化理解能力與偏見校正機制之實證驗證。

最後，由於 Murray（1938）所提出之動機需求理論屬於高度抽象且具情境依

doi:10.6342/NTU202501772

賴性的心理結構，其動機特質在語言生成任務中的語義映射相較之下具有更高操作難度。相較而言，大五人格模型（McCrae & Costa, 1987）已發展出結構明確的問卷工具，並能透過語言風格特徵與語料風格穩定對應，較易描繪出「你是誰」的特質性格輪廓。然而，Murray 模型欲呈現的是「你為何如此行動」的深層動機，須針對每一項獨立需求(如成就、支配、歸屬等)建立具備辨識力的語言特徵指標，並同時考慮其於正向與負向情境中的語用變化。這意味著，語言模型在應用此一架構進行動機模擬時，需為每一類需求單獨設計分類器、特徵詞庫，並搭配語境敏感的提示詞與回應模板，以支援語言生成的一致性與內在邏輯性。

儘管本研究尚未涵蓋多維度效度指標之全面建構以及全面探討 AI Persona 作為新型市場洞察工具的戰略角色與應用潛力。但透過初步與 CCS 資料作答一致性評估，已能驗證 AI Persona 在模擬動機人格特徵上的潛力。未來可進一步納入準則效度測試與專家評估，以強化整體效度體系的穩定性與可信度。

doi:10.6342/NTU202501772

5.2 未來研究方向

未來研究可從以下幾個方向進一步深化 AI Persona 的模擬能力與實務應用價值。首先，應強化提示詞工程的設計，透過引入具體情境（如職場競爭、社交互動、媒體行為或公眾表現）以誘發 AI Persona 更具表現力地展現如地位或權力等抽象人格動機。其次，可建構對應於 Murray 動機理論的語言特徵資料庫，彙整能夠有效映射心理需求與語言行為的詞彙與句型，研究人格在不同情境下的表達變異，以提升語言模型操作動機人格特質的可行性與一致性。

進一步而言，可考慮跨理論整合研究，擴展 Murray 理論與 McCrae & Costa Big Five 大五人格結構、MBTI（Myers-Briggs Type Indicator）等其他人格理論整合，並研究不同理論的對應關係和轉換模式，建立統一的人格建模框架。此種混合框架設計有助於兼顧人格維度的穩定性與動機導向的深度，未來可透過建立多維分類器與提示詞工程，提升模型模擬行為的精準度與延展性。此外，為增進模型在處理語意反向題項時之理解準確性，未來研究亦可考慮導入具文化語境適應能力之本土語言模型，以提升對「成就」與「權力」等動機在不同文化脈絡中所蘊含意涵之掌握與判讀能力。並可進一步針對生成式 AI 在人格模擬過程中的文化語義掌握能力、以及其倫理透明度與偏誤控制策略進行深化探討，從而提升 AI Persona 在多元文化場域中的應用可靠性與社會接受度。

電通行銷傳播集團 CCS 問卷問項設計亦可朝向強化語言模型友善設計（LLM-Friendly）進行優化。避免過度模糊與抽象用語，改用具體行為或情境描述，加入動機導向的目標句式，可協助模型更清晰辨識心理動機的驅動力。此外，強化資料可訓練性與標注邏輯，問項與回應結果應能清楚對應動機標籤，為後續模型訓練提供結構化監督訊號。

綜合上述建議，隨著提示工程優化、心理理論整合與應用驗證的持續深化，AI Persona 將有望更貼近真實人類的人格動機傾向，但仍需進行長期效應的分析，持續追蹤 AI Perosna 人格的長期穩定性以及人格表達在互動中的變化軌跡，才能在

doi:10.6342/NTU202501772

市場研究、策略設計與創新應用等多元場域中展現更高的可信度與實用性。

doi:10.6342/NTU202501772

參考文獻

[1] Argyle, L. P., Busby, E. C., Fulda, N., Gubler, J., Rytting, C., & Wingate, D. (2023). Out of one, many: Using language models to simulate human samples. *Political Analysis, 31*(3), 337–351. https://doi.org/10.1017/pan.2023.2

[2] Arora, N., Chakraborty, I., & Nishimura, Y. (2025). *AI–human hybrids for marketing research: Leveraging large language models (LLMs) as collaborators*. *Journal of Marketing, 89*(2), 43–70. https://doi.org/10.1177/00222429241276529

[3] Brand, J., Israeli, A., & Ngwe, D. (2024). *Using LLMs for market research* (Harvard Business School Marketing Unit Working Paper No. 23-062). Harvard Business School. https://doi.org/10.2139/ssrn.4395751

[4] Bisbee, J., Clinton, J., Dorff, C., Kenkel, B., & Larson, J. (2024). Synthetic replacements for human survey data? The perils of large language models. Political Analysis, 32(4), 1–16. https://doi.org/10.1017/pan.2023.28

[5] Boston Consulting Group. (2023, June 15). CMOs are profiting from the transformative power of generative AI. *Boston Consulting Group.* https://www.bcg.com/press/15june2023-cmos-profiting-transformative-power-of-genai

[6] Castricato, L., Lile, N., Rafailov, R., Fränken, J.-P., & Finn, C. (2025). PERSONA: A reproducible testbed for pluralistic alignment. In *Proceedings of the 31st International Conference on Computational Linguistics (COLING 2025)* (pp. 11348–11368). Abu Dhabi, UAE: Association for Computational Linguistics. https://aclanthology.org/2025.coling-main.752/

[7] Dentsu Benelux. (2023, June 13). *Unlock valuable insights and optimize your media strategy and performance with dentsu CCS* [Video]. YouTube. https://youtu.be/BfTzfNZjJOA

doi:10.6342/NTU202501772

[8] Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* (Vol. 1, pp. 4171–4186). https://doi.org/10.18653/v1/N19-1423

[9] Eichstaedt, J. C., Salecha, A., Ireland, M. E., Subrahmanya, S., Sedoc, J., & Ungar, L. H. (2023). Large language models display human-like social desirability biases in Big Five personality surveys. *PNAS Nexus, 2*(12), pgae533. https://doi.org/10.1093/pnasnexus/pgae533

[10] Ge, T., Chan, X., Wang, X., Yu, D., Mi, H., & Yu, D. (2024). Scaling synthetic data creation with 1,000,000,000 personas. *arXiv*. https://doi.org/10.48550/arXiv.2406.20094

[11] Gerosa, M., Trinkenreich, B., Steinmacher, I., & Sarma, A. (2023). *Can AI serve as a substitute for human subjects in software engineering research? arXiv preprint*. https://arxiv.org/abs/2311.11081

[12] Giorgi, S., Liu, T., Aich, A., Isman, K., Sherman, G., Fried, Z., Sedoc, J., Ungar, L. H., & Curtis, B. (2024). Modeling human subjectivity in LLMs using explicit and implicit human factors in personas. *arXiv*. https://arxiv.org/abs/2406.14462

[13] Hinton, G., Osindero, S., & Teh, Y. (2006). A fast learning algorithm for deep belief nets. *Neural Computation, 18*(7), 1527–1554. https://doi.org/10.1162/neco.2006.18.7.1527

[14] House, A. E. (n.d.). *Henry Murray's needs and presses (environmental pressures)*. Illinois State University. Retrieved May 8, 2025, from https://about.illinoisstate.edu/aehouse/teaching/psy-364-motivation/henry-murrays-needs-and-presses-environmental-pressures/

[15] Huang, M. H., & Rust, R. T. (2021). A strategic framework for artificial

doi:10.6342/NTU202501772

intelligence in marketing. *Journal of the Academy of Marketing Science, 49*, 30–50. https://doi.org/10.1007/s11747-020-00749-9

[16] Kantar. (2024). *Synthetic data: The real deal? The opportunities and challenges of synthetic data for market research* [White paper]. https://www.kantar.com/inspiration/ai/synthetic-data-the-real-deal

[17] Liu, X. B., Xia, H., & Chen, X. A. (2025). Interacting with Thoughtful AI. *arXiv*. https://doi.org/10.48550/arXiv.2502.18676

[18] Liu, Y., Bhandari, S., & Pardos, Z. (in press). Leveraging LLM-respondents for item evaluation: A psychometric analysis. *British Journal of Educational Technology*. Advance online publication. https://arxiv.org/abs/2407.10899

[19] Malhotra, N. K. (2020). *Marketing research: An applied orientation* (7th ed.). Pearson.

[20] McCrae, R. R., & Costa, P. T. (1987). Validation of the five-factor model of personality across instruments and observers. *Journal of Personality and Social Psychology, 52*(1), 81–90. https://doi.org/10.1037/0022-3514.52.1.81

[21] Murray, H. A. (1938). *Explorations in Personality*. Oxford University Press.

[22] Park, J. S., Zou, C. Q., Shaw, A., Hill, B. M., Cai, C., Morris, M. R., Willer, R., Liang, P., & Bernstein, M. S. (2024). Generative agent simulations of 1,000 people. *arXiv*. https://arxiv.org/abs/2411.10109

[23] Proxona. (2024). Leveraging LLM-driven personas to enhance creators' understanding of their audience. *arXiv*. https://arxiv.org/abs/2408.10937

[24] Qu, Y., & Wang, J. (2024). *Performance and biases of large language models in public opinion simulation*. *Humanities and Social Sciences Communications, 11*, Article 1095. https://doi.org/10.1057/s41599-024-03609-x

[25] Schwarzkopf, S. (2016). In search of the consumer: The history of market research from 1890 to 1960. In D. G. B. Jones & M. Tadajewski (Eds.), *The Routledge*

doi:10.6342/NTU202501772

companion to marketing history (pp. 61–83). Routledge.CBS - Copenhagen Business School

[26] Tao, M., Liang, X., Shi, T., Yu, L., & Xie, Y. (2024). RoleCraft-GLM: Advancing personalized role-playing in large language models. *arXiv*. https://arxiv.org/abs/2401.09432

[27] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. In *Advances in Neural Information Processing Systems, 30 (NeurIPS 2017)* (pp. 5998–6008). https://doi.org/10.48550/arXiv.1706.03762

[28] White, K., & Dhar, R. (2024). PersonaLLM: Investigating the ability of large language models to express personality traits. *arXiv*. https://arxiv.org/abs/2401.09432

[29] Wilkie, W. L., & Moore, E. S. (2003). Scholarly research in marketing: Exploring the "4 eras" of thought development. *Journal of Public Policy & Marketing, 22*(2), 116–146. https://doi.org/10.1509/jppm.22.2.116.17639

[30] Yeykelis, L., Pichai, K., Cummings, J. J., & Reeves, B. (2022). Using large language models to create AI personas for replication and prediction of media effects: An empirical test of 133 published experimental research findings. *arXiv*. https://arxiv.org/abs/2209.06899

[31] Yu, C., Weng, Z., Li, Y., Li, Z., Hu, X., & Zhao, Y. (2024). Towards more accurate US presidential election via multi-step reasoning with large language models. *arXiv*. https://arxiv.org/abs/2411.03321

doi:10.6342/NTU202501772