

國立成功大學

工程科學系

碩士在職專班論文

基於生成式 AI 方法開發的行動學習與導覽系統

**Developing a Mobile Learning and Tour-guiding System based on  
Generative AI Approach**

研究生：郭富凱

指導教授：陳牧言 博士

中華民國 113 年 6 月

## 摘要

由於目前市面上的多模態大型語言（Multimodal Large Language Model, MLLM）模型框架無論是訓練、微調、運行的成本都十分高昂，本研究提出了一套新穎、低能源消耗、低運行/微調成本的新架構，能夠在單板機上離線運行。本研究使用量化（Quantization）方法載入大型語言模型（Large Language Model, LLM）作為系統的核心大腦並使用了檢索增強生成（Retrieval-Augmented Generation）加強特定領域的知識，同時整合了 You Only Look Once（YOLO）物件識別模型作為視覺解碼器，並融入了 OpenAI 的 Whisper 作為語音輸入與 Linux 內建的 eSpeak 作為語音輸出，實現了 MLLM 多模態的思考、視覺、聆聽和說話的能力，且能使用中文語音互動，本研究將此系統實作於 Jetson Orin NX 上。接著，本研究由淺入深地探討了雲端運算、邊緣運算乃至嵌入式 AI、TinyML 的概念，以確立本研究系統的定位。接著探討 LLM、MLLM、通用人工智慧以及與物連網之間的關係，確定了本研究提出系統的重要性。

本研究將此套系統應用於行動學習與數位導覽領域，並鎖定了臺南孔廟作為系統驗證的場域，將受測者分為實驗組與對照組，實驗組由本研究提供的設備進行導覽，而對照組則由真人導覽員帶領，兩組各經歷了前測、後測，確定了本研究所能帶來的學習成效相較於真人導覽員提高了 18.0%。

並且鑒於目前專門探討 AI 使用意圖相關的研究模型，只有社交導向的 AIDUA（Artificially Intelligent Device Use Acceptance）以及任務導向的 T-AIA（Task-oriented AI Acceptance）。並無法完全適用於生成式人工智慧（Generative Artificial Intelligence, GAI）設備。也因此，本研究基於理性行為理論（Theory of Reasoned Action）提出了一套創新的研究模型，同時做為評估本系統成效的重要依據，探討 GAI 的使用意圖（Generative Artificially Intelligent Use Acceptance, GAIUA），共使用了六個構面。包括，解決問題、感知人性、回應內容、類人互動、滿意度、使用意圖等。研究結果顯示，「感知人性」和「類人互動」是影響「滿意度」和「使用意圖」的關鍵構面。其中，「類人互動」對於提升使用者滿意度和使用意圖的影響最為顯著。

**關鍵詞：**嵌入式 AI、多模態大型語言模型、物件辨識、行動學習、導覽系統、GAI 使用意圖。

## **ABSTRACT**

### **Developing a Mobile Learning and Tour-guiding System based on Generative AI Approach**

Author : Fu-Kai Kuo

Advisor : Mu-Yen Chen

Department of Engineering Science, College of Engineering

National Cheng Kung University

## **SUMMARY**

Given the high costs associated with training, fine-tuning, and operating existing Multimodal Large Language Model (MLLM) frameworks on the market, this study proposes a novel, low-energy consumption, and low-cost framework for running and fine-tuning. This framework can operate offline on Single Board Computers (SBCs). This system is applied to mobile learning and digital tour domains, with Tainan Confucius Temple chosen as the verification site. Considering the lack of research models specifically exploring the intention of using GAI, this study proposes an innovative research model based on the Theory of Reasoned Action. This model serves as an important basis for evaluating the effectiveness of the system and explores the Generative Artificially Intelligent Use Acceptance (GAIUA).

## **INTRODUCTION**

Artificial Intelligence (AI) has achieved unprecedented accomplishments in various application fields. Recently, Generative Artificial Intelligence (GAI), which has become a focal point of research and application, is seen as a stepping stone towards Artificial General Intelligence (AGI). In response to the impact of GAI on education, UNESCO has proposed that it is time to prepare the next generation of students for the future. AI will become an increasingly important part of their lives, and UNESCO recommends incorporating GAI as an educational tool into curriculum policies.

In recent years, especially after the emergence of GPT-4, there has been growing interest in expanding the multimodal capabilities of Large Language Models (LLMs). Multimodal Large Language Models (MLLMs) use LLMs as their brain, extending their ability to process additional information such as images, videos, and audio to perform multimodal tasks.

Additionally, due to the complexity of GAI tasks, these models are typically large-scale and computation-intensive, requiring powerful centralized computing facilities (i.e., cloud servers) to handle user requests.

However, because general mobile devices or computers have limited processing power, instructions to LLMs are usually processed in the cloud. This is the standard procedure for ChatGPT and most chat applications supporting LLMs. This reliance on network resources may result in response delays, financial costs, and potential information security concerns, which are undesirable.

Fortunately, since last year (2023), significant progress has been made in LLM-related research. On one hand, more compact LLMs, such as Gemma-2B and Falcon-1B, have been released. On the other hand, various compression algorithms, such as 4-bit and even 1-bit quantization, have been developed, effectively reducing the memory usage of LLMs. This increases the likelihood of deploying LLMs on edge devices.

This study explores a novel and cost-effective solution, demonstrating how to efficiently deploy an innovative and affordable MLLM on mobile devices, potentially addressing all the aforementioned issues. Moreover, this solution is applied to educational fields such as mobile learning and digital tours.

## **MATERIALS AND METHODS**

This study uses quantization methods to load a Large Language Model (LLM) as the core brain of the system and employs Retrieval-Augmented Generation to enhance domain-specific knowledge. Additionally, it integrates the You Only Look Once (YOLO) object detection model as a visual decoder and incorporates OpenAI's Whisper for voice input and Linux's built-in eSpeak for voice output. This implementation enables the MLLM to think, see, listen, and speak multimodally, and it supports interaction in Chinese. The system is implemented on Jetson Orin NX.

This system is applied to mobile learning and digital tour domains, with Tainan Confucius Temple chosen as the verification site. Test subjects are divided into an experimental group and a control group. The experimental group uses the device provided by this study for tours, while the control group is led by human tour guides. Both groups undergo pre-tests and post-tests, confirming the learning outcomes.

Considering the current research models focusing on AI use intention, only the socially-oriented AIDUA (Artificially Intelligent Device Use Acceptance) and the task-oriented T-AIA (Task-oriented AI Acceptance) exist. These models are not fully applicable to GAI devices. Therefore, this study proposes an innovative research model based on the Theory of Reasoned Action as an important basis for evaluating the effectiveness of the system, exploring the Generative Artificially Intelligent Use Acceptance (GAIUA) with six dimensions. These dimensions include problem-solving, perceived humanity, response content, anthropomorphic interaction, satisfaction, and use intention.

## **RESULTS AND DISCUSSION**

This study verifies the practical usability of this MLLM framework. It also identifies potential research limitations encountered in real-world applications, providing valuable insights for future researchers. The study believes that on the path towards AGI, it has explored a novel and distinctive model, pointing out a different direction for future research.

The learning outcomes provided by this study's device improved by 18.0% compared to human tour guides. Additionally, the five path hypotheses of GAIUA were validated through the feedback scales from the experimental group, confirming that all hypotheses proposed in this study were substantiated.

## **CONCLUTION**

This study proposes a novel, low-power, embedded AI-based MLLM framework capable of offline execution on Single Board Computers (SBCs). This framework offers an alternative perspective in the current era of soaring training and operational costs for large models, potentially providing a lower-cost MLLM solution for users with limited resources.

The study verifies the application of this GAI system, which supports Chinese voice interaction, in mobile learning and digital tour domains, demonstrating its potential for enhancing learning outcomes and user experience.

Furthermore, this study introduces an innovative research model, Generative Artificially Intelligent Use Acceptance (GAIUA), to explore the intention of using GAI. The study believes that as GAI continues to evolve, this research model can be applied in the future to measure various new types of GAI services.

# 目錄

摘要 .....	i
目錄 .....	vi
表目錄 .....	ix
圖目錄 .....	xi
第一章、緒論 .....	1
1.1 研究背景與動機 .....	1
1.2 研究目的 .....	3
1.3 研究範圍 .....	5
1.4 論文架構 .....	6
第二章、文獻探討 .....	7
2.1 物聯網與通用人工智慧 .....	7
2.1.1 物聯網、雲端與邊緣運算 .....	7
2.1.2 邊緣人工智慧 .....	8
2.1.3 嵌入式人工智慧與微型機器學習 .....	8
2.1.4 邁向通用人工智慧 .....	9
2.2 新時代的學習與導覽 .....	18
2.2.1 行動學習 .....	18
2.2.2 數位導覽 .....	19
2.2.3 生成式人工智慧的應用 .....	20
第三章、系統設計與建置 .....	22

3.1 實驗地點 .....	22
3.1.1 實驗地點介紹 .....	22
3.1.2 實驗資料集 .....	23
3.2 實驗準備 .....	25
3.2.1 模型訓練與資料預處理 .....	25
3.2.2 實驗設備 .....	28
3.3 系統開發 .....	30
3.3.1 系統啟動流程 .....	32
3.3.2 使用者可執行的指令 .....	33
3.4 實驗準備結果 .....	34
3.4.1 物件辨識模型訓練 .....	34
3.4.2 向量資料庫預處理 .....	39
3.4.3 小節整理 .....	39
第四章、使用者學習成效與 GAIUA 模型探討 .....	40
4.1 實驗流程 .....	40
4.1.1 實驗對象 .....	40
4.1.2 實驗講解 .....	41
4.1.3 前測 .....	41
4.1.4 導覽 .....	42
4.1.5 後測 .....	47
4.1.6 量表 .....	48
4.2 研究模型 .....	49
4.2.1 理性行為理論 .....	49



4.2.2 生成式人工智慧的使用意圖研究模型 .....	50
4.2.3 構面 .....	51
4.2.4 路徑假說 .....	54
4.2.5 量表詳細題項 .....	58
第五章、統計分析與研究假說結果 .....	60
5.1 受測者族群分析 .....	60
5.1.1 對照組 .....	60
5.1.2 實驗組 .....	61
5.2 學習成效 .....	62
5.2.1 實驗組 .....	62
5.2.2 對照組 .....	62
5.3 量表 .....	63
5.3.1 構面信度與效度 .....	63
5.3.2 區別效度 .....	64
5.3.3 因素負荷量與共線性 .....	65
5.3.4 路徑係數 .....	69
第六章、研究貢獻與未來展望 .....	76
6.1 研究貢獻 .....	76
6.2 研究限制 .....	77
6.3 未來展望 .....	78
參考文獻 .....	79

## 表目錄

表 2-1：BREEZE-7B 與其它 LLM 的成效比較[31] .....	11
表 2-2：不同參數量大小的 LLM 於不同精度下量化後所需的記憶體空間[75] .....	16
表 3-1：本研究圖像資料集 .....	24
表 3-2：實驗準備-模型訓練與資料預處理-硬體環境.....	25
表 3-3：實驗準備-模型訓練與資料預處理-軟體環境.....	25
表 3-4：物件辨識模型訓練參數.....	27
表 3-5：實驗準備-實驗設備-硬體環境 .....	29
表 3-6：實驗準備-實驗設備-軟體環境 .....	30
表 3-7：實驗準備-實驗設備-使用軟體 .....	30
表 3-8：物件辨識物件中英對照 .....	34
表 3-9：物件辨識模型訓練結果 .....	39
表 3-10：向量資料庫處理時間 .....	39
表 4-1：前測問題清單 .....	41
表 4-2：觸控螢幕顯示日誌輸出範例 .....	44
表 4-3：量表題項分布 .....	48
表 4-4：量表詳細題項清單 .....	59
表 5-1：對照組族群分析 .....	60
表 5-2：實驗組族群分析 .....	61
表 5-3：實驗組學習成效 .....	62
表 5-4：對照組學習成效 .....	62

表 5-5：構面信度與效度 .....	64
表 5-6：區別效度.....	65
表 5-7：因素負荷量與共線性.....	68
表 5-8：直接效果.....	71
表 5-9：間接效果.....	72
表 5-10：「滿意度」的總和效果 .....	74
表 5-11：「使用意圖」的總和效果.....	75

## 圖目錄

圖 1-1：論文架構.....	6
圖 2-1：邊緣運算至 TINYML 的上下關係[59] .....	8
圖 2-2：先訓練出 FM，再由微調導向不同任務的過程[63].....	10
圖 2-3：FM 的微調方式[64] .....	10
圖 2-4：視覺語言模型訓練方式[8] .....	12
圖 2-5：ONE-STAGE 與 TWO-STAGE 物件辨識比較[67] .....	13
圖 2-6：YOLO 演進歷程[66] .....	13
圖 2-7：精度型別格式的比較[74] .....	15
圖 2-8：RAG 流程[80] .....	17
圖 3-1：臺南孔廟空拍圖[101].....	22
圖 3-2：臺南孔廟平面圖 .....	23
圖 3-3：CVAT 標註介面[102] .....	26
圖 3-4：物件辨識模型訓練流程圖 .....	26
圖 3-5：向量資料庫處理流程圖 .....	27
圖 3-6：JETSON ORIN NX 配件總覽[103] .....	28
圖 3-7：JETSON ORIN NX 組裝成品 .....	29
圖 3-8：系統流程.....	31
圖 3-9：物件辨識混淆矩陣 .....	35
圖 3-10：物件辨識損失、精度、召回率和平均精度 .....	36
圖 3-11：物件辨識性能指標曲線 .....	37

圖 3-12：物件辨識模型預測結果 .....	38
圖 4-1：實驗流程.....	40
圖 4-2：導覽路線.....	42
圖 4-3：實驗組導覽詳細流程.....	43
圖 4-4：實驗組實際導覽情形.....	44
圖 4-5：對照組實際導覽情形.....	47
圖 4-6：理性行為理論[104].....	49
圖 4-7：生成式人工智慧的使用意圖研究模型 .....	50
圖 4-8：恐怖谷理論[109].....	52
圖 5-1：路徑係數.....	69
圖 6-1：程式待命時的記憶體占用 .....	77
圖 6-2：與 LLM 互動時的記憶體占用 .....	78

# 第一章、緒論

## 1.1 研究背景與動機

人工智慧（Artificial Intelligence, AI），在各種應用領域取得了前所未有的成就。而在近期成為研究和應用的焦點的生成式人工智慧（Generative Artificial Intelligence, GAI），更是邁向通用人工智慧（Artificial General Intelligence, AGI）的墊腳石[1]。GAI 是 AI 的一個特定領域，旨在模仿人類以不同的形式產生內容，從最初的影像、音訊、文字，甚至 3D 物件。隨著近年的蓬勃發展，各種不同類型的應用璀璨而生，如文字轉影像（Text-to-Image）、文字轉影片（Text-to-Video）、語音轉文字（Speech-to-Text, STT）、聊天機器人（Chatbot）等。而隨著 ChatGPT 代表的大型語言模型（Large Language Model, LLM）問世[2]，GAI 產出的內容（Artificial Intelligence Generated Content, AIGC）開始逐漸被大眾廣泛應用。

為因應 GAI 對於教育的衝擊，聯合國教科文組織（UNESCO）提出，現在是時候讓下一代學生為未來做好準備，AI 將成為他們生活中越來越重要的一部分，並建議將 GAI 作為教育工具納入課程政策中[3]。歐盟執委會（European Commission）也於近年發布了一份指南，其結論是：AI 在教育領域不再是遙遠的未來，它已經在改變教育工作者的工作方式及孩子的學習方式。它將使教育環境更具回應性及成為個人化輔導的主要內容，且其影響力將在未來進一步擴大[4]。也因此人們可以預見未來的 GAI 將更無處不在，而 LLM 將在其中扮演重要的角色[5]，如何將 LLM 開枝散葉應用在不同環境並且擁有多模態的能力想必是未來的重中之重。

近年來，尤其在 GPT-4[6]出現後，人們對擴展 LLM 的多模態能力越來越感興趣。多模態大型語言模型（Multimodal Large Language Model, MLLM）將 LLM 作為大腦，擴展了其處理影像、影片和音訊等額外資訊的能力來執行多模態任務[7]。在這種情況下，LLM 的模組化訓練越來越受到關注[8-11]。而這種訓練通常需要將 LLM 部署額外的視覺輸入投影層（Visual Input Projection）串接視覺編碼器（Vision Encoder），然後將大量影像與文字進行預訓練以實現跨模態對齊[12]，這種方式儘

管有效，但相當昂貴。例如，BLIP2[8]的預訓練，在 1.29 億個影像文字對上消耗超過 100 個 GPU 小時。而 LLaVA-13B[9]則在訓練期間對整個 LLM 進行了全面的參數微調，導致訓練時間和中間儲存開銷顯著增加。更重要的是，目前的解決方案對於 LLM 原本的參數空間將會產生巨大變化，這些微調方案將不可避免地削弱 LLM 原本的自然語言處理能力（Natural Language Processing）[12]。也因此，尋求微調參數以外對齊文字與影像的解決方案，將是一個重要的課題。

此外，由於 GAI 的任務複雜性，此類模型通常規模龐大、運算量高，需要強大的集中式運算設施（即雲端伺服器）來處理使用者的請求。然而，繁重的運算會消耗大量能源。當請求量很大時，僅依賴雲端的運算能力還可能導致高延遲[13]。以 ChatGPT 為例，2023 年 1 月，它每天處理超過 1300 萬個請求[14]。GPT-3 是一個包含 1750 億個參數的大語言模型[15]，需要超過 350GB 的 VRAM 來運行模型，其每月消耗的電量，據估計相當於 17.5 萬名丹麥人每月的用電量[14]。將如此過度集中的計算框架部署在雲端伺服器上，不經濟、不環保、不可持續且成本效率低，將造成迫在眉睫的環境問題[13]。

然而，由於一般的行動設備或電腦的處理能力有限，對 LLM 的指令通常仍是在雲端處理，這是 ChatGPT 和大多數支援 LLM 的聊天應用程式的標準流程。但這種需要使用網路資源的方式可能會帶來回應延遲、財務開銷，甚至可能引發資訊安全的疑慮，這是不被希望的[13]。例如，在網路連線能力有限的地區、網路流量嚴格監控的環境或一些特殊的網路隔離環境當中（如工廠），運算就必須在設備上完成[16]。為了緩解這些挑戰，人們提出了邊緣運算（Edge Computing）的人工智慧物聯網（Artificial Intelligence of Things, AIoT）概念，使 AIGC 能在本地生成。同時，為了使一項服務成為每個公民生活中的一部分，必須接受小尺寸、高能效、離線和精確性的考驗[17]。從幾年前推出的「低功耗影像辨識挑戰賽」（Low-power Image Recognition Challenge）[18]競賽中也可以看出實現這些目標的緊迫性。而這些目標不僅有吸引力，且在物聯網（Internet of Things, IoT）、機器人、自動駕駛、無人機監控等多個領域甚至是必要的[17]。然而，雲端運算與邊緣運算在計算資源上的根本性差異[19]，會導致此目標在實現上的困難。

值得慶幸的是，去年(2023)以來 LLM 相關的研究取得了長足的進步。一方面，參數量更緊湊的 LLM 被發布，如 Gemma-2B[20]和 Falcon-1B[21]。另一方面，各種壓縮演算法，如 4bit[22]甚至 1bit[23]量化也被發布，能有效減少 LLM 的記憶體占用。讓人們能更有機會能將 LLM 部屬在邊緣設備中。

因此，本研究探索出了一套新穎且經濟實惠的解決方案，示範如何將一套創新且成本低廉的 MLLM 有效地部屬在行動設備中，可望解決以上提出的所有問題。並且，本研究嘗試將此解決方案應用在行動學習、數位導覽等教育的領域中。

人工智慧在教育領域的應用雖已有 40 年之久[24]，但有研究認為 GAI 的出現才開始對全球教育產生了更巨大的衝擊[25]。同時，也有研究坦言沒有人真正知道推陳出新的 GAI 將會走向何方[24]。對此，本研究將貢獻一己之力探索 GAI 於教育上的創新應用。本研究相信行動式的 GAI 具有革新學習、導覽與教育方式的巨大潛力，將是邁向 AGI 的重要一步。

## 1.2 研究目的

本研究提出了一套創新且成本低廉的 MLLM 有效的部屬在單板機 (Single-board Computer, SBC) 上，展示了 GAI 在嵌入式系統上本地運行的可能性。這種方法不僅可行，而且在能源和成本效率上具有優勢，且有望成為邁向 AGI 的重要一步。

為了達到此目的，本研究提出的 MLLM 達到了以下八點目標：

- (1) 「可移動性」：本研究使用 Nvidia 發行的 Jetson Orin NX 16GB 開發板作為研究設備，使用 Jetson 系列開發板的優勢在於其支援 CUDA 的 GPU 具有加速深度學習運算的能力。其主要特點包括外形小巧、重量輕和低功耗，特別適合應用於攜帶式環境[26]。
- (2) 「低成本視覺編碼器」：本研究拋棄傳統的跨模態對齊訓練，轉而選擇訓練成本較低的 YOLO(You Look Only Once)作為 MLLM 的視覺編碼器。本研究將此視覺編碼器的功能定在將辨識出的物件傳送給 LLM 使用。而使用 YOLO 的優勢在於，YOLO 是相當成熟的 One-Stage 物件辨識模型，相較於



Two-Stage 的物件辨識模型，能更快產出辨識結果[27]。並且，從事研究 YOLO 的社群蓬勃發展，使其容易得到迭代支持[28]。

- (3) 「低微調成本」：本研究使用了 LLM 的技術－檢索增強生成（Retrieval-Augmented Generation, RAG），透過相似性搜尋（Similarity Search）從外部的向量資料庫（Vector Database）檢索文本塊（Chunk），使 LLM 能輕鬆理解未曾學習過的特定領域知識[29]，而不是使用微調參數這種耗費龐大資源的方式[30]。
- (4) 「中文互動」：本研究使用由 MTK Research 團隊提出的 Breeze-7B 模型[31]，Breeze-7B 的基礎建立在 Mistral-7B 模型[32]上，旨在透過微調改進 LLM 使用繁體中文的能力，並縮小了開源模型和 GPT-3.5 等專有模型於理解繁體中文上的差距。
- (5) 「語音輸入」：本研究使用了 OpenAI 團隊提出的 Whisper[33]來作為 MLLM 的 STT 模型，該模型預設支援了 99 種語言，並且依模型參數量大小分為 Tiny、Base、Small、Medium、Large 等模型級距，甚至有研究微調了 Whisper 使其能識別台語[34]。本研究實測起碼要使用 Base 等級的模型，才能盡量減少幻聽的出現，同時維持良好的低記憶體占用。
- (6) 「語音輸出」：本研究使用 Linux 內建的 eSpeak 來做為 MLLM 的文字轉語音（Text-to-Speech, TTS）模組，進行語音輸出與使用者互動。
- (7) 「低能源消耗」：本研究運行在 SBC 上，相較於 Jetson Orin NX 16GB 開發板的功耗為 10W 至 25W，一般顯示卡以 4060Ti 16GB 為例，功耗約是在 70W 至 165W 之間。本研究僅需使用行動電源供電，而不需專門的電源供應器。
- (8) 「離線環境」：本研究能完全運行在離線環境。

本研究將此系統應用在行動學習、數位導覽的領域中，探索行動式 GAI 設備位於日益成長的教育改革需求中，可展現出的學習成效及使用意圖。

探討使用意圖的重要性在於，有研究指出，儘管與 AI 機器互動似乎是一種不可避免的趨勢，但並不是所有人都準備好接受它[35]。有研究結果表明，當發生緊急

情況或複雜問題時，客戶仍然更喜歡人類勞動力[36]。另有研究發現，消費者不想依賴 AI 演算法來執行通常由人類完成的任務，儘管 AI 演算法在這些工作中表現優於人類[37]。還有研究發現，如果消費者發現 AI 演算法曾出錯，他們對 AI 演算法的依賴感會直接降低[38]。例如，消費者更願意選擇人類醫生而不是 AI 醫生，因為他們認為 AI 醫生更可能忽視個體的獨特性[39]。而在旅遊、飯店業的背景下，遊客抱怨 AI 服務失敗，因為它們無法靈活處理知識範圍外的複雜的問題[40]。因此，許多研究表明，儘管許多服務提供者建議客戶使用 AI 設備，但客戶最終仍可能尋求人工服務。

綜上所述，探討 AI 設備使用意圖的重要性便不言而喻。然而，目前相關的模型，只有社交導向的「AIDUA(Artificially Intelligent Device Use Acceptance)」[41]以及任務導向的「T-AIA(Task-oriented AI Acceptance)」[35]。只是這兩種研究模型都是針對一般的 AI 服務，並無法完全適用於 GAI 設備。也因此，鑒於目前沒有相關研究針對 GAI 設備提出研究模型，本研究便提出一套創新的研究模型，探討「GAI 的使用意圖（Generative Artificially Intelligent Use Acceptance, GAIUA）」，同時做為評估本系統成效的重要依據。

總結而言，本研究主要的貢獻有三個面向：

- (1) 提出了一套創新且低成本的 MLLM 框架並置於 SBC 上。
- (2) 比較了一般導覽員與 GAI 設備帶給使用者的學習成效。
- (3) 提出了一套創新的研究模型 GAIUA，探討使用者對 GAI 設備的使用意圖。

## 1.3 研究範圍

本研究盡可能要求完整嚴謹，但因時間、精力及個人能力有限，因此在實驗範圍與系統表現上將承受部分限制，具體說明如下：

- (1) 實驗主體選擇：臺南孔廟成立於 1665 年，是明鄭時期台灣的最高學府[42]，境內有不少特殊且難得保存至今、充滿歷史的建築物。本研究選擇了其中四

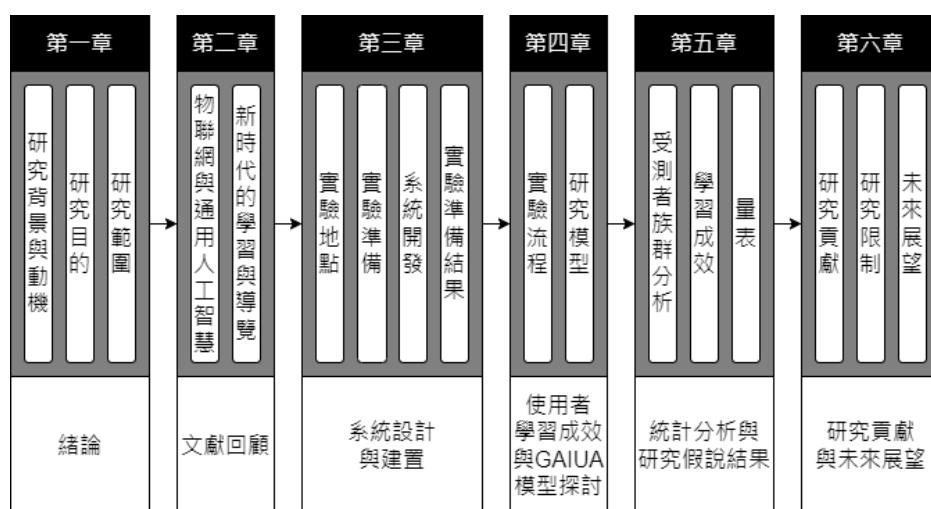
樣建築物：全臺首學、明倫堂、文昌閣、泮宮石坊，作為驗證本研究行動學習、數位導覽的標的物。

- (2) 系統穩定性：由於系統的記憶體占用遊走在滿載邊緣，故時常於使用時發生記憶體不足（Out Of Memory, OOM）的情況而觸發 Killed，使用者要繼續使用便須等待重新啟動程式，載入 MLLM 約需等待 43 秒。
- (3) 系統反應速度：由於 SBC 的處理速度，MLLM 於下指令後，須等待約 20 秒才會開始出現文字。
- (4) 系統幻聽：由於現場較吵雜的環境，藍芽耳機麥克風會偵測到許多異音，故有時會請使用者重複下達指令，直至偵測正確。

然而，本研究的目的旨在探索此創新 MLLM 框架的可行性，同時探討學習成效以及持續使用意圖，故系統的品質與整合性問題並非本研究的第一考量，期待未來技術的進步能夠更完善使用者的體驗。

## 1.4 論文架構

本研究之論文架構，可參考圖一-1。



圖一-1：論文架構

## 第二章、文獻探討

### 2.1 物聯網與通用人工智慧

本節將從物聯網的角度延伸至未來趨勢的通用人工智慧，並承先啟後地探討。

#### 2.1.1 物聯網、雲端與邊緣運算

近年來，有越來越多的設備連上網路，也因此產生了大量的數據，這種模式被稱為物聯網（Internet of Things, IoT）。IoT 的發展帶動了全球大量硬體設備/感測器的生產和應用，如智慧家庭、智慧電網、智慧交通、智慧製造等[43]。這些硬體設備/感測器通常具有感知周圍物理環境資訊並轉化為數據的能力，而這些數據通常都傳輸到雲端運算（Cloud Computing）平台進行運算與儲存[44]。

雲端運算的概念最早是由 John McCarthy 於 1955 年提出[45]。然而，這個概念直到 2006 年的搜尋引擎大會（SES San Jose 2006）上，由 Google CEO Eric Schmidt 再次提出[46]才逐漸普及，其代表著集中運算服務，整合了分散式運算、負載平衡、平行運算、網路儲存、虛擬化等技術。然而，隨著 IoT 在人們的生活中普及和發展，海量的數據隨之產生。Cai et al. (2016) 指出，大多數感測器獲得的數據只有三分之一是有用的[47]，將這些無價值的資料傳輸至雲端勢必會造成資源及頻寬巨大的浪費。Wan et al. (2022) 的研究闡明，目前雲端運算無法滿足部分系統需快速響應的需求，並且持續增長的 IoT 給雲端頻寬帶來了巨大的壓力，而這也可能導致資訊安全和隱私等問題的出現[48]。

為了解決上述挑戰，名為邊緣運算（Edge Computing）的方式，於近年引起了廣泛關注，然而事實上，邊緣運算的概念早於 1998 年便由 Akamai 所提出[49]。其核心理念就是將原本由雲端處理的資料運算和儲存作業，下放給靠近資料源的終端設備。這樣的轉變助於減少資料傳輸及回應的時間成本，也減輕了頻寬的壓力，同時還實現了去中心化[50]。

### 2.1.2 邊緣人工智慧

由於邊緣設備的運算能力不斷增強，近年來運算任務由雲端中心轉移到邊緣設備已逐漸成為一個新的趨勢[51]。而在網路邊緣執行 AI 模型的推理，具有無須網路傳輸、低延遲和增強隱私的優點，並且可以進行及時決策，這對於醫療保健、自動駕駛汽車、智慧城市、工業物聯網等 AI 應用是至關重要的[52]。此外，IEEE 更將「邊緣人工智慧（Edge AI）」定為 2020 年的 12 大技術趨勢之一[53]。而隨著 Edge AI 的發展，對教育領域的影響也逐漸擴大，Wang et al. (2021) 的研究表明，人工智慧可以提供學生自適應的學習資源，並根據學習者的表現提供評估和建議，明顯地提高了學生的學習成效[54]。

### 2.1.3 嵌入式人工智慧與微型機器學習

事實上，邊緣層的設備類型可謂是非常多樣化，從具有較高延遲但運算資源豐沛的硬體到超低延遲但運算資源十分有限的硬體都囊括其中。因此，以定義來說，AI 應用在硬體資源較少的設備上，如單晶片（System-on-Chip, SoC）上運行的 AI 應用，被稱為嵌入式 AI（Embedded AI）[55]，而當硬體資源受到更多限制的狀況下（可以運行在功耗 1mW 下的設備），如微控制器單元（Microcontroller Unit, MCU）上運行 AI 應用，則稱為微型機器學習（Tiny Machine Learning, TinyML）[56, 57]。由上到下的關係可以參考圖 二-1。使用邊緣設備不僅有助於減少雲端的流量，同時有機會成為在大眾日常生活中傳播 AI 應用的先驅[58]。

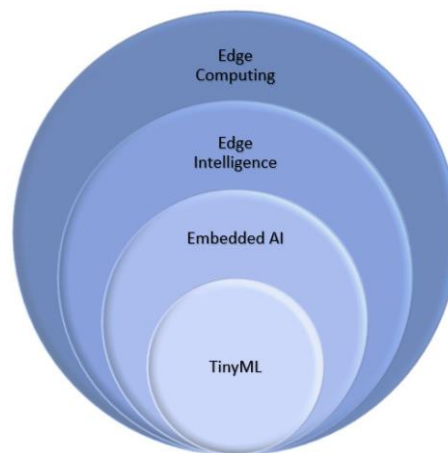


圖 二-1：邊緣運算至 TinyML 的上下關係[59]

然而，在本研究使用的硬體是 Nvidia 的 Jetson Orin NX，運行時的功耗大幅地超越了 TinyML 的定義，因此停留在嵌入式 AI 的範疇中。選擇此設備的原因在於其 16GB 的隨機存取記憶體（Random-access Memory, RAM），可以支援大部分的 AI 應用。例如：物件辨識（Object Detection）、語音轉文字（Speech-to-Text, STT）、大型語言模型（Large Language Model, LLM）等。並且，Jetson 系列開發板為市面上最適用於執行 AI 應用的單板機（Single-board Computer, SBC）之一，其具有 CPU-GPU 異質架構[60]，並且支援 CUDA 可以加速計算複雜的 AI 任務。同時，它小巧的外形、輕量化的設計和低功耗，使其非常適合應對於重量及功耗受限的移動式場景[17]。

## 2.1.4 邁向通用人工智慧

AI 在過去十年呈高速發展的趨勢，其成功主要可歸因為三個原因[61]：

- (1) 可用的多模態訓練資料不斷增加，如近年 IoT 設備/感測器大量的增長。
- (2) AI 基礎設施的增強，如更先進的晶片製程等。
- (3) 生成式人工智慧（Generative Artificial Intelligence, GAI）等 AI 模型的進步 [62]。

近年來，通用人工智慧（Artificial General Intelligence, AGI）已成為產官學界越來越流行的話題，並且在不同領域逐漸發酵，如農業、生物學、醫療保健、地理等皆開始進行 AGI 的應用研究。這種趨勢在 OpenAI 推出 ChatGPT 並大獲成功後尤為明顯[62]。

### 2.1.4.1 大型語言模型

LLM 的成功，代表著人們邁向 AGI 又更近了一步。其起源於 Google 於 2017 年提出的 Transformer 架構[25]，這種創新架構基於並行自注意力機制的神經網路而構想出來，擺脫過去使用的 RNN 或 CNN 等。此後，越來越多的研究人員不再訓練特定任務的 AI 模型，轉而開始尋求方法，使用大規模的單一模態資料預訓練大型模型。這種模型後來又被稱為基礎模型（Foundation Model, FM）[63]。而 LLM 則是 FM 經過微調的過程後所產出，如 ChatGPT 或 Breeze-7B[31]等，其關係可以參考圖 二-2。

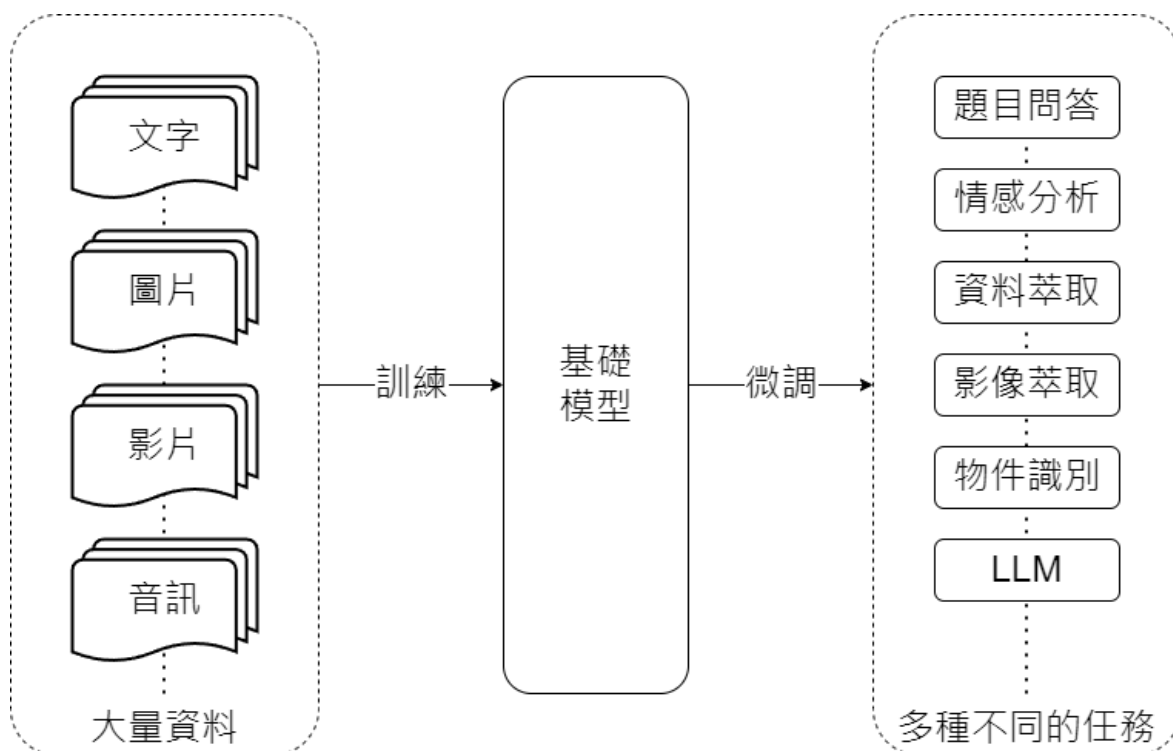


圖 二-2：先訓練出 FM，再由微調導向不同任務的過程[63]

微調過程，如 InstructGPT 的強化學習方式[64]，可以參考圖 二-3，其中「代理人/模型」為前述的 FM。

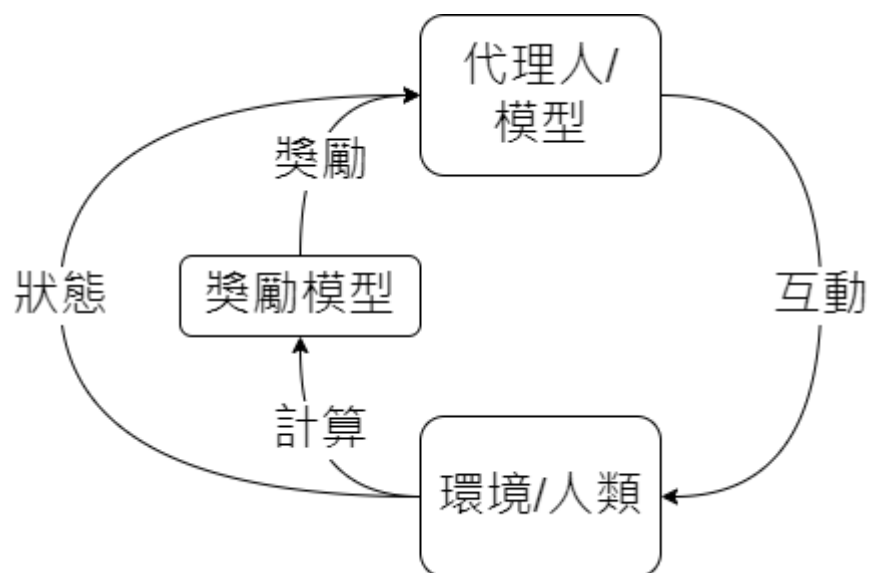


圖 二-3：FM 的微調方式[64]

然而，由於大模型任務的複雜性，其繁重的運算會消耗大量能源。值得慶幸的是，去年（2023）以來 LLM 相關的研究取得了長足的進步，參數量更緊湊的 LLM

被發布。例如，Gemma-2B[20]、Falcon-1B[21]和 Breeze-7B[31]等。讓系統能以較不耗費資源的方式運行。

其中，Breeze-7B[31]是為繁體中文任務量身打造的 LLM，由 MediaTek Research 團隊提出，建立在 Mistral-7B 的基礎上。使用了 650 GB 的繁體中文資料及 7,000 個 H100 小時才將 Mistral-7B 微調為 Breeze-7B-Base-v1\_0。其與其它支援中文的語言模型比較，可參考表 二-1，其評分標準為將模型的回應交由 GPT-4 打分數[6]。本研究因在臺灣驗證實驗成效，故使用此模型作為本研究的 LLM。雖然 Qwen1.5-7B-Chat 的平均分數較高，然而其並非專門支援繁體中文的 LLM。

表 二-1：Breeze-7B 與其它 LLM 的成效比較[31]

LLM 模型	STEM	資訊 萃取	推理	數學	編碼	角色 扮演	寫作	人文 學科	平均
GPT-3.5-Turbo	7.8	6.1	5.1	6.4	6.2	8.7	7.4	9.3	7.1
Breeze-7B-Instruct-v1_0	7.8	5.2	4.2	4.2	4.1	7.6	5.9	9.1	6
Qwen1.5-7B-Chat	9	5.6	4.7	2.8	3.7	8	8	9.4	6.4
Yi-6B-Chat	7.3	2.7	3.1	3.3	2.3	7.2	5.2	8.8	5
Taiwan-LLM-13B-v2.0-chat	6.1	3.4	4.1	2.3	3.1	7.4	6.6	6.8	5
Taiwan-LLM-7B-v2.1-chat	5.2	2.6	2.3	1.2	3.4	6.6	5.7	6.8	4.2

#### 2.1.4.2 多模態大型語言模型

為了邁向 AGI，Dou et al. (2023) 指出 AGI 需有以下特性：感知環境、使用工具、記憶歸納、學習事物、邏輯推理、策略規劃、設立目標、感知情感、多模態互動、計算能力、創造性[61]。其中多模態互動的特性，與目前研究多模態大語言模型（Multimodal Large Language Model, MLLM）的基礎不謀而合，它要能同時處理文字、圖像、視訊、音訊等多種資料模態。

而目前較多的研究導向，為使用文字與圖像兩種模態的 MLLM。如 BLIP-2[8]、LLaVA-13B[9]、MiniGPT-4[11]等。本研究大致歸納出這類型 MLLM 的訓練模式[65]，如圖 二-4。



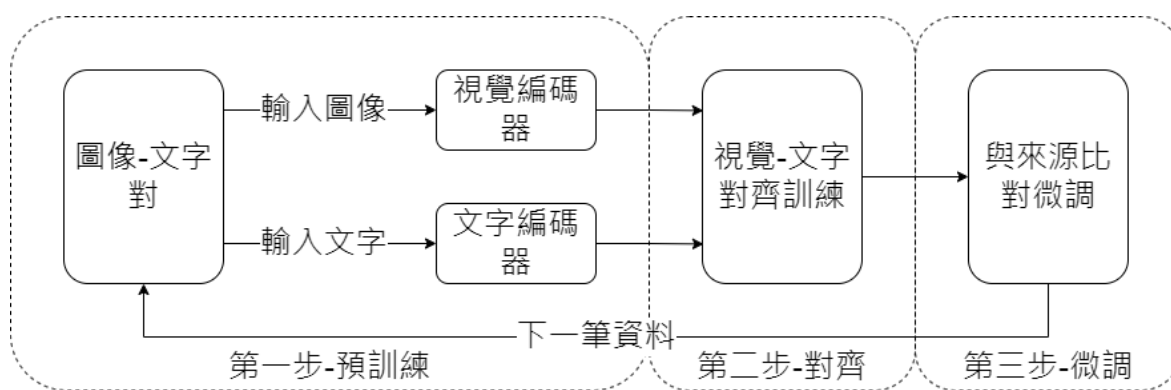


圖 二-4：視覺語言模型訓練方式[8]

然而，目前這種訓練方式儘管有效，卻相當昂貴。例如，BLIP2[8]的預訓練，在 1.29 億個圖像文字對上，使用十六張 A100 訓練超過 100 個小時。MiniGPT-4[11]的預訓練，在 500 萬個圖像文字對上，使用四張 A100 訓練 10 個小時。而 LLaVA-13B[9]則在訓練期間對整個 LLM 進行了全面的參數微調，導致訓練時間和中間儲存開銷顯著增加。更重要的是，目前的解決方案對於 LLM 原本的參數空間將會產生劇烈變化，將不可避免地削弱 LLM 原本的自然語言處理能力（Natural Language Processing, NLP）[12]。也因此，尋求微調參數外對齊文字與圖像的解決方案，將是一個重要的課題。同時，這種訓練方式所需的龐大算力資源，也並非一般民眾所負擔的起。

本研究嘗試使用外掛視覺模組的方式達到類似 MLLM 的效果，所需的訓練成本便可依模組而各別拆開。並且，使用 YOLO（You Only Look Once）作為系統的視覺編碼器，並將此視覺編碼器的功能定在將辨識出的物件傳送給 LLM 使用。由於目前較少關於物件辨識與 LLM 整合的研究，本研究希望能在在此做出貢獻。

即時物件偵測（Real-time Object Detection）演算法發展到今日，已成為眾多應用中的關鍵元件，涵蓋在農業、交通、醫療等不同領域[66]。在不同的物件偵測方法中，YOLO 框架因其辨識、訓練速度與其準確性的卓越平衡而脫穎而出，其使用的 One-Stage 辨識方法，相較於 Two-Stage 的物件辨識方法，能夠更快速地識別圖像中的目標[67]。One-Stage 與 Two-Stage 物件辨識流程上的區別可參考圖 二-5。

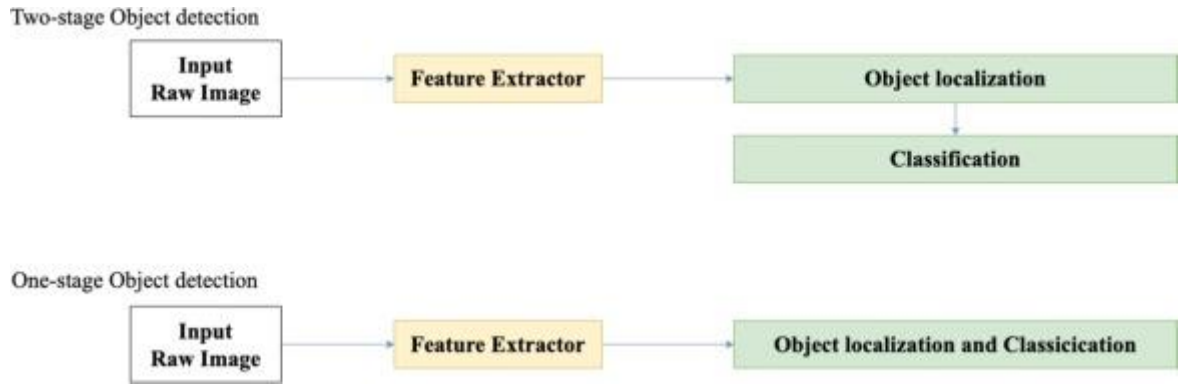


圖 二-5：One-Stage 與 Two-Stage 物件辨識比較[67]

YOLO 自誕生以來，已經經歷了多次迭代，每次迭代都建立在先前版本的優勢上，從最初的 YOLO v1 發展到近年的 YOLO v8，如圖 二-6。本研究使用的便是於 2023 年提出的 YOLO v8。然而至截稿日前，YOLO 公布了 v10 版本[68]。

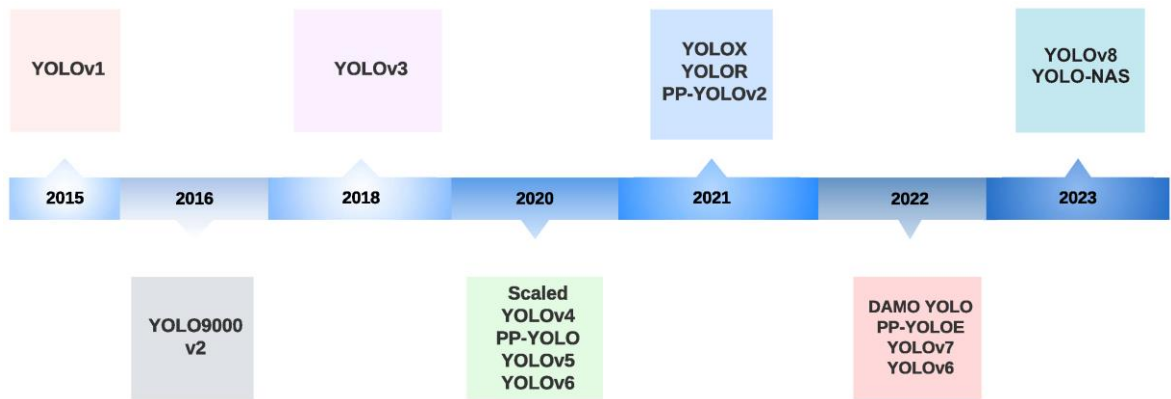


圖 二-6：YOLO 演進歷程[66]

#### 2.1.4.3 通用人工智慧

AGI 意味著更先進和能適應多場景的 AI 系統[61]，這樣的系統應擁有類人的認知能力，能無縫切換使用場景，還要能動態處理並整合來自不同模態的資料，並將其置於上下文中，始能了解事件的前因後果，再透由設備與人類進行互動。更精確地說，其必須要能從互連的感測器、設備和系統中吸收大量異質資料，從中分析並提取相關價值後，始能理解並做出相應的決策與回饋。這將是更強大、更靈活的 AI 解決方案[61]。

Dou et al. (2023) 指出，AGI 的關鍵功能應包括：遷移學習、多模態理解、推理能力、自我監督學習、互動性、小樣本學習等[61]。

其中多模態理解與互動性，是本研究想重點強調的部分。互動，尤其語音互動的重要性，在於口語是人類最自然的溝通方式。Scott et al. (2019) 指出，如果語音助理（Voice Assistant, VA）能夠做到與電腦或智慧型手機等其他設備相同的事情，消費者更傾向於與 VA 互動[69]。並且，口語互動同時也是一種壓倒性的社交行為[70]。消費者在與 VA 交談時可以形成一種社會臨場感（Social Presence），而這種社會臨場感便是指使用科技設備時，能接收到的溫暖、社交性和親密感[71]。引發社會臨場感的同時，可以改善消費者對新技術的體驗並引發積極的行為產生[72]。VA 擁有兩個顯著的類人特徵（說和聽）可以導出這種社會臨場感。

因此，本研究使用了語音互動的方式來驅動本系統。在聽的層面，本研究使用 OpenAI 的 Whisper 作為系統的語音理解模型[33]。而在說的層面，本研究使用 Linux 內建的 ESpeak 作為系統的語音生成模塊。這兩樣工具，將使使用者能直接使用語音與系統進行互動。

Dou et al. (2023) 指出 IoT 設備注定是未來 AGI 應用的主要載體[61]，然而 AGI 和 IoT 的整合仍面臨著許多挑戰。其中最關鍵的一點在於，運行 AGI 需要龐大的算力，然而 IoT 設備通常運算能力有限、記憶體有限、能源也有限（若設計為電池供電）。這在面對 AGI 時，正是一個先天條件的挑戰。也因此除了在邊緣執行 AI 應用外，將計算卸載到雲端處理也是一種辦法。然而，正如 2.1.1 小節所述，使用雲端運算也有其相應的挑戰要面對（如：即時響應、資訊安全、資訊隱私等）。對此本研究相信，能在邊緣運算、離線使用的 AGI 可以實現更好的通用性，但要將 AGI 整合到 IoT 設備中，並且適應各種不同應用的場景，還需要仰賴技術的進步。

值得慶幸的是，這幾年發展出了更小的 AI 模型、壓縮模型的量化技術（Quantization）和能適應多場景的檢索增強生成（Retrieval-Augmented Generation, RAG），可望克服以上的難題，同時這些技術也是本研究所採用的。

#### 2.1.4.4 量化

在 GAI 的時代，模型的參數量於近年來不斷地擴大，其需要龐大的運算資源，使用的能源也很可觀。以 ChatGPT 為例，其包含了 1750 億個參數量[15]，部署這種量級的模型進行推理，需要超過 350GB 的記憶體來運行模型，也就是需要至少五個

80 GB 的 A100 GPU 才能順利運行，其每月消耗的電量，據估計相當於 17.5 萬名丹麥人的每月用電量[14]。最近的研究還估計了運行大型 AI 模型可能產生了驚人的碳排放[73]。這對於近年來社會尋求的永續發展而言，已成為一個負面的問題。此外，這種量級的模型，在資源受限的 IoT 設備或行動裝置上運行，顯然是更不可行的。這對於實現 AGI 的願景來說，無疑是一個沉重的挑戰。也因此近年來發展出了量化（Quantization）技術，試圖解決以上問題。

量化，是指透過降低數值精度來壓縮模型，進而減少推理時所需的算力和記憶體空間。一般來說，LLM 通常使用 32 位元浮點值（float32, FP32）型別來表示權重（Weight）、激勵函數（Activation function）與梯度（Gradient）等。然而，這種高精度型別需要大量的記憶體空間和運算資源。因此透過量化，將這些值轉換為較低精度的型別。例如，16 位元浮點數（float16, FP16）、8 位元整數（Integer8, INT8）、4 位元整數（Integer4, INT4）[22]甚至 1 位元（1 bit）[23]等。從而減少對記憶體和運算能力的需求，便是一種實際的作法。精度型別格式可參考圖 二-7。

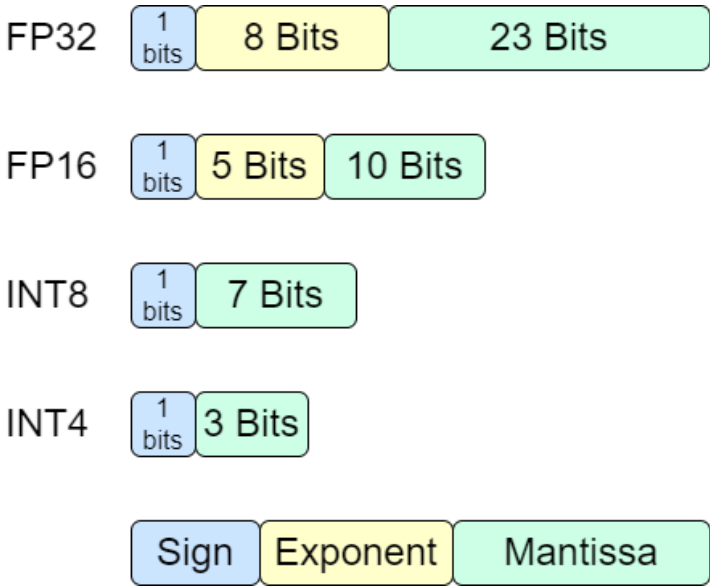


圖 二-7：精度型別格式的比較[74]

量化主要可分為以下兩種方式進行：

- (1) 量化感知訓練（Quantization Aware Training, QAT）：在 QAT 中，量化被整合到模型訓練的過程中。這種方法使 AI 模型能夠在訓練期間適應較低精度的型別，從而增強其於推理時，處理量化造成的精度損失的能力。

(2) 訓練後量化 (Post Training Quantization, PTQ)：在 PTQ 中，量化為每次使用模型前才進行。這種方式的優點在於其能簡易實現模型壓縮，以及無須耗費額外的訓練成本。然而，在量化過程中，PTQ 可能會帶來一定程度的精度損失。經由 PTQ 載入的 LLM，其減少的記憶體空間用量，可以參考表 二-2，此為實際運行在 A100 GPU 上的 LLM 量化紀錄[75]。

表 二-2：不同參數量大小的 LLM 於不同精度下量化後所需的記憶體空間[75]

LLM 模型	NormalFloat4 精度	Ineteger8 精度	Float16 精度
Falcon-7B	5.72	8.71	14.50
Falcon-40B	26.40	44.52	80.85
Llama2-7B	4.58	7.82	13.53
Llama2-13B	8.83	14.20	26.23
Llama2-70B	40.23	70.44	-

因此，近年來有不少改善 PTQ 精度損失與壓縮效率的研究相繼被提出。例如，AWQ (Activation-aware Weight Quantization) [76]、GPTQ (Generative Pretrained Transformer Quantization) [77]和 BNB (Bits-and-bytes) [22]等。

其中，本研究採用的是 BNB 方法，其能利用了一種改良自 INT4 的 4 位元正規浮點數 (NormalFloat4, NF4) 型別，其理論上最適合表示常態分佈。此外，BNB 還能在顯著減少記憶體佔用 (以 4 位元精度儲存量化權重) 的同時，保持了接近全精度的效能 (以 16 位元或 32 位元的精度執行計算)，確保計算穩定性和壓縮效率[22]。

#### 2.1.4.5 檢索增強生成

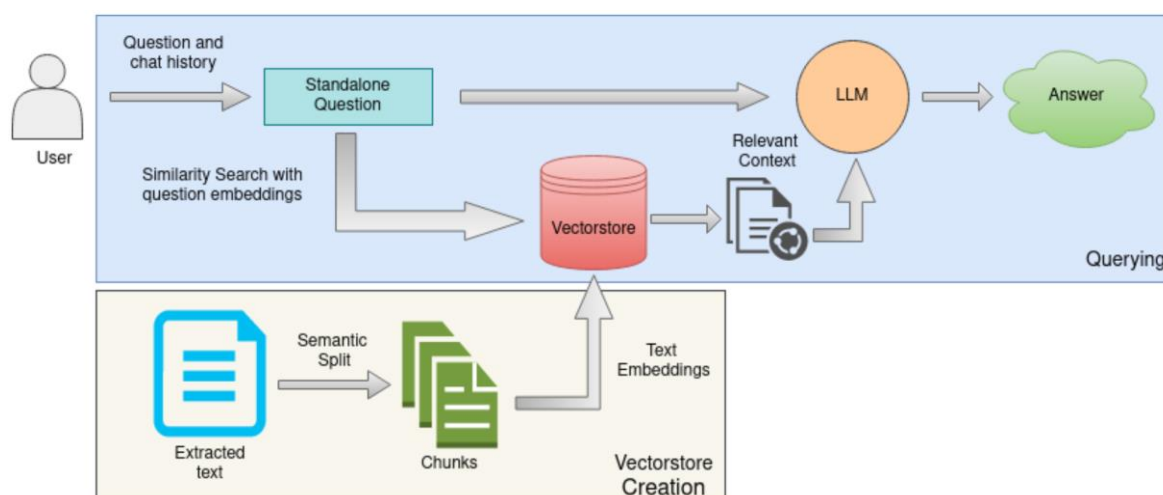
在已訓練好的 LLM 上，要使其能理解最新的資訊、未曾學習過的知識、企業內部的專有領域知識[78]，從而提升回應能力，減少當處理超出其預訓練資料的查詢而出現的幻覺 (Hallucination) [79]。目前有三種方式，可以使已訓練好的 LLM 學習新知識：

- (1) 微調 (Fine-tune)
- (2) 低秩適應 (Low-Rank Adaptation, LoRA)

(3) 檢索增強生成 (Retrieval-Augmented Generation, RAG)。

其中前兩種都需要另外進行訓練，而 RAG 則無需，是成本相對較低的方式[30]。

RAG 的使用方式為：先從資料來源（如：HTML、PDF、CSV 等）中取出文字（Extracted Text），接著透過句子轉換器（Sentence Transformer）進行語意分割（Semantic Split）切為塊（Chunk），再將處理完成的 Chunk 轉換為嵌入（Embedding）將其儲存至向量資料庫（Vector Database）中。以便在 LLM 查詢時，能透過查詢詞彙轉換的 Embedding 在向量資料庫進行相似性搜尋（Similarity Search）對應的值（Relevant Context）提供給 LLM 參考。RAG 執行的詳細流程方式可以參考圖二-8。



圖二-8：RAG 流程[80]

藉由這種方式引用外部知識，RAG 能有效減少了產生與事實不符的內容。然而，使用這種方式的 LLM，其回答的品質也取決於資料進行切塊時的分法，以及相似性搜索的機制限制[81]。

## 2.2 新時代的學習與導覽

### 2.2.1 行動學習

幾十年來，科技進步為教育環境注入了新的活水，教育工作者努力利用新技術創造新的教學方式，使創新的學習方法不斷湧現[82]。

新媒體聯盟（New Media Consortium）於 2019 年的高等教育小組報告指出[83]，有六項技術有潛力對教育產生實質的影響，特別是在教學、學習方法、教師工作組織和設計等方面。這六項技術包括：行動學習（Mobile Learning）、分析技術（Analytics Technologies）、混合實境（Mixed Reality）、AI、區塊鏈（Blockchain）和虛擬助理（Virtual Assistants）。

其中，行動學習目前已成為一種流行且有效的教育方法，其定義於近年來有越來越精確的趨勢。Keegan (2005) 將其狹隘的定義為使用小型電腦、PDA 或手機進行的學習[84]。Crompton (2013) 將其定義為透過社交環境和使用行動裝置（如：手機、平板電腦）的互動在多種環境中學習，在各種環境下進行學習，涉及了不同的個體和內容[85]。近年，Goksu (2021) 則將其定義為「一種學習形式，使個人能透過自己或協作進行學習，使用攜帶式設備透過數位互動來獲取知識」[86]。

Bhati and Song (2019) 的研究結果表明，使用行動設備對正規教育環境中的學習，產生了積極影響。如：其發現可以使用行動設備鼓勵學生們進行協作並提供學生們自我評估的機會[87]，還可以增強學生學習技能的成效[88]。這些發現表明了行動設備在融入教育時，能對學習體驗產生正面影響。

Timms (2016) 指出，AI 在行動學習上的應用應探索不同的可能性，而非僅只透過手機、平板、電腦作為媒介[89]。如 AI 與機器人和 IoT 設備的結合可以提供更多社交互動可能，吸引並保持學生的注意力，還可以提供個人化的指導[89]。

最後，由於行動學習概念的產生，現今的教育不會僅侷限在傳統的學校裡，使用行動設備在非正式場所（學校之外）學習也已成了一個趨勢[90]。而這種自我導向的學習形式，可以發生在每個人人生中的任何時刻，因此「終身學習」、「全方

位學習」的概念，在行動設備蓬勃發展的年代是非常重要的[90]。例如，於自然景點、博物館或歷史名勝（如孔廟）中進行數位導覽（Digital Tour Guide）獲得知識，便也是一種行動學習模式。

### 2.2.2 數位導覽

古往今來，人們參觀許多蘊含知識量的地點，有導遊的陪伴是至關重要的，因為其能解釋各個地方的自然、人文或藏品的重要性，受其導覽可能可以帶來難忘的體驗[91]。然而，導遊本身亦可能會因為重複性的任務而感到疲倦、心煩意亂從而無聊分心。從而導致每次導覽的結果大相逕庭，甚至於他們對某些地方的解釋還可能存在偏頗的個人意見[92]。若發生這樣的事，其服務品質便可能在一定程度上破壞遊客的導覽體驗。

因此從上個世紀以來，伴隨著新科技的發展，迭代出現了許多新的方式能進行數位導覽，從而使遊客能實現自主導遊（Autonomous Guided Tour），而無須依賴人力。例如，在現場掃描 QRCode 獲取相關資訊、使用 XR 設備於現場或遠端導覽景點，甚至在部分場域還有機器人能協助進行導覽。

Umam et al. (2022) 指出，新科技的應用，特別是基於機器人技術的系統，可以提高參觀者捕捉資訊的能力。這樣的機器人稱為導遊機器人。其可以持續提供遊覽服務，直到電池耗盡[93]。

這類型的導遊機器人，不只擁有導遊的能力，還可以識別人臉、以及語音功能與遊客互動塑造社會臨場感[94]。然而，目前的導遊機器人也碰到了一些窘境，值得本研究借鑑。除了完全不依靠人類協助獨立運行的可靠性令人質疑外，另一個問題於不同的機器人研究中，被不約而同地提出，那就是網路問題，其具體狀況如下[94-96]：

- (1) 景點內所有位置的訊號強度不一。
- (2) 機器人在多個節點間漫遊，有時會失去連線。
- (3) 網路不安全。



此問題無論是使用 XR 設備或是智慧型手機這樣的行動設備，只要運算是在雲端，都會遇到一樣的狀況。因此，如本研究前述，使設備能運行在離線環境、邊緣運算的重要性，便不言而喻。尤其目前的 GAI 服務（如 ChatGPT）皆以雲端運算為主。然而，這也給了本研究使用本地運行的嵌入式 AI 作為載體開發的 GAI 服務一個機會。

### 2.2.3 生成式人工智慧的應用

Wong et al. (2023) 指出，ChatGPT 可以訪問龐大的資源網路，獲取更新的資訊，並提供在準確性和資訊深度上超越任何人類導遊的全面知識[97]。它可以為遊客提供世界各地自然、人文或藏品的準確介紹，無論時間、地點或語言。這種自主導遊模式代表了依賴人工的傳統旅遊模式將發生顛覆性的改變。GAI 能使遊客能夠掌控自己的旅行體驗。這種個人化的導遊模式不僅可以提高旅行品質，還可以幫助遊客創造旅行回憶[97]。正如另一份研究所指出的，在 GAI 的支持下建立這種自我主權身分（Self-Sovereign Identity, SSI）可以大大提高使用者的整體價值和滿意度[98]。

Salinas-Navarro et al. (2024) 指出，在教育中使用 GAI 可以建立一個帶有支持性並且具吸引力的學習環境，以補充傳統的教學方法[99]。

Santos (2023) 將 GAI 應用在 STEM 教學上，研究結果表明其優勢包括：互動性、客製化、多學科知識、批判性思考、解決問題、創造力和協作能力等[100]。除了能促進動態對話外，還能提供豐富的跨學科知識與個人化內容。將其納入 STEM 教育可以塑造體驗式學習（Experiential Learning）的環境，使學生可以在引人入勝的學習環境中提升解決問題的能力、養成批判性思維和增加創造力，為他們在 STEM 領域取得成功做好準備[100]。

因此，本研究也將分別比較人類導覽員與 GAI 帶給學生的學習成效。這種比較的目的在於評估 GAI 在教育環境中的應用潛力，特別是在提供即時回饋、個人化學習體驗以及知識整合方面的優勢。隨著科技的進步和 GAI 應用的普及，理解這種技術在教育中的實際效用變得尤為重要。透過比較，我們能夠更全面地了解這兩種方式在知識傳遞、學生參與度和學習成果上的差異。這不僅有助於探索 GAI 在教育領

域的最佳應用實踐，也能為教育政策制定者和實踐者提供實證支持，從而推動教育模式的創新和進步。

## 第三章、系統設計與建置

### 3.1 實驗地點

臺南孔廟（Tainan Confucius Temple）建於 1665 年，是臺灣最早的孔廟，也是台灣歷史最悠久的儒家教育機構[42]。孔廟位於臺南市中心，原為明鄭時期鄭成功的兒子鄭經所建，旨在推廣儒家學說和教育。廟內設有全台首座官辦學校「明倫堂」，象徵著台灣教育的起源。孔廟建築保存良好，展示了中國傳統建築風格，並定期舉辦祭孔大典，吸引眾多遊客參觀。它不僅是台灣文化和教育的重要象徵，也是研究台灣歷史和儒家思想的重要場所。

#### 3.1.1 實驗地點介紹

臺南孔廟位於臺南市中西區南門路 2 號，圖 三-1 為臺南孔廟的空拍圖。



圖 三-1：臺南孔廟空拍圖[101]

圖 三-2 為臺南孔廟的建築平面圖，黑色標線為建築群，綠色虛線為步道，而本次實驗選擇黑底紅框內的四座臺南孔廟的建築物進行建模。分別是，全臺首學、明倫堂、文昌閣、泮宮石坊。

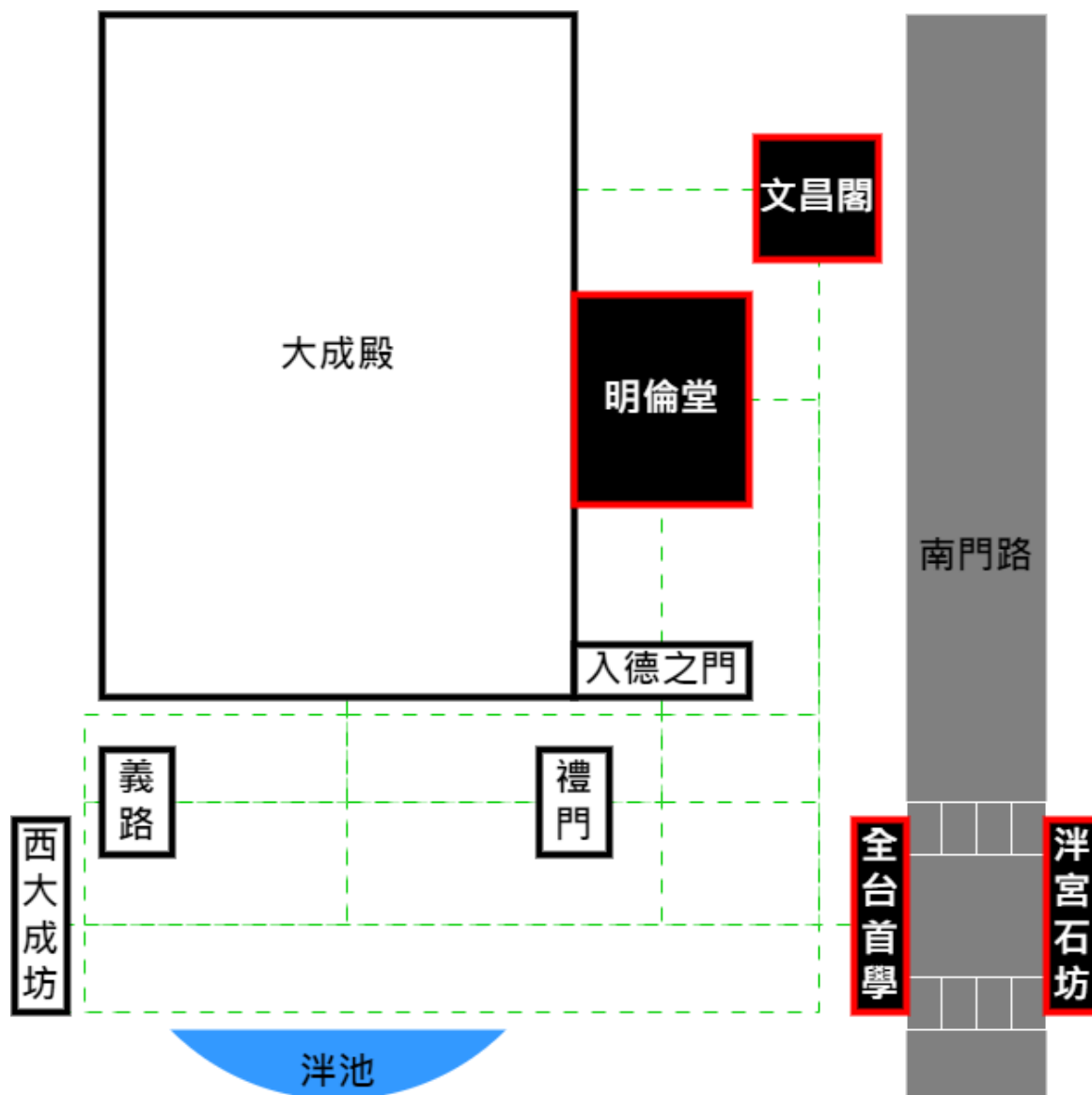


圖 三-2：臺南孔廟平面圖




### 3.1.2 實驗資料集

由於坊間並無專為臺南孔廟建築群設立的圖像及文本資料集，本研究使用的資料皆取材自 Google 圖片、Google 搜尋。

### 3.1.2.1 圖像

本研究進行物件辨識訓練的來源圖像皆取材於 Google 圖片，並依照四種建築物各整理出了數量不同的資料集，圖像總共有 243 張，請參考表 三-1。

表 三-1：本研究圖像資料集

圖像範例		
	全臺首學 (First Academy of Taiwan)	明倫堂 (Ming-Lun Hall)
圖像數量	80	51
圖像範例		
	文昌閣 (The Pavilion of the God of Literature)	泮宮石坊 (The Paifang)
圖像數量	43	69

### 3.1.2.2 文本

本研究進行 RAG 時，向量資料庫的文本來自 20 個介紹孔廟歷史典故、建築特色的網頁，皆皆取材自 Google 搜尋。

### 3.2 實驗準備

本節主要探討本研究實驗所需的物件辨識模型訓練、RAG 向量資料庫中的文字資料預處理以及預備實驗設備。

#### 3.2.1 模型訓練與資料預處理

##### 3.2.1.1 硬體環境

表 三-2：實驗準備-模型訓練與資料預處理-硬體環境

硬體類型	硬體名稱
CPU	AMD Ryzen 5 5600 6-Core, 12-Thread
GPU	Nvidia GeForce RTX 4070 Ti SUPER 16GB
CUDA Cores (GPU)	8448 Cores
Memory	32 GB

##### 3.2.1.2 軟體環境

表 三-3：實驗準備-模型訓練與資料預處理-軟體環境

軟體類型	軟體名稱
OS	Windows 10 Professional
CUDA Version	12.1
Python Version	3.10.11
PyTorch Version	2.0.1
LangChain Version	0.1.8
PyCharm Version	2023.3.4

##### 3.2.1.3 物件辨識模型訓練

本研究使用 Intel 的 CVAT(Computer Vision Annotation Tool)[102]對 243 張臺南孔廟圖像進行標註。標註介面如圖 三-3。



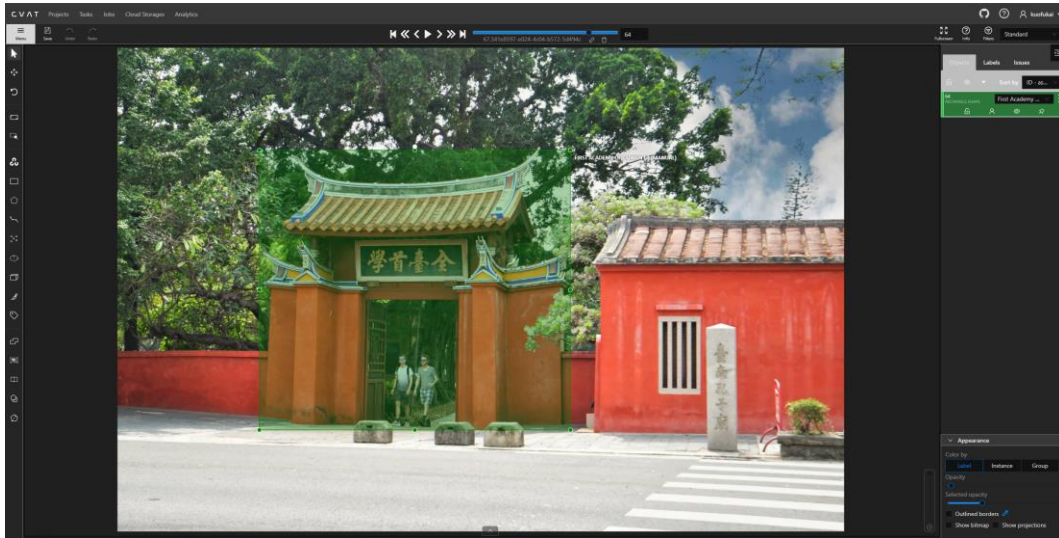


圖 三-3：CVAT 標註介面[102]

標註完畢後，將資料集切分為訓練、測試、驗證，並依序把比例設置為 8:1:1(使用 Autosplit)，再來將 Epoch 設置為 500 並使用 ultralytics/yolov8n.pt 為基底進行訓練。訓練流程可參考圖 三-4。完整訓練參數可參考表 三-4。

這裡使用 yolov8n.pt 的原因為，這是主要負責 YOLOv8 開發的官方 ultralytics 提供的 YOLOv8 最小參數量模型，特別適合應用於資源受限的移動式設備上。

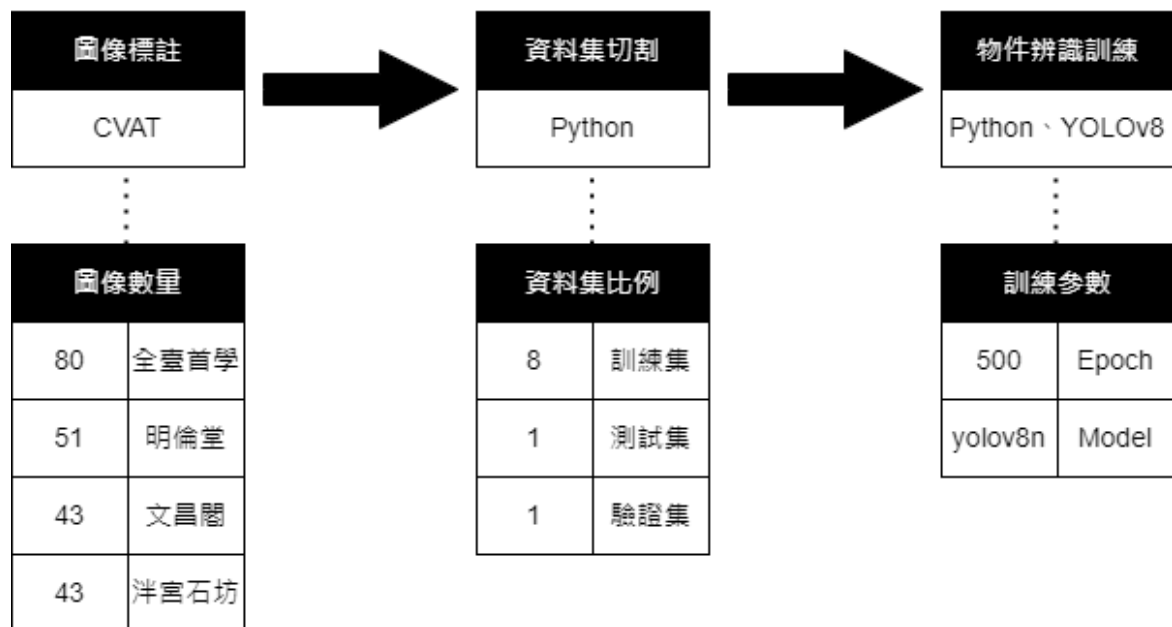


圖 三-4：物件辨識模型訓練流程圖

表 三-4：物件辨識模型訓練參數

參數選項	參數值	參考來源
訓練輪數 (Epoch)	500	試錯法
學習率 (Learning Rate)	0.01	預設
圖像大小 (Image Size)	640	預設
批次大小 (Batch Size)	16	預設
優化器 (Optimizer)	AdamW	預設

### 3.2.1.4 向量資料庫預處理

本研究進行 RAG 時使用的向量資料庫，資料來源自 20 個介紹孔廟歷史典故、建築特色的網頁。本研究使用 LangChain 內的套件以輕易達成此目的，步驟如下：

- (1) 使用 AsyncChromiumLoader 載入 20 個介紹孔廟歷史典故、建築特色的網頁。
- (2) 使用 CharacterTextSplitter 將網頁的文本以 1024 Bytes 的尺寸切割為 Chunk。
- (3) 使用 HuggingFaceEmbeddings 載入 sentence-transformers/paraphrase-multilingual-MiniLM-L12-v2(支援多語言的模型)將 Chunk 轉換為 Embedding。
- (4) 將結果存入向量資料庫 FAISS。

完整流程可參考圖 三-5。

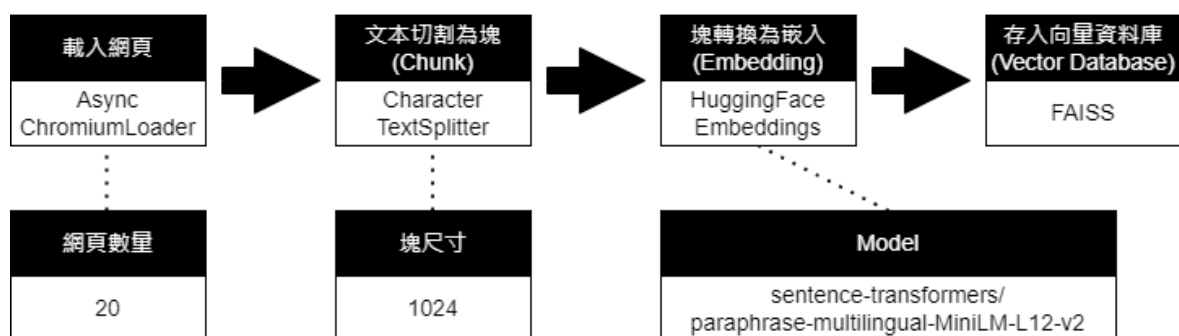


圖 三-5：向量資料庫處理流程圖



### 3.2.2 實驗設備

#### 3.2.2.1 單板機硬體配置

本研究使用了 Yahboom 的 Jetson Orin NX 的進階套裝[103]，參考圖 三-6。本研究主要使用了其中的 IMX219 攝影機作為影像輸入、觸控螢幕輸出程式執行日誌、鋁製機殼保護單板機。



圖 三-6：Jetson Orin NX 配件總覽[103]

本研究還額外使用了 Corsair 的 HS70 藍芽耳機麥克風作為使用者與程式之間的互動設備、ZMI 的 25000mAh Power Bank No.20 作為供電來源，其可提供單孔 USB-C 100W 的最大供電。

最終，將觸控螢幕置於上層，單板機裝入鋁製機殼後置於中層，行動電源置於下層。組裝起來用於實驗的成品，可參考圖 三-7。



圖 三-7：Jetson Orin NX 組裝成品

### 3.2.2.2 硬體環境

表 三-5：實驗準備-實驗設備-硬體環境

硬體類型	硬體名稱
CPU	8-core NVIDIA Arm® Cortex A78AE v8.2 64-bit CPU, 2MB L2 + 4MB L3
GPU	NVIDIA Ampere GPU with 32 Tensor Cores
CUDA Cores (GPU)	1024 Cores
Memory	16GB 128-bit LPDDR5, 102.4 GB/s

### 3.2.2.3 軟體環境

表 三-6：實驗準備-實驗設備-軟體環境

軟體類型	軟體名稱
SDK	Jetpack 5.1.1
OS	Linux Ubuntu 20.04
CUDA Version	11.8
Python Version	3.8.10
PyTorch Version	2.1.0
OpenCV Version	4.5.3
LangChain Version	0.1.8

### 3.2.2.4 使用軟體

表 三-7：實驗準備-實驗設備-使用軟體

軟體類型	軟體名稱
LLM	Breeze-7B[31]
Object Detection	YOLOv8
Speech-to-Text	Whisper[33]
Text-to-Speech	eSpeak

## 3.3 系統開發

以下將完整講解系統執行的流程，本研究將分為系統的啟動流程以及使用者在系統啟動完畢後能執行的指令，分別簡介，完整的系統流程圖可參考圖 三-8。

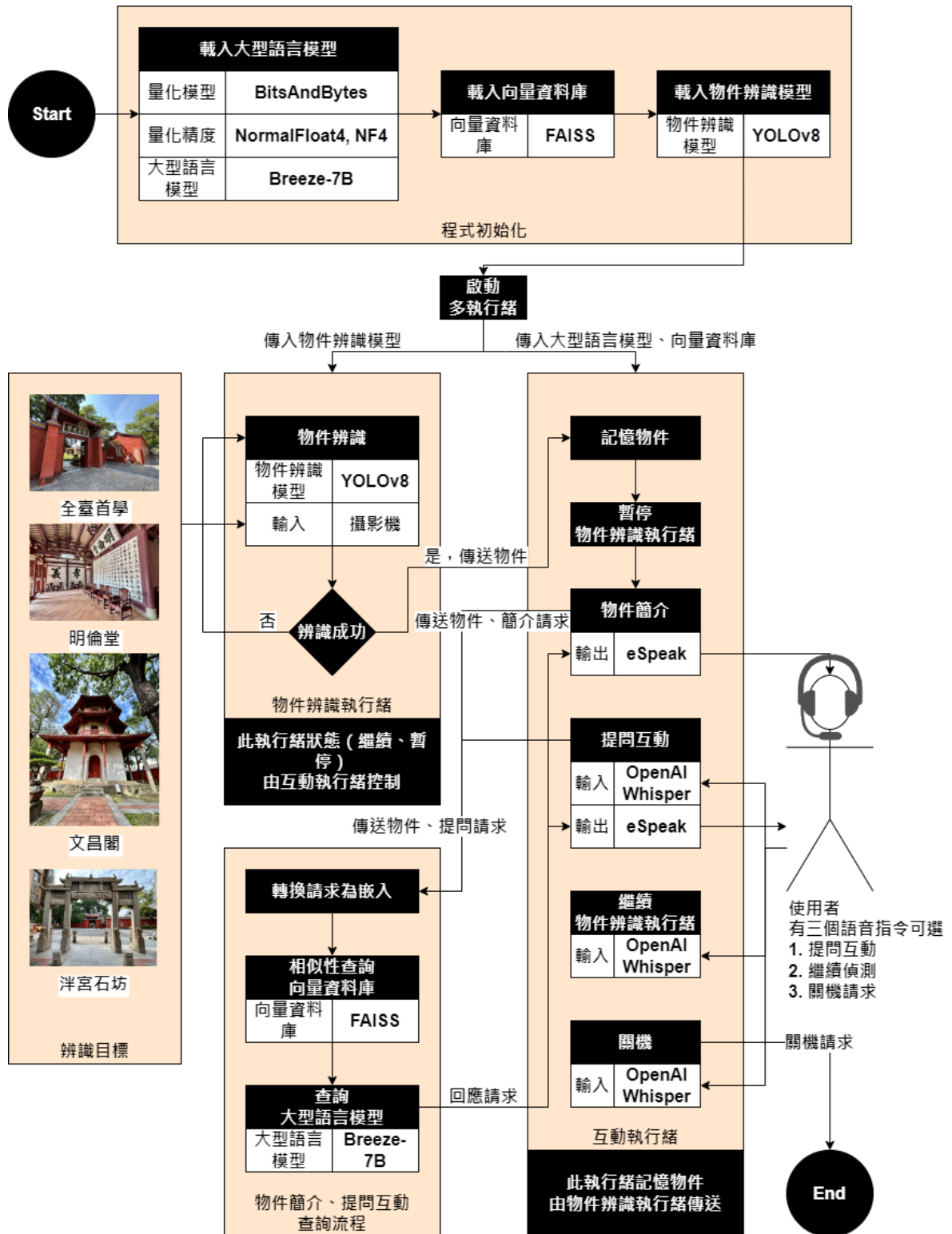


圖 三-8：系統流程

### 3.3.1 系統啟動流程

#### 3.3.1.1 程式初始化

- (1) 首先載入預訓練好的物件辨識模型，
- (2) 再來經由 BitsAndBytes 量化套件[22]使用 NormalFloat4 精度載入 Breeze-7B 大型語言模型
- (3) 接著再載入向量資料庫 FAISS 以預備執行 RAG。
- (4) 接著系統會啟動兩個執行緒：物件辨識執行緒、互動執行緒。  
★使用執行緒的原因是希望物件辨識推論與大型語言模型推論可以分開來運行，尤其 LLM 推論速度較慢，使用執行緒使其不會影響彼此的推論運行。

#### 3.3.1.2 啟動多執行緒

- (1) 物件辨識執行緒：  
動態接收攝影機傳回的畫面，並執行物件辨識：
  1. 若無辨識新物件，則繼續辨識。
  2. 若有辨識到物件，則將辨識物件傳送給互動執行緒
- (2) 互動執行緒：  
持續透過 Whisper 監聽使用者麥克風、接收物件辨識執行緒傳送來的物件，並交由 LLM 推論介紹：
  1. 接收並記憶物件辨識執行緒傳遞之物件。
  2. 暫停物件執行緒  
★此處暫停是因使用者須先與當前之物件互動，為避免於 LLM 推論過程中送來新物件，打斷原本的互動，因而暫停物件辨識執行緒。
  3. 將「物件」與「介紹」提示詞指令（Prompt）傳遞給 LLM 查詢方法：
    - a. 將提示詞指令「請簡短快速簡介」轉換為 Embedding。

- b. 在向量資料庫 FAISS 進行 Embedding 相似性搜尋（Similarity Search）。
- c. 將提示詞指令「請簡短快速簡介」、FAISS 查詢結果以及預設的 LLM 提示詞（如下），傳送給 Breeze-7B 模型。

您是一名導遊，遊客都只聽得懂中文，眼前看到的建築物，請您回覆精確、簡短，不要提到問題外的事物。 提示：[context] 問題：[question]	
context	FAISS 查詢結果
question	請簡短快速簡介

- d. 將查詢結果傳送回互動執行緒，並由 eSpeak 發聲至使用者耳機。

### 3.3.2 使用者可執行的指令

在系統啟動完畢後，使用者能透過中文語音互動，執行三種指令。

#### (1) 提問互動：

1. 對著麥克風先講「Hi」再講出要問的問題。  
 ★此處需要先講「Hi」是因擔心麥克風收到雜訊導致誤以為雜訊為使用者提問。
2. 系統此時會複述使用者的問題，請使用者回覆「Yes」或「No」。  
 ★此處系統會複述使用者問題，是擔心語音辨識錯誤。
3. 將「物件」與「問題」提示詞指令（Prompt）傳遞給 LLM 查詢方法：
  - a. 將「問題」提示詞指令轉換為 Embedding。
  - b. 在向量資料庫 FAISS 進行 Embedding 相似性搜尋（Similarity Search）。
  - c. 將提示詞指令「問題」、FAISS 查詢結果以及預設的 LLM 提示詞（如下），傳送給 Breeze-7B 模型。

您是一名導遊，遊客都只聽得懂中文，眼前看到的建築物，請 您回覆精確、簡短，不要提到問題外的事物。 提示：[context] 問題：[question]	
context	FAISS 查詢的 Chunk
question	問題

d. 將查詢結果傳送回互動執行緒，並由 eSpeak 發聲。

- (2) 繼續偵測：
- 對著麥克風講出「請繼續」，即可繼續物件辨識執行緒進行偵測。
- (3) 關機請求：
- 對著麥克風講出「關機」，即可退出程式。

3.4 實驗準備結果

3.4.1 物件辨識模型訓練

由於 YOLO 在標註物件時，使用中文會出現亂碼，故本研究將臺南孔廟內的四個標的建築物皆翻譯成英文，中英對照可參考表 三-8。以下圖片，基本上使用英文名稱。

表 三-8：物件辨識物件中英對照

中文名稱	英文翻譯
全臺首學	First Academy of Taiwan
明倫堂	Ming-Lun Hall
文昌閣	The Pavilion of the God of Literature
泮宮石坊	The Paifang

3.4.1.1 混淆矩陣

混淆矩陣的結果，可參考圖 三-9。

「First Academy of Taiwan」類別的樣本共有 14 個，其中 12 個被正確分類，2 個被錯誤分類為其他類別。「Ming-Lun Hall」類別的樣本共有 6 個，全部被正確分類。「The Pavilion of the God of Literature」類別的樣本共有 7 個，全部被正確分類。「The Paifang」類別的樣本共有 8 個，全部被正確分類。其中 7 個被正確分類，1 個被錯誤分類為其他類別。

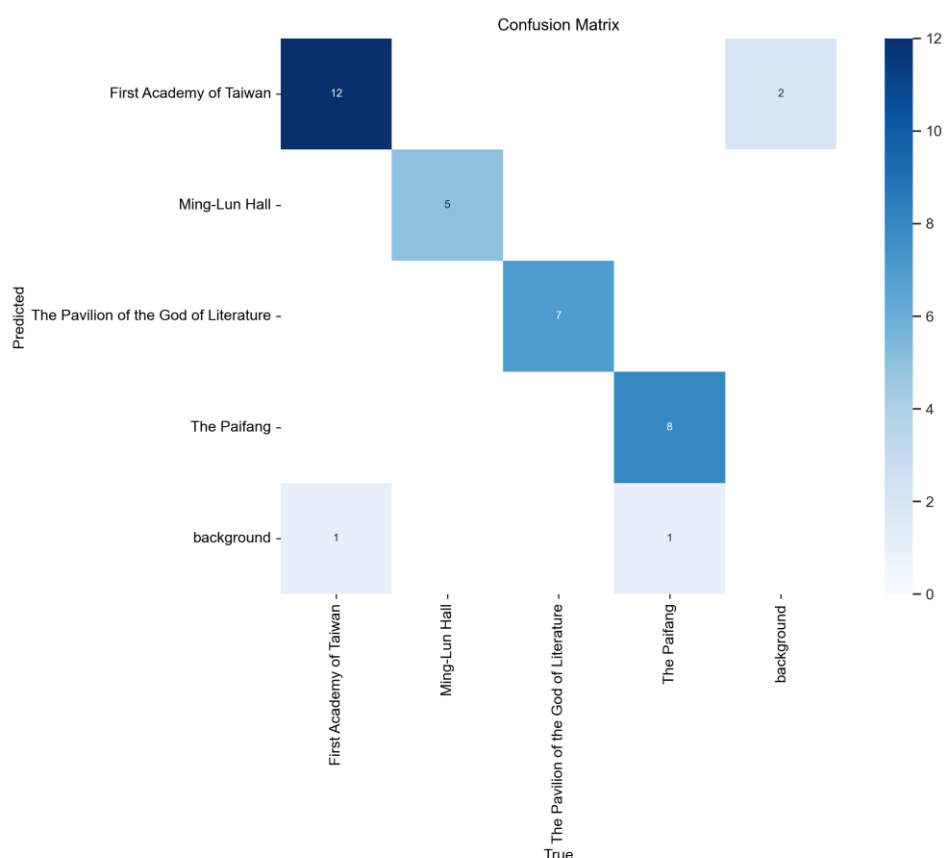


圖 三-9：物件辨識混淆矩陣

### 3.4.1.2 損失、精度、召回率和平均精度的分析

訓練和驗證過程中的結果，可參考圖 三-10。包括，邊界框損失（Box Loss）、分類損失（CLS Loss）、DFL 損失、精度（Precision）、召回率（Recall）和平均精度（mAP）。仔細分析這些結果可以看出，隨著 Epoch 的增加，所有損失指標、精度和 mAP 均呈現收斂趨勢。訓練過程的收斂表明模型在逐步學習如何減少誤差並提高其預測準確性。



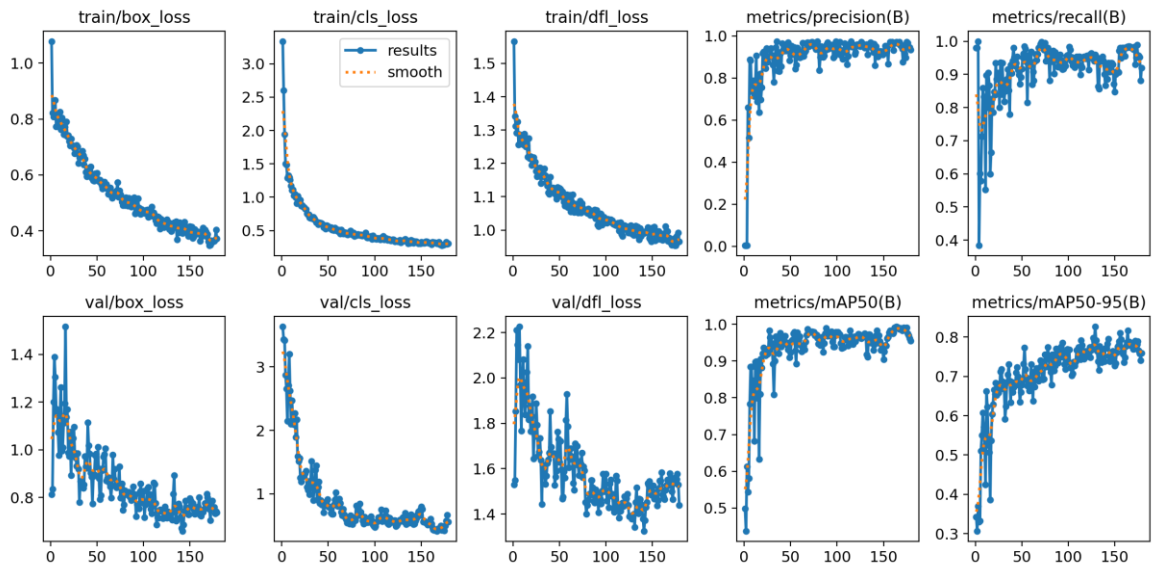


圖 三-10：物件辨識損失、精度、召回率和平均精度

### 3.4.1.3 模型性能指標曲線分析

指標曲線，可參考圖 三-11，展示了模型的性能變化，協助理解模型的檢測能力：

#### (1) 精度 - 召回曲線（Precision-Recall Curve）：

整體來看，模型在保持高精度的同時，也能夠維持較高的召回率。模型在所有類別上的  $mAP@0.5$  為 0.971，顯示了其良好的整體檢測性能。

#### (2) 召回率 - 信度曲線（Recall-Confidence Curve）：

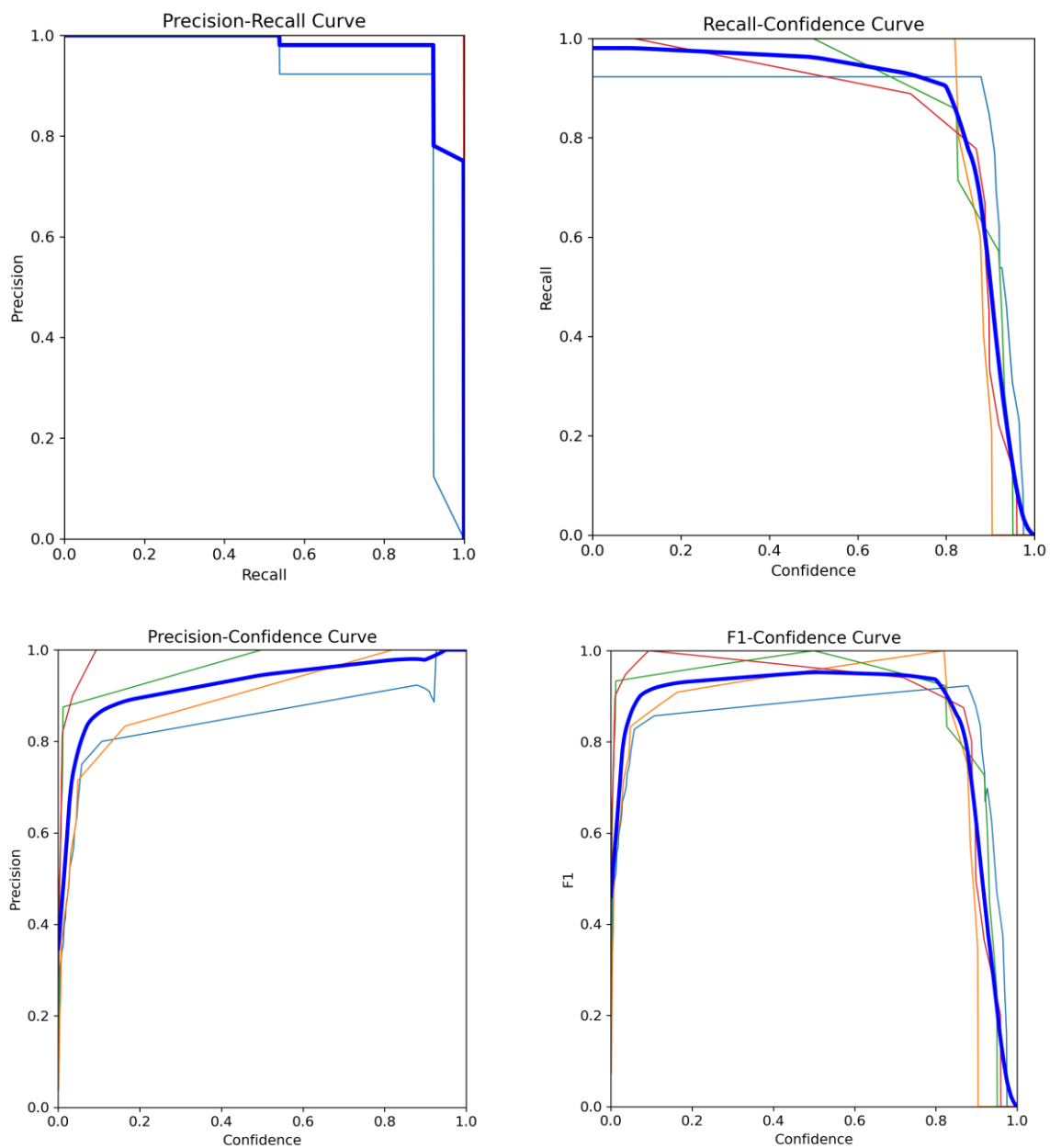
隨著信度的增加，模型的召回率逐漸下降。這表明在高信度下，模型可能會錯過一些目標，但能夠更準確地檢測到剩餘的目標。

#### (3) 精度 - 信度曲線（Precision-Confidence Curve）：

隨著信度的增加，模型的精度逐漸上升，並在高信度時保持穩定，表明模型在高信度下的預測更加準確。

#### (4) F1 - 信度曲線（F1-Confidence Curve）：

隨著信度的增加，F1 分數先上升後保持穩定，再逐漸下降。這反映了模型在不同信度下的綜合性能。



- all classes
- First Academy of Taiwan
- Ming-Lun Hall
- The Pavilion of the God of Literature
- The Paifang

圖 三-11：物件辨識性能指標曲線

### 3.4.1.4 模型預測結果

模型預測結果，可參考圖 三-12，可以看到大部分標的建築物都能正常辨識。

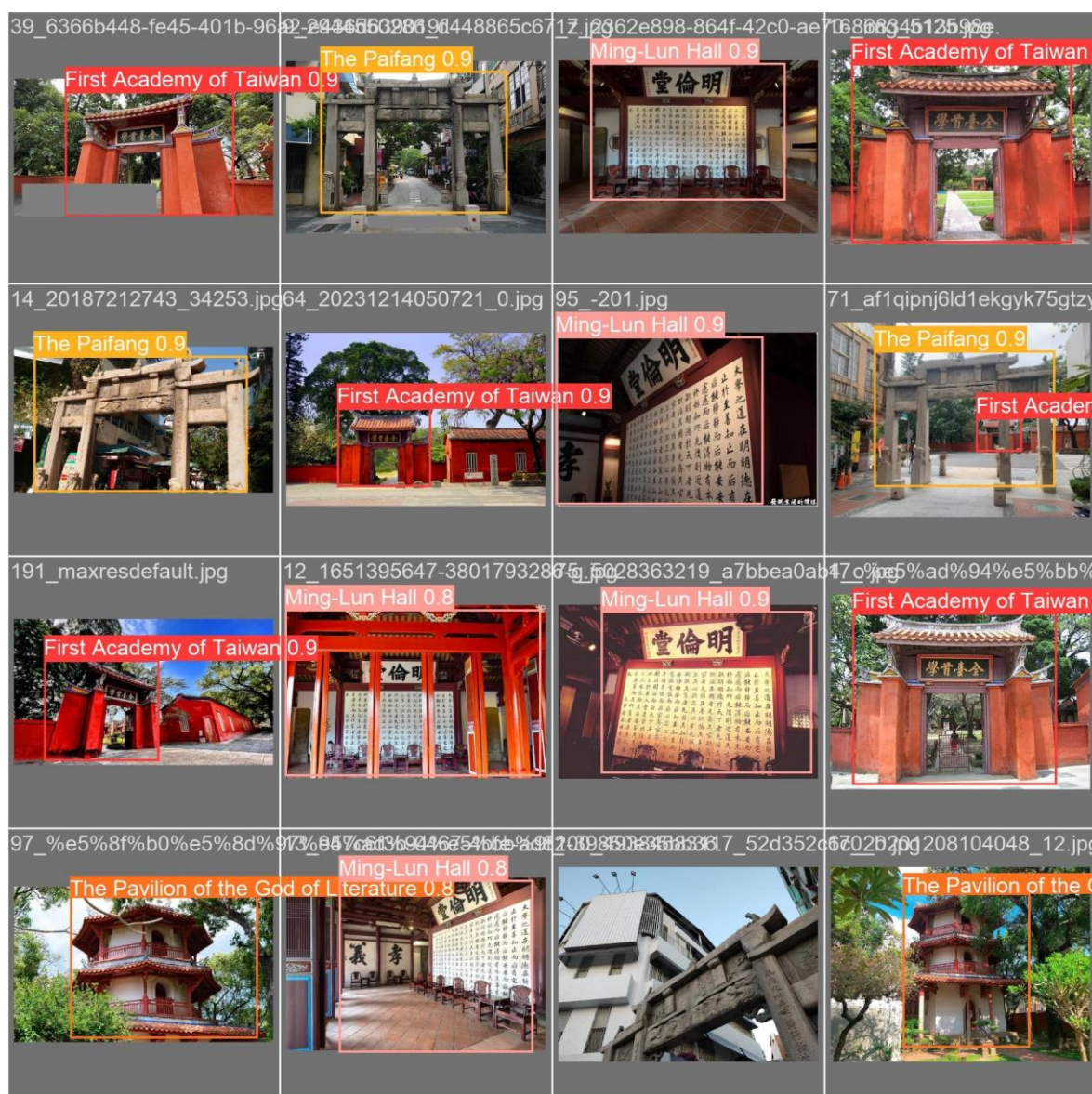


圖 三-12：物件辨識模型預測結果

### 3.4.1.5 綜合結果

訓練時長為 4.92 分鐘，訓練停止於 179 Epochs，但可提早停止於 129 Epochs。完整訓練結果可參考表 三-9。由表中可以看到物件辨識的精度都在 0.85 以上。平均精度更高達 0.948。

表 三-9：物件辨識模型訓練結果

類別	圖片數量	實例數量	精度	召回率	mAP@50	mAP@50-95
全部	31	34	0.948	0.959	0.971	0.826
全臺首學	31	13	0.866	0.923	0.900	0.806
明倫堂	31	5	0.924	1.000	0.995	0.920
文昌閣	31	7	1.000	0.990	0.995	0.756
泮宮石坊	31	9	1.000	0.924	0.995	0.823

### 3.4.2 向量資料庫預處理

向量資料庫預處理所花費的時間，可參考表 三-10。

表 三-10：向量資料庫處理時間

載入網頁	文本切割為 Chunk	Chunk 轉換為 Embedding	存入向量資料 庫 FAISS	總執行時長
86.72 秒	0.46 秒	0.23 秒	0 秒	87.41 秒

### 3.4.3 小節整理

無論是物件辨識訓練或向量資料庫預處理，在一般的電腦配置上，執行時間都能壓的非常短。本研究物件辨識訓練的時間小於五分鐘，而向量資料庫預處理的時間則在一分半內。也就是說，若目標資料集完備，本系統框架有極大的可能性能快速並大量地部屬在不同的目標上使用。例如：景點、植物、動物等。

## 第四章、使用者學習成效與 GAIUA 模型探討

### 4.1 實驗流程

本章將聚焦在實驗設計於探討使用者學習成效與 GAIUA 模型上，本研究將受測者共分為實驗組與對照組。使用者學習成效將透過受測者的前測與後測進行對比驗證，而研究模型的部分則是實驗組在實際體驗過設備導覽後透過量表進行測量。接下來，本研究將依據實際的實驗流程，由實驗對象的挑選、對受測者進行實驗講解、前測、實際導覽流程、後測、量表等，進行各項的簡介。實驗總流程，可參照圖四-1。

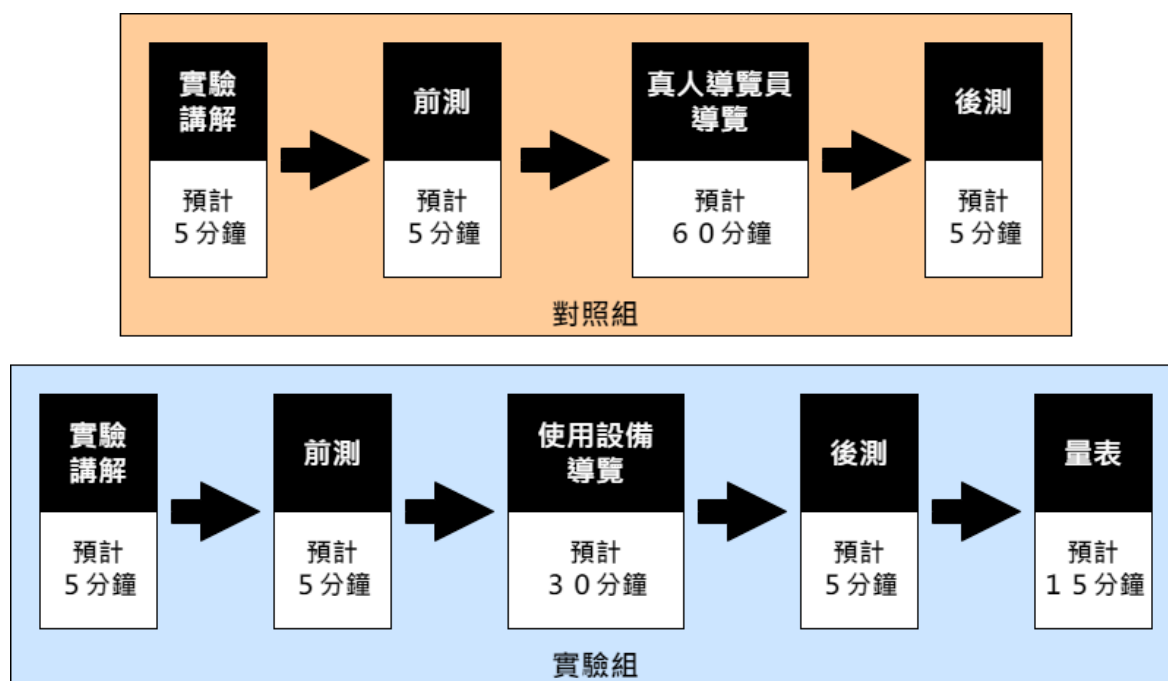


圖 四-1：實驗流程

#### 4.1.1 實驗對象

##### 4.1.1.1 實驗組

實驗組對象為隨機訪問臺南孔廟現場之遊客。

#### 4.1.1.2 對照組

對照組對象為報名參加每週日由臺南觀光旅遊局推出「米其林三星之旅」之「府城歷史散步-孔廟線」遊客。一梯次遊客最多 20 人。

#### 4.1.2 實驗講解

實驗講解為向實驗組與對照組個別講解，實驗的目的、實驗的流程等相關資訊。對照組會特別講解設備使用說明。

#### 4.1.3 前測

前測主要是驗證受測者於導覽前的知識水平，本研究總共提出十題與本研究使用的臺南孔廟內的四個標的物（全臺首學、明倫堂、文昌閣、泮宮石坊）之相關問題，問題清單可參考表 四-1。

表 四-1：前測問題清單

全臺首學成立於何時	<input type="checkbox"/> 1865 <input type="checkbox"/> 1665 <input type="checkbox"/> 1765 <input type="checkbox"/> 1965
明倫堂的大門稱為	<input type="checkbox"/> 入德之門 <input type="checkbox"/> 智慧之門 <input type="checkbox"/> 尊師之門 <input type="checkbox"/> 文化之門
明倫堂的地位相當於當時的	<input type="checkbox"/> 最高學府 <input type="checkbox"/> 最高行政機關 <input type="checkbox"/> 最高司法機關 <input type="checkbox"/> 最高法院
明倫堂最初的用途	<input type="checkbox"/> 教室 <input type="checkbox"/> 行政辦公 <input type="checkbox"/> 判決 <input type="checkbox"/> 監督執法
文昌閣的另一個名字	<input type="checkbox"/> 天守閣 <input type="checkbox"/> 魁星閣 <input type="checkbox"/> 萬世閣 <input type="checkbox"/> 寶生閣
文昌閣最初為何時建造	<input type="checkbox"/> 日治 <input type="checkbox"/> 清領 <input type="checkbox"/> 民國 <input type="checkbox"/> 荷蘭
文昌閣從一樓到三樓的形狀分別為	<input type="checkbox"/> 圓方八 <input type="checkbox"/> 圓八方 <input type="checkbox"/> 方圓八 <input type="checkbox"/> 八圓方
泮宮石坊的建造的材料源自	<input type="checkbox"/> 蘇州花崗岩 <input type="checkbox"/> 江西青石 <input type="checkbox"/> 四川大理石 <input type="checkbox"/> 河北砂岩
泮宮石坊象徵	<input type="checkbox"/> 郡縣之學 <input type="checkbox"/> 文化的交流 <input type="checkbox"/> 商業的繁榮 <input type="checkbox"/> 國家的權力
為了開闢道路，於何時將泮宮石坊與孔廟之間用道路隔開	<input type="checkbox"/> 日治 <input type="checkbox"/> 明鄭 <input type="checkbox"/> 荷蘭 <input type="checkbox"/> 民國

#### 4.1.4 導覽

實驗組與對照組導覽路線一致。

路線順序皆為全臺首學、明倫堂、文昌閣、泮宮石坊。可參考圖 四-2。

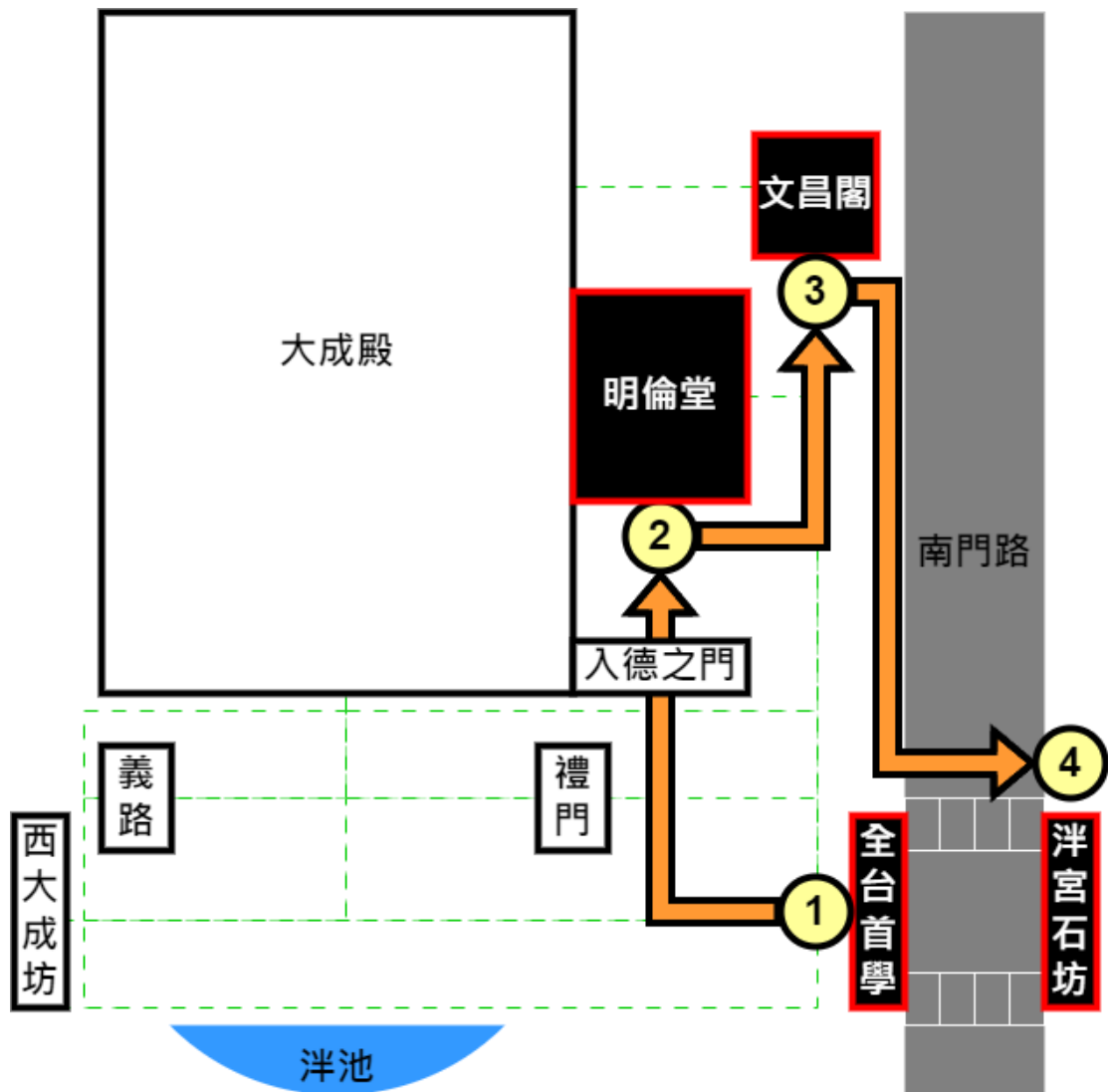


圖 四-2：導覽路線

##### 4.1.4.1 實驗組

實驗組導覽全程由實驗人員陪同，同時使用同一設備之受測者為一至五人不等。實驗方式為受測者戴上藍芽耳機麥克風，並提著設備分別由圖 18 的 1 走到 4。以受測者為角度的詳細導覽流程可參考圖 四-3。實際情景導覽可參考圖 四-4。觸控螢幕顯示的日誌輸出範例參考表 四-2。



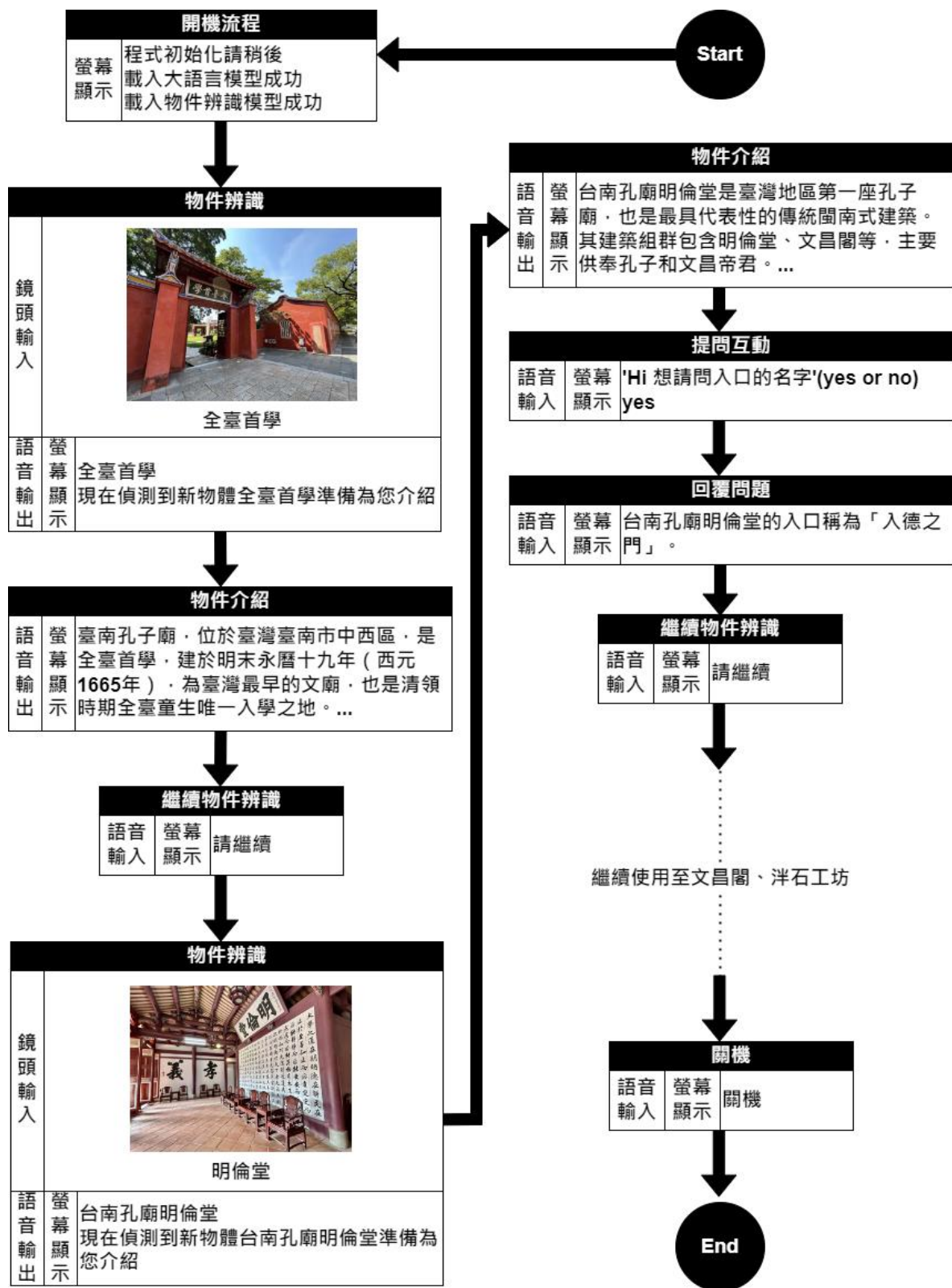


圖 四-3：實驗組導覽詳細流程





圖 四-4：實驗組實際導覽情形

表 四-2：觸控螢幕顯示日誌輸出範例

```
jetson@ubuntu:~/thesis$ ./activate.sh
```

程式初始化請稍後

載入大語言模型成功

載入物件辨識模型成功

影像輸入



全臺首學

現在偵測到新物體全臺首學準備為您介紹

臺南孔子廟，位於臺灣臺南市中西區，是全臺首學，建於明末永曆十九年（西元 1665 年），為臺灣最早的文廟，也是清領時期全臺童生唯一入學之地。該文廟於

民國 72 年（1983 年）12 月 28 日公告為祠廟類國家一級古蹟。臺南孔廟主要建築包括大成殿、文昌閣、子祿閣和子聖閣等，為木結構、木構架的建築物，具有歷史、文化價值。

Ask a question (or say '關機' to exit) :

請繼續

影像輸入



台南孔廟明倫堂

現在偵測到新物體台南孔廟明倫堂準備為您介紹

台南孔廟明倫堂是臺灣地區第一座孔子廟，也是最具代表性的傳統閩南式建築。其建築組群包含明倫堂、文昌閣等，主要供奉孔子和文昌帝君。明倫堂為全臺的最高學府，在明鄭時期的臺南市發揮了重要教育和文化功能。在歷史變遷與戰爭摧殘中，明倫堂的建築部分毀損，目前的建築風貌主要是 1917 年的重修後的面貌，維持清代的風格和規制。

Ask a question (or say '關機' to exit) :

'Hi,想請問入口的名字'(yes or no)

yes

台南孔廟明倫堂的入口稱為「入德之門」。

Ask a question (or say '關機' to exit) :

請繼續

影像輸入



## 孔廟文昌閣

現在偵測到新物體孔廟文昌閣準備為您介紹

臺南孔廟文昌閣，又稱「魁星樓」，是孔廟建築組群中唯一的樓塔式建築。位於臺南孔廟明倫堂左後方，建築由圓形、方形臺基而至一樓方形、二樓圓形、三樓八角形。二樓供奉文昌帝君，三樓奉祀魁星。昌化年（1713）時，陳瓚在新建文昌閣表示：「有志之士，無急求名於世，而務積學於己，亦無徒乞靈于神，而務常操其未放之心。」文昌閣保存戰前初期的風貌，具有歷史意義，是臺南孔廟最具代表性的傳統閩南式建築。

Ask a question (or say '關機' to exit) :

'Hi,想請問這裡是何時蓋的'(yes or no)

yes

文昌閣是臺南孔廟的一部分，它在乾隆四十二年（1777 年），知府蔣元樞加建的時候，在大成殿西邊增設了府學署（文昌閣）。

Ask a question (or say '關機' to exit) :

請繼續

影像輸入



## 泮宮石坊

現在偵測到新物體泮宮石坊準備為您介紹

泮宮石坊是臺南孔廟的重要建築，位於孔廟東大門外的出入口。它代表郡縣之學，是大成殿東北方的建築。泮宮石坊的材質是泉州的花崗岩，具有悠久歷史且是全臺唯一的石坊。建築設計模仿木建築，有鴟尾雀替、瓦當滴水飛簷等裝飾。後半部八隻石獅都非常可愛。

Ask a question (or say '關機' to exit) :

'Hi,想請問這裡的變化'(yes or no)

yes

泮池是臺南孔廟大成坊的建築物，它在日治時期為了開闢南門路而切割遷移至現今府中街口。在臺南孔廟，泮宮石坊是唯一的石坊，其材質為泉州的花崗岩。在



歷史上，爰池的重要性來自孔子受封為宣王後的地位，因此依諸侯規制修建爰池，供新入學生員進行入泮儀式。

爰池的北側原有的櫺星門毀壞後未重修，已不存，但在西元4月，百年老榕因受褐根病感染，倒塌而壓毀緊鄰的古蹟紅磚牆和「思樂泮水」石碑。專家推測這磚牆可能是清代孔廟的櫺星門的遺構。目前，臺南市文資處正在擴大遺構發掘面積，並邀請專家學者討論未來保存展示的方式。

#### 4.1.4.2 對照組

對照組導覽全程由「府城歷史散步 - 孔廟線」的真人導覽員引導介紹。實際情景可參考圖 四-5。



圖 四-5：對照組實際導覽情形

#### 4.1.5 後測

後測主要是驗證受測者於導覽後的知識水平，並與在導覽前進行對比，最終結果可以驗證經過導覽後（真人導覽員、使用設備）的學習成效。與前測是同樣的十題，但選項經過亂序處理。

#### 4.1.6 量表

本研究的量表題項採用李克特五點量表，並只有實驗組需填寫，在後測結束後接著進行。依據本研究「探討 GAI 的使用意圖」研究模型的六個構面。包括，解決問題（Problem-Solving）、感知人性（Perceived Humanness）、回應內容（Response Content）、類人互動（Human-like Interaction）、滿意度（Satisfaction）、使用意圖（Using Intention）等。共提出 23 個問題，題項分布可參考表 四-3。詳細的題項內容會在下一節探討完研究模型後再提出。

表 四-3：量表題項分布

探討構面	解決問題	感知人性	回應內容	類人互動	滿意度	使用意圖
題項數量	4	4	4	4	4	3

## 4.2 研究模型

目前探討 AI 使用意圖相關的研究模型，只有社交導向的 AIDUA（Artificially Intelligent Device Use Acceptance）[49]以及任務導向的 T-AIA（Task-oriented AI Acceptance）[43]。只是這兩種研究模型都是針對一般的 AI 服務，並無法完全適用於 GAI 設備。

也因此，鑒於目前沒有相關研究針對 GAI 設備提出研究模型。本研究便提出一套創新的研究模型，同時做為評估本系統成效的重要依據，探討 GAI 的使用意圖（Generative Artificially Intelligent Use Acceptance, GAIUA）。

並且，本研究使用的 GAIUA 是由理性行為理論（Theory of Reasoned Action, TRA）所推演而出。因此以下會先介紹 TRA，再介紹其如何影響到 GAIUA 的設計。

### 4.2.1 理性行為理論

TRA 於 1977 年由 Ajzen 與 Fishbein 共同提出，其是一個在社會心理學中廣泛使用的模型，關注有意識行為的決定因素[104]。TRA 的架構可參考圖 四-6。

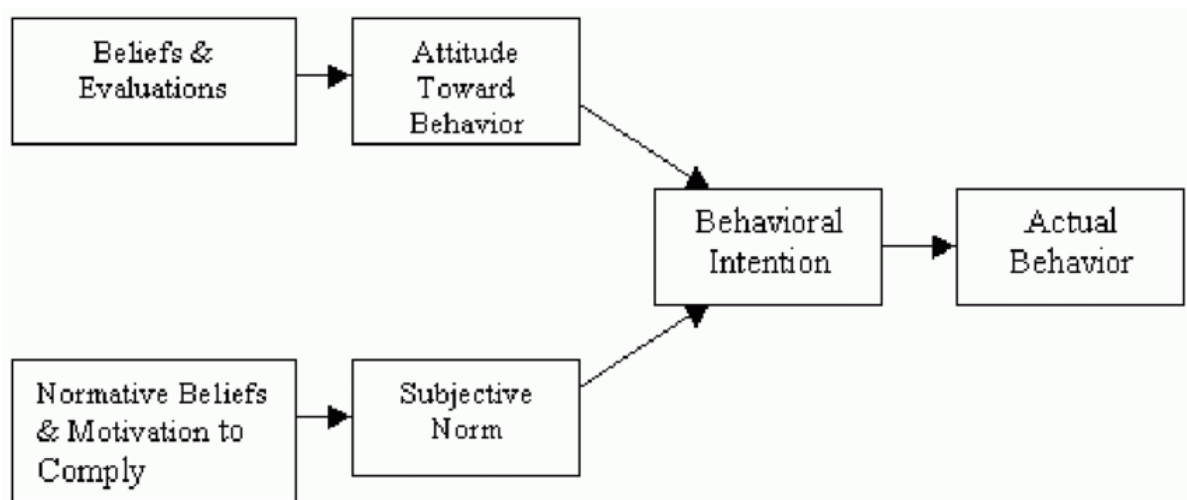


圖 四-6：理性行為理論[104]

在 TRA 的架構中，信念（Beliefs）被定義為個人對於目標行為結果的主觀感受、評價（Evaluations）被定義為個人對於目標行為結果的主觀評價。對行為的態度（Attitude Toward Behavior）被定義為個人對執行目標行為的正面或負面感受。

TRA 認為，一個人對某行為的態度是由其對於執行該行為結果的顯著信念乘以對這些結果的評價所決定的。

在 TRA 的架構中，規範信念（Normative Beliefs）被定義為參考特定個人或群體的預期，與其遵守這些預期的動機（Motivation to Comply）。主觀規範（Subjective Norm）指的是個人對於重要的人認為他應該或不應該執行特定行為的看法。TRA 認為，個人的主觀規範是由其的規範信念乘以其遵守這些預期的動機所決定的。

最後，TRA 認為一個人執行特定行為（Actual Behavior）的表現是由其行為意圖（Behavioral Intention）決定的，也就是一個人執行特定行為意圖的強度，而行為意圖是由該人對於行為的態度（Attitude）和主觀規範（Subjective Norm）所共同決定的。

#### 4.2.2 生成式人工智慧的使用意圖研究模型

本研究使用的研究模型，主要是參照 TRA 進行設計，其中主觀規範分支由於與本研究目的不符，因此 GAIUA 的設計主要參照 TRA 的上層路線（信念與結果評估、行為態度、行為意圖）並依照 GAI 的特徵，設計出 GAIUA 的架構，完整架構可參考圖 4-7。

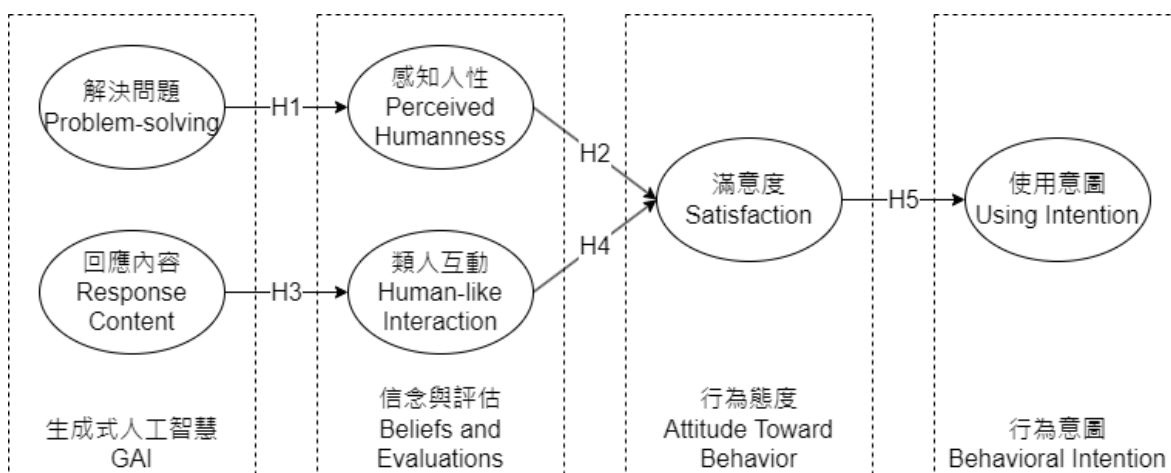


圖 4-7：生成式人工智慧的使用意圖研究模型

GAIUA 使用的構面主要有六個：解決問題（Problem-Solving, PS）、感知人性（Perceived Humanness, PH）、回應內容（Response Content, RC）、類人互動

( Human-like Interaction, HI ) 、滿意度 ( Satisfaction, SA ) 、使用意圖 ( Using Intention, UI ) 等。其中，使用意圖、滿意度在探討使用意圖的研究中，已是被廣泛使用的構面。因此將傾向於探討其餘的構面在 GAI 上的重要性。

本節將先探討構面的定義，接著將在推演構面之間路徑的假說。

### 4.2.3 構面

#### 4.2.3.1 解決問題

「解決問題」是評估 GAI 在各種應用場景中的一個重要構面，直接影響其在面對複雜問題時的表現和用戶接受度。Brandtzaeg and Følstad (2017) 的研究表明，客戶使用 AI 機器人的主要目的是提高生產力[105]。Hsiao and Chen (2022) 將「解決問題」定為 AI 能準確、立即為每個客戶解決問題的能力，增強使用者的信任，進而增加使用意願[106]。Wu et al. (2023) 指出，GAI 的出現對勞動市場的影響是巨大的，AIGC 能夠在短時間內生成大量內容，而人工創作通常需要更長的時間[14]。

Daugherty et al. (2023) 的研究發現，像 ChatGPT 這樣的 GAI 應用，可以使大部分的職業將原本的工作任務細化拆分，將部分任務交由 GAI 處理，只保留高價值的任務由人類親自執行[107]。他們發現，業務代表大部分的工作可以分解為十三種任務。其中有四項任務可以由 GAI 實現完全自動化。另有五項任務可由 GAI 與人類協作執行。這些是 GAI 出現後的重大變革，企業應圍繞 GAI 重新設計工作，以期能從中獲取最大優勢[107]。

#### 4.2.3.2 感知人性

相較於「感知人性」，在形容一個 AI 機器人類人時，「擬人化」( Anthropomorphism ) 是一個更常出現的字彙。在 AIDUA 研究模型中也是使用這個字彙作為構面[49]。Bartneck et al. (2009) 指出，擬人化是指感知到的人類相似程度，如自我意識、人類外表和情感[108]。

日本機器人科學家 Masahiro Mori 於 1970 年提出「恐怖谷」理論，隨著機器人的外觀和動作變得越來越像人類，人類對機器人的情感反應變得越來越積極和富有



同理心，然而到達一個點之後，反應很快就會變成強烈的排斥。隨著外表和動作與人類的區別越來越小，情緒反應才會再次變得積極並且接近對待人類的同理心的水平[109]。可參考圖 四-8。

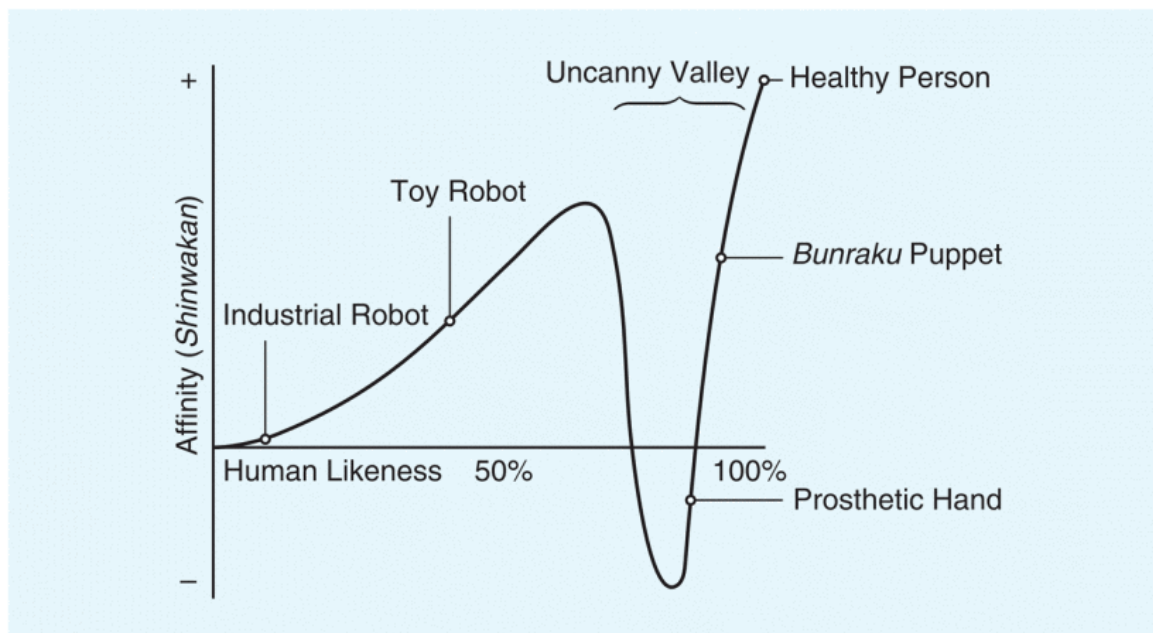


圖 四-8：恐怖谷理論[109]

然而，考量到本研究提出的系統以及目前大多數的 GAI 應用（如 ChatGPT），基本上都是使用一般的使用者介面（User Interface），而並不是像機器人一樣擁有擬人化的外觀。故本研究認為在探討 AI 擬人化的程度時，「感知人性」是更適合作為探討目前 GAI 人性的字彙。以 ChatGPT 來說，期能展現出的人性特徵，包括能承認自己的錯誤、挑戰不正確的論述、拒絕回覆不適當的指令等[6]。

Li et al. (2022) 指出，隨著 AI 技術的發展，人類與 AI 的互動可能變得更加複雜，因為這些互動融合了人際關係中典型的關係和社會層面[110]。也因此 Ma and Huo (2023) 指出，雖然社群媒體皆稱讚 ChatGPT 的思維模式近似人類，然而這裡的擬人化實際上是一種技術進步，而不是一種真正意義上的自我意識[111]。Li et al. (2022) 的研究發現，與感知人性程度低的消費者相比，感知人性程度高的消費者更傾向於與 AI 電話銷售員產生共鳴，對這些 AI 實體也持有更好的態度，並且與它們的互動時間更長[110]。

因此，探討人性的概念對於研究人類與 AI 的互動是至關重要的。如同 Rapp et al. (2021) 研究綜述，人們在賦予聊天機器人人性時，可能會考慮不同的維度，並依賴不同的線索來判斷對話者是人類還是機器。除了語言層面上的因素，如語法、打字方式、合理性和語言風格之外，他們還重視某些心理和互動因素，如撒謊的能力、幽默感、高水平的知識交流、意識、原創性、態度和全面性等[112]。

Ma and Huo (2023) 指出，雖然 ChatGPT 僅僅是一個人工智慧應用，但它有能力生成與人類寫作非常相似的回應。這使得 ChatGPT 的回應和答案充滿了人性，也使得用戶對 ChatGPT 的期望有所提升。這樣子的類人回應以及對用戶的體貼友好，使得與 ChatGPT 的互動變得輕鬆、愉快[111]。從而能增加其使用意圖。

#### 4.2.3.3 回應內容

近年來，人工智慧生成內容（Artificial Intelligence Generated Content, AIGC）迅速發展，由 GAI 演算法生成滿足命令的文本、圖像和影片，其目標是使人們創作的過程更加高效和便捷，從而更快地產出高品質的內容。技術上來說，AIGC 生成的過程通常包含以下兩步驟[113]：

- (1) 從人類下達的命令中提取其實際意圖。
- (2) 根據提取的意圖訊息生成對應內容。

然而，儘管像 ChatGPT 這樣的工具能夠生成符合事實且合理的內容[6]，但有時，其也會輸出反事實、荒謬且充滿幻覺的回應[114]。並且，這種狀況不僅發生在 NLP 領域，在電腦視覺（Computer Vision）領域也存在類似的問題。例如，Stable Diffusion 在生成人類的手部圖像存在著困難，其無法正確地描繪手指的數量[115]。Stackoverflow 甚至採取了禁止用戶使用 AIGC 作為答案的政策，以減少被不準確和偏頗的內容淹沒的風險[116]。

也因此，近年甚至出現了提示工程（Prompt Engineering）這樣的生成內容優化方法[117]。提示（Prompt）是下給 GAI 的指令，其可用於添加生成的規則，以確保生成內容的品質。例如，ChatGPT 可以在使用時通過具體的規則和指南來增強其生

成的內容[6]、Stable Diffusion 在使用時可以添加負向提示詞，來減少 GAI 生成時繪製不良的部分[115]。

由此可見 AIGC 在使用 GAI 時所代表的重要性，本研究將此構面的名稱定為「回應內容」，便是由於 AIGC 須經由人類下達指令觸發，而 GAI 回應時所產生的內容。回應的內容與指令下達者的期望是否相符，便是 GAI 相當重要的課題。

#### **4.2.3.4 類人互動**

個體之間的溝通和互動是社會存在的特徵之一，「類人互動」指的是 GAI 能夠模仿人類的語言和行為模式，進行自然、流暢的互動。Araujo (2018) 指出在不同服務產業的背景下，AI 設備的成功取決於消費者和 AI 進行互動，使消費者得到所需的服務之餘，也滿足了 AI 設備的建置意義[118]。然而，自然的互動被認為是人類彼此之間獨有的。

因此，有幾項研究重點調查消費者在 AI 與人類員工之間互動的差異比較。Hill et al. (2015) 表明，消費者雖願意與 AI 機器人交流，但來往的訊息較短[119]。Mou and Xu (2017) 的研究則指出，消費者與 AI 機器人交流時，較少進行自我揭露[120]。然而，Ciechanowski et al. (2019) 的研究結果表明，若 AI 機器人展現出類人互動的特質時，消費者往往願意花更多時間互動[121]。同樣的結果可以在社交護理機器人上觀察到。Fan et al. (2017) 的研究指出，老年用戶更傾向於與展現出類人互動特質的 AI 機器人交流，其可以充當教練或指南，吸引和鼓勵老年用戶完成一系列的治療任務，以增強他們的身體機能、認知功能以及健康狀況[122]。Go and Sundar (2019) 的研究指出，如果使用者認為 AI 機器人與自己有相似的對話模式，他們更可能會善意地評價該次互動[123]。

#### **4.2.4 路徑假說**

##### **4.2.4.1 解決問題對感知人性有正面影響 (H1)**

Kemeny (1955) 表示「一個正常的人就像一台通用機器。如果有足夠的時間，他可以學會做任何事情，沒有確鑿的證據表明人與機器之間存在本質差距。對於每一

項人類活動，我們都可以設想一個機械對應。」[124]，將人類與機器關聯起來，是打從 AI 的概念一出現就已存在。McCarthy et al. (1956) 於 1956 年舉辦的達特茅斯夏季人工智慧研究計劃（Dartmouth Summer Research Project on Artificial Intelligence）目標為「研究嘗試找到讓機器使用語言、形成抽象和概念、解決目前由人類處理的各類問題，以及自我改進的方法。」[125]。也就是說從一開始，AI 的目標就是整合不同研究領域以使機器實現人類智慧的雄心壯志。例如語言、視覺和解決各類型的問題。

直到 1970 年代，AI 首次被描述為能夠「完美模擬人類心智能力」的思考機器[126]。而在 IBM 發布 Deep Blue，一個戰勝國際象棋世界冠軍的 AI 後，公眾給 AI 的擬人化修辭由模擬人類心智開始轉變為直接比較 AI 和人類的智能差距，這樣的比較在 AlphaGo 擊敗圍棋世界冠軍李世乭（Lee Sedol）後達到頂峰。直到 GAI 的到來。其模仿人類創作的文本、音樂、圖像甚至影片的能力，使其更容易受到擬人化。GAI 與專為特定任務設計的傳統 AI 不同，其能夠處理多模態、跨情境、通用的任務。因此，社會將 GAI 視為「不同於以往的顛覆性事物」（something disruptive with no ties with the past）以 Bory (2019) 的話詮釋[127]。

可以看到隨著 AI 系統的逐步發展，圍繞 AI 出現的描述與定義也在逐漸改變。如前面章節所述，Ma and Huo (2023) 指出，雖然社群媒體皆稱讚 ChatGPT 的思維模式近似人類，然而這裡的擬人化實際上是一種技術進步，而不是一種真正意義上的自我意識[111]。

綜上所述，AI「解決問題」的能力與科技的進步密不可分，其能力的進步還能影響到公眾看待 AI 的態度，由 1970 年代的模擬人類開始，在 2000 年左右直至近十年又走向超越人類。直到 GAI 的到來，又因其擬人的多模態能力，使人類又開始感知到 AI 中的人性。

鑒於 AI「解決問題」的能力提升與公眾從 AI 中「感知人性」會因而改變的微妙關係。本研究假設「解決問題」能對「感知人性」產生正面的影響。具體來說，假說 H1 的內容為：「解決問題對感知人性有正面影響」。

#### 4.2.4.2 感知人性對滿意度有正面影響 (H2)

目前已有許多研究證實 GAI 的「感知人性」(或擬人化)對於「滿意度」有正面積極的作用，Ma and Huo (2023)的研究表明，使用者與聊天機器人互動後獲得的滿意度取決於他們理解人性、對突發事件的感知以及類人的回應能力[111]。Schuetzler et al. (2020) 的研究指出，當用戶與聊天機器人互動，收到聊天機器人理解他們的回應會讓他們感到被傾聽，這種 GAI 展現的感知人性能激發用戶的社會認知，從而使他們更有可能繼續與聊天機器人互動[128]。

Faruk et al. (2023) 在探討大學生對 ChatGPT 接受度的研究指出，感知人性會顯著影響 ChatGPT 的使用。其能創造一個良好的使用者體驗環境，從而影響使用者使用這些服務的傾向[129]。

綜上所述，本研究假設「感知人性」對於「滿意度」有正面積極的作用。具體來說，假說 H2 的內容為：「感知人性對滿意度有正面影響」。

#### 4.2.4.3 回應內容對類人互動有正面影響 (H3)

雖然目前尚無研究針對「回應內容」與「類人互動」做出研究假說，但能從 GAI 相關的研究推導出端倪。AIGC 已成為人們工作和生活的重要組成部分，人類透過 GAI 可以產生逼真的文字、圖像和其他類人的輸出（包括音樂、詩歌、繪畫等），這種模式目前正在改變許多不同的產業。例如，人類和 GAI 的互動可以在藝術創作中形成有效的協作夥伴關係[130]、利用基於 LLM 的會話代理（Conversation Agent, CA）還可以作為自閉症患者的互動工具[131]。

在 Choi et al. (2024) 的自閉症研究中指出，自閉症患者能透過簡單地向 CA 下指令，在面臨意想不到的日常挑戰時可以隨時存取大量資訊。此外，CA 還可以分析自閉症患者的個人特徵和使用意圖，以提供更個人化的回應內容。憑藉這些能力，CA 有望幫助自閉症患者獨立應對日常生活，從日常任務、自我照顧到社交互動。Choi et al. (2024) 指出，參與者向 ChatGPT 詢問了一系列生活問題，涵蓋了諸如如何處理突然出現的蟑螂、如何限制在遊戲上的過度支出、如何搭配衣服的顏色、如何戰勝懶惰以及討論戒菸策略等主題[131]。如此生活化的互動模式彰顯著 GAI 個人化的

「回應內容」能與使用者間的「類人互動」有著積極的關係。最後 Choi et al. (2024) 的研究結果表明，ChatGPT 有邏輯性和說服力的「回應內容」給自閉症患者留下了深刻的印象，從而產生了高度的信任。他們甚至認為 ChatGPT 比人類更聰明，因其語法和拼字完美無缺[131]。

Wang et al. (2020) 研究發現，使用者與 AI 之間的互相理解有望帶來更成功的互動[132]。隨著 AI 技術的不斷進步，未來期望看到人類與 AI 之間更加複雜和個人化的類人互動。

綜上所述，本研究假設 GAI 的「回應內容」會對「類人互動」產生正面積極的影響。具體來說，假說 H3 的內容為：「回應內容對類人互動有正面影響」。

#### **4.2.4.4 類人互動對滿意度有正面影響 (H4)**

Schuetzler et al. (2020) 的研究指出，當 CA 能夠保持自然流暢的對話流程時，互動被認為是更具吸引力的。能夠理解使用者輸入並做出適當回應的 CA 被認為是好的溝通者，並且能讓人們感覺到他們的意見被聽到。相反，當 CA 給予不恰當的回應並且無法保持流暢的對話時，人們就會分心並感覺沒那麼投入，導致與系統的互動效率降低[128]。也因此，Schuetzler et al. (2020) 的研究結果表明，人們認為具有較高會話技能的聊天機器人比較低的聊天機器人更具人性化與吸引力[128]。

Cohn et al. (2024) 的研究發現，相比於純文字介面，參與者認為當他們聽到系統說話時，系統提供的資訊更準確[133]。這項發現與自動駕駛的研究一致，Waytz et al. (2014) 的研究指出，在自動駕駛車輛系統中人類聲音的存在能增加使用者信任[134]。

如同第二章所述，互動，尤其語音互動的重要性，在於口語是人類最自然的溝通方式。Scott et al. (2019) 指出，如果語音助理 (Voice Assistant, VA) 能夠做到與電腦或智慧型手機等其他設備相同的事情，消費者更傾向於與 VA 互動[69]。並且，口語互動同時也是一種壓倒性的社交行為[70]。消費者在與 VA 交談時可以形成一種社會臨場感 (Social Presence)，而這種社會臨場感便是指使用科技設備時，能接收到的溫暖、社交性和親密感[71]。引發社會臨場感的同時，可以改善消費者對新技術

的體驗並引發積極的行為產生[72]。VA 擁有兩個顯著的類人特徵（說和聽）可以導出這種社會臨場感。

綜上所述，本研究假設 GAI 的「類人互動」會對「滿意度」產生正面積極的影響。具體來說，假說 H4 的內容為：「類人互動對滿意度有正面影響」。

#### 4.2.4.5 滿意度對使用意圖有正面影響（H5）

「滿意度」是對產品或服務達到或超出預期時的評價，這些評價源自於使用者在使用該產品時的服務體驗。Jaiswal et al. (2022) 的研究表明，「滿意度」在決定技術使用的方面發揮著重要影響[135]。Mofokeng (2021) 指出，高滿意度往往會促進持續使用並降低技術放棄的可能性[136]。Reim et al. (2020) 表示，在 AI 的背景下，當使用者發現 AI 工具令人滿意時，他們更有可能將其納入日常工作中，並依賴它來完成平時的各種任務[137]。因此，預計如果 GAI 能夠有效滿足使用者的需求，他們將在未來持續使用它。

故本研究假設 GAI 的「滿意度」會對「使用意圖」將產生正面積極的影響。具體來說，假說 H5 的內容為：「滿意度對使用意圖有正面影響」。

#### 4.2.5 量表詳細題項

依照前述小節的構面探討以及路徑假說，GAIUA 使用的構面主要有六個：解決問題（Problem-Solving, PS）、感知人性（Perceived Humanness, PH）、回應內容（Response Content, RC）、類人互動（Human-like Interaction, HI）、滿意度（Satisfaction, SA）、使用意圖（Using Intention, UI）等。本研究提出共 23 個題項來量測 GAIUA 模型，可參考表 四-4。

表 四-4：量表詳細題項清單

次序	代碼	構面	題項
1	PS1	解決問題	我認為 AI 提供的回答符合我的需求
2	PS2	解決問題	我認為 AI 的回答內容充分且具體
3	PS3	解決問題	我認為 AI 能夠提供我需求的實用資訊
4	PS4	解決問題	我認為 AI 在提供回答時展示出了對於當前詢問話題的理解
5	PH1	感知人性	我認為 AI 具有類似人類對於問題理解的能力
6	PH2	感知人性	我認為 AI 能模仿人類進行對話交流的表現
7	PH3	感知人性	我認為 AI 有能力在談話中顯示出類似人類的思考模式
8	PH4	感知人性	我認為 AI 具有足夠的語言豐富度以模仿人類
9	RC1	回應內容	我認為 AI 提供的回答符合我的需求
10	RC2	回應內容	我認為 AI 的回答內容充分且具體
11	RC3	回應內容	我認為 AI 能夠提供我需求的實用資訊
12	RC4	回應內容	我認為 AI 在提供回答時展示出了對於當前詢問話題的理解
13	HI1	類人互動	我認為 AI 能理解並適應我們的交談
14	HI2	類人互動	我認為 AI 能迅速且恰當地回應我的問題
15	HI3	類人互動	我認為與 AI 的交流，充滿情境的參與感和互動性
16	HI4	類人互動	我認為與 AI 互動時，感覺像是在與真人導覽員進行交談
17	SA1	滿意度	我對 AI 機器人的整體滿意度很高
18	SA2	滿意度	我對 AI 機器人的使用體驗感到開心
19	SA3	滿意度	我認為 AI 機器人在解答我的問題與執行任務時，表現達到預期
20	SA4	滿意度	我認為 AI 機器人在理解我的指令或問題時的準確性，令我感到滿意
21	UI1	使用意圖	我會推薦其他人使用這個 AI 機器人
22	UI2	使用意圖	我在未來導覽類似景點時，會繼續使用 AI 機器人服務
23	UI3	使用意圖	我未來會繼續使用 AI 機器人來完成學習與導覽



## 第五章、統計分析與研究假說結果

### 5.1 受測者族群分析

本研究採用非機率便利抽樣法來收集數據，因其在方便性、時間成本方面具有優勢。並使用 Tally Forms (<https://tally.so/>) 製作線上問卷，於導覽前後開放填寫。

#### 5.1.1 對照組

對象為報名參加每週日由臺南市政府觀光旅遊局推出的「米其林三星之旅」之「府城歷史散步 - 孔廟線」遊客。一梯次遊客最多 20 人。該導覽路線不僅包含孔廟，還涵蓋了孔廟周遭的景點，也因此雖然孔廟導覽的部分約 1 小時，但導覽總時長長達約 3 小時。本研究共參加六梯次，受測者為 104 人。族群分析可參考表 五-1。

表 五-1：對照組族群分析

類別	選項	占比	數量
出生性別	男	45%	47
	女	55%	57
年齡背景	10 歲以下	0%	0
	10 歲～19 歲	9%	9
	20 歲～29 歲	32%	34
	30 歲～39 歲	23%	24
	40 歲～49 歲	11%	11
	50 歲～59 歲	17%	18
	60 歲～69 歲	2%	2
	70 歲～79 歲	4%	4
	80 歲以上	2%	2
此景點遊玩經驗	0 次	31%	32
	1 次～3 次	54%	56
	4 次（含）以上	15%	16

### 5.1.2 實驗組

對象為隨機訪問臺南孔廟現場之遊客，同時間同設備的受測者約為 1 至 5 人小組，實驗時長約為 1 小時，調查時間為 2024/05/11(六)起至 2024/05/30(四)止，總共實驗 107 人。族群分析可參考表 五-2。

表 五-2：實驗組族群分析

類別	選項	占比	數量
出生性別	男	60%	64
	女	40%	43
年齡背景	10 歲以下	2%	2
	10 歲～19 歲	2%	2
	20 歲～29 歲	21%	22
	30 歲～39 歲	33%	35
	40 歲～49 歲	7%	8
	50 歲～59 歲	22%	24
	60 歲～69 歲	11%	12
	70 歲～79 歲	1%	1
	80 歲以上	1%	1
此景點遊玩經驗	0 次	40%	43
	1 次～3 次	50%	54
	4 次（含）以上	10%	10
行動導覽操作經驗 （如掃描 QR Code）	有	24%	26
	無	76%	81
平時有無使用 AI 工具 （如 Siri）	有	49%	52
	無	51%	55
有無接觸過大語言模型 （如掃描 ChatGPT）	有	47%	50
	無	53%	57

## 5.2 學習成效

為驗證受測者的學習成效，本研究共提出十題選擇題，滿分十分，由受測者於導覽前後進行前後測小考。結果確認參與者無論是由真人導覽員帶領或者是由實驗設備導覽，學習成效皆能有所提升。表 16 及表 17 分別展示了實驗組及對照組的參與人數（N）、前後測平均成績、成績標準差、平均分數成長及成長幅度。

### 5.2.1 實驗組

實驗組的受測者為 107 人，後測成績平均提升 2.58 分，成長幅度為 34.2%。平均成績及成長幅度大幅超過對照組的原因推測為，實驗組的人通常是特地來參觀孔廟，並且導覽時程約在一小時內結束，結束後馬上就能填小考，因此學習印象較深刻。實驗組學習成效可參照表 五-3。

表 五-3：實驗組學習成效

實驗組	N	成績標準差	平均成績	平均分數成長	平均成長幅度
前測成績	107	1.99	4.96	-	-
後測成績	107	2.14	7.54	+2.58	+34.2%

### 5.2.2 對照組

對照組的受測者為 93 人，後測成績平均提升 0.96 分，成長幅度為 16.2%。平均成績及成長幅度不如實驗組的原因推測為，對照組的人通常不是只來參加孔廟導覽，並且真人導覽整體時程較長，導致學習印象反倒不如實驗組只參加孔廟導覽來得深刻。對照組學習成效可參照表 五-4。

表 五-4：對照組學習成效

對照組	N	成績標準差	平均成績	平均分數成長	平均成長幅度
前測成績	93	2.31	5.01	-	-
後測成績	93	2.56	5.98	+0.97	+16.2%

## 5.3 量表

本研究為驗證本論文提出的研究模型 GAIUA，使用了 SmartPLS 4.0 軟體檢定假說[138]。並且，本研究使用了「偏最小平方法的結構方程模型（Partial Least Squares Structural Equation Modeling, PLS-SEM）」驗證「構面信度與效度」、「區別效度」、「因素負荷」與「共線性」。而在「路徑係數」上，本研究使用「拔靴法（Bootstrapping）」進行 5,000 次重覆抽樣，驗證「直接效果」和「間接效果」。

### 5.3.1 構面信度與效度

#### 5.3.1.1 實驗指標

本研究使用的信度、效度指標，包括「Cronbach's Alpha 值」、「組合信度」和「平均方差萃取量」，以下進行解釋：

- (1) 「Cronbach's Alpha 值（Alpha）」：此指標由 Cronbach (1951) 提出，是衡量問卷內部一致性的重要指標[139]。一般來說，若 Alpha 值大於 0.7，表示該構面的信度可被接受；若 Alpha 值在 0.8~0.9 之間，表示該構面的信度很好；若 Alpha 值大於 0.9，則顯示該構面的信度十分優秀；然而，若 Alpha 值小於 0.7，則顯示該構面的信度可能需要注意。
- (2) 「組合信度（Composite Reliability, CR）」：用來評估量表的內部一致性信度。CR 值的計算方式考慮了每個觀測變數的負荷量，比 Alpha 值更為精確。Fornell and Larcker (1981) 認為 CR 值大於 0.7 是可接受的信度標準[140]，而大於 0.8 則表示信度良好。
- (3) 「平均方差萃取量（Average Variance Extracted, AVE）」：此指標用來評估構面的聚合效度，即各題項對其構面的代表性。Bagozzi and Yi (1988) 建議 AVE 值應大於 0.5 表示該構面具有良好的聚合效度[141]，說明大部分的變異可以通過該構面來解釋。

### 5.3.1.2 實驗結果

本研究所有的構面 Alpha 值均超過 0.85 且 CR 值皆高於 0.9，這表明本問卷具有很高的內部一致性且題項間的相關性高，能夠準確反映每個構面所代表的潛在變數。再來，本研究所有的構面 AVE 值均高於 0.7，這表明每個構面的測量指標具有很強的代表性，也意味本問卷具有很高的聚合效度。

本實驗分析出的高信度、效度，顯示本問卷的設計和實驗數據的採集是可靠且有效的，能夠準確反映每位受測者的真實感受和行為意圖。構面信度與效度的完整結果可參考表 五-5。

表 五-5：構面信度與效度

構面	Alpha 值	CR 值	AVE 值
類人互動	0.911	0.937	0.789
感知人性	0.924	0.946	0.814
解決問題	0.884	0.921	0.744
回應內容	0.860	0.906	0.707
滿意度	0.937	0.955	0.840
使用意圖	0.896	0.935	0.828

## 5.3.2 區別效度

### 5.3.2.1 實驗指標

「區別效度」通常用來檢查不同構面之間是否具有顯著差異。此指標根據 Fornell and Larcker (1981) 所提出的準則來評估[140]，也就是每個構面的 AVE 平方根值應大於各構面之間的相關性係數，這樣才能證明構面具有良好的區別效度。

### 5.3.2.2 實驗結果

實驗結果表明，本研究中的對角線值皆大於對角線以外的值，可參考表 五-6，如表中對角線粗體字的部分，這表明本研究的所有構面皆判別有效。例如，類人互動的 AVE 平方根值是 0.888，顯著大於與其他構面的相關性係數（如與感知人性的

0.838)。並且本實驗大部分相關性係數均小於 0.85，這表明本研究模型的構面之間均有良好的區別效度。

表 五-6：區別效度

區別效度	類人互動	感知人性	解決問題	回應內容	滿意度	使用意圖
類人互動	<b>0.888</b>					
感知人性	0.838	<b>0.902</b>				
解決問題	0.844	0.773	<b>0.862</b>			
回應內容	0.628	0.663	0.799	<b>0.840</b>		
滿意度	0.805	0.755	0.799	0.677	<b>0.917</b>	
使用意圖	0.740	0.739	0.702	0.661	0.878	<b>0.910</b>

### 5.3.3 因素負荷量與共線性

#### 5.3.3.1 實驗指標

「因素負荷量」用來衡量每個觀測變數（題項）在其所屬構面中的解釋力，而「變異數膨脹因子」則用來檢查多重共線性的問題。以下進行解釋：

- (1) 「因素負荷量（Factor Loading）」：通常表示每個測量題項（變項）與所屬構面（潛在變數）之間的相關程度。Hair (2019) 認為若係數大於 0.7，表示該題項與構面之間具有較強的相關關係，並且該題項對構面具較強的解釋力，而這種相關性通常會被認為是顯著的；若係數低於 0.7，則表示該題項與構面之間的相關性較弱，並且這種相關性被認為是不顯著的[142]。
- (2) 「變異數膨脹因子（Variance Inflation Factor, VIF）」：用來檢測回歸分析中的多重共線性。它量化了題項和其他構面之間存在的共線性，而共線性將導致回歸係數的不穩定性。簡單來說，VIF 值代表某個題項與其他構面間的相關性，是否會讓回歸模型中該題項的影響力變得不可靠。根據 James et al (2013) 的定義，若 VIF 值=1：說明題項沒有共線性，即這個題項與其他構面之間沒有相關性，不會影響回歸結果的穩定性；若  $1 < \text{VIF 值} < 5$ ：則表明該題項存在中等程度的共線性，意味著題項和其他構面之間有一定的相關性，

但這種程度通常是可接受的；若 VIF 值  $> 5$ ：則代表題項存在高度共線性，說明題項和其他構面之間的相關性很強，這會讓回歸模型中的該題項的影響力變得不可靠，可能導致結果失真；若 VIF 值  $> 10$ ：則代表題項存在非常高的共線性，在這種情況下，題項和其他構面之間的相關性過高，會嚴重影響回歸模型的穩定性，通常會認為是需要調整的嚴重問題[143]。

### 5.3.3.2 實驗結果

- (1) 「解決問題 (Problem-Solving, PS)」：構面的因素負荷量均高於 0.8，這表示所有題項都能很好地解釋該構面。其中，題項「我認為 AI 在回覆我的問題時的表現符合我的期待」的因素負荷量最高，達到 0.910，顯示該題項對構面的貢獻最大。VIF 值方面，所有題項的 VIF 值均低於 5，表明不存在顯著的多重共線性問題。
- (2) 「感知人性 (Perceived Humanness, PH)」：構面的因素負荷量均高於 0.8，且大部分題項的因素負荷量接近或超過 0.9，特別是「我認為 AI 有能力在談話中顯示出類似人類的思考模式」的因素負荷量最高，為 0.924。這表明這些題項對該構面的解釋力很強。VIF 值雖然相對較高，但仍在可接受範圍內，表明多重共線性問題不大。
- (3) 「回應內容 (Response Content, RC)」：構面的因素負荷量均超過 0.7，其中題項「我認為 AI 提供的回答符合我的需求」和「我認為 AI 的回答內容充分且具體」的因素負荷量最高，均為 0.878。這顯示這兩個題項對構面的解釋力較強。VIF 值均低於 3，顯示多重共線性問題輕微。
- (4) 「類人互動 (Human-like Interaction, HI)」：構面的因素負荷量均高於 0.8，且大部分題項的因素負荷量接近或超過 0.9，尤其是「我認為與 AI 的交流，充滿情境的參與感和互動性」的因素負荷量最高，達到 0.926，表明該題項對構面的解釋力最強。VIF 值方面，部分題項的 VIF 值接近 4，但整體仍在可接受範圍內。

(5) 「滿意度 (Satisfaction, SA)」：構面的因素負荷量在 0.896 至 0.932 之間，顯示各題項對該構面的解釋力都很強，特別是「我對 AI 機器人的使用體驗感到開心」的因素負荷量最高，為 0.932。VIF 值則較高，但均在 5 以內，表明多重共線性問題不顯著。

(6) 「使用意圖 (Using Intention, UI)」：構面的因素負荷量均高於 0.8，其中「我在未來導覽類似景點時，會繼續使用 AI 機器人服務」的因素負荷量最高，達到 0.932，顯示該題項對構面的解釋力最強。VIF 值均低於 4，表明多重共線性問題不大。

綜上所述，所有構面的題項因素負荷量均超過 0.7，顯示每個題項對其所屬構面的解釋力都很強。同時，VIF 值均低於 5，表明多重共線性問題在本研究中並不顯著，這些結果進一步支持了本研究提出的研究模型具有良好的信度和效度。完整實驗結果可參考表 五-7。



表 五-7：因素負荷量與共線性

構面	代碼	題項	因素負荷量	VIF
解決問題	PS1	我認為 AI 能識別並回應我的需求	0.833	2.139
	PS2	我認為 AI 在回覆我的問題時的表現符合我的期待	0.910	3.378
	PS3	我認為 AI 能針對我的問題，給出完美的答案	0.895	3.081
	PS4	我認為 AI 在導覽過程中展現的豐富知識令人讚賞	0.808	1.991
感知人性	PH1	我認為 AI 具有類似人類對於問題理解的能力	0.885	2.634
	PH2	我認為 AI 能模仿人類進行對話交流的表現	0.890	2.929
	PH3	我認為 AI 有能力在談話中顯示出類似人類的思考模式	0.924	3.926
	PH4	我認為 AI 具有足夠的語言豐富度以模仿人類	0.910	3.641
回應內容	RC1	我認為 AI 提供的回答符合我的需求	0.878	2.759
	RC2	我認為 AI 的回答內容充分且具體	0.878	2.712
	RC3	我認為 AI 能夠提供我需求的實用資訊	0.832	2.080
	RC4	我認為 AI 在提供回答時展示出了對於當前詢問話題的理解	0.770	1.498
類人互動	HI1	我認為 AI 能理解並適應我們的交談	0.855	2.748
	HI2	我認為 AI 能迅速且恰當地回應我的問題	0.895	3.129
	HI3	我認為與 AI 的交流，充滿情境的參與感和互動性	0.926	4.277
	HI4	我認為與 AI 互動時，感覺像是在與真人導覽員進行交談	0.875	3.399
滿意度	SA1	我對 AI 機器人的整體滿意度很高	0.909	3.793
	SA2	我對 AI 機器人的使用體驗感到開心	0.932	4.436
	SA3	我認為 AI 機器人在解答我的問題與執行任務時，表現達到預期	0.930	4.372
	SA4	我認為 AI 機器人在理解我的指令或問題時的準確性，令我感到滿意	0.896	3.451
使用意圖	UI1	我會推薦其他人使用這個 AI 機器人	0.909	3.074
	UI2	我在未來導覽類似景點時，會繼續使用 AI 機器人服務	0.932	3.540
	UI3	我未來會繼續使用 AI 機器人來完成學習與導覽	0.889	2.273

### 5.3.4 路徑係數

#### 5.3.4.1 實驗指標

本研究使用以下幾個指標來評估模型中的路徑關係：「標準差」、「路徑係數」、「T 統計量」和「機率值」。以下進行解釋：

- (1) 「標準差 (Standard Deviation, SD)」：衡量數據的離散程度。在模型中，它反映了估計係數的不確定性。
- (2) 「路徑係數 (Path Coefficient)」，又稱「標準化迴歸係數 (Standardized Regression Coefficients,  $\beta$ )」：表示變數之間的影响強度，數值範圍在-1 到 1 之間，絕對值越大表示影响越強。
- (3) 「T 統計量 (T-statistic, T)」：檢驗  $\beta$  值是否顯著，通常 T 值大於 1.96 (對應於 95% 的置信水平) 表示  $\beta$  值顯著。
- (4) 「機率值 (Probability Value, P)」：顯著性水平，通常 P 值小於 0.05 表示該路徑的影响顯著。

圖 五-1 呈現各路線的  $\beta$  值。GAIUA 使用的構面主要有六個：解決問題 (Problem-Solving, PS)、感知人性 (Perceived Humanness, PH)、回應內容 (Response Content, RC)、類人互動 (Human-like Interaction, HI)、滿意度 (Satisfaction, SA)、使用意圖 (Using Intention, UI) 等。

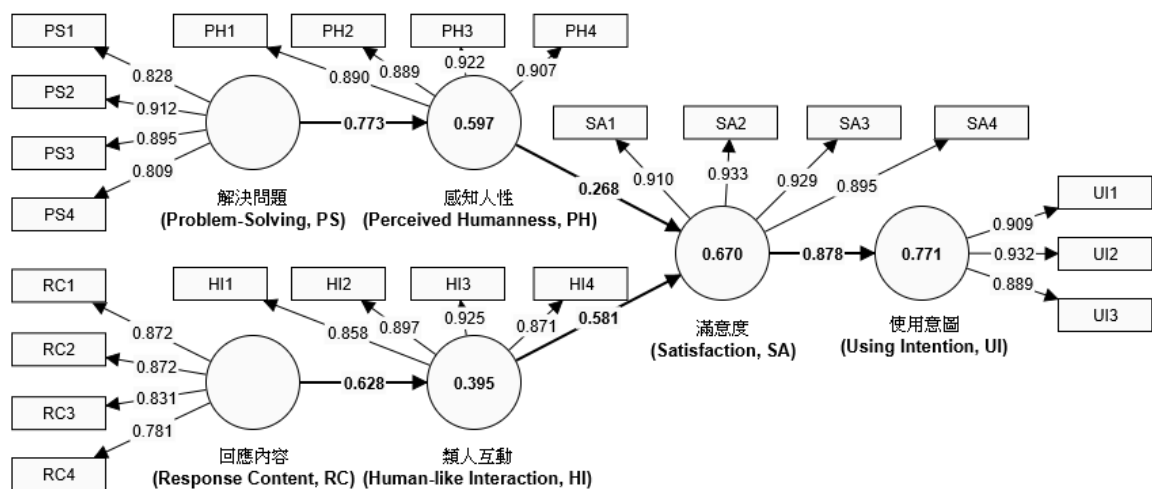


圖 五-1：路徑係數

#### 5.3.4.2 直接效果

直接效果探討各路徑假說的標準差、 $\beta$ 值、T值和P值的表現，以確認假說是否成立。

- (1) H1「解決問題→感知人性」：SD為0.047， $\beta$ 值為0.773，T值為16.434，P值為0.000。這些結果顯示「解決問題」對「感知人性」有顯著且強烈的正向影響。SD較小，表示數據分佈較集中， $\beta$ 值顯示影響強度很高，T值遠大於1.96，表明 $\beta$ 值在統計上極其顯著，P值為0.000，更進一步證實了該路徑的顯著性。因此，「解決問題→感知人性」假說成立。
- (2) H2「感知人性→滿意度」：SD為0.115， $\beta$ 值為0.268，T值為2.331，P值為0.010。「感知人性」對「滿意度」有顯著的正向影響，但影響程度相對較低。SD稍大，表明數據分佈略有分散， $\beta$ 值顯示影響中等強度，T值大於1.96，表示影響顯著，P值為0.010，證明該路徑在統計上具有顯著性。因此，假說H2成立。
- (3) H3「回應內容→類人互動」：SD為0.066， $\beta$ 值為0.628，T值為9.496，P值為0.000。這表明「回應內容」對「類人互動」有顯著且較強的正向影響。SD適中，數據分佈較集中， $\beta$ 值顯示影響強度較高，T值遠大於1.96，表示 $\beta$ 值在統計上極其顯著，P值為0.000，進一步證實了該路徑的顯著性。因此，「回應內容→類人互動」假說成立。
- (4) H4「類人互動→滿意度」：SD為0.104， $\beta$ 值為0.581，T值為5.591，P值為0.000。這說明「類人互動」對「滿意度」有顯著且強烈的正向影響。SD較大，表示數據分佈較分散， $\beta$ 值顯示影響強度較高，T值大於1.96，表明 $\beta$ 值影響顯著，P值為0.000，證明該路徑在統計上具有顯著性。因此，「類人互動→滿意度」假說成立。
- (5) H5「滿意度→使用意圖」：SD為0.025， $\beta$ 值為0.878，T值為35.698，P值為0.000。這表明「滿意度」對「使用意圖」有非常強烈且顯著的正向影響。SD極小，數據分佈非常集中， $\beta$ 值顯示影響強度極高，T值遠大於1.96，表

示  $\beta$  值在統計上極其顯著，P 值為 0.000，進一步證實了該路徑的顯著性。因此，「滿意度→使用意圖」假說成立。

總結來說，所有假設路徑的  $\beta$  值均顯示出正向影響，且 P 值均小於 0.05，表示所有假設均得到顯著支持。這些結果進一步驗證了本研究的模型，表明各構面之間的關係具有高度的顯著性和影響力，支持了本研究提出的研究模型，也就是說本研究提出的假說全數成立。完整結果可參考表 五-8。

表 五-8：直接效果

代碼	路徑	SD	$\beta$ 值	T 值	P 值	假說
H1	解決問題→感知人性	0.047	0.773	16.434	0.000	成立
H2	感知人性→滿意度	0.115	0.268	2.331	0.010	成立
H3	回應內容→類人互動	0.066	0.628	9.496	0.000	成立
H4	類人互動→滿意度	0.104	0.581	5.591	0.000	成立
H5	滿意度→使用意圖	0.025	0.878	35.698	0.000	成立

#### 5.3.4.3 間接效果

間節效果探討各路徑對「滿意度」和「使用意圖」的間接影響程度，這些結果可以協助理解中介變數在解釋構面之間關係中的作用。

- (1) 「解決問題→感知人性→滿意度」的路徑：SD 為 0.094， $\beta$  值為 0.207，T 值為 2.208，P 值為 0.014，這表明「解決問題」通過「感知人性」對「滿意度」有顯著的間接影響。T 值大於 1.96，P 值小於 0.05，顯示該中介效應的顯著性。
- (2) 「解決問題→感知人性→滿意度→使用意圖」的路徑：SD 為 0.084， $\beta$  值為 0.182，T 值為 2.164，P 值為 0.015。這表明解決問題通過「感知人性」和「滿意度」對「使用意圖」有顯著的間接影響。T 值大於 1.96，P 值小於 0.05，顯示該中介效應的顯著性。
- (3) 「感知人性→滿意度→使用意圖」的路徑：SD 為 0.103， $\beta$  值為 0.235，T 值為 2.294，P 值為 0.011。這顯示「感知人性」通過「滿意度」對「使用意圖」

有顯著但相對較弱的間接影響。T 值大於 1.96，P 值小於 0.05，顯示該中介效應顯著。

(4) 「回應內容→類人互動→滿意度」的路徑：SD 為 0.076， $\beta$  值為 0.365，T 值為 4.819，P 值為 0.000，這顯示「回應內容」通過「類人互動」對「滿意度」有顯著且較強的間接影響。T 值大於 1.96，P 值小於 0.05，顯示該中介效應的顯著性非常高。

(5) 「回應內容→類人互動→滿意度→使用意圖」的路徑：SD 為 0.068， $\beta$  值為 0.320，T 值為 4.704，P 值為 0.000，這表明「回應內容」通過「類人互動」和「滿意度」對使用意圖有顯著的間接影響。T 值大於 1.96，P 值小於 0.05，顯示該中介效應的顯著性非常高。

(6) 「類人互動→滿意度→使用意圖」的路徑：顯示出 SD 為 0.091， $\beta$  值為 0.510，T 值為 5.621，P 值為 0.000。這表明「類人互動」通過「滿意度」對「使用意圖」有顯著且較強的間接影響。T 值大於 1.96，P 值小於 0.05，顯示該中介效應顯著性極高。

綜上所述，所有間接路徑的  $\beta$  值均顯示出正向影響，且 P 值均小於 0.05，表明這些中介效應均得到顯著支持。這些結果進一步證實了中介變數（如滿意度和類人互動）在解釋構面之間關係中的重要性，並支持了本研究的研究模型。完整結果可參考表 五-9。

表 五-9：間接效果

序號	路徑	SD	$\beta$ 值	T 值	P 值
(1)	解決問題→感知人性→滿意度	0.094	0.207	2.208	0.014
(2)	解決問題→感知人性→滿意度→使用意圖	0.084	0.182	2.164	0.015
(3)	感知人性→滿意度→使用意圖	0.103	0.235	2.294	0.011
(4)	回應內容→類人互動→滿意度	0.076	0.365	4.819	0.000
(5)	回應內容→類人互動→滿意度→使用意圖	0.068	0.320	4.704	0.000
(6)	類人互動→滿意度→使用意圖	0.091	0.510	5.621	0.000

#### 5.3.4.4 總和效果

在總和效果分析中，除了考慮直接效果外，我們還需要考慮間接效果，這樣才能全面了解各構面之間的總影響。本研究將針對 GAIUA 中，最重要的兩個構面「滿意度」與「使用意圖」的總和效果進行詳細探討：

##### (1) 「滿意度」的總和效果分析：

##### 1. 「感知人性」對「滿意度」的影響：

- a. 直接效果：SD 為 0.115， $\beta$  值為 0.268，T 值為 2.331，P 值為 0.010，顯示顯著的正向影響。
- b. 間接效果：透過「解決問題→感知人性→滿意度」的 SD 為 0.094， $\beta$  值為 0.207，T 值為 2.208，P 值為 0.014。
- c. 總和效果：0.475 (0.268 + 0.207)，顯示「感知人性」對「滿意度」的總影響顯著。

##### 2. 「類人互動」對「滿意度」的影響：

- a. 直接效果：SD 為 0.104， $\beta$  值為 0.581，T 值為 5.591，P 值為 0.000，顯示顯著的正向影響。
- b. 間接效果：透過「回應內容→類人互動→滿意度」的 SD 為 0.076， $\beta$  值為 0.365，T 值為 4.819，P 值為 0.000。
- c. 總和效果：0.946 (0.581 + 0.365)，顯示「類人互動」對「滿意度」的總影響非常顯著。

##### 3. 小結：

根據以上分析，「感知人性」和「類人互動」對於「滿意度」均有顯著的正向影響。其中，「類人互動」的總效果 (0.946) 遠高於「感知人性」 (0.475)，顯示在提升使用者滿意度方面，「類人互動」的影響更為顯著。「滿意度」的總和效果清單可參考表 五-10。

表 五-10：「滿意度」的總和效果

效果	路徑	SD	T 值	P 值	$\beta$ 值
直接效果	感知人性→滿意度	0.115	2.331	0.010	0.268
	類人互動→滿意度	0.104	5.591	0.000	0.581
間接效果	解決問題→感知人性→滿意度	0.094	2.208	0.014	0.207
	回應內容→類人互動→滿意度	0.076	4.819	0.000	0.365

(2) 「使用意圖」的總和效果分析：

1. 「滿意度」對「使用意圖」的影響：
  - a. 直接效果：SD 為 0.025， $\beta$  值為 0.878，T 值為 35.698，P 值為 0.000，顯示顯著的正向影響。
  - b. 總和效果：0.878，「滿意度」對「使用意圖」有非常強烈且顯著的正面影響。
2. 「感知人性」對「使用意圖」的間接影響：
  - a. 間接效果 1：透過「感知人性→滿意度→使用意圖」的 SD 為 0.103， $\beta$  值為 0.235，T 值為 2.294，P 值為 0.011。
  - b. 間接效果 2：透過「解決問題→感知人性→滿意度→使用意圖」的 SD 為 0.084， $\beta$  值為 0.182，T 值為 2.164，P 值為 0.015。
  - c. 總和效果：0.417 (0.235 + 0.182)，顯示「感知人性」通過多條路徑對「使用意圖」的間接影響顯著。
3. 「類人互動」對「使用意圖」的間接影響：
  - a. 間接效果 1：透過「類人互動→滿意度→使用意圖」的 SD 為 0.091， $\beta$  值為 0.510，T 值為 5.621，P 值為 0.000。
  - b. 間接效果 2：透過「回應內容→類人互動→滿意度→使用意圖」的 SD 為 0.068， $\beta$  值為 0.320，T 值為 4.704，P 值為 0.000。
  - c. 總和效果：0.830 (0.510 + 0.320)，顯示「類人互動」通過多條路徑對「使用意圖」的間接影響非常顯著。

#### 4. 小結：

根據以上分析，「滿意度」對於「使用意圖」有非常強烈且顯著的正向影響，總效果為 0.878。「感知人性」和「類人互動」對於「使用意圖」的間接影響也顯著，其中「類人互動」的總效果（0.830）高於「感知人性」（0.417），顯示在影響使用意圖方面，「類人互動」的影響更為顯著。「使用意圖」的總和效果清單可參考表 五-11。

表 五-11：「使用意圖」的總和效果

效果	路徑	SD	T 值	P 值	$\beta$ 值
直接效果	滿意度→使用意圖	0.025	35.698	0.000	0.878
間接效果	感知人性→滿意度→使用意圖	0.103	2.294	0.011	0.235
	解決問題→感知人性→滿意度→使用意圖	0.084	2.164	0.015	0.182
	類人互動→滿意度→使用意圖	0.091	5.621	0.000	0.510
	回應內容→類人互動→滿意度→使用意圖	0.068	4.704	0.000	0.320



## 第六章、研究貢獻與未來展望

### 6.1 研究貢獻

本研究主要的研究貢獻可分為三個面向：

- (1) 本研究提出了一套新穎的、低能耗的、基於嵌入式 AI 的 MLLM 框架，可於 SBC 上離線執行。有望在現今訓練及運行成本都大漲的大模型時代，提供另一種不同的框架思考，並可望帶給只有一般資源的用戶，一條較低成本的 MLLM 解決方案。同時，在實驗過程中，本研究再度確認了此套 MLLM 框架的微調/訓練成本十分低廉，包含 RAG 的預處理以及 YOLO 的訓練，皆於五分鐘內完成。也就是說，只要資料齊全，極易於快速部署至不同的場域，而不需經歷傳統較耗時且耗能源的大模型微調。此外，本研究將設備投入在臺南孔廟中實地使用，驗證了此 MLLM 框架實際意義上的可用性。也確認了於實際場域應用時，可能會遇到的研究限制，可供後續的研究者參考。本研究相信，在邁向 AGI 的道路上，本研究探索出了一套新穎且與眾不同的模式，為後人指明了一條不同的道路。
- (2) 本研究驗證了此 GAI 系統應用中文語音互動於行動學習與數位導覽領域，所能帶來的學習成效與使用體驗。並且，就學習成效的層面還與真人導覽員進行對比。實驗證明，本系統對於學習成效的提升相較於真人導覽員多提升了 18.0%。為未來使用 GAI 進行教育創新的研究者們，提供了一個寶貴的經驗。
- (3) 本研究提出了一套創新的 GAIUA 研究模型，用來探討 GAI 的使用意圖，總共使用了六個構面。包括，解決問題、感知人性、回應內容、類人互動、滿意度、使用意圖。本研究不僅探討了這六個構面及提出了在其之間的假說，還在後續的實驗中證實了 GAIUA 的所有假說皆是成立的。研究結果顯示，「類人互動」對於提升使用者滿意度和使用意圖的影響最為顯著。本研究相信，隨著 GAI 的不斷發展，將來這套研究模型還能應用在測量各種不同類型的新 GAI 服務。

## 6.2 研究限制

本研究建置的系統於現場實驗時，碰到以下限制：

- (1) 反應速度慢：程式啟動速度約 43 秒（其中載入 LLM 約 27 秒）、物件簡介與互動等需要 LLM 查詢的行動約 20 秒。
- (2) 口音問題：Linux 內建的 eSpeak 對於中文的支援不友善，導致受測者時常聽不清楚系統的簡介與回應。
- (3) 收訊問題：Whisper 對於雜訊相當敏感，時常收音錯誤，推測是由於本研究使用的模型級別，因記憶體占用考量，只使用「Base」級的緣故（Whisper 依照模型參數量共分為：Tiny、Base、Small、Medium、Large）。並且由於臺南孔廟是熱門觀光景點，遊客數量眾多，實驗時只能請受測者不用麥克風時將其關閉，待需要互動時再打開。
- (4) 記憶體不足：由於 Jetson 單板機為 CPU-GPU 共享記憶體。單板機本身的 OS 也會占用記憶體，並且本系統同時使用 LLM、YOLO、Whisper 待命，導致 OOM(Out of Memory)的情況時常發生，使得受測者常得重新等待單板機重啟程式，間接影響使用體驗。程式待命時的記憶體占用，可參考圖 六-1。與 LLM 互動時的記憶體占用，可參考圖 六-2。

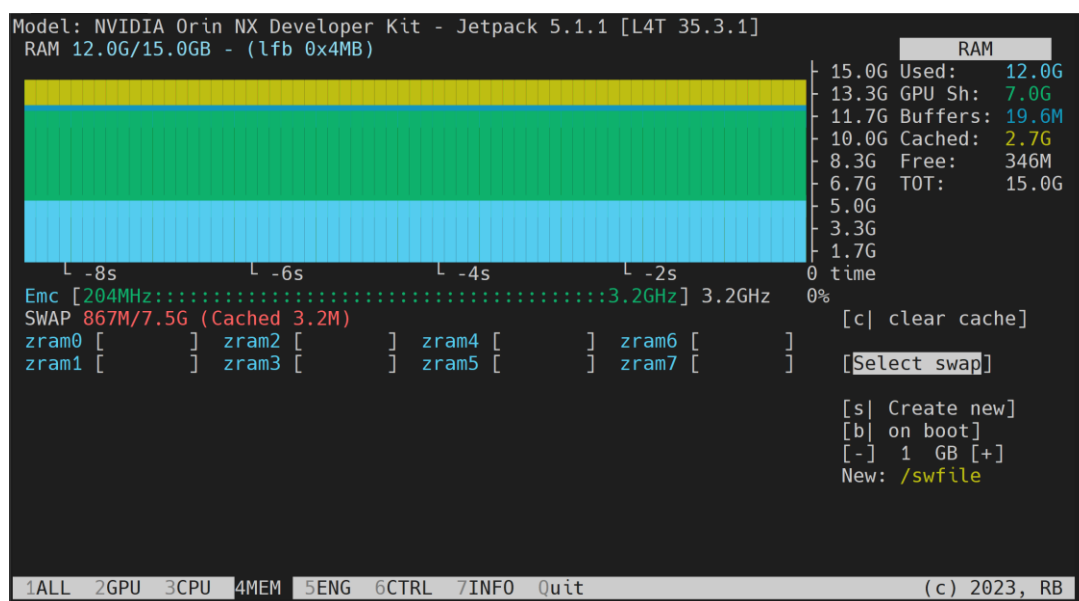


圖 六-1：程式待命時的記憶體占用

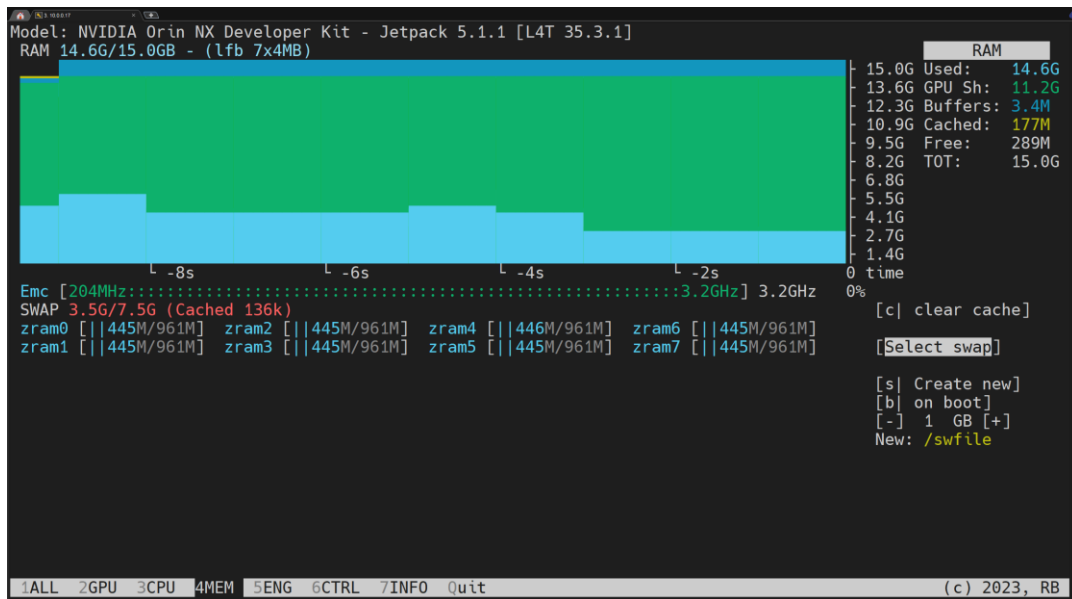


圖 六-2：與 LLM 互動時的記憶體占用

## 6.3 未來展望

- (1) 本研究提出了一套創新的 MLLM 框架，期望能達成了第一步拋磚引玉，但同時於實驗時也發現了許多系統整合性的問題待解決。冀望能透過未來軟/硬體的持續進步，使這些問題能迎刃而解
- (2) 根據許多受測者反應，皆希望若要進行行動學習與數位導覽，最棒的使用體驗應在智慧型手機上。主因為受測者們覺得實驗設備的體積太過龐大且需額外攜帶，而不是使用平常熟悉的裝置，導致使用體驗受影響。期許未來的研究者也許能以智慧型手機為方向進行研究。

## 參考文獻

- [1] N. Fei *et al.*, "Towards artificial general intelligence via a multimodal foundation model," *Nature Communications*, vol. 13, no. 1, p. 3094, 2022.
- [2] T. Wu *et al.*, "A brief overview of ChatGPT: The history, status quo and potential future development," *IEEE/CAA Journal of Automatica Sinica*, vol. 10, no. 5, pp. 1122-1136, 2023.
- [3] I. UNESCO, "Supporting teachers in back-to-school efforts. Guidance for policy-makers. International task force on teachers for education 2030; 2020," ed, 2021.
- [4] K. Masters, "Ethical use of artificial intelligence in health professions education: AMEE Guide No. 158," *Medical Teacher*, vol. 45, no. 6, pp. 574-584, 2023.
- [5] S. Grassini, "Shaping the future of education: exploring the potential and consequences of AI and ChatGPT in educational settings," *Education Sciences*, vol. 13, no. 7, p. 692, 2023.
- [6] J. Achiam *et al.*, "Gpt-4 technical report," *arXiv preprint arXiv:2303.08774*, 2023.
- [7] S. Yin *et al.*, "A survey on multimodal large language models," *arXiv preprint arXiv:2306.13549*, 2023.
- [8] J. Li, D. Li, S. Savarese, and S. Hoi, "Blip-2: Bootstrapping language-image pre-training with frozen image encoders and large language models," in *International conference on machine learning*, 2023: PMLR, pp. 19730-19742.
- [9] H. Liu, C. Li, Q. Wu, and Y. J. Lee, "Visual instruction tuning," *Advances in neural information processing systems*, vol. 36, 2024.
- [10] S. Huang *et al.*, "Language is not all you need: Aligning perception with language models," *Advances in Neural Information Processing Systems*, vol. 36, 2024.
- [11] D. Zhu, J. Chen, X. Shen, X. Li, and M. Elhoseiny, "Minigpt-4: Enhancing vision-language understanding with advanced large language models," *arXiv preprint arXiv:2304.10592*, 2023.
- [12] G. Luo, Y. Zhou, T. Ren, S. Chen, X. Sun, and R. Ji, "Cheap and quick: Efficient vision-language instruction tuning for large language models," *Advances in Neural Information Processing Systems*, vol. 36, 2024.
- [13] Y.-C. Wang, J. Xue, C. Wei, and C.-C. J. Kuo, "An overview on generative ai at scale with edge-cloud computing," *IEEE Open Journal of the Communications Society*, 2023.
- [14] J. Wu, W. Gan, Z. Chen, S. Wan, and H. Lin, "Ai-generated content (aigc): A survey," *arXiv preprint arXiv:2304.06632*, 2023.
- [15] T. Brown *et al.*, "Language models are few-shot learners," *Advances in neural information processing systems*, vol. 33, pp. 1877-1901, 2020.

- [16] H. Fassold, "Porting Large Language Models to Mobile Devices for Question Answering," *arXiv preprint arXiv:2404.15851*, 2024.
- [17] S. Mittal, "A survey on optimized implementation of deep learning models on the nvidia jetson platform," *Journal of Systems Architecture*, vol. 97, pp. 428-442, 2019.
- [18] S. Alyamkin *et al.*, "2018 low-power image recognition challenge," *arXiv preprint arXiv:1810.01732*, 2018.
- [19] Y. Shi, K. Yang, Z. Yang, and Y. Zhou, "Mobile edge artificial intelligence: Opportunities and challenges," 2021.
- [20] G. Team *et al.*, "Gemma: Open models based on gemini research and technology," *arXiv preprint arXiv:2403.08295*, 2024.
- [21] E. Almazrouei *et al.*, "The falcon series of open language models," *arXiv preprint arXiv:2311.16867*, 2023.
- [22] T. Dettmers, A. Pagnoni, A. Holtzman, and L. Zettlemoyer, "Qlora: Efficient finetuning of quantized llms," *Advances in Neural Information Processing Systems*, vol. 36, 2024.
- [23] S. Ma *et al.*, "The Era of 1-bit LLMs: All Large Language Models are in 1.58 Bits," *arXiv preprint arXiv:2402.17764*, 2024.
- [24] M. Fullan, C. Azorín, A. Harris, and M. Jones, "Artificial intelligence and school leadership: challenges, opportunities and implications," *School Leadership & Management*, pp. 1-8, 2023.
- [25] G. Cooper, "Examining science education in ChatGPT: An exploratory study of generative artificial intelligence," *Journal of Science Education and Technology*, vol. 32, no. 3, pp. 444-452, 2023.
- [26] V. Mazzia, A. Khaliq, F. Salvetti, and M. Chiaberge, "Real-time apple detection system using embedded systems with hardware accelerators: An edge AI application," *IEEE Access*, vol. 8, pp. 9102-9114, 2020.
- [27] Z. Zou, K. Chen, Z. Shi, Y. Guo, and J. Ye, "Object detection in 20 years: A survey," *Proceedings of the IEEE*, vol. 111, no. 3, pp. 257-276, 2023.
- [28] J. Terven, D.-M. Córdova-Esparza, and J.-A. Romero-González, "A comprehensive review of yolo architectures in computer vision: From yolov1 to yolov8 and yolo-nas," *Machine Learning and Knowledge Extraction*, vol. 5, no. 4, pp. 1680-1716, 2023.
- [29] Y. Gao *et al.*, "Retrieval-augmented generation for large language models: A survey," *arXiv preprint arXiv:2312.10997*, 2023.
- [30] A. Gupta *et al.*, "RAG vs Fine-tuning: Pipelines, Tradeoffs, and a Case Study on Agriculture," *arXiv preprint arXiv:2401.08406*, 2024.
- [31] C.-J. Hsu, C.-L. Liu, F.-T. Liao, P.-C. Hsu, Y.-C. Chen, and D.-S. Shiu, "Breeze-7B Technical Report," *arXiv preprint arXiv:2403.02712*, 2024.
- [32] A. Q. Jiang *et al.*, "Mistral 7B," *arXiv preprint arXiv:2310.06825*, 2023.

- [33] A. Radford, J. W. Kim, T. Xu, G. Brockman, C. McLeavey, and I. Sutskever, "Robust speech recognition via large-scale weak supervision," in *International Conference on Machine Learning*, 2023: PMLR, pp. 28492-28518.
- [34] Y.-C. Hsieh, K.-m. Lyu, and R.-Y. Lyu, "Taiwanese/Mandarin Speech Recognition using OpenAI's Whisper Multilingual Speech Recognition Engine Based on Generative Pretrained Transformer Architecture," in *Proceedings of the 35th Conference on Computational Linguistics and Speech Processing (ROCLING 2023)*, 2023, pp. 210-214.
- [35] Y. Yang, J. Luo, and T. Lan, "An empirical assessment of a modified artificially intelligent device use acceptance model—From the task-oriented perspective," *Frontiers in Psychology*, vol. 13, p. 975307, 2022.
- [36] A. Lommatzsch, "A next generation chatbot-framework for the public administration," in *Innovations for Community Services: 18th International Conference, I4CS 2018, Žilina, Slovakia, June 18-20, 2018, Proceedings*, 2018: Springer, pp. 127-141.
- [37] N. Castelo, M. W. Bos, and D. R. Lehmann, "Task-dependent algorithm aversion," *Journal of Marketing Research*, vol. 56, no. 5, pp. 809-825, 2019.
- [38] B. J. Dietvorst, J. P. Simmons, and C. Massey, "Overcoming algorithm aversion: People will use imperfect algorithms if they can (even slightly) modify them," *Management science*, vol. 64, no. 3, pp. 1155-1170, 2018.
- [39] C. Longoni, A. Bonezzi, and C. K. Morewedge, "Resistance to medical artificial intelligence," *Journal of Consumer Research*, vol. 46, no. 4, pp. 629-650, 2019.
- [40] I. P. Tussyadiah, F. J. Zach, and J. Wang, "Do travelers trust intelligent service robots?," *Annals of Tourism Research*, vol. 81, p. 102886, 2020.
- [41] D. Gursoy, O. H. Chi, L. Lu, and R. Nunkoo, "Consumers acceptance of artificially intelligent (AI) device use in service delivery," *International Journal of Information Management*, vol. 49, pp. 157-169, 2019.
- [42] M. I. Yao, "Re-imagining the 'Taiwanese' nation in the interpretation of the Chinese-oriented heritage," *Nations and Nationalism*, vol. 28, no. 4, pp. 1230-1248, 2022.
- [43] H. Hua, Y. Li, T. Wang, N. Dong, W. Li, and J. Cao, "Edge computing with artificial intelligence: A machine learning perspective," *ACM Computing Surveys*, vol. 55, no. 9, pp. 1-35, 2023.
- [44] W. Shi and S. Dustdar, "The promise of edge computing," *Computer*, vol. 49, no. 5, pp. 78-81, 2016.
- [45] M. Sajid and Z. Raza, "Cloud computing: Issues & challenges," in *International conference on cloud, big data and trust*, 2013, vol. 20, no. 13: sn, pp. 13-15.
- [46] Z. Xu and Y. Tian, "The development history and application of cloud computing," *Inform Recording Mater*, vol. 19, no. 8, pp. 66-67, 2018.
- [47] H. Cai, B. Xu, L. Jiang, and A. V. Vasilakos, "IoT-based big data storage systems in cloud computing: perspectives and challenges," *IEEE Internet of Things Journal*, vol. 4, no. 1, pp. 75-87, 2016.

- [48] S. Wan, S. Ding, and C. Chen, "Edge computing enabled video segmentation for real-time traffic monitoring in internet of vehicles," *Pattern Recognition*, vol. 121, p. 108146, 2022.
- [49] M. Satyanarayanan, "The emergence of edge computing," *Computer*, vol. 50, no. 1, pp. 30-39, 2017.
- [50] A. M. Ghosh and K. Grolinger, "Edge-cloud computing for Internet of Things data analytics: Embedding intelligence in the edge with deep learning," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 3, pp. 2191-2200, 2020.
- [51] Y. Shi, K. Yang, T. Jiang, J. Zhang, and K. B. Letaief, "Communication-efficient edge AI: Algorithms and systems," *IEEE Communications Surveys & Tutorials*, vol. 22, no. 4, pp. 2167-2191, 2020.
- [52] W. Li, H. Hacid, E. Almazrouei, and M. Debbah, "A comprehensive review and a taxonomy of edge machine learning: Requirements, paradigms, and techniques," *AI*, vol. 4, no. 3, pp. 729-786, 2023.
- [53] I. C. Society, "IEEE Computer Society's Top 12 Technology Trends for 2020," ed: IEEE Computer Society Washington, DC, USA, 2019.
- [54] Y. Wang, C. Liu, and Y.-F. Tu, "Factors affecting the adoption of AI-based applications in higher education," *Educational Technology & Society*, vol. 24, no. 3, pp. 116-129, 2021.
- [55] Z. Zhang and J. Li, "A review of artificial intelligence in embedded systems," *Micromachines*, vol. 14, no. 5, p. 897, 2023.
- [56] L. Dutta and S. Bharali, "Tinymml meets iot: A comprehensive survey," *Internet of Things*, vol. 16, p. 100461, 2021.
- [57] C. R. Banbury *et al.*, "Benchmarking tinymml systems: Challenges and direction," *arXiv preprint arXiv:2003.04821*, 2020.
- [58] H. A. Imran, U. Mujahid, S. Wazir, U. Latif, and K. Mehmood, "Embedded development boards for edge-AI: A comprehensive report," *arXiv preprint arXiv:2009.00803*, 2020.
- [59] A. Garcia-Perez, R. Miñón, A. I. Torre-Bastida, and E. Zulueta-Guerrero, "Analysing Edge Computing Devices for the Deployment of Embedded AI," *Sensors*, vol. 23, no. 23, p. 9495, 2023.
- [60] S. Mittal and J. S. Vetter, "A survey of CPU-GPU heterogeneous computing techniques," *ACM Computing Surveys (CSUR)*, vol. 47, no. 4, pp. 1-35, 2015.
- [61] F. Dou *et al.*, "Towards artificial general intelligence (agi) in the internet of things (iot): Opportunities and challenges," *arXiv preprint arXiv:2309.07438*, 2023.
- [62] C. Zhang *et al.*, "One small step for generative ai, one giant leap for agi: A complete survey on chatgpt in aigc era," *arXiv preprint arXiv:2304.06488*, 2023.
- [63] R. Bommasani *et al.*, "On the opportunities and risks of foundation models," *arXiv preprint arXiv:2108.07258*, 2021.

- [64] L. Ouyang *et al.*, "Training language models to follow instructions with human feedback," *Advances in neural information processing systems*, vol. 35, pp. 27730-27744, 2022.
- [65] Y. Li, Y. Du, K. Zhou, J. Wang, W. X. Zhao, and J.-R. Wen, "Evaluating object hallucination in large vision-language models," *arXiv preprint arXiv:2305.10355*, 2023.
- [66] J. Terven and D. Cordova-Esparza, "A comprehensive review of YOLO: From YOLOv1 to YOLOv8 and beyond," *arXiv preprint arXiv:2304.00501*, 2023.
- [67] M.-J. Tsai and S.-L. Peng, "QR code beautification by instance segmentation (IS-QR)," *Digital Signal Processing*, vol. 133, p. 103887, 2023.
- [68] A. Wang *et al.*, "YOLOv10: Real-Time End-to-End Object Detection," *arXiv preprint arXiv:2405.14458*, 2024.
- [69] P. Hu, Y. Gong, Y. Lu, and A. W. Ding, "Speaking vs. listening? Balance conversation attributes of voice assistants for better voice marketing," *International Journal of Research in Marketing*, vol. 40, no. 1, pp. 109-127, 2023.
- [70] S. K. Scott, "From speech and talkers to the social world: The neural processing of human spoken language," *Science*, vol. 366, no. 6461, pp. 58-62, 2019.
- [71] C. S. Oh, J. N. Bailenson, and G. F. Welch, "A systematic review of social presence: Definition, antecedents, and implications," *Frontiers in Robotics and AI*, vol. 5, p. 409295, 2018.
- [72] D. Grewal, S. M. Noble, A. L. Roggeveen, and J. Nordfalt, "The future of in-store technology," *Journal of the Academy of Marketing Science*, vol. 48, pp. 96-113, 2020.
- [73] A. S. Luccioni, Y. Jernite, and E. Strubell, "Power hungry processing: Watts driving the cost of ai deployment?," *arXiv preprint arXiv:2311.16863*, 2023.
- [74] B. Rouhani *et al.*, "A Microsoft custom data type for efficient inference," 2020.
- [75] S. Roy, "Understanding the Impact of Post-Training Quantization on Large-scale Language Models," *arXiv preprint arXiv:2309.05210*, 2023.
- [76] J. Lin, J. Tang, H. Tang, S. Yang, X. Dang, and S. Han, "Awq: Activation-aware weight quantization for llm compression and acceleration," *arXiv preprint arXiv:2306.00978*, 2023.
- [77] E. Frantar, S. Ashkboos, T. Hoefler, and D. Alistarh, "Gptq: Accurate post-training quantization for generative pre-trained transformers," *arXiv preprint arXiv:2210.17323*, 2022.
- [78] K. VM, H. Warriar, and Y. Gupta, "Fine Tuning LLM for Enterprise: Practical Guidelines and Recommendations," *arXiv preprint arXiv:2404.10779*, 2024.
- [79] Y. Zhang *et al.*, "Siren's song in the AI ocean: a survey on hallucination in large language models," *arXiv preprint arXiv:2309.01219*, 2023.
- [80] M. A. Habib, S. Amin, M. Oqba, S. Jaipal, M. J. Khan, and A. Samad, "TaxTajweez: A Large Language Model-based Chatbot for Income Tax Information In Pakistan Using Retrieval Augmented Generation (RAG)," in *The International FLAIRS Conference Proceedings*, 2024, vol. 37.



- [81] F. Cuconasu *et al.*, "The power of noise: Redefining retrieval for rag systems," *arXiv preprint arXiv:2401.14887*, 2024.
- [82] C. Wang, X. Chen, T. Yu, Y. Liu, and Y. Jing, "Education reform and change driven by digital technology: a bibliometric study from a global perspective," *Humanities and Social Sciences Communications*, vol. 11, no. 1, pp. 1-17, 2024.
- [83] A. Bryan *et al.*, "NMC Horizon Report: 2019 Higher Education Edition. louisville," CO: *EDUCAUSE*, 2019.
- [84] D. Keegan, "The incorporation of mobile learning into mainstream education and training," in *World Conference on Mobile Learning, Cape Town*, 2005, vol. 11, pp. 1-17.
- [85] H. Crompton, "The benefits and challenges of mobile learning," *Learning and leading with technology*, vol. 41, 2013.
- [86] I. Goksu, "Bibliometric mapping of mobile learning," *Telematics and Informatics*, vol. 56, p. 101491, 2021.
- [87] A. Bhati and I. Song, "New methods for collaborative experiential learning to provide personalised formative assessment," *International Journal of Emerging Technologies in Learning*, vol. 14, pp. 179-195, 2019.
- [88] M. Al-Emran, V. Mezhuyev, and A. Kamaludin, "Towards a conceptual model for examining the impact of knowledge management factors on mobile learning acceptance," *Technology in Society*, vol. 61, p. 101247, 2020.
- [89] M. J. Timms, "Letting artificial intelligence in education out of the box: educational cobots and smart classrooms," *International Journal of Artificial Intelligence in Education*, vol. 26, pp. 701-712, 2016.
- [90] I. Roll and R. Wylie, "Evolution and revolution in artificial intelligence in education," *International Journal of Artificial Intelligence in Education*, vol. 26, pp. 582-599, 2016.
- [91] S. Huang, B. Weiler, and G. Assaker, "Effects of interpretive guiding outcomes on tourist satisfaction and behavioral intention," *Journal of Travel Research*, vol. 54, no. 3, pp. 344-358, 2015.
- [92] Y. Reisinger and C. Steiner, "Reconceptualising interpretation: The role of tour guides in authentic tourism," *Current issues in tourism*, vol. 9, no. 6, pp. 481-498, 2006.
- [93] F. Umam, F. Adiputra, A. Dafid, and S. Wahyuni, "Autonomous museum tour guide robot with object detection using tensorflow learning machine," in *2022 IEEE 8th Information Technology International Seminar (ITIS)*, 2022: IEEE, pp. 274-281.
- [94] S. Wang and H. I. Christensen, "Tritonbot: First lessons learned from deployment of a long-term autonomy tour guide robot," in *2018 27th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, 2018: IEEE, pp. 158-165.
- [95] A. D. Diallo, S. Gobe, and V. Durairajah, "Autonomous tour guide robot using embedded system control," *Procedia Computer Science*, vol. 76, pp. 126-133, 2015.

- [96] S. Rosa *et al.*, "Tour guide robot: a 5G-enabled robot museum guide," *Frontiers in Robotics and AI*, vol. 10, p. 1323675, 2024.
- [97] I. A. Wong, Q. L. Lian, and D. Sun, "Autonomous travel decision-making: An early glimpse into ChatGPT and generative AI," *Journal of Hospitality and Tourism Management*, vol. 56, pp. 253-263, 2023.
- [98] I. Carvalho and S. Ivanov, "ChatGPT for tourism: applications, benefits and risks," *Tourism Review*, vol. 79, no. 2, pp. 290-303, 2024.
- [99] D. E. Salinas-Navarro, E. Vilalta-Perdomo, R. Michel-Villarreal, and L. Montesinos, "Using Generative Artificial Intelligence Tools to Explain and Enhance Experiential Learning for Authentic Assessment," *Education Sciences*, vol. 14, no. 1, p. 83, 2024.
- [100] R. P. d. Santos, "Enhancing chemistry learning with chatgpt and bing chat as agents to think with: A comparative case study," *arXiv preprint arXiv:2305.11890*, 2023.
- [101] Google, "Google earth homepage [online]," 2005.
- [102] B. Sekachev, M. Nikita, and Z. Andrey, "Computer vision annotation tool: a universal approach to data annotation," *Intel [Internet]*, vol. 1, 2019.
- [103] YAHBOOM. "Jetson Orin NX SUB Board with 16GB RAM Superior Kit." <https://category.yahboom.net/products/jetson-orin-nx?variant=45177043059004> (accessed June 4th, 2024).
- [104] I. Ajzen and M. Fishbein, "Attitude-behavior relations: A theoretical analysis and review of empirical research," *Psychological bulletin*, vol. 84, no. 5, p. 888, 1977.
- [105] P. B. Brandtzaeg and A. Følstad, "Why people use chatbots," in *Internet Science: 4th International Conference, INSCI 2017, Thessaloniki, Greece, November 22-24, 2017, Proceedings 4*, 2017: Springer, pp. 377-392.
- [106] K.-L. Hsiao and C.-C. Chen, "What drives continuance intention to use a food-ordering chatbot? An examination of trust and satisfaction," *Library Hi Tech*, vol. 40, no. 4, pp. 929-946, 2022.
- [107] P. Daugherty, H. Wilson, and K. Narain, "Generative AI will enhance—not erase—customer service jobs," *Harvard business review*. <https://hbr.org/2023/03/generative-ai-will-enhance-not-erase-customer-service-jobs>. Accessed, vol. 25, 2023.
- [108] C. Bartneck, D. Kulić, E. Croft, and S. Zoghbi, "Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots," *International journal of social robotics*, vol. 1, pp. 71-81, 2009.
- [109] M. Mori, "The uncanny valley: the original essay by Masahiro Mori," *Ieee Spectrum*, vol. 6, pp. 1-6, 1970.
- [110] S. Li, A. M. Peluso, and J. Duan, "Why do we prefer humans to artificial intelligence in telemarketing? A mind perception explanation," *Journal of Retailing and Consumer Services*, vol. 70, p. 103139, 2023.
- [111] X. Ma and Y. Huo, "Are users willing to embrace ChatGPT? Exploring the factors on the acceptance of chatbots from the perspective of AIDUA framework," *Technology in Society*, vol. 75, p. 102362, 2023.

- [112] A. Rapp, L. Curti, and A. Boldi, "The human side of human-chatbot interaction: A systematic literature review of ten years of research on text-based chatbots," *International Journal of Human-Computer Studies*, vol. 151, p. 102630, 2021.
- [113] Y. Cao *et al.*, "A comprehensive survey of ai-generated content (aigc): A history of generative ai from gan to chatgpt," *arXiv preprint arXiv:2303.04226*, 2023.
- [114] H. Zhang, Y. Li, F. Ma, J. Gao, and L. Su, "Texttruth: an unsupervised approach to discover trustworthy information from multi-sourced text data," in *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2018, pp. 2729-2737.
- [115] R. Rombach, A. Blattmann, D. Lorenz, P. Esser, and B. Ommer, "High-resolution image synthesis with latent diffusion models," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2022, pp. 10684-10695.
- [116] S. Wang, D. Lo, and L. Jiang, "An empirical study on developer interactions in stackoverflow," in *Proceedings of the 28th annual ACM symposium on applied computing*, 2013, pp. 1019-1024.
- [117] J. White *et al.*, "A prompt pattern catalog to enhance prompt engineering with chatgpt," *arXiv preprint arXiv:2302.11382*, 2023.
- [118] T. Araujo, "Living up to the chatbot hype: The influence of anthropomorphic design cues and communicative agency framing on conversational agent and company perceptions," *Computers in human behavior*, vol. 85, pp. 183-189, 2018.
- [119] J. Hill, W. R. Ford, and I. G. Farreras, "Real conversations with artificial intelligence: A comparison between human-human online conversations and human-chatbot conversations," *Computers in human behavior*, vol. 49, pp. 245-250, 2015.
- [120] Y. Mou and K. Xu, "The media inequality: Comparing the initial human-human and human-AI social interactions," *Computers in Human Behavior*, vol. 72, pp. 432-440, 2017.
- [121] L. Ciechanowski, A. Przegalinska, M. Magnuski, and P. Gloor, "In the shades of the uncanny valley: An experimental study of human-chatbot interaction," *Future Generation Computer Systems*, vol. 92, pp. 539-548, 2019.
- [122] J. Fan *et al.*, "A robotic coach architecture for elder care (ROCARE) based on multi-user engagement models," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 25, no. 8, pp. 1153-1163, 2016.
- [123] E. Go and S. S. Sundar, "Humanizing chatbots: The effects of visual, identity and conversational cues on humanness perceptions," *Computers in human behavior*, vol. 97, pp. 304-316, 2019.
- [124] J. G. Kemeny, "Man viewed as a machine," *Scientific American*, vol. 192, no. 4, pp. 58-67, 1955.
- [125] J. McCarthy, M. L. Minsky, N. Rochester, and C. E. Shannon, "A proposal for the dartmouth summer research project on artificial intelligence, august 31, 1955," *AI magazine*, vol. 27, no. 4, pp. 12-12, 2006.
- [126] S. Natale and A. Ballatore, "Imagining the thinking machine: Technological myths and the rise of artificial intelligence," *Convergence*, vol. 26, no. 1, pp. 3-18, 2020.

- [127] P. Bory, "Deep new: The shifting narratives of artificial intelligence from Deep Blue to AlphaGo," *Convergence*, vol. 25, no. 4, pp. 627-642, 2019.
- [128] R. M. Schuetzler, G. M. Grimes, and J. Scott Giboney, "The impact of chatbot conversational skill on engagement and perceived humanness," *Journal of Management Information Systems*, vol. 37, no. 3, pp. 875-900, 2020.
- [129] L. I. D. Faruk, R. Rohan, U. Ninrutsirikun, and D. Pal, "University Students' Acceptance and Usage of Generative AI (ChatGPT) from a Psycho-Technical Perspective," in *Proceedings of the 13th International Conference on Advances in Information Technology*, 2023, pp. 1-8.
- [130] M. Mazzone and A. Elgammal, "Art, creativity, and the potential of artificial intelligence," in *Arts*, 2019, vol. 8, no. 1: MDPI, p. 26.
- [131] D. Choi *et al.*, "Unlock Life with a Chat (GPT): Integrating Conversational AI with Large Language Models into Everyday Lives of Autistic Individuals," in *Proceedings of the CHI Conference on Human Factors in Computing Systems*, 2024, pp. 1-17.
- [132] D. Wang *et al.*, "From human-human collaboration to Human-AI collaboration: Designing AI systems that can work together with people," in *Extended abstracts of the 2020 CHI conference on human factors in computing systems*, 2020, pp. 1-6.
- [133] M. Cohn *et al.*, "Believing Anthropomorphism: Examining the Role of Anthropomorphic Cues on Trust in Large Language Models," in *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*, 2024, pp. 1-15.
- [134] A. Waytz, J. Heafner, and N. Epley, "The mind in the machine: Anthropomorphism increases trust in an autonomous vehicle," *Journal of experimental social psychology*, vol. 52, pp. 113-117, 2014.
- [135] D. Jaiswal, V. Kaushal, A. Mohan, and P. Thaichon, "Mobile wallets adoption: pre-and post-adoption dynamics of mobile wallets usage," *Marketing Intelligence & Planning*, vol. 40, no. 5, pp. 573-588, 2022.
- [136] T. E. Mofokeng, "The impact of online shopping attributes on customer satisfaction and loyalty: Moderating effects of e-commerce experience," *Cogent Business & Management*, vol. 8, no. 1, p. 1968206, 2021.
- [137] W. Reim, J. Åström, and O. Eriksson, "Implementation of artificial intelligence (AI): a roadmap for business model innovation," *Ai*, vol. 1, no. 2, p. 11, 2020.
- [138] C. M. Ringle, S. Wende, and J.-M. Becker, "SmartPLS 4. Oststeinbek: SmartPLS," *Retrieved March*, vol. 13, p. 2023, 2022.
- [139] L. J. Cronbach, "Coefficient alpha and the internal structure of tests," *psychometrika*, vol. 16, no. 3, pp. 297-334, 1951.
- [140] C. Fornell and D. F. Larcker, "Evaluating structural equation models with unobservable variables and measurement error," *Journal of marketing research*, vol. 18, no. 1, pp. 39-50, 1981.
- [141] R. P. Bagozzi and Y. Yi, "On the evaluation of structural equation models," *Journal of the academy of marketing science*, vol. 16, pp. 74-94, 1988.
- [142] J. F. Hair, "Multivariate data analysis," 2009.

- [143] G. James, D. Witten, T. Hastie, and R. Tibshirani, *An introduction to statistical learning*. Springer, 2013.