國立台東大學

資訊工程學系碩士班

碩士論文

AI輔助學習：基於Unity和ComfyUI的互動式創意生成教育工具

AI-Assisted Learning: An Interactive Creativity Generation Educational Tool Based on Unity and ComfyUI

研 究 生 : 胡修銘

指導教授 : 賴盈勳 博士

狄旻

目錄

[第一章 緒論 3](#_Toc183983167)

[1-1 研究背景與動機 3](#_Toc183983168)

[1-2 研究目的 3](#_Toc183983169)

[1-3 研究流程 4](#_Toc183983170)

[1-4 研究限制與範圍 4](#_Toc183983171)

[第二章 文獻探討 5](#_Toc183983172)

[2-1 AR智慧眼鏡 5](#_Toc183983173)

[2-2 ComfyUI 6](#_Toc183983174)

[2-3 生成式AI 6](#_Toc183983175)

[2-4 選用的ComfyUI模型 7](#_Toc183983176)

[2-4-1 TripoSR 7](#_Toc183983177)

[2-4-2 Flux 7](#_Toc183983178)

[2-4-3 SF3D 8](#_Toc183983179)

[2-4-5 SDXL 8](#_Toc183983180)

[第三章 流程設計與架構 10](#_Toc183983181)

[3-1 AR眼鏡前端與使用者介面 10](#_Toc183983182)

[3-2 ComfyUI 與WorkFlow 設置 10](#_Toc183983183)

[3-3 3D物件生成模型 10](#_Toc183983184)

[3-4 Unity API串接方式 10](#_Toc183983185)

[第四章 研究結果 11](#_Toc183983186)

[4-1 AR眼鏡選用 11](#_Toc183983187)

[4-2 3D生成模型選擇 11](#_Toc183983188)

[4-3 節點參數設置 11](#_Toc183983189)

[4-4 模型生成速度 11](#_Toc183983190)

[第五章 結論與建議 12](#_Toc183983191)

[5-1 合適的AR眼鏡款式 12](#_Toc183983192)

[5-2 選用模型的差異及生成時間突破 12](#_Toc183983193)

[5-3 參數或額外功能性的節點設置 12](#_Toc183983194)

[參考文獻資料 13](#_Toc183983195)

第一章 緒論

1-1 研究背景與動機

近年來人工智慧技術快速發展，特別是生成式AI的興起，如ChatGPT等大型語言模型的出現，不僅改變了人們的生活方式，也為教育領域帶來重大影響。在此背景下，培養學習者的AI素養已成為現代教育的重要課題。AI素養不僅包含對AI技術的基本認知與應用能力，更強調培養學習者具備AI倫理意識、批判性思考能力，以及在AI時代中所需的問題解決能力。

而智慧眼鏡作為一種新興的穿戴式科技，具備擴增實境(AR)功能，能夠將虛擬資訊無縫地整合到現實環境中，為學習者提供即時的視覺回饋與互動體驗。結合生成式AI與智慧眼鏡的創新教學模式，不僅能讓學習更具互動性和沉浸感，更能通過實時的AI輔助，幫助學習者更直觀地理解AI技術的應用場景與影響。

本研究計畫將致力於建立一個整合生成式AI與智慧眼鏡的創新學習環境。研究將聚焦於設計適合學生的AI素養的應用工具，透過智慧眼鏡提供的混合實境體驗，讓學習者能夠：

1. 直接體驗並操作生成式AI的各種應用場景
2. 理解Gen AI生成過程與其對參數對結果的影響
3. 培養對AI技術的正確認知與相關工具使用能力
4. 發展利用AI工具解決實際問題的能力

本研究將採用混合研究方法，結合量化與質性數據分析，評估此創新教學方式對提升學生AI素養的成效。研究成果預期能為科技教育領域提供新的教學模式參考，同時為培養具備AI素養的新世代人才提供有效的教育方案。此研究主題高度符合當前教育部推動的數位轉型與科技創新教育政策，也呼應了培養具備未來關鍵能力人才的重要目標。

通過此研究，我們期望能建立一個可推廣的AI素養教育模式，培養學生在AI時代所需的核心素養，包括：AI技術認知能力、AI倫理思辨能力、AI應用創新能力，以及跨領域整合能力。這不僅有助於提升學生的科技素養，更能為未來AI時代培育具備關鍵競爭力的人才。

1-2 研究目的

本研究基於生成式AI與智慧眼鏡的結合潛力，設定以下研究目的：

探索智慧眼鏡在室內設計領域的創新應用模式 透過結合生成式AI的capabilities，開發適合智慧眼鏡運行的生成式AI使用流程，本研究以室內設計輔助系統為應用情境。

研究重點著重在ComfyUI的生成式流程以及眼鏡上呈現3D物件效果的比較，為室內設計工作帶來效率提升。

評估ComfyUI框架在智慧眼鏡平台的整合效能 研究ComfyUI節點式架構與智慧眼鏡的整合方式，探討如何優化工作流程，以及在資源受限的智慧眼鏡平台上實現高效能的AI模型運算。同時評估不同模型在實際應用場景中的表現，為未來相關研究提供參考基準。

建立智慧眼鏡結合生成式AI的應用設計準則透過實際開發經驗，歸納出適合智慧眼鏡的生成式AI應用設計原則，包含使用者體驗考量、效能優化策略，以及如何善用ComfyUI的優勢特性。這些準則將有助於未來相關應用的開發。

1-3 研究流程

本研究依循下列步驟進行探討：

前置研究階段

1. 蒐集並分析智慧眼鏡、生成式AI及ComfyUI的相關文獻與技術資料
2. 評估現有智慧眼鏡硬體規格與運算能力
3. 研究ComfyUI框架的功能特性與擴充性

系統開發階段

1. 設計適合智慧眼鏡的使用者介面
2. 建立ComfyUI工作流程
3. 開發API整合介面
4. 進行效能優化與測試

實驗驗證階段

1. 執行系統功能測試
2. 進行效能評估
3. 收集使用者回饋
4. 歸納研究發現與建議

1-4 研究限制與範圍

 硬體限制

本研究僅針對特定型號的智慧眼鏡(Jorjin plus)進行測試，受限於目前智慧眼鏡的運算能力、電池續航力與散熱性能。同時，網路連線品質可能影響系統運作效能。

 軟體限制

研究主要基於ComfyUI框架進行開發，使用其支援的生成式AI模型。在功能方面，著重於2D/3D物件的生成與視覺化呈現，不包含複雜的物理模擬或即時渲染功能。

 應用範圍

本研究以室內設計輔助為主要應用情境，探討智慧眼鏡結合生成式AI的可行性。研究成果雖可作為其他領域的參考，但可能需要因應不同場景需求進行調整。

 技術限制

目前生成式AI模型的運算需求較高，在智慧眼鏡等資源受限設備上的效能表現可能受到影響。同時，模型的生成品質與速度也需要在實際應用中取得平衡。

第二章 文獻探討

2-1 AR智慧眼鏡

擴增實境 (AR) 是一種新興的體驗形式，它透過將電腦生成的內容與特定地點和或活動連結，來增強現實世界 (RW) 的體驗。近年來，AR 應用程式已變得更加便攜，並在行動裝置上廣泛使用 (NMC, 2011) 。AR 內容可以透過多種方式觀看，最初是透過網路攝影機觀看快速回應 (QR) 碼的網路應用程式。另一個觀看選項是使用頭戴式顯示器 (HMD)。使用者戴著 HMD (AR眼鏡) 時，可以在 HMD 螢幕上看到數位內容，並透過螢幕看到真實環境 (或透過連接的相機顯示在螢幕上) (Yuen, S et al，2011)。擴增實境 (AR) 技術現今已廣為人知，並被廣泛應用於醫療保健、餐旅、教育、旅 遊、文化、軍事、建築、設計、工程、遊戲、娛樂等領域 (Hsieh, M. C，2018)。像是其中土耳其醫療產業中醫師使用 AR 智慧眼鏡(ARSG) 的情況及其採用率的研究實，研究結果顯示，要瞭解醫師採用 ARSG 的原因，需要考慮許多因素，包含有用性、受易用性、相容性、提醒難易性和語音辨識的影響，而易用性受易學性、醫學教育難易程度、外部影響和隱私性的影響 (Basoglu, N. A，2018)。其ARSG 的開發者應該注意醫療保健的特定需求，以提高 ARSG 在醫療保健環境中的使用率和普及率 (Basoglu, N. A，2018)。而ARSG 可以被視為電腦系列的新成員，它們在螢幕和互動功能方面與手持電腦和其他穿戴式裝置截然不同 (Wasik，2013、Muensterer et al，2014)。ARSG 的主要優勢在於能夠在不中斷正在進行的活動的情況下虛擬化線上資訊 (Armstrong et al， 2014、Monroy et al，2014、Moshtaghi et al，2015)。

AR 眼鏡的硬體和軟體技術都在不斷發展，例如顯示螢幕的清晰度和解析度、硬體的耐用性、系統軟體與醫院資料庫的整合等。可能會結合 GPS、藍牙、相機、感測器和人臉辨識功能，並與醫院雲端資料庫連線，讓醫護人員更輕鬆地了解病人的詳細資訊。也可以與 MR 技術結合，用於遠距醫療教育，例如遠端手術刀訓練、癌症治療和護理教育訓練 (Hsieh, M. C，2018)。然而 AR 眼鏡在未來發展中仍然面臨一些挑戰，例如隱私問題，因為AR 眼鏡可以輕鬆地捕捉使用者和非使用者的資料，醫療領域尤其需要注意保護病人的隱私、顯示螢幕的清晰度和解析度、硬體的耐用性等還需要進一步提升、使用體驗還不夠好，例如長時間使用可能會造成眼睛疲勞 (Armstrong et al， 2014、Monroy et al，2014、Moshtaghi et al，2015)，價格還比較高，這限制了它的普及率 (Wasik，2013)。

總而言之，AR 眼鏡的發展前景非常廣闊，但要實現大規模應用，還需要克服一些挑戰。隨著技術的進步和應用領域的擴展，相信 AR 眼鏡未來會在各個領域發揮越來越重要的作用。

2-2 ComfyUI

ComfyUI 是一款功能強大、操作簡便的 AI 生成插件，其基於節點式的介面設計，讓使用者可以自由地組合各種功能模組和節點，創建出複雜的 AI 繪圖流程。支援 Stable Diffusion 1.0、2.0、XL 等模型，並且具有高度的擴展性和相容性，可以整合各種插件和模型，滿足使用者不同的需求。(羅禎俋，2024)

工作流節點中也有許多在 AI繪圖常見的正、反提詞 (Prompt)、採樣方法 (Sampling Methods) 及採樣步驟 (Steps)、重繪值(Classifier-free guidance，簡稱 CFG) 與 Controlnet等插件運用 (徐志溢，2023)。

不僅可以在具有 VRAM 的 GPU 設備上進行高效運算，也可以在沒有 GPU 的情況下進行運算，這為使用者提供了更大的靈活性。提供了豐富的功能，例如文生圖、圖生圖、圖片微調、局部調整、遮罩、細節修正、影片製作等等，可以滿足使用者從簡單到複雜的各種 AI 繪圖需求，並且可離線使用，使用者只需要將套件與模型下載後即可在本地端運行，不用透過對外網路 (羅禎俋，2024)。

CotroNet

ControlNet 是一種神經網絡架構，可作為大型模型和預訓練模型中的圖像擴散模型的基本框架。它允許對圖像生成過程進行精確控制，例如佈局、形狀、姿勢和位置，而這些僅通過文字提示難以準確表達 (羅禎俋，2024)，在 ComfyUI 中，ControlNet 可以與 Stable Diffusion 模型結合使用，並且包含各種檢測模型和處理器，例如 Canny 邊緣檢測、深度檢測和 OpenPose 人體姿態檢測 (Gal, R et al，2024)。

ControlNet 的工作原理：

當 ControlNet 處於啟動狀態時，它會鎖定正在運行的 Stable Diffusion 模型，並利用預先構建的編碼層和卷積神經網絡進行條件控制，同時確保不會破壞生成過程中的降噪過程 (Zhang, L et al，2023)。

ControlNet 的應用：

控制角色形態 : ControlNet 可以通過分析深度、姿態和邊緣訊息，將圖片以形態轉繪的方式提供給 AI 繪圖，確保角色形態的穩定性和一致性

生成更精確的圖像： 使用 ControlNet 可以有效地控制圖像生成過程，避免僅使用文字提示時可能出現的隨機性和不確定性。

製作特定鏡頭： 在製作特寫鏡頭等需要精確控制細節的場景時，ControlNet 可以利用深度和線條信息作為物體判斷基準，幫助 AI 繪圖更準確地生成目標圖像 (羅禎俋，2024)。

ComfyUI 中 ControlNet 的優勢：

主要優勢之一是它能夠根據用戶提示生成更精確的圖像。 通過提供額外的指導，ControlNet 可以幫助減少僅依賴文字提示時可能出現的隨機性和不確定性。這對於需要精確控制圖像構圖和結構的任務特別有用，例如在特定場景中生成特定鏡頭或控制角色姿勢 (Gupta, Y et al，2024、Yuan, Z et al，2024、Ghosh, D et al，2024)。

ComfyUI 中 ControlNet 的限制：

儘管 ControlNet 是一個強大的工具，但它也有一些局限性。 例如，它可能難以處理涉及複雜相機運動的場景，例如手持拍攝或扭曲效果。 這是因為 ControlNet 主要關注圖像中物體的結構和位置，而相機運動會顯著改變這些因素。 此外，由於 AI 圖像生成模型在識別遠處物體方面仍然面臨挑戰，因此即使使用 ControlNet 也可能難以準確地生成包含遠處物體的圖像。 最後，在使用 AI 圖像生成模型進行影片後製時，可能會出現由於權重偏差導致生成失敗的問題。 這些限制表明，儘管 ControlNet 可以顯著提高圖像生成的準確性和穩定性，但它仍然是一個正在發展的技術，在某些應用中可能存在挑戰 (Gal, R et al，2024)。

2-3 生成式AI

生成式 AI，也稱為生成式人工智慧（generative artificial intelligence），是一種透過機器學習和深度學習技術，依託數位化圖像數據庫，運用演算法生成圖像的技術。 生成式 AI 可應用於生成文本、圖像、影音、音樂等內容，這些內容也被稱為 AIGC（AI generated content），即人工智慧生成內容 (羅禎俋，2024) 。AIGC 指的是利用先進的 GAI 技術生成的內容 ，而不是由人類作者創作的內容。AIGC 可以自動化內容創作，在短時間內創作大量內容。舉例來說，ChatGPT 是一個由 OpenAI 開發用於建構對話式 AI 系統的語言模型，它可以有效地理解 並以有意義的方式回應人類語言的輸入。另一個例子是 DALL-E，這是另一個由 OpenAI 開發的GAI 模型，它能夠在幾分鐘內根據文字描述創作出獨特且高品質的圖像 ( A. Ramesh，2022)。GAI 的發展歷程可以追溯到早期的生成模型，如 GAN 和 VAE1。近年来，随着 Transformer 模型的出現，GAI 技術得到了快速發展，並應用於更廣泛的領域包括藝術、廣告和教育( N. Anantrasirichai，2021、 J. Kietzmann et al，2018、 M. Kandlhofer et al，2016 )。

在高齡化社會和醫護人員不足的背景下，科技創新被視為改善醫療體系效能和提升照護品質的關鍵。 其中，生成式 AI 被認為是具有潛力的技術之一，可以用於輔助高齡照護，例如提供情感支持、安全監控和娛樂等 (李奉爵，2024)。

生成式 AI也能應用於教育領域，為教學帶來創新和突破，例如客製化學習體驗、互動式學習環境、自動化評量和回饋、提升教師效率、激發創造力和想像力等 (羅禎俋，2024)

未來AIGC將是機器學習的一個重要的研究領域，包括醫療保健、金融服務、自動駕駛汽車和科學發現以及增強 GAI 模型的推理能力，使其能夠更好地理解和處理複雜的任務 ( S. Reddy et al，2020、 Y. Qi et al，2018、 S. Grigorescu et al， 2020、 Y. Gil et al，2014、 R. Taylor et al，2022)。

2-4 選用的ComfyUI模型

2-4-1 TripoSR

TripoSR 是一個基於 Transformer 架構的 3D 重建模型，能夠從單一影像快速生成 3D 網格。 該模型在 NVIDIA A100 GPU 上的推論時間不到 0.5 秒。 TripoSR 建立在大型重建模型 (LRM) 的基礎上，並在資料處理、模型設計和訓練技術方面進行了多項改進 (Dmitry Tochilkin1, 2024)。

TripoSR 的核心組件包括一個影像編碼器、一個影像到三平面解碼器和一個基於三平面的神經輻射場 (NeRF)。 首先，輸入的 RGB 影像會被編碼成特徵向量。 然後，解碼器將這些特徵向量轉換成三平面特徵表示，用於描述 3D 物體的形狀和外觀。 最後，NeRF 模組根據三平面特徵渲染出新的視圖，並用於訓練模型 (Dmitry Tochilkin1, 2024)。

為了提高模型的性能，TripoSR 在資料收集方面進行了兩項改進：資料策展和一致的渲染設定。 資料策展的目標是從 Objaverse 資料集中選擇一個精心策展的高品質子集，以減少低品質資料對模型訓練的影響。 一致的渲染設定則確保所有訓練影像都使用相同的相機參數和光照條件生成，從而提高資料的一致性。TripoSR 的模型和訓練方面也進行了一些改進，包括使用感知損失函數 (LPIPS) 來提升重建品質。 LPIPS 是一種用於衡量兩張影像之間感知相似度的指標，能夠更好地捕捉人類視覺系統對影像差異的感知 (Dmitry Tochilkin1, 2024)。在公開資料集上的評估結果顯示，TripoSR 在 3D 重建品質和速度方面都優於其他開源模型。 在 GSO 和 OmniObject3D 資料集上進行的量化比較表明，TripoSR 在 Chamfer Distance 和 F-score 等指標上都取得了最佳性能。

TripoSR 也具有較高的計算效率，能夠在 0.5 秒內從單一影像生成 3D 網格。 與其他快速 3D 重建方法相比，TripoSR 在重建品質和速度之間取得了良好的平衡 (Dmitry Tochilkin1, 2024)。然而，TripoSR 也存在一些缺點。 與使用基於網格訓練的 SF3D 模型相比，TripoSR 的重建精度較低。 此外，在處理低動態範圍 (LDR) 輸入時，TripoSR 輸出的顏色可能與輸入影像不完全匹配。 TripoSR 模型的另一個限制是它沒有輸入姿態和內在條件，因此模型會被鼓勵去猜測物體的尺度，這可能導致尺度預測不準確 (Hanwen Jiang,2024)。

總體而言，TripoSR 是一種快速且高效的單一影像 3D 重建模型，它在 3D 生成領域取得了顯著進展。

2-4-2 Flux

Flux模型專注於即時3D生成：

* 動態場景生成
* 即時渲染優化
* 光影效果模擬
* 物件互動處理

2-4-3 SF3D

SF3D 是一種從單張圖像快速重建高品質帶紋理和 UV 展開的 3D 模型的方法，並具有去光照的材質屬性。 與大多數現有方法不同，SF3D 是專門為網格生成而訓練的，它採用一種快速的 UV 展開技術，能夠快速生成紋理，而不是依賴頂點顏色。 該方法還可以學習預測材質參數和法線貼圖，以提高重建的 3D 網格的視覺品質。 此外，SF3D 整合了一個去光照步驟，可以有效地去除低頻照明效果，確保重建的網格可以在新的照明條件下輕鬆使用 (Mark Boss，2024)。

SF3D 的主要目標是解決現有快速 3D 重建模型中的一些關鍵問題，同時保持快速生成速度（在 H100 GPU 上只需 0.5 秒），這些關鍵問題包括 :

光照烘焙 : 現有技術通常將陰影烘培到材質中，SF3D 透過整合明確的照明和使用球面高斯函數 (SG) 的可微分著色模型，提出了分解照明和反射屬性的方法。

頂點著色 : 大多數 3D 生成模型使用頂點著色來表示物體紋理，這會導致生成的 3D 資源在遊戲等應用中使用效率低下。SF3D 提出一種高度可並行化的快速基於盒子投影的 UV 展開技術，以實現 0.5 秒的生成時間。

步進立方體偽影 : 前饋網路通常會創建立體表示，例如 Triplane-NeRFs，然後使用步進立方體 (MC) 演算法將其轉換為網格。 MC 會導致「步進」偽影，可以透過增加體積解析度來減少，但這會帶來巨大的計算開銷。相比之下，SF3D 使用更高效的架構來處理更高解析度的三平面圖，並使用 DMTet 生成網格，並具有學習到的頂點位移和法線貼圖，從而產生更平滑的網格表面。

缺乏材質屬性 : 以前的前饋技術生成的模型在使用不同的照明渲染時，通常看起來很暗淡。 這主要是由於輸出生成中缺乏明確的材質屬性，這會影響光的反射。 為了應對這一問題，SF3D 預測非空間變化的材質屬性。

透過這些改進，SF3D 可以從單張圖像生成高品質的 3D 網格，並具有平滑形狀、分離的材質屬性、可在0.5秒內生成及較小的物件資源(小於1MB)(Mark Boss，2024)。

2-4-5 SDXL

SDXL (Stable Diffusion XL)模型具備：

* 高解析度圖像生成
* 改進的細節表現
* 更好的構圖能力
* 多樣化的風格控制

這些模型的選擇基於以下考量：

* 運行效能要求需符合AR眼鏡的硬體限制
* 生成效果需滿足室內設計應用需求
* 推理速度需達到實時互動的要求
* 模型整合難度需考慮開發資源限制

各模型特性比較：

1. 運算效能

TripoSR：★★★★☆ (較低資源需求)

Flux：★★★☆☆ (中等資源需求)

SF3D：★★☆☆☆ (較高資源需求)

SDXL：★★★☆☆ (中等資源需求)

1. 生成質量

TripoSR：★★★★☆ (優秀的3D重建質量)

Flux：★★★★☆ (良好的即時渲染效果)

SF3D：★★★★★ (最佳的結構重建能力)

SDXL：★★★★★ (頂級的圖像生成質量)

1. 推理速度

TripoSR：★★★★★ (最快)

Flux：★★★★☆ (較快)

SF3D：★★★☆☆ (中等)

SDXL：★★★☆☆ (中等)

第三章 流程設計與架構

3-1 AR眼鏡前端與使用者介面

3-2 ComfyUI 與WorkFlow 設置

3-3 3D物件生成模型

3-4 Unity API串接方式

第四章 研究結果

4-1 AR眼鏡選用

4-2 3D生成模型選擇

4-3 節點參數設置

4-4 模型生成速度

第五章 結論與建議

5-1 合適的AR眼鏡款式

5-2 選用模型的差異及生成時間突破

5-3 參數或額外功能性的節點設置

參考文獻資料

[聚焦AR/VR產品走向與應用趨勢 | 電腦與通訊](https://ictjournal.itri.org.tw/xcdoc/cont?sid=0M257582562602272358&xsmsid=0M208578644085020215)

[AR智慧眼鏡帶動Micro LED需求大爆發， 2026年產值達4,100萬美元|數位時代 BusinessNext](https://www.bnext.com.tw/article/70331/-micro-led-ar-smart-glasses--explode0629)

Gal, R., Haviv, A., Alaluf, Y., Bermano, A. H., Cohen-Or, D., & Chechik, G. (2024). ComfyGen: Prompt-Adaptive Workflows for Text-to-Image Generation. arXiv preprint arXiv:2410.01731.

Zhao, Y., Cheng, Y., Ding, S., Fang, Y., Cao, W., Liu, K., & Cao, J. (2024, June). Magic Camera: An AI Drawing Game Supporting Instantaneous Story Creation for Children. In Proceedings of the 23rd Annual ACM Interaction Design and Children Conference (pp. 738-743).

Xue, X., Lu, Z., Huang, D., Ouyang, W., & Bai, L. (2024). GenAgent: Build Collaborative AI Systems with Automated Workflow Generation--Case Studies on ComfyUI. arXiv preprint arXiv:2409.01392.

徐志溢（2023）。 基於 Stable Diffusion 模型和 LoRA 優化訓練生成人臉表情

羅禎俋. (2024). AI 工具 ComfyUI 輔助 2D 人型角色動畫應用於各景別之研究. 嶺東科技大學數位媒體設計系碩士班學位論文, 1-66.

基於角色情感互動與主動式照護的生成式 AI 模型能力研究; Role-Based Emotional Interaction and Proactive Care Capabilities of Generative AI Models. 2024. PhD Thesis. 國立中央大學.

Huan-Yi Chen、Jinn-Tsong Tsai, Ph.D. (2023). 生成式AI在生成演算法與資料結構文本之相似度與超參數性能研究Research on the Similarity and Hyperparameter

Performance of Generative AI in Generating Algorithm and Data Structure Texts

Li-Heng Chen、Ding-Ming Wang (2024). 運用生成式 AI 工具對研究生學術研究能力影響之實證研究An Empirical Study of Generative AI as a Research Tool for Graduate Students

Tochilkin, D., Pankratz, D., Liu, Z., Huang, Z., Letts, A., Li, Y., ... & Cao, Y. P. (2024). Triposr: Fast 3d object reconstruction from a single image. arXiv preprint arXiv:2403.02151.

Jiang, H., Huang, Q., & Pavlakos, G. (2024). Real3D: Scaling Up Large Reconstruction Models with Real-World Images. *arXiv preprint arXiv:2406.08479*.

Yu, X., Peng, B., Galley, M., Gao, J., & Yu, Z. (2023). Teaching Language Models to Self-Improve through Interactive Demonstrations. arXiv preprint arXiv:2310.13522.

Ma, K., & Chung, J. (2024). A Research on AI Generated 2D Image to 3D Modeling Technology.

Fu, J., Fu, S., & Grierson, M. (2024). Coral Model Generation from Single Images for Virtual Reality Applications. arXiv preprint arXiv:2409.02376.

Xu, J., Cheng, W., Gao, Y., Wang, X., Gao, S., & Shan, Y. (2024). Instantmesh: Efficient 3d mesh generation from a single image with sparse-view large reconstruction models. arXiv preprint arXiv:2404.07191.

Boss, M., Huang, Z., Vasishta, A., & Jampani, V. (2024). Sf3d: Stable fast 3d mesh reconstruction with uv-unwrapping and illumination disentanglement. arXiv preprint arXiv:2408.00653.

Zhai, X., Chu, X., Chai, C. S., Jong, M. S. Y., Istenic, A., Spector, M., ... & Li, Y. (2021). A Review of Artificial Intelligence (AI) in Education from 2010 to 2020. Complexity, 2021(1), 8812542.

 Xue, X., Lu, Z., Huang, D., Ouyang, W., & Bai, L. (2024). GenAgent: Build Collaborative AI Systems with Automated Workflow Generation--Case Studies on ComfyUI. arXiv preprint arXiv:2409.01392.

Gal, R., Haviv, A., Alaluf, Y., Bermano, A. H., Cohen-Or, D., & Chechik, G. (2024). ComfyGen: Prompt-Adaptive Workflows for Text-to-Image Generation. arXiv preprint arXiv:2410.01731.

Zhao, Y., Cheng, Y., Ding, S., Fang, Y., Cao, W., Liu, K., & Cao, J. (2024, June). Magic Camera: An AI Drawing Game Supporting Instantaneous Story Creation for Children. In Proceedings of the 23rd Annual ACM Interaction Design and Children Conference (pp. 738-743).

Yuen, S. C. Y., Yaoyuneyong, G., & Johnson, E. (2011). Augmented reality: An overview and five directions for AR in education. Journal of Educational Technology Development and Exchange (JETDE), 4(1), 11.

Basoglu, N. A., Goken, M., Dabic, M., Ozdemir Gungor, D., & Daim, T. U. (2018). Exploring adoption of augmented reality smart glasses: Applications in the medical industry.

Hsieh, M. C., & Lee, J. J. (2018). Preliminary study of VR and AR applications in medical and healthcare education. J Nurs Health Stud, 3(1), 1.

Gu, J. (2024). A Survey on Responsible Generative AI: What to Generate and What Not. arXiv preprint arXiv:2404.05783.

Cao, Y., Li, S., Liu, Y., Yan, Z., Dai, Y., Yu, P. S., & Sun, L. (2023). A comprehensive survey of ai-generated content (aigc): A history of generative ai from gan to chatgpt. arXiv preprint arXiv:2303.04226.

Anantrasirichai, N., & Bull, D. (2022). Artificial intelligence in the creative industries: a review. Artificial intelligence review, 55(1), 589-656.

 Kietzmann, J., Paschen, J., & Treen, E. (2018). Artificial intelligence in advertising: How marketers can leverage artificial intelligence along the consumer journey. Journal of Advertising Research, 58(3), 263-267.

 Kandlhofer, M., Steinbauer, G., Hirschmugl-Gaisch, S., & Huber, P. (2016, October). Artificial intelligence and computer science in education: From kindergarten to university. In 2016 IEEE frontiers in education conference (FIE) (pp. 1-9). IEEE.

 Reddy, S., Allan, S., Coghlan, S., & Cooper, P. (2020). A governance model for the application of AI in health care. Journal of the American Medical Informatics Association, 27(3), 491-497.

Qi, Y., & Xiao, J. (2018). Fintech: AI powers financial services to improve people's lives. Communications of the ACM, 61(11), 65-69.

 Grigorescu, S., Trasnea, B., Cocias, T., & Macesanu, G. (2020). A survey of deep learning techniques for autonomous driving. Journal of field robotics, 37(3), 362-386.

Gil, Y., Greaves, M., Hendler, J., & Hirsh, H. (2014). Amplify scientific discovery with artificial intelligence. Science, 346(6206), 171-172.

Taylor, R., Kardas, M., Cucurull, G., Scialom, T., Hartshorn, A., Saravia, E., ... & Stojnic, R. (2022). Galactica: A large language model for science. arXiv preprint arXiv:2211.09085.

Wei, J., Wang, X., Schuurmans, D., Bosma, M., Xia, F., Chi, E., ... & Zhou, D. (2022). Chain-of-thought prompting elicits reasoning in large language models. *Advances in neural information processing systems*, *35*, 24824-24837.

Wei, J., Wang, X., Schuurmans, D., Bosma, M., Xia, F., Chi, E., ... & Zhou, D. (2022). Chain-of-thought prompting elicits reasoning in large language models. *Advances in neural information processing systems*, *35*, 24824-24837.

Zhang, Z., Zhang, A., Li, M., Zhao, H., Karypis, G., & Smola, A. (2023). Multimodal chain-of-thought reasoning in language models. arXiv preprint arXiv:2302.00923.

Schramowski, P., Tauchmann, C., & Kersting, K. (2022, June). Can machines help us answering question 16 in datasheets, and in turn reflecting on inappropriate content?. In Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency (pp. 1350-1361).

Wang, J., Liu, X. G., Di, Z., Liu, Y., & Wang, X. E. (2023). T2iat: Measuring valence and stereotypical biases in text-to-image generation. arXiv preprint arXiv:2306.00905.

Dathathri, S., Madotto, A., Lan, J., Hung, J., Frank, E., Molino, P., ... & Liu, R. (2019). Plug and play language models: A simple approach to controlled text generation. arXiv preprint arXiv:1912.02164.

Touvron, H., Martin, L., Stone, K., Albert, P., Almahairi, A., Babaei, Y., ... & Scialom, T. (2023). Llama 2: Open foundation and fine-tuned chat models. arXiv preprint arXiv:2307.09288.

Zhang, F., Sun, Z., & Chen, Q. (2024). Research on Interior Intelligent Design System Based On Image Generation Technology. Procedia Computer Science, 243, 690-699.

Zhang, L., Rao, A., & Agrawala, M. (2023). Adding conditional control to text-to-image diffusion models. In Proceedings of the IEEE/CVF International Conference on Computer Vision (pp. 3836-3847).

Gupta, Y., Jaddipal, V. V., Prabhala, H., Paul, S., & Von Platen, P. (2024). Progressive knowledge distillation of stable diffusion xl using layer level loss. arXiv preprint arXiv:2401.02677.

Yuan, Z., Liu, Y., Cao, Y., Sun, W., Jia, H., Chen, R., ... & Sun, L. (2024). Mora: Enabling generalist video generation via a multi-agent framework. arXiv preprint arXiv:2403.13248.

Ghosh, D., Hajishirzi, H., & Schmidt, L. (2024). Geneval: An object-focused framework for evaluating text-to-image alignment. Advances in Neural Information Processing Systems, 36.

Due, B. L. (2014). The future of smart glasses: An essay about challenges and possibilities with smart glasses (Vol. 1, pp. 1-21). København, Denmark: Centre of Interaction Research and Communication Design, University of Copenhagen.

Wasik, B. (2013). Why wearable tech will be as big as the smartphone. Retrieved April, 26, 2014.

Muensterer, O. J., Lacher, M., Zoeller, C., Bronstein, M., & Kübler, J. (2014). Google Glass in pediatric surgery: an exploratory study. International journal of surgery, 12(4), 281-289.

Armstrong, D. G., Rankin, T. M., Giovinco, N. A., Mills, J. L., & Matsuoka, Y. (2014). A heads-up display for diabetic limb salvage surgery: a view through the google looking glass. Journal of diabetes science and technology, 8(5), 951-956.

Monroy, G. L., Shemonski, N. D., Shelton, R. L., Nolan, R. M., & Boppart, S. A. (2014, February). Implementation and evaluation of Google Glass for visualizing real-time image and patient data in the primary care office. In Advanced Biomedical and Clinical Diagnostic Systems XII (Vol. 8935, pp. 166-174). SPIE.

Moshtaghi, O., Kelley, K. S., Armstrong, W. B., Ghavami, Y., Gu, J., & Djalilian, H. R. (2015). Using Google Glass to solve communication and surgical education challenges in the operating room. The Laryngoscope, 125(10), 2295-2297.