

# 國家科學及技術委員會

## 114年度大專學生研究計畫申請書

一、綜合資料：

申請條碼：114CFD2500085



申請人 【學生】	姓 名	張詠翔	身分證號碼	L12598****
	就 讀 學 校、 科 系 及 年 級	中山醫學大學醫學資訊學系 3 年級	電 話	0968412637
	學 生 研 究 計 畫 名 稱	D-MMFusion：用於人工智慧輔助失智症診斷的多模態特徵融合		
	研 究 期 間	自114年7月1日至115年2月底止，計8個月		
	計 畫 歸 屬 司 別	工程處		
	研究學門代碼及名稱	E4105 -- 人工智慧與計算智慧		
	上年度曾執行本會大專學生研究計畫	是（計畫編號：NSTC 113-2813-C-040-039-B）		
指導教授	姓 名	曾明性	身分證號碼	J12041****
	服 務 機 構 及 科 系(所)	中山醫學大學醫學資訊學系		
	職 稱	教授	電 話	(04)2473-0022-12214
補助經費	項 目	金 額	說 明	
	研 究 助 學 金	48,000元	每月補助研究助學金新臺幣6,000元，8個月計新臺幣48,000元。	
	耗材、物品、圖書及雜項費用	20,000元	依研究計畫實際需求擇優補助，每一計畫最高以補助新臺幣20,000元為限。	
	合 計	68,000元		
本人瞭解並同意遵照學術倫理規範；本計畫無違反學術倫理				

表C801

## 二、研究計畫內容（以 10 頁為限）：

### （一）摘要

本研究計畫針對現行失智症診斷方法的主觀判斷偏差、診斷成本高及效率低等問題，提出基於多模態大模型（MM-LLMs, Multimodal Large Language Models）的「**D-MMFusion：用於人工智慧輔助失智症診斷的多模態特徵融合**」人工智慧診斷輔助研究方案，結合醫學影像與臨床文本，以提升診斷準確度與自動化程度。

本研究方法始於多模態數據預處理與特徵提取，以及文本標準化與詞嵌入處理，以確保數據的一致性與融合性。其後將進行 MM-LLMs 訓練，本研究選用 LLaVA 與 BLIP-2 模型來做測試訓練，採用 InfoNCE Loss 進行對比學習，評估指標包括 Recall@K、BLEU 與 FID 等。訓練過程將使用 Liger-Kernel 作為加速訓練方案，透過高效 Triton 核心程式加速模型訓練，減少 GPU 記憶體使用，提高計算效率，並結合 Glows.ai 雲端算力平台進行大規模數據處理，以提升運算吞吐量並降低運行成本。本計畫期望透過多模態 AI 技術提升失智症診斷準確率，為臨床醫學與人工智慧領域的交叉應用提供創新方法，並改善現有診斷方法的效率與可行性。

### （二）研究動機與研究問題

失智症為全球性公共健康的挑戰，隨著人口老齡化，其發病率持續上升，根據台灣失智症協會的推估，到了 2065 年時更將會超過 89 萬人，然而傳統診斷方法主要依賴於影像判讀與臨床訪談，存有主觀判斷偏差、診斷成本高及效率低等問題。此外，單一模態數據往往無法提供足夠的疾病特徵，導致診斷準確率受限。

隨著人工智慧與深度學習技術的發展，多模態學習已成為改善醫學影像分析與自然語言處理的重要方法。多模態大模型（MM-LLMs, Multimodal Large Language Models）透過結合影像與文本資訊，將可提高模型對於醫學數據的理解與預測能力，從而提升診斷精確度與自動化程度。

本計畫提出「**D-MMFusion：用於人工智慧輔助失智症診斷的多模態特徵融合**」，旨在針對失智症影像與臨床文本進行數據預處理與特徵提取，確保不同模態數據的一致性與可比性，同時比較多種視覺模態編碼器，評估其在醫學影像分析中的表現，再利用預訓練語言模型對臨床文本進行處理，提取關鍵診斷詞與醫療事件，並利用詞嵌入技術進行數值化表示。最終，本研究將採取雲端算力平台以及針對 MM-LLMs 的訓練加速方案來測試多項多模態大模型。透過解決上述研究問題，本研究期望能提供一種創新的多模態 AI 診斷輔助方案，提升失智症診斷的準確性與可行性，為臨床醫學與人工智慧領域的交叉應用提供有價值的參考。「D-MMFusion：用於人工智慧輔助失智症診斷的多模態特徵融合」研究流程示意圖如圖1所示。

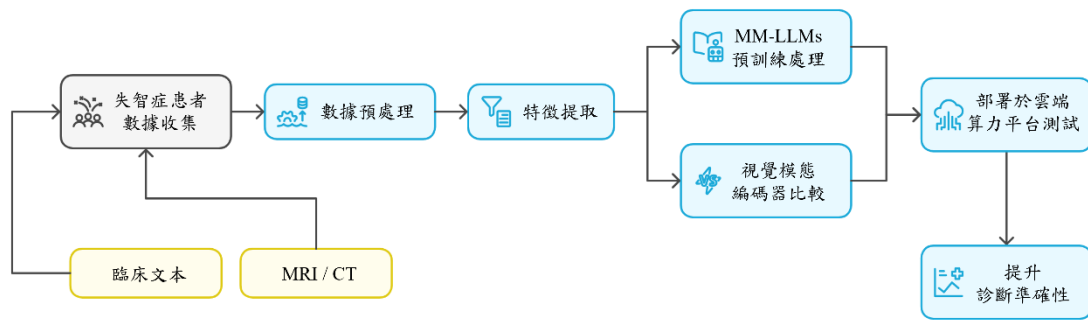


圖 1、「D-MMFusion」研究流程示意圖

### (三)文獻回顧與探討

大型語言模型做為輔助工具廣泛應用於各場域專業，有許多醫學相關的應用被提出，本計畫參考多篇文獻所提及的內容，取其特長後引入本研究，將做為計畫中基於多模態大模型（MM-LLMs）的實驗方案。

#### 1. 失智症資料集之相關研究

進行失智症檢測有助於及早發現疾病，減少症狀惡化的可能性，提高生活品質，針對失智症檢測，已有多項公開醫療研究資料庫。

ADNI (Alzheimer's Disease Neuroimaging Initiative) [1]: 自2004年起，便已開始收集與分析關於失智症的腦部MRI影像數據和隨訪數據，其目標是為了讓醫療論文研究可以在長期追蹤中開發出新的診斷與治療方式。

OASIS (Open Access Series of Imaging Studie): 此資料庫開放多項神經影像資料，支持神經科學和臨床研究，包括不同年齡層次下大腦組織的 MRI 影像。其中包含在2007年，Marcus 學者等人進行彙編，初始資料集由 416 名年齡在 18 至 96 歲之間的受試者的橫斷面集合構成之「OASIS-1」[2]; 在2010年，Marcus 學者等人進行彙編，包括 150 名年齡在 60 至 96 歲之間的受試者的縱向收集，並使用臨床癡呆評定量表 (CDR) 將受試者描述為非癡呆或患有極輕度至輕度阿茲海默症所構成之「OASIS-2」[3]; 在2019年，LaMontagne 學者等人進行彙編，由1098名參與者的MRI和PET成像以及相關臨床數據所構成之「OASIS-3」[4]; 在2020年，Koenig 學者等人自1990年至2018年收集之正常老化、訓練、測試和臨床隊列的1287名參與者，由研究或臨床患者記錄於2020年組成之「OASIS-4」[5]。

#### 2. 影像特徵提取視覺編碼器之相關研究

為了深入探索大腦組織變化的關鍵模式，本研究將分別選用「NFNet」、「ViT」與「CLIP ViT」三種不同的視覺模態編碼器。

NFNet (Normalizer-Free ResNets) [6]: 在2021年，Brock 學者等人所發表，是一種不使用 Batch Normalization、網路架構基於 ResNet 的模型，並且還提出了自適應梯度修剪 (Adaptive Gradient Clipping, AGC)，使用更大的 batch size 和更強的 data augmentations 進行訓練。NFNet模型架構圖如下圖2所示。

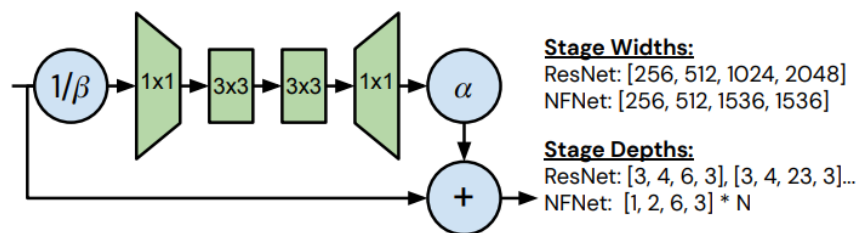


圖 2、NFNet模型架構圖

ViT (Vision Transformer) [7]: 在2017年, Alexey 學者等人所發表, 是一種基於Transformer架構的模型, 主要將影像劃分為多個區塊, 以此進行全局特徵採集, 對於大量資料進行預訓練並轉移到多個中型或小型影像辨識基準時, ViT相比於卷積神經網絡, 將取得更優異的效果。ViT模型架構圖如下圖3所示。

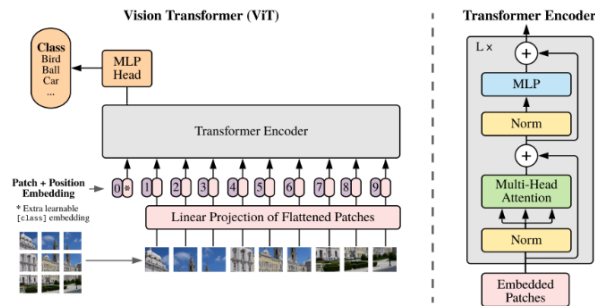


圖3、ViT模型架構圖

CLIP ViT (Contrastive Language-Image Pre-training Vision Transformer) [8]: 在2021年, 由Alec 學者等人所發表, 主要利用對比學習方法, 將影像與文本嵌入至同一特徵空間, 以確保影像與文字的關聯性。CLIP ViT模型架構圖如下圖4所示。

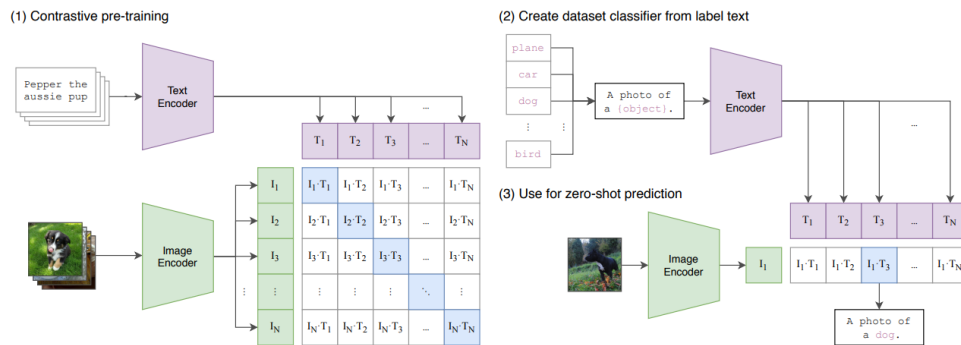


圖4、CLIP ViT模型架構圖

### 3. 針對MM-LLMs發展進程之研究

在2017年, Vaswani 學者等人首先提出了大型語言模型的架構[9], 其網路架構是基於 Seq2Seq + self-attention mechanism堆疊而成的Transformer 模型, 而後於2023年, Touvron 學者等人首度將大型語言模型LLaMa2作為開源技術開放給任何人使用[10], 其開源易調參數的性質為大型語言模型生態系統帶來重大進展。

其後為有效利用預先訓練的大型語言模型 (LLM) 和視覺模型的能力, 於2023年, Li 學者等人發表了BLIP-2[11], 之中所提出的Q-Former將可連接預訓練的視覺模型和大語言模型, 以此完成高效的多模態預訓練。BLIP-2 與 Q-Former 視覺語言表徵學習目標的模型架構圖如下圖5所示。

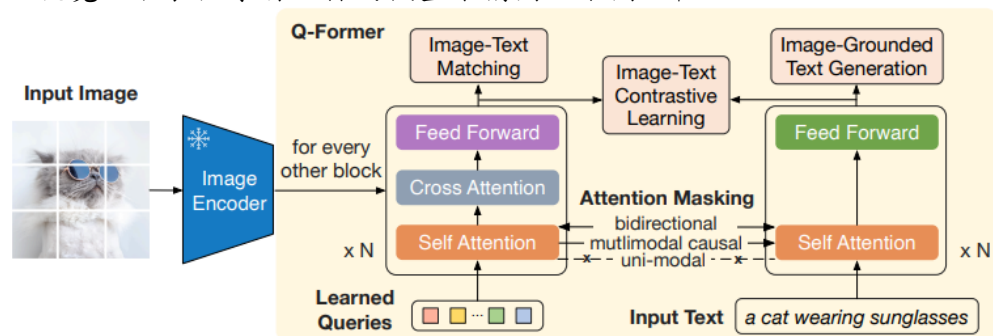


圖 5、BLIP-2 與 Q-Former 視覺語言表徵學習目標的模型架構圖

於2024年Liu 學者等人所發表了LLaVA：基於LLMs的多模態預訓練模型[12]。並可透過視覺指令進行微調訓練，使模型能夠理解和生成與視覺相關的文本內容。LLaVA 網路架構圖如下圖6所示。

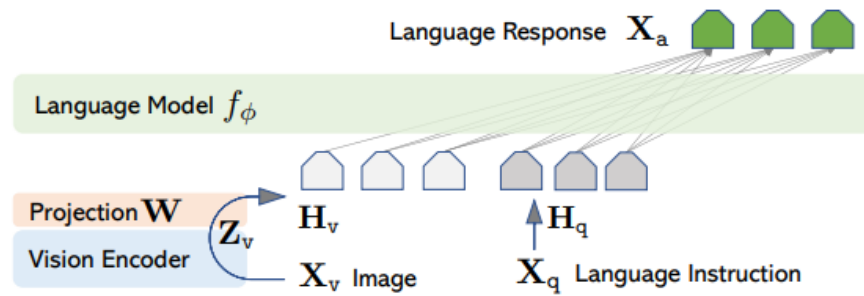


圖6、LLaVA 網路架構圖

#### (四)研究方法與步驟

「D-MMFusion：用於人工智慧輔助失智症診斷的多模態特徵融合」的實驗的設計流程主要分為「失智症之多模態數據預處理與特徵提取」、「訓練 MM-LLM 與評估優化」與「採用 Liger-Kernel 模型訓練加速方案」三大項目，「D-MMFusion：用於人工智慧輔助失智症診斷的多模態特徵融合」研究方法流程圖如圖 7 所示。

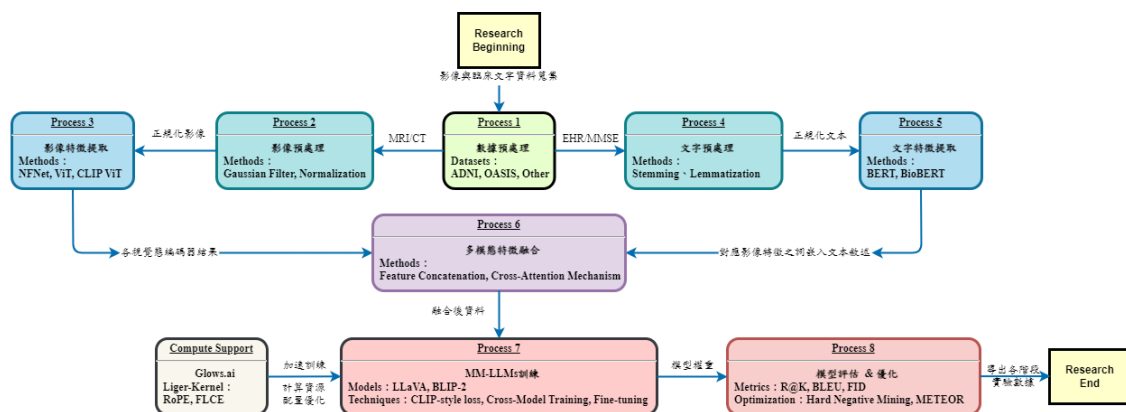


圖 7、「D-MMFusion」研究方法流程圖

##### 1. 針對多模態數據之失智症資料預處理與特徵提取

本階段聚焦於影像與臨床文字資料的收集、處理及特徵提取，確保數據的一致性與高效融合，以提升後續模型的準確度。

本階段研究步驟涵蓋影像與文字資料的統一預處理，以及對應的特徵提取方法。本研究的影像數據來源將從ADNI與OASIS等公開醫療研究資料數據庫收集腦部MRI或CT影像與病患電子病歷，針對影像數據將使用NFNet、ViT與CLIP ViT將影像特徵編碼提取，文字數據則是影像所對應之醫生診斷報告及問卷評估結果，並篩選符合研究需求的患者數據，以確保數據來源的多樣性與完整性。最終將使影像與文字建立對應關係，以確保同一位病患的MRI 影像與其診斷報告對應，以便模型學習影像與語義之間的關聯。本研究數據將採取統一多模態數據預處理將分為「影像預處理」與「文字預處理」，多模態數據預處理流程圖如圖8所示。





圖 8、「D-MMFusion」多模態數據預處理流程圖

- I. 影像預處理：首先將對每張影像進行灰階歸一化，確保所有影像數據的亮度與對比度一致，再根據影像品質，選擇是否應用高斯濾波、Wavelet 轉換進行降噪，提升影像清晰度。
  - II. 文字預處理：移除非必要標點、統一醫學術語(如 ICD-10 編碼標準化)。再使用自然語言處理 (NLP) 中的 Stemming、Lemmatization 等技術統一詞彙[13]，而後轉換為 JSON 等結構化文件格式。
- 將資料預處理後，將進行多模態特徵提取，分為「影像特徵提取」以及「文字特徵提取」，多模態特徵提取流程圖如圖9所示。



圖9、「D-MMFusion」多模態特徵提取流程圖

- I. 影像特徵提取：任務目標為提取大腦結構變化的關鍵模式，如腦室擴大、皮質萎縮等特徵。將測試不同視覺模態編碼器 (Vision Encoder)，比較其效果：
    - i. NNet：針對醫療影像的高效 CNN 模型。
    - ii. Vision Transformer (ViT)：基於 Transformer 的影像處理架構。
    - iii. CLIP ViT：利用對比學習 (Contrastive Learning) 確保影像文本關聯。
  - II. 文字特徵提取：採用最後採用 BERT [14] 與 BioBERT [15] 預訓練語言模型對臨床文本進行語義理解，提取診斷關鍵詞，使其文字敘述對應至影像特徵，並通過詞嵌入 (Word Embedding) 轉換為數值表示。
- 最終將進行多模態特徵融合，為了確保影像與文字特徵的高效結合，本研究將測試「特徵拼接 (Feature Concatenation)」與「交叉注意力機制 (Cross-Attention Mechanism)」兩種多模態融合策略，前者將直接將影像與文字，透過 MLP 進行分類或預測，而後者則讓影像與文字的特徵相互關聯，確保模型學習到影像與語義之間的依存關係。

## 2. 訓練 MM-LLMs 與評估

在完成數據處理與特徵提取後，本研究將使用多模態大模型 (MM-LLMs) 進行影像與文字的聯合學習，以確保模型能夠從多模態數據中提取出有價值的資訊。針對模型選擇，本研究將先測試並比較 LLaVA (Large Language and Vision Assistant) 和 BLIP-2 (Bootstrapping Language-Image Pretraining) 兩種架構的 MM-LLMs，並進行微調任務。訓練方式則採用 CLIP-style loss (InfoNCE Loss) 讓模型學習影像與文字之間的匹配關係[16]，透過正樣本 (Positive Pairs)

和負樣本（Negative Pairs）來提高模型的區分能力。再來將採取交叉模態訓練（Cross-Modal Training），透過 Transformer 進行影像-文本交互，確保影像和文字內容能互相影響。訓練過程將利用 Glows.ai 雲端算力平台，提供H100、RTX-4090等高效能GPU進行大規模訓練。「D-MMFusion：用於人工智慧輔助失智症診斷的多模態特徵融合」MM-LLMs訓練示意圖如圖10所示。

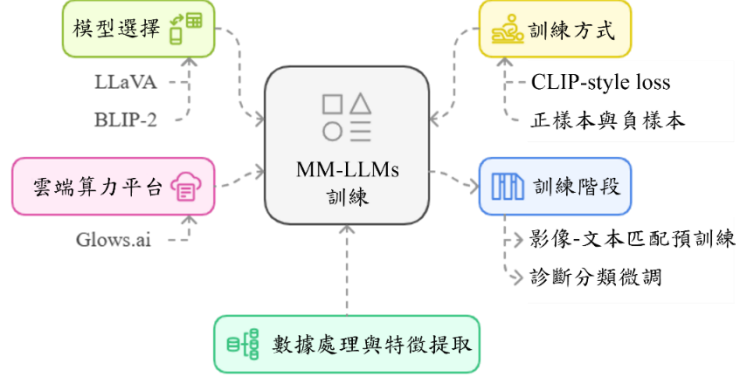


圖10、「D-MMFusion」MM-LLMs訓練示意圖

模型訓練完成後，將進行模型評估與優化任務，包含「模型評估指標」與「優化方法」。在模型評估方面，針對影像-文本匹配（Image-Text Matching），會使用 Recall@K（R@K）來測試模型能否正確檢索匹配的文本，若模型匹配準確率較低，則增加 Hard Negative Mining 提高對比學習效果。R@K評估指標公式如(1)所示。

$$Recall@k = \frac{TP@k}{TP@k + FN@k} \quad (1)$$

在文本生成（Text Generation）方面，主要以 BLEU（Bilingual Evaluation Understudy），以及METEOR（Metric for Evaluation of Translation with Explicit ORdering）做為評估指標，以進一步評估文本品質。BLEU評估指標公式如(2)所示；METEOR評估指標公式如(3)所示。

$$BLEU = BP \cdot \exp \left( \sum_{n=1}^N w_n \log p_n \right) \quad (2)$$

$$BP = \begin{cases} 1 & \text{if } c > r \\ e^{(1-r/c)} & \text{if } c \leq r \end{cases}$$

$$F = \frac{(\alpha^2 + 1)P}{R + \alpha P}, \quad \text{Penalty} = \gamma \left( \frac{chunks}{unigrams\_matched} \right)^\theta \quad (3)$$

$$Meteor = (1 - \text{Penalty}) \cdot F$$

$p$ ：準確率， $r$ ：召回率， $\alpha$ ：平衡權重，

$unigrams\_matched$ ：候選與參考匹配量，

$chunks$ ：相匹配詞組數量， $\gamma, \theta$ ：調整懲罰強度參數(根據實驗設置)

此外，在影像與文字融合評估上，本研究將採用 FID（Frechet Inception Distance）來衡量生成的文本是否自然，以及其與原始數據分布的相似程度。FID評估指標公式如(4)所示。



$$FID(x, g) = |\mu_x - \mu_g|^2 + T_r(\Sigma_x + \Sigma_g - 2\sqrt{\Sigma_x \Sigma_g}) \quad (4)$$

$\mu_x$ ：真實圖像分佈的特徵均值， $\mu_g$ ：生成圖像分佈的特徵均值，

$\Sigma_x$ ：真實圖像分佈的特徵covariance matrix，

$\Sigma_g$ ：生成圖像分佈的特徵covariance matrix，

$T_r$ ：矩陣對角線元素總和， $\gamma, \theta$ ：調整懲罰強度參數(根據實驗設置)，

$unigrams\_matched$ ：候選與參考匹配量， $chunks$ ：相匹配詞組數量

### 3.採用 Liger-Kernel 模型訓練加速優化方案

Liger-Kernel為一組高效Triton核心程式[17]，旨在提升大型語言模型(LLM)訓練的效率。其主要透過「運算融合(operation fusion)」和「輸入分塊(input chunking)」等技術，將多個運算整合為單一核心操作，減少記憶體存取次數，從而提升訓練速度並降低GPU記憶體使用量。透過Liger-Kernel可將訓練吞吐量平均提高約20%，同時減少平均約60%的GPU記憶體使用量。其中，在Kernels內所涵蓋的演算法包含RoPE (Rotary Position Embedding)，將位置編碼與查詢和鍵的旋轉嵌入結合，減少計算開銷，於本研究中處理大量批次訓練MM-LLM資料時能夠降低計算複雜度並提高效率，幫助模型更精準的處理患者病歷記錄(EHR)與MRI報告對應關係等長序列醫學文本，以確保模型能夠更準確地學習影像與文字之間的多模態關聯。RoPE (Rotary Position Embedding) 公式如(5)所示：

$$y = R^{d\theta, m} x \quad (5)$$

$R^{d\theta, m}$ ：旋轉矩陣， $y$ ：輸出向量， $x$ ：輸入向量，

$d$ ：向量的維度(輸入向量的長度)，

$\theta$ ：模型特定參數(定義旋轉形式)，

$m$ ：當前token的位置索引

另外，FLCE (Fused Linear CrossEntropy) 演算法則是透過避免同時生成logits張量來減少記憶體消耗，以提高計算效率，並降低GPU的記憶體使用量，於本研究中將可加速文本提取效率，減少醫學影像與病歷對應生成過程中的記憶體瓶頸，使得模型可以支持更長的文本輸入。FLCE (Fused Linear CrossEntropy) 公式如(6)所示：

$$L = - \sum_i t_i \log(y_i) \quad (6)$$

$L$ ：交叉熵損失值(表示預測結果與真實標籤之間的差距)，

$t_i$ ：真實標籤的 one-hot 編碼(對應於第*i*個類別的標籤值)，

$y_i$ ：模型對第*i*個類別的預測概率

為確保Liger-Kernel的高效運行，本研究預計將整合於Glows.ai雲端算力平台進行實驗，利用雲端算力調度(Compute Resource Optimization)透過動態算力分配技術，確保記憶體與計算資源的最優化配置，使得本研究在訓練測試中，採取Glows.ai + Liger-Kernel方案相較於傳統相較於使用地端訓練，將大幅節省整體訓練時間，並減少GPU記憶體佔用空間，大幅降低運行成本。Glows.ai + Liger-Kernel方案選項如圖11所示。

Instance Type		OS					
GPU		CPU		Windows		Linux	
GPU Model		GPUs	GPU RAM	Avail Unit	vCPUs	RAM	Price/h
<input type="radio"/>	NVIDIA GeForce RTX 4090	1	24 GB	10	10 ⓘ	48 GB	6.500 Credits
<input checked="" type="radio"/>	NVIDIA H100 80GB HBM3	1	80 GB	8	16 ⓘ	224 GB	35.000 Credits

Official Image

Snapshot

Selected -

img-4blk9qr6

Liger Kernel

Ubuntu20.04 CUDA12.2,  
Python3.11, NVCC,  
JupyterLab/VNC supported

img-6jq1yqgo

Ubuntu24.04 CUDA11.8

Ubuntu 24.04 CUDA11.8

img-9vl7bq1j

GEMMA2-WebUI

WebUI GEMMA2 9B

img-j3lomp8e

GEMMA2

ComfyUI GEMMA2

## (六)需要指導教授指導內容

在模型訓練階段，本計畫將針對LLaVA與BLIP-2進行測試與微調，並透過InfoNCE Loss與交叉模態訓練來提升模型的診斷能力，因此期望能與指導教授討論訓練策略，根據其專業經驗，調整模型的超參數設定，優化多模態特徵融合技術，提升影像與文本匹配的準確率，並確保診斷輔助系統的可靠性與穩定性。

在計算資源優化方面，本計畫預計採用Liger-Kernel加速模型訓練，並結合Glows.ai雲端算力平台，以減少GPU記憶體使用量並提高計算效率。為確保訓練過程的順利進行，期望能與指導教授討論高效能計算技術的適用性，分析不同雲端計算平台的優勢與限制，並針對批次處理與資源調度策略進行調整，以降低運行成本並提高運算吞吐量。

計畫完成後，期望能與指導教授共同探討本研究的可改進之處，進一步優化診斷輔助系統的架構，並將研究成果整理成論文，投稿至醫學與人工智慧領域的國際期刊與研討會，讓本計畫的研究成果接受外界學者的檢視與指導，促進學術交流，並為人工智慧在臨床醫學的應用提供實質貢獻。

## (七)參考文獻

- [1] Petersen, Ronald Carl, et al. "Alzheimer's disease Neuroimaging Initiative (ADNI) clinical characterization." *Neurology* 74.3 (2010): 201-209.
- [2] Marcus, Daniel S., et al. "Open Access Series of Imaging Studies (OASIS): cross-sectional MRI data in young, middle aged, nondemented, and demented older adults." *Journal of cognitive neuroscience* 19.9 (2007): 1498-1507.
- [3] Marcus, Daniel S., et al. "Open access series of imaging studies: longitudinal MRI data in nondemented and demented older adults." *Journal of cognitive neuroscience* 22.12 (2010): 2677-2684.
- [4] LaMontagne, Pamela J., et al. "OASIS-3: longitudinal neuroimaging, clinical, and cognitive dataset for normal aging and Alzheimer disease." *medrxiv* (2019): 2019-12.
- [5] Koenig, Lauren N., et al. "Select Atrophied Regions in Alzheimer disease (SARA): An improved volumetric model for identifying Alzheimer disease dementia." *NeuroImage: Clinical* 26 (2020): 102248.
- [6] Brock, Andy, et al. "High-performance large-scale image recognition without normalization." *International conference on machine learning*. PMLR, 2021.
- [7] Dosovitskiy, Alexey. "An image is worth 16x16 words: Transformers for image recognition at scale." *arXiv preprint arXiv:2010.11929* (2020).

- [8] Radford, Alec, et al. "Learning transferable visual models from natural language supervision." *International conference on machine learning*. PMLR, 2021.
- [9] Vaswani, A. "Attention is all you need." *Advances in Neural Information Processing Systems* (2017).
- [10] Touvron, Hugo, et al. "Llama 2: Open foundation and fine-tuned chat models." *arXiv preprint arXiv:2307.09288* (2023).
- [11] Li, Junnan, et al. "Blip-2: Bootstrapping language-image pre-training with frozen image encoders and large language models." *International conference on machine learning*. PMLR, 2023.
- [12] Liu, Haotian, et al. "Visual instruction tuning." *Advances in neural information processing systems* 36 (2024).
- [13] Cambria, Erik, et al. "Sentiment analysis is a big suitcase." *IEEE Intelligent Systems* 32.6 (2017): 74-80.
- [14] Devlin, Jacob. "Bert: Pre-training of deep bidirectional transformers for language understanding." *arXiv preprint arXiv:1810.04805* (2018).
- [15] Lee, Jinhyuk, et al. "BioBERT: a pre-trained biomedical language representation model for biomedical text mining." *Bioinformatics* 36.4 (2020): 1234-1240.
- [16] Patashnik, Or, et al. "Styleclip: Text-driven manipulation of stylegan imagery." *Proceedings of the IEEE/CVF international conference on computer vision*. 2021.
- [17] Hsu, Pin-Lun, et al. "Liger kernel: Efficient triton kernels for llm training." *arXiv preprint arXiv:2410.10989* (2024).

### 三、耗材、物品、圖書及雜項費用：

(一) 凡執行研究計畫所需之耗材、物品、圖書及雜項費用，均可填入本表內。

(二) 說明欄請就該項目之規格、用途等相關資料詳細填寫，以利審查。

(三) 依研究計畫實際需求擇優補助，每一計畫最高以補助新臺幣 20,000 元為限。

金額單位：新臺幣元

項目名稱	說明	單位	數量	單價	金額	備註
1.雲端平台算力租借	Glows.ai雲端算力 調度需求	Credit	5,000	3.2	16,000	Credit為雲平台 計價單位
2.雜支	大型固態硬碟、學 習資料等	批	1	4,000	4,000	
合 計					20,000	

表C803

畢業總成績：

[illegible]

最低畢業學分數：128 (必修學分數71 學分)

- 中山醫學大學  
Chung Shan Medical University  
教務處  
Academic Affairs Office



畢業總成績：

[illegible]

最低畢業學分數：128 (必修學分數71 學分)

1. 抵免 # 校際選課 @ 暑修 & 免修 @ 跨校雙主修輔系 \* 不及格 @ 成績未送出 P 核定通過 P 核定不通過
2. 抵免 & 校際選課學分、成績不列入當學期學分及平均成績，惟列入畢業修業總學分，不列入畢業總成績。
3. 暑修學分及成績不列入當學期學分及平均成績，惟列入畢業修業總學分及畢業總成績。
4. 跨校雙主修輔系修讀科目學分及成績列入當學期與畢業學分及成績。



工業技術研究院

Industrial Technology  
Research Institute

# 其他與計畫相關有助於審查之文件

作者：張詠翔

就讀學校：中山醫學大學

就讀學系：醫學資訊學系

2025/02/15



1.

自我介紹



2.

系統開發能力



3.

校內外合作計畫



4.

競賽參與



5.

論文發表



# 個人簡歷—自我介紹



## ◆ 人格特質

好奇心



積極主動



抗壓性高



承擔責任



## ◆ 專業課程修習

- C語言
- 網頁程式設計
- 深度學習
- Python與程式設計
- 作業系統
- 資料庫系統
- 演算法

## ◆ 實習經歷

時間	實習單位	職位
2024年7月至2024年8月	點睛科技股份有限公司	軟體開發暑期Intern
2024年12月	中山醫學大學	智慧醫療暨大數據分析實驗室 Research Assistant



# 個人簡歷—系統開發能力 (1/7)

程式語言能力	實作案例	貢獻項目
深度學習程式設計： <b>Python</b> 深度學習程式集： <b>PyTorch、TensorFlow、Keras、Transformers、NLTK</b>	「大專學生研究計畫-TransMbD失智監護輔助流程」	題目發想、需求擷取、模型訓練、系統撰寫、語言模型研究
	「中山醫學大學-護理治療溝通技巧聊天機器人系統」	需求擷取、系統撰寫
全端程式設計： <b>HTML5、PHP、CSS、JavaScript、MySQL</b> 網頁設計框架： <b>Bootstrap</b>	「兒童髖關節偵測輔助系統，DDH - Detection Assistive System」	需求擷取、分析與設計、網頁撰寫、模型訓練、系統撰寫
行動軟體程式設計： <b>Arduino C、SQLite、Java、Kotlin</b> 嵌入式系統： <b>Arduino、MaixPy</b>	「點睛科技-AI Exercise 智慧手錶」	需求擷取、分析與設計、系統撰寫
	「OverWatch」輔助道路安全穿戴式裝置	需求擷取、硬體電路開發
	「SunBone向日骨健」骨質疏鬆預防輔助系統	題目發想、模型訓練、硬體電路開發
	「一圖勝千言」溝通交流輔助系統	需求擷取、網頁撰寫

# 個人簡歷—系統開發能力 (2/7)

- ◆ **專案實作能力 (1/2)** - 以下為實作專案及其應用之能力，包含模型訓練、網頁前後端撰寫、需求擷取、系統分析與設計、邊緣運算穿戴式裝置開發

## 點睛科技 實習經歷

1

**個人簡歷—系統開發能力 (4/7)**

◆ 專案實作能力-AI Exercise 智慧手錶：點睛科技

- 主要開發技術：Android Studio(Kotlin、Java)、Firebase、Figma

本專案主要為開發一款WearOS App，透過智慧手錶的裝置感測器及SDK，可追蹤心率、睡眠品質、卡路里消耗、步數等健康數據，幫助用戶了解身體狀況，維持健康生活。因此針對對長者的各項活動，透過串接API，將手錶上監測數據回傳至點睛科技內部網站內進行客製化分析，以推演出長者健康狀態，同時推播訓練訊息。




**AI Exercise 智慧手錶**

## AI模型結合穿戴式裝置 競賽電子組 亞軍

2

**個人簡歷—系統開發能力 (5/7)**

◆ 專案實作能力-「OverWatch」輔助道路安全穿戴式裝置

- 主要開發技術：建置EfficientNet模型、控制OLED顯示面板輸出

聖胡安常因無法避免外來影響導致車輛發生，因此開發這項穿戴式裝置。本裝置由可部署AI模型的嵌入式系統為主要開發板組成，我將訓練好的EfficientNet聲音辨識模型部署至開發板中，該模型可辨識環境中的救護車聲、警車聲、消防車聲及呼喊聲等外界聲音，系統若辨識到這些聲音時，便會透過顯示面板提醒使用者須注意周圍環境。本裝置目前已讓聖胡安實際配戴，並確實減少了30%的事故發生機率。




**輔助道路安全穿戴式裝置**





# 個人簡歷—系統開發能力 (3/7)

- ◆ **專案實作能力 (2/2)** - 以下為實作專案及其應用之能力，包含模型訓練、網頁前後端撰寫、需求擷取、系統分析與設計、邊緣運算穿戴式裝置開發

中山醫學大學  
彰化基督教醫院

3

工業技術研究院  
Industrial Technology  
Research Institute

個人簡歷—系統開發能力 (6/7)

◆ 專案實作能力-兒童髖關節偵測輔助系統「DDH - Detection Assistive System」

• 主要開發技術：Python、Flask、OpenCV、PyTorch  
框架、HTML5、CSS、JavaScript、PHP、WAMP  
Server、MySQL、SQLite

本專案使用與彰化基督教醫院合作提供的  
兒童髖關節影像作為訓練資料集，經過特徵比  
度增強後，進行特徵點標記連線。使用者輸入影  
像後，系統將自動計算Alpha角與Beta角，並生  
成對應的臨床，為醫師提供一套完整的影像輔助  
判讀流程，從而提升診療效率。

功能示意圖

系統架構圖

五權基金會工業技術研究院 2024/11/30 5

兒童髖關節偵測輔助系統

中山醫學大學 護理系

4

工業技術研究院  
Industrial Technology  
Research Institute

個人簡歷—系統開發能力 (7/7)

◆ 專案實作能力-護理治療溝通技巧聊天機器人系統

• 主要開發技術：Python、NLTK、MongoDB

本專案開發一種具有提示和評分功能的治療溝通  
技巧聊天機器人，用於護理實習。護理實習生可以  
使用聊天機器人在實習前進行複習或解決實習中遇到  
的疑問。此外，該聊天機器人具有結構化查詢語言  
(SQL) 資料庫和評分功能，允許每個護理實習生重複  
練習並記錄每次練習的狀態和分數。

系統運作示意圖

系統架構示意圖

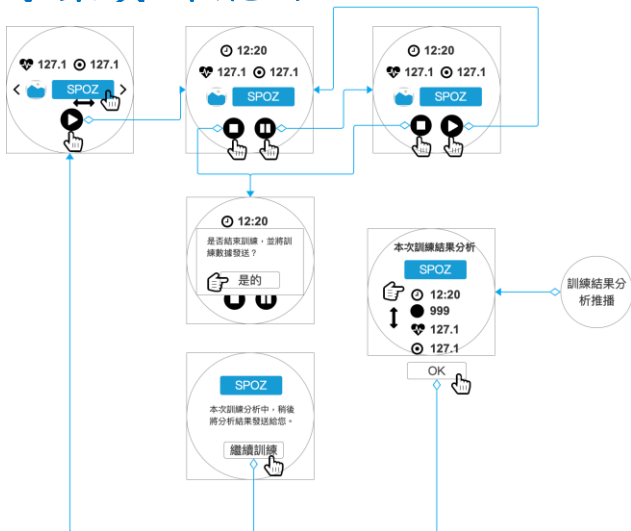
五權基金會工業技術研究院 2024/11/30 5

護理治療溝通技巧聊天機器人系統

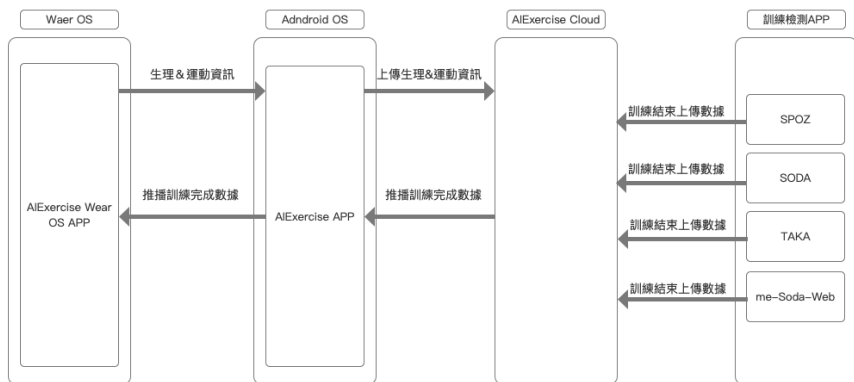


# 個人簡歷—系統開發能力 (4/7)

## ◆ 專案實作能力-AI Exercise 智慧手錶：點睛科技



AI Exercise 整體流程圖



AI Exercise 系統架構說明

- 主要開發技術：Android Studio(Kotlin、Java)、Firebase、Figma

本專案主要為開發一款WearOS App，透過智慧手錶的裝置感測器及SDK，可追蹤心率、睡眠品質、卡路里消耗、步數等健康數據，幫助用戶了解身體狀況，維持健康生活。因此將針對長者的各項活動，透過串接API，將手錶上監測數據回傳至點睛科技內部網站內進行客製化分析，以推理出長者健康狀態，同時推播訓練訊息。



AI Exercise Watch  
Prototype

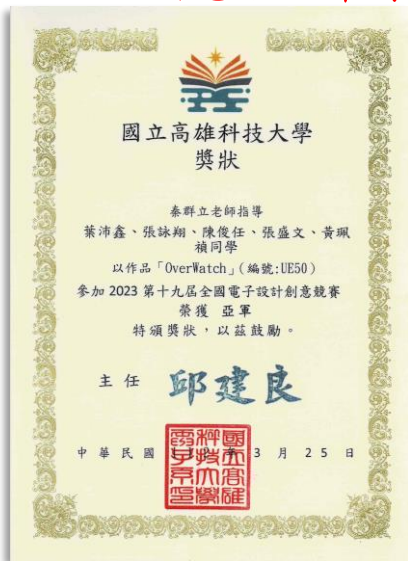


# 個人簡歷—系統開發能力 (5/7)

## ◆ 專案實作能力-「OverWatch」輔助道路安全穿戴式裝置



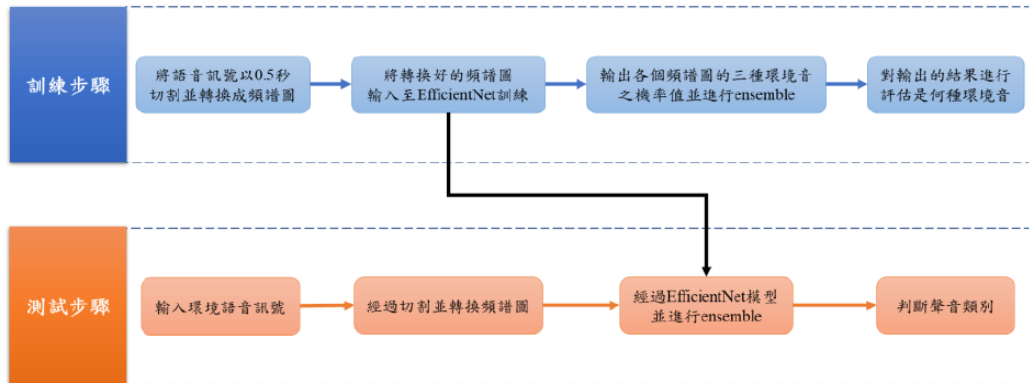
輔助道路安全穿戴式裝置  
2023第19屆全國電子設計創意競賽



電子組亞軍

- 主要開發技術：建置EfficientNet模型、控制OLED顯示面板輸出

聾朋友常因無法聽見外界聲響導致車禍發生，因此開發這項穿戴式裝置。本裝置由可部署AI模型的嵌入式系統為主要開發板組成，我將訓練好的EfficientNet聲音辨識模型部署至開發板中，該模型可辨識環境中的救護車聲、警車聲、消防車聲及呼喊聲等外界聲音，系統若辨識到這些聲音時，便會透過顯示面板提醒使用者須注意周遭環境。本裝置目前已讓聾朋友實際配戴，並確實減少了30%的事故發生機率。



# 個人簡歷—系統開發能力 (6/7)

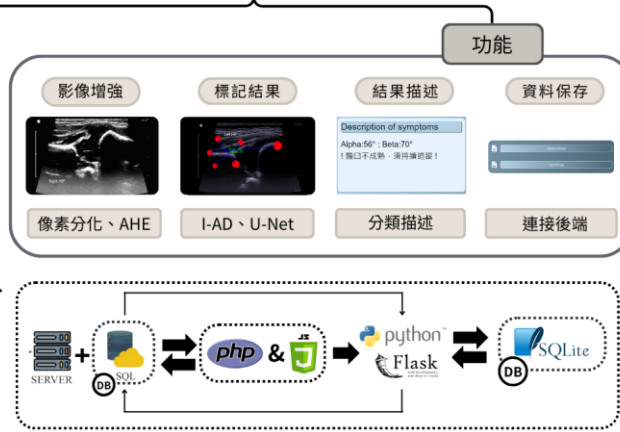
## ◆ 專案實作能力-兒童髖關節偵測輔助系統「DDH - Detection Assistive System」

DDH - DAS  
兒童髖關節偵測輔助系統

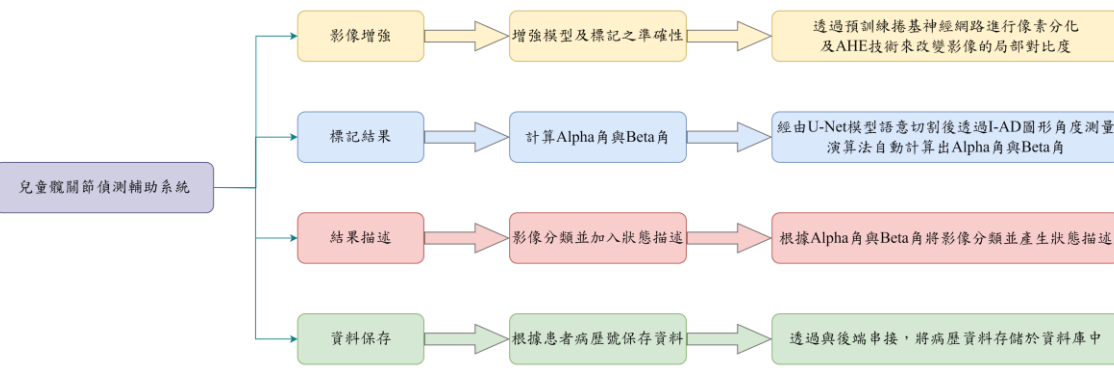


- 主要開發技術：Python、Flask、OpenCV、PyTorch  
框架、HTML5、CSS、JavaScript、PHP、WAMP  
Server、MySQL、SQLite

本專案使用與彰化基督教醫院合作提供的兒童髖關節影像作為訓練資料集，經過特徵對比度增強後，進行特徵點標記連線。使用者輸入影像後，系統將自動計算Alpha角與Beta角，並生成對應的描述，為醫師提供一套完整的影像輔助判讀流程，從而提升診療效率。



功能示意圖



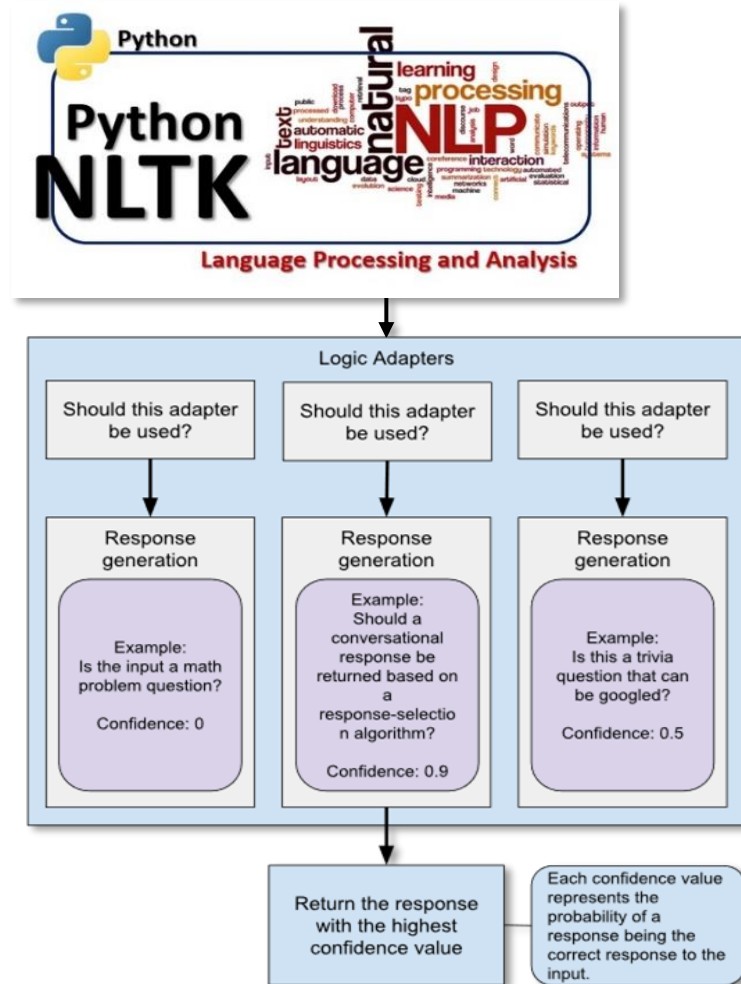
系統架構圖





# 個人簡歷—系統開發能力 (7/7)

## ◆ 專案實作能力-護理治療溝通技巧聊天機器人系統



護理實習治療溝通技巧聊天機器人  
系統技術示意圖

- 主要開發技術：Python + NLTK、MongoDB

本專案開發一種具有提示和評分功能的治療性溝通技巧聊天機器人，用於護理實習。護理實習生可以使用聊天機器人在實習前進行複習或解決實習中遇到的疑問。此外，該聊天機器人具有結構化查詢語言（SQL）資料庫和評分功能，允許每個護理實習生重複練習並記錄每次練習的狀態和分數。

### 系統運作示意

AI護理導師：歡迎使用個案護理照護教學系統

AI護理導師：請您輸入學號

護生：9198068

AI護理導師：請您輸入欲查詢的個案，例如：個案1st、個案2nd、個案11th、個案12th、個案21st、個案22nd

護生：個案21st

AI護理導師：妄想型思覺失調症、智能不足及自閉傾向，此次會談目的是希望了解個案對入院原因的認知程度。

AI護理導師：本次共有 3 個訓練情境，請依序進行訓練

AI護理導師：第 1 個訓練情境，

護生：「OO 你好，我們現在有 20 分鐘的時間進行會談，今天想和你聊一聊你知不知道為什麼自己會來住在這裡。」

個案21：「恩。(沒有眼神接觸)」

AI護理導師：請選擇護生回覆，甲-(ngpp-1-1).說說看在來這裡住之前，發生了什麼事情？乙-(ngpp-1-1).你是怎麼進來這裡的呢？

護生：甲-(ngpp-1-1).說說看在來這裡住之前，發生了什麼事情？

AI護理導師：恭喜您！答對了！請選擇其中運用的技巧，甲-(ngpp-1-2).面質，乙-(ngpp-1-2).開放性問句，丙-(ngpp-1-2).給予認知，丁-(ngpp-1-2).簡述語意

護生：乙-(ngpp-1-2).開放性問句

AI護理導師：恭喜您！答對了！這裡使用的是開放性問句。



# 個人簡歷—校內外合作計畫

## ◆ 校內外合作計畫表

單位	計畫名稱
國科會	大專生研究計畫 -113-2813-C-040-039-B- 「TransMbD 失智監護輔助流程」
國科會	整合型研究計畫學習型助理 - 「促進聽覺障礙者多面向溝通的新世代科技」

單位	系統名稱
中山醫學大學 智慧健康暨資訊安全實驗室 彰化基督教醫院	「兒童髖關節偵測輔助系統」系統開發及維護
中山醫學大學 智慧健康暨資訊安全實驗室	「 <u>SunBone</u> 向日骨健 骨質疏鬆預防輔助系統」系統開發及測試
中山醫學大學 生醫訊號與健康照護實驗室 護理學系	「護理治療溝通技巧聊天機器人系統」系統維護及開發
中山醫學大學 (前)健康照護實驗室	「 <u>一圖勝千言</u> 溝通交流輔助系統」系統開發及測試





# 個人簡歷—競賽參與

◆ **競賽表現** - 曾參與**多場資訊競賽**，並獲得**多項佳績** (第一名\*2、第二名\*1、第三名\*1、佳作\*2)

主辦單位	競賽名稱	組別	名次
教育部	2024全國大專校院智慧創新暨跨領域整合創作競賽	數位永續科技組	佳作 值得注目獎
高雄大學工學院	2024教育部智慧晶片系統應用創新專題實作競賽	智慧健康	入選獎
亞洲大學	2024永續智慧創新黑克松競賽	柏拉罕：智慧共生	第一名
亞洲大學	2024永續智慧創新黑克松競賽	智齡科技：翻轉居家服務	第一名
高雄科技大學	2023第19屆全國電子設計創意競賽	電子組	第二名
高雄科技大學	2023第19屆全國電子設計創意競賽	資通組	第三名
高雄科技大學	2023第19屆全國電子設計創意競賽	智慧大數據及行動App組	佳作



# 個人簡歷—論文發表



## 論文發表紀錄

01

### 2023 International Conference on Fuzzy Theory and Its Applications (oral presentation)

#### Cognitive-Inclusive Communication: An Integrated Text and Graphics Generator for Improved Readability

作者：Yong-Xiang Chang, Chiun-Li Chin, Zhong-Ting Zhemg, Geng-Kai Wong, Su-Juan Chen, Guan-Tsen Liu, Pei-Hsin Chang, Zhe-Cheang Liu, Pei-Chen Huang, Pei-Xin Ye

02

### 2023 International Conference on Fuzzy Awareness Science and Technology (oral presentation)

#### DDA-SVM: A Dynamic Difficulty Adjustment SVM Model for Electronic Cognitive Training Games

作者：Yi-Wei Huang, Yong-Xiang Chang, Guang-Tao Lin, Pei-Chen Huang, Che-Cheng Liu, Kun-Zhe Zhou, Chiun-Li Chin





工業技術研究院

Industrial Technology  
Research Institute

# 其他與計畫相關有助於審查之文件

作者：張詠翔

就讀學校：中山醫學大學

就讀學系：醫學資訊學系

# THANK YOU FOR ATTENTION!!

2025/02/15