

## 數位仿生與傳承：利用語意辨識技術及大型語言模型達成文字使用習慣複製

許皓鈞<sup>1\*</sup>、黃暉哲<sup>2</sup>、陳秉佑<sup>3</sup>、楊凱丞<sup>4</sup>、戴文諺<sup>5\*</sup>、黃政嘉<sup>6</sup>  
<sup>123456</sup> 國立臺灣科技大學資訊管理系

<sup>1</sup>Hsu.haojiun@gmail.com、<sup>2</sup>wythehuang@icloud.com、<sup>3</sup>yoyochen2000@gmail.com、  
<sup>4</sup>benson9204215@gmail.com、<sup>5</sup>d11109101@mail.ntust.edu.tw、  
<sup>6</sup>jhengjia.huang@mail.ntust.edu.tw

### 摘要

隨著台灣獨居人口逐年增加，孤獨感逐漸成為一個潛在的社會問題，尤其在 20-44 歲年齡層中，超過 50% 未婚，單人戶數超過 300 萬。失去摯愛的痛苦常難以釋懷，而多數人渴望保留與逝去親友的情感聯繫，以此支撐面對未來的勇氣。在此背景下，本系統《數位仿生與傳承》應運而生，旨在協助使用者保存和重溫逝去親友的記憶。這是一個對話生成式 AI 工具，用戶可上傳逝者的聊天記錄，系統將透過對話生成技術和情感分析，訓練出一個模擬該對象的模型，讓使用者得以“與”他們對話，以緩解心靈創傷並支持情感上的復原。本篇論文提供 WebApp 形式，用戶可將聊天數據上傳至後台，系統利用機器學習、自然語言處理和情感分析等技術，重現目標對象的語言風格、情感特質和對話習慣，並生成高度擬真的互動體驗。這套系統不僅幫助使用者回憶摯愛，也能在心理治療和文學創作等領域應用，讓使用者在重溫過往對話的過程中獲得安慰與陪伴。

**關鍵詞：**生成式 AI、WebApp、機器學習、自然語言處理、情感分析、Llama3 模型、語言風格模擬

---

\* 通訊作者 (Corresponding author.)

# Digital Bionics and Heritage: Achieving Text Usage Habit Replication through Semantic Recognition Technology and Large Language Models

Hao-Jiun Hsu<sup>1\*</sup>, Wei-Che Huang<sup>2</sup>, Ping-Yu Chen<sup>3</sup>, Kai-Cheng Yang<sup>4</sup>, Wen-Yan Dai<sup>4</sup>, Jheng-Jia Huang<sup>5</sup>

<sup>123456</sup>National Taiwan University of Science and Technology, Department of Information Management

<sup>1</sup>Hsu.haojiun@gmail.com, <sup>2</sup>wythehuang@icloud.com, <sup>3</sup>yoyochen2000@gmail.com,

<sup>4</sup>benson9204215@gmail.com, <sup>5</sup>d11109101@mail.ntust.edu.tw,

<sup>6</sup>jhengjia.huang@mail.ntust.edu.tw

## Abstract

With the growing number of individuals living alone in Taiwan, loneliness has increasingly become a significant social issue. Among people aged 20-44, over 50% are unmarried, and single-person households now exceed 3 million. The pain of losing a loved one is often difficult to overcome, and many people long to maintain an emotional connection with their departed loved ones as a source of strength for facing the future. In response to this, the "Digital Bionics and Legacy" system was developed to help users preserve and revisit memories of lost loved ones. This conversational AI tool allows users to upload chat records with the departed, enabling the system to train a model that simulates the communication style of the lost loved one. Through dialogue generation technology and sentiment analysis, users can "converse" with them, alleviating emotional pain and supporting the healing process. This paper introduces the system as a WebApp, allowing users to upload chat data to the backend. The system then uses machine learning, natural language processing, and sentiment analysis, to replicate the linguistic style, emotional nuances, and conversational habits of the target person, creating a highly realistic interactive experience. This system not only helps users reminisce about loved ones but also has potential applications in fields such as psychotherapy and literary creation, providing comfort and companionship as users revisit past conversations.

**Keywords: Generative AI, WebApp, Machine Learning, Natural Language Processing, Sentiment Analysis, Llama3 Model, Language Style Simulation**

## 壹、前言

### 1.1 研究動機

2022年3月，美國精神醫學會正式將「延續性哀傷」(Prolonged Grief Disorder, PGD) 納入《精神疾病診斷與統計手冊》[5,7]。延續性哀傷是一種持續且過度延長的哀傷狀態，通常在至親的非自然死亡或驟逝（如自殺或意外事故）後發生。根據研究，孩童在事件發生約六個月後、成人約一年後可能進入此狀態。美國精神醫學會列出了幾項延續性哀傷的警訊，包括自我認同崩潰、不願接受事實、難以回歸日常生活和強烈的孤獨感。許多患者還會出現對酒精、香菸等物質的依賴，甚至有自殺傾向，顯示出延續性哀傷對個人和社會的深遠影響。

根據流行病學研究，約有10%的失親者可能罹患延續性哀傷，台灣的統計數據則顯示，全台單獨生活戶數已超過300萬，占家庭型態的34%；其中，20至44歲未婚者超過一半。這樣的社會結構使孤獨感成為重要的國家問題，尤其是獨居者在面對親人驟逝時，特別容易陷入延續性哀傷，進一步影響心理健康和日常功能[11,12]。儘管世界各國積極延長壽命，但自然死亡和事故死亡難以避免。突如其來的離世往往給家庭帶來衝擊，因此許多人希望能保留逝者的語言風格和文字習慣，以維持與摯愛的情感聯繫。

Xyngkou 和 Siriaraya [9]的研究指出，透過模擬逝者的聊天機器人，親人可以在回憶逝者的同時，適應現狀，並獲得面對未來的勇氣。這類聊天機器人還被證明對憂鬱症患者有幫助，能有效緩解症狀[1]。同時，隨著科技的進步，OpenAI 在2024年2月發布的 Sora 影音生成模型[3]，為生成情感豐富的數位內容帶來新的契機。結合 Sora 逼真的影音生成技術，本研究所創造的情緒豐富模型將能應用於影視市場，為影視創作帶來更多可能性，使得數位孿生技術在情感聯繫和心理健康方面發揮更大潛力。

### 1.2 研究目的

本研究旨在透過對話紀錄或聲音、影像資料，分析個人的語言風格和角色特徵，以建立一個能重現個人特質的模型。這樣的模型不僅能讓使用者在與模擬出的角色互動中懷念摯愛，提供情感上的支持，也能成為文學及影音創作者的寶貴資源，創建更真實且多樣化的角色，增強作品世界的互動體驗。為實現此目標，本研究將運用先進的自然語言處理技術和人工智慧模型，專注於逝者的語言紀錄分析，以重現其語言習慣。透過這個模型，單獨生活或有心理需求的人可以與已逝摯愛的數位模擬角色互動，從中緩解孤獨感和悲傷，在情感聯繫中找到慰藉。這項技術不僅是一種情感支持工具，也為文學及影視創作帶來更多角色塑造和情感表現的可能性，為創作者提供更豐富的靈感和創作手

法。

## 貳、文獻探討

### 2.1 機器學習 (Machine Learning, ML)

機器學習是一門科學，旨在開發演算法及統計模型。開發出的演算法及模型可以讓電腦根據原始訓練資料所整理出的模式去推理並執行任務，其中不需要人類去給予任何指令。常見的演算法如隨機森林法 (Random forest) [2]、支持向量機 (SVM) [4]、以及模仿人類大腦所研發出的深度學習 (Deep Learning) [6] 等等。

### 2.2 自然語言處理 (Natural Language Process, NLP)

自然語言處理 (Natural Language Process, NLP) 是一種機器學習技術，目的為讓只有數字的電腦可以理解複雜的人類語言並給予回應，往下細分可分為自然語言理解 (Natural Language Process, NLU) 以及自然語言生成 (Natural Language Generation, NLG)。如林子平 [10] 研究的聊故事機器人即是自然語言處理的應用之一。

### 2.3 情感分析 (Sentiment Analysis, SA)

情感分析 (Sentiment Analysis, SA) 指的是使用自然語言處理 (NLP) 技術來判斷內容的情緒傾向，如正面、負面或中性。透過機器學習對被模仿者過去的文字聊天記錄進行分析，可以推測特定語句的情緒指標，並運用自然語言生成技術模擬被模仿者在當下情境中的情感。本研究的核心目標是達到高擬真度，因此文本中情感的真實性必須得到充分體現，情感分析技術因此成為本研究的重要關鍵。

### 2.4 大型語言模型 (Large Language Model, LLM)

大型語言模型 (LLM) [8] 是一種具有超過千億參數的深度學習模型，通過大量文本數據的訓練，使其能夠從龐大的知識庫中進行識別、匯總、預測及生成文字等操作。其中最知名的例子是由 OpenAI 推出的聊天式預訓練轉換器 ChatGPT。大多數 LLM 首先使用通用且大量的數據進行預訓練 (Pre-training)，之後根據不同應用需求進行微調 (Fine-tuning)。本研究將採用相同的方法，並計劃對 Meta 開發的 LLaMA3 模型進行微調，以滿足研究需求。

## 2.5 Large Language Model Meta AI 3 (LLaMA 3)

LLaMA 3 是 Meta 於 2024 年 4 月 推出的開源大型語言模型，參數量高達 700 億，並使用 15 萬億 tokens 訓練數據。相比上一代，LLaMA 3 提升了語言理解和指令遵循能力，應用範圍涵蓋語言處理、代碼生成等。該模型延續了開源和商業化支持，為開發者提供更強大的工具，推動 AI 技術進一步發展。對於本研究系統開發非常重要。

## 參、方法

### 3.1 訓練資料

本研究採用台灣市占率最高的通訊軟體 LINE 作為聊天記錄資料的主要來源，因其匯出聊天資料檔的功能簡單明瞭，方便易用。初期資料將由研究小組成員間的對話記錄提供，並通過問卷徵求更多自願者上傳聊天記錄，進一步擴充數據量，以投入模型訓練，使模型訓練更為完整。

### 3.2 實驗前處理

本研究的系統採用切詞、斷詞等技術對聊天室資料的原始檔案進行資料清理與格式轉換，並通過關鍵詞檢查，確保使用者提供的資料不含腥羶色或不雅言論。系統會分析其中的情緒與語氣，適當過濾並保留無傷大雅的語助詞，以避免過度刪減影響生成結果的相似度，使結果更符合需求。

### 3.3 模型應用

由於從無到有地訓練完整的大型語言模型需耗費大量時間與資源，採用開源的大型語言模型進行微調是更為妥當且資源節省的选择。本研究已採用上文提到的 LLaMA 2 進行微調，使其能模仿指定被模仿者的用字習慣，達成更真實的語言模擬效果。

### 3.4 架構設計

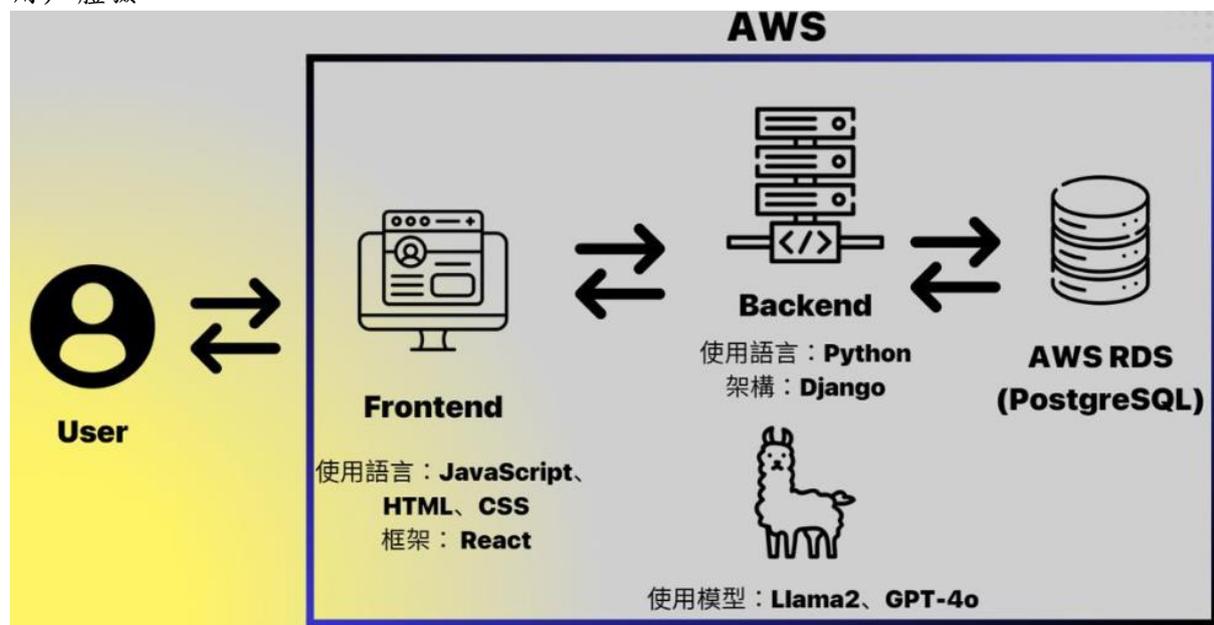
本系統的架構設計(如圖一)包含生成文字的自我評分機制。在生成文字傳送給使用者之前，系統進行自我評分，評估項目涵蓋情感表達、用詞恰當性及模仿精確度，以確保生成內容符合要求。若評分結果達標，系統便將內容傳送給使用者；若未達標則進行重新生成。



圖一 架構設計圖

### 3.5 系統架構與結果

本系統架構(如圖二)採用前後端分離的傳統架構，前端開發語言為 HTML、CSS 和 JavaScript，並以 React 框架建構用戶界面，以確保良好的互動性和使用者體驗。後端部分則使用 Python 語言，搭配 Django 框架進行伺服器端的邏輯處理和資料管理。本系統的核心模型基於 Llama3，從而精確模擬目標對象的語言風格，使對話更具個人化和真實感。為達成高度仿真的對話體驗，系統採用了 NLP 技術，對使用者上傳的文本資料進行深度語義分析，提取其中的語氣、用詞風格和情緒基調等特徵，使系統能夠更貼近真實的語言表達與情感傳遞。此架構和技術選擇共同確保了系統的精確性、靈活性與用戶體驗。



圖二 系統架構圖

## 肆、結論

本研究成功創建了一個陪伴使用者的虛擬角色，作為理想互動對象，提供情感豐富的對話體驗。系統以網站為互動介面，使用者可於平台上上傳資料進行訓練，後台透過 Celery 分散式處理技術來微調 LLaMA 3 模型，生成具個人語言風格的模型，達成數位傳承。該系統能夠學習情感表達與語意習慣，模擬目標對象的溝通模式，營造如同真人互動的情境，滿足使用者的情感需求，帶來心靈上的慰藉。系統具備擴展至多種應用場景的潛力，無論是作為教育工具提升溝通技巧和自信，還是輔助文學及影視創作中的角色模擬，都能增強真實感。同時，系統的應用已延伸至醫療領域，與醫療機構合作，協助患者紓解延緩性哀傷症狀（如憂鬱和焦慮），達到情緒緩解的療效。互動平台的核心功能包括對話記錄匯入與訓練，使用者上傳對話記錄後，系統會持續訓練模型，最終透過網站呈現生動、逼真的互動體驗，實現本研究的主要目標。

## 參考文獻

- [1] L. Athota, V.K. Shukla, N. Pandey, and A. Rana. "Chatbot for Healthcare System Using Artificial Intelligence," in *2020 8th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO)*. 2020.
- [2] L. Breiman, "Random Forests," *Machine Learning*, 2001. 45(1): p. 5-32.
- [3] Brooks, T., B. Peebles, C. Homes, W. DePue, Y. Guo, L. Jing, D. Schnurr, J. Taylor, T. Luhman, E. Luhman, C.W.Y. Ng, R. Wang, and A. Ramesh, *Video Generation Models as World Simulators*. 2024.
- [4] N. Cristianini, and E. Ricci, "Support Vector Machines," in *Encyclopedia of Algorithms*, M.-Y. Kao, Editor. 2008, Springer US: Boston, MA. p. 928-932.
- [5] G. Holly Prigerson, Ph.D., Jeffrey Bridge, B.S., Paul K. Maciejewski, Ph.D., Laurel C. Beery, B.S., Robert A. Rosenheck, M.D., Selby C. Jacobs, M.D., M.P.H., Andrew J. Bierhals, M.D., M.P.H., David J. Kupfer, M.D., and David A. Brent, M.D., "Influence of Traumatic Grief on Suicidal Ideation Among Young Adults," *American Journal of Psychiatry*, 156(12):1994-5, 1999 Dec.
- [6] I.H. Sarker, "Deep learning: A comprehensive overview on techniques, taxonomy, applications and research directions, Taxonomy, Applications and Research Directions," *SN Computer Science*, 2021. 2(6): p. 420.
- [7] Sekowski, M. and H.G. Prigerson, "Associations between symptoms of prolonged grief disorder and depression and suicidal ideation," *The British journal of clinical psychology*, 2022. 61(4): p. 1211-1218.
- [8] H. Touvron, L. Martin, K. Stone, P. Albert, A. Almahairi, Y. Babaei, N. Bashlykov, S. Batra,

- P. Bhargava, and S. Bhosale, “Llama 2: Open foundation and fine-tuned chat models,” *arXiv preprint*, arXiv:2307.09288, 2023.
- [9] A. Xyngkou, P. Siriaraya, A. Covaci, H.G. Prigerson, R. Neimeyer, C.S. Ang, and W.-J. She, “The ‘Conversation’ about loss: Understanding how chatbot technology was used in supporting people in grief,” in *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 2023, Association for Computing Machinery: Hamburg, Germany. p. Article 646.
- [10] 林子平, “基於大型語言模型應用指示詞打造無程式碼對話系統平台-以聊故事機器人為例,” in 資訊工程學系. 2023, 國立中央大學:桃園縣. p.49.
- [11] 黃惠聆, 單身社會時代來臨 近六成台灣女性積極理財 亞太區最高, 2023; Available from: <https://www.ctee.com.tw/news/20230821700803-439803>.
- [12] 親友過世走不出傷痛? 醫教當事人、陪伴者應對「延續性哀傷障礙」。 2022; Available from: <https://gooddoctorweb.com/post/1440>.
- [13] <https://tnten.co/learning/llama-3-by-meta-ai/>