



Make Oracle Bones Great Again：川普遇上甲骨文 AI
Make Oracle Bones Great Again: Trump Meets Oracle AI

作者：王祁偉、王彤云、陳靖宜、楊宛霓、姚佩甄

系級：外文二、中文二、中文二、中文二、會計二

學號：D1356421、D1362757、D1390205、D1362727、D1380438

開課老師：曾智義

課程名稱：生成式人工智慧的人文導論

開課系所：人工智慧自然語言技術學程

開課學年： 113 學年度 第 2 學期

中文摘要

本研究的目的是在於回應古文字學習門檻過高、教材缺乏互動性與趣味性的問題，嘗試運用人工智慧技術，降低學習者進入古文字世界的障礙，並探索如何以數位化與互動化的方式推動教育普及與文化傳承。研究團隊希望能透過影像辨識、語音合成與生成式內容創作，使古文字不再只是專業領域的研究對象，而能成為一般學習者易於接觸且富有吸引力的學習資源。

在研究過程與方法上，我們採用模組化設計，將系統劃分為七個步驟，涵蓋從環境建置到成果展示的完整流程。首先，建立深度學習與生成式 AI 的基礎環境，並進行 API 連結與硬體檢測，以確保系統可順利運行。接著，透過影像辨識技術，將甲骨文、金文等古文字字形與現代漢字比對，並輔以語音生成模組，提供字形演變與語義的語音導讀。為了提升趣味性，研究亦引入生成式 AI，創造古文字相關的文本、笑話與插畫，並進一步利用影音生成與合成技術，製作成數位繪本與短片。整體方法結合了深度學習影像辨識、TTS 聲音克隆、文本創作、插畫生成與影音後製，展現跨領域的整合應用能力。

研究結果顯示，本系統成功將古文字轉化為兼具「可視、可聽、可互動」的多模態學習資源，能幫助學習者更快理解字形與意義，並藉由趣味化內容提升學習動機與記憶效果。此成果不僅適用於中文母語者，也能作為僑生與外籍學習者的輔助資源，未來更可應用於教育機構、博物館及語言學習平台。整體而言，本研究不僅是一項技術實驗，更是一項結合教育、文化與科技的應用實踐，展現人工智慧於人文教育與文化傳播上的潛力。

關鍵字：

人工智慧、古文字、文化傳承、影像辨識、語音合成

Abstract

The purpose of this study is to address the challenges of high entry barriers, lack of interactivity, and limited engagement in ancient character learning. By applying artificial intelligence technologies, the research seeks to lower the obstacles for learners entering the world of ancient characters and to explore how digitalization and interactivity can promote educational accessibility and cultural transmission. The research team aims to employ image recognition, speech synthesis, and generative content creation so that ancient characters are no longer confined to the domain of specialists but can become accessible and appealing learning resources for general learners.

In terms of process and methods, the study adopts a modular design, dividing the system into seven steps that cover the complete workflow from environment setup to final presentation. First, a foundation for deep learning and generative AI is established, along with API integration and hardware testing, to ensure smooth system operation. Next, through image recognition technologies, ancient scripts such as oracle bone inscriptions and bronze inscriptions are compared with modern Chinese characters, supplemented by a speech synthesis module that provides audio narration on character evolution and meaning. To enhance engagement, the study incorporates generative AI to create texts, jokes, and illustrations related to ancient characters, and further applies video generation and editing techniques to produce digital picture books and short films. The overall methodology integrates deep learning-based image recognition, TTS voice cloning, text generation, illustration creation, and multimedia post-production, showcasing the capacity for interdisciplinary applications.

The results indicate that the system successfully transforms ancient characters into multimodal learning resources that are “visual, audible, and interactive,” helping learners more quickly grasp both form and meaning while boosting motivation and retention through engaging content. These outcomes are valuable not only for native Chinese speakers but also for overseas Chinese and foreign learners, and the system holds potential for application in educational institutions, museums, and language learning platforms. Overall, this research is not only a technical experiment but also an applied practice that integrates education, culture, and technology, demonstrating the potential of artificial intelligence in the fields of humanities education and cultural dissemination.

Keyword :

Artificial Intelligence (AI) 、 Cultural Heritage 、 Image Recognition 、 Speech Synthesis

目 次

壹、議題與問題說	3
一、研究動機.....	3
二、研究目的.....	4
三、問題構想.....	5
四、小結.....	6
貳、模型及技術運用說明	7
一、程式操作影片.....	7
二、各單元介紹.....	7
三、小結.....	11
肆、執行結果	11
一、Cell 5 輸出之結果實例.....	11
二、Cell 6 輸出之結果實例.....	11
三、Cell 7 輸出之結果實例.....	11
陸、未來展望	13

壹、議題與問題說明

一、研究動機

古文字不僅是漢字的早期形態，更深刻承載了數千年來中華文化的思想體系、審美觀念與歷史演進軌跡。從甲骨文的卜辭記錄到金文的宗族記事，古文字體系內蘊豐富的社會脈絡與文化意涵，其不僅見證了漢字演變的過程，更提供我們理解古人世界觀、宗教信仰、社會組織及價值取向的重要視角。漢字作為中華文明最核心的文化載體之一，其形體、結構與意義的生成與變化，實際上映照了古代社會的發展脈絡與人文精神。因此，學習古文字不僅是語言溯源的探索，更是培養歷史意識、深化文化理解與建立人文素養的關鍵途徑。

然而，當前古文字的學習門檻偏高，主要資源多集中於高等教育與專業學術領域。對於一般學習者而言，甲骨文與金文在字形結構上與現代漢字有著極大的差異。甲骨文作為中國最早的成系統文字，字形或取象於自然萬物，或用符號指抽象事物，與後世楷書的方整結構迥異，雖然具有高度研究價值，卻也造成學習難度。加之相關教材多偏重於專業訓詁與考釋，充斥大量古文字學術術語，往往使初學者感到晦澀難懂，進而在學習初期就喪失動力。

此外，現有古文字的教學與學習資源主要以紙本文獻與傳統課程為主，雖然能夠維持學術研究的嚴謹性，但缺乏互動性與趣味性。尤其在資訊化時代，大多數學習者已習慣於透過數位媒介獲取知識，若教材仍停留於靜態圖像與文字解說，容易使人產生距離感，甚至覺得古文字學習與日常生活完全脫節。這種脫節，使得古文字的教育長期侷限於少數研究圈層之中，難以普及至社會大眾。其結果不僅是學習人口的縮減，也使得古文字所承載的文化價值逐漸失去社會關注，形成一種學術孤立的現象。

進一步而言，古文字的教育困境不僅反映在知識傳遞的形式上，也涉及學習心態的培養。許多學習者在接觸古文字時，往往僅視其為考古學或古典文獻學的專業工具，而忽略了古文字背後所蘊含的文化趣味與審美價值。例如，甲骨文中的「日」「月」「水」「山」等字形，往往具有高度的圖像性，若能適切轉化為學習媒材，將能激發學習者的想像力與探索興趣。但在現行教材中，這類「圖像—符號—意義」的轉換往往被過度學術化處理，而缺乏親切、生動的呈現方式，致使一般學習者難以在情感與認知層面建立連結。

面對這樣的侷限，如何結合現代科技，以突破傳統教學模式，成為當前值得關注的課題。人工智慧技術的快速發展，為古文字學習的普及化提供了新的契機。近年來，圖像辨識、深度學習與自然語言處理等技術已廣泛應用於文字識別與教育場域，例如現代文字的手寫辨識、語音轉譯以及智慧教學系統等，均展現了人工智慧在輔助人類學習上的巨大潛能。若能將這些技術引入古文字領域，將可協助學習者突破字形辨識與語義理解的瓶頸，降低入門難度，並提供更直觀、互動的學習體驗。

因此，本研究希望透過人工智慧技術的導入，建立一套能輔助學習者辨識、理解並應用古文字的工具。透過 AI 的圖像識別功能，可以自動將甲骨文或金文的字形與現代漢字進行比對，幫助學習者快速建立字形與意義的聯繫；透過語義分析與知識圖譜的結合，則能將單一字詞置於更完整的文化與歷史脈絡中加以呈現；再配合互動式介面設計，能讓使用者在操作過程中獲得即時回饋與引導，提升學習效率與參與感。藉此，不僅能突破古文字學習「艱澀、專業、難以入門」的刻板印象，也能使古文字教育真正走出學術殿堂，走向更廣泛的社會大眾。

總而言之，古文字作為中華文化的重要資產，其價值不應僅被少數研究者掌握，而應透過現代科技的輔助，使更多人能夠接觸、理解並欣賞其中蘊含的智慧與美感。本研究的動機，正是基於「降低門檻、擴大參與」的理念，藉由 AI 技術為古文字教育創造新的可能，推動其普及化與大眾化，並進一步促進文化傳承與人文素養的提升。

二、研究目的

為了降低學習障礙、提升學習趣味性與互動性，本專題嘗試結合人工智慧技術，設計一個集古文字辨識、語音導讀與生成式內容為一體的互動學習程式，使古文字「看得懂」、「聽得到」、「玩得起」，進而促進其教育價值與文化推廣潛力。

透過本專題的設計與應用，我們期望實現以下三項目標：

（一）針對以中文為母語的學習者，透過視覺化與語音化的解釋方式，促進其對古文字的文化理解與認同，加強與自身語言根源的連結。

（二）針對學童或尚在學習中文的僑生，以更生動活潑的方式呈現古文字的字形演變與語義內涵，作為國語文教育的輔助資源，為中文學習奠定更深厚

的文化基礎。

(三) 針對非華語背景的外籍使用者，以趣味性與互動性為核心，結合遊戲化的學習體驗，引導其初步認識與興趣建立，進而作為學習中文與理解東亞文字文化的起點。



圖一

對此，本專題在研究動機與研究目的的基礎上，將實際建構一套結合人工智慧技術的古文字互動學習系統。此系統不僅在技術層面上實現了古文字的辨識、語音生成與內容創造，更在教育應用層面上嘗試突破傳統古文字教材的侷限，提供一種更具趣味性、互動性與延展性的學習方式。以下將就專題主要內容加以說明：

(一) 圖片輸入與漢字辨識功能

使用者可透過上傳圖片的方式，將含有古文字的圖片（如甲骨拓片、青銅器銘文影像）匯入程式。系統結合深度學習的圖像辨識技術，能自動判讀出其中的古文字，並將之與現代漢字對應。這項功能的核心意義在於降低古文字的入門障礙。以往學習者必須依賴專業字典或釋文資料，耗費大量時間比對，而平台能透過 AI 的快速辨識縮短學習距離，使初學者能在短時間內對古文字有清晰直觀的理解。

(二) AI 語音解說與多模態學習

在辨識文字的基礎上，本研究進一步整合語音生成技術。使用者不僅能看到古文字與對應的現代漢字，還能聽到系統所生成的語音解說。語音內容涵蓋字形演變、基本語義、相關文化背景等，並以淺顯易懂的方式呈現，適合不同年齡與程度的學習者。這樣的設計突破了傳統紙本教材單一文字資訊的限制，讓學習過程同時結合「視覺」與「聽覺」，促進多模態的知識吸收，也更符合現代數位學習者的習慣。

(三) 生成式內容：古文字笑話

為了提升學習趣味性與互動感，本平台設計了生成古文字笑話的功能。系統透過語言模型，將古文字的字形特徵或語義內涵與幽默元素結合，生成簡短的笑話內容。例如，以「馬」字的甲骨文為靈感，生成與其甲骨文字形相關的幽默語句。這些笑話不僅能緩和學習過程的嚴肅氛圍，也能讓學習者在輕鬆的語境中記憶古文字，達到寓教於樂的效果。更重要的是，這種創新設計提供了一種古文字「活化」的方式，使之不再僅是歷史遺存，而是能夠與當代學習者的日常語境產生連結。

(四) 笑話圖象化與繪本化

在生成笑話的基礎上，進一步將其圖象化、繪本化。透過圖像生成技術，系統能根據笑話的內容創造相應的插畫，將文字轉化為視覺故事。當這些插畫進一步被組織成繪本形式時，學習者便能透過閱讀圖文並茂的故事，更直觀地理解古文字的形義特色與文化內涵。這一功能特別適合學童或語言初學者，因為圖像化能減少文字解釋上的抽象感，使學習過程更具吸引力。對於非華語背景的外籍學習者而言，繪本化的內容更能跨越語言隔閡，以視覺符號建立學習基礎。

(五) 教育與文化推廣的潛力

上述成果的整合，展現出 AI 技術在古文字教育中多層次的應用可能性。一方面，透過辨識與語音解說，此系統能幫助學習者快速建立基礎知識；另一方面，藉由笑話生成與繪本化，則能提升學習的趣味性與互動性。這樣的多元設計使 AI 不僅是一套工具，更是一種教育理念的實踐：即將古文字學習轉化為富含娛樂性與文化性的數位體驗。其潛在影響不僅限於學術領域，也能拓展至國語文教育、文化推廣，甚至國際華語教學的場域。

綜上所述，本專題希望透過人工智慧的導入，為古文字教學提供新的可能性。其

核心價值在於突破傳統學習框架，讓古文字「看得懂」、「聽得到」、「玩得起」，進而讓更多學習者在親近與愉快的氛圍中理解並欣賞古文字之美。

三、問題構想

如前述，在當代數位人文研究與文化教育的脈絡中，古文字作為中國文化的重要符號系統，承載了豐富的歷史、語言與審美價值，其保存與傳承具有重要意義。然而，現有古文字學習資源主要以靜態查詢系統或專業教材為主。這類資源雖具備完整的字形資料、詞源解釋及學術深度，但在學習情境的設計上往往缺乏互動性與生活化應用。對於一般學習者，特別是初學者而言，甲骨文、金文等字形異於現代漢字系統，呈現出高度抽象化特徵，導致辨識困難與學習門檻偏高。即使學習者對古文字抱有興趣，也可能因缺乏直觀的導引與解說而產生距離感，進而影響學習動機與持續性。

針對上述問題，本專題提出設計一套結合人工智慧技術的互動學習系統。系統核心理念在於透過影像辨識、語音導讀以及生成式內容，將古文字由單一的靜態符號轉化為可讀、可聽、可互動的學習媒介。其目的在於降低學習障礙、提升學習趣味性，並促進古文字教育的普及與文化傳承。



圖二

本系統主要規劃三個模組，說明如下：

(一) 古文字影像辨識

古文字影像辨識模組為系統的基礎功能。使用者只需上傳古文字圖片，系統即可自動判定該字所屬的文字類型（如甲骨文、金文、篆書等），並輸出對應的現代楷書字形、標準字音、基本字義以及筆畫數。此模組旨在提

供學習者對古文字的直接認知，讓使用者能快速了解每個字的結構及基本意涵，為後續語音導讀與繪本生成奠定知識基礎。與傳統專業查字方式相比，本模組降低了學習門檻，無需專業背景即可掌握基本字形與意義。

(二) 語音導讀

在完成基礎辨識後，系統提供語音導讀功能，使文字不再僅依賴視覺呈現。語音內容包括該字的標準發音、基本字義及相關文化背景說明，幫助使用者加深理解與記憶。此模組亦支援趣味化語音選擇，使用者可以選擇不同語音風格或模擬特定角色的聲音，以增加互動性與學習參與感。語音導讀建立在模組一提供的字形字義基礎上，形成從視覺認知到聽覺理解的連貫學習流程。

(三) 生成式數位繪本

生成式數位繪本模組為系統的進階功能，其內容與語音導讀模組中的趣味化元素（如笑話或小故事）相互關聯。系統依據使用者在模組一中辨識出的字形與字義，以及模組二所選擇的語音風格，自動生成對應的短篇故事或插畫情境，形成互動式數位繪本。此模組將學習過程由單純的認知與聽覺理解，延伸至視覺化與情境化的多模態學習，促進學習者對古文字的記憶與文化理解。透過三個模組的遞進設計，使用者可從基礎字形字義出發，經由語音導讀，再延伸至互動故事與圖像繪本，形成完整且連貫的古文字學習體驗。



圖三

四、小結

綜合而言，本專題設計之系統試圖在專業性與大眾可近性之間建立橋梁。影像辨

識、語音導讀及生成式數位繪本三大模組整合，能將古文字從高門檻的研究材料轉化為具有互動性、生活化的學習資源。系統特色包括降低學習門檻，使初學者可迅速進入學習狀態；提升互動性與參與感，使學習過程更符合現代數位學習者習慣；強化文化教育推廣，使古文字知識能在教育場域及公眾領域廣泛應用。本系統旨在突破傳統靜態資料庫限制，使古文字不僅是研究對象，亦成為可親近、可互動的學習媒介，進而兼顧知識傳授與文化傳承雙重目標。

貳、模型及技術運用說明

一、[程式操作影片](#)

二、各單元介紹

在本專題系統的設計中，我們依循模組化與層層遞進的邏輯，將整體開發流程劃分為七個 Cell，每一單元皆承擔特定功能，並在技術上與方法論上相互銜接。此一設計不僅有助於降低開發過程中的複雜度，也確保了研究成果能夠以可擴充、可追溯的方式推進。以下將逐一概述各 Cell 的角色與任務。

Cell 1 的核心在於「環境建置與依賴安裝」，其功能為安裝並配置深度學習框架、語音合成與圖像生成工具，以及其他輔助資源。此一階段確立了運算環境的相容性與穩定性，使後續模組能在統一的基礎上運行。

Cell 2 則進一步處理「外部資源連結與資料管理」。其主要內容包括 API 金鑰的設置、運算設備檢測，以及數據掛載與預處理。此一模組確保研究系統能與外部模型交互，並且能靈活運用硬體資源，為資料輸入與模型推理建立橋梁。

Cell 3 屬於「圖像資料處理與古文字辨識」階段。該模組主要負責接收使用者輸入的古文字圖片，透過深度學習模型進行特徵擷取與字形比對，並輸出對應的現代楷書字形、字音、字義及筆畫數。此模組的重點在於將高門檻的專業知識轉譯為直觀可讀的資訊，為後續多模態擴展奠定基礎。

Cell 4 的任務是「語音合成與導讀生成」。在辨識結果的基礎上，系統能以語音方式提供字音與字義解說，並透過 TTS 技術支援個人化或特定風格的語音模擬。此舉既擴展了學習的多感官路徑，也降低了文字輸入的門檻，使學習過程更具互動性。

Cell 5 則將功能延伸至「趣味化文本生成」。基於前一模組的語音輸出，系統能生成與古文字相關的笑話、故事或短篇文本，強化學習者的情境感與參與度。此模組標誌著系統由基礎知識傳遞轉向文化敘事與娛樂互動。

Cell 6 的焦點在於「圖像與繪本生成」。其設計使先前生成的文本內容能夠透過影像化處理轉化為插圖或數位繪本，並與語音導讀相結合，形成多模態的學習資源。此模組進一步提升了古文字學習的可視化與情境化程度，拓展了應用場域。

Cell 7 則屬於「整合與展示」階段。它負責將前述各模組的輸出進行串聯與排布，生成完整的互動學習成果，例如影音繪本、互動展示頁面或可分享的數位內容。此模組的設計目的在於確保系統功能能夠以一體化的形式呈現，並具備可供教學、研究與推廣使用的應用價值。

整體而言，七個 Cell 的規劃呈現出由環境建置、資料輸入，到識別、生成，再到整合展示的邏輯鏈條。此種架構不僅符合系統開發的技術需求，也體現了研究在方法論上的層次性與系統性，為古文字數位學習的創新應用提供了完整的技術路徑。接下來的章節將針對各 Cell 的設計細節與運行機制進行逐一說明，以進一步展現本研究系統的整體運作與教育潛能。

(一) Cell 1: 環境設置與依賴安裝

本單元旨在搭建和配置運行多功能 Notebook 所需的軟件環境與依賴庫。技術與模型概覽如下：

- 包管理器：pip，用於安裝與管理 Python 庫。
- 核心計算庫：PyTorch (包含 torch、torchvision、torchaudio)，作為深度學習計算的基礎框架，對後續的語音合成 (TTS)、圖像生成及視頻生成模型至關重要。通常會指定較新版本以支持最新功能。
- 語音合成 (TTS) 相關：f5-tts (用於高質量聲音克隆)、gTTS (Google Text-to-Speech Python 庫，作為簡單英文 TTS 或備選)、openai-whisper (主要用於自動語音辨識，但可引入音頻處理依賴)。
- 圖像生成 (Diffusers) 相關：diffusers (Hugging Face 庫，用於運行各種擴散模型)、transformers (Hugging Face 核心庫，提供多種預訓練模型與 Pipeline)、accelerate (簡化在不同硬件上運行 PyTorch 代碼)、compel 與 invisible-watermark (用於 Prompt 加權與水印處理)。
- 大型語言模型 (LLM) 相關：google-generativeai (Google Gemini API 的 Python SDK，用於古文字辨識、文本生成與 Prompt 創建)。
- 視頻處理相關：moviepy (用於視頻編輯與合成)、imageio 與 imageio-ffmpeg (MoviePy 依賴，用於圖像與視頻讀寫)。

Make Oracle Bones Great Again：川普遇上甲骨文 AI

- 其他輔助庫：Pillow(圖像處理)、numpy(數值計算)、soundfile/wavio(音頻文件處理)、requests(網絡請求)、gradio(用於 UI)、datasets、pandas 等。
- 系統工具：ffmpeg，用於音視頻編解碼。

此單元執行升級 pip、卸載舊版本庫、安裝/升級核心與輔助庫、檢查版本與 F5-TTS 命令行工具可用性。首次成功執行後，通常需重新啟動 Colab 執行階段以確保所有依賴正確加載。

(二) Cell 2: API 金鑰、設備檢查、Drive 掛載與數據準備

本單元負責項目運行前的初始化，主要操作包括：

- API 交互：配置 Google Gemini API，並透過 Colab Secrets 讀取安全的 API 金鑰。
- 硬件加速：利用 PyTorch (torch.cuda) 檢測與選擇 GPU 或 CPU 作為計算設備。
- 雲存儲整合：掛載 Google Drive 以訪問存儲的參考音頻文件。
- 數據集處理：使用 Hugging Face datasets 讀取與處理數據集(如 tuenguyen/trump-speech-dataset-tts)，但核心 F5-TTS 參考音頻主要來源於 Drive。

本單元設定 API 金鑰、檢查 GPU 可用性、初始化模型緩存變數、掛載 Drive，並定義參考音頻路徑與可選的 Hugging Face Dataset。

(三) Cell 3: 核心功能函數定義

本單元集中定義項目中所有核心功能函數，涵蓋以下技術：

- Google Gemini API 調用：古文字辨識、文本評估與修正、單人與多角色故事或笑話生成、靜態圖像與視頻 Prompt 生成。
- 文本處理與解析：古文字辨識結果解析、多角色對話腳本解析及正則表達式清理。
- 語音合成 (TTS)：F5-TTS 聲音克隆及 gTTS 作為備選。
- 圖像生成 (Diffusers)：Stable Diffusion 或其他模型生成靜態圖像，包含模型加載、緩存與顯存優化。

Make Oracle Bones Great Again：川普遇上甲骨文 AI

- 動態視頻生成 (Zeroscope via Diffusers)：加載模型並基於文本生成短視頻幀序列。
- 視頻/音頻編輯 (MoviePy)：靜態圖像與音頻合成視頻、多音頻片段拼接、多視頻片段合成完整音軌。
- 其他技術：圖像處理 (Pillow)、文件與系統操作 (os、shutil、pathlib、subprocess)、Colab 特定功能 (google.colab.files)、輸出顯示 (IPython.display)。

本單元提供圖像處理與上傳、古文字辨識與解析、LLM 內容創作、多媒體元素生成、語音合成、靜態與動態圖像生成，以及音視頻合成與處理的完整工具函數，為後續主處理流程提供基礎支持。

(四) Cell 4: 初始化全局狀態變數

本單元主要初始化或重置在多個處理流程中需要共享的全局變數，例如：

- current_char_info (存儲當前古文字辨識結果)
- last_recognition_text (上一次辨識返回文本)
- uploaded_image_pil、image_filename、image_data (上傳圖像相關信息)
- loaded_pipeline、loaded_pipeline_name (Diffusers 模型緩存)
- device (全局計算設備確認)

此單元確保主處理循環開始前擁有乾淨初始狀態。

(五) Cell 5: 主處理循環 - 古文字探索與多媒體創作

本單元為項目的交互與處理中樞，集成 Cell 3 的各項功能，主要步驟如下：

1. 語言選擇 (中文或英文)
2. 主要聲音風格選擇
3. 圖像上傳與古文字辨識
4. 描述流程：生成文字描述、語音合成、圖像 Prompt 生成及靜態圖像生成，並可將音頻與圖像合成短視頻
5. 單角色笑話流程 (可選)：生成角色笑話文本、語音與圖像，並合成短視頻

Make Oracle Bones Great Again：川普遇上甲骨文 AI

6. 多角色對話式笑話流程(可選):生成多角色對話文本、逐句語音合成、音軌拼接及圖像生成，最終合成短視頻
7. 循環與退出：控制新會話或退出程序

此單元實現從古文字圖像輸入到多媒體內容輸出的完整交互流程。

(六) Cell 6: AI 說書人 - Zeroscope 動態視頻創作

本單元專注於生成包含動態視頻片段的故事影片，主要技術與功能包括：

- 核心視頻生成模型：Zeroscope (通過 Diffusers Pipeline)
- 用戶交互與流程控制 (語言選擇、聲音風格、故事類型)
- 內容生成：單人旁白故事或多角色對話式故事
- 語音處理：逐句 TTS 生成與音軌拼接
- 視頻幀處理與合成：將 Zeroscope 生成的圖像幀序列導出為 MP4，並與旁白音軌合成最終視頻

Cell 6 創建動態故事影片，提供更具動感的視覺敘事體驗。

(七) Cell 7: AI 說書人 - 多圖靜態影片創作

本單元通過多張靜態圖像與完整旁白音頻呈現故事，功能概覽如下：

1. 語言、聲音與故事類型選擇
2. 故事文本生成 (單人旁白或多角色對話)
3. 旁白音頻生成
4. 多圖像 Prompt 生成：將故事文本分解為關鍵場景並生成圖像 Prompt
5. 圖像迭代生成與用戶選擇
6. 靜態圖視頻合成：將選定圖像與旁白音頻合成 MP4 視頻
7. 輸出與清理：顯示最終視頻並清理臨時文件

Cell 7 提供與 Cell 6 不同的視覺風格，專注於透過靜態圖像序列輔助故事敘述。

三、小結

Make Oracle Bones Great Again：川普遇上甲骨文 AI

綜上所述，本專題將七個 Cell 按照功能劃分並依序串連，形成一個由基礎環境建置、資料輸入與處理，到語音與文本生成，再至圖像化與成果整合的完整技術流程。此一設計不僅降低了古文字學習的門檻，也使得原本靜態的知識資源能以多模態的方式呈現，兼具學術性與趣味性。透過此流程，我們得以將古文字的字形、語音、語義與文化內涵轉化為更具體、可感知的數位學習體驗，進一步提升其教育應用與文化推廣的可能性。

參、Python Coding

一、完整 Python Coding 及說明

[Github 連結](#)

肆、執行結果

一、Cell 5 輸出之結果實例



圖四 [介紹影片](#)



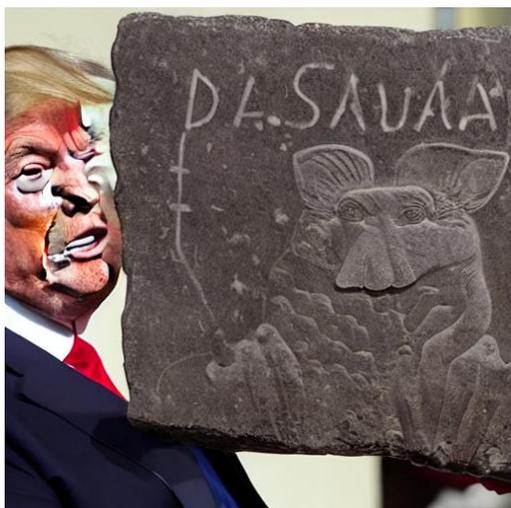
圖五 [笑話影片](#)

二、Cell 6 輸出之結果實例



圖六 [川普與歐巴馬對話影片](#)

三、Cell 7 輸出之結果實例



圖七



圖八



圖九



圖十

[川普與歐巴馬對話繪本影片](#)

伍、專題介紹影片

[專題介紹影片](#)

陸、未來展望

本專題以古文字辨識與互動學習為核心，旨在探索人工智慧於文字教育與文化傳播領域的應用可能性。團隊期望藉由系統的持續優化，逐步發展出一套能實際運用於教育機構、博物館與語言學習平台的成熟工具，使古文字不僅能夠以數位方式保存，更能透過創新科技實現教育普及與文化傳承。換言之，本專題不僅是技術實驗，更是一項結合教育、文化與科技的應用實踐。

在技術層面，未來的發展可朝向導入更高階的深度學習模型，特別是針對古文字特徵所設計的影像辨識架構，以增強系統對模糊、不規則與殘缺字形的處理能力。同時，資料庫建置亦將是重要方向之一，若能與中央研究院漢字資料庫或國際漢學研究機構合作，擴充字形、書體、歷史語源等多維度資訊，將能顯著提升辨識結果的精確度與學術權威性，並為後續應用奠定穩固基礎。

Make Oracle Bones Great Again：川普遇上甲骨文 AI

在語音模組的應用上，本專題亦著眼於多語言與方言的融入，例如日語音讀、韓國漢字音、客語與閩南語等，期望讓不同語言背景的學習者皆能透過語音輔助理解漢字文化。除此之外，語音模組未來亦可擴充為訓練工具，幫助使用者學習正確發音，兼具文化導覽與語言學習雙重功能，進一步提升互動性與教育價值。

在實際應用的層面，本專題的系統若能開發為跨平台應用程式（App/Web），將能廣泛推廣於語文教育現場、文史研究者社群與一般大眾的日常使用。同時，博物館亦可將此系統與實體展覽結合，作為智慧導覽與教育推廣的輔助工具，增添觀展體驗的深度與趣味性。更進一步而言，這套系統亦可作為漢字文化對外交流的重要載體，藉由互動式與數位化的方式，將東亞文字的美感與歷史意涵推廣至國際舞台。

展望未來，本專題亦可嘗試整合擴增實境（AR）、虛擬實境（VR）以及手寫辨識等技術，使學習者能透過掃描拓片、碑刻與銘文等方式，在虛實融合的環境中探索漢字文化。透過持續的資料庫擴充、模型優化與介面調整，本專題所建構之系統不僅能深化教育層面的應用，更可成為文化傳承與科技創新的重要橋樑，體現人工智慧於人文教育領域的潛力。

參考文獻

無