組別：202628

專題名稱：基於人工智慧之影片字幕自動生成系統

一、指導老師：鄭煜輝老師

二、組　　員：趙建騏（11030112）

三、系統環境：

(一) 軟體：Windows、Python、FFmpeg、Gradio、Whisper 語音辨識模型、Git

(二) 硬體：電腦（具備 GPU 加速之顯示卡佳）、音效輸出入裝置

(三) 通訊設備/協定：TCP/IP、HTTP / WebSocket（Gradio 網頁介面傳輸）

四、系統功能與特色：

(一) **功能**

(1) 影音音訊自動擷取與格式轉換：支援常見影片與音訊格式上傳，透過後端自動抽離音軌並標準化為音訊檔案

(2) AI 多語言語音轉文字： 整合 Transformer 架構語音辨識模型，自動轉錄語音內容並進行時間軸（Timestamp）對齊與標點補齊。

(3) 標準字幕檔生成與匯出： 自動生成標準 SRT 字幕檔案供使用者預覽與下載

(4) 影片字幕內嵌（燒錄）： 可勾選將生成之字幕直接嵌入影片畫面中，輸出含字幕的 MP4 影片檔案

(5) Web 圖形化操作介面： 提供直覺的拖曳上傳、即時處理進度提示與一鍵生成下載功能

(二) **特色**

1. 一鍵全自動化流程：串接音訊擷取、AI 推論、文本處理至字幕嵌入，大幅降低傳統人工聽寫、切分與對齊的時間與人力成本
2. 兼具高抗噪與多語言辨識：採用強健之深度學習模型架構，在口語對話與日常環境音下均具備良好的辨識率與可讀性
3. 本機端推論與高擴展性模組架構：採分層模組化設計（介面、處理、推論、生成），兼顧隱私性，未來可彈性擴充多語言翻譯或語意摘要功能