

# AI vs. 人類文本偵測器

AI B 基校士語言學應用有分計  
一個基於語言學特徵的機器學習專案

「強大的生成能力，  
伴隨著巨大的驗證需求。」

許巧琳 (Hsu-chiao-lin) | 專案展示

# 我們為何需要驗證文本的來源？



## 學術誠信 (Academic Integrity)

確保學生提交的作品為原創，維護教育公平性。



## 防範假資訊 (Misinformation)

偵測由 AI 大規模生成的假新聞或惡意內容，避免其快速擴散。



## 保護原創性 (Authenticity)

在 AI 內容普及的時代，分辨並保留人類獨有的創造力與表達方式。

# 專案目標：建立一個清晰、可解釋的分類器

輸入：任何原始文本



AI/人類文本偵測器



輸出：一個機率分數

人類撰寫: 15%

AI 生成: 85%

15%

論文、電子郵件、文章等

清楚標示文本來源的可能性

# 系統架構：一個從輸入到預測的完整流程



# 核心洞察：解構 AI 與人類寫作的「文字指紋」

AI 和人類在寫作時會留下不同的統計學特徵。我們的模型便是學習辨識這些細微的差異。

**句子長度 (Sentence Length)**

**詞彙多樣性 (Vocabulary Diversity)**

**標點符號與寫作的「突發性」 (Punctuation & Burstiness)**

# 特徵分析：AI 與人類寫作風格的顯著差異

## 句子長度 (Sentence Length)

AI：句子長度趨於一致且均勻。



人類：長短句交錯，變化顯著，更具節奏感。



## 詞彙多樣性 (Vocabulary Diversity)

AI：傾向使用「安全」的常用詞彙，詞彙多樣性較低（Type-Token Ratio 較低）。

我們可以這個因為所以  
但是，一些很有和

人類：詞彙選擇更豐富，會使用較不常見的詞語。

我們 可以 這個 詞彙 差異 空白  
因為 所以 但是 多樣性 豐富 交連  
安全 突發性 很 有 特徵 顯著 交錯  
傾向 規律 模式 思考 點 聚集

寫作「突發性」(Burstiness)

AI：句構和標點符號的使用模式較為規律。



人類：寫作風格更具「突發性」，思考與表達的模式會反映在句子結構上。



# 模型的選擇：為何從邏輯斯迴歸開始？

選擇一個簡單、強大且可解釋的模型作為堅實的基礎。



## 高效率 (Efficiency)

訓練速度快，能夠進行即時推論 (real-time inference)，適合部署為線上應用。



## 可解釋性 (Interpretability)

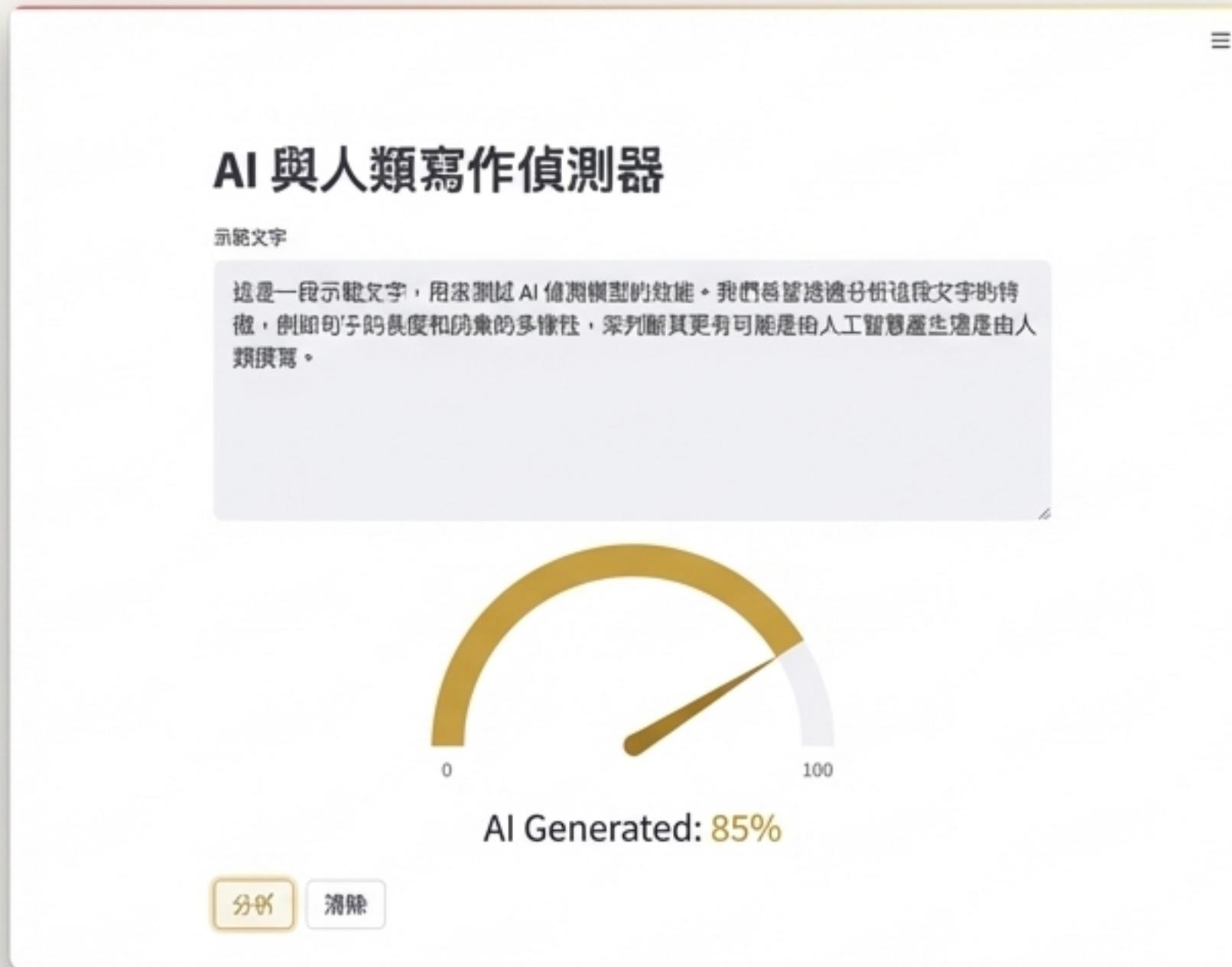
模型的係數 (coefficients) 能清楚顯示哪些特徵或詞彙最可能觸發「AI」的判斷，易於理解與除錯。



## 強大的基準 (Strong Baseline)

在嘗試如 BERT 等更複雜的 Transformer 模型之前，建立一個可靠的效能基準線。

# 實時演示：一個簡潔、直觀的偵測工具



## 介面

採用 Streamlit 打造，提供簡單的文字框介面。

## 反饋

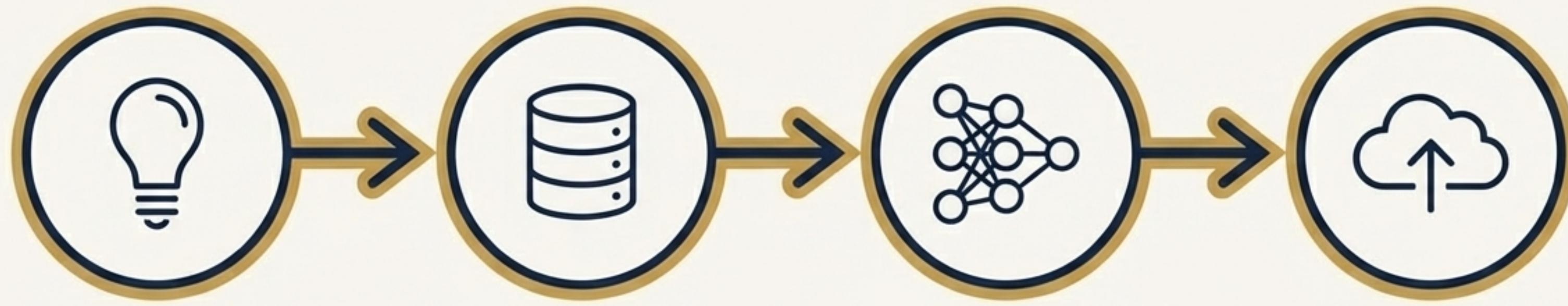
即時提供 AI/人類的機率分析。

## 線上應用程式網址

<https://aihumandetector.streamlit.app/>



# 專案工作流程：從概念到部署的四個階段



**1. 構想 (Idea) :**  
問題定義與專案概念  
化

**2. 資料 (Data) :**  
數據集的收集與預處  
理

**3. 模型 (Model) :**  
訓練邏輯斯迴歸分類  
器

**4. 部署  
(Deployment) :**  
使用 Streamlit Cloud  
將應用程式部署上線

# 成果總結與未來展望

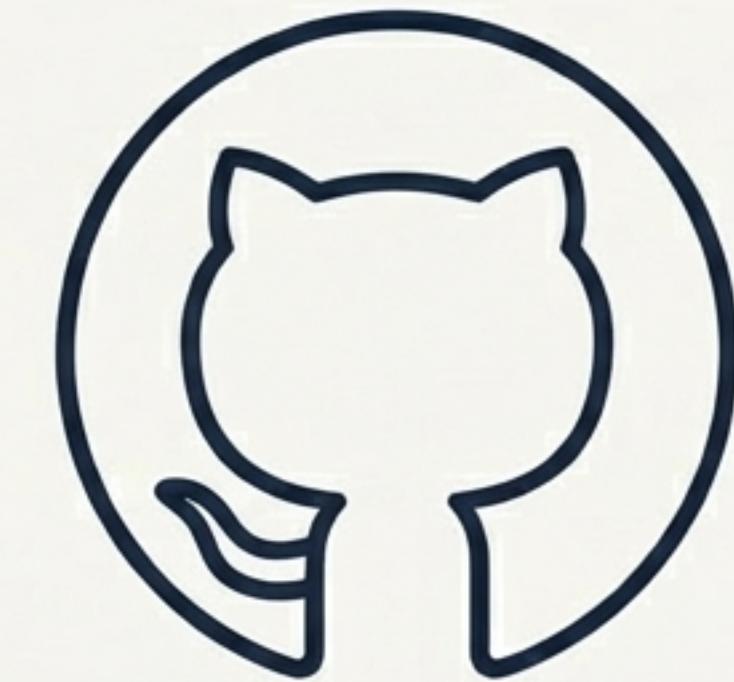
## 已達成就 (Achievements)

-  成功建立一個端到端的偵測流程。
-  部署了一個使用者友善的互動介面。

## 未來優化方向 (Next Frontiers)

-  整合深度學習模型 (如 BERT/RoBERTa) 以提升準確度。
-  擴展訓練資料集，以增強模型的泛化能力 (generalization)。

# 深入了解：程式碼與技術細節



GitHub Repository:  
<https://github.com/Hsu-chiao-lin/ai-human-text-detector>



感謝您的閱讀。