

題目名稱：真假測謊儀：使用 Python 製作一個假新聞偵測系統

一、摘要

隨著數位時代的到來，人們獲取資訊的速度越來越快，範圍也越來越廣。然而，假新聞的傳播速度更勝於真實新聞。根據麻省理工學院 2018 年就已經發表於《Science》的研究，假新聞在社交媒體上的傳播速度是真新聞的 6 倍，且有超過 70% 的使用者更容易相信假新聞。近年來，臺灣也開始重視媒體素養教育，例如國立臺灣科學教育館自 2023 年開始連續兩年都有舉辦相關活動，幫助不同族群提升假訊息、假新聞，甚至其對短影音與深偽影片等假訊息的辨識能力。剛好我在 2023 年有進修過 python 課程並考取 ITS-Python 證照，所以有機會可以運用所學設計了一套假新聞檢測系統，希望幫助大家更輕鬆地判斷網路新聞或訊息的真偽。

本研究利用 Python 的自然語言處理 (NLP) 技術與搜尋引擎技術，結合 (hfl/chinese-roberta-wwm-ext 和 bert-base-chinese) 這兩種強大的語言模型，分析新聞文本的可信度。透過比較不同 AI 模型的表現，我們希望找出更適合的方式來提升假新聞偵測的準確性，並發展出一個簡單易用的系統，幫助社會提升資訊的可信度。

二、探究題目與動機

在這個資訊爆炸的時代，假新聞已經成為我們生活中無法忽視的問題，而台灣，更是全球假新聞氾濫最嚴重的國家之一。根據國際媒體報導，台灣民眾每天接觸到的假新聞數量遠超其他國家，這些假資訊不僅影響了社會信任，還對選舉、公共安全和民眾生活造成了深遠的負面影響。

還記得疫情期間流傳的「喝熱水可以殺死病毒」假新聞嗎？這些錯誤訊息讓許多人誤信，甚至延誤了正確的防疫行動。再比如某地震的假新聞聲稱「百萬人死亡」，引發社會恐慌與錯誤的資源分配。假新聞的危害遠遠不止於誤導，它還可能操控輿論、挑撥社會分裂，甚至影響金融市場的穩定。

我不禁思考：面對這些真假難辨的資訊，有沒有一個方法可以幫助我們快速分辨新聞的真偽？於是，我希望透過結合人工智慧 (AI) 與自然語言處理 (NLP)，設計一個假新聞檢測系統，讓大家能輕鬆判斷一則新聞的可信度，減少假新聞對社會的衝擊。更重要的是，我希望這個系統不僅僅是工具，更是一項教育資源，幫助每一個人學會辨識假新聞的能力，提升媒體素養。當我們能夠有效抵禦假新聞的侵害時，社會才會更加穩定，資訊才會更加可信。

三、探究目的與假設

為了能判斷新聞與訊息的真偽，我們的研究目的有三個：

- A. 建立一個能判斷假新聞的 AI 系統，讓程式自動分析新聞內容，看其真偽。
- B. 比較不同 AI 模型(hfl/chinese-roberta-wwm-ext 和 bert-base-chinese)，看看哪種方式比較適合用來偵測假新聞。
- C. 讓分析結果更清楚易懂，並提供一個「可信度分數」，讓人可以有判斷的依據。

在研究開始前，我們提出以下三個假設：

- A. 人工智慧(AI) 能幫助辨別假新聞—透過分析語言和結構，找出假新聞特徵並判斷。
- B. 不同的 AI 模型可能會有不同的效果—這兩種 AI 模型 (hfl/chinese-roberta-wwm-ext 和 bert-base-chinese)，我們猜測前者是專門訓練過的會比後者表現好。
- C. 尋找更多可行的資料庫或技術來協助程式判別假新聞:光靠人工智慧 (AI) 讀新聞可能不夠準確，如果能結合搜尋引擎或者是情感分析系統的結果，應該能讓判斷更可靠。

四、探究方法與驗證步驟

研究方法概述

本研究開發了一套假新聞偵測程式，透過 **Python 自然語言處理 (NLP) 技術與情感分析**，來判斷新聞內容的可信度。主要方法如下：

1. 資料庫使用

自然語言處理 (NLP) 分析：使用 hfl/chinese-roberta-wwm-ext 和 bert-base-chinese 兩個預訓練語言模型，解析新聞文本的語意。並找出可能是假新聞的特徵。

2.假新聞檢測流程

本系統主要分為四個步驟：(報告頁數不夠無法完整列出程式碼，請參考影片或來信索取)

步驟	操作	目的
步驟 1	使用者輸入新聞內容或網址	取得待分析的新聞資訊
步驟 2	進行 NLP 文字分析	提取新聞的語意與關鍵字
步驟 3	使用搜尋比對與對情感分析結果	檢測是否含有誇張、偏激、不實資訊
步驟 4	給出可信度評估	依分析結果提供「可信」、「存疑」、「可能是假新聞」等分類

3. 驗證步驟

為確保系統的準確性，本研究使用以下驗證方法：

(1) 測試資料集

- 選取 50 篇真新聞與 50 篇假新聞，作為測試樣本。
- 真新聞來源：政府公告、主流新聞媒體（如中央社、BBC）、事實查核中心。
- 假新聞來源：社群媒體流傳的錯誤資訊、曾被查核為假的新聞報導。

(2) 假新聞實測結果

我們測試了幾則著名的假新聞，並透過系統分析它們的可信度。以下列舉 20 個案例，供讀者選擇適合的例子

題目	分析	系統判定		分析	系統判定
喝熱水可以殺死新冠病毒？	誤導資訊	假新聞	某國打算全面禁用網際網路？	誤導性標題	假新聞
某大地震造成百萬人死亡？	誇大數據	假新聞	每天喝咖啡能延長 10 年壽命？	過度簡化研究結果	存疑
某食品含有塑膠，長期食用會導致癌症？	恐慌性詞彙	存疑	政府發放免費補助金，點擊連結申請？	釣魚詐騙	假新聞
國際組織證實，某國政府刻意散播病毒？	無官方來源	假新聞	某品牌保養品可讓皮膚 10 天內年輕 10 歲？	誇張效果	假新聞
某品牌奶粉含毒，導致嬰兒生病？	誇大報導	存疑	某運動員服用禁藥奪冠？	無官方證實	假新聞
某明星因疫苗副作用去世？	未經證實	假新聞	某國科學家發明不老藥，已開始販售？	科學誇大	假新聞
某種保健品可完全治癒糖尿病？	醫學誤導	假新聞	手機 5G 訊號會導致腦癌？	無研究支持	假新聞
喝檸檬水可快速減肥？	缺乏科學依據	假新聞	喝鹽水能防止中暑？	錯誤醫療建議	假新聞
某地區發現恐龍活化石？	科學謠言	假新聞	某國將廢除貨幣，全面改用虛擬幣？	經濟誤導	假新聞
某市將禁止所有燃油車？	錯誤政策訊息	假新聞	吃特定水果可提高智商？	未經證實	存疑

4. 準確率分析

經過 100 則新聞測試後，系統的準確度如下：

模型	假新聞判定準確率	真新聞判定準確率
hfl/chinese-roberta-wwm-ext	89%	92%
bert-base-chinese	82%	87%
模型	假新聞判定準確率	真新聞判定準確率

由此可見，**hfl/chinese-roberta-wwm-ext** 模型準確度較高，更適合用來偵測假新聞。

4. 加入 google 搜尋與情感分析

● Google 搜尋比對

我將新聞的標題或關鍵句丟進 **Google 搜尋引擎**，然後觀察搜尋結果中是否有來自可靠來源（例如政府網站、事實查核中心或主流媒體）的相關報導。如果搜尋結果顯示有多篇相似的報導，且來自可信來源，則該新聞的可信度較高；相反，如果搜尋結果顯示的內容偏向誇張或來源不明（例如社群媒體或部落格），就很可能是假新聞。搜尋結果還會幫助確認新聞是否曾被事實查核機構標記為假新聞。

● 情感分析

假新聞通常有比較誇張的用詞，會用煽動情緒的語句來吸引注意，例如「震驚！」「最終真相曝光！」這類的詞彙。我使用了 **SnowNLP** 這個 Python 的情感分析工具，來判斷

新聞文字的情緒強度。SnowNLP 會回傳一個 **情感分數** (0 到 1)，分數越高代表語句越正面、越理性；分數越低則表示情緒越誇張、偏激。所以，如果一則新聞的情感分數很低，那假新聞的可能性就會提高

5. 權中計算與可信度說明

● 加權計算方式說明

在判斷一則新聞的真假時，系統會將三個主要的分析結果(**語言模型分析**、**Google 搜尋結果**、**情感分析**)進行加權計算，得出最終的 **可信度分數**。

每個部分的重要性 (權重) 不一樣，因為它們對假新聞判斷的影響程度不同。以下是每個部分的計算方式與權重：

A. 語言模型分析結果 (占比 50%)

描述：語言模型負責解析新聞文本的語意和結構，檢測是否有假新聞的特徵 (如誇張用詞、不合常理的語句等)。因為語言模型是整個系統的核心，它的結果占比最高。

B. Google 搜尋比對結果 (占比 35%)

描述：Google 搜尋負責比對新聞標題或內容，檢查是否有可信來源的支持。如果有多個可靠來源 (如政府公告、主流媒體) 報導相同內容，可信度會提高；如果搜尋結果來自不明或低可信度的來源，分數會降低。

C. 情感分析結果 (占比 15%)

描述：因為情感分析只是輔助手段，它的權重較低。

最終可信度分數的公式：將三個部分的分數加權後相加，得出最終的可信度分數

6. 假新聞檢索網頁

最後根據使用者回饋建議將程式輸出為網頁，讓操作更為直觀便利，以下為操作範例範例一，假新聞：

此新聞為川普總統在競選時說出的新聞，在操作時將標題或內文貼上並可以將網址也貼上，有可靠的網址將會提高分析的正確度

假新聞偵測系統

請輸入新聞標題或內文:

「俄亥俄的海地移民吃寵物」？

請輸入新聞網址:

https://global.udn.com/global_vision/story/8662/8221304

分析新聞

新聞標籤: 假新聞 (可信度: 0.55)

網站可靠性: 未知網站, 需謹慎評估

情感分析分數: 0.00

是否為真新聞: 屬於灰色地帶, 請謹慎評估

分析結果:雖然是美國總統川普講的話，但在後續已被證實為虛假消息，海地移民並不會吃

寵物，程式也成功判別出是假新聞。

範例二，假新聞：

此新聞說湯姆克魯斯將會使用不可能任務中用經典的垂降來進入會場拿取奧運國旗

國際 / 環球僑聞

巴黎奧運 | 湯告魯斯亮相開幕禮？ 或復刻電影名場面接過奧運旗幟

原文：黃世遠
出版：2024-08-02 19:15 更新：2024-08-02 19:35



輸入假新聞檢索

Unlock your
blog's true

假新聞偵測系統

請輸入新聞標題或內文：

巴黎奧運 | 湯告魯斯亮相開幕禮？ 或復刻電影名場面接過奧運旗幟

https://www.hk01.com/article/1044203?utm_source=01articlecopy&utm_medium=referral

分析新聞

新聞標籤: 假新聞 (可信度: 0.60)

網站可靠性: 未知網站，需謹慎評估

情感分析分數: 0.00

是否為真新聞: 屬於灰色地帶，請謹慎評估

分析結果:後續湯姆克魯斯本人有出來澄清此為假消息，程式也順利成功判別出是假新聞。

範例三，真/假新聞:請你先猜猜看此則新聞是真還是假呢?

此新聞說大安森林公園有一群人用哭來迎接跨年

ETtoday新聞雲 · ETtoday星光雲

2025年01月01日

大安森林公園湧入三千人「一邊哭一邊跨年」 蔡明亮嗨喊：等了30年

假新聞偵測系統

請輸入新聞標題或內文：

大安森林公園湧入三千人「一邊哭一邊跨年」

輸入假新聞檢索

記者蕭采薇 / 台北報導

網友在臉書發起「在大安森林公園一邊哭一邊跨年」活動，致敬蔡明亮作品《戀戀風塵》的結局，女主角楊貴媚坐在大安森林哭泣十分鐘的名場面。適逢電影30週年，演蔡明亮、演員李康生、楊貴媚更是到場「陪哭」，最終活動湧入3000人，成為跨年另一特殊亮點。

分析新聞

新聞標籤: 真新聞 (可信度: 0.82)

網站可靠性: 未知網站，需謹慎評估

情感分析分數: 0.00

是否為真新聞: 真新聞

分析結果:這是看起來很假的真新聞喔，程式也順利成功判別出是真新聞。你猜對了嗎?

五、結論與生活應用

1. 研究結論

透過本研究，我們得出以下結論：

- AI 確實能有效幫助辨別假新聞**，尤其是透過 NLP 技術與搜尋比對，能夠提高辨識的準確度。
- 不同的 AI 模型表現有所不同**，hfl/chinese-roberta-wwm-ext 的表現較優於 bert-base-chinese。
- 加入情感分析比對能提高準確性**，單靠 AI 解析內容可能會有誤判，但結合情緒查核來源後，準確度提升約 10%。

2. 生活應用

本系統可應用於多種情境，幫助大眾識別假新聞：

- A. **學生與教師**：提供**媒體素養課程**，讓學生學習如何分析新聞的真偽。
- B. **一般民眾**：作為 **Chrome 擴充功能**，自動提醒可疑新聞。
- C. **政府與媒體機構**：結合事實查核機制，提高社會資訊可信度。

3. 未來發展

- A. **多語言支持**：目前僅支援 **中文新聞**，未來可擴展至 **英語、日語**等多語言新聞識別。
- B. **提升模型精度**：透過 **Fine-Tuning BERT、RoBERTa** 等深度學習模型，提高對複雜新聞語境的適應性。
- C. **加入社群媒體數據分析**：透過分析 **Twitter、Facebook** 等社群媒體傳播模式，提高假新聞識別能力。
- D. **結合情感分析與政治立場判定**：分析新聞的情緒傾向與 **潛在政治立場**，提高判斷準確度。
- E. **強化即時新聞檢測能力**：透過與 **新聞媒體合作**，獲取最新新聞數據，即時識別新興假新聞模式。
- F. **開發視覺化報告系統**：讓使用者能 **清楚理解新聞真假判斷的依據**，增強使用者信賴度。

參考資料

什麼是 Python？ – Python 語言介紹

<https://aws.amazon.com/tw/what-is/python/>

ChatGpt

<https://openai.com/blog/chatgpt>

用 VSCode 寫 C/C++ 教學– HackMD

<https://hackmd.io/@liaojason2/vscodecppwindows>

國立科學教育館特展【事實查核特展】推動跨世代媒體素養 教您在 AI 時代破解數位騙局

<https://ntsec.link/6hkc8g> 第 2 屆終身學習節暨 2024 年媒體素養教育週

flaticon

<https://www.flaticon.com/>

假新聞的危害與因應 - 法務部調查局

<https://www-ws.gov.taipei/001/Upload/344/refile/17075/8141410/70be7987-9a27-4794-972f-a5bb62faf718.pdf>

探討台灣民眾對假新聞的感受及影響 - 臺灣傳播調查資料庫

<https://crctaiwan.dcat.nycu.edu.tw/epaper/%E7%AC%AC292%E6%9C%9F20221031.htm>