

## 摘要

當今 21 世紀為數位媒體崛起的時代，通訊技術的突破，讓人類獲得資訊的方式十分多元，例如：社群軟體。大眾傳播媒體的運作形式已經打破傳統的框架，影響事物的層面無遠弗屆，目前社群媒體結合資訊科技已達到前所未見的發展。然而假新聞的傳播藉機甚囂塵上，對於社會產生持續性的不良影響，加上近年來新冠疫情爆發，假新聞議題逐漸受到重視。政府、民間團體、媒體業者紛紛自發性成立事實查核組織，調查網路流傳的訊息與新聞的真實性，當發現假新聞或假消息時，查核組織將進行查證並盡全力阻止虛假資訊混淆閱讀者的視聽，嘗試維護與培養公民媒體識讀的能力。本研究應用三份英文新聞資料集(KaggleFN、ISOT、CCPD)與一份繁體中文新聞資料集(TFC)進行實作，使用針對文本資料的大數據分析方法：自然語言處理(Natural Language Processing, NLP)，結合自監督式學習的語言模型(BERT)。投入英文與繁體中文假新聞檢測的研究，根據假新聞檢測模型的分類績效評估當作指標，分別找出適用於英文和繁體中文的假新聞檢測模型，最後部署模型建立事實查核系統。本研究除了在自然語言處理應用到不同語言進行研究之外，並落實機器學習模型在實務上的應用。

**關鍵字：**假新聞檢測、自然語言處理、自監督式學習

