

東海大學統計學系學術演講

時 間：115 年 9 月 22 日（星期二下午 14:10~15:00）

地 點：管理學院新大樓 M240

主講人：余清祥 教授（國立政治大學 統計系）

講 題：大數據與文字分析（Big Data and Text Analytics）

Abstract

政治大學統計學系余清祥教授以「大數據與文字分析」為題，系統性探討在生成式人工智慧快速發展的趨勢下，文字探勘與作者風格辨識的演進與應用。近年來，以大型語言模型為代表的 AI 工具廣泛應用於內容製作、新聞編寫及學術研究，隨之引發關於文本真實性、學術倫理與人機分工的多元思辨。本次演講從統計學與資料科學的視角出發，首先介紹西方經典研究如圖靈字彙估計、《聯邦黨人文集》作者考證，以及中文學界針對《紅樓夢》作者爭議與金庸小說字彙特徵之分析脈絡；隨後進一步借鑑「生物多樣性」的量化思維（如相異字比例、資訊熵與 Simpson 指數），將非結構化的自然語言轉化為具備解釋力與鑑別度的結構化指標。演講核心以臺灣專利摘要與 ChatGPT 生成內容之比對為實證範例，藉由探索性資料分析（EDA）進行關鍵變數篩選，比較統計模型（如羅吉斯迴歸）與機器學習、深度學習（如 BERT）的分類效能。分析發現統計方法在確保準確度的同時，兼具模型簡約、運算效率及實質詮釋價值的特色，在面對大數據與人工智慧浪潮下，為分析策略與溝通思維提供具體且客觀的參考。

敬請公告，歡迎參加！

爾後演講公告將於本系網站刊載

<http://stat.thu.edu.tw/>