假新聞在社群媒體中之操作模式初探
-以 ChatGPT 為例

作者/曾柏元

提要

一、**ChatGPT作為先進的自然語言處理模型，擁有文本生成能力。然而，其自動生成文章的功能也引發假新聞傳播的嚴重問題，特別是在政治與社會議題上尤為顯著。此系統能模擬真實新聞，使虛假資訊更加具迷惑性，增加辨識難度，對新聞媒體與公眾認知產生負面影響。**

二、**ChatGPT生成的假新聞透過社群媒體傳播，已成為認知戰中的重要工具。利用網路多層節點與病毒式擴散策略，使假新聞在多國內影響「政治、經濟、軍事和社會心理」，對公眾的理性判斷及政府高層決策構成威脅。**

三、**我國應對ChatGPT生成假新聞策略應強調以下幾點：建立獨立監管機構、開發假新聞檢測工具、推動媒體識讀教育，並加強法律監管，上述措施均有助於防止生成式AI技術擴散假新聞，確保資訊真實性。**

關鍵字：ChatGPT、社群媒體、假新聞、網路宣傳

前言

隨著生成式人工智慧技術的蓬勃發展，ChatGPT已逐漸成為社群媒體中強大的文本生成工具。然而，其應用不僅促進內容創作自動化，也使假新聞生成與傳播更為隱蔽且高效。因ChatGPT能迅速生成高度擬真，且具有說服力的新聞報導、媒體貼文與評論，使得假新聞在媒體平台上迅速擴散，成為全球資訊生態中重大挑戰。

在社群媒體中，ChatGPT的假新聞生成方式呈現兩種顯著特徵：首先，ChatGPT生成文本可模仿專業媒體報導語氣與結構，使其具有高度擬真性，從而降低受眾警惕。此種假新聞內容常以精確語言表達與可信敘事結構，達到輿論誤導目的。其次，藉自動生成大量相似內容，以操縱社群媒體輿論風向，藉由重複發布同一觀點，營造出一致的輿論氛圍，影響公眾對特定議題看法，甚至干預公共討論走向，從而達到假新聞推動者的特定政治、經濟、軍事以及社會心理目的。

此種自動生成假新聞模式，其生成速度與社群傳播範圍較為廣泛性，因快速內容生成、生產與傳播模式，使得現有社群媒體監管機制難以有效應對，若仍以傳統的內容審核與假新聞偵測手段顯得落後。因此，為遏制ChatGPT在假新聞生成中的負面影響，應從假新聞內容生成、生產與傳播等過程予以杜絕，所以政府應提升技術偵測與篩選能力，更需結合嚴格的社群媒體監管政策與法律框架，及強化公眾的媒體素養教育來提升辨識能力，確保資訊的真實性與維護社會信任。

「ChatGPT」與假新聞之關係

Chat Generative Pre-trained Transformer（簡稱ChatGPT）[1]是種經常出現在新聞報導中的大型語言模型（LLM）工具，此系統由OpenAI開發，是作為自然語言處理技術的一部分，它雖為使用者提供極大便利性，但同時也帶來潛在傳播風險。如同美國事實查核研究機構NewsGuard共同執行長克羅維茲（GordonCrovitz）所言：「ChatGPT可能成為網絡上最強大的錯誤訊息工具。」[2]，此句使ChatGPT與假新聞傳播兩者關係更為密切。

一、ChatGPT發展背景

ChatGPT是由馬斯克（Elon Musk）參與創立的OpenAI基金會所研發(如圖1)其發展始於2018年的GPT-1.0，並於2022年11月推出ChatGPT-3.5，創下上線五日內吸引超過百萬使用者登入下載運用。至次年1月底，使用數已突破1億。[3]此系統在此發展階段不斷透過反覆訓練與生成內容回饋，在巨量資料中自主學習，進以轉變為具創造性生成模型。[4]

ChatGPT是種聊天機器人，採用生成式預訓練轉換器（Generative Pretrained Transformer, GPT），進行大數據的自然語言處理分析。[5]在基於GPT系列模型下，隨著各代模型參數提升，以促進對話生成技術實際應用，對社會帶來重大影響。自2018年推出GPT-1以來，OpenAI逐步發布GPT-2、GPT-3、ChatGPT、

1 GPT 的全名是「Generative Pre-trained Transformer」，中文意為「生成式預訓練變換模型」。GPT的第一個字母「G」是 Generative 的字首，其意思是「生成式」。「P」是Pre-trained的字首，其意思是「預訓練」，「T」是Transformer的字首，直接翻譯是「轉換器」。參考黃仲宏，〈活用生成式人工智慧擘劃機器人自動化的發展〉，《機械工業雜誌》(竹東)，財團法人工業技術研究院，第485期，2023年8月，頁10-11。

2 田孟心，〈研究揭露：ChatGPT如何助長虛假訊息？〉，《天下雜誌》，2023年2月16日，https://www.cw.com.tw/article/5124734，檢索日期：2024年8月11日。

3 曾敏禎，〈中國版的ChatGPT，鸚鵡學舌？畫虎類犬？〉，《國防安全研究院》，2023年2月15日，https://indsr.org.tw/focus?uid=11&typeid=30&pid=575，檢索日期：2024年7月24日。

4 王亞珅、李強、石戈，〈ChatGPT對社交機器人技術發展的影響分析〉，《無人系統技術》(北京市)，海鷹科技情報研究所，第6卷第2期，2023年，頁95-102。

5 董慧明，〈生成式技術發展對國家安全的影響與挑戰〉，《清流雙月刊》(臺北)，法務部調查局，第48期，2023年11月，頁4。

ChatGPT-3.5及ChatGPT-4等版本(如表1)致力於為使用者提供更流暢且具上下文理解之對話體驗。ChatGPT的到來也標誌著人工智能技術應用進入新階段，也為GPT-4奠定後續基礎。

圖 1 馬斯克（Elon Musk）與 OpenAI

資料來源：J.FrSebastião，Elon Musk tried to take control of the company behind ChatGPT，https://www.menosfios.com/wp-content/uploads/2023/03/elon_musk_targets_openai_and_chatgpt.webp，檢索日期：2024年7月24日。

表 1 ChatGPT 發展歷程表

項次	名稱	時間	訓練資料量	參數量內容
一	GPT-1	2018年	5GB	具備1.5億筆參數的預訓練模型，採用自迴歸生成方式，通過序列化預測下一個詞，開啟生成式AI基礎探索
二	GPT-2	2019年	40GB	具備1.75萬億筆參數。其龐大規模使其能在多樣化語境中生成高質量語言內容，成為生成式AI領域的領航者
三	GPT-3	2020年	45TB	當時最先進的GPT模型，具有1.75萬億筆參數，是最大語言模型。
四	ChatGPT	2020年	45TB	ChatGPT基於GPT-3的技術，是專門針對對話生成應用，具備強大語境理解和回應能力，並在各類應用場景中獲得廣泛運用。
五	ChatGPT-3.5	2022年	無	ChatGPT-3.5是ChatGPT的改進版本，針對對話生成進行優化，並進以調整提高回應準確性和針對性，支援最多2,048個token的輸入。
六	GPT4	2023年	無	GPT-4提升付費版本，具有優化模型架構，能支援最多8,192個字元輸入，並提供高效和精準內容生成能力。

資料來源：邱銘傳，〈ChatGPT〉，http://ielab.ie.nthu.edu.tw/IIE2023_all/IIE_TA_AI_ChatGPT.pdf，檢索日期：2024年7月10日。

資料說明：GPT各版本主要差異集中在兩個方面：預訓練資料規模和參數量大小。隨著訓練資料和參數量增長，模型生成能力得以提升，使其能更精確地處理多樣化且複雜應用場景，從而顯著增強其應對各種任務能力。參閱黃仁志，〈生成式AI的應用、風險與對應政策〉，《國際前瞻》(臺灣)，中華經濟研究院，第208期，2023年7月，頁81。

二、ChatGPT功能與運作簡介

ChatGPT定義「基於深度學習的自然語言處理模型」。此系統透過深度學習技術，具備理解人類語言能力，能生成文本與對話。[6]主要依賴大量可存取之數位化資訊，提供使用者平台，使任何人能與其互動，並予以搜尋內部知識搜尋、解析查詢內容，生成完整回應。[7]其核心功能涵蓋搜尋功能、對話互動、創意寫作、資料整理、程式設計與教育學習、翻譯語言及數據分析等(如表2)。

表 2 ChatGPT 之多功能應用與效益表

項次	功能	敘述說明
一	搜尋功能	搜索專有名詞、歷史事件、天氣狀況等，迅速提供精確資訊。
二	對話互動	能夠理解上下文，生成連貫、相關且具洞察力回答，提升互動體驗。
三	創意寫作	協助生成故事創意與情節大綱，創建角色背景，編寫對話及場景描述，並提供文學風格示例，克服寫作障礙。
四	資料整理	分析長篇文件，進行條例和分類，甚至將資訊製成表格，便於系統化呈現。
五	程式設計	解釋程式語言、語法和概念，提供代碼示例，幫助解決編程問題，並建議代碼優化方案與解釋數據結構。
六	教育學習	幫助學生理解複雜概念，解答疑問，並提供個性化學習計畫，促進學習進度。
七	翻譯語言	翻譯短文與段落，解釋習語與俚語，提供語法與語言使用指導，甚至矯正發音。
八	數據分析	協助解釋數據分析結果，提供統計概念解釋，並建議適合特定數據集分析方法，增強數據處理能力。

資料來源：城市撞球館，〈ChatGPT的10大功能！讓您的AI體驗更精彩！〉，2024年9月1日，https://www.104net.net/API/modules/news/article.php?storyid=8，檢索日期：2024年7月9日。

再者，使用ChatGPT僅要登入ChatGPT網頁版進入後，點擊「Sign up」按

6 沙珮琦，〈你的祕密，ChatGPT 全知道？新工具背後隱藏的資安風險〉，《科技魅癮》，2023年6月9日，https://www.charmingscitech.nat.gov.tw/post/worldview10-chatgpt，檢索日期：2024年7月24日。

7 吳維雅，〈生成式AI的善與惡（一）：ChatGPT為何成為當代顯學？〉，《鳴人堂》，2023年5月8日，https://opinion.udn.com/opinion/story/120817/7149512，檢索日期：2024年7月24日。

鈕，即可以快速註冊，執行步驟如下(如圖2)：

(一)首先，使用者需登入OpenAI網站，並進入ChatGPT專屬頁面。

(二)點擊「Create New Chat Model」按鈕，選擇所需創建的模型類型（如GPT-4.0）。

(三)使用者可選擇透過OpenAI API執行對話生成，或下載OpenAI提供Python套件以便在本地環境中運行模型。

(四)在獲得API存取權限後，可使用API端點傳送對話請求，通常需提交包含聊天歷史及指令輸入參數，以促使模型生成相應回應。

(五)模型生成回應將透過API回傳，使用者可根據回應內容進行呈現或後續處理，並將結果傳遞至最終用戶。

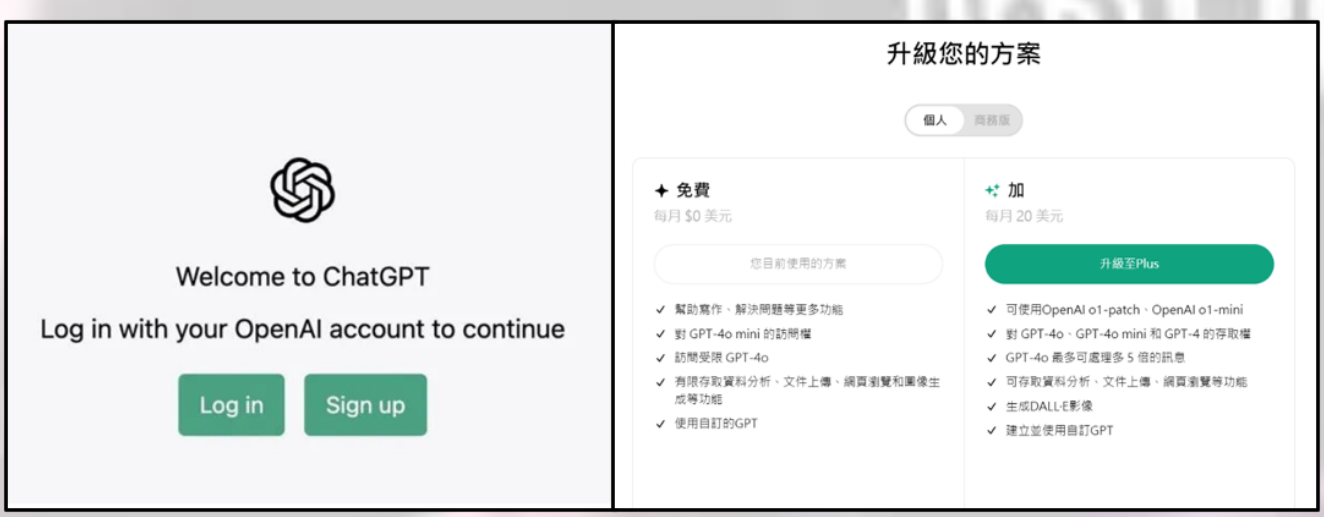

圖 2 ChatGPT 網頁版登入及方案

資料來源：chatgpt，https://openai.com/chatgpt/，檢索日期：2024年7月19日。

該系統藉大量非監督式學習所得的文本資料，理解單詞和短句語義結構，將其轉換成數學向量，形成生成自然語言文本模型。整體而言，ChatGPT操作流程可總結為「輸入→編碼→解碼」。[8]均屬於生成式人工智慧生成內容，即AIGC（AI-Generated Content）。[9]

(一)輸入處理(Input Processing)：

使用者首先輸入文本資料，如問題、指令或評論，該輸入經過一系列「預處理」步驟，將其轉換為模型能理解的語義格式。預處理目的是確保

8 周秉誼，〈深度學習與ChatGPT〉，《國立臺灣大學技資中心電子報》，2023年6月20日，https://www.cc.ntu.edu.tw/chinese/epaper/home/20230620_006503.html，檢索日期：2024年7月25日。

9 葉繼元、郭衛兵，〈生成式人工智慧參與學術評價的反思〉，《中國社會科學評價》(北京市)，中國社會科學院，第1期，2024年6月11日，頁37。

輸入一致性，並為後續編碼階段奠定基礎。

(二)編碼轉換(Encoding Transformation)：

模型將預處理過之文本進行編碼，轉換為向量表示，此過程依賴於「Tokenization」技術，即將文本劃分為單詞或詞元，更將其轉換為模型可處理數字形式。該轉換為後續語義理解和文本生成提供結構化數據格式，使模型能夠更有效地解析輸入內容。

(三)解碼生成(Decoding Generation)：

基於輸入編碼數據，模型利用多層Transformer架構中的自注意力機制（Attention Mechanisms）和前饋神經網絡來處理序列數據。此過程旨在捕捉文本各部分之間語義關聯，並根據以上關聯性預測下一個可能詞元。隨著詞元逐步生成，最終產生連貫且完整回應。由於模型經過大量文本數據訓練，能夠模仿人類語言模式，生成流暢且符合上下文文本內容。

三、ChatGPT與假新聞之關聯性

假新聞（Fake News）問題相當複雜，涉及文字、圖像和影片等多種形式，尤其書面內容更難以識別。自從ChatGPT自動生成寫作功能演化後，新聞業面臨倫理挑戰，無需記者即可對各議題實施操控，節省機構資源與時間。[10]然而，由於ChatGPT依賴非監督式數據訓練，其生成文本可能會嵌入不可靠、模糊甚至虛假資訊，在各領域問題上均會有潛在影響性。(如表3)

表3 ChatGPT與假新聞潛在影響

項次	種類	敘述說明
一	經濟	虛假商業新聞透過影響市場動態和投資決策，以謀取特定商業利益，可能加劇市場波動，並損害投資者信心。
二	政治	關於政治人物、政黨及選舉的虛假訊息，旨在操縱輿論，引導選民錯誤認知，影響民主選舉公正性與合法性。
三	民生	利用社會謠言和虛假新聞以提高點擊率和流量，這些訊息往往加重社會不安與焦慮，破壞社會信任結構。
四	醫學	在醫療與健康領域，特別是公共衛生危機期間，錯誤訊息傳播可能誤導公眾對疾病、疫苗和治療認知，從而損害公共健康。
五	歷史	故意扭曲歷史事實或推廣修正主義史觀，以達到特定政治或意識形態目的，以改變公眾歷史認知。
六	科學	曲解科學研究結果或選擇性引用數據，以支持特定政策議程或商業利益，從而誤導大眾對科學事實理解。
七	軍事	境外勢力通過認知作戰及資訊戰製造虛假軍事訊息，試圖瓦解國內外軍事決策者信心，並擾亂社會穩定與國家安全。

資料來源：作者自行整理

10 AIContentfy. (2023, November 6). ChatGPT and the future of news: Automation and AI. *AIContentfy*.https://aicontentfy.com/en/blog/chatgpt-and-future-of-news-automation-and-ai (檢索日期：2024年9月19日)。

過去受限於人力與成本，假新聞(訊息)生成較為困難且分類眾多(如圖3)。

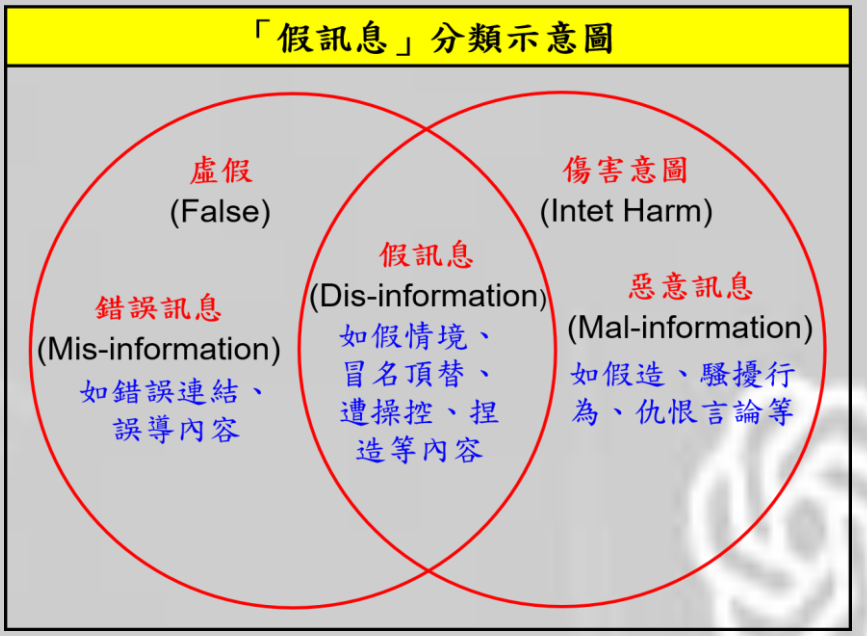

圖 3 假訊息分類示意圖

資料來源：UNESCO, Journalism, Fake News & Disinformation: Handbook for Journalism Education and Training, 2018/9/17, p44

現今卻可透過更少人力實施改寫，其形式區分為「虛假上下文」（False context）、「冒名內容」（Imposter content）、「操縱內容」（Manipulated content）及「偽造內容」（Fabricated content）等四種。[11]且如今ChatGPT每月僅需付20美元使用費，[12]便可從數十億篇文章與數據資料庫中生成類似真人撰寫之評論報告，以提升邏輯結構與可信度，使真假更難以分辨。[13]《Scientific Advances》期刊指出，該系統就如「雙面刃」，既能生成易於理解之準確資訊，也可創造具吸引力的虛假資訊，提升迷惑和欺騙性，使假新聞更難識別。[14]另外，2023年3月中央大學資訊電機學院教授蔡宗翰也有所述，GPT-4在生成文本時可能編造非事實性內容，導致誤導和負面影響。[15]

11 施達妮、顏妤恬，〈數位時代的假新聞〉，《漢學研究通訊》(臺北市)，國家圖書館，第147期，2018年8月，頁8。

12 陳品潔，〈chatGPT紅遍全球大量文本彈指生成訊息虛實難辨〉，《卓越新聞電子報》，2023年3月16日，https://feja.org.tw/69079，檢索日期：2024年7月25日。

13 Dekens, N. (2023, July 7). How ChatGPT could spread disinformation via fake reviews. Spiceworks. https://www-spiceworks-com.translate.goog/tech/artificial-intelligence/guest-article/how-chatgpt-could-spread-disinformation-via-fake-reviews/?_x_tr_sl=en&_x_tr_tl=zh-TW&_x_tr_hl=zh-TW&_x_tr_pto=sc (檢索日期：2024年9月19日)。

14 James Laird. (2023, June 29).ChatGPT Makes Spotting Fake News Impossible for Most People. https://tech-co.translate.goog/news/chatgpt-spotting-fake-news-impossible?_x_tr_sl=en&_x_tr_tl=zh-TW&_x_tr_hl=zh-TW&_x_tr_pto=sc(檢索日期：2024年7月29日)。

15 臺灣科技媒體中心，〈ChatGPT回答錯誤？如何不被生成式AI假訊息詐騙？｜ChatGPT使用須知8件事〉，《未

從上述可見，ChatGPT技術便利性，已帶來假新聞風險。根據2024年6月牛津大學「路透新聞學研究所」(Reuters Institute for the Study of Journalism)發布《數位新聞報告》(Digital News Report)，針對47個國家調查顯示，ChatGPT應用確實間接加速假新聞擴散，[16]上述更證實兩者之間的緊密聯繫。

ChatGPT 與假新聞生成之潛在角色與風險

俗話說：「水能載舟，亦能覆舟」，現代戰爭趨勢已轉向大軍未動，爭訊先啟，而ChatGPT生成式AI技術不同於傳統聊天機器人，其一旦應用於軍事上便可被武器化，因此類技術會加劇「資訊迷霧」現象。[17]

在此背景下，我國對假新聞攻擊的現狀深感憂慮。根據美國蘭德公司(RAND Corporation)於2023年9月所發表《防務一號》（Defense One）報告顯示，中共對臺灣虛假訊息攻擊已發展為系統性策略，非偶發性行為。目前正積極研究生成式AI工具（如ChatGPT）藉以操控國際輿論，進行深層認知影響。[18]儘管此類軟體在中共境內遭受限制，但仍可翻牆技術進行操作。同時，也在2023年3月推出「文心一言」(ERNIE Bot)軟體，展現積極布局野心。[19]

一、ChatGPT在假新聞生成中之角色

ChatGPT作為一種生成式人工智慧技術，潛藏極高生成假新聞風險。其生成過程缺乏透明度，使其成為不當用途下的「遊戲規則改變者」。因此，在假新聞生成與傳播中具有不可忽視之影響力，能輕易創造煽動且說服力強之假新聞報導、社論、博客文章或社交媒體貼文等。此內容形式與風格幾乎與真實新聞無異，可達到「迷惑性、氾濫性與快速傳播性」高度，對輿論操控上具備前所未有的效果。

(一)虛假內容─迷惑性：

ChatGPT生成內容速度遠超越傳統新聞製作流程，使用者可於網頁對話問框中提出任一問題，便可由網路資訊獲取訓練數據，在短短數秒內生成令人信服內容，可謂是「正經八百─胡說八道」。若此技術被特定人士所

來城市》，2023年3月26日，https://futurecity.cw.com.tw/article/2990?rec=i2i&from_id=3143&from_index=8，檢索日期：2024年7月29日。

16 劉文瑜，〈研究：受眾憂AI產製內容為新聞機構帶來新挑戰〉，《中央通訊社》，2024年6月17日，https://www.cna.com.tw/news/aopl/202406170026.aspx，檢索日期：2024年8月1日。

17 顏翩翩，〈ChatGPT帶來人工智慧新衝擊資策會科法所：應關注其風險層級〉，《資策會科技法律研究所》，2023年3月6日，https://www.cna.com.tw/postwrite/chi/336122，檢索日期：2024年8月1日。

18 Honrada, G. (2023, September 12). China ramps up AI-powered campaign against Taiwan. *The Geopolitics*. https://thegeopolitics-com.translate.goog/china-ramps-up-ai-powered-campaign-against-taiwan/?_x_tr_sl=en&_x_tr_tl=zh-TW&_x_tr_hl=zh-TW&_x_tr_pto=sc(檢索日期：2024年8月1日)。

19 曾敏禎，〈中國公布《生成式AI管理辦法》——推進器或緊箍咒？〉，《國防安全雙週報》(臺北)，財團法人國防安全研究院，2023年5月12日，頁57-58。

利用，可輕易插入虛構細節誤導受眾，[20]從而產生「幻覺」[21]效果。可從加拿大新布藍茲維大學(University of New Brunswick, UNB)聖約翰分校實驗心理學博士生喬丹麥克唐納(MacDonald)研究發現，ChatGPT產生300則文章中，共有32.3%產生幻覺。[22]此項研究揭示生成式AI在創造令人信服內容的同時，亦伴隨著潛在錯誤風險。

(二)降低成本─氾濫性：

ChatGPT應用大幅降低假新聞的邊際成本，幾乎達到零，顯著壓低在社交媒體上宣傳成本。此種低成本生產極大地提升假新聞擴散速度與影響範圍，可迅速蔓延至更廣泛受眾。在此背景下，訊息的有效控制與監管正面臨前所未有挑戰。[23]在美國喬治城大學（Georgetown University）研究員喬什·古德斯坦（Josh A. Goldstein）於《美國之音》受訪指出：「語言模型能顯著壓低大規模生成特定文本之成本，未來或將促使更多針對個人或量身定做的政治宣傳活動的出現」，此現象加深假新聞在數位化時代對社會影響之複雜性。[24]

(三)強化速度─傳播性：

ChatGPT在全球傳播中影響力尤為顯著，尤其可應用於新型欺騙形式中，例如：量身定制宣傳活動，使有意圖操縱輿論的人更易操弄假新聞。當其應用於社交媒體平台或論壇時，能生成極具真實感之社論、貼文或推文，並自動創造大量看似來自不同用戶的支持性評論。一旦在網路平台(Twitter、Discord、Telegram、4chan和Reddit)上被討論、擴散時，就像是真實新聞被討論一般，快速滲透到大眾輿論場域。[25]

---

20 劉芮菁，〈ChatGPT就是「一本正經地胡說八道」！杜奕瑾：AI生成科技恐讓假訊息更氾濫〉，《今周刊》，2023年3月27日，https://www.businesstoday.com.tw/article/category/183027/post/202303270036/，檢索日期：2024年8月4日。

21 「AI幻覺」（hallucination）被定義為：AI系統（如ChatGPT）生成看似真實，但不對應任何現實輸入的感官體驗。包括視覺、聽覺或其他類型的幻覺。此種現象在如生成模型等AI系統中較為常見。參考Alkaissi, H., & Mcfarlane.（2023, February 19） ChatGPT: Implications in scientific writing. Cureus, 15. https://doi.org/10.7759/cureus.35179 (檢索日期：2024年9月19日)。

22 Dolan, E. W. (2024, April 14). ChatGPT hallucinates fake but plausible scientific citations at a staggering rate, study finds. *PsyPost.* https://www-psypost-org.translate.goog/chatgpt-hallucinates-fake-but-plausible-scientific-citations-at-a-staggering-rate-study-finds/?_x_tr_sl=en&_x_tr_tl=zh-TW&_x_tr_hl=zh-TW&_x_tr_pto=sc (檢索日期：2024年8月4日)。

23 Endert, J. (2024, March 26). Generative AI is the ultimate disinformation amplifier. *DW*. https://akademie-dw-com.translate.goog/en/generative-ai-is-the-ultimate-disinformation-amplifier/a-68593890?_x_tr_sl=en&_x_tr_tl=zh-TW&_x_tr_hl=zh-TW&_x_tr_pto=sc (檢索日期：2024年8月12日)。

24〈ChatGPT技術恐成中國假訊息新利器？〉，《Rti中央廣播電臺》，2023年2月22日，https://www.rti.org.tw/news/view/id/2159936，檢索日期：2024年8月10日。

25 Silverberg, D. (2023, February 14). Could AI swamp social media with fake accounts? *BBC*. https://www-bbc-com.translate.goog/news/business-64464140?_x_tr_sl=en&_x_tr_tl=zh-TW&_x_tr_hl=zh-TW&_x_tr_pto=sc&_x_tr_hist=true (檢索日期：2024年8月10日)。

上述凸顯ChatGPT在假新聞生成中所扮演的認知核心角色。可巧妙地操控人類心理脆弱性，透過設計精妙之虛假訊息，深刻影響受眾之認知與情感反應。關鍵在於，此種認知操控不僅影響個人情感，還有潛力影響高層決策過程，對國家政治、經濟、軍事與社會穩定構成深遠威脅。

二、ChatGPT在社群媒體中之策略

在資訊快速發展脈絡下，全球訊息傳播領域面臨外部變革，特別是生成式人工智慧（如ChatGPT）應用，此模式如同1949年夏農（Claude Shannon）與韋佛（Warren Weaver）所出版《傳播的數位理論》(The Mathematical Theory of Communication)，它是以一個線性過程，強調訊號傳輸的傳播模型(資料源→傳送器→路徑→接收器→終點)。就像訊息生成、產生及傳播模式一般，使資訊傳遞更有效率。因此從上述訊息傳遞，可深入探究ChatGPT技術在各階段中的運作機制及其對資訊生態帶來的影響。

(一)訊息生成：它是傳播過程起點。此階段強調如何利用ChatGPT技術針對特定受眾設計資訊。此技術不僅能夠針對不同情感和文化背景實施調整，還可根據即時情緒變化，進而影響受眾。此過程中的訊息設計，往往帶有強烈地心理操控意圖。

(二)訊息生產：訊息生成後，緊接著是訊息生產階段，即將生成內容轉化為具體的媒體型態。例如新聞報導、部落格文章、社群媒體貼文甚至多媒體內容（如自動產生的圖像、視訊字幕等）等多種形式呈現，從多種管道來覆蓋更廣泛受眾。

(三)訊息傳播：訊息生產的最終目的在於實現訊息傳播，此階段的核心在於透過傳播管道，藉由多種平台整合，使訊息迅速覆蓋到不同受眾群體，達到更廣泛的傳播效果。

在此三階段結構基礎下，若ChatGPT能在社群媒體中策略地應用，其訊息的生成、產生與傳播效率與規模將會大幅提升，這過程顯然超越過去傳統新聞傳播模式所能達到速度與影響範圍。不僅對大眾的資訊素養產生不利影響，更失去決策者的判斷能力，間接造成政治、經濟、社會和軍事的穩定性構成重大威脅，其策略模式敘述如下(如圖4)：

(一)訊息過載與多平台整合

在高度互聯之社群媒體環境中，以上平台成為聚集網路言論核心樞紐，訊息散播呈現出洗版式高頻串流特性。與過去依賴被動傳遞之假新聞不同，現今透過ChatGPT生成虛假內容不僅侷限於大型社交平台，還能迅速滲透至如Medium、Quora等小型論壇及分眾網站。透過多渠道得以在廣泛受眾

中迅速散布。其互鏈策略顯著提升訊息觸及率，在短時間內引發資訊轟炸與過載效應，強化多平台輿論操控。[26]

(二)全球化與多生成語言

更具威脅性的是，ChatGPT憑藉其多語言生成能力，能夠將虛假訊息快速翻譯成50多種語言，從而顯著擴展其全球影響力，並加強對各國輿論操控。此種策略尤以推動特定政治敘事為核心，影響多國及區域輿論動態。從OpenAI報告指出，來自中共「垃圾偽軍」（Spamouflage）利用OpenAI技術分析社群媒體趨勢，並用多種語言（如中、韓、日、英）生成並傳播虛假資訊，更深化全球資訊操控影響力。[27]

(三)病毒式傳播與擴散

ChatGPT生成內容具高度個性化處理能力，能對不同受眾進行量身定制，實現大規模內容複製與轉傳。再藉由設置大量殭屍帳號和運營內容農場[28](Content Farms），使虛假資訊得以擴散。根據測試結果，此類虛假新聞可信度平均約12.22%。[29]然而，操控者可利用此策略精準針對特定受眾，提升顯著影響力。如同2023年7月，印第安納大學(Indiana University)研究揭示，社群平台上有超過1,100個由殭屍帳號組成宣傳網路，疑似由ChatGPT生成內容進行運作。[30]

(四)學術虛假與操縱

ChatGPT具備生成高度擬真學術研究與報告能力，能模仿學術論文結構、語言風格及研究方法，並製造虛構數據內容。此不存在之數據可悄然收錄於知名權威機構或學者研究成果中，從而提高可信度。一旦此類虛假研究經由主流媒體或社群平台傳播，極易被學術機構或政策制定者引用，更大幅擴大傳播範圍，對學術誠信與決策機制產生嚴重威脅。

(五)同溫層效應與輿論引導

在民主社會中，社群平台承載資訊散布的重要功能，而ChatGPT運

26 孔德廉、劉致昕，〈寫手帶風向不稀奇：AI產文、侵入私人LINE群，輿論軍火商已全面升級〉，《報導者The REPORTER》，2020年1月6日，https://www.twreporter.org/a/information-warfare-business-weapons，檢索日期：2024年8月11日。

27 譯陳昱婷，〈OpenAI：中俄等組織以AI技術試圖操縱全球輿論〉，《中央通訊社》，2024年5月31日，https://www.cna.com.tw/news/ait/202405310074.aspx，檢索日期：2024年8月23日。

28 「內容農場」指的是為了增加流量並賺取網路廣告收益而建立的網站，其內容通常缺乏原創性，真實性難以保證。此類網站不進行內容管理，經常使用侵權、抄襲或改寫的手段生產文章，導致大量摻雜虛假或浮濫訊息。目的是提升點擊率以謀取商業利益，甚至出於政治目的對政府施政造成衝擊。

29 王維菁、廖執善、蔣旭政、周昆璋，〈利用AI技術偵測假新聞之實證研究〉，《中華傳播學刊》(臺北市)，臺灣傳播學會，第39期，2021年6月，頁59。

30 洪欣慈，〈雲端影響力／救社群亂象？AI是把雙面刃〉，《聯合新聞網》，2024年7月8日，https://udn.com/news/story/124068/8079752，檢索日期：2024年8月25日。

作模式能將單一虛假特定議題迅速推向社群媒體。透過點擊、分享以及轉發等互動行為，快速獲得大眾可見度，營造出似乎廣泛支持某特定的假象。後續隨議題熱度攀升，使輿論走向遭受引導，從而形成同溫層、過濾氣泡或帶風向效應，進以加劇社會兩極分化，此行為對社會穩定構成嚴峻挑戰。[31]

簡言之，政治宣傳核心在於藉由反覆傳遞資訊，最終改變受眾認知。在網路議題推動下，訊息以「點─線─面」方式擴散，這也改變假新聞生成與傳播模式，也為未來戰爭型態帶來前所未有挑戰與不確定性。此時中共也已將資訊戰、輿論戰、心理戰與認知戰相互結合，採取「建構專業團隊→滲透輿論→奪取話語權→掌控輿論」四步走戰略。一旦付諸實施，生成假新聞將與真實新聞無縫融合，形成高度可信虛假輿論環境。

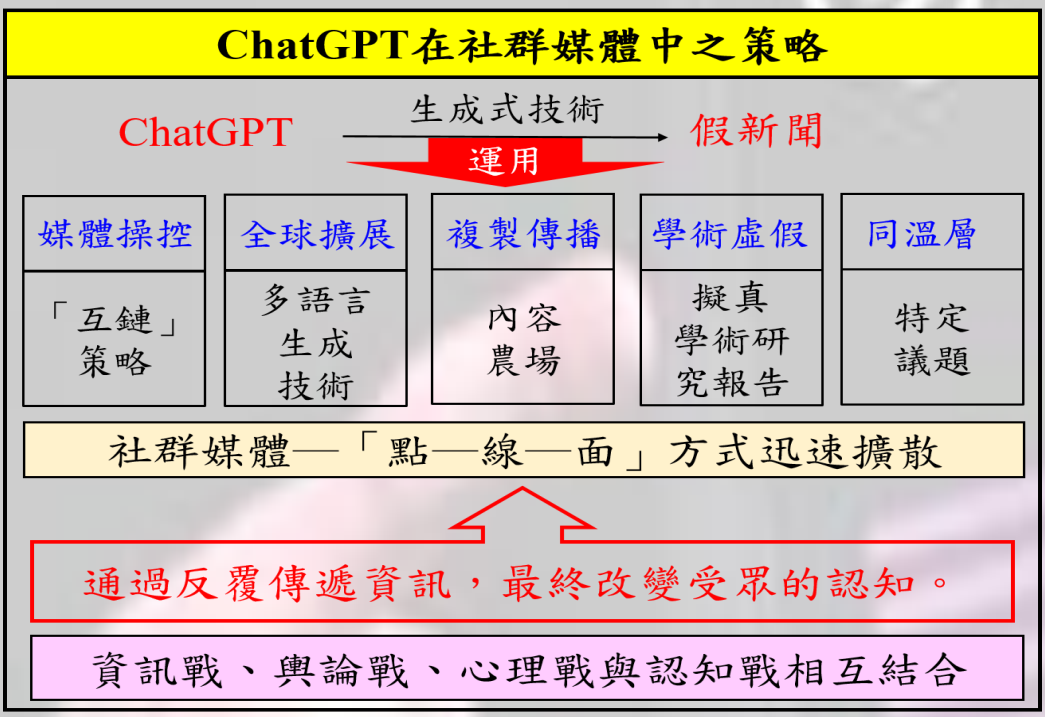

圖 4 ChatGPT 在社群媒體中之假新聞應用

資料來源：作者自行繪製

三、ChatGPT生成假新聞之事件分析

根據麻省理工學院（Massachusetts Institute of Technology, MIT）研究顯示，在推特（Twitter）上之假新聞散播速度是真實新聞的六倍。[32]一則假新聞在短短10小時內可傳播至1,500名用戶。而如今ChatGPT此類生成式AI技術應用已

31 鄭樺，〈ChatGPT成中共操控輿論工具 美智庫：恐用於臺灣總統選舉〉，《大紀元》，2023年9月12日，https://www.epochtimes.com/b5/23/9/12/n14072240.htm，檢索日期：2024年9月12日。

32 蔡琮浩，〈從德國、新加坡管理網路假訊息散播探討立法方向〉，《立法院法制局專題研究報告》(臺北市)，立法院，2020年2月，頁3。

成為資訊與認知操縱工具。

從2023年之政治風險諮詢公司歐亞集團（Eurasia Group）在《全球風險報告》中，將生成式人工智慧形容為「大規模擾亂性武器」（Weapons of Mass Disruption），此預測也漸在當前媒體環境中逐步驗證。[33]自2023年5月起，OpenAI宣布成功阻止五起試圖秘密操控輿論的行動，此行動主要由「俄羅斯、中共、以色列和伊朗」等境內私人企業策劃，旨在藉由虛假訊息影響公眾輿論，進而對區域局勢產生深遠影響，[34]而以上推動虛假訊息的行為者稱為「媒體操縱者」。[35]此類事件凸顯社群媒體在加劇「資訊碎裂化」中的作用，並進以反映了「碎片化時代」下，資訊過載與技術進步的雙重挑戰。

在此現象下，根據2024年9月《世界經濟論壇全球風險報告》分析指示，已成為全球短期內最嚴峻的風險之一，特別是在社會兩極化加劇，假新聞影響力僅次於極端氣候變化與國際武裝衝突，成為威脅全球社會穩定的核心因素。鑒於此潛在風險，本文將從「政治、經濟、軍事及社會心理」四個層面分析：

(一)政治層面分析

生成式假新聞已對全球多個民主國家構成挑戰。假訊息以其高度擬真的特性，不僅干擾選舉過程，影響政策制定，還可能引發國內政治動盪與對立，最終削弱公眾對民主制度的信任，並對政治穩定構成威脅。臺灣作為一個民主國家，特別容易受到此類假訊息侵擾，尤其在選舉期間更是虛假訊息攻擊的主要目標。

此外，極權國家如中共、俄羅斯與伊朗，利用生成式AI來操控輿論，擴大假訊息在國際上的影響力，特別在左右政治選舉上。蘭德公司的政策分析家莫小龍（Nathan Beauchamp-Mustafaga）曾指出，中共正在積極研究如何使用生成式AI技術，來操控全球輿論，並試圖在2024年臺灣總統選舉中加以應用，[36]以左右選民情緒來影響選舉結果。上述主要藉由情緒操控方式，增加社會對立分化，構成政治穩定長期威脅。此類干預行為，不僅損害民主制度正常運行，也在擴大政治不穩定性。

(二)經濟層面分析

33〈ChatGPT技術恐成中國假訊息新利器？〉，《Rti中央廣播電臺》，2023年2月22日，https://www.rti.org.tw/news/view/id/2159936，檢索日期：2024年8月10日。

34 譯張曉雯，〈濫用ChatGPT影響美大選 OpenAI封鎖伊朗團體帳號〉，《中央通訊社》，2024年8月17日，https://www.cna.com.tw/news/aopl/202408170096.aspx，檢索日期：2024年9月19日。

35 黃哲斌，〈如何與假新聞共處？現代公民必備的資訊抗體〉，《天下雜誌》，2019年3月27日，https://books.cw.com.tw/article/438，檢索日期：2024年9月19日。

36〈美智庫蘭德公司示警：中國生成式AI 可能擾亂臺灣2024總統大選〉，《自由時報》，2023年9月12日，https://news.ltn.com.tw/news/world/breakingnews/4426409，檢索日期：2024年10月28日。

除了在政治領域影響外，ChatGPT生成式技術在經濟領域也產生必然的作用。可憑藉強大運算能力，生成更逼真之金融新聞，一旦金融市場遭受此類假新聞的衝擊，會使投資者基於虛假資訊做出錯誤決策，從而加劇市場的不確定性與波動。如同2019年5月24日中共科大訊飛股價事件。[37]

此案例反映出生成式假訊息對經濟活動造成破壞性影響，尤其是在高度依賴即時數據與速度的金融市場中更為突顯。一旦錯誤訊息進行交易，將加劇市場波動性，進而擴大經濟體系的不穩定性，不僅影響短期市場運行，還對全球經濟秩序造成長期穩定性挑戰。[38]

(三)軍事層面分析

在軍事領域上，ChatGPT生成式技術已被應用於「認知作戰」中，可藉由構擬虛假敘事，模擬真實的戰場情境或戰報，從而影響公眾和決策者判斷。例如，俄羅斯運用ChatGPT生成虛假報導，聲稱烏克蘭戰爭中的「惡行」是西方媒體虛構，此資訊在社交媒體上廣泛傳播，進以擴大影響力。[39]

針對我國而言，中共長期應用虛假訊息進行輿論戰，特別是針對美國對臺政策提出各種錯誤觀點，例如「拋棄論」、「實力論」和「亂源論」等。主要在干擾戰略決策，使決策者對戰略形勢誤判。[40]因此，未來臺海衝突中，此技術將會大大提升資訊戰風險。現今更運用銳實力（Sharp Power）滲透至我國的資訊體系，以增加對手戰略脆弱性(如圖5)。此認知作戰策略或錯誤輿情不僅削弱軍隊應變能力，對整體軍事行動構成多層次挑戰。[41]

(四)社會心理層面分析

最後，在社會心理層面上，ChatGPT生成式技術所生成的虛假訊息對群體情緒和社會穩定性有負面影響。據臺灣事實查核教育基金會於2023年進行調查顯示，83%的臺灣受訪者在過去一年內接觸過假訊息，且90%的受訪者認為假訊息對社會的影響非常嚴重。[42]

---

37 黃心怡，〈AI造謠文引發科大訊飛股價下跌？文心一言反駁〉，《財聯社》，2023年5月25日，https://www.cna.com.tw/news/acn/202305250079.aspx，檢索日期：2024年10月28日。

38 高隆樺、李佳靜，〈企業面對AI聊天機器人崛起的資安風險與策略〉，《臺灣期貨雙月刊》(臺北市)，台灣期貨交易所，第81期，2024年6月17日，頁15。

39 羅世宏，〈散播虛假資訊、影響公眾認知：以國家力量進行的訊息操弄〉，《臺灣適時查核中心》，2024年10月17日，https://tfc-taiwan.org.tw/articles/11122，檢索日期：2024年10月28日。

40 劉兆隆，〈認知作戰的理論與應對之道〉，《清流雙月刊》(臺北市)，法務部調查局，第45期，2023年5月，頁9。

41 林政榮，〈中國大陸認知作戰對我政治作戰六戰之啟示〉，《空軍學術雙月刊》(臺北市)，空軍司令部，第697期，2023年12月，頁52。

42 臺灣事實查核中心，〈【2023假訊息年度大調查】愈闢謠反而「信者恆信」？政治澄清效力因政治傾向「打折扣」〉，《臺灣事實查核中心》，2023年5月19日，https://tfc-taiwan.org.tw/articles/9154，檢索日期：2024年9月20日。

根據海德特(Jonathan Haidt)對「社群直覺模式」[43]敘述，當ChatGPT與社交媒體結合時，假訊息能迅速引發「群體極化」效應。此訊息能讓不同群體在高情緒下更加堅定己見，並排斥異見，最終加劇社會內部分裂與對立。[44]例如，中共可在選舉期間散布假新聞，使在政治高度敏感的臺灣百姓產生效應，加劇社會兩極化使其對立，便可擾亂選舉活動。

綜上所述，生成式假新聞事件已在全球範圍內構成了全方位威脅(如表4)。臺灣作為一個民主社會，面臨的挑戰尤為嚴峻。為應對上述問題，我國必須制定更加精細且前瞻的政策，才能有效減少此類技術對國家穩定與發展帶來的負面效應。

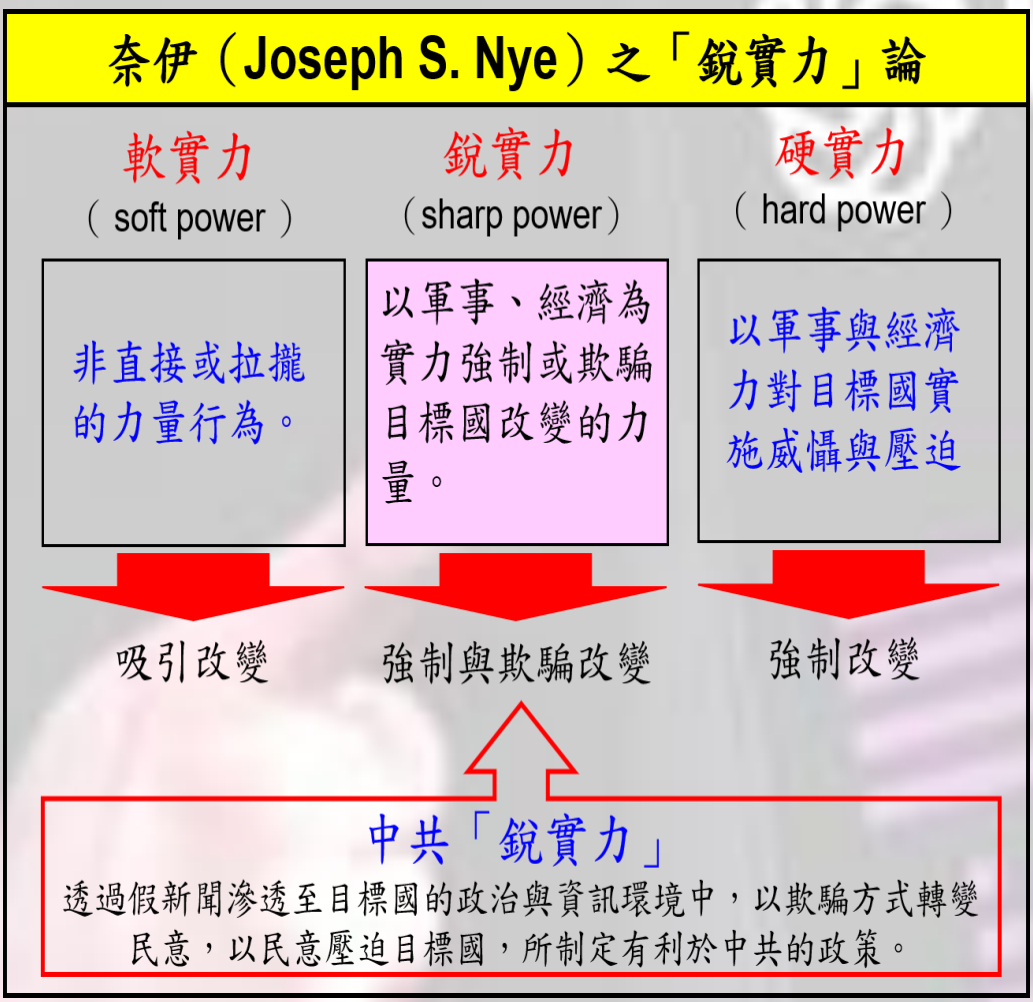

圖 5 中共「銳實力」

資料來源：參考劉文斌，〈中共「銳實力」運用中的假新聞防制〉，《清流雙月刊》(臺北市)，法務部調查局，第17期，2018年9月12日，頁12。

43 海德特的「社群直覺模式」是指大腦在認知過程中，產生一個現象。當在認知人事物中，相比於繁瑣的理性思考過程，大腦更傾向較簡單思路；簡言之，也就是「直覺」會先於「理性」。

44 林照真，〈新聞，在轉捩點上：數位時代的新聞轉型與聚合〉，(臺北：聯經)，2017年5月，頁52-53。

表 4 國內、外 ChatGPT 生成假新聞事件敘述及影響

項次	類 型	事 件 敘 述	具 體 影 響
一	政治	在 2023 年 5 月 NewsGuard 報導，使用 ChatGPT 生成的假新聞增加至 50 件，由「CelebritiesDeaths.com」網站錯誤報導，此標題為「拜登逝世、賀錦麗出任代理總統，上午 9 時將發表談話」。	透過生成虛假政治新聞並進行多語言全球傳播，干擾選舉過程與政策討論，更加影響國內外政治動態和決策。
		2024 年 8 月，路透社報導稱，代號「風暴-2035」行動使用 ChatGPT 生成並聚焦於多個政治議題(美國大選、以巴衝突、以色列參與奧運)，藉社群媒體進行大規模傳播。	
二	經濟	2019 年 5 月 24 日，生成式虛假消息導致科大訊飛股價盤中暴跌 9%，雖後來證實消息不實，但已對市場造成重大影響。	假新聞針對金融市場及企業聲譽，導致投資者行為改變和經濟決策波動，從而增加經濟不確定性，並引發信任危機。
		2023 年 8 月，我國金融監管會參考國際主要監管機構及組織的 AI 指導原則，制定「金融業運用人工智慧（AI）之核心原則與相關推動政策」草案，並推出八項相關配套措施，如有問題將對金融機構進行問責。	
三	軍事	2023 年 9 月，根據「蘭德公司」威廉斯（Heather Williams）指出中共正研究何以利用 ChatGPT 及類似 AI 技術，創造「合成資訊」，編造虛假內容以影響全球受眾認知。	軍事領域中的假新聞被用作「認知戰」工具，減低國內軍事信心和社會凝聚力，或透過快速生成虛假戰況報告來誤導敵方戰略決策。
		2024 年 10 月，OpenAI 發布的最新研究報告揭露，俄羅斯某些機構利用 ChatGPT 生成虛假文章，聲稱烏克蘭戰爭中的「惡行」是西方媒體所編造的，並使用 DALL-E 生成虛構的戰爭圖像，以吸引更多人點擊和轉發。	
		「TNewsNetwork」的網站，則發出一則關於烏俄戰爭，數千名俄軍喪生的故事，但完全未經查證。	
四	社會心理	2023 年 4 月，中共甘肅省透過百度搜尋引擎出現假新聞，報導稱：「甘肅省一列火車撞上工人，造成九人死亡。」同時有 21 個帳號發布相同文章，瀏覽量達 1.5 萬次。	此類重複發布與擴散行為引發「回音室效應」，降低公眾信任感，並加深社會焦慮與恐懼情緒。

資料來源：譯張曉雯，〈濫用 ChatGPT 影響美大選 OpenAI 封鎖伊朗團體帳號〉，《中央通訊社》，2024 年 8 月 17 日，https://www.cna.com.tw/news/aopl/202408170096.aspx，檢索日期：2024 年 9 月 19 日，張曉雯，〈美智庫：中國擬運用生成式 AI 操縱各界對臺看法〉，《中央通訊社》，https://www.cna.com.tw/news/aopl/202309120057.aspx，檢索日期：2024 年 10 月 28 日。

我國對生成式假新聞內容之建議

ChatGPT的廣泛應用在各行各業迎來顯著變革，但同時也帶來前所未有的安全隱患。從著名政治學者巴柏（B.Barber）1998提出的《潘朵拉劇情》（Pandora Scenario）認為，社群網路的開放性猶如潘朵拉盒子一般，尤其是「數位落差」（Digital Divide）進以惡化社會不平等。此悲觀預測在生成式AI的語境下尤為應驗，主因在於假新聞的快速生成、生產及傳播。[45]

在2023年7月29日臺北舉行的「調查通報與事實查核工作坊」上，臺灣大學資訊管理系副教授孔令傑於「AI時代的真假挑戰」講座中，強調面對生成式假新聞問題，應遵循理解生成機制、識別訊息來源、確認可信度三大核心原則。[46]以上原則不僅是為了保護接收者的知情權，也是為國人保障言論自由的同時，有效抵禦假新聞的侵蝕。因此，政府應對此威脅實施重視。基此，筆者為有效應對ChatGPT生成式假新聞帶來的挑戰，必須針對其內容生成、生產及最終傳播等三階段，提出具體防範策略。

一、假新聞生成階段之技術與控制

生成式假新聞源於AI技術驅動的訊息生成過程，此虛假內容往往具有高度誤導性並且能迅速擴散。因此，在生成階段需採取有效技術防範措施至關重要

(一)開發生成式假新聞檢測工具

推動國內、外技術產學合作政策，促進開發生成式假新聞檢測工具，如國內有中央研究院與成功大學等多家學者對假訊息辨識技術上均有研究。中央研究院團隊更運用人工智慧與機器學習技術，開發一套假新聞偵測系統，能夠自動判讀訊息的標題、內文、時間及來源等資訊。[47]此外，聯邦學習（Federated Learning）也是另一種技術解決方案。此技術允許不同機構在不共享資料的的前提下進行AI模型訓練，藉由國家或企業間的模型訓練合作，以提升假訊息辨識的準確度，不僅能降低資料標記成本，還能保護隱私，也可被快速識別並及時處理。[48]

同時，在打擊假新聞的過程中，AI假訊息辨識模型或軟體設置，須考量政治立場、宗教信仰、多元文化及價值觀等多層次之平衡及約束性，需要求系統

45 劉兆隆，〈臺灣大眾媒體假消息散布的政治效應〉，《中國地方自治》(台灣)，中國地方自治學會，第76卷2期，2023年2月，頁30-31。

46 馬麗昕，〈【2023調查報導與事實查核工作坊】AI時代的真假挑戰學者：民眾應了解AI、辨認訊息來源〉，《臺灣事實查核中心》，2023年7月29日，https://tfc-taiwan.org.tw/articles/9405，檢索日期：2024年9月19日。

47 張庭軒，〈辨識假訊息之科技研究分析〉，《工程技術研究發展司》，科技部109年度科技行政自行研究報告，民國109年12月，頁5。

48 AIF Editor，〈聯邦學習是什麼？近來備受討論的機器學習技術介紹〉，《知勢》，2022年7月8日 ，https://edge.aif.tw/book-federated-learning/，檢索日期：2024年12月5日。

在偵測假訊息時，能對訊息的正確性理解，避免單一標準的偏狹誤判或忽略重要訊息的情況發生。

(二)借鑒GPTZero的成功經驗

美國普林斯頓大學華裔學生愛德華·田(Edward Tian)於2023年開發的GPTZero程式，此程式展現自動化識別AI生成內容潛力，主要檢測文本的「困惑性」（Perplexity）和「突發性」（Burstiness）兩項指標，並分別對其評分，根據統計學特徵來確定，文本是由人工智慧還是人類編寫，且後續更持續研發「Origin」以提升識別內容的精準度，可是謂杜絕假新聞成功案例之一。[49]

(三)應用區塊鏈技術驗證新聞真偽

Polygon公司營運長邁克爾·布蘭克(Michael Blank)表示：「只要把內容上傳至區塊鏈上，就可驗證該內容是由某個人或品牌所創作」。[50]區塊鏈技術也是防範假新聞手段之一，主要因具備去中心化及不可篡改特性，其運作模式是將新聞內容上傳至區塊鏈，並透過多個節點共同驗證，可確保訊息來源的真實性，並有效防止內容篡改與假訊息傳播。

儘管區塊鏈技術具備防偽優勢，但仍存在辨識效率較低的問題。為提升運作效率，可將AI技術與區塊鏈相互結合，藉AI快速分析歷史數據預測能力，再與區塊鏈技術協同，提升其智能化程度，還能增強系統適應性，實現精準追蹤、預測與處理效率，從而有效遏制假訊息傳播。

二、假新聞生產之監管與審查

若單一從源頭管制，僅依僅靠技術層面(如區塊鏈或GPTZero程式等)仍然無法根除假新聞，因現今手段眾多，成本低廉。因此在生成假新聞之後，訊息會藉各種媒體平台轉化為具體內容，進行加工與再生產。此階段重點在於加強技術監管和內容審查，以防止虛假訊息擴散。

(一)建立專責生成式AI監管機構

應設立專門監管機構，負責針對生成式AI技術管理與風險評估。該機構應具備多層次的監管能力，包括數據篩選、內容審查以及技術風險之預警和應對。目前如臺灣事實查核中心（Taiwan FactCheck Center,TFC）雖已在假新聞防治方面發揮一定作用，但面對生成式AI的挑戰仍顯不足。因此，

49 美漪，〈ChatGPT剋星來了！22歲工程師發明GPTZero，靠「文字困惑度」抓包AI文章〉，《商周》， 2024年1月25日，https://www.businessweekly.com.tw/international/blog/3011599，檢索日期：2024年9月19日。

50 Sisley，〈區塊鏈可能是AI深偽爭議、AI 假新聞的最佳解答？〉，《INSIDE》，2024年3月15日 ，https://www.inside.com.tw/article/34478-blockchain-could-be-the-best-solution-to-deepfake-controversies-and-ai-fake-news， 檢索日期：2024年9月19日。

新設立監管機構應補充不足之處，確保生成式新聞內容能夠被全方位監控。

(二)設立跨部門協調的假新聞防治小組

成立跨部門的假新聞防治小組，有助於加強政府各部門間的協作，快速應對虛假訊息。該小組應依據「快速、公開、結構化」的原則，及時發布澄清訊息，防止假新聞擴散，並在必要時採取法律行動。歐盟成立的假新聞防治小組便是成功案例，每年總結策略與技術，並推廣採用AMITT框架（Adversarial Misinformation and Influence Tactics and Techniques），藉由多部門協作，有效抑制謠言的傳播並推動反制策略。[51]

(三)引進自動化事實查核工具及促進合作機制

馬克吐溫曾說—「當真相還在穿鞋時，謊言已走遍半個世界」。這些在在顯示，面對資訊量的急劇增長，依賴傳統手動查核方式在處理假新聞的議題，多半是緩不濟急，已無法及時應對生成式假新聞的威脅。因此，新聞機構引進自動化事實查核工具是有其必要性，並綿密與第三方事實查證機構(如美國的《PolitiFact》與西班牙的《Maldita》以及臺灣的蘭姆酒吐司、臺灣事實查核中心及MyGoPen等)以及社群媒體(平台)合作，才能提高查核效率。以上手段均應結合資料探勘技術進行多層次審核，確保虛假訊息在廣泛傳播前即被識別並阻止。

三、假新聞傳播之教育與監管

假新聞一旦透過社交媒體及其他傳播途徑擴散，將會影響廣泛受眾。為了有效抑制虛假訊息的擴散，必須在傳播階段採取強有力的公眾教育措施和立法管控作為。

(一)推動媒體素養教育的普及與深化

媒體素養教育是提高公眾抵禦假新聞能力的基礎性措施，媒體素養不應僅限於學校教育，而應普及至全社會，並納入終身學習的框架中。只有具備批判性思維和科技素養的公民，才能在面對虛假訊息時做出正確判斷。公眾教育計畫應強調如何辨識訊息真偽，提升整體社會的媒體識讀能力。

(二)融合批判性思維與科技素養

媒體素養教育應與批判性思維和科技素養訓練相結合。具體來說，應在教育過程中引入如「5W思考法」（Who、What、When、Where、Why）等工具，幫助受眾獨立分析訊息來源與內容真實性，減少受到假新聞影響

51 王仁甫，〈資安就是「國安 3.0」打假護真數發部責無旁貸〉，《yahoo 新聞》，2024 年 2 月 18 日，https://tw.news.yahoo.com/%E6%80%9D%E6%83%B3%E5%9D%A6%E5%85%8B-%E5%81%87%E6%96%B0%E8%81%9E%E9%98%B2%E6%B2%BB%E4%B9%8B%E9%81%93-010113352.html，檢索日期：2024 年 10 月 29 日。

之風險。此類教育計畫應全面推廣至社會各階層，從根本上提升公眾的資訊鑑別能力。

(三)制定生成式AI相關法規並強化監管

面對生成式AI技術帶來的挑戰，必須加強法律監管並制定相應的政策框架。在保障言論自由的前提下，應設立包括透明度要求、內容審查標準以及責任追溯機制在內的法律規範。可參考歐盟《數位服務法》（Digital ServicesAct）等國際立法範例，強化技術開發者和數位平台在偵測與防範假新聞方面的責任，確保生成式AI技術的應用符合法律與道德標準。

(四)促進國際間立法合作應對全球假新聞挑戰

假新聞的傳播具有跨國性，單靠一國之力難以有效應對。國際間應加強協作，推動全球一致的規範標準。參考歐盟《人工智能法》（AI Act）等國際規範，我國應積極參與國際立法合作，推動生成式AI技術的全球法律框架，確保跨國假新聞的防控得以有效執行，並促進數位平台的公信力。

上述措施除了可辨識「具敏感性的議題」與「偏激的情緒用詞」等假新聞生成外，也可針對以圖搜尋工具查證圖片的來源與歷史，自行查核資訊的真偽。最普遍的管道是使用Google網頁版的以圖搜圖功能，能快速方便辨識真假，在國外的Online Exif Viewer與Metapics等搜圖網站，都是被大眾廣為通用的識別網站。以上為技術方面，但最主要還是需藉教育和法律多方面著手，以全面應對生成式假新聞對社會、政治和經濟穩定帶來的威脅。

結論

ChatGPT迅速發展下，假新聞在社群媒體中的傳播模式日益複雜，並變得愈發難以監控和防範。社群媒體之去中心化特性，結合高度傳遞速度與廣泛影響力，為假新聞擴散提供理想條件。生成式AI雖顯著提升訊息生成效率，但也同時為不實訊息產生與散播開啟更多可能性。因此，此現象進行系統性研究，尤其是ChatGPT更能深入揭示假新聞在社群媒體平台上運作邏輯與傳播機制。

生成式假新聞往往以具吸引力標題及情緒化內容來激發點擊，此訊息借助ChatGPT的自動化生成能力，大大擴展假新聞規模與覆蓋面。此技術不僅增加識別假新聞難度，其自然語言生成功能也使假新聞在形式上顯得更加真實與可信，進以強化在社群媒體上散布效果。

再者，ChatGPT推動假新聞傳播模式暴露社群媒體平台在內容監管上之局限性。由於平台監控能力的有限性，加上使用者對資訊需求急劇增長，假新聞在此過程中難以有效追蹤與遏制。不僅影響到輿論環境透明性，還對民主制度

穩定構成潛在威脅。因此，解決假新聞在社群媒體中擴散問題，需從多層面共同努力。首先，須發展專門針對生成式AI偵測與防範技術，通過AI監控工具進行即時檢測。其次，政府與科技公司需加強合作，制定明確的政策框架和法律規範，對生成內容使用進行管理。此外，社會各界應提升公眾媒體素養與批判性思維，培養公民對假新聞辨識和抵抗能力。

總結來看，假新聞在社群媒體中的操作模式，既揭示生成式AI技術挑戰，也突顯全球資訊治理迫切性。未來若要實現透明、公正且安全資訊環境，還須通過技術創新、政策規範與教育普及全方位努力，建立更安全的全球資訊治理體系，方能有效應對ChatGPT在假新聞傳播中潛在威脅。

參考文獻

中文部分

一、書籍

林照真，〈新聞，在轉捩點上：數位時代的新聞轉型與聚合〉，(臺北：聯經)，2017年5月。

二、期刊專題

(一)王亞珅、李强、石戈，〈ChatGPT對社交機器人技術發展的影響分析〉，《無人系統技術》(北京市)，海鷹科技情报研究所，第6卷第2期，2023年。

(二)王維菁、廖執善、蔣旭政、周昆璋，〈利用AI技術偵測假新聞之實證研究〉，《中華傳播學刊》(臺北)，臺灣傳播學會，第39期，2021年6月。

(三)林政榮，〈中國大陸認知作戰對我政治作戰六戰之啟示〉，《空軍學術雙月刊》(臺北市)，空軍司令部，第697期，2023年12月。

(四)施達妮、顏妤恬，〈數位時代的假新聞〉，《漢學研究通訊》(臺北)，國家圖書館，第147期，2018年8月。

(五)高隆樺、李佳靜，〈企業面對AI聊天機器人崛起的資安風險與策略〉，《臺灣期貨雙月刊》(臺北)，臺灣期貨交易所，第81期，2024年6月17日。

(六)張庭軒，〈辨識假訊息之科技研究分析〉，《工程技術研究發展司》，科技部109年度科技行政自行研究報告，民國109年12月。

(七)曾敏禎，〈中國公布《生成式AI管理辦法》—推進器或緊箍咒？〉，《國防安全雙週報》(臺北)，財團法人國防安全研究院，2023年5月12日。

(八)黃仲宏，〈活用生成式人工智慧擘劃機器人自動化的發展〉，《機械工業雜誌》(竹東)，財團法人工業技術研究院，第485期，2023年8月。

(九)葉繼元、郭衛兵，〈生成式人工智慧參與學術評價的反思〉，《中國社會科學評價》(北京市)，中國社會科學院，第1期，2024年6月11日。

(十)董慧明，〈生成式技術發展對國家安全的影響與挑戰〉，《清流雙月刊》(臺北)，法務部調查局，第48期，2023年11月。

(十一)劉文斌，〈中共「銳實力」運用中的假新聞防制〉，《清流雙月刊》(臺北)，法務部調查局，第17期，2018年9月12日。

(十二)劉兆隆，〈臺灣大眾媒體假消息散布的政治效應〉，《中國地方自治》(臺灣)，中國地方自治學會，第76卷2期，2023年2月。

(十三)劉兆隆，〈認知作戰的理論與應對之道〉，《清流雙月刊》(臺北)，法務部調查局，第45期，2023年5月。

(十四)蔡琮浩，〈從德國、新加坡管理網路假訊息散播探討立法方向〉，《立法院法制局專題研究報告》(臺北)，立法院，2020年2月。

三、網路：

(一)〈ChatGPT技術恐成中國假訊息新利器？〉，《Rti中央廣播電臺》，2023年2月22日，https://www.rti.org.tw/news/view/id/2159936。

(二)〈美智庫蘭德公司示警：中國生成式AI 可能擾亂臺灣2024總統大選〉，《自由時報》，2023年9月12日，https://news.ltn.com.tw/news/world/breakingnews/4426409。

(三)AIF Editor，〈聯邦學習是什麼？近來備受討論的機器學習技術介紹〉，《知勢》，2022年7月8日 ，https://edge.aif.tw/book-federated-learning/。

(四)Sisley，〈區塊鏈可能是AI深偽爭議、AI 假新聞的最佳解答？〉，《INSIDE》，2024年3月15日 ，https://www.inside.com.tw/article/34478-blockchain-could-be-the-best-solution-to-deepfake-controversies-and-ai-fake-news。

(五)周秉誼，〈深度學習與ChatGPT〉，《國立臺灣大學技資中心電子報》，2023年6月20日，https://www.cc.ntu.edu.tw/chinese/epaper/home/20230620_006503.html。

(六)孔德廉、劉致昕，〈寫手帶風向不稀奇：AI產文、侵入私人LINE群，輿論軍火商已全面升級〉，《報導者The REPORTER》，2020年1月6日，https://www.twreporter.org/a/information-warfare-business-weapons。

(七)王仁甫，〈資安就是「國安3.0」 打假護真數發部責無旁貸〉，《yahoo新聞》，2024年2月18日，https://tw.news.yahoo.com/%E6%80%9D%E6%83%B3%E5%9D%A6%E5%85%8B-%E5%81%87%E6%96%B0%E8%81%9E%E9%98%B2%E6%B2%BB%E4%B9%8B%E9%81%93-010113352.html。

(八)臺灣事實查核中心，〈【2023假訊息年度大調查】愈闢謠反而「信者恆信」？政治澄清效力因政治傾向「打折扣」〉，《臺灣事實查核中心》，2023年5月19日，https://tfc-taiwan.org.tw/articles/9154。

(九)臺灣科技媒體中心，〈ChatGPT回答錯誤？如何不被生成式AI假訊息詐騙？｜ChatGPT使用須知8件事〉，《未來城市》，2023年3月26日，https://futurecity.cw.com.tw/article/2990?rec=i2i&from_id=3143&from_index=8。

(十)田孟心，〈研究揭露：ChatGPT如何助長虛假訊息？〉，《天下雜誌》，2023年2月16日，https://www.cw.com.tw/article/5124734。

(十一)吳維雅，〈生成式AI的善與惡（一）：ChatGPT為何成為當代顯學？〉，《鳴人堂》，2023年5月8日，https://opinion.udn.com/opinion/story/120817/7149512。

(十二)沙珮琦，〈你的祕密，ChatGPT 全知道？新工具背後隱藏的資安風險〉，《科技魅癮》，2023年6月9日，https://www.charmingscitech.nat.gov.tw/post/worldview10-chatgpt。

(十三)洪欣慈，〈雲端影響力／救社群亂象？AI是把雙面刃〉，《聯合新聞網》，2024年7月8日，https://udn.com/news/story/124068/8079752。

(十四)美漪，〈ChatGPT剋星來了！22歲工程師發明GPTZero，靠「文字困惑度」抓包AI文章〉，《商周》，2024年1月25日，https://www.businessweekly.com.tw/international/blog/3011599。

(十五)馬麗昕，〈【2023調查報導與事實查核工作坊】AI時代的真假挑戰學者：民眾應了解AI、辨認訊息來源〉，《臺灣事實查核中心》，2023年7月29日，https://tfc-taiwan.org.tw/articles/9405。

(十六)陳品潔，〈chatGPT紅遍全球大量文本彈指生成訊息虛實難辨〉，《卓越新聞電子報》，2023年3月16日，https://feja.org.tw/69079。

(十七)黃心怡，〈AI造謠文引發科大訊飛股價下跌？文心一言反駁〉，《財聯社》，2023年5月25日，https://www.cna.com.tw/news/acn/202305250079.aspx。

(十八)曾敏禎，〈中國版的ChatGPT，鸚鵡學舌？畫虎類犬？〉，《國防安全研究院》，2023年2月15日，https://indsr.org.tw/focus?uid=11&typeid=30&pid=575。

(十九)黃哲斌，〈如何與假新聞共處？現代公民必備的資訊抗體〉，《天下雜誌》，2019年3月27日，https://books.cw.com.tw/article/438。

(二十)劉文瑜，〈研究：受眾憂AI產製內容為新聞機構帶來新挑戰〉，《中央通訊社》，2024年6月17日，https://www.cna.com.tw/news/aopl/202406170026.aspx。

(二十一)劉芮菁，〈ChatGPT就是「一本正經地胡說八道」！杜奕瑾：AI生成科技恐讓假訊息更氾濫〉，《今周刊》，2023年3月27日，https://www.busi

nesstoday.com.tw/article/category/183027/post/202303270036/。

(二十二)鄭樺，〈ChatGPT成中共操控輿論工具 美智庫：恐用於臺灣總統選舉〉，《大紀元》，2023年9月12日，https://www.epochtimes.com/b5/23/9/12/n14072240.htm。

(二十三)顏翩翩，〈ChatGPT帶來人工智慧新衝擊資策會科法所：應關注其風險級〉，《資策會科技法律研究所》，2023年3月6日，https://www.cna.com.tw/postwrite/chi/336122。

(二十四)羅世宏，〈散播虛假資訊、影響公眾認知：以國家力量進行的訊息操弄〉，《臺灣事實查核中心》，2024年10月17日，https://tfc-taiwan.org.tw/articles/11122。

(二十五)譯張曉雯，〈濫用ChatGPT影響美大選 OpenAI封鎖伊朗團體帳號〉，《中央通訊社》，2024年8月17日，https://www.cna.com.tw/news/aopl/202408170096.aspx。

(二十六)譯陳昱婷，〈OpenAI：中俄等組織以AI技術試圖操縱全球輿論〉，《中央通訊社》，2024年5月31日，https://www.cna.com.tw/news/ait/202405310074.aspx。

英文部分

(一)AIContentfy. (2023, November 6). ChatGPT and the future of news: Automation and AI. *AIContentfy*. https://aicontentfy.com/en/blog/chatgpt-and-future-of-news-automation-and-ai。

(二)Alkaissi, H., & McFarlane SI. （2023, February 19） Artificial hallucinations in ChatGPT: implications in scientific writing. *Cureus*. https://doi.org/10.7759/cureus.35179。

(三)Dekens, N. (2023, July 7). How ChatGPT could spread disinformation via fake reviews. *Spiceworks*. https://www-spiceworks-com.translate.goog/tech/artificial-intelligence/guest-article/how-chatgpt-could-spread-disinformation-via-fake-reviews/?_x_tr_sl=en&_x_tr_tl=zh-TW&_x_tr_hl=zh-TW&_x_tr_pto=sc。

(四)Dolan, E. W. (2024, April 14). ChatGPT hallucinates fake but plausible scientific citations at a staggering rate, study finds. *PsyPost*. https://www-psypost-org.translate.goog/chatgpt-hallucinates-fake-but-plausible-scientific-citations-at-a-staggering-rate-study-finds/?_x_tr_sl=en&_x_tr_tl=zh-TW&_x_tr_hl=zh-TW&_x_tr_pto=sc。

(五)Endert, J. (2024, March 26). Generative AI is the ultimate disinformation amplifier. *DW*. https://akademie-dw-com.translate.goog/en/generative-ai-is-the-ultimate-

disinformation-amplifier/a-68593890?_x_tr_sl=en&_x_tr_tl=zh-TW&_x_tr_hl=zh-TW&_x_tr_pto=sc。

(六)Honrada, G. (2023, September 12). China ramps up AI-powered campaign against Taiwan. *The Geopolitics*. https://thegeopolitics-com.translate.goog/china-ramps-up-ai-powered-campaign-against-taiwan/?_x_tr_sl=en&_x_tr_tl=zh-TW&_x_tr_hl=zh-TW&_x_tr_pto=sc。

(七)James Laird. (2023, June 29).ChatGPT Makes Spotting Fake News Impossible for Most People. https://tech-co.translate.goog/news/chatgpt-spotting-fake-news-impossible?_x_tr_sl=en&_x_tr_tl=zh-TW&_x_tr_hl=zh-TW&_x_tr_pto=sc。

(八)Silverberg, D. (2023, February 14). Could AI swamp social media with fake accounts? BBC. https://www-bbc-com.translate.goog/news/business-64464140?_x_tr_sl=en&_x_tr_tl=zh-TW&_x_tr_hl=zh-TW&_x_tr_pto=sc&_x_tr_hist=true。

作者簡介

曾柏元上校，陸軍軍官學校90年班、陸軍步兵學校正規班339期、陸軍指揮參謀學院101年班、國立政治大學戰略與國際關係研究所碩士。曾任排、連、營長、教官，現任國防大學教官。