法學院 論著
《高大法學論叢》
第 18 卷第 2 期（3/2023），頁 57-108

台灣假訊息管制的未來展望
－以規範與科技的互動為核心

李岳軒*、林志潔**

摘要

台灣在經歷 2018 與 2020 兩次大選、以及 2021 年 Covid-19 疫情的挑戰後，針對假訊息的修法、執法以及相關政策的修訂，有了初步的方向與成果。另外在「假訊息監理科技」方面的研究，也開始有所進展。惟假訊息的管制涉及言論自由等基本人權，不可不慎。本文認為，科技與法規政策，同為數位時代管制假訊息的工具，兩者如何相輔相成，則仰賴規範制定者的智慧。並且，在理解科技作為監理工具的可能性之後，必須更進一步地思考此類科技被濫用的可能。依循此脈絡，本文針對我國假訊息監理，提出兩個核心觀點：「立法者（Lawmakers）與程式設計

* 國立陽明交通大學科技法律學院博士生；美國 Duke 大學法學碩士。
負責題目的選定、架構的訂定、理論基礎的建立及論述、文獻的蒐集與分析。
**國立陽明交通大學科技法律學院特聘教授；美國 Duke 大學法學博士。
負責細部內容的鋪陳及論述、相關文獻的蒐集與分析與編輯修改。

接受刊登日期：12/27/2022
責任校對：葉蕙禎、容珮瑜

師（Codemakers）的跨領域對話與合作。」以及「建立與假訊息監理科技相關的科技治理或倫理規範。」

本文由分析假訊息議題的嚴重性與複雜性切入，點出目前我國修法所帶有的「保護法益」觀點。於第三章引入科技的觀點，例舉國內外針對假訊息監理科技與人工智慧的相關研究，呈現科技監管假訊息的可能性以及隱含的言論自由等基本權侵害風險。於第四章探討規範制定者與監理科技的互動模式，並借鑒歐盟於 2021 年頒布的《人工智慧法律調和規則草案》，並以「風險評估」、「保護法益與基本權衡平」兩個面向，分析我國在未來針對此類假訊息監理科技，訂定倫理規範與政策的必要性以及可能的參考方向。

The Future of Misinformation Regulations in Taiwan

-The Interaction between Regulation and Technology

Yueh-Hsuan Lee[***]、Chih-Chieh Lin[****]

Abstract

After two major elections in 2018 and 2022, and the 2021 COVID-19 pandemic, Taiwan has developed its primary policies and regulations toward misinformation. Besides legal frameworks, there were also researches on fighting misinformation with technologies and artificial intelligence. However, regulating fake news is constantly in conflict with basic human rights such as freedom of speech. This article argues that both law and technology are tools for managing misinformation. Also, avoiding malicious uses of regulatory technologies of misinformation is also a critical issue. Therefore, forming and developing a regulatory policy that integrates law and technology relies on the regulator's wisdom and understanding of technologies. In this regard, this paper proposes two core goals: "promoting the interaction between law-makers and code-makers" and "establishing governance policies or ethic codes for regulatory technologies of misinformation."

*** S.J.D. Student, School of Law, National Yang Ming Chiao Tung University; LL.M, School of Law, Duke University, U.S.A.

**** Professor, School of Law, National Yang Ming Chiao Tung University; S.J.D, School of Law, Duke University, U.S.A.

This article begins with an analysis of the severity and complexity of the misinformation problem, following by reviewing Taiwan's current "legal-interest focus" legislative actions and amendments. Chapter 3 will include the technology point of the view, giving examples of current research and future possibilities for using technology to combat misinformation. Chapter 4 will further analyze the possible in relation to ways technology and regulation that interact with each other, and the possible conflict within the interaction. This article will use Europe Union Commission's latest proposal act on artificial intelligence as an example and conclude that ethical code and technology governance is essential for Taiwan to further develop its strategies on combatting misinformation.

台灣假訊息管制的未來展望

－以規範與科技的互動為核心

李岳軒、林志潔

目錄

壹、前言－假訊息的多方治理框架與科技和法律的互動

貳、假訊息的成因與影響

一、經濟層面

二、社會層面

三、國安層面

四、小結

參、AI 在處理假訊息問題中扮演之角色

一、人工為主的模式：以臉書（Meta）為例

二、以 AI、機器人為主的模式

三、我國以 AI 監管假訊息的技術

四、管制者與假訊息監理科技的互動

肆、使用假訊息監理科技的隱憂與科技倫理問題

一、假訊息監理科技的潛在問題

二、歐盟《人工智慧法律調和規則草案》

伍、我國假訊息監理科技規範的未來展望

一、風險分類

二、保護法益與基本權衡平

三、小結
陸、結論

關鍵字：假新聞、假訊息、監理科技、人工智慧、言論自由

Keywords: Fake News, Misinformation, Regulatory Technology, Artificial Intelligence, Freedom of Speech

壹、前言－假訊息的多方治理框架與科技和法律的互動

「假訊息」（或稱假消息、假新聞）這一名詞近年來躍入了大眾的視線中，從 2016 美國總統大選開始，此議題就獲得了各界空前的關注度[1]，各國也開始著手進行相關的對策或討論。而我國，在所謂的「衛生紙之亂[2]」以及 2018 年九合一大選時，疑似遭受境外勢力利用假訊息操弄選情[3]等事件後，也促使了政府對於假訊息管制議題做出正面的回應，開始了一系列的法規修正，包含《刑法》、《社會秩序維護法》、《公職人員選舉罷免法》、《災害防救法》、《傳染病防治法》等十多部，以及以納入「網際空間」管制為目的而修正的《國安法》等[4]。大規模的修法使得我國假訊息與謠言的管制散佈於諸多不同法規範之中，呈現多頭馬車之現象[5]。2021 年中隨著台灣 Covid-19 疫情的爆發，假訊息對防疫工作以及社會秩序產生的影響，以及執法單位大

---

[1] David O. Kleina & Joshua R. Wuellera, *Fake News: A Legal Perspective,* 20 J. Internet L. 1, 6 (2017).

[2] 自由時報（04/18/2019），打擊假消息！扼阻衛生紙之亂 刑法修法重罰，https://news.ltn.com.tw/news/society/breakingnews/2763257（最後瀏覽日：02/17/2022）。

[3] Chris Horton, Specter of Meddling by Beijing Looms Over Taiwan's Elections, The New York Times, at https://www.nytimes.com/2018/11/22/world/asia/taiwan-elections-meddling.html (last visited 02/17/2022).

[4] 羅承宗（2019），〈虛假訊息與法律管制－我國現況與建議〉，《台灣法學雜誌》，369 期，頁 52-60。

[5] 王服清（2020），〈假消息：謠言或不實訊息的規範競合關係－以衛生紙之亂為例〉，《台灣法學雜誌》，390 期，頁 72-73。

規模地偵辦此類案件，再一次地使各界意識到此議題的重要性[6]。在民間，自 2016 年以來，許多以闢謠、事實查核與澄清假訊息為目的而成立的組織，如雨後春筍般出現，例如台灣事實查核中心、Cofactst 真的假的、Mygopen、蘭姆酒吐司等。而大型跨國科技公司如 Meta、Google、LINE 也各自有針對假訊息問題所做出的相關對策。

儘管各國已意識到假訊息的嚴重性，並開始研擬反制的措施，但具體的手段與執行方式等，仍處於發展階段，且具體效果亦尚不明確。而當政府介入假訊息、等問題的治理，可能造成對憲法言論自由的保障的威脅，甚至有造成寒蟬效應的疑慮[7]。因此，有論者指出，隨著全球網路治理所發展出來的多方利害關係人治理（Multistakeholder Governance），應該可以運用在假訊息治理上，使政府、立法者、科技業者、平台業者與民間組織共同來探討與規劃假訊息治理的框架[8]。因此，除了立法處罰層面的討論外，考量到資訊社會以及網際網路時代對於假訊息傳播的推波助瀾[9]，以及如何更有效率地遏止假訊息的流傳，目前已有許多單位開始研究以資訊科技、機器人甚至人工智慧（Artificial Intelligence）來處理假訊息問題[10]。一旦此種「以科技對抗科

---

[6] 鏡周刊（09/12/2021），假訊息趁本土疫情爆發亂竄 檢警 4 個月偵辦 636 件移送 489 人，https://www.mirrormedia.mg/story/20210912edi008/（最後瀏覽日：02/17/2022）。

[7] 張文貞（2019），〈2018 年憲法發展回顧〉，《臺大法學論叢》，48 卷特刊，頁 1534-35。

[8] Vidushi Marda & Stefania Milan, *Wisdom of the Crowd: Multistakeholder Perspectives on the Fake News Debate,* Internet Pol'y Rev. Series, Annenberg Sch. of Comm. 1-5 (2018).

[9] 施達妮著，顏妤恬譯（2018），〈數位時代的假消息〉，《漢學研究通訊》，37 卷 3 期，頁 9-10。

[10] iThome，MIT 與 Qatar 科學家打造 AI 系統，辨識假消息從來源著手，

技」的監理模式成為趨勢，則需要仰賴各個不同領域的專家對話與合作，以創造跨領域的假訊息監理機制。其中，立法者（Lawmakers）與軟體工程師（Codemakers）的持續與有效的對話，是其中重要的一環[11]。規範制定者在討論如何制定假訊息管制政策時，必須要對科技的可能性有一定程度的理解，方能制定出更有效、全面的政策。我國近年來雖然已經有針對假訊息防治進行的修法，惟目前的修法方式仍屬於傳統上「發現問題便制定規範與罰則」的模式，來針對各種不同的假訊息訂定相關處罰規範。在新興科技出現後，未來的相關修法以及政府政策的制定不應止步於此，而是應該逐步將科技的觀點導入立法過程中。

而人工智慧究竟係以何種方式來防範假訊息？其與依靠人力來識別假訊息的差異與優劣何在？在新興科技出現後，又能帶給規範制定者何種啟示？本文將以上述議題之探討為核心，梳理資訊時代下假訊息問題以及相關的防治手段；例舉實際已經投入運用、或是正在研發的技術，探討 AI 作為管制決策一環時的優劣以及可能性。另外，AI 與科技創新所伴隨的倫理問題，在近年來亦受到各國重視。例如歐盟已在 2021 年發布《人工智慧法律調和規則草案》，而我國科技部也曾在 2019 年頒布「人工智慧科研發展指引」，行政院也將在 2022 年成立「數位發展部」。有鑑於我國目前對於 AI 倫理規範的討論仍屬稀缺，本文將以歐盟最新的 AI 草案作為借鏡，討論假訊息監理科技可能涉及的基本權侵害問題，以及我國規範制定者在未來可能的因應方式。

承上，有鑑於新興科技將於未來成為假訊息治理的一環，本

---

https://www.ithome.com.tw/news/126302（最後瀏覽日：02/17/2022）。

11 Vidushi Marda & Stefania Milan, *supra* note 8, at 11.

文著重探討此類「假訊息監理科技」的可能性，以及分析此類科技和人工智慧，未來被投入運用在打擊假訊息問題時，政府作為政策與規範的制定者，應該如何做出回應。因此，本文提倡我國應朝向「帶有科技觀點的假訊息治理政策與規範制定」。在此議題上，近程目標應為「促進 Lawmakers 與 Codemakers 的對話」，而遠程目標則為「制定科技倫理或監理規範」。

貳、假訊息的成因與影響

假訊息並非一個全新的概念，無論是商業、或是政治上的假訊息，都已經存在許久[12]，是名符其實的「老問題」。然而，這樣的「老問題」，到了數位時代，被賦予了「新樣貌」。除了網路所帶來的匿名性與訊息傳播速度與廣度的提升，社群網站的發展，更加劇了假訊息的發展[13]。而新興科技如生成對抗網路（Generative Adversarial Network）、機器學習、社群網站的演算法等技術，也在過去數年間成為數位時代假訊息問題的一環。以下就假訊息的成因以及對於社會的影響與衝擊，簡單分類成經濟、社會與國安三個層面分析之。

一、經濟層面

在外國針對假訊息對經濟造成影響的討論不多，但對於台灣而言，在 2018 年初的「衛生紙之亂」，當時媒體相繼報導衛生

[12] Lili Levi, *Real "Fake News" And Fake "Fake News"*, 16 First Amend. L. Rev. 232, 233 (2018).

[13] Louis W. Tompros et al., *The Constitutionality of Criminalizing False Speech Made on Social Networking Sites in A Post-Alvarez, Social Media-Obsessed World*, 31 Harv. J. L. & Tech. 65, 66-67 (2017).

紙價將上漲三成的新聞，在國內引發衛生紙囤積與搶購潮[14]。讓政府與國內民眾均意識到在虛擬世界流傳的假訊息，如何影響現實世界的經濟秩序，進而導致了刑法以及相關法規的修正[15]。這項經驗，讓我們得以一窺謠言或假訊息對於市場經濟可能造成的衝擊，以及提供各界思考如何在未來避免此類情形重演，以及預防類似謠言造成更大規模的經濟衝擊。另外，假訊息亦造成可觀的社會成本，因為製造一篇假訊息所需的成本，往往遠低於製作一篇真實新聞、或澄清一篇假訊息所需耗費的人力、時間與金錢[16]。

不僅如此，假訊息背後常常代表的是龐大的經濟利益，而社群平台注重點擊率的廣告分紅機制，又助長了此一現象。例如紐約時報曾報導，2016 年美國總統大選時，一位民眾花了 15 分鐘寫了一篇關於希拉蕊偽造選票的假訊息，讓他透過 Google 廣告的分紅機制，賺進了 5000 美元，也因此鼓勵他繼續撰寫更多類似內容來獲取龐大的報酬[17]。這類以聳動標題或假訊息來吸引用戶眼球的文章，又被稱為「誘餌式標題（Click Bait）」，其背後最大的誘因，也是受到點擊率與廣告經濟效應的驅使[18]。社群平台或搜尋引擎強調用戶「點擊率市場」的機制，恰好讓假訊息在這樣的「供給、需求」模型下，朝向更具規模化經營的方向。

---

[14] ETtoday 新聞雲（02/23/2018），衛生紙 3 月起要漲價了！漲幅高達 10%～30%，民湧賣場囤貨，https://www.ettoday.net/news/20180223/1118030.htm#ixzz5rfSin8is（最後瀏覽日：02/17/2022）。

[15] 同前註 2。

[16] Alexandra Andorfer, *Spreading Like Wildfire: Solutions for Abating the Fake News Problem on Social Media via Technology Controls and Government Regulation*, 69 Hastings L.J. 1409, 1423-1424 (2018).

[17] Andrew Higgins et al., Inside a Fake News Sausage Factory: 'This Is All About Income', The New York Times, at https://www.nytimes.com/2016/11/25/world/europe/fake-news-donald-trump-hillary-clinton-georgia.html (last visited 02/17/2022).

[18] Lili Levi, *supra* note 12, at 246.

二、社會層面

假訊息對於社會造成的影響相當廣泛，如前所述，這些假訊息造成的「結果」，往往又是導致更多假訊息出現的「原因」。例如，當社會上假訊息的問題層出不窮時，可能使人民對於主流、傳統媒體的信賴程度下跌，導致正確報導的能見度與影響力降低。有外國研究指出，在這種情形下，人民將更可能受到來源不明的假訊息影響[19]。在探討假訊息與社會問題時，必定會提及的概念之一，便是在 2016 年被牛津辭典選為年度詞彙的「後真相（Post-Truth）[20]」。此一詞彙的定義為「訴諸情感及個人信念，較陳述客觀事實更能影響輿論的情況。（Circumstances in which objective facts are less influential in shaping public opinion than appeals to emotion and personal belief.[21]）」或以更白話的方式來說，便是指整個資訊社會與公共討論空間，由傳統的「事實勝於雄辯」，轉向「雄辯勝於事實[22]」。此一詞彙雖存在已久，但在 2016 年美國總統大選期間被大量使用[23]。假訊息訴諸使用者自身情感層面，無論是驚奇、恐懼、厭惡或是感動，都是使假訊息的流傳速度、廣度都遠大於真新聞的原因[24]。

而使用者對於資訊的偏見，亦是假訊息之所以有高度影響力

---

19 Ari Ezra Waldman, *The Marketplace of Fake News,* 20 U. Pa. J. Const. L. 845, 851 (2018).

20 Word of the Year 2016, Oxford Dictionaries, at https://languages.oup.com/word-of-the-year/word-of-the-year-2016 (last visited 02/17/2022).

21 *Id.*

22 AM730 報（10/17/2016），曾鈺成，同營護短，http://archive.am730.com.hk/column-333314（最後瀏覽日：02/17/2022）。

23 Word of the Year 2016, *supra* note 20.

24 胡元輝（2018），〈造假有效、更正無力？第三方事實查核機制初探〉，《傳播研究與實踐》，8 卷 2 期，頁 46。

的肇因之一。人類對於資訊的攝取，以及信賴程度並非完全客觀，而是會受到自己本身的喜好、信仰與偏見影響[25]。而「資訊的來源」對人們來說，也有一定程度的重要性，例如一個美國共和黨的支持者，面對關於希拉蕊涉及兒童性買賣的假訊息澄清資訊時，該澄清資訊是被刊登在對共和黨或民主黨友善的媒體，將會大幅度地影響澄清資訊對該民眾的效力[26]。這種「資訊偏食」的現象，在社群網站演算法推波助瀾之下，又成就了一個新興用語：「同溫層（英文稱作回音室『Echo chamber』）」。

社群網站如 Meta 公司旗下的 Facebook（下稱臉書），向來奉行一套守則：為了獲得最多的點擊率，必須讓用戶看到「他們喜歡的東西」。它們用這套邏輯來設計平台的運作、廣告的投放以及背後的演算法。亦即，在用戶頻繁地以按讚或分享等方式與特定觀點、類型或意識形態的資訊互動時，便會在社群平台上形成一個「同溫層」，裡面充斥更多類似觀點與類型的資訊[27]。同溫層最大的問題是造成訊息接收的單一化，如臉書的資料科學家們在 2015 年針對 1000 萬個美國臉書使用者的研究指出，雖然平均而言，美國自由派與保守派的臉書使用者，有大概 20%的臉書好友在另一個陣營，但有一個相當顯著的現象，即雙方都幾乎不會點擊另一陣營臉書好友所分享的訊息[28]。

這個現象對於探討如何有效地澄清假訊息，至關重要。因為

---

[25] Lili Levi, *supra* note 12, at 315.

[26] *Id.*

[27] Walter Quattrociocch 著，鍾樹人譯（2017），〈同溫層效應蔓延中〉，《科學人雜誌》，185 期，http://sa.ylib.com/MagArticle.aspx?Unit=featurearticles&id=3609（最後瀏覽日：02/17/2022）。

[28] 王宏恩，〈誰會相信假消息？該怎麼對抗假消息？行為科學的啟示〉，菜市場政治學，http://whogovernstw.org/2017/03/19/austinwang23/（最後瀏覽日：02/17/2022）。

同溫層與偏見等現象，容易讓假訊息提供者進行操作。首先，假訊息提供者在同溫層中凝聚使用者的信賴，使假訊息更為可信；此外，也可以針對同溫層外的主流媒體、甚至事實查核中心進行攻擊，讓政府、媒體或查核中心對於假訊息的批判，在同溫層群體中無效化[29]。這種把自己操作成被國家機器、對手政黨打壓的「弱勢方」的手法，在同溫層之中相當有效，並且使得查核澄清工作雪上加霜[30]。因此有論者認為，在貿然開始進行假訊息管制，或發布澄清訊息前，必須注意這些「逆火效應」，否則將會讓假訊息的防治工作產生反效果[31]。

另外，假訊息亦可能直接影響社會秩序，甚至人民的安全。如美國著名的「披薩門（Pizzagate）」事件，在選舉期間關於希拉蕊與民主黨員進行兒童人口販運，並將華盛頓特區一間批薩店做為據點的假訊息，導致一名男子持槍闖入該披薩店，聲稱要救出這些兒童，而造成店家與民眾的恐慌[32]。

三、國安層面

假訊息所帶來的另一個問題，尤其對許多民主國家而言，便是境外勢力對於本國的滲透。2017 年 1 月，美國情報體系（U.S. Intelligence Community）公開指稱俄羅斯官方與非官方組織，在 2016 美國總統大選期間，以各種方式影響了選舉，並打擊公眾對於民主程序的信心[33]。而台灣在 2018 年九合一大選時，種種

---

29 Alexandra Andorfer, s*upra* note 16, at 1417.
30 *Id.*
31 胡元輝，同前註 24，頁 46-50。
32 David O. Kleina & Joshua R. Wuellera, *supra* note 1, at 5.
33 Fake News, Free Speech, And Foreign Influence, Human Rights First, at https://www.humanrightsfirst.org/sites/default/files/Disinformation-Brief-March-2018.pdf (last visited 02/17/2022).

跡象指出，選舉的過程似乎受到境外勢力（中國），以媒體、假訊息操縱與影響[34]。這樣的擔憂並非陰謀論，而是真實正在發生的現象，如瑞典哥德堡大學所主持的 V-Dem 計劃，在 2019 年的一項調查指出，台灣在 2018 年全球國家「遭受外國假資訊攻擊」的程度中，名列第一[35]。至此，假訊息的危害從單純的傳遞錯誤訊息，上升到了國與國之間「資訊戰」的一環，對於民主制度與國家安全均構成威脅。而從 2019 年開始，我國各界均開始針對此一議題做出回應，希望能強化我國在資訊戰之下的自我防衛能力，例如政府推動的國安法修正[36]，以及「反紅媒」集會等[37]。

在科技的發展下，境外資訊站所使用來製造假訊息的技術也推陳出新，例如在去年台灣地方立委罷免投票的過程中，便有研究者發現一系列有高度可能由中國主導的「假新聞 Youtube 頻道」，不斷地製作針對特定候選人的不實指控，並以「機器人主播」的方式播送給台灣閱聽人[38]。這些頻道透過機器的文字轉語音技術，快速且大量製作這類報導並獲得觸及率，以達到影響選舉結果的目標[39]。

---

[34] The New York Times, s*upra* note 3.

[35] 菜市場政治學，〈台灣「接收境外假資訊」嚴重程度被專家評為世界第一＋V-Dem 資料庫簡介〉，https://whogovernstw.org/2019/04/12/whogovernstw9/（最後瀏覽日：02/17/2022）。

[36] 羅承宗，同前註 4。

[37] 中央社（06/23/2019），凱道反紅媒遊行 總統：提高對中國滲透媒體醒覺，https://www.cna.com.tw/news/firstnews/201906230106.aspx（最後瀏覽日：02/17/2022）。

[38] 王宏恩，〈中國 Youtube 假主播罷 Q 全面啟動〉，思想坦克，https://voicettank.org/%E4%B8%AD%E5%9C%8Byoutube%E5%81%87%E4%B8%BB%E6%92%AD%E7%BD%B7q%E5%85%A8%E9%9D%A2%E5%95%9F%E5%8B%95/（最後瀏覽日：06/21/2022）。

[39] 同前註。

四、小結

（一）不同法益保護與管制密度之關聯性

上述分析呈現出假訊息的影響並非單一，而是針對社會上各種層面的法益，均可能造成威脅。理解並歸納假訊息所侵害的各種法益，有助於立法者在考量言論自由等基本權保障的前提下，合理的管制假訊息。假訊息並非全新的議題，全世界的法體系早已有針對假訊息的處罰，例如誹謗、證券交易中的散布流言等規範，其本質就是在對抗假訊息的法益侵害，但因侵害程度有別（個人法益的名譽侵害、社會法益的市場秩序侵害），法規範亦給予程度不同的管制密度與處罰。又例如上述提及的新興的資訊戰行為，係針對國家安全（國家法益）的侵害，針對資訊戰行為的立法規範密度，以及言論自由等基本權在其中所需做出的退讓，程度必然與個人法益的誹謗罪有所不同。

綜觀我國近年來針對假訊息管制的新修法，可說是以不同法益的侵害作為立法管制的標準。《社會秩序維護法》、《災害防救法》、《傳染病防治條例》等法規的率先修正，給予散布假訊息者不同程度的處罰，顯示了立法者對於法益保護的重視程度有別，以及對假訊息管制與言論自由的權衡。本文之重點不在討論傳統以增訂規範與處罰治理假訊息的立法模式，而是著重於討論人工智慧治理假訊息的可能性，以及資訊時代下帶有網路治理、科技治理的規範展望。惟上述假訊息所侵害法益的多元性、差異性，以及我國立法者目前所採取的以法益保護觀點為核心的規範模式，對於此議題仍有其重要性。因為這樣的修法進程，一定程度代表了立法者的價值判斷以及態度，這對於本文所提出的兩個目標「Lawmakers 與 Codemakers 的對話」以及「制定 AI 倫理規範」均相當關鍵，本文第四章將有更進一步的討論。

（二）科技成為假訊息問題的催化劑

透過上述分析假訊息的成因，以及在社會上可能造成的各種衝擊與影響，除了協助吾人理解假訊息的嚴重性與危害外，也呈現出新興科技如演算法、機器學習，已經成為假訊息問題的導因與催化劑。因此如何對抗這些新興科技所帶來的社會問題，便是數位時代假訊息治理問題的重要一環。社群網站的演算法創造出的同溫層效應，使人們更難以接觸真相、搜尋平台的廣告與分潤機制，鼓勵一般人投入全球的假訊息製造業、境外勢力得以運用先進的技術如機器人主播、甚至深度偽造技術（Deepfake），在更短時間內製作出有更高可信度的假訊息，發動資訊戰。

假訊息問題存在已久，但卻是在網際網路、資訊科技以及社群網站等技術發展之下，才在近年獲得了爆炸性的成長。因此，面對資訊社會所造成的問題，有論者指出，應該以資訊科技，也就是人工智慧、機器人與演算法，加以回應[40]。近年來，以人工智慧與大數據協助人類進行資料分析、決策判斷等工作已非新聞，惟運用在假訊息的治理上，目前仍屬發展階段 。但不可否認的，機器人或人工智慧等技術在今日，已經被或多或少地運用在假訊息的治理上。因此，「假訊息監理科技」的探討，亦應成為假訊息多方管制框架的一環。以下本文將簡介目前假訊息監理科技的發展與可能性切入，進而在後續章節分析監理科技本身可能帶來的問題，以及其與法律的互動方式。

[40] 鏡傳媒（08/09/2018），終結假消息？AI 辨識方法再升級，https://www.mirrormedia.mg/story/20180809mit001/（最後瀏覽日：02/17/2022）。

參、AI 在處理假訊息問題中扮演之角色

假訊息有許多潛在的危害，而如何避免或降低這些危害，是目前各國都在努力的目標，亦有提出許多解方。例如美國提倡的網路與社群平台自治、德國以較強力的罰則作為督促平台業者的手段等[41]。而本文認為，這些討論都是假訊息多方利害關係治理框架的一環，惟如前所述，以「科技對抗科技」作為假訊息的治理，或許相對立法處罰，不失為一個更加直接且有效的手段。為聚焦重點，本文將討論範圍放在「AI 人工智慧、機器人」等資訊科技應對假訊息的方法與可能性，以及這些科技發展與法律原則、理論以及立法技術如何互動等議題。本章節便由實際案例出發，探討以 AI 作為處理假訊息的工具，與傳統人工的差異與優劣勢何在，以及目前在實務上被運用的態樣，與未來發展的方向。

一、人工為主的模式：以臉書（Meta）為例

Meta 執行長 Mark Zuckerberg 原先認為自家平台上的假訊息並沒有太大的問題。但在 2016 年美國總統大選後，尤其是美國參議院針對臉書、Google 與 Twitter 的假訊息問題作出聽證會後，他的態度有了轉變，認為臉書應該「更加重視此問題[42]」。在此之後，臉書發布了一連串的新政策，針對平台上的假訊息控

---

41 何吉森（2018），〈假消息之監理與治理探討〉，《傳播研究與實踐》，8 卷 2 期，頁 13-14。

42 Kurt Wagner, Mark Zuckerberg admits he should have taken Facebook fake news and the election more seriously: ‘Calling that crazy was dismissive and I regret it,’ vox, at https://www.vox.com/2017/9/27/16376502/mark-zuckerberg-facebook-donald-trump-fake-news (last visited 02/17/2022).

管作出改革。其中一項作法，便是在貼文的右上角新增了一個「標籤（Flag）」功能，用戶可以透過這個功能，來回報他們認為疑似有虛偽不實內容的新聞或文章[43]。當獲取足夠的回報時，臉書便會將爭議的文章，送交與其合作的第三方事實查核機構，由這些機構以他們內部的流程進行人工事實查核，並將結果反饋[44]。臉書在收到反饋後，便會將被確認為假訊息的爭議文章，標示為「Disputed」，使用者仍可以分享與轉貼，但在轉貼時會收到相關的提醒字樣，且該文章會自動被臉書的演算法調整到靠後的順位，減少使用者接觸到該資訊的機會[45]。台灣有著全球最高的臉書活躍用戶比例，因此臉書也想在未來把這套依靠第三方事實查核組織的假訊息治理模式運用到台灣[46]。

仰賴人工手動進行查核當然有其好處，尤其是在現今 AI 尚無法準確判別部分假訊息時，人工查核更能準確識別「諷刺」、「觀點」等容易與假訊息混淆之內容[47]。而與臉書合作的事實查核中心，均經過國際事實查核（International Fact-Checking Network，簡稱 IFCN）組織之認證，該組織從 2014 年以召開會議、簽署查核守則等方式不斷精進事實查核工作[48]。因此，也可以認為臉書此舉為尊重專業之作法，讓最理解假訊息的人與組織來進行假訊息的查核工作。但相反地，此種仰賴人工的作法亦存在下列幾項缺點：

---

43 Alexandra Andorfer, *supra* note 16, at 1415.
44 *Id.*
45 *Id*. at 1415-1416.
46 自由時報（06/19/2019），最重刪除粉專、帳號！FB 在台開啟「假消息」事實查核機制，https://3c.ltn.com.tw/news/37126（最後瀏覽日：02/17/2022）。
47 Alexandra Andorfer, *supra* note 16, at 1416.
48 胡元輝，同前註 24，頁 54。

1. 效率不彰：人工查核最大的缺點就是速度較慢，而且若有大量的社群舉報，往往會造成查核工作量無法負荷[49]。

2. 仰賴閱聽人舉報：臉書此項設計，在最前端仰賴用戶主動以標籤方式檢舉貼文，但社群平台的用戶對於假訊息的判斷能力為何？有研究指出，甚至連美國的大學生，對於網路上的信息來源與可信度，都不一定具備足夠的判斷能力[50]，廣大的社群用戶是否具備辨識假訊息的能力，實有疑問。

3. 逆火效應：承上，雖然前一章節所述之管制假訊息的逆火效應，在使用人工或機器人治理的情形均有可能發生。但臉書此處更加仰賴用戶自行檢舉，若用戶受到前述偏見或同溫層效應之影響，反而更有可能產生反效果，使正確的新聞暴露在「檢舉攻勢」風險中，增加查核機構的工作量[51]。

二、以 AI、機器人為主的模式

（一）以 AI 偵測假訊息

鑒於上述人工審查的幾項缺點，便有論者提出以機器人或 AI 來辨識假訊息，且已經有實踐的可能。例如西維吉尼亞大學的媒體與電腦科學計畫，進行了以 AI 偵測假訊息的研究。他們透過機械學習技術（Machine Learning），讓 AI 分析文章的內文，並給出一個代表真偽程度的分數，以此來幫助使用者判斷訊息的真假，並且公開評分的標準及過程，達到透明化的效果[52]。

---

[49] Alexandra Andorfer, *supra* note *16*, at 1418.

[50] Camila Domonoske, Students Have 'Dismaying' Inability to Tell Fake News from Real, Study Finds, NPR, at http://www.npr.org/sections/thetwo-way/2016/11/23/503129818/study-finds-students-have-dismaying-inability-to-tell-fake-news-from-real (last visited 02/17/2022).

[51] Alexandra Andorfer, *supra* note 16, at 1418.

[52] *Id.*

另外又如 Twitter 在 2019 年宣布併購的英國 AI 研究公司 Fabula AI，也是讓 AI 透過深度學習來識別假訊息，不過與前者不同的是，Fabula AI 將重點擺在識別假訊息傳播的「方式」而非「內容」，因此就算內容經過加密無法識別，仍得以透過 AI 揪出假訊息[53]。

另外一個例子，是由微軟與亞利桑那州立大學所合作進行的一項研究計畫，其目的在於設計出一套可以偵測假訊息的 AI 系統[54]。但因 AI 在識別新聞內容真偽的技術上仍有所限制，該研究團隊便以社群網站上使用者與各種資訊互動的數據集結，並以此來訓練他們所開發的演算法[55]。其結合了社群網站的大數據，以及社會學的理論，歸納出幾項識別假訊息的重要信號[56]。為了給予各項不同的信號在系統中不同的權重，該團隊用少量以人工標註的資料作為乾淨的資料，配合大量充滿雜訊的其他資料，來開發並訓練出一套計算使用者在社群網站上所產生的各種變數的權重的系統[57]。

該團隊發現，對於以使用行為判斷一個文章是否為假訊息，有三項最重要的變數：1.使用者的情緒（Sentiment-based），如在與某個文章互動的行為（如轉貼、留言）中，使用者以「我不是很相信…」、「我蠻同意的…」等留言，將分別代表負面及正

---

53 數位時代（06/04/2019），Twitter 併購 AI 新創打擊假消息，另類演算邏輯讓加密內容也能揪出，https://www.bnext.com.tw/article/53516/twitter-buys-fabula-ai-for-machine-learing-on-fake-news（最後瀏覽日：02/17/2022）。

54 Kai Shu et al., Leveraging Multi-Source Weak Social Supervision for Early Detection of Fake News (2020), arXiv, available at https://arxiv.org/abs/2004.01732 (last visited 02/17/2022).

55 *Id.* at 2-3.

56 *Id*. at 7-9.

57 *Id*.

面的情緒信號。當一篇文章有許多負面的信號，或是一篇文章的情緒信號相當兩極時，它就有較高的機率是假訊息[58]。2.使用者的偏見指數（Bias-based），以社群網站上的大數據來計算使用者是否具有高度偏見，對於判斷假訊息也相當重要，因為這些具有高度偏見的使用者所分享的訊息，有更高的機率不是真實的[59]。3.使用者的信用指數（Credibility-based），該團隊主要以分析使用者在網路上的可疑活動（例如大型的群聚通常代表機器人的行為或惡意活動），來給予其信用評等，此類用戶分享的文章也會較容易被系統評價為不可信任[60]。在該研究團隊初步的實驗結果中，此系統與臉書的自然語言演算法結合後，在 PolitiFact 的新聞資料庫中運行時，其識別假訊息的準確度高達 82%[61]。

如前所述，以目前的技術而言，讓 AI 透過深度學習，在一次次的資料分析與運算中不斷進化，其判斷假訊息的準確率甚至可以超越人工[62]。曾有研究指出，某些 AI 識別假訊息的準確率已經來到 80%，而人工識別僅有 66%[63]。但此種完全依靠 AI 與演算法的方式，仍有幾項缺點。首先，儘管有許多研究正在進行，但目前仍沒有真正能投入市場使用的假訊息機器人，Google AI 專家紀懷新曾經表示，其最大的難題，在於「定義」假訊息[64]。

---

58 *Id*.
59 *Id*. at 8-9.
60 *Id*.
61 *Id*. at 11.
62 Tom Simonite, Humans Can't Expect AI to Just Fight Fake News for Them, WIRED, at https://www.wired.com/story/fake-news-challenge-artificial-intelligence/ (last visited 02/17/2022).
63 *Id*.
64 中時電子報（07/04/2018），打擊假消息靠 AI？Google AI 專家紀懷新：現在還很 難 ， https://www.chinatimes.com/realtimenews/20180704002130-260410?chdtv（最後瀏覽日：02/17/2022）。

其次的難題，便是我們是否、或應否將管制言論市場的工作，交由這些矽谷大公司的 CEO 以及工程師[65]？例如 Google 開發的偵測網路仇恨性言論的 Perspective 計畫便遭受批評，因其將往往將非主流的意見偵測為「具冒犯性的」，這種以演算法的結果來「淨化」公共討論的方式是否得當，值得深思[66]。

（二）以 AI 支援事實查核工作

除了識別假訊息以外，AI 與演算法也被運用在事實查核工作中，以增進事實查核的效率。例如由德州大學與杜克大學共同研發的演算法 ClaimBuster，便是試圖結合自然語言、大數據運算等技術來協助事實查核組織，進行事實查核工作的系統[67]。該系統於事實查核的各階段，包含蒐集與篩選文章，比對查核資料庫與協助推播等，都有對應的演算法來協助查核的人力[68]。例如，其中的 ClaimSpotter，就是以自然語言所訓練的 AI，其主要功能是在每日成千上萬的網路文字庫中，篩選最值得被查核的文章[69]。此系統背後的原理為，「越接近事實陳述，而非意見表達的文字，越有查核的價值[70]。」其並非艱深的道理，該團隊認為，如果一段文字是屬於意見表達，就代表其背後並沒有絕對的事實可加以驗證，也就無法判斷其「真假」，將不具備查核的意義。依循此一原理，該研究團隊以美國從有紀錄以來的選舉辯論內容做為資料，以人工方式將每一個句子標註其更偏向事實陳述或是

65 Alexandra Andorfer, s*upra* note 16, at 1421.
66 *Id.* at 1420-1421.
67 Naeemul Hassan et al., ClaimBuster: The First-ever End-to-end Fact-checking System, 10 Proc. of the VLDB Endowment 1945, 1946-47 (2017), available at http://www.vldb.org/pvldb/vol10/p1945-li.pdf.
68 *Id.* at 1946.
69 *Id.*
70 *Id.*

意見表達，再用標註的結果訓練演算法，使其得以識別具有高度事實陳述的句子與報導，藉此找出當日最值得查核的新聞[71]。該系統也連結到各大事實查核組織的資料庫，並自動比對其偵測到的內容，避免傳送已經被查核過的重複報導，節省工作量並增加效率[72]。

（三）以 AI 對抗其他新興技術

在網路時代下，AI 其實是一把雙面刃[73]；因為在作為應對假訊息問題解方的同時，AI 亦可能被運用來產生假訊息。美國華盛頓大學的一項研究中，研究者為了應對在未來可能越發嚴重的 AI 假訊息問題，嘗試開發出了一套專門製造假訊息的 AI 系統：Grover [74]。透過大量新聞資料以及深度學習所訓練的 Grover，其產出的宣傳內容或假訊息，在實驗中被證實甚至比人類所寫的假訊息更具備說服力[75]。該團隊甚至架設了一個網站來展示該 AI 的能力[76]。只要使用者輸入文章的部分資訊如作者名、標題或內容，Grover 就可以自動生成一篇假訊息[77]。當然，該研究團隊的終極目標並非研發一款假訊息製造機，而是試圖處理與其相同運作原理並可能在未來被惡意使用的 AI，而研究團隊發現，Grover 就是識別這些 AI 產生假訊息最強大的利器[78]。

---

71 *Id*. at 1946-47.
72 *Id*. at 1947.
73 *See* A. K. Cybenko & G. Cybenko, AI and Fake News, 33 IEEE Intelligent Sys. 1, 1-5 (2018), available at https://ieeexplore.ieee.org/document/8567972(last visited 02/17/2022).
74 Rowan Zellers et al., Defending Against Neural Fake News (2019), arXiv, available at https://arxiv.org/abs/1905.12616 (last visited 02/17/2022).
75 *Id*. at 5.
76 GROVER -A State of the Art Defense against Neural Fake News, at https://grover.allenai.org/ (last visited 02/17/2022).
77 *Id*.
78 Rowan Zellers et al., s*upra* note 74, at 6-7.

Grover 所使用的生成對抗網路（Generative Adversarial Network，簡稱 GAN），是機器深度學習中的一項技術，其基本原理為製作兩組 AI，一個不斷製造假的產物（如文章、圖片），另一個不斷去識別這些產物是否為電腦製作。在不斷互相對抗之下，兩種 AI 的能力都能在短期內獲得大幅提升，生成型的 AI 會更能製造出真偽難辨的產物，而識別型 AI 則是可以更準確的判斷該產物是否為機器所製造。Grover 透過類似的訓練，已經賦予該系統識別與其類似 AI 所產生的文章內容（在運用最大的運算能力與資料庫時，準確率甚至能超過 90%），這將能在未來有同類型的惡意 AI 出現時，投入實戰並發揮識別的功用[79]。

上述研究彰顯著新技術的出現，意味著對於傳統以政策、立法為主的治理模式之挑戰。近年來受到國際社會關注的「深度偽造（Deep Fake）技術，即為另一個典型的例子。除了以法律規範技術發展的準則以外，引進或研發相對應的偵測技術，使管制者有同等的武器去對抗被濫用的新興科技與 AI，亦為重要的一環[80]。

除了上述問題以外，有論者指出，這些平台在設計管制假訊息的手段（無論是依靠人力或機器），常常忽略了假訊息背後的驅動力除了「金錢」外，更多的是「權力（Power）」[81]。除此之外，這些以營利為主的平台業者也會面臨一項潛在衝突，即平台業者以販賣廣告維生，在假訊息議題浮上檯面以前，他們本來就或多或少地有在進行內容篩選與過濾，主要是針對不當內容或是仇恨性言論，而驅使他們執行此類政策的理由，是為了避免使

---

79 *Id.*

80 余和謙（2019），〈人工智慧之治理－以深度偽造為例〉，《科技法律透析》，31 卷 8 期，頁 70-71。

81 Ari Ezra Waldman, *supra* note 19, at 858.

用者感到不適而離開該平台，造成用戶數降低[82]。但假訊息有所不同，假訊息因為聳動、有趣的特性，往往受到使用者的關注、分享與互動，伴隨高點擊率而來的便是龐大的廣告收益，這反而是平台業者最樂見的結果[83]。

三、我國以 AI 監管假訊息的技術

提起台灣的「謠言破解 AI」，許多人的第一反應便是「美玉姨」。在美玉姨推出後，新聞媒體以「台灣鑑別假訊息的 AI 機器人」介紹她，使其在當時引起了一股旋風[84]。但美玉姨是否真的是前述定義下的 AI？其實不完全正確。美玉姨是由旅居香港的台灣工程師徐曦所創造，是通訊軟體 LINE 上面的機器人，用戶只要將美玉姨的帳號加入任何一個聊天群組，她便會自動偵測群組內的聊天內容，一旦發現疑似假訊息的內容，便會自動做出回應與張貼澄清資訊。如此看似非常神奇的「假訊息判別程式」，其實仍與前述以 AI 作為識別假訊息工具的技術，相差甚遠[85]。美玉姨這個機器人本身，其實不具備判別假訊息的能力，而是與 Cofacts 或台灣事實查核中心資料庫連結，當使用者轉貼網站或內容包含在事實查核組織已經被澄清的假訊息，美玉姨便

---

[82] *Id.*

[83] *Id.*

[84] BBC 中文網（02/14/2019），台灣 AI 機器人「美玉姨」的使命：甄別假資訊，挑戰網路謠言，https://www.bbc.com/zhongwen/trad/chinese-news-47208116（最後瀏覽日：02/17/2022）。

[85] 今周刊（01/02/2019），80 後女生 一手催生 LINE 打假「美玉姨」，https://www.businesstoday.com.tw/article/category/154769/post/201901020016/80%E5%BE%8C%E5%A5%B3%E7%94%9F%20%20%E4%B8%80%E6%89%8B%E5%82%AC%E7%94%9FLINE%E6%89%93%E5%81%87%E3%80%8C%E7%BE%8E%E7%8E%89%E5%A7%A8%E3%80%8D（最後瀏覽日：02/17/2022）。

會自動貼出澄清資訊[86]。因此，可以將美玉姨理解為一個設計上相當具有巧思的聊天機器人，讓使用者可以把她直接加入群組，在第一時間偵測假訊息資訊，但其本身僅為一個中介的角色，讓使用者可以更方便地接觸事實查核中心的資訊。

美玉姨是致力改善使用者經驗的機器人，而它的「養分」來源則為事實查核組織「Cofacts 真的假的」（下稱 Cofacts）所建立的假訊息查核資料庫。Cofacts 為民間組織「g0v 零時政府」旗下的一個計劃，致力以公眾協作力量與資訊科技來解決假訊息的問題[87]。在查核者端，Cofacts 並沒有針對查核者的身分做限制與驗證，只需要簡單登入後便可以成為查核者的一份子。以一個沒有受到大量資助的民間查核中心來說，為了運用社群自發的協作力量，如此做法可以降低一般人加入查核者的門檻。但也因此，相較於全職的事實查核工作人員（如台灣事實查核中心），其查核者的專業度較低。針對這部分，Cofacts 則是以網站上的教學，以及查核者的經驗分享（撰寫文章、編輯者聚會等）來彌補專業性。查核的過程並沒有嚴謹的標準作業程序（SOP），查核者可以透過任何方式（自身專業、資料查找甚至直接 Google 等）進行查核與填寫回報，進而連結到 Cofacts 的 Line 聊天機器人。這樣的設計，偏向前述臉書目前所採取的以人工審核為主、演算法與機器人為輔的假訊息治理模式。

我國目前仍未有真正以 AI 治理假訊息的本土技術投入市場，惟亦有相關研究正在進行中。例如近期由國立臺灣師範大學大眾傳播研究所王維菁教授所領導的研究團隊，便於 2021 年中

[86] 同前註。
[87] Cofacts 真的假的，https://cofacts.tw/（最後瀏覽日：02/17/2022）。

發表《利用 AI 技術偵測假新聞之實證研究》一文，簡述該團隊基於國際上的相關研究，試圖在台灣開發以自然語言、機器學習為技術基礎的偵測假訊息 AI[88]。該研究以 Pérez-Rosas 等人在美國曾進行的研究為基礎，以語言學的角度來訓練 AI 識別假新聞[89]。該 AI 的核心概念在於普通的新聞與「刻意以模擬、類似新聞的形式，來傳遞虛假及錯誤的訊息」的假新聞，在語意、詞彙、句法的使用上會有所差異[90]。

該團隊以傳統四大報「聯合報」、「自由時報」、「蘋果日報」、「中國時報」的網站作為「正常新聞」的資料蒐集範圍；從「怒吼」、「密訊」、「琦琦看新聞」、「寰球軍事網」四個假訊息內容農場網站所蒐集到的資料，作為「假新聞」，以此訓練其所設計的 AI，來測試 AI 識別兩種不同「新聞」的能力[91]。最後該團隊得出的結論是，由此種方式訓練的 AI，在判別一般新聞以及內容農場假新聞，所給出的平均新聞可信度為 89.95% 以及 33.71%[92]。此項結果顯示以自然語言訓練的 AI 的確可以透過新聞的撰寫方法與語句使用，在一定程度上區別來自四大報的新聞以及內容農場的新聞。

由此可知，我國若想發展更為成熟的假訊息 AI 技術，有一項前提便是資料庫的建構[93]。如前述的外國實例中，建構演算法

---

88 王維菁、廖執善、蔣旭政、周昆璋（2021），〈利用 AI 技術偵測假新聞之實證研究〉，《中華傳播學刊》，39 期，頁 43-70。

89 *See* Verónica Pérez-Rosas et al., *Automatic detection of fake news,* 27 Int'l Conf. on Computational Linguistics Proc. 3391 (2018).

90 王維菁、廖執善、蔣旭政、周昆璋，同前註 88，頁 49、53。

91 同前註，頁 52。

92 同前註，頁 59。

93 同前註，頁 65。

與 AI 的前提便是有足夠多的資料可以訓練機器，讓 AI 以深度學習的方式，不斷增進其準確度。美國目前在各大媒體以及事實查核組織的努力下，已經漸漸形成幾個主要的假訊息資料庫，並將此類資料庫提供給研究團隊使用。我國已有上述的事實查核組織的出現並且逐漸成熟，雖仍欠缺規模，但透過不斷累積，將有助於台灣、尤其是中文假訊息資料庫的建置，以利於未來相關技術的研究與發展。

四、管制者與假訊息監理科技的互動

本章節所舉例的各項研究與技術，顯示在假訊息的治理的各個層面上，人工智慧已經漸漸嶄露其可能性。近年來為了打擊假訊息，我國政府各部門亦有與 LINE、Google 等科技公司合作，透過這類科技公司進行跨部門、公私協力的假訊息管制交流[94]。政府意識到科技公司以及技術創新對於假訊息治理的重要性，並且願意在此議題上為 Lawmaker 與 Codemaker 的對話跨出第一步，值得吾人肯定。惟目前所謂「公私協力」的態樣，似乎仍僅侷限在政府以科技公司所提供的平台與技術傳遞澄清資訊，並增加官方資訊對民眾的觸及度等基本層面[95]。但對於本文提出的概念，「帶有科技觀點的假訊息治理規範制定」，仍未有具體作為。

如前所述，人工智慧、演算法等科技往往是把雙面刃，能載舟亦能覆舟。假訊息治理涉及言論自由保障以及國家、經濟、社

---

94 洪貞玲、羅世宏、胡元輝，〈台灣如何對抗不實訊息-跨部門合作模式分析〉，《台灣如何對抗不實資訊－跨部門合作模式分析》報告發表會，優質新聞發展協會主辦，2021 年 5 月 6 日，頁 6-22。

95 同前註，頁 21-22。

會層面等法益的衡平，不可不慎。在這樣的特性下，管制者的角色就越發重要，且其態度必須更加積極。在了解科技作為假訊息監理科技的可能性後，本文認為，在 Codemaker 不斷研究科技「載舟」的可能性時，管制者（Lawmaker）必須將目光放遠，想像「覆舟」的可能性，並且先行做出應對。但 AI 的倫理規範，卻是我國目前在因應此類可能影響人民基本權的新興科技時，仍欠缺的一環。以下，本文將以歐盟 2021 年《人工智慧法律調和規則草案》為借鑒，探討假訊息監理科技的倫理問題，以及提出對我國在未來制定 AI 倫理規範時的建議。

肆、使用假訊息監理科技的隱憂與科技倫理問題

一、假訊息監理科技的潛在問題

（一）言論自由的侵害與限制

假訊息的管制手段造成人民基本權（尤其是憲法所保障的言論自由）的限縮，以及管制的界線探討，一直都是假訊息的管制的討論中一項充滿挑戰性的難題。例如美國自 2016 總統大選以來，假訊息問題便受到高度的重視。而在假訊息氾濫與管制的背後，美國憲法第一修正案所保障的言論自由也正在受到挑戰。依據過去美國法院在實務上的見解，於判斷「不實的言論」是否應該受到憲法言論自由保障時，其仍是傾向於給予一定程度之保障，而非將其歸納進傳統上被歸類為不受保障的言論（如猥褻 obesinity）或受較低度保障的言論種類（如商業性言論 commercial speech）[96]。此項見解至今還未被最高法院推翻，如

[96] 詳細可參 *United States v. Alvarez*, 567 U.S. 709 (2012) 與 *Texas v. Johnson*, 491

2020 年的 *WASHLITE v. Fox News* 案中，福斯新聞被告散布假訊息，包含與疫情相關的假新聞。地方法院採取美國最高法院向來傾向保護言論自由的見解，認為「錯誤的言論」不在憲法所禁止的言論範圍之內，換言之，錯誤的言論仍應受到保護（與猥褻、挑釁等種類的言論不同）。

而德國在 2018 年率先因應數位時代假訊息問題而推出《網路執行法》（NetzDG），更進一步地在網際網路言論自由保障的脈絡下，掀起了數位時代下政府（規範制定者）、社群平台（受規範客體，同時作為網路言論空間的管制者）以及使用者三方間法律關係的討論[97]。該法案在推出之際，便因高額的罰款與對大型平台業者課予的下架義務，以及對網際網路言論自由可能造成的限縮，而遭受合憲性的質疑[98]。在我國方面，近年亦有論者指出，我國最被頻繁使用的假訊息處罰條文如《社會秩序維護法》第 36 條，亦存在主、客觀構成要件不夠明確，保護法益被擴張解釋而導致實務上的濫用情形[99]。

而如本文前述，就假訊息的問題而言，科技或人工智慧的管制在未來或許將比立法管制更加直接、有效率，但同時這些管制假訊息的「監理科技」其本身對於社會以及人民的言論自由等基本權，是否也產生了一定程度的風險，而應該被管制？或至少被相關的科技倫理規範所限縮？例如美國前總統川普（Donald

U.S. 397, 414 (1989)等判決。

97 相關討論可參考蘇慧婕，〈正當平台程序作為網路中介者的免責要件：德國網路執行法的合憲性評析〉，《國立臺灣大學法學論叢》，49 卷 4 期，頁 1915-1977（2020）。

98 同前註，頁 1937-1945。

99 楊智傑，〈美國不實言論之言論自由保障〉，《國立中正大學法學集刊》，71 期，頁 172-180（2021）。

Trump）的支持者在 2021 年初，根據川普的一則影片聲明，佔領美國國會山莊的暴動事件，在事件發生後，臉書與推特均迅速決定封鎖川普的帳號，無疑是對這位向來被認為「推特成癮」的總統，展現了平台業者與科技的力量[100]。當平台業者、科技公司甚至有能力封鎖美國總統的言論時，我國應將此案例作為反思，思考言論自由作為民主國家的重要基本權應如何受到保障，不會因管制科技受到限縮甚至產生寒蟬效應。因此，如何在科技、人工智慧倫理規範中，納入基本權保障之觀點，在鼓勵監理科技創新競爭的同時，避免犧牲人民的言論自由，便是假訊息監理科技在未來必須面對的問題之一。

（二）科技濫用與不透明

除了上述的言論自由限縮的風險以外，此類監理科技的開發者、使用者本身該如何受到「監督」，亦可能成為問題。例如，政府或執法單位得否使用「假訊息偵測科技」來進行打擊不實訊息的執法？其界限何在？又該如何受到監督？退一步言，就算這類監理科技不是由公權力所掌控，而是由私部門開發與使用，在網際網路時代下作為守門人的大型科技公司，其可能造成的科技濫用與使用者權利侵害，或許不亞於公權力的執法，而如何對公權力或大型私部門使用這類監理科技的方式與範圍作出限縮，是將來我國在制定相關政策時必須考量的重點。

以美國為例，雖然美國憲法向來對於第一修正案言論自由保護的解釋，認為僅有政府行為（State action）限縮人民的言論自由時，才有憲法上的權利可以主張。但近年來開始有案件挑戰此

[100] Nitasha Tiku et al., Twitter bans Trump's account, citing risk of further violence, The Washington Post, at https://www.washingtonpost.com/technology/2021/01/08/twitter-trump-dorsey/ (last visited 02/17/2022).

一觀點，認為過大的社群平台作為網際網路的守門員，其實已經具備準政府組織（quasi-state actors）的地位，而應該受到憲法所限制。例如 2020 年的 *Prager University v. Google LLC* 一案中，美國非營利組織 Prager University 的 Youtube 帳戶的某些影片被 Youtube 認為是受限制的內容而下架，因此認為母公司 Google 此舉是違反了他們的言論自由。第九巡迴上訴法院則採取法院對於憲法第一修正案一貫的見解，認為只有「政府」會違反言論自由，而 Google 屬於私人企業，不在憲法第一修正案的規範之中。在同年的 *Freedom Watch v. Google* 一案中，法院亦重申相同立場，且駁斥原告認為大型平台業者已經成為準政府組織（quasi-state actors）的主張。

本文認為，對於濫權的防止，最好的方式是建立監督與問責的機制。而監督與問責，又建立在「科技的使用」與「科技本身」的透明度上。惟監理科技，與大部分的人工智慧與複雜的演算法類似，其本質上有另一個問題是演算法的不透明性，也就是機器深度學習技術創造出的黑盒子（Black box）效應，往往使創造者都無法清楚說明 AI 的運作原理與參數，此時建立在透明度上的 AI 問責機制如何執行，便成為一大問題[101]。

二、歐盟《人工智慧法律調和規則草案》

關於這類監理科技的開發與使用應該遵循何種些準則，方能避免上述濫用以及基本權侵害的問題，本文認為，歐盟於 2021 年推出的新法，應可供我國參考。歐盟執行委員會（EU Commission）於 2021 年中頒布了《人工智慧法律調和規則草

[101] 王維菁、廖執善、蔣旭政、周昆璋，同前註 88，頁 65。

案》（Proposal for a Regulation Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts）（下稱《歐盟人工智慧草案》）。該草案是奠基於歐盟《一般資料保護規範》（GDPR）以及於 2020 年初發布的《人工智慧白皮書》（White Paper on Artificial Intelligence），主要的目標有四點：

1. 確保在歐盟市場上使用的人工智慧系統是安全的，並符合歐盟現行法律中對於基本權利的保障以及歐盟價值觀；

2. 確保法律的明確與穩定，以促進人工智慧的投資和創新；

3. 在基本權保障與安全要求層面，加強適用於人工智慧技術的法規治理與執行。

4. 促進合法、安全和值得信賴的人工智慧單一市場的發展，防止市場碎片化[102]。

該草案以風險評估為核心，將不同種類與用途的 AI，依照其對社會、經濟、安全與基本權利的影響，區分為「不可接受的風險」、「高風險」、「低風險」三個種類，並給予不同程度的倫理規範與監管密度[103]。該草案一旦通過後，將會使歐盟在未來的人工智慧治理上，不再完全仰賴科技產業自治，而是受到一定程度的官方監督，以促進良性的科技發展進程。該草案涉及層面甚廣，本文無意一一詳述與評析。惟從台灣假訊息管制科技、以及管制者與科技產業互動的觀點來看，本文認為該草案有幾項特

---

102 Eur. Parl. Doc. (COM 52021PC0206) 3 (2021), available at https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX:52021PC0206 (last visited 02/17/2022).

103 Vera Lúcia Raposo, May I have some artificial intelligence with my human rights? About the recent European Commission's Proposal on a Regulation for Artificial Intelligence, KSLR EU Law Blog (May. 24, 2021), at https://blogs.kcl.ac.uk/kslreuropeanlawblog/?p=1569 (last visited 02/17/2022).

點，值得我國在未來制定帶有科技觀點的假訊息管制政策時，作為參考：（1）對基本權的重視與保障、（2）以風險評估為核心，將 AI 進行分類、（3）政府能否使用此類管制科技。以下就此三點詳加探討。

（一）基本權利之保障

《歐盟人工智慧草案》中，從立法理由到規範本身，均不斷重申基本權保障的重要性，且其對於基本權的理解，來自《歐盟基本權利憲章》（Charter of Fundamental Rights）[104]。而基本權利憲章的第 11 條，便是對於言論自由、表意自由的保障[105]。假訊息氾濫固然會對一個社會造成諸多不利影響，惟過度或不當的管制假訊息手段，亦會對言論自由造成限縮甚至侵害[106]。我國憲法第 11 條保障人民的言論自由與出版自由，且以傳統三權分立（或五權分立）的角度而言，立法者以規範管制假訊息時，會受到憲法的約束。

（二）AI 分類與監理

歐盟的人工智慧草案的其中一項特色，便是以風險評估作為核心，並且試圖將所有可能出現或運用在歐盟市場內的 AI 進行分類，並給予不同的規範密度。這樣的分類是否完備、合理，不在本文的討論範圍，但本文認為該草案提供了我國未來在制定科技倫理規範上一個可參考的重要觀點：「科技的創新伴隨著風險，但該風險隨著科技的種類與目的，則會有程度上的差異，而倫理規範制定者便是要在創新與風險控管中取得平衡。」

---

104 Eur. Parl. Doc. (COM 52021PC0206), *supra* note 102, at 11.

105 2012 O.J. (C 326, 26.10.2012), available at https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:12012P/TXT (last visited 02/17/2022).

106 張文貞，同前註 7，頁 1535。

該草案將 AI 區分為「不可接受的風險」（Unacceptable Risk AI）、「高風險（High-risk AI）」、「低風險（Low-risk AI）」三個種類[107]。其中「不可接受的風險」的種類規範於第 3 章第 5 條（Title III article 5），其涵蓋範圍包括（1）以操控人類行為為目的的 AI（第 1 項 a 與 b 款）、（2）協助公權力進行社會分數評比（Social Scoring）的 AI（第 1 項 c 款）、（3）協助公權力在公共場所進行即時的生物識別偵測的 AI（如司法單位使用人臉識別技術）（第 1 項 d 款）[108]。基於操縱型 AI 對於弱勢族群的侵害以及政府掌握的監控型 AI 對於社會大規模監視（Mass surveillance）造成集權的隱憂，該草案認為此類 AI 風險過高，應該被完全禁止進入歐盟市場[109]。

而「高風險」AI，根據該草案的第 6 條，以及附件 III（Annex III），含蓋諸多涉及產品安全，以及基本權利的 AI 系統[110]。該草案自第 9 條以下有相當大的篇幅，在課予這類 AI 許多嚴格的要求以及義務，包含透明性、人為監督、準確性、健全性與資料安全以及通報義務等，並且要求在其進入市場前進行可行性評估[111]。

「低風險」AI 的部分，則是除了仰賴研發者的自律以外，亦有部分技術的透明與揭露規範。其中值得一提的是，該草案在第 52 條 3 款的透明性要求部分特別提出，「深度偽造技術 AI」（Deep Fake），在使用上必須清楚揭露，避免造成錯誤訊息傳遞；惟若是由公權力使用的「偵測 Deep Fake 技術的 AI」，則

107 Vera Lúcia Raposo, *supra* note 103.
108 Eur. Parl. Doc. (COM 52021PC0206), s*upra* note 102 at 43-45.
109 *Id.*
110 *Id.* at 45-46, Annexes 4-5.
111 *Id.* at 46-68.

不在此限[112]。在深度偽造技術的管制已成為假訊息治理一環的今日，歐盟的此項規範恰好呼應了本文所提出的幾個重點：（1）以 AI 來對付 AI，將是未來假訊息管制的一種形式（2）假訊息治理技術亦需要倫理規範（3）假訊息治理技術有可能為政府所用。

在科技迅速發展的今日，《歐盟人工智慧草案》以 AI 的任務與目的，進行風險評估與分類的作法，對於我國在未來假訊監理科技發展及監理機制的過程中，具有一定的參考價值。以本文曾在第三章舉例的假訊息監理科技而言，雖然將其統稱為「假訊息監理科技」，但觀其 AI 在「偵測」、「篩選」、「協助事實查核」的任務與目的，均不相同。以基本權（言論自由）的保障觀點而言，其所蘊含的風險高低亦不相同，例如偵測、屏蔽訊息的 AI 對於言論自由所產生的風險便可能高於蒐集、篩選查核資訊的 AI。因此，如何以我國基本權保障的角度，評估此類科技發展或投入使用後產生的風險，將是規範制定者必須面對的考驗。

（三）政府能否使用假訊息監理科技

而回到本章節之初提及的問題，當假訊息監理科技越發成熟後，是否得為政府或公權力所用？尤其是當許多種類的假訊息在我國已經被刑罰化的今日，公權力得否運用此類技術偵查如《傳染病防治法》第 63 條的散布疫情假訊息的犯罪？針對此問題，《歐盟人工智慧草案》似乎也無法給出明確的解答。惟如前所述，該草案的核心在於基本權保障與風險評估，因此若認為政府使用特定的 AI 將會侵害人民的基本權，該 AI 便很有可能被歸

112 *Id*. at 69.

類在「高風險」或是「不可容許」的類別中[113]。例如，對於大規模監控人民的技術，如即時的人臉識別系統，該法案便嚴格禁止，僅在幾種有重大公益性的例外下允許執法單位使用，例如搜尋犯罪被害人（包含失蹤兒童）、緊急避難、防止與偵查特定重大犯罪等[114]。

雖然在歐盟有公民團體認為這樣的分類仍不足以保障基本人權，大規模監控技術應該被全面禁止[115]。該草案仍提供了許多關於政府監管以及人權保障的價值取捨，可以作為我國在未來訂定相關倫理規範的參考。本文在此提出的初步想法是必須探究假訊息監理科技的目的、影響，並且將其劃分種類。因為不是每一種技術都必然不得由政府掌握，例如協助政府推播官方澄清資訊的 AI。但監測、刪除或犯罪偵查的 AI 技術，由於對於人權所造成的風險較高，則應該受到倫理規範更高程度的監管，抑或是禁止之。

伍、我國假訊息監理科技規範的未來展望

行政院已經確定將在 2022 年成立「數位發展部」，其本質上是為了因應全球數位化浪潮、協助產業發展而進行行政機關的組織改組，將目前分散在各部會有關通訊、資訊、資通安全、網路及傳播五大業務整合並統一管理[116]。而其中由行政院資安處

---

113 *Id.* at 43-46.

114 *Id.* at 43-44.

115 EU: New proposal on artificial intelligence must protect human rights, Article 19, at https://www.article19.org/resources/eu-artificial-intelligence-and-human-rights/ (last visited 02/17/2022).

116 詳見《數位發展部組織法》，https://law.moj.gov.tw/News/NewsDetail.aspx?msgid=167188（最後瀏覽日：

（在數位發展部成立後將升格為資通安全署）所推動的「資安卓越中心」（Cyber Security Center of Excellence（CCoE），亦預計於同年上路[117]。而 CCoE 旗下三大實驗室之一的「網路數據分析實驗室」，其主要任務便是注重社交工程威脅，包含社群網站假訊息以及深度偽造技術的偵測、識別與反制科技研究[118]。由此可知，除了學術單位以及民間科技產業的研究，公家機關亦準備更加積極投入假訊息監理科技的研發。而本文所提倡的觀點，包含「規範制定者與科技研發者的對話」、「建立反映我國基本人權價值觀的倫理規範或監理機制」等，在此脈絡下就顯得更加重要。

科技部於 2019 年發布「人工智慧科研發展指引」，並提出 AI 發展的三大核心價值：（1）以人為本：AI 發展應促進人類福祉，尊重基本權利與人性尊嚴；（2）永續發展：AI 發展應注重人類、社會、環境的平衡與永續發展；（3）多元包容：應創造包容多元價值觀的 AI，並促進跨領域對話與科技普惠[119]。由此三大核心價值，又衍生出八項指引：共榮共利、公平性與非歧視性、自主權與控制權、安全性、個人隱私與數據治理、透明性與可追溯性、可解釋性、問責與溝通[120]。這些指引與近年世界各

02/17/2022）。

117 大紀元（09/24/2021），資安即國安？台數位發展組改案還躺立院，https://www.epochtimes.com/b5/21/9/24/n13257960.htm（最後瀏覽日：02/17/2022）。

118 同前註。

119 科技部（2019），人工智慧科研發展指引，https://www.most.gov.tw/most/attachments/53491881-eb0d-443f-9169-1f434f7d33c7（最後瀏覽日：02/17/2022）。

120 同前註。

國所提倡的 AI 倫理指引類似，在 AI 與科技研發的過程中有其重要性。惟與歐盟提出的人工智慧草案相比，我國科技部提出的此項指引，僅為概括性的上位指導原則。

相較於歐盟的人工智慧治理已經由自律轉向一定程度的法規監理體制，我國目前的發展指引並沒有公權力的介入作為後盾，且對於細部規範仍仰賴企業或研發者自律。本文認為，假訊息監理科技作為新興科技的一種，有對言論自由等基本權侵害之虞，應受到有系統性的監督與限制。我國未來若要制定假訊息監理科技的治理與倫理規範，不僅要參考歐盟之規範，將 AI 本身的種類與風險進行分類，亦應將我國目前針對假訊息立法採取的「保護法益」觀點納入，將監理科技所對抗的假訊息種類以及嚴重程度也進行分類。進而劃定倫理規範與治理的密度，以及透明度等要求。以下就本文之觀點，將假訊息監理科技，透過「風險分類」以及「保護法益」兩個觀點，說明我國未來界定監理科技的分類與適用範圍可以參考的方向。

一、風險分類

參考前述《歐盟人工智慧草案》的相關規定，假訊息監理科技涵蓋了各種人工智慧、演算法與新興科技等不同態樣。且如本文第三章所例示，各項假訊息監理科技的功能也大不相同，有的旨在協助事實查核，有的則被訓練在識別假訊息或是屏蔽假訊息上。未來在倫理規範的制訂上，我國應依照各個演算法的潛在功能以及規模，參考前述《歐盟人工智慧草案》的相關規定，以風險分析的架構來分類各種假訊息監理科技。

以本文第三章節例舉的各項技術為例，如各項偵測假訊息的 AI，其本身對於網路空間的言論自由必造成一定程度的限制，再配合社群網站的篩選式演算法的過濾與屏蔽，使其具有較高度的

風險，而應該受到較高程度的倫理規範監督。又如 Grover 此類訓練來對抗製作假訊息 AI 的 AI，其本身的演算法也是製作假訊息的利器，有被盜取或惡意濫用的風險。而如 ClaimBuster 這種以蒐集網路資料並支援事實查核機構的演算法，其對於言論自由限制，以及濫用的風險性應相對較小。

除了演算法與 AI 本身的風險有所不同外，「誰」握有該假訊息監理科技的開發、使用與主導權，也與其對社會可能蘊含的風險高低息息相關。如前述歐盟對於人臉辨識系統是否可為公權力所用的討論，我國未來在制定相關規範時也必須注重假訊息監理科技的使用者，畢竟一個偵測假訊息的 AI 是由民間第三方查核機構，或由政府與執法單位所掌控，其背後所蘊含的風險大不相同。另外，在大型社群網站成為網際網路的守門員後（或如前所述，有論者主張其已成為準政府組織體），其在特定議題與領域的影響力其實不亞於政府。而如何評價其使用這類假訊息監理科技對社會可能帶來的風險，亦是立法者以及科技人必須共同努力探究的問題。

二、保護法益與基本權衡平

如何在人工智慧或科技監理中納入基本權利或言論自由保障，其中一項可行的做法，便是在制定相關政策時，納入本文第二章所提及的，以保護法益區分管制言論種類的觀點。依照我國目前的修法進程，在不違反憲法的前提下，立法者作為民意的體現，在衡平言論自由與其他人民或國家的重要利益後，已經做出其對於哪些種類言論應該受到管制的價值判斷。而這樣的價值判斷，作為我國對於言論自由的理解而套用在 AI 或科技倫理規範的制定上，不失為一個重要的參考指標。換言之，以假訊息監理

科技而言，除了監理科技本身要做出風險分類，其所對應到的假訊息種類也必須有所區隔，作為判斷使用該監理科技的正當化基礎。本文認為，對於監理科技的風險分類會影響開發與使用該技術之人或組織體的義務，而對於假訊息種類的分類，則會影響假訊息監理科技的「使用範圍」。

本文認為，假訊息監理科技要依據其本身風險程度的不同，而受到科技治理或倫理規範的限縮，但「高風險監理科技」被運用在對抗「高法益侵害」的假訊息時，可以有更高的正當性基礎。如前述歐盟對人臉辨識系統的觀點，若我國立法者在基於民意基礎，已經將疫情假訊息、涉及國安的境外資訊戰、選舉不實訊息等假訊息態樣，歸類於高法益侵害而課予較高處罰刑度時，運用高風險的偵測或下架假訊息的 AI 或演算法來監理這類假訊息就有更高的正當性，甚至有讓政府針對特定領域的假訊息使用這類科技加以監理的可能。當然必須重申的是，對於高風險 AI 所應具備的義務要求，如透明度與監督的義務，並不因此有所改變。

其實許多大型社群網站目前已經採取此種針對假訊息分類而使用不同程度的技術對應的政策。例如臉書在其《社群守則》中，便特別揭露平台會強制下架的不實訊息種類，包含「對他人會造成立即身體危害的訊息」、「與健康有關的不實訊息」、「干預選民投票或人口普查」以及「變造的影片與 Deepfake 影像」等[121]。而其他種類的不實訊息，則是回歸以「減少推播」或「標記」的方式處理，並不會直接屏蔽[122]。本文對於這樣的方

[121] Facebook 《社群守則》，Meta，https://transparency.fb.com/zh-tw/policies/community-standards/misinformation/（最後瀏覽日：06/24/2022）。

[122] 同前註。

針表示理解與贊同，惟對於這些「特別有害的」假訊息所作出的政策與分類，是屬於美國公司 Meta 的價值判斷，其是否有反映出我國的民情以及民意，或是我國規範制定者與國民的價值判斷是否能反過來影響跨國企業 Meta 的社群守則，本文抱持高度懷疑。

因此，本文認為，由政府介入，以倫理或科技治理規範創設具有強制力的政策，使我國針對假訊息治理政策與立法的價值判斷，能適度地影響科技公司與平台業者的標準，使其制定符合當地情況的社群守則，是我國在未來假訊息治理政策的重要一環。這種適度影響，除了直接訂定法規禁止某些訊息的傳播，亦可透過強制這類社群平台成立由本土專家以及各方利害關係人組成的委員會，定期審視與修正專屬於本土的社群守則，並檢視平台的執法標準等方式達成。

三、小結

綜上所述，本章節針對我國未來假訊息監理科技的相關治理政策所提出的各項觀點，可以彙整如下：

1. 假訊息監理科技開發者與使用者的義務

假訊息監理科技涉及言論自由等基本權，其開發者與使用者應該受到具有一定強制力的科技治理政策，或至少是倫理政策所拘束，不應該完全仰賴私部門或科技公司的自律。相關的義務可以參考歐盟的人工智慧草案，包含透明度、人工監督、準確性、可行性、健全性、資料安全、通報義務以及建立良好的救濟制度等。但並非所有的假訊息監理科技都必須受到相同強度的義務所限制，因此，監理科技的分類就相當重要。

2. 假訊息監理科技的風險分類

呈上所述，我國未來在透過科技治理與倫理政策課予各項義務時，並非應該將所有的假訊息監理科技一視同仁，而應該依照風險管理模式依照其技術的性質、使用予掌握該技術的人或組織，作出不同層級的分類。越高風險的技術，其使用的範圍應該越受到限縮，而受規範的密度也應該更高。例如低風險的事實查核輔助 AI，或許只需要透過科技開發者的自律。但高風險的偵測或屏蔽型技術，應該受到高度透明性以及監督的要求，例如要求其組成外部委員會、定期產出執行與透明性報告、甚至要求其建立健全的救濟機制等。

3. 假訊息監理科技的適用範圍

除了上述的各項義務要求外，越高風險的假訊息監理科技，其所能使用的範圍應該越加受到限縮。而限縮的標準應可依照我國目前的法益保護立法脈絡，區分不同假訊息所造成的侵害程度來加以判斷。在立法者認為具有高度侵害的法領域，允許較高程度風險的假訊息監理科技介入並控管，作為其正當化基礎。而大型平台業者也應該依據我國立法者或利害關係人的價值判斷，定期修正其針對我國使用者與網路環境的社群守則，以反映不同社會以及不同時空脈絡間之差異。

本文理解科技創新與監理，向來是不斷拉扯的過程，尤其在我國目前 AI 科技尚在起步階段，過度嚴格的監理將不利於創新發展。但假訊息管制涉及民主國家言論自由基本權保障問題，前述對於歐盟的新草案以及假訊息監理科技的分析，對於我國在制定假訊息治理政策或法規時，仍有參考的價值。科技的發展日新月異，因此針對監理科技的相關規範也必須是浮動且必須不斷進行調整的。這都需要仰賴科技人、法律人以及所有利害關係人的不斷溝通與對話，方能在保障人民基本權的前提下，最有效率地以科技對抗這些同樣在不斷推陳出新的假訊息。

陸、結論

資訊科技與資訊社會助長了假訊息問題，但資訊科技也提供了我們治理假訊息的手段。目前最主要以 AI 作為治理假訊息工具的態樣，是以人工進行事實查核，再配合 AI、機器人與演算法，使檢舉的過程更加簡便、查核的結果能更有效地傳遞。而將 AI 真正運用在識別假訊息的工作，目前亦有許多單位正進行相關研究，或許在不久的將來，便能投入市場開始運作。但筆者認為，AI 畢竟只是工具，無論其發展多完善、準確率多高，在治理假訊息此議題上，有許多最根本的問題仍不能忽略，在專注於工具的效能與精確度上時，仍需要考量前述的假訊息成因與影響。例如，如何使澄清資訊有效突破同溫層效應？如何有效地突破黑盒子效應而建立問責機制？在以 AI 進行「假訊息的抑制」抑或是「促使真實訊息的傳播」時，如何使假訊息監理技術符合基本人權的保障？此均為在技術發展以及管制方式建構的過程中，必須不斷被重新提出與檢視的關鍵問題。

而以管制者的觀點，在新技術出現以後，應重新審視目前修法的方向，使法規範的建構與科技能更有效的連結；在修法過程，或是訂定科技發展監理規範、倫理規範時，納入更多具有前瞻性的觀點。例如政府、科技公司與事實查核機構，在假訊息治理框架中所扮演的角色、言論自由保護與社群網路治理的衡平與取捨等。另外也需要在未來規範制定過程中，以「法律如何與科技互動」為切入點，考量更多面向的議題。例如政府作為管制者，在技術發展的過程中「如何」或「應否」進行公私協力、提供民間單位資源，其角色可能相當敏感。例如，政府提供的資源可以加速技術的發展，但如何在政府介入的情形下維持技術發展

的中立、非黨派？在協助研發後，政府得否、如何運用相關管制假訊息的技術，而不違反言論自由的保障？當管制者的力量被新興科技加以提升後，該如何防止其濫權，這些都是不可避免的問題。因此，本文以假訊息管制科技的態樣與可能性為核心，導出「公私協力、對話」的必要性，以及「科技治理或倫理規範」對於假訊息管制與人權保障的重要性，期待我國未來在規劃政策或修法時，能有更全面、前瞻的觀點。

參考文獻

一、中文部分

Walter Quattrociocch 著，鍾樹人譯（2017）。〈同溫層效應蔓延中〉，《科學人雜誌》，185 期，載於 http://sa.ylib.com/MagArticle.aspx?Unit=featurearticles&id=3609（最後瀏覽日：02/17/2022）。

王宏恩（2017）。〈誰會相信假消息？該怎麼對抗假消息？行為科學的啟示〉，載於菜市場政治學 http://whogovernstw.org/2017/03/19/austinwang23/（最後瀏覽日：02/17/2022）。

王宏恩（2021）。〈中國 Youtube 假主播罷 Q 全面啟動〉，載於思想坦克 https://voicettank.org/%E4%B8%AD%E5%9C%8Byoutube%E5%81%87%E4%B8%BB%E6%92%AD%E7%BD%B7q%E5%85%A8%E9%9D%A2%E5%95%9F%E5%8B%95/（最後瀏覽日：06/21/2022）。

王服清（2020）。〈假消息：謠言或不實訊息的規範競合關係－以衛生紙之亂為例〉，《台灣法學雜誌》，390 期，頁 71-94。

王維菁、廖執善、蔣旭政、周昆璋（2021）。〈利用 AI 技術偵測假新聞之實證研究〉，《中華傳播學刊》，39 期，頁 43-70。

何吉森（2018）。〈假消息之監理與治理探討〉，《傳播研究與實踐》，8 卷 2 期，頁 1-41。

余和謙（2019）。〈人工智慧之治理－以深度偽造為例〉，《科技法律透析》，31 卷 8 期，頁 52-72。

林志潔（2021）。〈數位時代下的不實資訊－刑事法的觀點〉，《科技部國外短期研究報告書》（未公開發表），頁 3-21。

施達妮著，顏妤恬譯（2018）。〈數位時代的假消息〉，《漢學研究通訊》，37 卷 3 期，頁 7-13。

洪貞玲、羅世宏、胡元輝（2021）。〈台灣如何對抗不實訊息－跨部門合作模式分析〉，《台灣如何對抗不實資訊－跨部門合作模式分析》報告發表會，優質新聞發展協會主辦，2021 年 5 月 6 日，頁 6-22。

科技部（2019），人工智慧科研發展指引，載於 https://www.most.gov.tw/most/attachments/53491881-eb0d-443f-9169-1f434f7d33c7（最後瀏覽日：02/17/2022）。

胡元輝（2018）。〈造假有效、更正無力？第三方事實查核機制初探〉，《傳播研究與實踐》，8 卷 2 期，頁 43-73。

張文貞（2019）。〈2018 年憲法發展回顧〉，《臺大法學論叢》，48 卷特刊，頁 1503-1545。

菜市場政治學（2019）。〈台灣「接收境外假資訊」嚴重程度被專家評為世界第一＋V-Dem 資料庫簡介〉，載於菜市場政治學 https://whogovernstw.org/2019/04/12/whogovernstw9/（最後瀏覽日：02/17/2022）。

楊智傑（2021）。〈美國不實言論之言論自由保障〉，《國立中正大學法學集刊》，71 期，頁 121-192。

羅承宗（2019）。〈虛假訊息與法律管制－我國現況與建議〉，《台灣法學雜誌》，369 期，頁 47-62。

蘇慧婕（2020）。〈正當平台程序作為網路中介者的免責要件：德國網路執行法的合憲性評析〉，《國立臺灣大學法學論叢》，49 卷 4 期，頁 1915-1977。

三、英文部份

Andorfer, Alexandra (2018). *Spreading Like Wildfire: Solutions for Abating the Fake News Problem on Social Media via Technology Controls and Government Regulation.* Hastings Law Journal, 69, 1409-1431.

Charter of Fundamental Rights of the European Union (C 326, 26.10.2012), available at https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:12012P/TXT (last visited 02/17/2022).

Cybenko, A. K. & Cybenko, G., AI and Fake News. IEEE Intelligent System, 33, 1-5 (2018), available at https://ieeexplore.ieee.org/document/8567972 (last visited 02/17/2022).

Domonoske, Camila, Students Have 'Dismaying' Inability to Tell Fake News from Real, Study Finds, NPR, at http://www.npr.org/sections/thetwo-way/2016/11/23/503129818/study-finds-students-have-dismaying-inability-to-tell-fake-news-from-real (last visited 02/17/2022).

EU: New proposal on artificial intelligence must protect human rights, Article 19, at https://www.article19.org/resources/eu-artificial-intelligence-and-human-rights/ (last visited 02/17/2022).

Fake News, Free Speech, And Foreign Influence, Human Rights First, at https://www.humanrightsfirst.org/sites/default/files/Disinformation-Brief-March-2018.pdf (last visited 02/17/2022).

GROVER -A State of the Art Defense against Neural Fake News, at https://grover.allenai.org/ (last visited 02/17/2022).

Hassan, Naeemul et al., ClaimBuster: The First-ever End-to-end Fact-checking System. Proceedings of the VLDB Endowment, 10, 1945-1948 (2017), available at http://www.vldb.org/pvldb/vol10/p1945-li.pdf (last visited 02/17/2022).

Higgins, Andrew et al., Inside a Fake News Sausage Factory: 'This Is All About Income', The New York Times, at https://www.nytimes.com/2016/11/25/world/europe/fake-news-donald-trump-hillary-clinton-georgia.html (last visited 02/17/2022).

Horton, Chris, Specter of Meddling by Beijing Looms Over Taiwan's Elections, The New York Times, at https://www.nytimes.com/2018/11/22/world/asia/taiwan-elections-meddling.html (last visited 02/17/2022).

Kleina, David O. & Wuellera, Joshua R. (2017). *Fake News: A Legal Perspective.* Journal of Internet Law, 20, 1-10.

Levi, Lili (2018). *Real "Fake News" And Fake "Fake News"*, First Amendment Law Review, 16, 232-327.

Marda, Vidushi & Milan, Stefania (2018). *Wisdom of the Crowd: Multistakeholder Perspectives on the Fake News Debate*. Internet Policy Review series, Annenberg School of Communication, 1-24.

Pérez-Rosas, Verónica et al. (2018). *Automatic detection of fake news.*

Proceedings of the 27th International Conference on Computational Linguistics, 27, 3391-3401.

Proposal for a REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL LAYING DOWN HARMONISED RULES ON ARTIFICIAL INTELLIGENCE (ARTIFICIAL INTELLIGENCE ACT) AND AMENDING CERTAIN UNION LEGISLATIVE ACTS (COM 52021PC0206) 3 (2021), available at https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX:52021PC0206 (last visited 02/17/2022).

Raposo, Vera Lúcia, May I have some artificial intelligence with my human rights? About the recent European Commission's Proposal on a Regulation for Artificial Intelligence, KSLR EU Law Blog (May. 24, 2021), at https://blogs.kcl.ac.uk/kslreuropeanlawblog/?p=1569 (last visited 02/17/2022).

Shu, Kai et al., Leveraging Multi-Source Weak Social Supervision for Early Detection of Fake News, arXiv, 1-17 (2020), available at https://arxiv.org/abs/2004.01732 (last visited 02/17/2022).

Simonite, Tom, Humans Can't Expect AI to Just Fight Fake News for Them, WIRED, at https://www.wired.com/story/fake-news-challenge-artificial-intelligence/ (last visited 02/17/2022).

Tiku, Nitasha et al., Twitter bans Trump's account, citing risk of further violence, The Washington Post, at https://www.washingtonpost.com/technology/2021/01/08/twitter-trump-dorsey/ (last visited 02/17/2022).

Tompros, Louis et al. (2017). *The Constitutionality of Criminalizing False Speech Made on Social Networking Sites in A Post-Alvarez,*

Social Media-Obsessed World. Harvard Journal of Law & Technology, 31, 65-109.

Wagner, Kurt, Mark Zuckerberg admits he should have taken Facebook fake news and the election more seriously: 'Calling that crazy was dismissive and I regret it', vox, at https://www.vox.com/2017/9/27/16376502/mark-zuckerberg-facebook-donald-trump-fake-news (last visited 02/17/2022).

Waldman, Ari Ezra (2018). *The Marketplace of Fake News.* The University of Pennsylvania Journal of Constitutional Law, 20, 845-870.

Word of the Year 2016, Oxford Dictionaries, at https://languages.oup.com/word-of-the-year/word-of-the-year-2016 (last visited 02/17/2022).

Zellers, Rowan et al. (2019). *Defending Against Neural Fake News*. arXiv, 1-21, available at https://arxiv.org/abs/1905.12616 (last visited 02/17/2022).