

罪，相反地，他還是一件關於澳洲儲備銀行（Reserve Bank of Australia）國際賄賂醜聞的吹哨者²。胡德市長立即委請律師寄發律師信，要求OpenAI更正錯誤，否則提出誹謗訴訟。

在特利教授或者海德市長的案例，AI是否該負誹謗責任？

壹、美麗的法律新世界

人類社會試圖發展人工智慧（Artificial Intelligence, AI）已有相當長的歷史，早在二戰後隨著電子計算機的發展即已逐步展開，並且在過去這十年得到驚人的進展³。但真正讓一般社會大眾真正有感到的，或許還是始於2022年年底發布的ChatGPT，以及隨後發表的Bard、文心一言這一類的聊天機器人。聊天機器人的問世，猶如平地一聲雷，社會大眾突然驚覺到：原來AI的iPhone時刻已經到來⁴。

隨著AI更大幅度地躍入社會大眾的生活，AI相關的法律議題再度成為法律界最熱門的話題之一，像是AI產出與智慧財產權的關係、AI對個人資料安全與隱私的影響、AI的產品責任、AI的不實言論、深度造假與假新聞、AI演算法歧視⁵與反歧視政策……等。與以往的討論不同，我們現在探討如何處理這些法律議題時，有一種強烈的急迫感與既視感，「未來式」已經變成了「現在進

行式」。

在眾多AI法律議題中，本文以不實言論為核心，試圖探討以聊天機器人為主的AI（以下均以AI稱之），在回覆不實言論而可能傷害他人名譽的情況下，是否構成誹謗或其他的法律責任，並探索未來可能的法律因應之道。

貳、現行關於誹謗的法律規定

以美國普通法⁶的原則來說，原告如果想在誹謗的民事求償勝訴，原告必須證明四個要件：一、被告出版或散布一個關於原告的陳述予第三人；二、該陳述是錯誤或虛假的；三、被告至少要達到過失（negligent）的程度；以及四、確實造成原告名譽上的傷害。

如果被陳述的對象是政府官員或公眾人物的話，光證明被告是過失還不夠，原告還必須證明被告有「真實惡意」（actual malice），證明被告在陳述誹謗的言論時，知道該言論是虛假或是對該陳述是否虛假毫不在意或極其漠視⁷。此時原告的舉證義務也被提高：在一般民事訴訟中，原告需要達到的舉證程度是證據優勢（Preponderance of evidence），但要成立對政府官員或公眾人物的誹謗，原告的舉證必須達到清晰可信的程度（clear and convincing proof）。

考慮AI的不實言論是否構成誹謗責

任，第一個需要思考的是：有沒有真實惡意或過失存在？

AI本身能不能成立真實惡意或過失，取決於AI是否像人類一樣有主觀的決策過程，或是像人類一樣有「注意」的能力。但AI只是一種指令系統，經由閱讀大量的資料來學習，然後以人類語言模式來回答問題。從這個角度來看，AI應不存在像人類一樣的主觀決策過程，也很難被認為對特定事物有注意的能力。而且，假設AI有像人類一樣「注意」的能力，AI的合理注意行為標準是什麼？是與人類相比，還是與其他的AI相比？如果是與其他的AI相比，其他AI的注意程度是什麼，目前有沒有一個合理的標準？這些問題現階段都沒有答案，從而就算我們認定AI有像人類一樣「注意」的能力，恐怕也無法輕易推導出AI有過失的結論。

那麼開發AI的企業或程式設計師，有沒有真實惡意或過失？AI依照所學習的資料按照演算法來編輯回覆，除非原告能證明在AI回答過程中，開發公司或程式設計師曾特別指示AI如何回覆或使用特定資料進行回覆，或者開發公司或程式設計師有意識或知悉AI正在對特定事項進行不實陳述，否則恐怕無法輕易認定開發公司或程式設計師有真實惡意或過失⁸。

從而，當原告要主張AI的不實言論構成誹謗時，現行法律下要證明誹謗主

觀要素存在，其實有一定的難度。

在客觀要件上，要成立「誹謗」責任，原告需要證明自身因誹謗言論遭受損害。不過，當AI回答問題時，通常只有很少的觀眾，經常就是只有使用者本身，而且不同的使用者向AI提出同一個問題時，它可能會給出不同的答案，甚至是同一個使用者在不同時間提相同的問題也會得到不同的答案。而且，觀眾真的會因為AI答覆而對原告產生負面評價嗎？還是會一笑置之地說AI又在胡言亂語了？以ChatGPT來說，其使用者介面清楚地標示警語：「ChatGPT可能會提供關於人物、地點或事實的不正確消息」（ChatGPT may produce inaccurate information about people, places, or facts.）；如果問ChatGPT某件事是否為真，它也經常強調自己不具備查證事實的能力。被告可能因此會抗辯，使用者清楚知道AI言論不可盡信⁹，原告並未真正遭受名譽的損害。基於以上這些理由，即使AI確已做出不實言論，原告要證明損害的範圍跟程度不是沒有困難，使得原告要在誹謗訴訟獲勝的難度增加。

參、互動式電腦服務提供者？

除了不具備「誹謗」的主客觀要件外，AI的開發企業還可援引其他的法律主張。

Angle

演算法是否會導致誹謗責任，其實不是全新的法律議題，Google過去即因為提供搜索引擎服務而面臨數次有關「誹謗」的訴訟。面對這類型的訴訟，Google在美國法下有一個很強的理據：美國傳播端正法（Communications Decency Act (2016)）第230條。該法條規定，「任何互動式電腦服務的提供者或使用者（provider or user of an interactive computer service）不因為另一位資訊內容提供者（information content provider）提供的消息而被視為是該消息的散布者或言論者」。「互動式電腦服務提供者」與「資訊內容提供者」不同，後者的工作是負責創造或開發資訊，並藉由網路或其他互動式電腦服務將所創造或開發的資訊提供出去；相較之下但互動式電腦服務提供者只是資訊流通的管道，不能僅因為其散布了別人發布的言論就構成誹謗。據此，AI的開發企業可能會主張其只是類似美國傳播端正法第230條下的互動式電腦服務提供者，即使AI答覆的內容涉及了不實言論，充其量也只是資訊的導管，不該因此就構成誹謗。

不過，援引美國傳播端正法第230條的主張不是毫無弱點。這好比，當Google的搜索引擎具備自動完成（auto-complete）或自動建議（auto suggestion）的演算法功能時，恐怕就難以主張絕對不負任何言論責任。例如2012年一名澳洲醫生即在加州對Google

提出誹謗訴訟，理由是在Google搜索引擎輸入他的名字時，自動更正的演算法會建議「破產」這個關鍵字，即使這名澳洲醫生沒有破產，Google仍拒絕修正其自動建議的功能。雖然這個案子最終以澳洲醫師撤告做終¹⁰，但許多類似的訴訟在歐洲做出不利Google的認定，像是法國法院曾判決Google成立誹謗，因為當用戶在搜索引擎輸入某家保險公司名稱後，它會自動建議在該公司名稱後增加「騙子」兩字，雖然Google抗辯自動生成建議是基於演算法而非人為決定，但該論述不被法國法院採信¹¹。極其類似地，德國法院也曾做出裁決，如果Google搜索引擎提供的自動完成建議具誹謗性質，那麼Google必須在收到請求後予以清除¹²。

如果今天AI答覆的不實言論是直接引用網路資訊，那麼AI的開發企業要主張互動式電腦服務提供者，有一定的道理；但如果不實言論是AI自動生成或合成的內容，像是本文開頭那篇從不存在的華盛頓郵報新聞，那麼主張其只是互動式電腦服務提供者而不負任何言論責任，恐怕也不是沒有挑戰的空間。當然，對AI的開發企業來說，AI什麼時候直接引用網路資訊，什麼時候則會自動生成或合成新資訊，這恐怕根本不是其能預見或控制的，要其為AI的所有言行負責實在是生命中不可承受之重。但放任AI合成不實言論，並以「互動式電腦服務

提供者」為由允許相關人士毋需承擔任何責任，恐怕也不是社會大眾能夠接受的法律政策。

但不管如何，在現行的法律規範下，不實言論的受害人要以誹謗訴訟獲得救濟，確實有許多懸而未決的法律問題尚待釐清。

肆、使用AI言論者仍應負責

不過，當AI生成不實言論後，如果使用者引用這些不實言論，那麼本文作者認為使用者本身仍有誹謗責任之適用。畢竟，當使用者引用AI不實言論時，其本身就在做成一種言論，使用者不論是機構或個人都應該有注意的能力，AI提供的資料不應成為免責的理由（至少在AI還未建立起公信力時還不是），尤其在AI開發公司已經公開強調AI提供的資訊不一定正確的情況下，引用轉述AI言論的機構或個人應更有警覺。

對於引用AI言論是否構成誹謗或可主張免責這一點，雖然目前還沒有相關訴訟，但我們可以從其他涉及自動系統的訴訟中得到一些指引。今年（2023）在一場關於特斯拉自動輔助駕駛系統的美國訴訟中，陪審團認定車禍事故歸因於駕駛分心而非自動駕駛功能。該名駕駛在2019年使用自動輔助駕駛系統時發生車禍，但特斯拉的使用手冊已經提醒車主不要在市區啟用自動輔助駕駛，且

車禍發生時駕駛的手並沒有放在方向盤上¹³，因此陪審團判定特斯拉免負賠償責任。換句話說，陪審團認定車主仍要掌控好駕駛工作，而非將全部的駕駛工作都交給自動輔助駕駛¹⁴。這場特斯拉自動駕駛輔助系統的訴訟，或可提供我們某程度的參照：使用AI所提供的資訊來製造或生成言論，使用者仍應就該言論負責任，而非將所有責任交給AI。

伍、產品責任？

如果誹謗的求償不成立，那麼不實言論受害人是否可以主張AI有瑕疵，從產品責任的角度來請求賠償？

如果主張「產品責任」，原告或許可因為「無過失責任」的制度免去證明故意或過失的負擔，因而增加勝訴的可能性。不過，產品責任補償的傷害類型通常是生命、身體、健康或財產損害，像是名譽受損這一類人格權的損害，並不直接屬於產品責任傳統上想要保護的範圍。

因應AI的應用，已經有政府機構意識到修改現行產品責任制度的必要性，像是歐洲議會在2022年9月已經公布修改歐盟會員國民事訴訟規則的提案¹⁵，新提案從降低資訊不對稱及建立因果關係兩方面著手，試圖降低AI產品的產品責任訴訟求償者的舉證負擔。在通過前述提案的同一時間，歐洲議會也通過另一

修改產品責任制度的提案，將AI系統以及AI驅動的產品明確納入無過失責任的產品責任制度涵蓋範圍¹⁶。然而，即使有這些提案，不實言論的受害者除非生命、身體、健康或財產同時受到損害，否則恐怕目前仍無法藉由「產品責任」制度直接取得賠償。

陸、我們需要新制度

從以上的分析可知，如果AI做出不實的言論而導致有人名譽受損，現行的法律制度似乎無法完全妥善給予救濟。

要降低AI不實言論對個人人格的傷害，現階段我們需要大聲疾呼與提醒社會大眾：AI的言論不完全可信，必須要謹慎使用AI所提供的資訊。不過，發出警示這種做法無法給予個案正義，畢竟仍可能有人相信AI，抑或雖不相信、但仍可能被其影響對個案人士的態度。再者，質疑AI可性度的警示長久下來對社會是一種傷害：AI所提供的真實資訊也將因此不被社會輕信，這除了將導致AI的功用大打折扣外，整體社會的交易成本也將因此大為增加。面對AI，很顯然地，我們需要新的法律制度。

為了規範AI的使用與發展，在ChatGPT這一類聊天機器人的問世前，歐盟其實已經在2021年提出人工智慧法（Artificial Intelligence Act）草案，這是一套基於風險分類的總體管理架構，如

果AI服務被分類為高風險（像是自動駕駛、由機器人協助手術的AI軟體、或為了徵才目的而進行履歷篩選的軟體），則企業必須遵守一系列的限制與要求，包含應對AI系統維持適當的人力監督、保存詳細完整的文件紀錄、確保高程度的安全與準確性，並應持續受到AI理事會（Artificial Intelligence Board）的監督¹⁷。不過，ChatGPT的問世迫使歐盟重新改寫人工智慧法的內容，因為這一類的AI能用於多種不同的使用用途，不僅限於特定目的¹⁸，增加了管理與監督AI的複雜與難度。

針對AI不實言論所產生的社會與法律問題，一篇名為「具真實性的AI」¹⁹論文即主張，對於AI系統的真实性標準應該要比人類更高，這是因為人類社會中能控制或抑制人類說謊或虛假行為的因素並不直接適用於AI，例如人類會為了名譽或社會地位而遵守乎社會規範，但AI並不在乎名譽或社會地位。該文指出，對AI適用較高真实性標準的成本比在人類社會低，例如我們有辦法在AI系統正式部署之前給予認證（但我們無法在人被「部署」前予以認證），或者我們能夠紀錄AI的所有言論以供未來裁決（adjudicate）用。為了確保並規範AI的真实性，該文提議了兩種解決方法：一是在AI系統部署啟用前，由認證者（certifier）對AI進行測試，如果AI能通過真實性的測試，發給認證（certification）；

另一則是在AI系統部署啟用後，記錄所有AI言論與使用者對話，在有爭議時由裁決者（adjudicator）來裁決AI是否違反真實性。不管是認證或裁決，都需要有一個第三方機構來承擔相關的工作，這是一個全新的職責跟功能，可能由一個全新的第三方機構負責，也可以是現有的機構承擔新的功能跟角色。

雖然「具真實性的AI」這篇文章並未直接處理AI適用「誹謗」責任的問題，但該文試圖更根本地提升AI言論的真實性與可信度，不論該文提議的方法是否獲得支持跟贊同，至少這為人類社會解決AI帶來的制度挑戰提供了樂觀的開始。♣

註釋

- * 本文為作者個人一般性觀點，非作者對具體個案的法律意見或建議，亦不代表事務所立場。
1. Pranshu Verma & Will Oremus, *What happens when ChatGPT lies about real people?* *The Washington Post*, 5 Apr. 2023, <https://www.washingtonpost.com/technology/2023/04/05/chatgpt-lies/> (last visited on June 5, 2023).
 2. Sophia Khatsenkova & Natalie Huet, *Mayor mulls defamation lawsuit after ChatGPT falsely claims he was jailed for bribery*, *Euronews*, 8 Apr. 2023, <https://www.euronews.com/next/2023/04/07/why-does-chatgpt-make-things-up-australian-mayor-prepares-first-defamation-lawsuit-over-it> (last visited on June 5, 2023).
 3. Cade Metz著，王曉伯譯，AI製造商沒說的祕密：企業巨頭的搶才大戰如何改寫我們的世界？，時報文化，2022年。
 4. Sharon Goldman, *Nvidia will bring AI to every industry, says CEO Jensen Huang in GTC keynote: "We are at the iPhone moment of AI"*, *Venture Beat*, 21 Mar. 2023, <https://venturebeat.com/ai/nvidia-will-bring-ai-to-every-industry-says-ceo-jensen-huang-in-gtc-keynote-we-are-at-the-iphone-moment-of-ai/> (last visited on June 5, 2023).
 5. 此為筆者個人最關心的AI法律議題，日後有機會為文討論。
 6. 因各國法律對誹謗責任的規定不同，美國尚有各州州法差異，為方便探討，本段討論以美國普通法為準。且本段討論暫時不考慮管轄問題以及AI言論是否構成美國憲法第一修正案保護言論的問題。
 7. 依據 *New York Times v. Sullivan* (1964)以及*Curtis Publishing Com v. Butts* (1967)案。
 8. Dan Ray, *AI Defamation and Libel*, *Nolo*, <https://www.nolo.com/legal-encyclopedia/artificial-intelligence-defamation-and-libel-is-anyone-liable.html> (last visited on June 5, 2023).
 9. 據報導，一位紐約州律師因為引用ChatGPT所提供但實際並不存在的6個案例，而可能面臨紀錄處分，見

- Kathryn Armstrong, *ChatGPT: US lawyer admits using AI for case research*, BBC News, 28 May 2023, <https://www.bbc.com/news/world-us-canada-65735769> (last visited on June 5, 2023); 當這類的新聞越多，沒有人會相信AI言論真實性的抗辯可能會越來越有說服力。
10. Seth C. Lewis, Amy Kristin Sanders & Casey Carmody, *Libel by Algorithm? Automated Journalism and the Threat of Legal Liability*, 96(1) JOURNALISM & MASS COMMUNICATION QUARTERLY 60-81 (2019).
 11. Danny Goodwin, *Google Fined \$65K in France for 'Crook' Autocomplete Suggestion*. *Search Engine Watch*, 10 Jan. 2012, <https://www.searchenginewatch.com/2012/01/10/google-fined-65k-in-france-for-crook-a-utocomplete-suggestion/> (last visited on June 5, 2023).
 12. BBC News, *Germany tells Google to tidy up auto-complete*, BBC News, 14 May 2013, <https://www.bbc.com/news/technology-22529357> (last visited on June 5, 2023).
 13. 自動輔助駕駛爭議上法庭 特斯拉初戰全身而退，自由時報電子報，2023年4月22日，<https://news.ltn.com.tw/news/world/breakingnews/4278695>（最後瀏覽日：2023年6月5日）。
 14. Abhirup Roy, Dan Levine & Hyunjoo Jin, *Tesla wins bellwether trial over Autopilot car crash*, Reuters, 22 Apr. 2023, <https://www.reuters.com/legal/us-jury-set-decide-test-case-tesla-autopilot-crash-2023-04-21/> (last visited on June 5, 2023).
 15. European Commission, *Proposal for a Directive of the European Parliament and of the Council on adapting non-contractual civil liability rules to artificial intelligence (AI Liability Directive)*, EUR-Lex, 28 Sep. 2022, <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52022PC0496> (last visited on June 5, 2023).
 16. European Commission, *Proposal for a Directive of the European Parliament and of the Council on liability for defective products*, EUR-Lex, 28 Sep. 2022, <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52022PC0495> (last visited on June 5, 2023).
 17. European Commission, *Regulatory framework proposal on artificial intelligence*, EUR-Lex, 29 Sep. 2022, <https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai> (last visited on June 5, 2023).
 18. Gian Volpicelli, *ChatGPT broke the EU plan to regulate AI*. *POLITICO*, 3 Mar. 2023, <https://www.politico.eu/article/eu-plan-regulate-chatgpt-openai-artificial-intelligence-act/> (last visited on June 5, 2023).
 19. Owain Evans, Owen Cotton-Barratt, Lukas Finnveden, Adam Bales, Avital Balwit, Peter Wills, Luca Righetti, William Saunders, *Truthful AI: Developing and governing AI that does not lie*, Cornell University, 13 Oct. 2021, <https://arxiv.org/pdf/2110.06674.pdf> (last visited on June 5, 2023).

關鍵詞：人工智慧，誹謗，真實惡意，演算法

DOI：10.53106/279069731708

（相關文獻☛月旦知識庫 www.lawdata.com.tw；
更多裁判分析☛月旦法律分析庫 lawwise.com.tw）