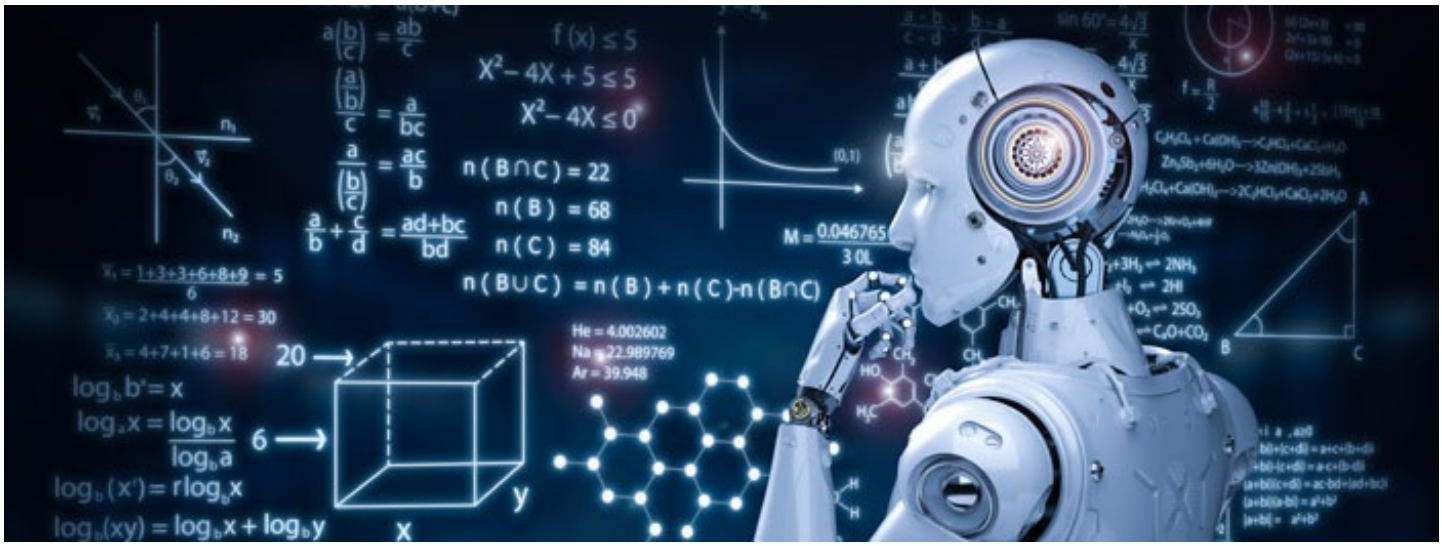
落實完善數位資料管理機制，有助於降低AI歧視及資料外洩風險



落實完善數位資料管理機制，
有助於降低AI歧視及資料外洩風險

資訊工業策進會科技法律研究所
2023年07月07日

近年來，科技快速發展，AI(人工智慧)等技術日新月異，在公私部門的應用日益廣泛，而且根據美國資訊科技研究與顧問公司Gartner在2023年5月發布的調查指出，隨著由OpenAI開發的ChatGPT取得成功，更促使各領域對於AI應用的高度重視與投入[1]，與此同時，AI歧視及資料外洩等問題，亦成為社會各界的重大關切議題。

壹、事件摘要

目前AI科技發展已牽動全球經濟發展，根據麥肯錫公司近期發布的《生成式人工智慧的經濟潛力：下一個生產力前沿(The next productivity frontier)》研究報告指出，預測生成式AI(Generative AI)有望每年為全球經濟增加2.6兆至4.4兆的經濟價值[2]。同時在美國資訊科技研究與顧問公司Gartner對於超過2500名高階主管的調查中，45%受訪者認為ChatGPT問世，增加其對於AI的投資。而且68%受訪者認為AI的好處大於風險，僅有5%受訪者認為風險大於好處[3]。然而有社會輿論認為AI的判斷依賴訓練資料，將可能複製人類偏見，造成AI歧視問題，而且若程式碼有漏洞或帳戶被盜用時，亦會造成資料外洩問題。

貳、重點說明

首先，關於AI歧視問題，以金融領域為例，近期歐盟委員會副主席Margrethe Vestager強調若AI用於可能影響他人生計的關鍵決策時，如決定是否能取得貸款，應確保申請人不受性別或膚色等歧視[4]，同時亦有論者認為若用於訓練AI的歷史資料，本身存有偏見問題，則可能導致系統自動拒絕向邊緣化族群貸款，在無形之中加劇，甚至永久化對於特定種族或性別的歧視[5]。

其次，關於資料外洩問題，資安公司Group-IB指出因目前在預設情況下，ChatGPT將保存使用者查詢及AI回應的訊息紀錄，若帳戶被盜，則可能洩露機敏資訊。據統計在2022年6月至2023年5月間，在亞太地區有近41000個帳戶被盜，而在中東和非洲地區有近25000個帳戶被盜，甚至在歐洲地區也有近17000個帳戶被盜[6]。另外在2023年3月時，ChatGPT除了發生部分用戶能夠檢視他人聊天紀錄標題的問題外，甚至發生個人資料外洩問題，即用戶可能知悉他人的姓名、電子郵件，付款地址，信用卡到期日及號碼末四碼等資料[7]。

參、事件評析

對於AI歧視及資料外洩等問題，應透過落實完善數位資料治理與管理機制，以降低問題發生的風險。首先，在收集訓練資料時，為篩選適合作為模型或演算法基礎的資料，應建立資料評估或審查機制，減少或避免使用有潛在歧視問題的資料，以確保分析結果之精確性。

其次，不論對於訓練資料、分析所得資料或用戶個人資料等，均應落實嚴謹的資料保密措施，避免資料外洩，如必須對於資料進行標示或分類，並依照不同標示或分類，評估及採取適當程度的保密措施。同時應對於資料進行格式轉換，以無法直接開啟的檔案格式進行留存，縱使未來可能不慎發生資料外洩，任意第三人仍難以直接開啟或解析資料內容。甚至在傳送帳戶登入訊息時，亦應採取適當加密傳送機制，避免遭他人竊取，盜取帳戶或個人資料。

財團法人資訊工業策進會科技法律研究所長期致力於促進國家科技法制環境完善，於2021年7月發布「重要數位資料治理暨管理制度規範（Essential Data Governance and Management System，簡稱EDGS）」，完整涵蓋數位資料的生成、保護與維護，以及存證資訊的取得、維護與驗證的流程化管理機制，故對於不同公私部門的AI相關資料，均可參考EDGS，建立系統性數位資料管理機制或強化既有機制。

本文同步刊登於TIPS網站（https://www.tips.org.tw）

[1]Gartner, *Gartner Poll Finds 45% of Executives Say ChatGPT Has Prompted an Increase in AI Investment* (May 3, 2023), https://www.gartner.com/en/newsroom/press-releases/2023-05-03-gartner-poll-finds-45-percent-of-executives-say-chatgpt-has-prompted-an-increase-in-ai-investment (last visited June 30, 2023).

[2]McKinsey, *The economic potential of generative AI: The next productivity frontier* (June 14, 2023), https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/the-economic-potential-of-generative-AI-the-next-productivity-frontier#introduction (last visited June 30, 2023).

[3]Gartner, *supra* note 1.

[4]Zoe Kleinman, Philippa Wain & Ashleigh Swan, *Using AI for loans and mortgages is big risk, warns EU boss* (June 14, 2023), https://www.bbc.com/news/technology-65881389 (last visited June 30, 2023).

[5]Ryan Browne & MacKenzie Sigalos, *A.I. has a discrimination problem. In banking, the consequences can be severe* (June 23, 2023), https://www.cnbc.com/2023/06/23/ai-has-a-discrimination-problem-in-banking-that-can-be-devastating.html (last visited June 30, 2023).

[6]Group-IB, *Group-IB Discovers 100K+ Compromised ChatGPT Accounts on Dark Web Marketplaces; Asia-Pacific region tops the list* (June 20, 2023), https://www.group-ib.com/media-center/press-releases/stealers-chatgpt-credentials/ (last visited June 30, 2023).

[7]OpenAI, *March 20 ChatGPT outage: Here's what happened* (Mar. 24, 2023),https://openai.com/blog/march-20-chatgpt-outage (last visited June 30, 2023).

你可能會想參加

→ 【2023科技法制變革論壇】AI生成時代所帶動的ChatGPT法制與產業新趨勢
→ 製造業及技術服務業個資保護及資安落實－經濟部工業局112年企業個人資料保護暨資訊安全宣導說明會
→ 【已額滿】2023科技研發法制推廣活動—科專個資及反詐騙實務講座
→ 供應鏈資安國際法制與政策趨勢分享會
→ 「跨域數位協作與管理」講座活動
→ 新創採購-政府新創應用分享會
→ 【實體】數位發展部數位經濟相關產業個資安維辦法說明會（南部場）
→ 【線上】數位發展部數位經濟相關產業個資安維辦法說明會（南部場）
→ 【線上場】113年「新創採購機制及鼓勵照護機構參與推動」說明會
→ 【北部場】113年「新創採購機制及鼓勵地方政府參與推動」說明會
→ 【南部場】113年「新創採購機制及鼓勵地方政府參與推動」說明會
→ 商業服務業個資保護宣導說明會
→ 113年新創採購-照護機構獎勵說明會
→ 【南部場】113年「新創採購機制及鼓勵地方政府參與推動」說明會
→ 【北部場】113年「新創採購機制及鼓勵地方政府參與推動」說明會
→ 【中部場】113年「新創採購機制及鼓勵地方政府參與推動」說明會
→ 【實體】2024科技研發法制推廣活動— 科專個資及反詐騙實務講座
→ 【直播】2024科技研發法制推廣活動— 科專個資及反詐騙實務講座
→ 【臺北場】113年度新創採購-招標作業廠商說明會
→ 【臺中場】113年度新創採購-招標作業廠商說明會
→ 【高雄場】113年度新創採購-招標作業廠商說明會
→ 數位發展部數位產業署113年資訊服務業者個資安維辦法宣導說明會

梁景濠
副法律研究員 編譯整理

上稿時間：2023年07月

文章標籤

人工智慧 資訊安全 金融科技

推薦文章