

AI 系統的資安 更有急迫性

文／黃光彩（臺灣量子安全協會理事長）

近期人工智慧風動整個科技產業，自從今年六月Computex 各大巨頭來台，黃仁勳的 Nvidia 極力推動 AI Foundry 的一股潮流。AI 的實際應用相當廣泛，農業甚至是醫療產業均適用。但是 AI 的劣勢在於資安風險，各國都積極制定風險控管的標準及測試方法，我們將討論一系列的活動，並提供將風險降至最低的建議策略，以推廣安全、符合道德使用的 AI 模型。

AI 系統在安全性、隱私和安全性方面面臨許多挑戰。包括：

- 1) 數據洩漏和濫用。2) 系統被攻擊。3) 偏見和歧視 等。
- 1.數據洩露和濫用：**AI 系統需要大量數據來訓練和運行，這些數據可能包含敏感的個人信息。如果這些數據被洩露或濫用，可能會導致身份盜竊、財務欺詐等問題。
- 2.系統被攻擊：**隨著 AI 系統變得越來越複雜和自動化，它們也變得更容易受到網絡攻擊，攻擊者可能會利用這些系統來做出有害的決策。
- 3.偏見和歧視：**AI 系統可能會在決策過程中引入偏見，導致對某些個人或群體的不公平待遇。例如，亞馬遜曾經開發過一個 AI 招聘工具，但發現它對女性存在偏見。

事件的例子：

2017 年 Equifax 數據洩露事件：

信用報告機構 Equifax 發生了大規模數據洩露事件，導致約 1.43 億美國消費者的個人信息被洩露，包括社會安全號碼、出生日期、地址和駕照號碼。

2018 Facebook-Cambridge Analytica 事件：

Facebook 被曝出與數據分析公司 Cambridge Analytica 之間的數據共享醜聞。Cambridge Analytica 未經用戶同意，獲取了數千萬 Facebook 用戶的個人數據，並用於政治廣告和選舉活動

2013 Target 購物數據洩露事件：

零售巨頭 Target 發生數據洩露事件，黑客獲取了約 4000 萬顧客的信用卡和借記卡信息。這次洩露事件導致大量的金融欺詐和信用卡盜用問題，並對 Target 的聲譽造成了嚴重影響。

解決方法：

1. 數據保護措施：實施數據加密和安全存儲，並制定數據洩露應對方案，確保只有經授權的人員可以訪問敏感數據。
2. 隨著 AI 系統變得越來越複雜和自動化，它們也變得更容易受到網絡攻擊，攻擊者可能會利用這些系統來做出有害的決策。定期進行系統監控和安全評估，以識別和修復潛在的漏洞，可以通過對抗性訓練來提高系統對攻擊的抵抗力。
3. 在 AI 系統的開發過程中，從一開始就考慮隱私問題。這包括收集和保留必要的個人數據，並使用技術來保護數據中的個人身份。
4. AI 系統可能會在決策過程中引入偏見，導致對某些個人或群體的不公平待遇。例如，亞馬遜曾經開發過一個 AI 招聘工具，但發現它對女性存在偏見。確保 AI 系統的決策過程透明，並設立機制來檢測和糾正偏見，這樣可以提高系統的公平性和可信度。
5. 遵循相關的法律和倫理指南，確保 AI 系統的開發和使用符合道德標準和法律要求。

AI 包括一般的機器學習 (ML)，大型語言模型 (LLM)，生成式 AI 等領域，這些措施可以幫助減少 AI 系統在安全性、隱私和安全性方面的風險，並

確保它們能夠安全、負責任地運行。整個產業需要一起努力制定標準，實施驗證及確認 (verification & validation)，建立方法，收集實用所可能產生的問題，持續性的改進。這是一個值得持續關注的議題。



作者黃光彩博士，現為臺灣量子安全協會理事長，亞洲大學講座教授，曾是台師大校長、IBM 全球副總、企業董事。專長為 AI、知識管理、供應鏈金融、企業轉型等。專注於量子計算、後量子資安、高速運算的散熱系統從晶片到資料中心。