# AI 時代的網路安全

英美電影《模仿遊戲》（The Imitation Game）曾介紹圖靈於二戰期間協助盟軍破譯德軍密碼的真實故事，而圖靈當時所發明的密碼機即為現代電腦雛形。（Photo Credit: The Weinstein Company）

◆ 中興大學國際政治研究所副教授 — 譚偉恩

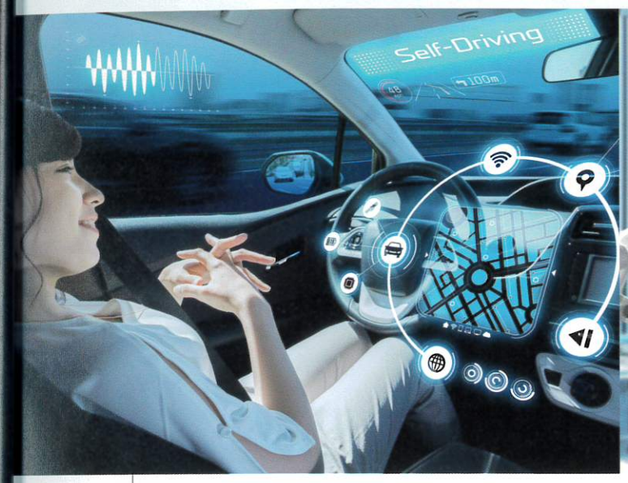
英國數學家圖靈（Alan Turing）於 1950 年經由測試發現，計算機在特定條件下可以和人類的心智思考相比擬，進而提出「人工智慧系統」（artificial intelligence system）的概念。[1]

隨著數理計算機的相關技術日臻成熟，人類社會的經濟、教育、醫療、托育長照、環境保護、交通運輸，以及公共行政和執法等，無不在一定程度上與圖靈提出的人工智慧（AI）鑲嵌在一塊兒。

[1] Alan Turing, "Computing Machinery and Intelligence," *Mind*, Volume LIX, Issue 236 (October 1950): 433-460.








AI 可以概分為兩種亞型：一種是與人類有互動的輔助型智能系統，旨在幫助人類更快更好地完成工作（左）；第二種亞型的 AI 不直接與人互動，多數是工廠中自動化生產的智能系統（右）。

## AI 的應用及其問題

市場上目前 AI 技術的開發主要是以創建「像人類一樣思考」的優等高階 AI 為主軸，藉由將計算機智能化，來分析客觀環境、學習特定事物，然後產出近似人類的理性判斷，但結果上更為精準。根據普華永道（PwC）的研究，大部分的 AI 可以概分為兩種亞型（subtype）：一種是與人類有互動的輔助型智能系統，旨在幫助人類更快更好地完成工作（例如停車）；此種類型的 AI 有時包括自我調適能力，可以配合使用者的實地需要做出情勢判斷，並在與人互動時進行自我學習和調整。第二種亞型的 AI 不直接與人互動，多數是工廠中自動化生產的智能系統；這種 AI 的工作範疇是固定的，很少會被增設新的工作項目，但在既有的工作內容中，其生產效率會透過自我學習而不斷提升。[2]

無論上面哪一種 AI 技術及其應用，計算機的功能與效率都會漸漸超越原始設計的智能水平，因此有可能對它的設計者、使用者，甚至是不特定的人群構成風險。詳言之，AI 在執行任務過程中的自我學習與資訊累積，讓它的適應能力與技術效能越來越純熟精準，以致有可能發生排除人類而獨自行動的風險。一個引起關注的例子是「AI 自動履歷篩選」；在美國已有高達 75% 以上的企業採用 AI 技術招募新人，

[2] 如果搭配物聯網系統，還可以輔助工廠生產線的管理事務。

取代傳統耗時的人資部門面試。然而，純熟精準的 AI 欠缺彈性，會將有額外才能或極具創意的求職者判定為資格不符，甚至有些企業的 AI 篩選系統會將主管核可的應聘者從名單中移除，導致企業最終痛失良才。[3]

AI 在應用上的另一個問題就是對於數據資料的取得和分析，這一部分與網路安全密切相關，有越來越多的犯罪是在網路上利用 AI 進行侵權和獲利。該如何因應，讓 AI 的總體效益大於潛在損害的結果，是 AI 技術與應用普及化的同時，難以迴避之挑戰。舉例來說，蒐集、分析和處理不特定多數人的某些資料是應用 AI 的關鍵環節。企業需要這些資料來進行 AI 的培訓，進而應用於廣告行銷和線上商務；國家需要這些資料作為政策擬定時的參考，或與人民互動交流意見，落實政策的風險溝通和施政彈性。由於透過 AI 蒐集和分析大數據變得越來越頻繁，掌握這些數據資料的使用者便取得了不對稱的資訊優勢，一旦用於犯罪，後果往往不堪設想。然而，對這些數據取得或使用的嚴格規範會減緩 AI 的發展，兩者間要如何平衡是各國正面臨的兩難困境。

## AI 對網路安全造成的威脅

AI 對網路安全的影響大約從 2017 年被各國正視，[4] 美國的 FBI 甚至針對犯罪組織使用 AI 的問題召開過專門會議。由

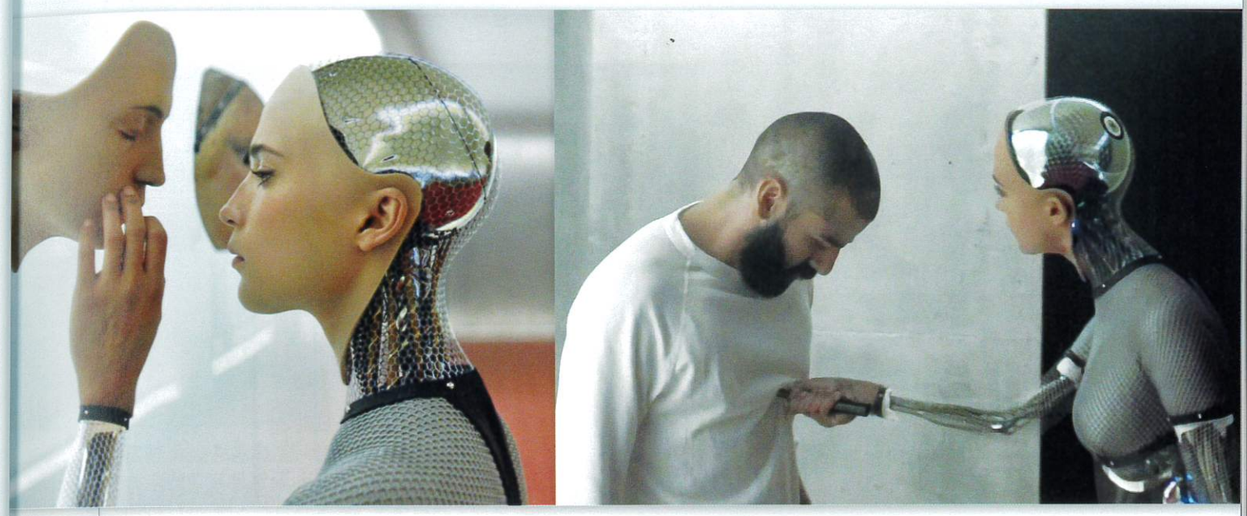
透過 AI 蒐集和分析大數據變得越來越頻繁，掌握這些數據資料的使用者便取得了不對稱的資訊優勢，一旦用於犯罪，後果往往不堪設想。

[3] Sarah K. White, "AI in Hiring Might Do More Harm than Good," *CIO*, (September 17, 2021), via at: https://www.cio.com/article/189212/ai-in-hiring-might-do-more-harm-than-good.html

[4] 一篇非常具有參考價值的論文是：Alex Wilner, "Cybersecurity and Its Discontinuities: Artificial Intelligence, the Internet of Things, and Digital Misinformation," *International Journal*, Vol. 73, No. 2 (June 2018): 308-316.







無論哪種 AI 技術，AI 功能都會漸漸超越原始設計的智能水平，因此有可能對它的設計者、使用者，甚至是不特定的人群構成風險。英國電影《人造意識》（Ex-Machine）即描述完美的 AI 機器人，最後卻殺掉設計者之情節。（Photo Credit: Universal Pictures）

於網路是一個虛擬空間，讓侵害權利的犯罪行為得以隱身其中，並藉助科技帶來之轉換效果對真實世界的秩序造成破壞。英國倫敦大學學院的報告指出，犯罪者透過 AI 技術破解密碼、複製人類語音，以及其他諸多的非法侵權技術。其中深偽技術（deepfake）被列為犯罪結合 AI 後對網路安全的首要威脅之一，因為這有可能讓人們對任何影音或視頻資訊的傳遞失去信任感，嚴重妨礙人類社會資訊交換與傳播的現狀。此外，上述報告也指出，運用 AI 的犯罪與傳統犯罪不同之處在於，它的犯罪效能可以在網路上被快速分享、重製與再現，甚至在犯罪組織的包裝下成為一種「服務」來銷售，以致國家司法機構難以有效抑制。[5]

此外，COVID-19 疫情爆發後，各國遠距工作人數大增，導致網路端點之間的聯繫暴露在風險中。許多企業或是智庫的分析報告均指出，資訊科技（IT）與營運科技（OT）已成為網路犯罪者的主要侵權對象，特別是數位支付及加密貨幣的攻擊事件或竊取行為明顯增加。由於犯罪者可以透過 AI 的協助來生產惡意軟體或非法取得個資，再將之出售給其他犯罪者來營利，暗網交易變得越來越熱絡。[6] 相較於過去，網路侵權犯罪多半是由專業的駭客為之，

[5] 詳見："AI-enabled Future Crime," https://www.ucl.ac.uk/jill-dando-institute/sites/jill-dando-institute/files/ai_crime_policy_0.pdf。

[6] Wytske van der Wagen and Wolter Pieters, "From Cybercrime to Cyborg Crime: Botnets as Hybrid Criminal Actor-Networks," *The British Journal of Criminology*, Vol. 55, No. 3 (May 2015): 578-595.

英國倫敦大學報告指出，犯罪者透過 AI 技術破解密碼、複製人類語音，以及其他諸多的非法侵權技術。美國電影《關鍵報告》（Minority Report）即描述凶嫌透過 AI 技術，將責任移花接木嫁禍予無辜者之情節。（Photo Credit: FOX）

但 AI 與暗網交易結合之後，資訊科技與營運科技會面臨更多元與廣泛的網路攻擊。顯然，AI 技術的普及化增加了我們正規生活中面臨威脅之風險，而這些風險在網路世代可歸類為兩大類：一、惡意軟體攻擊；二、涉及社交工程（social engineering）的技術性攻擊。

第一類可以說是犯罪者受惠於 AI 的最佳證明；由於 AI 在速度和效率方面的突出表現，讓犯罪者得以將之用以強化勒索軟體的破壞性，升級病毒避開防火牆、深入企業計算機網路，癱瘓運作並竊取重要資料。第二類是藉由 AI 技術編寫縝密的「故事」進行社交詐騙；犯罪者透過 AI 有系統地分析特定人士的網路使用慣性，再設計個人化的「故事」進行網路詐騙。數據安全專家 George Dvorsky 及 Brian Wallace 等人曾經指出，AI 是兩面刃，對駭客或有心犯罪人士而言，是絕佳的新一代武器。[7]

## 犯罪與相關風險之因應

許多國家已經發現，既存的法律規範很難對網路上的 AI 犯罪行為進行有效管制。或許也因為如此，私人性的網路安全措施相繼推出，例如較具代表性的阿西洛馬人工智慧原則（Asilomar AI Principles）。[8] 這個原則已獲得諸多業界人士的廣泛支持，在總共 23 項的原則性內容中，有幾個面向值得吾人注意。

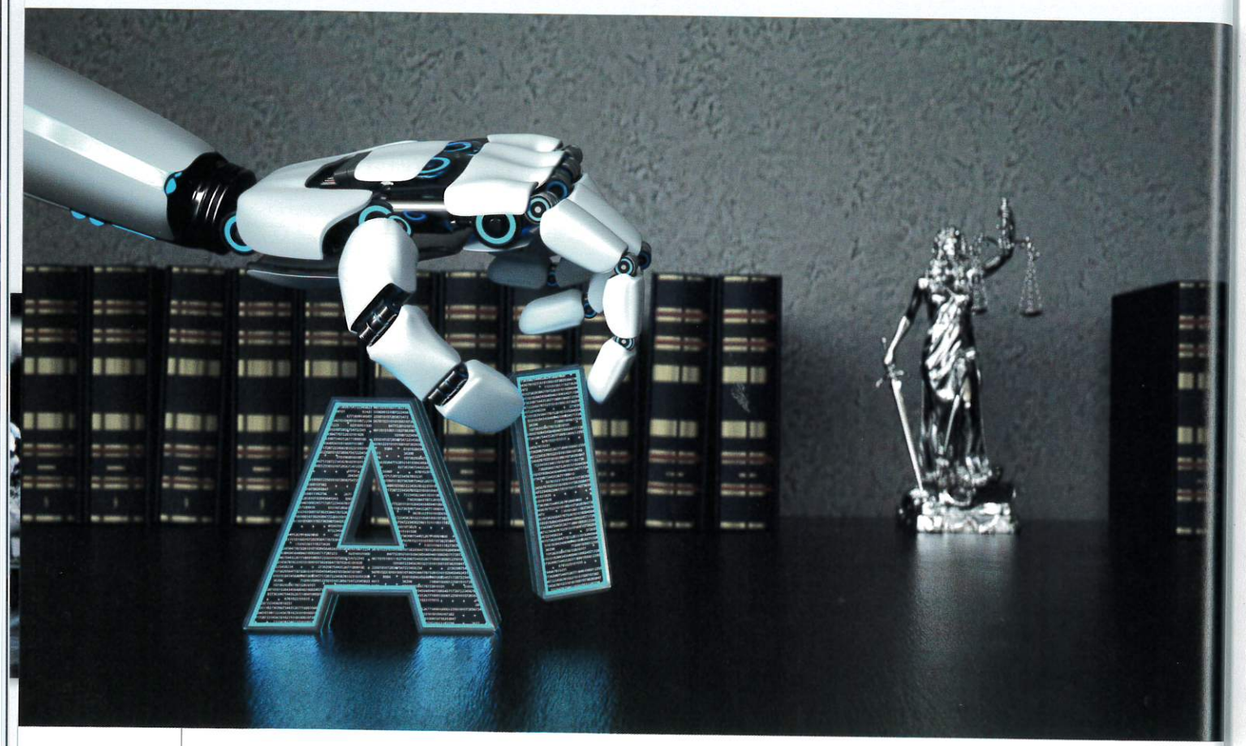
首先，AI 研發之目的與使用可能在不久的將來會成為立法時的考量。研發者有義務對自己 AI 系統承擔責任；由於 AI 的自主性是透過海量數據資料的學習而來，但 AI 對什麼樣的資料感到興趣卻是研發者「價值觀」的反映。鑑此，在設計之初就應明確化與公開 AI 的目的，並同時在手段（means）與目標（goals）上給予清楚的說明。根據此原則，具有攻擊性或使用目的不明確的 AI 或相關應用，日後在立法上就應受到高密度審查，若研發者無法清楚交代此類資訊，政府就不應核可。

第二，因為 AI 一定會建立屬於自己的獨立性，所以「未來的不可控」必然是會存在之風險，研發者與使用者都應該預見且提早預防此類風險。第 7 項原則中特別提到 AI 如果出現意外也必須要有透明性，即對於釀成損害的因果關係要明確呈現並客觀上歸責。此外，第 8 項的審判透明性強調「司法性的決策」不能逸脫人類社會合理之解釋範疇，並且最終要由人類的監管機構保留審核權。


圖 1　阿西洛馬人工智慧 23 條原則內容

[7] George Dvorsky, "Hackers Have Already Started to Weaponize Artificial Intelligence," *Gizmodo*, (September 11, 2017), via at: https://gizmodo.com/hackers-have-already-started-to-weaponize-artificial-in-1797688425

[8] 此原則的製訂背景與詳細內容可見：https://futureoflife.org/2017/08/11/ai-principles/。




AI 的治理不妨思考以法律賦予 AI 適當之權利與義務，使其對可能致生的風險或實害承擔責任。

最後，是因 AI 而產生的利益應予開放及盡可能公有化。在原則的第 23 項提及「公共利益」，認為 AI 的問世及應用要符合全人類的利益，而不是某一個國家或特定組織之利益。然而，「利益」的定義是什麼？這個問題的爭議性幾乎是不可能解決的，而定義不清楚的規範，無論是私人機構或公家部門，在執行上都會有困難，其最終的結果就是法律漏洞。

## 結語

不久的將來，人類社會現行的法律制度就會因為 AI 技術的發展和相關網路應用而大幅修正，其中網路犯罪的防治和相關侵權行為的歸責與賠償機制極需被處理。鑑於 AI 自我學習及自我調適後所可能產生的不確定風險，本文建議對於 AI 的治理不妨思考以法律賦予 AI 適當之權利與義務，概念上類似透過立法擬制給予 AI 有限或準法人的資格，使其對可能致生的風險或實害承擔責任。有別於傳統以自然人或法人為中心的立法，保障因 AI 的應用或商業化使用而發生之權利受損並提供救濟，是新一代治理規範的主旨。