當網頁愛上人工智慧

◆ 社團法人台灣 E 化資安分析管理協會、嘉義大學資訊工程系教授 — 王智弘

要在多管齊下的誘騙中全身而退，最好的防範方式就是讓自己隔絕在威脅之外；而人工智慧是否能幫忙，一眼就看穿惡人的把戲？

原來是場騙局

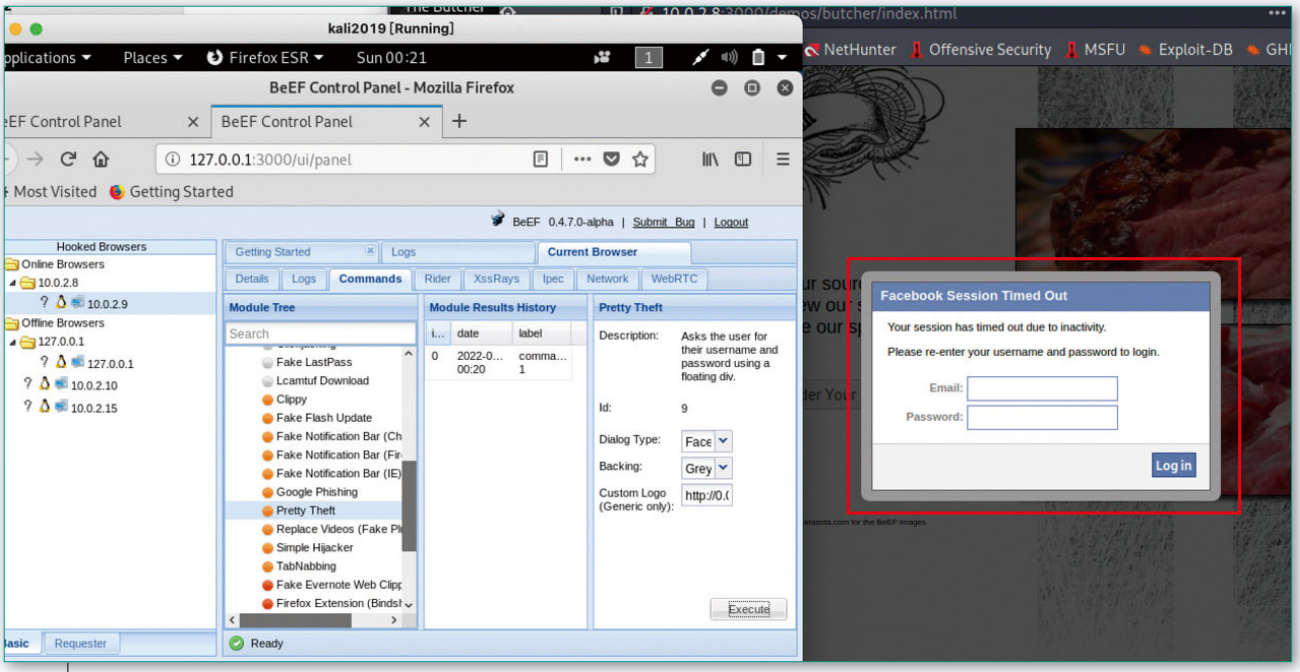
「盡信網路，不如無網路」，已成了現代人對於網路上充斥著太多假訊息，詐騙術無所不在的深沉無奈與抗議。以往享受於瀏覽網頁、沉浸在無論是文字知識的充實之樂，或是音樂影音的華麗饗宴，感受到無比的雀躍。現在卻得要處處防範、時時小心。深怕一個錯誤滑鼠的「click」，造成難以彌補的損失。在大量的影音互動所帶動的誘惑之下，詐騙的行為也因而開始升級。人們很難在多管齊下的誘騙之下能全身而退，最好的防範方式就是讓自己隔絕在這樣的威脅之外。然而，我們現今的科技足以支援這樣的服務嗎？哪些網站是有疑慮的？科技究竟能否幫我們忙，一眼就看穿惡人的把戲？

現今網路上充斥著大量假訊息，詐騙術無所不在；然而，亦有許多網站可能本身沒有惡意，但卻因為具有漏洞而遭駭客利用犯罪。

我們常聽到有一種駭客攻擊方法稱作「跨站腳本程式碼攻擊」（Cross-site Scripting, XSS），讓你看似網站是正常的，但卻是潛藏危機。這些網站可能本身是沒有惡意的，但卻因為具有漏洞（Vulnerability）而遭受了駭客栽贓的禍害。網路上種種迷幻的效果，讓人目不暇給，也讓人覺得眼見的內容竟然也非事實。譬如社交工程裡的釣魚（Phishing）手法，甚至是復刻整個網站內容以達到欺騙的目的。近期新聞便有關犯罪者製作假的銀行網站並傳送簡訊給被害者，由於太過仿真，使得好幾十人以上受騙，損失竟逾千萬元。在數位包圍下生活，我們對於實與虛、真與假、正本與副本的界線判定已退化，尋求外力協助是可以理解的想法。**「以科技解決科技所製造的問題」，看來是當前可能的藥方**，否則當有一天你發現了所有背後隱藏的攻擊程序，才驚覺，原來之前看到的那些亮麗的網路資訊，都只是個騙局。

欺騙花樣層出不窮

當你連上了惡意或是有漏洞的網站，它所能搞欺騙的花樣可謂千奇百怪。大家可能會想到的是，假的網站可能會盜取使用者的密碼。因此現在防範的方式類似透過一次性密碼（One-time Password, OTP），傳送簡訊到手機或 email 信箱。然而，實際上，駭客透過腳本程式碼，如 Java Script，可以變出許多不同的花樣，令人防不勝防。例如透過跳出式視窗（Popup Window）的社交工程方式，於網頁瀏覽



利用在 Kali Linux 中的 BeEF 工具進行漏洞利用（Exploitation）測試，出現 Session 過期的通知，誆騙使用者鍵入正確的密碼。（圖片來源：作者提供）

的時期跳出類似 Session 過期的通知，誆騙使用者鍵入正確的密碼。此外，還有多種不同型的攻擊運作，例如，透過啟動自動重新導向（Redirection）的方式或是修改 HREFs 的連線網址，讓使用者不自覺中連線到具有 Hook 的惡意網站；也有其他的手法像是開啟相機（Webcam）、播放聲音、偽造虛假的通知欄（Notification Bar）等。每個人在長期地接受這些攻擊，不禁要問，如何能還我一個乾淨的瀏覽空間，告訴我哪些網站可連，而哪些網站有安全疑慮呢？

黑名單與白名單

網站的安全評分是一直以來許多專家建議的方式。安全評分的方式透過許多綜合的指標來評估一個網站的安全性，也透過一些回報機制來登錄部分問題網站。我們可以從網路上查到許多這類的服務，包括像是針對釣魚網站的檢查，如趨勢科技。[1] 此外，Google 的「安全瀏覽」（Google Safe Browsing）每天也都會進行數十億個網站檢查，以找到可能的威脅。而像是 ScamAdviser [2] 則能夠檢測可能的釣魚及詐騙網站，相當具有準確性。另外，也有針對網站聲譽（Reputation）

[1] Trend Micro, https://global.sitesafety.trendmicro.com/

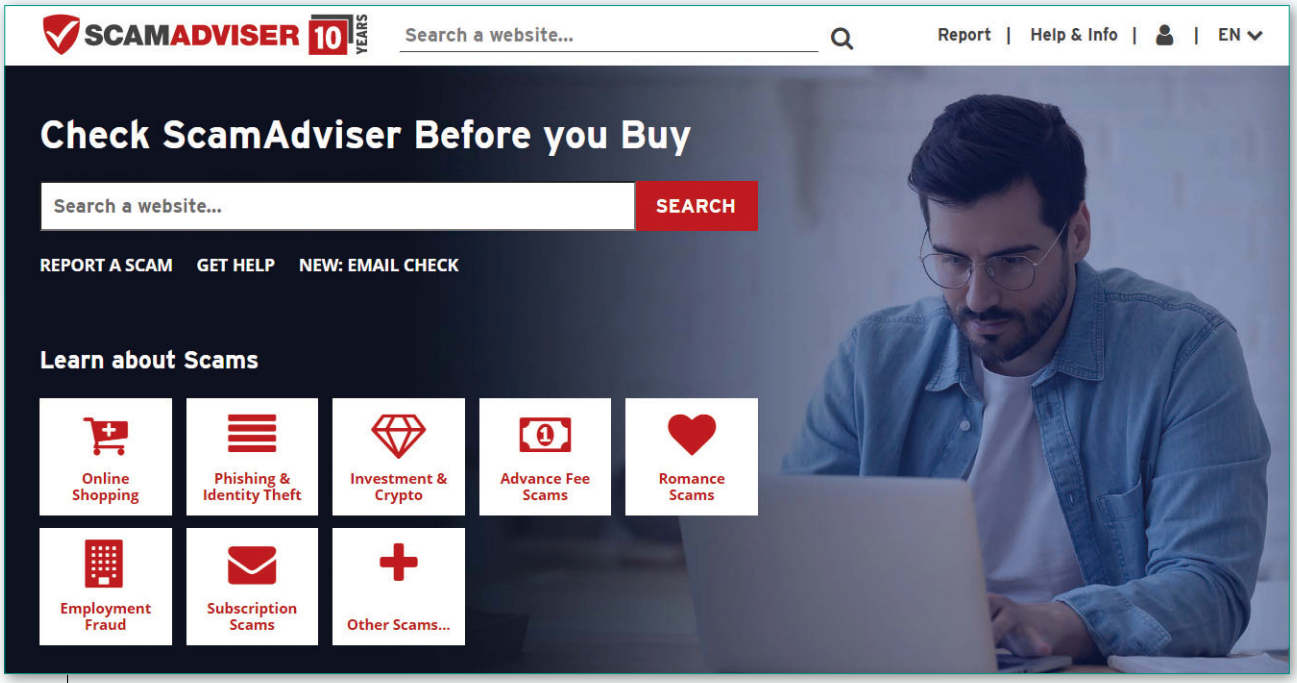
[2] https://www.scamadviser.com/

進行評分，如 URLVoid，[3] 能夠透過超過 40 個以上眾多不同的黑名單報告（Blacklist Report）資訊進行評估；亦有提供網域註冊（Domain Registration)， 從 whois 查詢網域資訊、Reverse DNS、ANS 以及位置資訊等。此外，著名病毒檢查網站 VirusTotal [4] 也可對於 URL 是否為惡意的情況進行檢查；而像是 Cisco Talos Intelligence [5] 也是一個相當知名的網站威脅分析工具。

上述的檢測服務，需要定期更新名單或是評估規則。因此雖基本上足夠使用，但難免也會有一些漏網之魚。此外，使用黑名單方法比較擔心的是因為檢測錯誤而導致用戶誤入有威脅的網站。另外一種方式則是建立白名單（Whitelist），只有被允許的網站或網域才能夠連上，其餘則進行攔阻。這樣做法安全性高，但對於用戶的限制也相對多，造成使用經驗與感受不佳。我們其實可以透過簡單的自救的方法，初步排除這些駭客的陷阱。

簡單自救方法

一、是否為安全加密連線？[6] 憑證（Certificate）是否有疑慮？



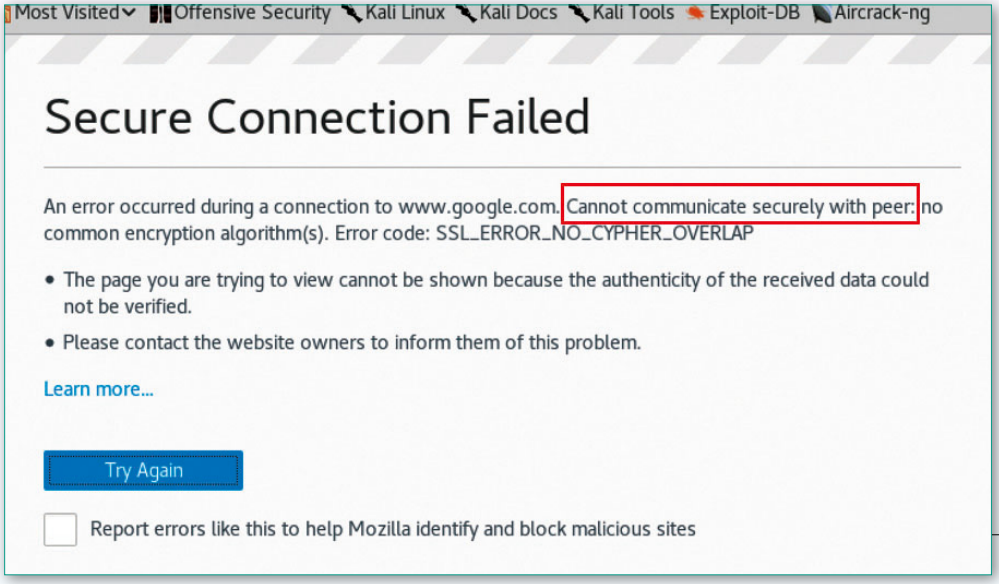
ScamAdviser 是一個免費的網站安全檢測服務，透過多種不同的指標來檢查網站是否安全可靠，使用者只要輸入網址就會顯示結果。（Source: https://www.scamadviser.com）

[3] https://www.urlvoid.com/

[4] https://www.virustotal.com/gui/home/url

[5] https://talosintelligence.com/

[6] 建立安全加密連線是保障資料的絕佳方案。我們連線的網站是否具備 TLS（Transport Layer Security）的安全機制雖然不是惡意網站判定的唯一方式，然而如果你的連線是正在進行帳號密碼的登錄，或是類似於購物網站要處理訂單或是信用卡資料的填寫，那加密與否就成了相當關鍵的問題。



Google 網站有強制安全傳輸的機制，連線若受到攻擊，會出現連線失敗的回應。（圖片來源：作者提供）

我們鍵入網址時，可能不會加上 https:// 或是 http://，但安全網站會將其轉換成 https 的安全連線。然而有項駭客的技術稱為 SSLStrip，可透過中間人攻擊，將原來要連線至 https 的重導向而映射到 http 連線，駭客因此能夠擷取重要的傳輸機密。而目前最新技術加上強制安全傳輸的機制（HTTP Strict Transport Security, HSTS），不允許跟網站之間進行無安全加密的傳輸，如此應可避免這類攻擊。此外，若遇到安全連線時憑證有問題的情況，如類似「您的連線不是私人連線」，或者是「網站的安全性憑證不可靠」等警告頁面，也請勿按下「仍要繼續」，以免引來隱藏風險而不自知。

二、睜大眼睛注意網址

我們在連線網站之前，通常將游標放在連線處，會出現連線的 URL 資訊。[7] 建議要注意 URL 的內容，以下有幾個簡單的判斷方式：

（一）故意與某些知名網站類似，但卻有一些差異，如 go0g1e，或是 rnicro.soft.com 之類的，讓使用者產生錯亂。

（二）縮短網址（Short URLs），例如，bit.ly、TinyURL 所提供的縮短網址服務，能夠取代長網址而使得連結的交換較為便利。然而由於這類短網址掩蓋了真正網址的諸多資訊，譬如真正的域名以及隱含的參數或檔名等，因此判斷良善或惡意並不容易。[8]

（三）網址前放置令人信賴名稱，如 google 後面再加上擴增的網域名。例如 http://login.google.com.

[7] 注意有些惡意透過 XSS 攻擊，其連線實際上是 Submit 按鈕以及一大串的填入資料，此時要避免與其連線。

[8] 基於過去許多安全的事件也因縮短網址而起，建議連線時仍要特別留意。

myphishing.com/welcome.html，上述顯然不是 google 的網站，但前面的域名卻又與 google 登入的名稱相同，藉以混淆視聽。[9]

（四）注意特殊字元，例如是否有類似 email 的 @ 符號，或是很多的點（dot）或斜線（/, slash）。譬如一般的網址其 dot 的數量大概為 3 個，如果過多，那麼可能會是有問題的網站，如上述 google login 的例子。

（五）查詢網域名稱註冊時間是否最近才建立；若是最近註冊，應考慮駭客為釣魚而建立的新網域。

（六）要特別留意連線的 URL 是否為 IP 而非網域名稱。

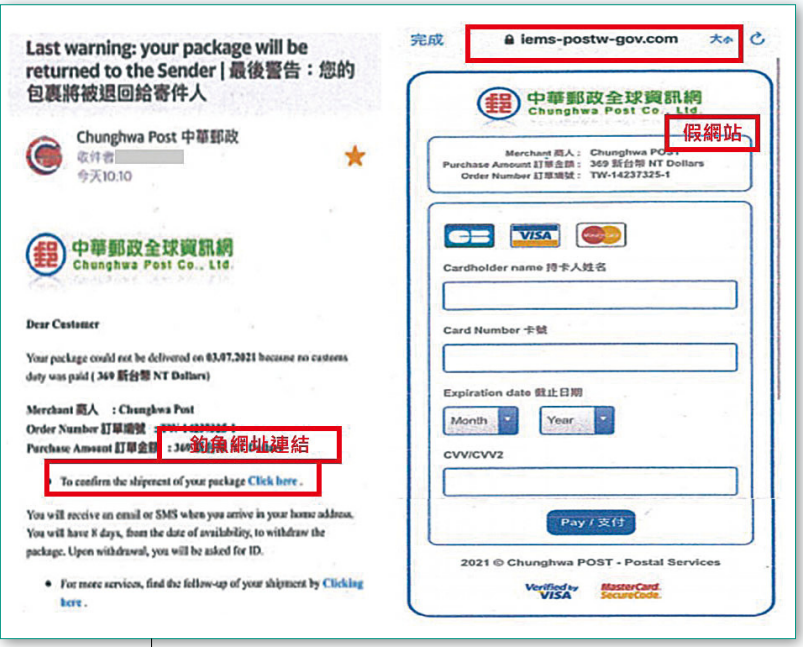
三、透過評分網站檢查後再連線

四、開網站後有問題，儘速離開

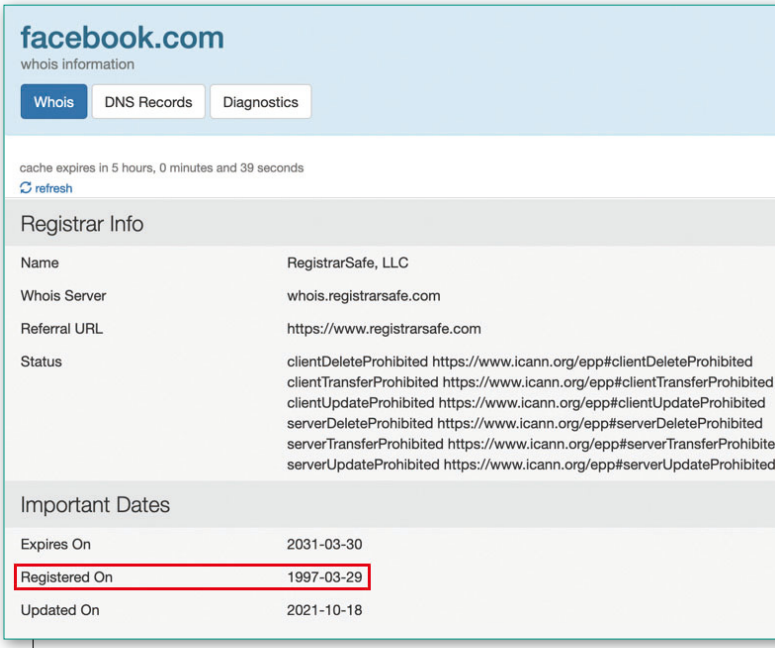
打開網站後要注意觀看內容，若有疑問，請儘速離開；然而有太多類型的攻擊是在一連線便進行，而且在極短時間內完成。

機器學習的可能與不可能

從上述的網址判別方法，可以思考透過更為自動的方式來進行。雖然目前已經有許多網站評分的服務，但若要找到潛藏的惡意網站，仍力有未逮。**透過機器學習（Machine Learning）的機制，以資料訓**



釣魚網站會故意採用與官方網站類似的網址，誘騙使用者登入，藉此竊取帳密資訊。（圖片來源：新北市政府警察局蘆洲分局，https://www.luzhou.police.ntpc.gov.tw/cp-1087-82938-23.html）



透過網域名稱註冊時間，可考慮是否為駭客為釣魚而建立的新網域；圖為 facebook.com 在 whois 所查詢的網域名稱註冊資訊。（圖片來源：作者提供）

[9] 注意 URL 長度，若過長，除了可能是上述的情況或是名稱編碼問題外，也可能是有一些惡意的參數輸入資料。

練方式替代人工制定規則，可能是對抗目前不斷激增且變異的惡意與釣魚網站的一個可選方案。

透過特徵（Feature）的篩選以及資料集的訓練，將會產生一個模型，[10] 該模型可儲存於雲端服務或是架設一臺代理伺服器（Proxy Server）以作為攔截檢查惡意連結，以及進一步深度檢測之用。圖 1 為可能的架構想法，表 1 則說明可能的特徵類型。

可以思考透過不同環境的訓練資料以強化情境分析。譬如有些惡意的連結來源是經由 Email，有一些是透過社群平臺，如 Facebook、Twitter 等，有些則是即時通訊如 Line、IG、Messenger 等，因此透過不同的訓練集或是模型參數，可以讓判斷更為精準，而若是對於網站有疑義，仍可經過一些深入的檢測模式（透過代理伺服器進行以避免用戶端身處險境）進行更為精準的判斷，提供用戶更好的安全監控及過濾服務。

網路安全與人工智慧之競合

面對科技，我們常會悠遊於它所帶來的便利，但也始終擔心它的負面效應。網

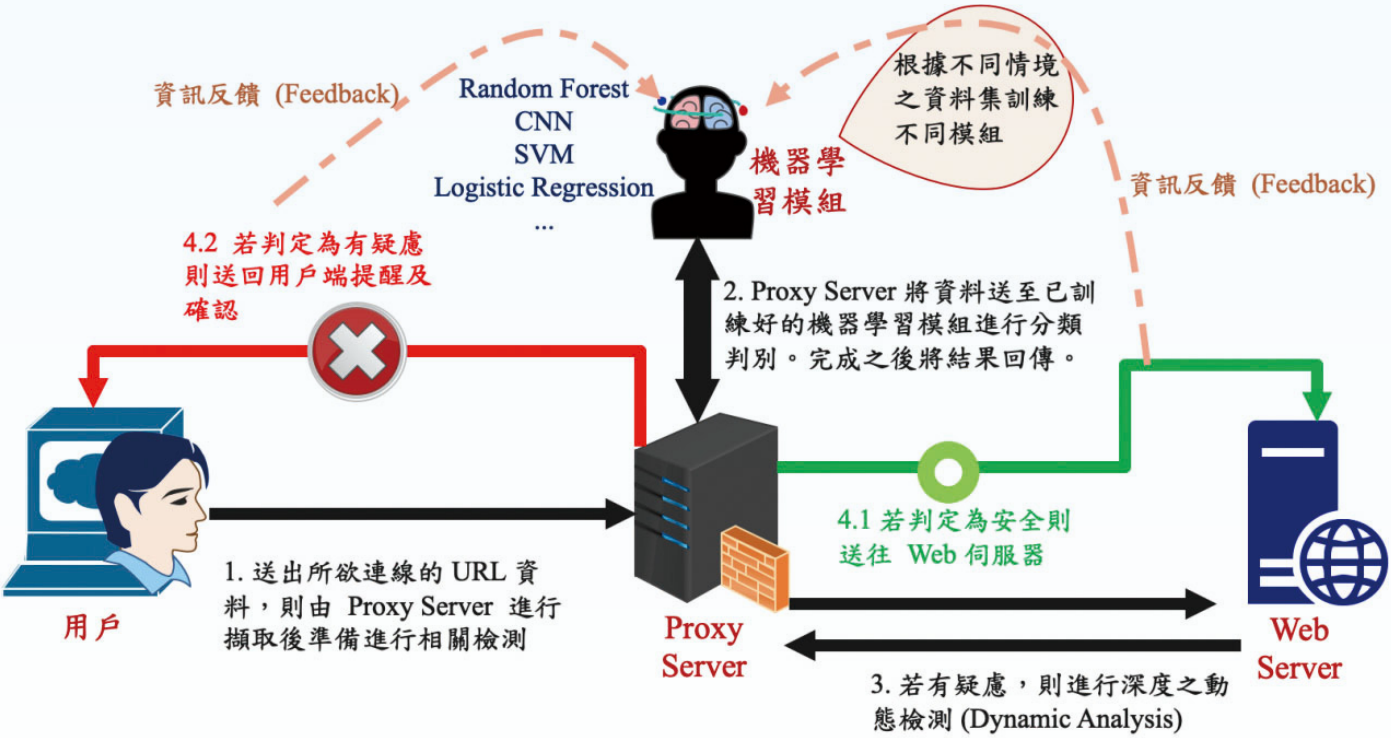


圖 1　整合機器學習的網頁安全性判別模式

[10] 如採用隨機森林（Random Forest）、卷積神經網路（Convolutional Neural Network, CNN）或其他機器學習模式。

表 1　網頁連結之安全性特徵例舉

安全性特徵	舉　例
從 URL 字面上所取得的特徵	URL 的長度、是否使用 IP 位址、是否使用縮短網址、是否有 @ 的符號、URL 中出現‘.’（dot）及‘/’（slash）次數、是否有前綴（Prefix）及後綴（Suffix）、是否使用 https 開頭、是否使用特殊的埠號（Port）等
URL 連接之網頁內容或行為	回傳網頁具有內部或外部連結的數量、是否使用跳出式視窗（Popup Window）、服務表單處理程序 Server Form Handler（SFH）是否為空白或是指向不同網域、是否啟動電子郵件服務傳遞資訊、是否載入大量外部網域之圖片、是否重導向等
網域及網站排名之相關特徵	網域名稱註冊距離現在時間、網站的名聲或排名、網站的流量大小等

路成癮、健康損害以及安全隱私的破壞都是我們所熟知的問題。然而，作為新一代科技人，我們要能夠掌握科技的脈動，要能駕馭科技而不是被科技所支配。當安全問題能夠假人工智慧之手而獲得更好的保障，這將是對抗惡意、詐欺等行為最佳的良藥解方。然而人工智慧也面臨自身系統被攻擊的問題，如最近非常熱門的研究議題—深偽技術（Deepfake)，把深度學習（Deep Learning）與偽造（Fake）結合在一起，這讓依賴人工智慧為安全判斷依據的防衛方法面臨不小的威脅。「道高一尺，魔高一丈」，看來這場網路安全與人工智慧之間的競合勢必還有一大段長路要走。

社團法人台灣 E 化資安分析管理協會（ESAM）

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。

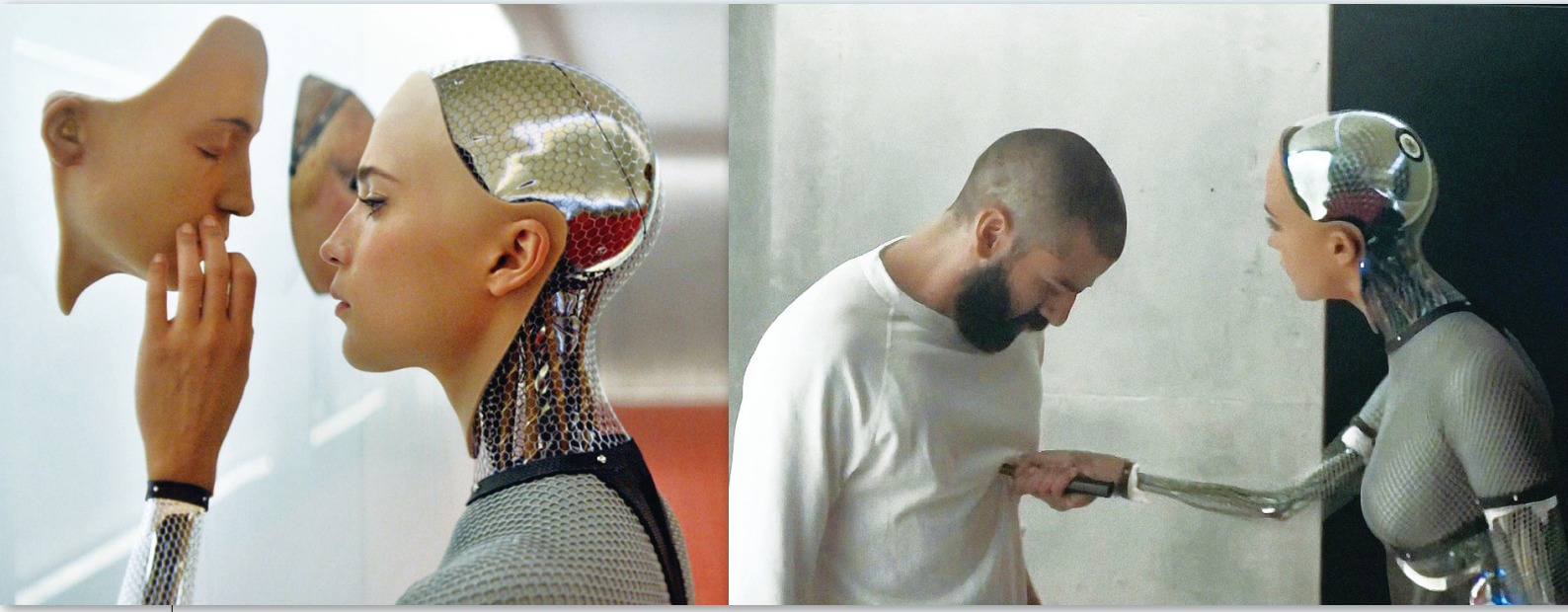
AI 對網路安全造成的威脅

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

[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

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。

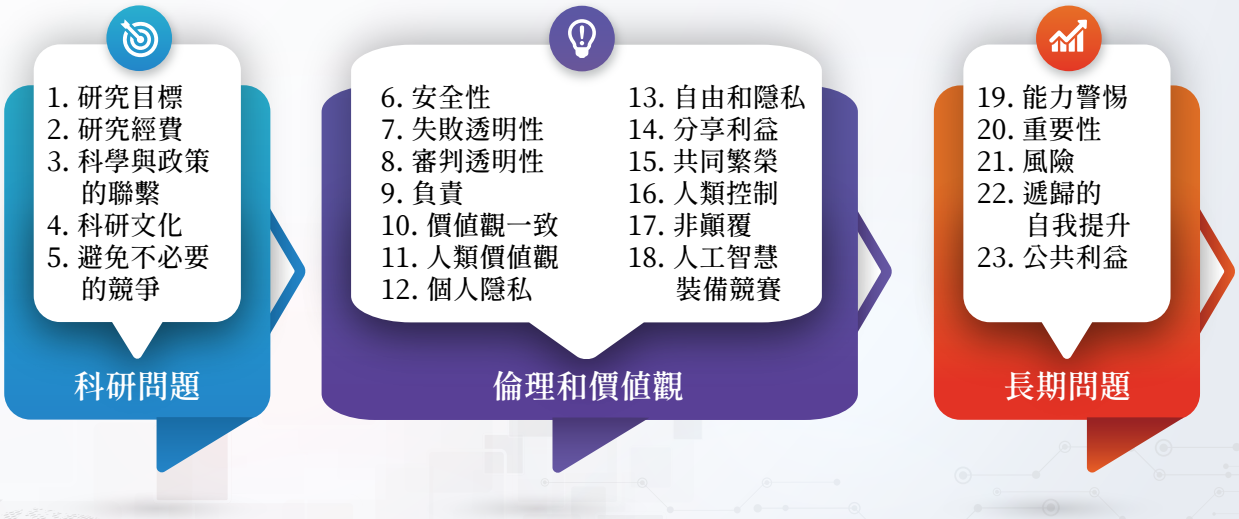


圖 1　阿西洛馬人工智慧 23 條原則內容

[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。