網路科技幽影之困擾--第三部分

2022 年 01 月 07 日
作者：艾迪.華警教授
翻譯：成振昊
戰略情報－區域顧問（亞洲事務總署）

人工智慧(AI)和機器學習(ML)中初顯的安全威脅

近年來，人工智慧（如圖像和語音辨識、智慧型機器人、語言任務和博弈等）取得了巨大的進步，人工智慧技術在不斷的發展和突破，朝著安全有益的方向探索。具有共同的全球利益、客觀目標、時刻關注並勇於挑戰的學者和決策者，通過研究和不斷合作處理短期和長期的技術安全帶來的影響，包括環境治理（如應對氣候變化，以及流行病應對）和生物風險。

人工智慧會發酵短期事件，如隱私、偏見、不平等、安全和保障等，從而影響了全球網路安全、數位化和核武器系統的威脅及發展趨勢。儘管人工智慧和機器人學習從方法策略角度影響了資料驅動型企業，然而安全問題始終是這些公司至關重要的環節。況且，在制定網路安全計畫時，常常需要做出及時調整，且花費巨大。但到了使用人工智慧和機器人學習的系統時，受威脅的規則有所更改。在即將到來的人工智慧時代，政府和行業仍沒有充分準備好控制和標準進行應對。

人工智慧/機器人學習的應用和結構變得更強大、更通用，和其他技術一樣有著類似的應用前景和阻礙。它們可能在許多領域已經優於人類本身-如果人工智慧/機器人學習替代了人類，它可能會在經濟、社會等方面帶來積極的轉變。但如同工業革命一樣，改革也有可能對現有網路行為造成威脅，可能為漏洞威脅、挑戰方面帶來新的場景及災難性的威脅。在直接的軍事對抗下（如間諜活動、心理戰和政治戰爭以及金融工具），國家和非國家對手的遭遇戰進一步加劇，這些對抗利用了我們開放經濟社會的伴隨而來的脆弱點（如基礎設施和社會凝聚力）。

儘管，現有系統經常出現無法預料的錯誤，遇到大量的技術困難和問題，這與反復的設計以及不斷出現的錯誤有關，這會將敵對挑戰（被貼上破壞威脅的標籤）在全球範圍內擴大……進而擾亂勞動力市場、改變長期形成的角色，甚至影響政治主張——先進的人工智慧架構將會成為經濟的關鍵，也是政府資產，它將具有適應性，既能在微觀上更精准，宏觀上更快速……人工智慧也使得網路攻擊和數位假情報更多的被犯罪分子利用，以集中在個人的新模式極大地改變了數位安全的威脅——和/或創造轉基因生物製劑。

隨著人工智慧系統進一步融入現代社會，並成為不可或缺的部分，其受到的攻擊將更具有突發性和系統易感性，進而對國家安全產生重大隱患。在人工智慧和機器學習系統中，有一些關鍵部分需要的複雜資料量更高，在這些資料中，人工智慧可以學習/訓練和推斷加上其模型本身的後續演算法和程式，生成預測、結果和見解——因此，它的技術結構也將提高敵方資訊處理的規模、精度和持久性，這將以三種途徑加劇其破壞性：

信息：人工智慧可以生成基於文本的內容，並操縱圖像、音訊和視頻，包括通過生成對抗網路(generative adversarial network, GAN)及強化學習(reinforcement learning, RL)進行深度偽造，這對區分真實、合法的資訊造成極大阻礙。

受眾：人工智慧可以創造出有著個體特徵、輪廓和大體形象的個人（即生成假角色），假角色具有偏好、行為和信念等，與錨定的特定受眾交互資訊。

媒介：人工智慧可以嵌入到平臺中，如通過位置演算法分類，以擴散破壞性資訊（控制和操縱數位資訊）。

在人工智慧/機器人學習系統擴展的進程中存在很大的模糊性、不適應性和分歧，與傳統網路攻擊中的“漏洞”或代碼中的人為錯誤不同，人工智慧攻擊是由底層演算法的固有限制導致的，目前還無法修復。社會趨向於人工智慧的安全性和建設性發展，以此為目標培養有領導力的研究人員、行業顧問，並從高校實驗室和科技公司中得到更多支持，包括推進先進的研究專案、基金，在機器人學習實驗室裡就國內和國際的戰略風險問題開展會議和討論。

網路攻擊及風險的加速出現

儘管以上只是人工智慧/機器人學習模型場景的一些演示範例，但在實際執行中，駭客可以在不同的單元上複製導致崩潰的程式，精准攻擊系統——大量的曲線和攻擊通過擴增成為可以進行網路攻擊的實體集合，再疊加每一步的機會軌跡，導致人工智慧/機器人學習在開發過程中，從錯誤的活動、有毒演算法和創建保密訓練資料集中學習如何干擾輸入，導致不可預見的損失。

惡意軟體在人工智慧時代能夠轉衍生出成千上萬的不同格式和方法，這些程式一旦載入到電腦系統上，比如多變的的惡意軟體，超過 90%的惡意可執行檔都是由此產生…深度強化學習演算法已經可以找到漏洞和隱藏的惡意軟體，並進行有針對性的攻擊。因此，人工智慧系統成為了一類新的攻擊目標，而在應對這類情況的前沿機構，如政府、商業公司和研究人員已經受到了如逃避、資料中毒、模型複製和利用傳統軟體缺陷開展的欺騙、操縱、損害等攻擊，並使人工智慧系統失效。

相比與傳統網路活動，與之相關但又不同的威脅是由於人工智慧系統的部署將很容易受到來自人工智慧強化領域的對抗性攻擊。公民社會、執法和/或軍事中的傳統的人工是否可以被人工智慧替代…另外，生物技術在科學創新的推動下，目前已經實現可程式設計，如基因編輯工具,它會引領一個新的時代，一個人類可以編輯 DNA、合併龐大的計算能力和人工智慧的時代。生物技術的創新可能會給最令人類文明困惑和無力解決的挑戰提供全新的解決方案，包括健康、食品生產和環境可持續性等，但與此同時，也會帶來極大的負面隱患。

人工智慧對公共安全和智慧社會的影響

目前，科學家和相關從業人員在監管、電腦犯罪、網路技術安全領域的合作已成為共識，為執法和數位社會中出現的突發事件和挑戰提供了全面的視角。人工智慧(AI)的出現，以及物聯網(IoT)設備的廣泛應用，通過推動文化變革形成多樣化產業，創造了相互關聯的智慧社會。隨著大資料應用的在複雜技術、個人監管的創新擴展成為一個全新的大資料應用流、容量中繼資料和平臺。這些都可以助力風險的防範，締造更安全的社區，但在公共安全和安保方面仍在惡意濫用的可能以及潛在的風險。

因此，網路罪犯（駭客）試圖成為電子人，人工智慧也不例外，他們的目標是在最短的時間內獲得更暴力的利潤，利用更多的受害者，創造多樣化的、創新的犯罪商業模式，加快和提高攻擊的成功，同時減少被逮捕的可能。因此，人工智慧維度具有更大的熟練度和自主性，充分展現了人性和技術的融合。政府對各種高科技工具的利用，可以通過已經部署和積極研究的

自動化和增強來提高作戰效率，如目前的全息通訊、智慧城市/智慧社會、面部識別系統、影片監控和搜索技術支援影片和圖像分析調查，偵查罪犯臉部照片的特徵等，阻止進一步犯罪，逮捕網路罪犯等。

所以，科學應用同時帶來複雜的挑戰，用非常有限的人力資源對幾乎無限的內容進行分析過濾，尤其是新一代的人工智慧，可以實現工具和技術的更廣泛配合，提供有效的監管、預防犯罪，更早地發現重大犯罪的信號，在犯罪行為發生的早期快速逮捕罪犯，取得更好的結果。

隨著人工智慧在現代化進程中的的廣泛應用，尤其在個人（模式識別）和軟體（演算法和電腦硬體）中，人工智慧將持續應用於刑事司法系統。這會導致其成為犯罪分子的攻擊目標。不像傳統的網路安全可被人工糾正並處理“漏洞”，以此阻止犯罪分子控制或操縱系統…相比之下,人工智慧的問題是由內部引起的，由此產生的人工智慧的攻擊，無法被“修復”或“修補”，它需要不同的工具和策略來保護核心演算法不被病毒感染。

研究暴露了各種各樣的問題，比如佩戴一副彩色眼鏡可以極大降低了人工智慧識別的準確性，或者通過染髮、語音、手語躲過執法檢測，總之，對目標的攻擊在持續升級。加密貨幣、帳戶劫持、資料盜竊或網路間諜以及恐怖主義等犯罪活動似乎都在持續增長，網路罪犯一直是對政府系統發起複雜攻擊的最新技術的早期採用者，人工智慧也不例外。他們利用其進行內外攻擊，“深度偽造”是目前作為人工智慧對內攻擊最出名的例子。

因此，政府希望通過人工智慧來擴展服務，比如在地方和國家政府網站上使用智慧“聊天機器人”來幫助民眾實現各種功能。執法機構仍然希望利用現有的資料來源輕鬆獲得資料，如利用手機、平板電腦、全球定位系統、無線通訊網路和其他如包含豐富資訊的接入點。從這些連接中產生的所有記錄都通過數位方式收集，執法機構可以更行之有效的進行分析並得出情報，推進複雜的調查工作。

國際警察協會（IPA）亞洲事務總署戰略情報部門記錄分析認為創新技術檢測將降低虛假資訊活動和敲詐勒索的風險，也同樣降低針對人工智慧資料集的嚴重威脅。分析就如何利用人工智慧來提供支持、降低這些威脅提出了建議，具體例子如下：

- 抓取文檔的惡意軟體可以更有效的開展攻擊。
- 勒索軟體攻擊，可以通過智慧目標進行逃避。
- 令人信任的社交軟體受大規模攻擊。
- 資料污染，在檢測規則中識別盲點。
- 逃避圖像識別和語音生物識別。

此外，針對安全的數位未來發展的更多的預測和參考如下:

- 利用人工智慧技術方面的前景作為打擊犯罪的工具，以抵禦未來的網路安全和治安監管。
- 通過不斷的研究、評估、訓練和實戰，促進防禦技術的改進和發展。
- 培育和發展安全的人工智慧設計和戰略架構。
- 全方位保護基礎設施，包括網路加密[協定資訊圖]、代理、防火牆和網路責任保險等，預防和保護未來的網路安全攻擊。
- 收集、交換和傳遞情報，以制定和實施精細化的流程，並深入瞭解人工智慧、機器人和相關技術，如預防犯罪、刑事司法、法治和新出現的安全威脅領域的綜合政策。

- 立法上減少對使用人工智慧解決網路安全問題的過多限制。
- 通過雲端、終端、電子郵件、物聯網和網路等途徑，增強的互聯、分享威脅的智慧解決方案，實現政府對企業和消費者的彈性管理，以獲得更好、更快的保護。
- 撬動公私合作，建立真正的夥伴關係，成立多學科專家小組。

總結

毫無阻礙和不計後果的熱情讓世界吸取了許多痛苦的教訓，科學技術暴露出了被煽動和執行的嚴重漏洞。人工智慧在社會的這些關鍵方面所發揮的不受約束的創造力，將為今後帶來許多亟待解決的威脅。在這些威脅中，政策制定者、當局和關鍵行業必須出臺強制性的安全合規問題，並由適當的監管機構在最佳時間時強制執行，既要避免扼殺創新，又要在這個快速變化的領域中防範這些風險。

然而，在這一挑戰中，公眾面臨著關鍵基礎設施安全、實體及公共部門網路安全的雙重挑戰。缺乏對網路問題的可預見性、無法持續追溯過去，都是最突出的問題。“威脅的停留時間”也在悄無聲息的對監管機構、政府機構和安全公司造成困擾。所以最終的問題是是，他們做得是不是還不夠?

據網路專家統計，每天約有 100 萬次潛在的網路攻擊，隨著移動和雲技術的發展，這個數字還會增加。為了降低增長，政府、執法部門、行業和企業一直在努力拓展其網路安全團隊。然而，為了準確地識別潛在的駭客和/或攻擊，網路安全團隊應該明確誰是網路罪犯、他們使用什麼技術以及將可以採取哪些措施保護和預防未來出現的網路犯罪。

政府和社交媒體公司之間的配合仍然是臨時的。此外，政府需要投入更多的注意力和資源來應對檢測、歸因和媒體的身份驗證等方面。因此，由於犯罪模式、政策和技術正在隨著智慧社會不斷變化，各國和國際執法機構都需要制定面向未來的立法框架;(a)重新審視預防犯罪;(b)調查、決策做出預防性監管;(c)防止或降低具有破壞性的網路攻擊，以及(d)確保安全執行，即：指揮、控制、通信和情報。

隨著這些人工智慧和機器學習科學工具的進化，人類文明的前沿技術可以實現對話交流。網路犯罪分子已經繞過了層層的網路安全控制和防禦，他們的成功之處在於採用各種不同程度的複雜方法或創新，從而在系統使用方面獲得優勢，而這些系統是可以模仿人類行為來執行特定的任務。因此，考慮到潛在的用途，網路犯罪分子會追求高投資及高回報，從而加速攻擊的數量，加強惡意服務，以保護匿名性，躲避執法機構，在這些地方，歸因和調查犯罪已經成為一大挑戰。

在世界範圍內，致力於確保安全和權益的超級智慧的社區正在蓬勃發展，但在先進的網路安全人工智慧系統的發展過程中，存在很大的不確定性和分歧。這種系統可能帶來好的結果，尤其在降低風險和威脅方面，也可能帶來更為複雜的負面影響——通過貫徹嚴格的安全意識，最終在安全問題、挑戰和解決方案上形成特有的管理流程，這需要全面、普遍的合作和以及對網路安全態勢的戰略的重視，以及持續的執行下去。

在國際警察協會亞洲事務總署中，我們延續了以價值觀教育這一傳統，在這裡通過有組織、有紀律及嚴謹的分析，創造有挑戰性、有意義的經驗。
