歐盟人工智慧法案已經通過 未來臺灣在AI資訊安全的影響與發展

總會諮詢顧問 廖建利

如今社會對於AI的應用可謂是毀譽參半。好的是，AI可以作為我們在生產應用上新的觀點與思維。壞的是AI也會被不肖人士拿來作為偽冒詐騙生成的應用。因此，盟理事會主席團與歐洲議會代表在2023年12月9日拍板全球第一個人工智慧協議草案Artificial intelligence(簡稱AIA)。該協議的目的是為了確保歐盟市場上AI系統的安全與可靠性。同時遵守基本人權和歐盟的價值。本期就讓我們初步的來分享人工智慧法案的重點精神，以及對於資訊安全的影響與發展。

人工智慧法案的立法的初衷是為了避免人工智慧過度的擴張，造成對於社會與經濟產生的危害，因此在人工智慧系統的定義上，採行以基於風險高低的方式來個別監管人工智慧系統。因此，除非你是特別用途，像是軍事國防或者研究領域，否則你所開發或者應用的AI系統風險越高，限制的規則將會越嚴格。簡單說，歐盟將AI的風險分為4個等級，分別是不可接受的風險、高風險、有限風險，以及低風險或無風險。如果你是使用AI生成的內容，像是ChatGPT、Midjoury、Co-Pilot這種低風險的AI應用，在對外發佈時，必須宣告該內容是透過AI生成的，讓接收者可以進一步判斷內容的可信度與否。如果你是使用高風

險的AI應用，如智慧法務系統、公共設施監控、自動駕駛、AI金融評估、AI醫療診斷，這種類型則必須在歐盟的境內。取得授權，遵守其歐盟的法規及義務才可以使用。

其他像是透過AI來進行認知作戰，或是網軍活動、社交評分、宗教或政治判斷等無特定目的，這種侵犯到個人隱私以及社會安定的應用的話，則是會被禁止的。具體風險的定義呢？至於罰款的部分，如果有企業違反AI法案的話，則會依照上一年的年報的營收的百分比來進行處罰。

使用被禁止的AI應用，則會罰款12億臺幣，或者7%的公司營收。違反人工智慧法案的義務，則會罰款5億臺幣或者3%的公司營收。因此！如果你的公司設立在歐盟區域且要發展高風險的AI應用的話，則必須先在歐盟設立的AI辦公室進行註冊，透過機關內的人工智慧監管沙箱(沙盒（英語： sandbox，又譯為沙箱）是一種安全機制，為執行中的程式提供隔離環境)來進行測試使用。

AI的發展及應用已經是顯而易見的事實，這次歐盟AI法案通過所傳達出來的背後的含義，除了讓AI的應用得到正向的約束外，也傳達了目前AI性能度的脆弱與隱憂。知名IT研究及顧問公司Gartner在2024年的科技趨勢中再次提到了TRISM這個概念。AI Trust、Risk and the Security Management說明了AI的發展上應該要具備3點特質。

第一要具備透明度以及可解釋性，避免當遇到與人命或者安全有關的判斷時，可以做出合理的判斷解讀。

第二要有Model-Ops的思維，在開發AI應用時要嚴格管理AI的模型開發以及其中的過程階段，並且檢查發布過程的所有部分，以確保完整性和合規性。

第三就是AI模型的資訊安全管理，確保在機器學習的模型操作的每一個階段，都要做到資訊安全。避免AI的數據遭到惡意的提取誘導或者感染等狀況。

試想，如果未來人人都使用AI，那麼評斷AI的好壞標準就在於AI本身是否可被信任了，因此在採用或者開發AI應用時，我們應該要朝向零信任的思維來下手，可以考慮與專業的資安廠商合作，導入合適的AI資安解決方案。

而對於員工在數位內容的認知判斷，也可以透過專業的資安服務廠商進行數位內容的安全意識培訓，讓員工掌握多方求證的方法與思維來確認數據的真假。市面上許多的知名的科技應用其實都有AI技術的影子，像是知名的新冠疫苗莫德納。就是透過AI應用進行基因序列的模擬，讓原本需要耗時10年的疫苗研發縮短為一年，就可以成功量產，在疫情期間做出了重大貢獻。

我們需要AI來幫助我們，我們更需要可信任且具安全風險的AI來穩固我們的發展。雖然這份AI的臨時協議的法規呢，目前並不適用於歐盟法律範圍之外的國家，但是有了這份法案作為基底。各國呢也會開始起草。對於AI相關的法案，達到相互監管制衡的作用，讓未來的社會與科技更加美好完善。