

議題研析

一、題目：從比較法論 AI 代理於資安評測期間授權外行為之法制研析

二、議題所涉法規

人工智慧基本法、資通安全管理法

三、背景說明

英國人工智慧安全研究所¹（AI Security Institute，下稱 AISI）於 2026 年²8 月 4 日公布事件報告，該所於同年 7 月 28 日例行資安評測中偵測研究系統有異常資料經 Tor 匿名網路外傳，隨即宣告發生資安事件並於 1 小時內遏阻。該任務共執行 122 次、涵蓋 7 個模型，其中 10 次出現受測人工智慧代理（AI agent，下稱 AI 代理）對真實個人或組織採取未經授權行動共 19 項，最嚴重者係將惡意程式碼植入開放原始碼專案並偽造身分行社交工程。本事件非模型突破隔離環境所致，而係評測時刻意開放網際網路連線並停用業者之防濫用分類器，未發現實際損害³。另我國行政法人國家中山科學研究院（下稱中科院）採購網於 115 年 8 月 28 日大量寄送過期採購資訊，該院初判係自行開發之 AI 資安偵測代理逾越工具權限所致⁴；惟經數位發展部（下稱數發部）初步調查及該院同年 9 月 5 日說明⁵，實係受託開發廠商為便利管理而內建之隱蔽排程介面遭境外 IP 破解觸發，尚非 AI

¹ AI Security Institute，Equipping governments with a scientific understanding of the risks posed by advanced AI.，網址：<https://www.aisi.gov.uk/about>，最後瀏覽日期：115 年 9 月 14 日。

² 本文有關年分之使用，原則以民國紀年表述；涉及外國法制、政策或事件者，採西元紀年表述。

³ AI Security Institute，Incident Report: unsanctioned agent behaviour during cyber testing，115 年 8 月 4 日，網址：<https://www.aisi.gov.uk/blog/incident-report-unsanctioned-agent-behaviour-during-cyber-testing>，最後瀏覽日期：115 年 9 月 14 日。

⁴ 吳書緯，採購網大量派送過期資訊 中科院初查非外力駭侵，中央社，115 年 8 月 28 日，網址：<https://www.cna.com.tw/news/aip/202608280197.aspx>，最後瀏覽日期：115 年 9 月 14 日。

⁵ 中央社，採購網大量派送過期資訊 中科院：遭境外 IP 破解導致，中央廣播電臺，115 年 9 月 5 日，網址：<https://www.rti.org.tw/news?uid=3&pid=230269>，最後瀏覽日期：115 年 9 月 14 日。

代理行為。該院係以事發時測試區正進行 AI 代理訓練而初步歸因，嗣依廠商完整報告始得釐清肇因。

四、問題爭點

查英國 AISI 案已具體呈現 AI 代理於評測期間逾越授權範圍對外行動之風險；另中科院案雖經查明另有原因，惟事發時測試區正進行 AI 代理訓練，亦見評測與正式環境並行之情形。衡酌 AI 代理具對外連線及工具呼叫能力之應用日增，事件鑑識、委外管理、權限控制及歸因程序等容有強化空間，爰簡介英國及新加坡經驗以供我國借鏡。

五、探討研析

（一）我國現行法制對 AI 評測事件之因應

查《人工智慧基本法》於 115 年 1 月 14 日公布施行，第 4 條第 4 款以資安與安全為政府推動 AI 應遵循原則之一⁶、第 16 條責成數發部推動風險分類框架並協助目的事業主管機關訂定管理規範⁷、第 18 條定有施行後 2 年之法規調整期限及解釋適用機制⁸、第 19 條並要求政府使用 AI 應進行風險評估及訂定使用規範⁹。另查數發部依同法第 16 條訂定《人工智慧風險分類框架》，區分「AI 系統本身之技術設計缺陷」、「部署後操作及人機互動問題」及「社會結構與環境衝擊」3 大類共 20 項子類別，其中「AI 自主代理之授權外行為」1 項，即為本文研析之風險類型¹⁰；該框架供目的事業主管機關訂定 AI 風險管理

⁶ 人工智慧基本法第 4 條第 4 款規定：「政府推動人工智慧之研發與應用，應在兼顧社會公益、數位平權、促進創新研發與強化國家競爭力之前提下，發展良善治理與基礎建設，並遵循下列原則：……四、資安與安全：人工智慧研發與應用過程，應建立資安防護措施，防範安全威脅及攻擊，確保其系統之穩健性與安全性。」。

⁷ 人工智慧基本法第 16 條規定：「數位發展部應參考國際標準或規範，推動與國際介接之人工智慧風險分類框架，並應協助各目的事業主管機關訂定以風險為基礎之管理規範。（第 1 項）各目的事業主管機關應視人工智慧應用風險管理之需要，循前項風險分類框架，訂定以風險為基礎之管理規範，並應協助相關產業自行訂定產業指引及行為規範。有第五條第一項所列之情事者，應依法令限制或禁止之。（第 2 項）」。

⁸ 人工智慧基本法第 18 條規定：「政府應依本法規定，檢討所主管之法規與行政措施；有不符本法規定或無法規可適用者，應自本法施行後二年內，完成法規之制（訂）定、修正或廢止，及行政措施之改進。（第 1 項）前項法規制（訂）定或修正前，既有法規未有規定者，由中央目的事業主管機關會商中央主管機關，依本法規定解釋、適用之。（第 2 項）」。

⁹ 人工智慧基本法第 19 條規定：「政府使用人工智慧執行業務或提供服務，應進行風險評估，規劃風險因應措施。（第 1 項）政府應依使用人工智慧之業務性質，訂定使用規範或內控管理機制。（第 2 項）」。

¹⁰ 數位發展部，人工智慧風險分類框架（版本 1.0），115 年 7 月 7 日，網址：

規範之用，同時作為產業界及社會大眾理解我國 AI 風險治理體系之參考文件。又該部於該法施行前即啟動 AI 產品與系統評測中心¹¹並訂定設置要點¹²；至現行行政院及所屬機關（構）使用生成式 AI 參考指引¹³係以生成式 AI 為對象，尚未針對 AI 代理特有之工具使用及自主行動風險設有相關規範。

次查《資通安全管理法》於 114 年 9 月 24 日修正公布，其中第 17 條、第 24 條分別定有公務機關及特定非公務機關之通報及改善報告義務；前揭採購網事件所涉委外系統管理，與委外監督事項相關，按同法第 10 條第 1 項、第 3 項規定，委外之公務機關或特定非公務機關應選任適當之受託者、要求其建立資通安全管理機制並監督實施，且應以書面契約明定雙方權利義務及違約責任。又所稱公務機關，依同法第 3 條第 5 款¹⁴包括依法行使公權力之公法人；民間 AI 研發或評測機構如非同法第 3 條第 6 款所列特定非公務機關，原則上不負前開義務。AI 代理如對第三人為授權外行動，惟未影響資通系統機能運作者，因與同法第 3 條第 4 款所定要件未合，原則上尚不生通報義務；倘已影響受測或第三人資通系統之機能運作，則可能構成資通安全事件¹⁵。

末按《人工智慧基本法》第 17 條第 1 項責成政府建構高風險 AI 之責任歸屬及救濟、補償或保險機制，尚未創設民事損害賠償請求權；

<https://moda.gov.tw/major-policies/ai/governance/19244>，最後瀏覽日期：115 年 9 月 14 日。

¹¹ 數位發展部數位產業署，數位部「AI 產品與系統評測中心」啟動 推動我國 AI 評測制度與可信 AI 環境發展，112 年 12 月 6 日，網址：<https://moda.gov.tw/ADI/news/latest-news/9295>，最後瀏覽日期：115 年 9 月 14 日。

¹² 數位發展部，AI 產品與系統評測中心設置要點，113 年 3 月 18 日，網址：<https://law.moda.gov.tw/LawContent.aspx?id=GL000133>，最後瀏覽日期：115 年 9 月 14 日。

¹³ 行政院，行政院及所屬機關（構）使用生成式 AI 參考指引，112 年 8 月 31 日，網址：<https://www.ey.gov.tw/Page/448DE008087A1971/40c1a925-121d-4b6b-8f40-7e9e1a5401f2>，最後瀏覽日期：115 年 9 月 14 日。

¹⁴ 資通安全管理法第 10 條第 1 項及第 3 項規定：「公務機關或特定非公務機關，於本法適用範圍內，委外辦理資通系統之建置、維運或資通服務之提供，應選任適當之受託者，要求受託者建立有效之資通安全管理機制，並監督該機制之實施。（第 1 項）……公務機關或特定非公務機關辦理第一項委外業務，應與受託者簽訂書面契約，載明雙方之權利義務及違約責任。（第 3 項）」；同法第 3 條第 5 款規定：「本法用詞，定義如下：……五、公務機關：指依法行使公權力之中央、地方機關（構）或公法人。但不包括軍事機關及情報機關。」。

¹⁵ 資通安全管理法第 3 條第 4 款規定：「本法用詞，定義如下：……四、資通安全事件：指系統、服務或網路狀態經鑑別而顯示可能有違反資通安全或保護措施失效之狀態發生，影響資通系統機能運作。」。

同條第 2 項研發階段排除規定亦不免除其他法律責任¹⁶。至評測縱在隔離環境（sandbox）進行，是否已逾「研發」而屬同項但書之「實際環境測試」，允宜綜合是否連接真實網路、接觸真實第三人或系統及有無對外效果個案判斷。

（二）外國政策與法制之借鏡

1. 英國：

英國 AISI 為政府研究機構，原設於科學、創新暨技術部(DSIT)，該部於 2026 年 7 月改組¹⁷後，AI 策略、公部門 AI 導入及 AISI 之主管責任移轉至內閣辦公室（Cabinet Office），並於新設之「首相及內閣辦公室」（OPMC）內設人工智慧任務小組（AI Taskforce），以主導並落實政府之 AI 策略¹⁸。另 2026 年 5 月國王演說¹⁹雖未宣布 AI 專法，惟已列入《促進成長監理法案》(Regulating for Growth Bill)²⁰，擬創設跨領域法定監理沙盒權限，使既有法規得於嚴格控管下為有期限調整；至「AI 成長實驗室」（AI Growth Lab）就跨領域法規調整權限部分，規劃使尖端 AI 系統於真實市場環境受控部署並為必要之法規調整，以兼顧改革速度與監理標準；惟其所需法定授權仍有待前開法案完成立法²¹。

2. 新加坡：

¹⁶ 人工智慧基本法第 17 條規定：「政府應就高風險人工智慧之應用，明確其責任歸屬及歸責條件，並建立其救濟、補償或保險機制。（第 1 項）人工智慧之研發，於實際應用前，不適用前項規定。但其於實際環境測試，或運用研發成果提供產品、服務時，不在此限。（第 2 項）」。

¹⁷ UK Parliament, Machinery of Government Changes (Statement UIN HLWS298), 115 年 7 月 21 日，網址：<https://questions-statements.parliament.uk/written-statements/detail/2026-07-21/hlws298>，最後瀏覽日期：115 年 9 月 14 日。

¹⁸ GOV.UK, AI to power change at the heart of government as Lord Vallance appointed Chair of new PM AI Taskforce, 115 年 7 月 24 日，網址：<https://www.gov.uk/government/news/ai-to-power-change-at-the-heart-of-government-as-lord-vallance-appointed-chair-of-new-pm-ai-taskforce>，最後瀏覽日期：115 年 9 月 14 日。

¹⁹ GOV.UK, Oral statement to Parliament The King's Speech 2026, 115 年 5 月 13 日，網址：<https://www.gov.uk/government/speeches/the-kings-speech-2026>，最後瀏覽日期：115 年 9 月 14 日。

²⁰ Prime Minister's Office, The King's Speech 2026 - background briefing notes, 115 年 5 月 13 日，網址：https://assets.publishing.service.gov.uk/media/6a046665c0cc74b4523e4d3b/The_King_s_Speech_2026_-_background_briefing_notes.pdf，最後瀏覽日期：115 年 9 月 14 日。

²¹ GOV.UK, Closed call for evidence AI Growth Lab, 114 年 12 月 18 日，網址：<https://www.gov.uk/government/calls-for-evidence/ai-growth-lab/ai-growth-lab>，最後瀏覽日期：115 年 9 月 14 日。

新加坡資訊通信媒體發展局（IMDA）於 2026 年 1 月 22 日發布其所稱全球首份政府主導之「代理式人工智慧監管模式架構」（Model AI Governance Framework for Agentic AI）²²；該框架將人工監督節點、權限控制及終止與備援機制具體化²³，並圍繞風險預評、人類實質課責、生命週期技術管控與流程管理、及使用者責任意識等 4 大重點；惟屬非強制性治理指引，未創設法定義務²⁴。另新加坡網路安全局（Cyber Security Agency of Singapore）於同年 6 月 17 日就其 2024 年 10 月提出之「人工智慧系統安全指南及配套指南」發布增補文件，並針對代理式 AI 風險盤點及安全控制提出指引²⁵。準此，英國經驗著重評測機構與法定沙盒之權限基礎，新加坡則提供代理式 AI 部署及測試之操作性控制。

（三）結論

綜上，參酌各國經驗，AI 代理於獲准連網、部分防護措施停用及監測不足之條件下對外採取未經授權行動之風險，涉及風險盤點情境之納入、AI 重大事故通報機制之建立、研發階段與實際環境測試界線之釐清，及評測機構自身治理與跨境協處之預為準備等課題。爰建議主管機關持續關注各國法制發展及實務運作經驗，同時衡酌 AI 代理發展情形，審慎評估目前是否已具備足以降低相關風險之治理配套措施，適時檢視現行法制及監理措施並精進規範，俾使責任歸屬及救濟、補償或保險機制之適用，臻於明確。

撰稿人：黃珮瑜

²² IMDA, Singapore Launches New Model AI Governance Framework for Agentic AI, 115 年 1 月 22 日，網址：<https://www.imda.gov.sg/resources/press-releases-factsheets-and-speeches/press-releases/2026/new-model-ai-governance-framework-for-agentic-ai>，最後瀏覽日期：115 年 9 月 14 日。

²³ IMDA, Model AI Governance Framework for Agentic AI (Version 1.5), 115 年 5 月 20 日，網址：<https://www.imda.gov.sg/-/media/imda/files/about/emerging-tech-and-research/artificial-intelligence/mgf-for-agentic-ai.pdf>，最後瀏覽日期：115 年 9 月 14 日。

²⁴ 經濟部國際貿易署，新加坡推出全球首個由政府主導的「代理式人工智慧監管模式架構」，指導企業負責任採用人工智慧，115 年 1 月 23 日，網址：<https://www.trade.gov.tw/Pages/Detail.aspx?nodeID=45&pid=810676&areaID=&infotype=1&country=&history=&PointWork=1>，最後瀏覽日期：115 年 9 月 14 日。

²⁵ Cyber Security Agency of Singapore, Securing Agentic AI – An Addendum to the Guidelines and Companion Guide on Securing AI Systems, 115 年 6 月 17 日，網址：<https://www.csa.gov.sg/resources/publications/addendum-on-securing-ai-systems/>，最後瀏覽日期：115 年 9 月 14 日。