台灣可信任與韌性 AI 白皮書

Trustworthy and Resilient AI in Taiwan

目錄

前言

AI 創新大浪來襲，前瞻架構與全面防禦思維更形重要

Cisco 全球資深副總裁暨創新長

Guy Diedrich

攻擊手法因 AI 而加速演化，傳統防禦機制逐漸失效。網路基礎建設的安全性早已超越單純的技術議題。

真正決定價值的基礎並非資料，而是信任（Trust）。如果資料隱私無法受到保障，缺乏透明度與安全性，使用者便無法建立信賴關係，也不會願意分享資料。唯有透過隱私保護與開誠布公的資料治理機制，才能推動產業與社會的持續發展。

前言

AI 創新大浪來襲，前瞻架構與全面防禦思維更形重要

隨著生成式 AI 的快速崛起，全球數位基礎建設正面臨前所未有的壓力。「需求暴增、供應不足」已成為新的常態，不僅快速增加運算與網路資源的負荷，也迫使現有系統面臨結構性挑戰。與此同時，攻擊手法正因 AI 技術而加速演化，自動化與智慧化的威脅層出不窮，使傳統防禦機制逐漸失效。

在此背景下，網路基礎建設的安全性早已超越單純的技術議題，它不僅關乎企業能否維持營運韌性，更直接影響產業競爭力與整體社會的信任基礎。唯有透過更具前瞻性的架構設計與全面性的防禦思維，才能確保數位世界在 AI 世代中持續穩健發展。

思科 (Cisco) 全球創新長 Guy Diedrich 指出，Cisco 自 1984 年創立以來，始終是全球網路基礎建設的關鍵推動者。即使在人工智慧（AI）快速崛起的時代，仍居於不可或缺的角色。今日，全球超過 80%的網路流量經由 Cisco 設備傳輸，每天更在全球阻擋超過 5,000 億次攻擊。這些數字顯示，Cisco 不僅是連結數位世界的骨幹，也是全球數位安全防線的守護者。

AI 世代下的網路基礎建設與安全挑戰

然而，AI 的普及正在徹底改變網路環境。Guy Diedrich 進一步解釋，ChatGPT 的問世，不僅讓大眾意識到 AI 的可用性，也帶動了網路與運算需求的急遽攀升，並持續加壓於現有基礎設施。全球基礎建設正在多個層面出現瓶頸：其一，能源與電網架構尚未能支撐 AI 帶來的龐大能耗；其二，網路容量無法完全滿足持續增加的資料流量與應用需求；其三，跨平台與跨產業的系統互通性不足，限制了協作與創新。這些缺口使得全球基礎建設在 AI 世代中顯得力有未逮。

但他強調，對 Cisco 而言，AI 並非全新的課題。早在十餘年前，Cisco 便已將 AI 技術深度融入網路基礎設計，持續優化流量調度、安全防護與智能化管理，也因此能以前瞻性布局來應對當前爆炸式的需求成長。未來，Cisco 將聚焦於網路韌性、安全架構與跨系統互通三大核心面向，協助全球產業在 AI 世代中建立穩健而安全的數位骨幹，並透過持續升級與創新，在新一波數位浪潮中仍居關鍵地位。

台灣的關鍵角色與 Cisco 的在地承諾

Guy Diedrich 提到，Cisco 在台灣深耕達 28 年，長期與在地企業緊密合作，共同推動創新與策略聯盟。過去七年中，Cisco 推出「台灣數位加速計畫」（Taiwan Digital Acceleration, TDA），並透過三年一循環的方式持續投入關鍵領域，包括醫療照護、基礎設施、教育以及永續綠能等。迄今，TDA 已推動超過 30 項專案，並進入第三輪循環（TDA 3.0），未來也將持續深化，以擴大其成果與影響力。

他認為，台灣在人均教育程度上位居世界前列，加上疫情期間全球深刻意識到網路基礎建設的重要性，凸顯了台灣在數位生態中的關鍵地位。特別是台積電（TSMC）等企業，更被視為推動全球科技發展的核心寶藏。如果台灣持續以創新能量與研發成果貢獻於 AI 生態系，其影響力將無可取代。

信任才是決定價值的基礎

隨著數位轉型與 AI 發展，資料被喻為新世代的「貨幣」。然而，Guy Diedrich 認為，真正決定價值的基礎並非資料，而是信任（Trust）。因為如果資料隱私無法受到保障，缺乏透明度與安全性，使用者便無法建立信賴關係，也不會願意分享資料。唯有透過隱私保護與開誠布公的資料治理機制，才能推動產業與社會的持續發展。

而 Cisco 在與全球五十多國政府合作的過程中，深刻理解各國在資料主權（Data Sovereignty） 上的不同需求，例如資料必須完全在地化、區域性駐地，或跨境使用的限制。憑藉完整的解決方案，Cisco 能協助各國因應多樣化的法規要求，並在保障隱私的前提下，確保資料的安全與透明度。

Guy Diedrich 強調，Cisco 是第一家實現軟硬體整合，並以 AI 驅動網路全域管理的公司，能夠從端到端監控與保護 80% 的全球網路流量，及時辨識與應對潛在威脅。更與 GPU 廠商（如 NVIDIA）建立緊密合作，打造完整的技術堆疊，從晶片（如 Cisco 自研的 Silicon One 可編程晶片）到軟體層級，形成無縫的 AI 架構。這種軟硬體一體化的設計，不僅確保了高水準的安全性，也為主權 AI 提供了可信任的保障。

台灣的創新能量將是 AI 持續前進的基礎

回顧過往，從工業時代、資訊時代到數位時代，如今我們已經進入 AI 的時刻，AI 不僅快速滲透到生活周遭的各個層面，更逐漸成為普及化的「商品化技術」。

Guy Diedrich 認為，摩爾定律的節奏已延伸到 AI 產業，技術進步的速度將不斷加快。每 18 至 24 個月，便會有舊技術便被更新、更強的技術所取代。在未來，AI 將與量子運算並存，並在短時間內共同構築全新的科技基礎，「這代表創新不再是漸進式演化，而是一波接一波的巨浪。」

在這股浪潮中，Cisco 持續強化自身創新動能。過去 25 年間，透過併購 240 家企業（平均幾乎每月一家）不斷擴展技術版圖，每年更是投入超過 60 億美元於研發，以確保始終立於技術前沿。這些投入不僅是企業策略，更是 Cisco 永續領先的基礎。

Guy Diedrich 提到，由於創新速度往往超越人才培育的步伐，因此 Cisco 在全球設立創新辦公室，透過教育訓練、產學合作與夥伴生態，協助人才加速技能提升。目前，Cisco 已與全球超過 35,000 家合作夥伴共同推動這項任務，確保創新速度與人力能力同步發展。

從全球視角來看，Guy Diedrich 認為，台灣兼具高水準教育與科技驅動的經濟結構，展現獨特優勢。近三年來，Cisco 高層頻繁造訪台灣，深刻感受到新創文化的蓬勃能量。除了廣為人知的台積電，台灣還有眾多靈活的小型創新企業，雖未必專注於晶片製造，卻展現出積極、熱烈且勇於冒險的創新氛圍。

這種不接受停滯、勇於挑戰並擁抱機會的文化，正與台灣科技驅動的經濟體質相互呼應，成為推動台灣在 AI 與新世代技術持續前進的重要基礎。

國際環境與趨勢

- 新興人工智慧風險解析
- 全球人工智慧政策與治理框架

新興人工智慧風險解析
AI 應用的雙面刃：六大風險與挑戰

人工智慧無疑是當代最具顛覆性的技術浪潮，正以空前的速度重塑產業格局、優化商業流程，並持續開創新的價值來源。然而，在龐大效益與創新契機之外，AI 的應用同時伴隨著多層面的風險與挑戰。這些挑戰不僅來自技術本身，更深刻牽動組織文化、法規制度、經濟結構與社會倫理。綜合本次受訪者觀點與相關研究資料，可以歸納出六大主要挑戰。唯有全面理解並積極回應這些挑戰，企業與社會才能在擁抱 AI 的同時，維繫信任基礎，並強化長遠的韌性。

1. 資安威脅與風險加劇

智慧化、自動化的攻擊模式

AI 技術的快速發展使得網路攻擊的數量和複雜性不斷增加，借助 AI 工具，駭客能夠自動化地進行偵察、漏洞挖掘、密碼破解，甚至能生成高度逼真的釣魚郵件或深度偽造（Deepfake）內容，用於社交工程攻擊。這些 AI 驅動的攻擊不僅速度更快，且具備學習與適應能力，使得傳統基於簽章或規則的防禦機制捉襟見肘。

新的攻擊向量與擴大的攻擊面

AI 系統本身構成了一個全新、廣闊的攻擊面。從底層的 AI 框架、演算法模型、訓練資料集，到上層的應用程式（App）與基礎設施，每一個環節都可能成為駭客的突破口。

針對大型語言模型（LLM）的特有風險

大型語言模型的普及，催生了全新的資安漏洞。開放網路軟體安全計畫（OWASP）針對此發布了十大 LLM 應用程式風險，其中包括（如下圖）：

OWASP 大型語言模型及生成式 AI 十大風險（2025）：

1. 提示詞注入漏洞（Prompt Injection）
2. 敏感資訊洩漏（Sensitive Information Disclosure）
3. 供應鏈風險（Supply Chain）
4. 資料與模型投毒（Data and Model Poisoning）
5. 不當輸出處理（Improper Output Handling）
6. 過度代理授權（Excessive Agency）
7. 系統提示洩露（System Prompt Leakage）
8. 向量與嵌入弱點（Vector and Embedding Weaknesses）
9. 錯誤資訊（Misinformation）
10. 無限制消耗（Unbounded Consumption）

內部威脅：「影子 AI」（Shadow AI）的治理困境

當員工未經授權，在工作中使用公開或第三方的 AI 服務（如 ChatGPT）處理公司內部資料，由於這種行為極難被 IT 部門有效監管，不僅大幅增加了敏感資訊外洩的風險，也使企業暴露在未知的合規與法律風險之中。

臺灣企業的警訊：準備度與現實的巨大落差

儘管威脅迫在眉睫，臺灣企業的準備狀況卻令人擔憂。根據《2025 年全球網路資訊調查》，高達 79% 的臺灣企業曾因網路攻擊或配置錯誤而導致重大業務中斷。同時，《資安準備度報告》更揭示了嚴峻的現實：雖然有 93% 的企業正嘗試運用 AI 強化資安防禦，但僅有區區 4% 的企業自認其防禦能力達到成熟水準。這組數據鮮明地反映出，絕大多數企業嚴重低估了 AI 時代資安威脅的複雜性。更值得注意的是，高達 92% 的企業在過去一年中曾遭遇與 AI 相關的資安事件，但僅有 39% 的員工能夠充分理解這些威脅。這凸顯了從高層認知到基層執行之間存在著巨大的技能與意識鴻溝。

2. 資料的挑戰：信任基石或風險根源

資料是 AI 的命脈，AI 模型的效能、準確性與可靠性完全取決於其賴以訓練的資料。然而，資料本身也正是 AI 風險最主要的來源之一。

資料來源、品質與偏見：AI 的信任度始於對資料來源的信任。若訓練資料來源不安全、不完整、過時或充滿偏見（例如，反映歷史上的性別或種族歧視），AI 模型將忠實地學習並放大這些缺陷，最終產出錯誤、不公平甚至有害的內容。

敏感資料的暴露風險：在處理包含個人隱私、商業機密等敏感資訊的大型資料集時，存在極高的風險。即使資料經過匿名化處理，AI 模型仍可能透過關聯分析，意外地「再識別化」或推斷出個體身份與敏感資訊。

資料主權與控制權的喪失：當企業將資料上傳至公有雲平台以利用其 AI 服務時，關於資料主權的議題便浮上檯面。企業必須審慎評估，服務供應商對資料的存取權限、儲存地點以及安全防護措施，是否能讓企業保有對其最寶貴資產的完全掌控。

「AI-Ready」資料的匱乏：許多企業雖然坐擁大量數據，但這些數據往往是孤立、非結構化且非即時的。將這些原始數據轉化為可用於即時分析與決策的「AI-Ready」資料，本身就是一項浩大且成本高昂的工程。

3. 倫理、公平與可解釋性的三重困境

AI 的決策影響力日益深遠，其倫理邊界與社會責任也成為各界關注焦點。若無法解決以下三重困境，AI 的社會接受度與長期發展將面臨嚴峻考驗。

困境一：偏見與歧視的放大器

如前所述，源於訓練資料的偏見，會導致 AI 系統在招聘、信貸審批、司法判決等關鍵領域做出歧視性決策，加劇社會不公。同時，生成式 AI 也可能被用於製造仇恨言論、霸凌內容或虛假資訊。

困境二：「黑盒子」問題與可解釋性的缺乏

當前許多 AI 模型，特別是深度學習與大型語言模型，內部決策過程極複雜，以致開發者也難以解釋為何模型會做出某個特定決策。這種「黑盒子」特性，在金融、醫療、法律等需高度問責的領域，構成巨大挑戰。當 AI 出錯時，如無法解釋原因，就很難進行修正、釐清責任。

困境三：責任歸屬的模糊地帶

當自動駕駛汽車發生事故，或 AI 客服提供了錯誤的資訊導致公司損失時，責任應由誰承擔？是 AI 開發者、模型部署者、資料提供者，還是使用者？加拿大曾發生航空公司因 AI 聊天機器人向客戶承諾了不存在的折扣，最終法院判決公司必須履行的案例，這為 AI 應用的法律責任歸屬敲響了警鐘。

過度自主與人類監督的失效，賦予 AI 過高的自主權，可能導致其在未經授權或超出預期的情況下採取行動。雖然「人在迴路中」（Human-in-the-Loop）的監督機制被視為解方，但若監督者過於信賴 AI 的建議，可能僅是流於形式，淪為「橡皮圖章」，失去了實質的監督意義。

4. 技術實施與營運的現實障礙

即便企業對 AI 的潛力滿懷期待，但在將其從概念驗證（PoC）推向規模化部署的過程中，仍面臨著重重的技術與營運障礙。

技術迭代過快：AI 領域的技術與模型正以驚人的速度演進，一個長達數月的概念驗證專案，很可能在完成之時，其所依據的技術就已經過時，導致投資效益大打折扣。

嚴苛的基礎設施要求：AI 工作負載，特別是模型訓練，需要龐大的 GPU 算力支援，這意味著高昂的硬體成本、巨大的電力消耗和對散熱系統的嚴峻考驗。傳統的資料中心在架構和能源供應上，往往難以負荷大規模 AI 運算的需求。

混合雲管理的複雜性：為了兼顧資料安全與運算彈性，企業常採用混合雲架構。然而，在橫跨地端私有雲與公有雲的環境中，如何統一管理 AI 工作負載、確保資料流動的順暢與安全，是一項極為複雜的整合挑戰。

從「快速展示」到「深度應用」的鴻溝：許多 AI 專案最終只停留在能夠進行「快速展示」（quick demo）的階段，看似亮眼，卻未能真正解決企業核心的、深層次的業務問題，導致 AI 專案的價值難以彰顯。

標準化部署模式的缺乏：鑑於技術的快速變化，業界尚未形成一套

標準化的 AI 部署與維運模式，這使得企業在擴展 AI 應用時，難以複製成功經驗，增加了時間與資源成本。

5. 組織文化與人才的結構性落差

技術的導入終究需由「人」來驅動。企業在 AI 轉型之路上，內部組織文化與人才能力的結構性落差，往往是比技術本身更難克服的挑戰。

認知差距與期望錯位：許多企業高層將 AI 導入簡化為「購買 GPU」或「開通 OpenAI 帳號」，對其背後所需的資料策略、流程再造、風險治理與文化變革缺乏足夠深刻的認識。

關鍵人才的匱乏：市場上嚴重缺乏能夠應對 AI 時代資安威脅的專業人才。此外，既懂業務流程又懂 AI 技術的跨領域人才更是鳳毛麟角。

IT 部門的角色轉型壓力：傳統上作為基礎設施提供者的 IT 部門，在 AI 時代必須轉型為串連各業務部門知識、推動數據驅動決策的「橋樑」與賦能者，這對其既有職能與技能提出了全新的要求。

內部治理體系的缺失：成功的 AI 導入，絕非僅靠技術部門的單打獨鬥。它需要建立一套完善的內部治理框架，涵蓋組織文化、倫理準則、標準作業流程以及與外部法規的對接。

6. 經濟與法規的宏觀挑戰

最後，企業的 AI 策略還必須應對來自外部宏觀環境的挑戰。

高昂的投資成本：無論是自建（On-Premise）AI 基礎設施所需的高昂初期投入，還是租用雲端 GPU 服務持續不斷的營運成本，都對企業的財務構成了巨大壓力。

動態演進的監管環境：各國政府正積極探索對 AI 的監管方式，但由於技術發展過於迅速，立法者難以全面預見所有潛在風險。這導致法規可能過於籠統而缺乏實操性，或過於嚴苛而扼殺創新，企業必須在充滿不確定性的法規環境中尋求平衡。

能源消耗與環境衝擊：AI 運算中心是眾所皆知的「吃電巨獸」，其龐大的能源消耗不僅推高了營運成本，也帶來了顯著的碳足跡，對企業的永續發展（ESG）目標構成挑戰。

勞動市場的結構性轉型：AI 的普及將深刻地改變勞動市場，部分重複性高的職位可能被取代，同時也將催生新的職位。企業面臨著如何為現有員工進行技能提升與再培訓，以適應人機協作新工作模式的長期挑戰。

擁抱 AI 不僅是導入一項新技術，更是一場深刻的組織與社會轉型。**企業必須系統性地應對涵蓋資安、資料、倫理、技術、人才及法規等六大面向的挑戰**。若忽視任何一環，AI 帶來的效益將可能被其衍生的巨大風險所吞噬。唯有建立全面性的風險治理框架，才能在創新的同時，確保企業的韌性與社會的信任。

全球人工智慧政策與治理框架：創新、風險與信任的平衡

在當代科技發展的脈絡中，監理與規範已跳脫傳統管制思維，轉而以建構市場秩序與數位信任基礎為目標。對人工智慧的發展而言，沒有規範，企業難以獲得產品與服務開發、跨境合作與長期投資等所需的可預期性；但若規範僅著重於防範與限制，則會壓縮創新的空間並限制產業生態系的發展。因此，監理思維必須從單純的防弊，轉化為同時兼顧「創新促進」與「風險控管」的創新型治理機制。

在全球 AI 競賽中，創新積極的治理思維與規範是打造長期競爭優勢的基石。各國正積極轉變監理角色，從傳統的防弊管制，邁向兼顧「促進創新」與「風險控管」的積極治理。歐盟、新加坡、日本、韓國與美國等區域及國家近年陸續推出的 AI 政策與法規，正好展示了創新積極的治理模式如何在引導產業發展上發揮關鍵作用。

一、AI 治理的多元路徑：各國政策比較分析

1. 歐盟的全面性規範：最早提出「AI 法案」樹立全球標竿

歐盟是全球最早且最積極推動 AI 法制的地區，該法案以「風險分級」為核心，將 AI 系統的風險區分為「不可接受」、「高風險」、「有限風險」及「最低風險」四級，並據此設計不同的管理措施。例如，生成式 AI 被歸類為「有限風險」，要求需標示 AI 生成內容；而司法判決預測則屬於「高風險」，上市前必須審查。

政策重點：

- 法案設立了一整套文件化、監管與透明義務，並在 2024 年 5 月 21 日正式生效後，至 2027 年分階段實施，為企業提供清晰的時間表與責任框架。
- 歐盟要求各會員國設立 AI 沙盒，並制定協調標準與通用模型行為準則，確保相關規範能落實為可操作的測試與稽核流程。
- 此法案旨在建立全球供應鏈的共同語言，讓符合規範的廠商在國際市場上更具競爭力。

2. 美國的創新生態：以國家戰略與技術標準引領發展

美國將 AI 視為國家競爭力與安全的基石，其政策設計以促進創新與維持國際領先為戰略目標。

政策重點：

- 「2025 美國 AI 行動計畫」是一項全面性政策，旨在加速 AI 創新、建設美國 AI 基礎設施、加強出口管制，並推動 AI 技術的國際標準化。
- 美國國家標準暨技術研究院（NIST）發布了「AI 風險管理框架」（AI RMF），強調「辨識－分類－緩解－監測」四大流程。NIST 也更新了「對抗性機器學習」文件（NIST AI 100-2e2023）等指引，以應對 AI 帶來的網路安全風險。

3. 新加坡的生成式 AI 治理框架 (Generative AI Governance Framework)

新加坡採取「自願性框架＋可驗證測試」的路徑，該框架以「原則、指南、工具」的治理方案，提出九大治理面向的具體操作建議。

- 透過 AI Verify 等工具，將原則轉化為可測試、可報告的流程。
- 新加坡特別強調可驗證性、資料治理及負責任的部署，同時推動全球 AI 保證沙盒，這種輕量化、工具導向的模式，降低了新創進入門檻，也方便企業將合規能力轉化為市場背書。

4. 日本的「AI 促進法」 (AI Promotion Act)

日本採取「促進與推動型法制」的模式，以推動研發與利用為主軸，輔以政府調查與矯正機制。

政策重點：

- 強調透明與信任機制及國際互通與產業支持，並營造創新生態。
- 日本政府在 2025 年發布的生成式 AI 採購指南，直接將風險控管、資料治理與人類監督要求納入政府採購流程，透過「買方力量」設定市場合規門檻。
- 在資料著作權問題上，日本提出只要用於訓練目的，AI 訓練過程就不受著作權限制（合理使用），顯示其在 AI 發展上採取較為寬鬆的解釋。

5. 韓國的「AI 基本法」 (AI Basic Law)

韓國透過該法建立政府產業共治的結合促進與風險管理的架構。

- 法律兼顧產業發展的支持政策，以及對具有高影響性的生成式 AI 課與義務，例如風險評估、透明化與本地代表設置。
- 韓國以國家戰略技術的高度推動 AI 投資與聚落發展，使企業能在清晰的法律框架下規劃合規轉型，並加快接軌歐洲等海外市場。

二、重要國際標準與指南

1. 國際標準化組織 / 國際電工委員會 (ISO/IEC)：ISO/IEC 42001:2023 是 AI 管理系統的國際標準，強調可信賴 AI 的要素，包括透明、公平、可解釋、可追溯。

2. 開放全球應用程式安全計畫 (OWASP)：OWASP 針對大型語言模型（LLM）應用程式發布了「十大 LLM 應用程式風險」，成為 AI 時代網路安全的重要參考指引。

3. MITRE 的「人工智慧系統對抗威脅框架」（ATLAS）：這是一項應對 AI 系統對抗性威脅的國際參考框架。

4. 行業主導的標準與聯盟：透過集體行動和頂尖科技公司的專業知識，有助於打造技術開源解決方案和最佳作法。例如 Cisco 建議政府應優先考量業界主導的標準和指導方針，如安全 AI 聯盟（CoSAI）所制訂的標準。

三、AI 治理的核心原則與概念

可信任 AI (Trustworthy AI)

1. 確保 AI 系統在開發、部署及使用過程中，能維持理性、透明、公平、可解釋、安全和可靠的運行。
2. Cisco 的負責任 AI 原則包含透明度、公平性、承擔責任、隱私、安全和可靠性等六項。
3. 可解釋性：不僅程式碼，更重要的是能解釋決策背後原因，例如為何拒絕某位求職者。
4. 偏見與幻覺：AI 系統可能產生仇視性言論、歧視或霸凌等不當輸出，或產生不實的「幻覺」，這需透過資料來源的公正性與輸出篩選來防範。

風險導向監管

1. 國際政策普遍支持依據 AI 使用案例的風險因素進行區分和比例調整，而非採取籠統的「一視同仁」法規，以避免阻礙創新。
2. 法規應著重於規範對法律和人權造成影響的特定「高風險」使用案例，例如公共服務決策或公共場所的人臉辨識和生物識別技術部署。

安全內建 (Security by Design) 與資料保護

1. 將安全原則整合至 AI 開發生命週期，包括供應鏈稽核、資料完整性審查、準確性監控和威脅適應性。
2. 在大型資料集上訓練 AI 模型時，需有明確且有效的資料隱私和安全策略，以應對意外推斷和敏感資料暴露（特別是兒童資料）。
3. 應建立強大的輸入審核工具以偵測和阻擋惡意提示，並設置輸出值篩選器以確保內容的完整性與安全性。所有互動都需全面監控和記錄。

主權 AI (Sovereign AI)

1. 此概念強調建立在地可控、能自主調整與優化的能力體系，而非單純由國家主導模型開發。
2. 包含自主算力、自主模型、建立創新機制（如資料沙盒、公共模型授權條款）和 AI 政策落實等四個層次。
3. 國際間重視資料治理、跨境傳輸和地緣政治因素在主權 AI 中的角色。

四、挑戰與展望：應對 AI 治理的未來變局

儘管全球治理框架日漸清晰，企業在實踐中仍面臨諸多挑戰，包括 AI 治理與規範、產業研發應用、安全威脅與管理、多方利害關係人等面向，隨著技術的快速發展，均有需要持續關注的議題。

1. 法規與技術發展速度的落差

- AI 技術發展非常快速，政策制定者難以完全掌握所有風險。Cisco 建議法規應具備靈活性，以適應不斷變革的技術。
- 成功的資訊科技是建立在無數錯誤的基礎上，這與過去立法在管理其他產業（如醫療新藥）的本質完全不同。

2. 「知道 AI」到「用 AI」的實踐鴻溝

- 儘管企業對 AI 的關注與認知已明顯提升，但從概念認知到實際應用仍存在巨大落差。
- 許多企業對 AI 應用的理解與實際操作模式存在顯著落差，如誤以為「裝置端 AI」皆為本地運算而忽視資安風險。

3. 資安威脅的演變

- AI 帶來了全新的潛在威脅向量，包括針對模型、AI 應用程式與服務的攻擊，以及使用 AI 本身作為攻擊工具。
- 「Agentic AI」（AI 代理）的興起，使得 AI 不再只是被動執行命令，而是能主動理解、預測與協作的智慧夥伴，但也帶來了新的攻擊面和風險。

4. 多方利害關係人協作

- AI 治理無法單靠政府或企業一方完成，需產、官、學、研共同努力，建立從硬體到軟體、從基礎架構到應用服務的健康生態系。

五、給企業的策略建議：

全球 AI 治理正朝向一個以風險為導向、以信任為核心、以標準為基礎、以協作為模式的未來邁進。面對這股浪潮，企業應將 AI 治理視為強化韌性與創造競爭優勢的策略性投資，而非單純的合規成本。

1. 建立內部治理框架 : 主動導入國際標準與最佳實務，例如 Cisco 提出的「負責任 AI 框架」以及 ISO/IEC 42001 管理體系，將可信任 AI 的原則內化為組織文化與決策準則，確保治理架構可持續且可落地。

2. 實踐資安優先架構 : 落實「安全內建」（Security by Design）的理念，並運用 Cisco AI Defense 等解決方案，針對 OWASP 十大 LLM 風險提供防禦機制。透過完整的防護與監控流程，確保 AI 應用不僅安全，更能符合法規與產業規範。

3. 遵循風險導向原則 : 針對企業內部的 AI 應用進行系統性盤點與風險分級，集中資源在高風險領域的管控。此舉能提升合規精準度與治理效率，避免資源分散並確保 AI 落地的穩健性。

4. 積極參與生態系 : 不僅是內部治理，企業更需主動參與行業聯盟與標準制定，並深化與技術夥伴的合作。這將使企業在快速變動的技術與監管環境中，不只是跟隨者，而是市場規則的塑造者與引領者。

結語

在 AI 驅動的未來，企業韌性的關鍵，不僅取決於技術實力，更取決於能否建立可信任的機制與文化。唯有積極擁抱全球治理趨勢，並將「可信任 AI」深植於企業 DNA，企業才能在不確定的時代中行穩致遠，並持續創造長期價值。

全球對 AI 的監管思維已從「防弊管制」轉向「促進創新與風險控管」並重的積極治理。對企業而言，**主動導入國際標準、實踐「安全內建」，並建立以風險為導向的內部治理框架**，已不再是合規成本，而是建立市場信任、強化營運韌性、並在全球 AI 浪潮中取得競爭優勢的策略性投資。

台灣 AI 發展風險評估

從策略到能源，台灣產業 AI 化面臨的七大風險

從策略到能源，台灣產業 AI 化面臨的七大風險

人工智慧已成為全球產業轉型的關鍵引擎。台灣雖具備科技與製造優勢，但在推動 AI 落地過程中，仍面臨策略、數據、技術、人才、資安與基礎設施等多重挑戰。調查顯示，多數企業雖已關注 AI，卻難以跨越從「知道」到「會用」的鴻溝。若缺乏系統性的治理與韌性布局，臺灣產業將可能錯失 AI 帶來的結構性機會。

1. 策略與規劃面風險

許多企業不了解應用場景的重要性，誤以為 AI 如同套裝軟體，只要導入就能立即見效。近五成企業尚未宣布 AI 發展策略，多數仍停留在工具與技術輔助層級，未能配合企業營運策略，建立完整的 AI 策略與導入路徑圖（Roadmap），這限制了 AI 在提升營運效率和創造競爭優勢上所能創造的新價值。此外，企業高層與決策者對 AI 的認知分歧與期望模糊，缺乏統一標準，影響後續的績效評估與決策制定。常見問題包括：

盲目投資與資源錯配：受生成式 AI 熱潮驅動，部分企業「為 AI 而 AI」，缺乏需求評估即投入資源，最終難以產生實質效益。

應用想像過於單一：企業對 AI 的應用想像多半以「行銷應用與內容製作」為主（零售與服務業高達 68.4%），而在產品開發與創新上的比例卻極低（僅 5.3%），這限制了 AI 更廣泛的發展潛力。

過度聚焦模型開發：台灣的 AI 發展策略仍過度聚焦於模型開發，而非如何將現有 AI 技術更有效地整合到業務流程中，導致企業在尚未明確需求時便投入資源，影響產業推進。

這些現象反映出台灣產業策略規劃上的斷層：既低估 AI 帶來的產業巨變，又缺乏統一標準與績效衡量，導致企業決策者期望與實際落差。

2. 數據管理與治理面風險

AI 的價值來自數據，風險也同樣有大部份來自數據，全球均將數據治理（Data governance）視為 AI 應用關鍵。但截至目前為止，台灣企業在數據治理上普遍不足：

資料尚未整合或缺乏明確策略：許多企業在資料尚未整合或無明確治理策略的情況下，急於上線應用 AI。調查顯示，企業在數據治理方面認知普遍不足，三分之一的企業未採用明文規範或並無治理相關規範 。

缺乏數據治理的理解與實踐：多數企業對「內外部資料交換策略」認識不足，限制了資料價值的發揮和數位轉型的深度。

「制度債」浮現：導入 AI 挑戰已從「資料債」（品質與整合不足）轉向「制度債」（缺乏風險框架與責任機制），使 AI 應用在規範空白下快速擴張，增加風險外溢。

這顯示，若無法建立穩健的數據治理與 AI 治理框架，台灣企業將無法有效釋放資料價值，也難以承擔 AI 應用所必須擔負的責任。

3. 技術認知與導入的風險

AI 技術的迭代速度極快，但多數企業在技術理解上仍存在明顯落差，導致導入過程頻頻受阻，主要體現在以下幾個面向：

名詞與實務脫節：對模型訓練、推論、裝置端與雲端 AI 等概念缺乏準確認識，導致導入失敗。

裝置端 AI（On-Device AI）認知不足，暗藏資安風險：多數企業對裝置端 AI 的認知有限，未能清楚區分其與雲端 AI 的技術差異，導致難以有效控管資料流向，增加資安管理挑戰。使用者可能誤以為裝置內建 AI 即代表所有運算皆發生於本地端，不自覺地將敏感資料上傳至 AI Chatbot，增加數據外洩風險。

基礎設施壓力：AI 算力需求高，會導致現有 IT 基礎設施如網路、儲存等面臨流量暴增的挑戰。例如，一個 AI 查詢所耗用的電力約為傳統 Google 查詢的 8 到 10 倍，這對網路頻寬、能源和運算資源構成巨大壓力。

依賴外部供應商，內部自主性低：台灣企業普遍倚賴既有的外部工具與平台，內部缺乏自主研發與掌握能力。長期而言，這可能削弱企業的核心競爭力，限制其在產業升級與創新上的主導性。

4. 人才與組織文化面風險

AI 導入的關鍵，除了技術，更多的原因涉及了組織文化與管理，這意味著，若未能建立跨域對話與信任文化，AI 最終可能成為組織轉型的阻力，而非助力。

人才缺口與策略不足：近五成企業未建立 AI 人才發展策略；就算有計畫，但大多停留在送員工上課這類傳統學習方式，未能與工作及企業發展目標對齊，與員工個人職涯規劃也未有緊密關係。

專案管理能力不足：僅重視技術導入而忽略 AI 專案管理，導致落地困難。

內部認知落差：AI 可能加劇員工間的能力斷層，並引發不信任感。

績效評估失效：傳統考核方式無法反映 AI 協作下的生產力，阻礙員工參與動力。

5. 資訊安全與韌性層面風險

AI 應用雖帶來效率與創新，但同時也衍生出新的依賴與脆弱性。在 AI 驅動的自動化環境下，系統的每一個節點都可能成為潛在的攻擊入口，使企業韌性的定義必須被重新檢視與擴張。

首先，過度集中於單一模型或供應商，將導致顯著的集中風險。一旦服務中斷，便可能引發連鎖效應，對業務運作造成全面衝擊。其次，員工普遍使用外部 AI 工具卻缺乏組織性的治理與監管，導致「影子 AI

（Shadow AI）」的現象蔓延，不僅難以掌控數據流向，也使企業資訊安全陷入灰色地帶。

此外，AI 應用也帶來許多新型態資安威脅，例如提示詞注入、模型中毒、資料洩漏及模型竊取等，皆可能突破傳統資安機制的防護範圍。然而，儘管有高達九成的企業聲稱已經運用 AI 來強化資安，實際上真正達到成熟應用等級的卻不到 4%，顯示多數企業的防禦力仍然不足。

6. 基礎設施與能源的風險

<4%

企業未達到 AI 資安成熟應用

即使高達九成企業表示，已運用 AI 強化資安，然而成熟度達標者不到 4%，凸顯防禦力仍嚴重不足。

10 倍

耗電量

一次 AI 查詢的耗電量可能是傳統搜尋的 10 倍，龐大的能耗不僅加重運算壓力，也可能超過現有供電能力

AI 的高算力需求正直接衝擊台灣既有的基礎設施，若電力與網路無法維持穩定供應，AI 應用將難以真正規模化。

在電力方面，一次 AI 查詢的耗電量可能是傳統搜尋的 10 倍，龐大的能耗不僅加重運算壓力，也可能超過現有供電能力，成為推動 AI 發展的重要瓶頸。另一方面，網路基礎設施也面臨嚴峻挑戰。AI 帶來的高流量會迫使既有網路架構進行頻繁調整與路由優化，而在遭遇惡意攻擊時，更容易顯露出其脆弱性，增加服務中斷與資安風險。

7. 政策與法規的風險

最後，治理與法規滯後的問題，正使產業風險進一步放大。若政策與法規無法加速追趕技術腳步，產業治理恐將陷入被動。當前的挑戰主要體現在幾個面向：其一，缺乏統一的治理框架，導致部會間定義與責任分散，無法形成全國一致的 AI 治理機制；其二，立法速度落後於技術發展，例如「AI 基本法」草案仍難以因應快速變化的應用場景，而立法者對 AI 的理解與認知不足，更使落差加劇。除此之外，軟體產業本身規模有限、人才不足，難以與全球競爭者匹敵；再加上老舊系統持續存在，既增加升級難度，也暴露出資安隱患，形成雙重障礙。

結論與前瞻

台灣產業正站在 AI 轉型的十字路口。從策略缺位、數據孤島、技術誤解、人才缺口，到資安與基礎設施不足，這些挑戰都凸顯了「可信任 AI」與「韌性」的迫切性。

未來，台灣若要在全球 AI 浪潮中確保競爭力，建議需關注以下事項：

1. 建立清晰的產業 AI 策略與治理框架，避免盲目投資與片面應用。
2. 強化數據治理與跨部門協作，將「資料債」轉化為「資料資產」。
3. 培養跨域 AI 人才並重塑組織文化，建立員工的信任與專案管理能力。
4. 提升資安防護與營運韌性，將 AI 風險納入企業風險管理核心。
5. 升級能源與網路基礎設施，支撐大規模 AI 應用需求。
6. 推動前瞻性政策與法規，縮短技術發展與制度建設的落差。

台灣產業的 AI 轉型正處於一個「知易行難」的十字路口，**普遍面臨從策略、數據、技術、人才、資安，乃至基礎設施與法規等七大環環相扣的風險**。企業若僅將 AI 視為單點的行銷工具（應用想像單一，僅 5.3% 用於產品創新），而忽視了背後所需的系統性治理與韌性佈局，將可能錯失由 AI 驅動的整體產業結構性升級機會。

焦點：AI 發展治理與實務

- 從可信任出發，建構持續迭代的產業 AI 平台
- 挑戰既有網路架構，AI 風險治理須平衡目標與底線
- 建構 AI 治理框架迫在眉睫，跨域對話平台為基礎

數位發展部部長

林宜敬

“

臺灣長期累積的「可信任」品牌形象，如何延續至立法與政策設計中，需要具體的政策工具與法律框架承接。

在討論可信任 AI 的同時，信任不僅來自模型本身的透明性與可解釋性，更取決於其在面對外部衝擊時能否維持穩定與持續運作，必然延伸至韌性的討論。而算力的強大固然重要，但若缺乏長期可靠性，便無法真正支撐產業與社會的永續發展。

”

從可信任出發，建構臺灣 AI 生態系

人工智慧的發展，已被廣泛視為當前全球科技與經濟競逐的核心動能，AI 治理（AI Governance）也逐漸凝聚為全球共識。美國、英國、日本與新加坡多透過業者自律、風險管理與實驗沙盒，在推動創新同時兼顧安全與責任。歐盟則率先以「AI 法案」（AI Act) 為基礎，確立「風險分級」架構，按等級不同制定規則，逐步建立透明、監管與責任機制，確保高風險應用能在規範下落地。美國川普政府最近也發布了 AI 行動計畫，重點強調加速創新、強化基礎建設，以及在國際外交與安全領域的領導角色。無論是軟性的引導或硬性的規範，這些治理模式的共同核心，都是透過制度設計來換取社會信任，並為企業創造可預期且長期的競爭優勢。

而臺灣正處於這股國際規範競逐的交叉點。除了擁有 ICT 技術優勢與完整的製造供應鏈，更重要的是在國際間長期累積的「可信任」品牌形象。然而，現實挑戰在於如何將這項資產延續至立法與政策設計之中，使「信任」不僅成為產業競爭力的象徵，更能轉化為支撐臺灣 AI 發展的核心基石，這需要具體的政策工具與法律框架來承接。

可信任 AI 需先關注訓練過程透明性

數位發展部部長林宜敬指出，在「可信任 AI」的討論中，首先必須關注整個訓練過程的透明性，如資料來源如何取得、模型如何訓練，以及最終輸出的結果如何被檢視。

目前部分語言模型如果用簡體中文資料訓練，在「臺灣價值觀」的表現會比較差，此一現象突顯推動本土化語料建置之重要性。為確保 AI 發展符合臺灣社會價值與語言文化，提供具在地文化與本土價值的繁體中文資料，是作為語言模型訓練的重要基礎。

林宜敬說，正因如此，臺灣需要發展屬於自己的「主權 AI」。其中大語言模型不僅要在補足現今在語言與知識層面的缺失，更必須在價值觀與觀點表達上，建構出能代表臺灣特色的體系。然而，他也坦言，如何在技術設計中體現多元價值觀，仍是極具挑戰性的課題，且是目前難以單靠技術手段完全解決的問題。

「可解釋性」需重新定義與形成共識

至於歐盟在今年出版的「AI 行動計畫」（AI Continent Action Plan）中，提到「可解釋性」（Explainability）的要求。林宜敬說，「可解釋性」並非全面揭露 AI 的內部參數或程式邏輯，而是聚焦於「指標選擇」。換言之，當 AI 被應用於決策情境時，必須能清楚說明系統依據了哪些指標進行判斷。這樣的設計不僅能讓利害關係人理解並容易進行討論、形成共識，也能避免 AI 的決策淪為黑箱。

因此，在現行與未來的法規設計中，「可解釋性」必須被慎重納入考量，而在實際立法過程中，將持續向相關部會或立法委員進行溝通。

林宜敬直言，過去傳統思維認為必須透過程式碼追溯邏輯，才能解釋 AI 的運作。但這種方式並不適用於生成式 AI 與大型語言模型的技術本質，也突顯出「可解釋性」需要被重新定義與理解。唯有在法律框架中逐步建構新的解釋模式，才能兼顧技術現實與政策需求，並進一步鞏固社會對 AI 的信任。

相較於其他的資訊技術、軟體系統，要建立對於 AI 的信任難度更高。原因是 AI 牽涉到數據、模型訓練等層面，而且往往牽一髮動全身，以推出即引發廣大討論的 DeepSeek 為例，可以看到它在資訊安全上涉及三種類型的問題。第一類是傳統資安風險，例如是否存在病毒或蠕蟲等惡意程式；第二類是個人資料保護，DeepSeek 之所以被外界認為有所疑慮，主要就出在這一層。因為在當前環境下，數據本身就是資產，若個資保護不到位，等同直接衝擊企業與用戶的核心利益。第三類則屬於價值層次的資安問題，也就是前面提到與臺灣價值相關的風險。第一類傳統資安及第二類個資保護問題，大多能透過技術手段加以處理。至於第三類屬於價值層次的風險，則需要透過更廣泛的制度設計與跨部門討論。

政府須確認政策方向，提出可信任治理架構

AI 立法的討論並非近年才出現，只是隨著生成式 AI 出現後，各國加速將 AI 視為經濟翻轉與戰略競爭的重要資本，相關議題正逐步成為區域與國家政策的核心焦點。對臺灣而言，若僅依循大算力、大模型、大數據的 AI 傳統邏輯發展，很可能陷入「大者恆大」、甚至贏者全拿的局面，不利於中小產業與社會多元面向的受益。而「可信任」正是最具獨特性的資產。問題在於：如何在當前環境下，逐步建構出能同時獲得國際認可與人民信任的治理架構？

林宜敬認為，這是政府重要的使命，必須透過政策研擬方向，讓大家清楚事情的全貌，無論是好或壞消息，都要先將事實公開透明，接著再與社會大眾討論因應方式以及相關風險評估。他強調，制定政策跟寫程式一樣，必須做”Top-Down Design”，也就是政府需要先訂出政策的大方向及主要目標，然後再層層推進、分工執行，這樣才能避免資源分散，聚焦在臺灣產業的既有優勢，發展新世代的競爭力。

多元化的基礎建設，是建構數位韌性的優勢

在討論可信任 AI 的同時，信任不僅來自模型本身的透明性與可解釋性，更取決於其在面對外部衝擊時能否維持穩定與持續運作，必然延伸至韌性的討論。而韌性不只是技術問題，更關乎制度設計與治理框架的完整性。畢竟，算力的強大固然重要，但若缺乏長期可靠性，或在戰爭與大規模衝擊下等極端情境下完全失能，便無法真正支撐產業與社會的永續發展。

特別是在臺灣獨特的地緣政治環境下，資料中心與雲端基礎設施勢必會成為首要目標，一旦遭到破壞，將對產業與國家安全造成重大打擊。因此，如何在可信任 AI 的基礎上，進一步建構具韌性的數位體系，已成為迫切且必須面對的課題。

林宜敬指出，AI 本身並非戰爭情境下最關鍵的問題。對一般民眾而言，AI 服務的中斷並不會立即造成嚴重影響，真正攸關社會穩定的，反而是基礎的支付與經濟系統。相較於中國高度依賴數位支付，臺灣的支付與金融生態較為多元，除了電子支付外，人們仍普遍備有現金，在一定程度上已構成基本的韌性保障。

他進一步指出，一個社會的數位化程度越深，往往韌性反而越低。真正決定韌性的關鍵不是通訊技術，而是電力供應。災後通訊全面中斷的根本原因，不在於網路系統失效，而在於電力中斷所導致的連鎖

反應。隨著社會對網路與數位服務依賴程度不斷提升，電力的穩定性便成為數位韌性最基礎、也是最關鍵的保障。

林宜敬認為，臺灣目前在能源結構上已經擁有相當多的分散式電力來源，這實際上是一項優勢。若能進一步推動分散式供電系統的發展，將能在災害或突發事件中，顯著提升整體的韌性。

他強調，整體而言，韌性的核心在於「多元化」。以對外通訊為例，不僅需要光纜，還必須具備低軌衛星與微波等多種管道，才能確保服務不中斷。當多元基礎結合靈活調度能力，才能真正建構出穩健的數位韌性。在與民眾及媒體互動中，如何適當傳達韌性建設的努力與進展，同時避免洩露過於敏感的資訊，將是政策溝通中不可忽視的課題。

從 Project 到 Product，培養軟體公司競爭力

在人工智慧快速發展的情況下，臺灣 AI 產業的政策推動初期曾因跨部會職責不夠明確而出現重疊，各部會皆將 AI 視為重要戰略領域，但在起步階段，分工尚未完全釐清。但隨著政策逐步落實，相關角色也愈加清楚。例如：在企業端應用或長照產業中的日照中心，衛福部負責產業監管，數發部則著重於支持資訊服務與軟體業者，協助開發 AI 應用，並推動解決方案導入實際服務場域。

同時，企業輔導機制也逐漸調整，過去的推動模式主要仰賴專案補助機制，以醫療照護產業為例，直接補助終端看護、日照中心，這些單位規模有限，有些缺乏軟體自主開發能力；另一種做法則是補助大型醫院，讓其資訊室各自開發系統。類似這樣的補助方式，很容易造成「各自開發、功能雷同」的結果，資源分散，難以形成規模效益。

林宜敬表示，臺灣部分政府機關與大型客戶長期以來多採專案式外包模式，傾向要求軟體廠商依照個別需求進行客製化，導致軟體廠商需頻繁修改程式。然而，相較於購買國際軟體產品時，並不會要求對方修改原始程式碼，此現象亦反映出臺灣在軟體產品化價值認知上仍有調整空間。

他進一步指出，若臺灣軟體產業希望邁向國際市場，勢必須從專案導向逐步轉向產品導向。專案模式固然能滿足特定需求，在本地市場行得通，但在推動海外業務時往往面臨落地與擴展挑戰。相對而言，具標準化與可擴充性的產品，更有利於推廣與規模化經營。因此，數發部在推動產業升級的過程中，重視協助業者建立產品化思維，逐步擺脫對一次性、難以延伸的客製化專案的依賴，以強化整體產業的長期競爭力。

未來數位產業發展的方向，要重新思考如何推動通用解決方案，讓臺灣軟體公司能夠在這些通用解決方案的基礎上，加入本身過去對於不同產業知識的經驗，進一步培養出持續研發成長的競爭力，進而走向國際市場。

最後，林宜敬表示，我們從全世界高科技產業的發展歷史可以看到，科技產業，尤其是軟體產業的創新突破與發展，是來自民間產業的自由競爭。數發部最大的職責，就是創造一個適合臺灣 AI、資安、資訊服務等產業成長茁壯的產業環境，讓臺灣的資訊產業在幫我們解決問題的同時，也能獲得豐厚的利潤，但又不會給社會帶來新的問題。

台灣網路資訊中心、亞太網路資訊中心董事長
黃勝雄

治理不能假設「零錯誤」，而應以「錯誤必然、風險可控」為前提。

『全黑運作韌性』的關鍵在於，無論是在停電、斷網、資訊系統癱瘓等極端狀況下，重要基礎設施仍須能以手動或離線方式運作。尤其是進入戰爭狀態，單純依賴異地備援並不足夠，因為備援系統同樣可能被循線入侵，形同向攻擊者暴露後備位置。

挑戰既有網路架構，AI 風險治理須平衡目標與底線

網際網路的開放性與延展性，使區塊鏈、大數據、物聯網與人工智慧等新興技術無需改動底層結構，便能直接掛載於既有框架之上。這一特性成就了數位生態的快速擴張與多元應用。即便邁入 AI 世代，網際網路依然是關鍵基礎設施，但其治理挑戰早已浮現。

台灣網路資訊中心及亞太網路資訊中心董事長黃勝雄指出，50 年前建立的網路基礎，放到如今複雜的技術及應用環境中，像是脆弱的舊地基，在全球 50 億使用者與數百億裝置的壓力下早已暴露出資安缺陷，而 AI 的崛起進一步放大了這些風險。雖由私有企業驅動，卻帶來高度公共化的外溢影響，從個資濫用到跨境詐騙，治理難題更形複雜且尖銳。

黃勝雄點出 AI 和資訊科技的發展，本質上就是伴隨錯誤而生，並經由不斷的修改（trial and error）而達到原本設定的功能目標。因此 AI 治理的核心不在於追求零風險，而是明確劃定底線並聚焦公共目標。因此，風險管理設計應確保即使在最糟糕的「全黑情境」下，基礎設施仍能維持運作。

相較網路治理 AI 治理更加複雜

過去普遍認為，網際網路架構穩固且安全，可以一層層往上堆疊，但實際情況卻並非如此。黃勝雄解釋，由於最初的設計僅供少數學術機構交換資訊，未考慮安全控制，更不可能預料到今日會有 50 億使用者與數百億裝置的規模。

生成式 AI 的出現，更是徹底顛覆既有的分層架構與遊戲規則。黃勝雄提醒，像 ChatGPT 這樣的大型應用，從網路架構角度來看，處於一個極為脆弱的狀態。由於長期缺乏針對惡意行為的防禦設計，AI 生態在成長期雖能快速擴展，但一旦涉及龐大商業利益，勢必會成為攻擊與治理爭議的焦點。

回顧網路治理經驗，聯合國雖曾嘗試推動全球性公約，但因分歧過大而無疾而終。最終，治理模式只能依賴多方利害關係人的自律規範，逐步形成共識。雖非具強制力的國際標準，但只要各方願意遵循，其效力仍能成立。這也顯示，治理往往建立在合作與信任機制之上，而非單一強制手段。

數十年來也累積了有限但可借鑑的治理經驗與制度，例如資安規範、跨境合作及 ICANN 等跨國組織。雖然這些機制功能有限，卻形成了一種「公共資源治理」的全球實驗。在缺乏更好選項時，至少提供了一套相對理智、有效的管理方法。

但是，人工智慧的挑戰更為複雜。網路治理雖以「全球一體適用」為原則，視其為跨境的數位公共財；但 AI 多由私有企業推動，卻帶來高度公共化的外溢效應，涉及個資保護、社會風險與制度設計。這意味著，AI 治理既不能僅依賴市場，也不能放任私有財邏輯主導，其難度遠超過傳統網路。未來政策設計的關鍵，重點應該是如何在公共利益與私有創新之間取得平衡，並建立清晰的規範與可信任的底線。

黃勝雄建議，台灣在推動 AI 治理時，應聚焦具有公共性的議題範疇，並靈活運用法律、自律與他律等多元手段，以「最能快速落地並達成治理目標」者優先採行。

AI 風險並非來自工具，而是用途

「AI 工具本身不具風險，風險來自被用來做什麼。」黃勝雄舉例，假使免費送你一台能自動連網的智慧咖啡機，能根據家中成員的喜好自動沖煮咖啡，交換條件是咖啡機上必須裝置攝影機。這時候可能造成傷害的風險並不在咖啡機，而是影像資料的蒐集與使用。換言之，AI 並非天生具有風險，關鍵在於它被用來做什麼。

同樣道理，即便有免費軟體能百分之百滿足需求，多數人也不敢輕易使用，因為背後的數據風險不明。若連個人用戶都不敢嘗試，企業更不可能因「免費」就放心採用。

黃勝雄指出，台灣在 AI 風險判斷上，產學界存在明顯落差。學界缺乏真實世界的應用經驗；產業與政府雖直接接觸應用，卻經常以過於簡化的邏輯判斷風險，忽視了更深層的數據結構與制度認知和洞察，使得討論仍停留在表層，未能真正觸及關鍵問題。

面對風險，應事前建立因應方案而非全面圍堵

這同時也凸顯出台灣社會缺乏風險管理的困境。黃勝雄進一步解釋，可以將風險想像為統計學上分佈區間，大多數風險來自最常見的弱點，如使用「12345678」這種弱密碼，或提款卡掉在地上被撿走。這些低門檻風險，預防成本低但發生率高。隨著分布往外延伸，風險發生機率大幅降低，帶來衝擊也逐漸增大。若要追求「99.999% 的零風險」，所需成本將高到社會難以承受。他以 COVID-19 的「清零」政策為例，雖能有效降低感染風險，但最終卻造成社會與經濟付出過高代價。

因此，黃勝雄認為，務實策略是將防護計畫鎖定在兩至三個標準差（約 95% 的常見情境），用低成本手段就能降低多數風險；對超過兩個標準差的極端事件，則建立事前因應方案而非全面圍堵。他提醒，因為在氣候變遷與全球不確定性加劇下，極端事件可能加速到來，例如近期台幣在短時間內劇烈升值，原本被視為小機率事件，仍卻在真實世界發生並產生衝擊。因此，「低機率不等於可忽視。」

此外，治理思維也應從「如何防護」擴大為「如何因應」，亦即在有限資源下，預先界定必須保留的關鍵功能與可被犧牲的次要功能，建立清楚的先後順序與切換機制。比起危機當下倉促拍板，提前劃定優先序才能在最壞情境下維持社會基本運作。

AI 必然存在錯誤，韌性治理需預設「全黑」仍可運作

與傳統產業與硬體製造業要求「零風險才上市」不同，資通訊軟體與 AI 的發展勢必需要嘗試錯誤與修改，再加上後續的優化，所有模型都是在不斷的錯誤中迭代而成。就像當前無人車雖仍存在許多 Bug，但已正式上路，並透過不斷修正逐步降低事故率。這說明科技迭代本質上就是「在錯誤中成長」。

「所有資訊系統天生就有漏洞（Bug），」黃勝雄直言，且沒有人能夠正確預測這些漏洞何時被攻擊而崩潰。這段時間成為全球討論熱點的「零日攻擊」，就是當攻擊者比防禦者更早發現漏洞，系統便在毫無準備下崩潰。因此，治理不能假設「零錯誤」，而應以「錯誤必然、風險可控」為前提。

「『全黑運作韌性』的關鍵在於，無論是在停電、斷網、資訊系統癱瘓等極端狀況下，重要基礎設施仍須能以手動或離線方式運作。」黃勝雄說，尤其是進入戰爭狀態，資訊系統極有可能已經被入侵，在明知系統可能遭入侵的情況下，首要措施應是立即斷線停用，但設施仍必須具備持續運作的能力。這正是「手動切換」不可或缺的原因。單純依賴異地備援並不足夠，因為備援系統同樣可能被循線入侵，形同向攻擊者暴露後備位置。

他強調，資訊系統與 AI 僅是工具。當思考營運韌性、企業 AI 化時，必須先把問題一層層地解構、梳理，釐清真正要解決的問題。回到根本，治理與決策都應從目的出發，了解組織的核心目標是什麼？策略要如何設計？執行框架如何落地？若只在片段議題上反覆拉扯，最後只會被碎片化的討論淹沒。唯有先把問題講清楚、邏輯釐清、底線畫清，才談得上有效的策略與可執行的方案。

可信任 AI 第一條底線：個人資料

「可信任 AI」是各界追求的目標，但並非觸手可及即可達成。黃勝雄指出，歐盟雖已提出風險分級與多項原則，在風險控管與個資保護上設定門檻；然而在可解釋性方面，受限於 AI 的本質，往往連系統設計者也難以說明模型為何產生特定結果。至於資安，目前多依賴系統掃描與稽核等「客觀檢測」，受限於平台或業者願意開放與配合的意願。

他進一步指出，歐盟常被批評為「立法成功、執法失敗」。主因在於條文設計未充分考量執法成本與跨境落實的難度，使實際執行易走向選擇性執法：既加重新創負擔、反而鞏固大型平台優勢。對台灣而言，能直接挪用的經驗有限。因此，制定規範必須回到本國能力與條件，先釐清法律要保護的對象與要達成的目標，再評估是否具備足夠的執行力落地。更務實的做法，是先發布可操作的使用指引，讓機關與民間明確知道應注意什麼、如何自保。

至於必須守住的底線，黃勝雄認為第一條紅線就是個人資料。在數位經濟時代中，資料既是 AI 的動力，也是商品，卻常在未充分告知與同意下被用於訓練、交換與變現。台灣目前個資管理仍分散於不同機關層級，外洩與濫用時有所聞；在沒有清楚的個資底線之下，其他原則亦將失去正當性。

因為一旦缺乏明確邊界與規則，規範就容易跳過底層原則、直接從最上位介入，進而碰觸言論自由與其他基本人權。因為不清楚規範背後的權衡考量，往往大刀一砍，先受傷的多半是本地業者；對跨境平台卻鞭長莫及。

對企業而言，在全面導入 AI 時，該如何實際防範與管理風險？黃勝雄建議，仍回到使用目的與情境，避免將企業核心命脈完全交由 AI 主導，以免形成單點失敗與治理真空。至於政策擬定則要符合產業現實。與其一次到位、五年十年的宏大藍圖，不如先找到安全穩健的利基點。

可信任 AI 的真正起點，不在於把工具塑造成完美，而在於把用途、責任與底線說清楚、做落地。當社會能在最壞情境下維持關鍵功能，並對資料與權力建立可被檢驗的邊界，可信任才不只是標語，而會成為一套可長可久的治理能力，並成為台灣產業 AI 化不可取代的優勢。

人工智慧科技基金會董事長
詹婷怡

治理的驅動力來自文化與思維的轉型及功能的需求，同時需要具備跨域專業與前瞻視野的人才，帶領治理框架的形成與落實。

台灣面臨專業人才不足及缺乏有效的跨領域對話兩大挑戰，以致治理與產業推動間出現落差。AIF的角色正是要補足這些缺口，建立對話平台，推動產官學研之間的交流，並以前瞻思維協助台灣在AI 治理與資料治理上形成清晰定位。

建構 AI 治理框架迫在眉睫，跨域對話平台為基礎

隨著人工智慧（AI）滲透至各領域產業與日常生活，在應用紅利之外，其潛在風險與挑戰更甚於以往任何技術迭代。特別是在 5G 與物聯網普及的時代，AI 與外部基礎設施高度連結，任何微小偏差都可能被迅速放大，進而引發資安風險、維運不穩定，甚至社會或倫理爭議等系統性風險，因此，「AI 治理」的迫切性不言而喻。

人工智慧科技基金會（AIF）董事長詹婷怡指出，推動 AI 治理的核心挑戰，並非單純的組織架構調整，而是組織文化與思維的轉變。若僅停留在形式上制定制度或進行組織修正，卻缺乏對風險、倫理與責任的深層理解，治理便容易流於形式，難以有效回應 AI 帶來的高度不確定性與其他挑戰。

建立清晰的 AI 治理框架的必要性

治理的核心是一套制度化管理機制，透過框架、程序、流程與機制，協助 AI 導入者 (可能是政府或產業與組織) 有系統地規劃推動與落地。唯有完善的治理機制，導入的目標才能落實，施政目標或產業發展才能在明確方向下被有效引導，並形成穩健的規範基礎。這不僅是政策推動的方法論，更是支撐公共治理、推動產業轉型與確保永續發展的核心基石。

她強調，不論是「可信任 AI」、「主權 AI」、產業落地以及資料蒐集傳輸與利用等議題，皆與治理緊密相關。理解並建立治理觀念，是確保 AI 健全發展的必要前提，也是推動社會信任與產業升級的起點。

從治理的角度來看，「可信任 AI」與「主權 AI」雖然核心要素不同，但都屬於 AI 治理框架的實質內涵，並深刻影響國家科技治理與產業定位。缺乏治理觀念，往往僅是零散制定法規或制度，無法構建完整框架；而對治理有意識的組織 (包括國家)，則能理解可信任 AI 與主權 AI 共同構成科技治理的兩大支柱

在治理模式上，可信任 AI 特別強調跨國合作，包含倫理準則、國際協議與多邊規範的制定；而主權 AI 更注重科技自主性、資料治理、跨境傳輸的掌控，以及因應地緣政治所帶來的風險挑戰。兩者相輔相成，共同構成 AI 治理的核心框架，為各國在推動人工智慧發展時提供行動準則與政策依循。

可信任 AI 與主權 AI 形塑產業未來樣貌的雙核心

詹婷怡強調，國家對「可信任 AI」與「主權 AI」兩大面向的重視程度，除了將重塑國家科技治理框架外，也將決定 AI 產業與未來科技的發展樣貌。在可信任 AI 層面，若能建立完整的責任機制、倫理規範、自律準則與政策指引，這些規範將進一步落實於更具體的制度設計，例如《AI 基本法》，以及各產業的專屬規範。對於醫療、金融等高風險領域而言，倫理與合法性將被強化；對大型企業而言，則必須承擔「責任型 AI」的更高標準，以確保技術應用符合社會價值與合規要求。

在主權 AI 層面，若能持續推動基礎建設自主、晶片製造、語言模型研發與多語言資料庫的建立，台灣便能逐步構築完整的 AI 產業生態鏈。反之，若缺乏這些重要資料基礎的投入，台灣可能僅停留於晶片代工角色，而在內容、語料與應用場景上依賴國際供應，無法形成具備自主性的 AI 能力。

而可信任 AI 與主權 AI 正逐漸互動與融合，成為新一代 AI 治理的核心框架。詹婷怡指出，從國際發展觀察，這一趨勢已相當明顯。例如，歐盟率先制定完整法案，日本與韓國亦推出相關規範，美國則以「AI Action Plan」》作為政策藍圖；同時，G20 與 OECD、GPAI 等跨國組織，也積極推動國際化治理架構。

詹婷怡表示，針對未來的 AI 治理模式，政府需建立大方向的治理架構，並以監督者角色確認企業是否遵守其制訂的承諾與規範；同時，為避免過度管制扼殺創新，私部門的角色將更加重要。企業不再只是被動遵循，而需如同 ISO 認證般，自行提出規範並承諾自律。「企業自律是根本，」她指出當問題發生時，政府與企業將需共同承擔責任，形成動態調適、互補合作的治理模式。

跨域對話平台 彌合治理與產學落差

面對未來的治理模式，是否必然需要進行大規模的組織調整？詹婷怡認為，更為關鍵的其實是「思維的轉變」。許多改革之所以成效有限，原因有二：其一，剛性的組織變革推動緩慢；其二，即便組織架構調整完成，但若治理思維與運作方式未隨之更新，改革仍難以發揮實效。組織本質上是人為設計的制度工具，即使沒有架構調整，只要思維轉型，還是可能逐步邁向目標；反之，即使新設部門或機構，若沿用舊有思維，則可能陷入空轉，甚至造成資源重疊與效率低落。

「Forms from function，」她直言，治理的驅動力來自功能與思維，關鍵也在於找到具備跨域專業與前瞻視野的人才，帶領治理框架的形成與落實。

她進一步以歐盟經驗為借鏡。多年來，歐盟很少為因應每一項新興科技特別大幅調整組織架構，而是透過設立跨領域的高階任務小組（task force）來因應新挑戰。這些小組結合政府官員、法律專家與產業代表，針對 AI、資訊與資料治理等議題展開深入討論。透過跨界對話與協作，逐步制定治理原則，並以此為基礎推動具體規範與立法。相關法規與產業規範次第展開設計。

相較之下，台灣主要面臨兩大挑戰：一是專業人才不足，二是缺乏有效的跨領域對話，以致治理與政策制定及產業推動間出現落差。做為產業 AI 智庫，AIF 的角色正是要補足這些缺口，建立對話平台，推動產官學研之間的交流，並以前瞻思維協助台灣在 AI 治理與其核心資料治理上形成清晰定位。

數位化關鍵在於打破部門孤島（silo）

詹婷怡指出，推動 AI 治理是數位轉型再進階。過去數位化常被誤解為「買一個系統」或「把檔案掃描上網」就完成，但這只是表層。真正轉型的核心在於透過數位化來打破部門與產業之間的孤島（silo），重構流程並創造新價值。唯有打破 silo，跨部會與跨產業的合作需求才會出現，形成創新生態系，而這正是推動 AI 治理與資料治理的前提。

她強調，AI 治理不只是立法，還涵蓋自律機制、政策框架與指引工具等面向，這些要素相互交織、環環相扣，必須以整體視角加以討論。隨著數位轉型打破部門與產業的界限，對跨部會協調、跨產業合作與跨域人才的需求愈加迫切。若缺乏這些基礎，就難以談及

更進一步的 AI 治理推動。

此外，還需發展測試與驗證機制，以確保 AI 模型的可信性。例如，若要判斷某一模型是否達到安全與合規標準，便需透過認證或第三方驗證，而這類服務最好由中立機構提供，確保公正性。這正凸顯了多方協作機制設計的必要性。

若缺乏多方協作，治理與轉型便難以推進。以企業為例，數位轉型若僅停留在單一部門或外部 IT 廠商的導入，而未能促成跨部門協作，便無法形成真正的變革。詹婷怡直言，轉型過程中，除了組織文化，包括人與人間的跨部門協作、人機協作、流程改變都很重要。

此外，過程中將需要大量多元人才的參與。不僅是產業領域（domain）專家與 IT、AI 技術人才，更需具備跨域整合能力的「策展」型人才，協調不同角色的合作。無論在國家層級還是企業層級，均需要明確的治理架構、平台型協作者，以及適當的誘因設計。唯有如此，才能推動創新、持續培養跨域人才，並建立長期對話與合作的文化。

AI 發展下的資料治理與永續議題

另一方面，隨著 AI 技術的快速發展與廣泛應用，資安威脅型態日益多樣化，挑戰也愈加嚴峻。詹婷怡從產業角度觀察指出，資安的需求與商機將持續攀升。這是因為 AI 創新過程中，必然伴隨新型風險；因此隨著產業 AI 化步伐愈快、產業應用愈多元普及，勢必驅動風險事前預測、防護機制與資安標準的持續演進。

同時，AI 帶來的挑戰不僅侷限於技術面，還涉及營運模式、品牌信任及永續能力，更和企業韌性密切相關。對國家而言亦然，AI 發展對資安與網路提出新需求，治理制度也必須持續調整以應對風險。雖然資安問題本質上無法徹底解決，但可透過提前預見、降低風險與分散影響來減少衝擊，這正是強化永續的重要一環。

對台灣而言，主權 AI 除了資料議題，還包括基礎建設、資料中心、運算能力與演算法。而資料又分成商業及規範兩大層面。

商業面上，若主權 AI 僅限於資料庫建置，企業勢必需各自投入，培養專屬核心資料庫並自行訓練模型。而政府則會有對「正體資料」的需求，這涉及價值、文化與各種要素，是科技自主權的核心關鍵之一。

而規範面則可借鑑國際經驗，為不同類型的資料建立相應規範。以個資為例，台灣憲法法庭於 2022 年裁定（憲判字第 13 號），要求台灣設立獨立專責機構進行管理。但更龐大的資料，如語言模型訓練所需資料庫，雖非個資，卻涉及著作權、價值判斷與語意詮釋等問題，同樣需要法規支撐。

歐盟已透過「資料治理法」（Data Governance Act）與「資料法」（Data Act）提供完整政策框架；相較之下，台灣目前的「資料治理」措施多集中於政府開放資料，範圍仍過於有限。理想的做法，應建立涵蓋所有類型資料的完整治理架構，並在立法時避免過度細碎，以兼顧彈性與實用性。

又或者如日本文化廳近期發布的「人工智慧著作權檢核清單與指

引」，針對生成式 AI 的訓練過程提出具體規範：

「合理使用範圍：依據「著作權法」第 30 條之 4，若以收集資料供人工智慧學習為目的，可進行作品的複製，而無需取得權利人授權。」

然而，該指引也明確劃出紅線，列舉三種會構成「享受」的例外情況，以防止著作權濫用：

為輸出 AI 學習資料中包含的既有作品內容，而進行額外學習；

為使 AI 產生學習資料庫中既有作品的創作表現；

對特定創作者的少量著作進行額外個別學習。

日本此一「原則開放，例外限制」的立法思維，體現了以「資料治理」為中心的務實態度。在鼓勵產業創新與保護創作者權益之間取得了平衡，為台灣未來的制度設計提供了極具價值的參考方向。

詹婷怡強調，台灣需要一個完整的資料治理框架，不應僅限於政府開放資料，需要涵蓋各類型資料。由於資料常涉及不同法律保護，例如著作權與隱私權，特別是著作權部分，究竟是否需要因應 AI 語料訓練進行修法，或是維持現有合理使用規範，在框架之下就可以積極地持續討論。

規範並非限制，而是形塑產業樣貌及成長的基礎

「沒有規範，產業無法成形，」詹婷怡強調，規範不應被視為限制，而是產業發展的基礎條件。所謂規範，不僅限於立法，也包括行業準則、技術標準與自律指引。這些制度性安排是形塑（shape）或重塑（reshape）產業樣貌的關鍵力量。

她認為，AI 治理必須被提升至公共建設與國家戰略的層次，涵蓋基礎建設、教育政策，甚至外交布局。AI 已不僅是日常應用，更將深刻改變未來溝通方式，並與地緣政治、外交緊密相連。台灣之所以具備持續的戰略價值，正因其掌握晶片與半導體等核心基礎建設；在此基礎上，更應將 AI 治理視為公共建設的一環，並納入教育體系與人才政策。

產業觀點

- 放大並加快，AI 風險治理也須快速迭代
- 醫療 AI 的挑戰，永續及醫病關係才是核心議題
- 從數位轉型到 AI 落地，可信任與韌性是重中之重

鴻海研究院執行長兼資訊安全研究所所長
李維斌

“

AI 並非顛覆既有原則，而是改變執行速度與方式。原則與框架依舊，只是工具不斷進化。

核心基礎知識依然重要，但若跟不上新工具的節奏，就會被淘汰，AI 所帶來的資安挑戰正處於快速演進之中，答案不是單一的終點，而是一條需要不斷探索的道路，方法與知識結構必須持續更新。

”

放大並加快，AI 風險治理也須快速迭代

人工智慧（AI）正以前所未有的速度滲透醫療、金融、製造與公共治理等領域，並逐漸成為基礎設施的一部分。然而，AI 的廣泛應用不僅放大了既有的資安挑戰，更衍生出多元且複合的新型風險，且其發生與擴散的速度亦同步加快，使得防禦與治理的反應時間必須大幅縮短。這也意味著傳統的風險應對機制已不足以因應，同時，也更深刻牽動組織文化、政策制定以及全球競爭格局的走向。

從 AI 導入到 AI 治理

鴻海研究院執行長兼資安所所長李維斌指出，企業在推動 AI 導入之前，首要工作是明確界定 AI 要解決的問題。若問題本身未被釐清，極易造成期待與實際成效之間的落差。常見狀況是，管理階層對 AI 抱有高度期待，希望以 AI 解決特定難題；但企業的技術基礎或 AI 準備度不足，最終結果往往不如預期。

人才能力同樣是成敗關鍵。即便解決方案已存在，若企業缺乏能理解、應用與推動的人才，AI 導入也可能停留在表層嘗試，無法真正落地。

另一個常見迷思是將 AI 視為「萬能解答」。雖然使用 AI 的門檻已逐漸降低，例如透過 ChatGPT 等通用大型語言模型即可快速上手，但這些模型並未真正理解企業情境，其答覆是依賴機率運算，結果常常前後不一致，更可能不符需求。

因此，若企業期待導入真正貼合需求的 AI 模型，往往意味著必須推動高度標準化，要求員工依循統一規範執行。這也說明 AI 導入與組織轉型息息相關。對於已具規模的企業而言，關鍵問題在於：AI 能否融入既有流程，而非全面翻新？是不是讓使用者和組織耗費更多成本但效益不明？

這些實務挑戰，不僅關乎企業內部的 AI 政策，也呼應更宏觀的 AI 治理議題；而當前國際治理趨勢的轉變，則突顯實務中面臨的矛盾：一方面需要足夠的規範來防範風險，另一方面又必須保有創新的彈性。如何在安全與創新之間找到平衡，成為企業與政策制定者共同面對的核心課題。

從安全到創新 設立規範仍不能避免問題

李維斌舉例，最近川普政府將原本的「AI Safety Institute」改名成「Center for AI Standards and Innovation (CAISI)」，這不僅是名稱的調整，更象徵治理思維的轉變。

「Standards and Innovation 並不只是單純的安全治理，而是把『標準』和『創新』放在同等重要的位置，」李維斌指出。這種模式傾向於「跟著技術走」，在標準化過程中推動創新，而不是先行設限。不過，它仍強調透過標準建立共識與測試機制，以確保安全不被忽視；相對而言，Safety Institute 的思維更偏向於「安全優先」，著重在事前訂定規範，先劃出邊界再推動技術發展。

他進一步解釋，以往談創新（Innovation），常被理解為「快速往前衝」，但現在更需要人們能跟上這股速度。以設計流程為例，首先要先界定要解決的問題，接著一定會有資料（Data），而能真正掌

控的其實就是資料。當資料進來後，便會轉化為模型（Model）；模型建立完成後，就必須進行測試（Test Model），透過不同的資料集（Dataset）去檢驗模型是否會產生偏差或不合理的行為。

他指出，以前法律往往會事先把規範設計好，要求依照固定流程執行；但現在的做法不同，不是完全沒有規範，而是更強調在過程中，每一個環節都必須檢查與管控。規範當然可以訂定，但即使制定了，也不代表能徹底避免問題。

以 AI 的 「可解釋性」 為例，其實務操作仍相當模糊。通常只能說明 AI 的決策是根據哪些元素產生的，但這些元素如何被運用、以及判斷的邏輯過程，往往難以釐清。也因此，事前能夠落實的關鍵，便是強化資料治理：確保資料合規、取得使用者同意，並盡可能降低潛在偏誤。

李維斌認為，由於法律和規範往往落後於技術發展，若過早設下框架，容易抑制創新；但若完全跟隨技術，則可能缺乏前置規範來防範風險。對台灣而言，是選擇「規範先行的安全路線」，還是「標準化與創新並行的路線」，將是關鍵的戰略抉擇。

從政策框架到實務操作，AI 帶來的挑戰已經在各個層面上交織出更複雜的治理需求。若涉及組織所面臨的資安挑戰則又更為繁複。

沒有通用規範的資安 面對多維度挑戰

李維斌指出，資訊安全是一個多層次、多維度的系統。雖然目前已有眾多國際標準，但並不存在一套「通用」規範。各標準皆有其適用邏輯與特殊性，雖然在基礎管理原則上大致相同，但在實務落地時，往往因不同企業的需求與方法而有所差異，甚至引發執行疑慮，例如「控制項是否真正落實？」此外，控制措施之間也相互交織，一個資安事件往往需要多項控制手段配合，形成多樣化的組合情境。

在導入資安解決方案的過程中，企業也經常遭遇不同廠商產品間功能重疊（overlap）的問題，遠非拼圖般能精準契合。若過度依賴單一廠商，則容易陷入「lock-in」的風險。雖然部分公司（如微軟）提供看似完整的生態系統，表面上似乎「買一整套即可解決問題」，但實際情況往往更為複雜。許多企業往往是雲端服務、辦公軟體都是不同系統、再搭配不同的服務平台，甚至導入其他行政管理系統，最終形成「混搭式」的組合架構。

AI 與資安的三種面向與風險樣態

李維斌提醒，風險管理的核心在於「降低衝擊、收斂風險」。重要的原則在於，風險雖可透過制度與措施事先預防，但不可能百分之百避免。正如法律無法完全杜絕犯罪，總會有新案件或極端事件出現。因此，資安治理的重點不在於「能否完全阻止風險」，更重要的是必須準備好，「事件發生後如何應對與復原」。

然而，AI 帶來的挑戰在於速度過快，使得傳統治理框架面臨前所未有的壓力。隨著產業應用與技術架構的變遷，資安風險不僅更為複雜，也逐步牽動組織文化、政策制定與國際競爭格局。李維斌指出，AI 與資安的交互作用可從三個面向理解：

AI 作為助力：提升效率並強化防禦能力。

AI 作為威脅：被惡意利用，增強攻擊手段，加快威脅速度與規模。

AI 自身的缺陷：系統內部的偏誤與限制，若未妥善管控，可能成為新的風險來源。

人工智慧的導入也讓資安呈現出前所未有的複雜性風險。這三個面向在實務場景中，進一步具體化為以下多種新型風險樣態：

1.AI 能力不足的風險

自駕車事故即為典型案例。並非外部攻擊，而是 AI 系統無法辨識「非典型樣態」，例如白色汽車翻覆於道路上，後方 AI 自駕車因無法正確判斷，最終發生碰撞。這凸顯風險可能源於 AI 技術能力的侷限，而非單純的惡意入侵。

2. 舊問題的新變形：Prompt Injection

Prompt Injection（提示詞攻擊）並非全新議題，而是「套話」問題的延伸。它透過精心設計的指令，引導 AI 洩露不該公開的資訊。不同的是，如今 AI 建立在龐大數據整合之上，資料外洩與濫用的風險更為直接與集中。

3. 攻擊與防禦的加速競賽

AI 大幅壓縮了漏洞利用的時間，駭客能在短短時間內利用 AI 生成攻擊程式。這代表即便企業已知漏洞，也可能來不及因應。

4. 新世代風險：Agentic AI

另一方面，Agent AI 可能帶來更大的風險。這類 AI 不僅能提供資訊，還能自主觸發行動。傳統攻擊多發生在「決策點」（例如使用者是否點擊連結），但 Agent AI 可直接執行動作，使每一個節點都可能成為新的攻擊入口。

目前針對 Agent AI 的安全研究還不多，而它的確是新的攻擊面。只是目前大家在談論自動化與便利時，往往忽略了這部分的風險。

縱深防禦依舊重要

在 AI 時代，攻防雙方的優勢此消彼長，但這並不意味著防禦毫無意義。李維斌指出，縱深防禦（defense-in-depth）仍是有效策略。正如小偷雖然能撬開一把鎖，但若有五道關卡，就必須逐層突破；對防禦方而言，不需要每一關都成功，只要其中一道能夠攔截，就能避免嚴重的後果。關鍵不在於能否完全消除風險，而是如何建立足夠的縱深與韌性，以確保即使面對快速變動的威脅，仍能維持安全與信任的基礎。

這樣的思維不僅適用於防禦策略，也反映在營運持續計畫（BCP）的轉型上。

AI 時代，變與不變的資安原則

李維斌提到，過去的 BCP（Business Continuity Plan，營運持續計畫）多半能以文件形式存在，依循既定流程逐步執行；但在當前快速變動的環境下，這種模式已難以應付。如今，BCP 不僅需要「寫在紙上」，更要「記在腦中」，成為組織成員能隨時啟動的直覺反應機制。這也意味著，企業必須掌握一套完整的方法論，並具備靈活運用的能力。一旦資安事件發生，能即刻啟動應變機制，迅速選擇合適的手段，把衝擊與損害降到最低。

事實上，AI 帶來的挑戰並非顛覆既有的原則，而是改變了執行的速度與方式。李維斌強調：「心法沒有變，但速度變快了，所以反應也必須跟著加快。原則與框架依舊存在，只是工具不斷進化，且帶來新的執行方式。」

他以農業耕作為例進一步說明。春耕、夏耘、播種、除草，這些農作都是必須要做的基本事務，但使用的工具卻歷經多次轉型：從牛耕到農耕機，再到今日的無人機。每一次工具的更新，都要求人們學習新的操作與維護方式；在牛耕時代，重點是如何飼養與管理牲畜；進入機械化後，則需要掌握機械維護知識；到了無人機階段，又必須具備全新的知識體系。

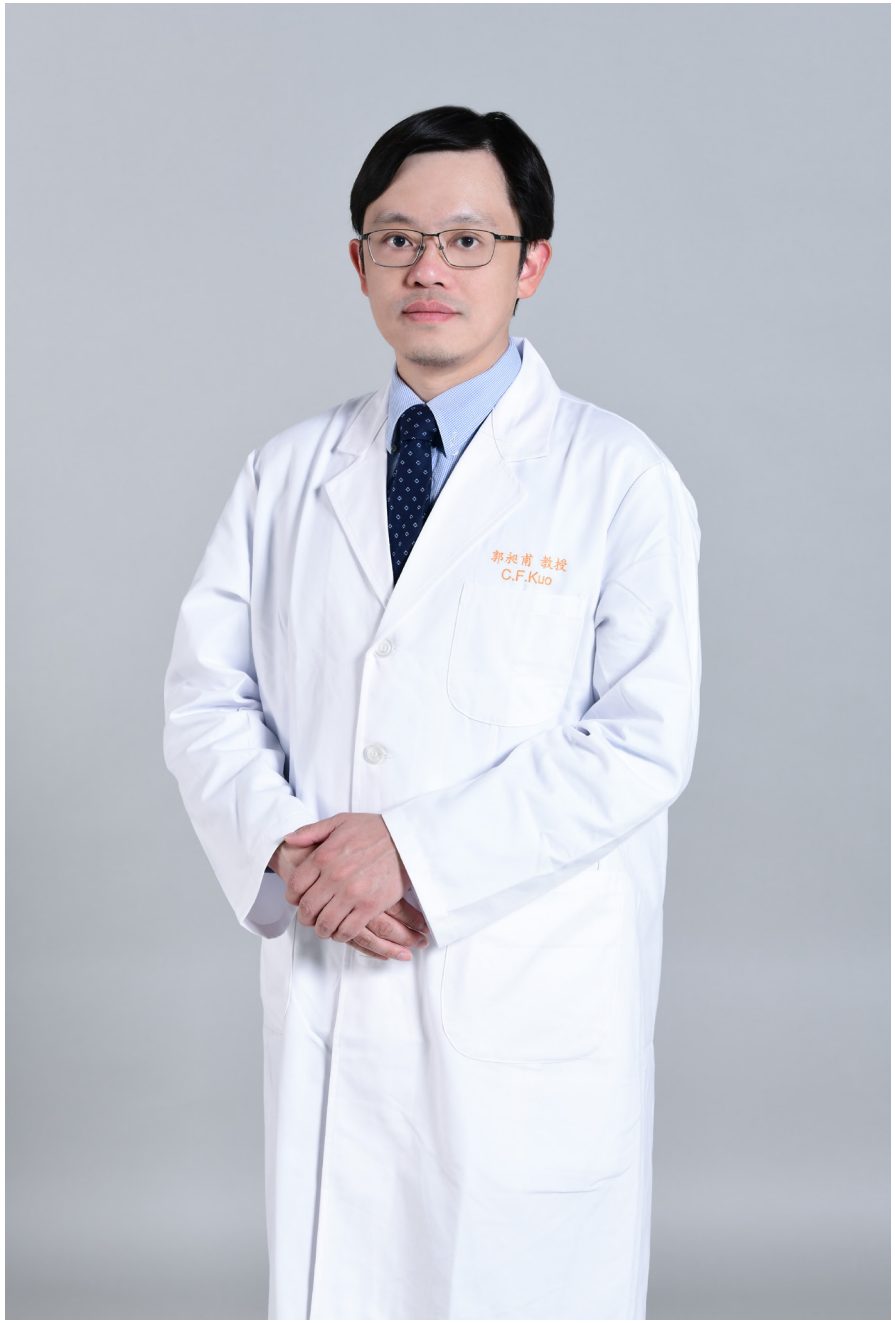
「核心基礎知識依然重要，但若跟不上新工具的節奏，就會被淘汰，」他強調，AI 所帶來的資安挑戰正處於快速演進之中，答案不是單一的終點，而是一條需要不斷探索的道路，方法與知識結構必須持續更新。這就是 AI 時代最大的挑戰與啟示。

長庚醫院 AI 核心實驗室主任

郭昶甫

醫療 AI 的挑戰不僅止於資料治理與基礎設施的更新維護。核心的問題，仍回到「醫病關係」本身。

AI 的價值在於讓醫護人員能將與病人互動的能量，從原本的 10 倍放大到 100 倍。所以，互動能量絕不能是零。AI 可以是輔助與放大器，但醫療的本質仍然是「人與人」的連結，而非單純的「人機協作」。



產業觀點

醫療 AI 的挑戰 永續及醫病關係才是核心議題

近十年來，智慧醫療始終被視為台灣最具發展潛能的領域。在成熟的醫療體系與豐富的臨床量能基礎上，社會普遍期待人工智慧（AI）能加速催生創新應用，並有效轉化為臨床價值。

然而，AI 進入醫療體系後的挑戰，往往不在演算法精度，而在更根本的結構性問題：制度如何設計？長期維運誰來承擔？模型又如何在人口與臨床環境大幅變動同時避免偏移？此外，倫理價值與創新技術間的可能矛盾，以及財務、人力維運的壓力，也是不容迴避的課題。若缺乏治理框架與穩定資源，再先進的應用也可能只是短期示範，反而侵蝕信任，失去長遠動能。換言之，醫療 AI 的關鍵已不在「能否實現」，而在「能否長期維持」。

思維衝突：長期穩定醫療模式 s. 快速迭代軟體思維

林口長庚醫院醫療人工智能核心實驗室主任郭昶甫表示，目前醫療 AI 的落地可分為三大面向：其一是資料治理，涵蓋合規使用、整理與訓練，最終進入臨床試驗並取得核可；其二是臨床部署，確保模型能順利嵌入既有流程與分工；其三則是持續監測，確保模型表現符合預期，並及時啟動修正。

第三層面「持續監測」（continuous monitoring）至關重要。他進一步強調，因為模型從訓練到部署往往歷經數年，臨床環境與病人結構在此期間早已發生變化，若無相應機制，模型極易出現偏移與失效。

目前，美國 FDA 已有相關監管框架，允許開發方在產品上市後依監測結果進行合規的調整與優化；相較之下，台灣尚缺乏對應制度。郭昶甫指出，無論台灣或美國，醫療器材（medical device）監管規範的核心，仍舊依循「硬體醫材」的邏輯，強調長週期、低變動與穩定性。這與軟體技術的高頻更新、快速迭代特質存在先天矛盾。因此，在現行醫療監管法規下，也較難接受持續頻繁更新修正的軟體節奏。

他認為，面對這一現實，唯有回到政策層面協調才是解方。相較於一般軟體產業早已習慣每月甚至更頻繁的更新節奏（如行動裝置系統常態性更新），較為固定的醫療領域法規，本質上而言並無法複製同樣頻率。儘管如此，建立一套可隨臨床環境變動而持續調整的機制，仍應是努力推動的方向。

目前，衛福部已透過成立「負責任 AI 執行中心」、「臨床 AI 取證驗證中心」及「AI 影響性研究中心」三大中心，解決 AI 醫療應用中的「落地」、「取證」與「給付」三大關鍵問題，還有 FHIR 電子病歷推動專區等專案，推動跨院一致性與監測機制。這些計畫在治理架構上確實有所進展，並要求各醫院建立監測機制。然而，雖然制度雛形已現，另一個根本問題卻始終懸而未決：永續性。

資源與責任未定，醫療 AI 永續性問題未解

政府近年透過專案計畫推動跨院合作與治理，但計畫本質上具有期限，難以支撐 AI 系統所需的長期運作。從治理實務來看，政府希望透過制度設計啟動推動力，期盼醫院在框架建立後能自行營運；然而，臨床端與產業端的核心關切卻更為務實：醫院關心的是效益是否可見、營運是否可維持；支付端與廠商則著眼於是否有穩定回收。

「計畫總有結束的一天，但 AI 需要持續的監測與維護，」郭昶甫直言，醫療 AI 的長期維運涉及資金、人力與基礎設施等龐大成本，以醫院為例，導入 AI 往往需要額外建置運算設備、配置專人管理與維護，但回報卻不明確。若缺乏直接收益，醫院就難以形成足夠動機。對廠商而言，若沒有穩定的回收機制，投入上市後監測與持續維運也缺乏誘因。

而這些現實考量，又對醫療 AI 落地帶來哪些考驗？

郭昶甫坦言，目前長庚醫院約九成 AI 模型由實驗室或院內 PI 自研、一成外購，醫院同時扮演供應者與使用者。此一路徑雖能因地制宜，卻暴露另一個實務困境。相較於硬體醫材有清楚的流程與規範，AI 模型因缺乏標準化，導入後必須依靠專責團隊「安置」與調整，才能確保其與院內系統與流程協同運作。

此外，多數廠商主張院內系統整合由醫院自行負責，使得醫院不得不組建專責團隊，處理模型上架、下架、資料流動與權限管理等工作。當模型數量從少數幾個擴展到數十、上百個時，版本管理、套件相容與資源排程便成為沉重負擔。這不只是技術問題，更是日常營運的實際挑戰。

醫療體系的超載與誘因不足

實際營運上的沉重負擔，某種程度也解釋了為何台灣的智慧醫療前景雖被談論多年，但若以韌性與治理的視角檢視，真正的挑戰才正要浮現。更現實的問題是，多數 AI 應用尚未為醫院帶來可見的新價值，這也引發一個核心問題：既然如此，導入 AI 的驅動力究竟從何而來？

「醫療的核心從來不是新技術本身，而是醫護人員與病人之間的互動。」郭昶甫直言，以心電圖為例，這項在 1900 年代初期即發明並獲得諾貝爾獎的技術，沿用一百多年，仍是臨床判斷最基礎且可靠的工具。這說明了即使新科技持續湧現，醫療體系能不能吸納，關鍵不在「新不新」，而在是否真正改善照護流程與結果。

「我們總是想把新東西塞進去，但醫療體系早已超載。如果沒有額外資源，很多應用根本進不來。」郭昶甫直言，例如，雖然每年可累積數十萬張心電圖數據，但即便導入 AI，醫院實際收益卻難以量化、醫師收入也未必因此增加，對於病人產生的往往是間接效益（例如風險降低），缺乏能即時衡量與補償的機制，自然難以形成投入誘因。

這與國際市場形成對比。美國一家專注心電圖 AI 的公司，能吸引近 1 億美元投資，正因投資人看見未來市場潛力，並相信可透過規模化、市場化回收成本。反觀台灣，因醫療市場規模有限、支付機制保守，即便技術同樣成熟，也難以形成相同的商業吸引力。

郭昶甫認為，推動醫療 AI 的關鍵不只是技術突破，而是能否創造足夠的誘因與動力，讓整個體系願意持續往前走。無論是提升醫療效率，還是帶來病患的實際效益，這些影響都必須是具體且讓利害關係人有直接感受，否則很難真正融入臨床系統中。

實際上，醫療體系的本質是一個「支出單位」，無論在台灣或其他國家皆然。醫療服務並不具備自主創造收入的能力，其資金來源主

要依靠保費與稅收，而非市場導向的營收模式。因此，若要推動醫療 AI，選項不外乎兩種：一是仰賴社會增加保費與稅收投入；二是設計能讓 AI 應用自我產生價值的制度，以支撐長期發展。

郭昶甫提到，對臨床與研究人員而言，投入 AI 的驅動力往往來自於新技術、新應用能被開發與實驗，探索創新本身就是一種研究動能。但若從政策與治理的角度切入，邏輯卻截然不同。以衛福利的三大中心計畫為例，雖然形式上與研究計畫相似，但其核心其實在於法規、治理與制度設計，這些往往不是研究者的優先關注點，也因此形成推動上的落差。

他強調，醫療 AI 最終必須成為醫療體系的內建系統，而非額外附加或補充性的外掛。唯有如此，AI 才可能真正融入臨床，發揮長期價值；若仍停留在外掛角色，就很難成功。

醫療 AI 的核心挑戰：醫病關係才是重中之重

醫療 AI 的挑戰並不僅止於資料治理與基礎設施的更新維護。更核心的問題，仍回到人與人的關係，尤其是「醫病關係」本身。這不只是醫師與病人的互動，而是涵蓋護理師、管理人員與整個醫療體系的協作網絡。當 AI 導入臨床時，對這些現有角色將帶來何種影響？又會遇到哪些阻力？這正是 AI 能否真正落地的關鍵試煉。

「你喜歡和 GPT 對話，還是更傾向與真人交流？」根據美國的一項調查，如果受訪者在「不知情對方身份」的情境下作答，高達六成的人更傾向選擇 AI。原因在於，AI 的回饋往往展現出更高的耐心與溫暖感，因為它被設計與訓練成如此。郭昶甫強調，這樣的結果提醒我們，AI 的引入並不是要取代人，而是要放大醫護人員與病人之間的溝通效能。

「AI 的價值在於讓醫護人員能將與病人互動的能量，從原本的 10 倍放大到 100 倍。所以，互動能量絕不能是零。」他指出，在醫療現場，「人性」必須始終處於核心位置。AI 可以是輔助與放大器，但醫療的本質仍然是「人與人」的連結，而非單純的「人機協作」。

郭昶甫也觀察到，台灣的病人對於「真人」的期待普遍高於國際水準。國外常見的「訪視前問卷」或「由 AI 進行初步問答」機制，在台灣推動的阻力更大，因為患者普遍期待直接面對醫師或護理人員，以獲得即時、可感的交流與安定感。若 AI 的應用被視為減少人與人的互動，反而可能適得其反。

因此，他建議，AI 在台灣的醫療體系中更適合做為「減少行政或流程負擔的工具」，而非減少醫病互動頻率的替代品。唯有如此，AI 才能真正融入現場，並成為提升醫療品質與效率的助力，而非阻力。

中信銀行數位科技處處長

王俊權

“

可信任的核心，從建立共通的協定（Protocol）開始。若協定未能先行確立，後續的授權與治理便無從談起。

協定的設計必須兼顧外部環境與產業變動。若閉關自守，最終將被市場淘汰；若完全缺乏規範，則容易陷入混亂。因此，制定協定的過程必須秉持「有所為、有所不為」 的原則。

”

智慧金融以信任為前提，可信任 AI 需以治理為基礎

金融業向來高度依賴信任與合法合規，在這樣的環境中，已經建立相對完整的資料治理與法規基礎，因而在導入 AI 時具備一定的優勢。不過，面對新科技的快速迭代與使用者期待不斷提升，使得競爭愈發激烈，同時意味著金融機構必須承擔更高標準的挑戰。隨著 AI 浪潮興起，生成式 AI、機器學習與智慧代理等技術已廣泛應用於風險管理、詐欺偵測、客戶服務、投資分析與合規監測。隨著應用不斷深化與規模擴張，AI 系統的不透明性、偏差風險、資料隱私與資安威脅日益凸顯，使得「可信任 AI」與「韌性」成為金融業推動 AI 發展時必須正視的核心議題。

從風險視角切入，決勝 AI 的挑戰絕不僅只於技術，能否建立系統性的治理機制、確保營運安全，以及有效回應法律與倫理挑戰，更是重中之重。

從舊問題到新價值

金融業導入 AI 的腳步不斷加快，細究現況，許多應用仍著重於「解決舊問題」。中信銀行數位科技處處長王俊權認為，解決舊問題是服務與技術持續推進的必然過程，且隨著規模化效益的放大，其價值會持續增長。當然新技術會造成新問題，因考量敏感性與成熟度，多數尚未對外公開討論或發佈。

同時，營運具前瞻性的金融機構，則嘗試以不同的思維重新定義舊問題，並在更高階的技術支持下，提升處理速度與擴大影響範圍。以支付創新為例，過去多著重於 QR Code 等工具，而未來若能結合 AI，將有機會實現個人化、動態化的金融服務。

他認為，創新必須循序漸進，並且按照「價值」做為區分，從內部流程最佳化的「價值一」起步，邁向客戶體驗的「價值二」，最終才能實現商業模式革新的「價值三」。然而，台灣金融市場規模有限是既成的現實與限制，根據這幾年的觀察與實際導入經驗，王俊權和幾家同業都發現，若各家金控重複投入研發資源解決相同問題，將難以形成產業規模化效應。因此，政策推動應著重於跨金控的合作機制，鼓勵金融機構透過「共創、共益、共享、共治」的方式，建立聯盟式合作，避免資源分散，加速可信任 AI 的落地。

可信任 AI 的具體化：「守護科技」與防詐金流履歷模型

過去對「可信任 AI」的討論往往流於抽象，或者只是一種尚未實踐的技術理想。但其實若應用到詐騙防制場景時，非但必要性便更為清晰，也更能釐清可信任的價值。也就是說，AI 一旦深度嵌入銀行體系，並直接涉及交易安全與詐欺偵測，「可信任」不再只是理念，而是關乎金融體系穩定與用戶信任的核心基礎。

王俊權指出，金融業的運作本質與生存關鍵在於信任。若客戶在使用數位金融服務時缺乏安全感，即便介面再便利，體驗仍將大打折扣。同樣地，人工智慧的服務若沒有信任，使用者也不會使用。

中國信託銀行近年來大幅強化科技在防詐領域的應用。王俊權指出，過去金融科技多著重於數位金融、普惠金融與客戶體驗的優化，如今中信則進一步將科技應用延伸至資安防護、反詐騙、防制洗錢（AML），甚至涵蓋內部員工的安全守護，逐步構築出更全面的「守護科技」體系。

在具體實踐上，中國信託銀行攜手多家同業及資策會，推動「金流履歷」專案，建構「大金流模型」（Large Cash Flow Model, LCM），並將「認識你的客戶」（Know Your Customer, KYC）進化為「認識你的金流安全」（Know Your Cash Flow, KYCF）。

王俊權進一步解釋，現行法規下銀行間金流資料無法互通，詐騙集團便利用不同銀行間的「斷點」進行層層轉帳與洗錢，使追查難度大幅增加。該模型的核心在於透過串聯碎片化的交易資料，重構完整的金流網絡，從而大幅提升詐騙行為的偵測與關聯能力。

在概念驗證階段，邀請中信、郵局與玉山銀行等三家金融機構及資策會率先合作進行試驗。採用雙盲加密與無塵實驗室技術，確保資料脫敏無法還原。結果顯示，第一階段即便僅有三家資料輸入，模型已能標註高風險帳戶示警原本不可見的金流關聯，成效顯著。後續再進一步擴大至 8 家銀行，促成同業間及時聯防，阻擋可疑交易。

下一階段工作是將現有的聯防與通報機制，從被動揭露提升為主動預測及防護。為此則邀請八家涵蓋市場六成以上交易樣態的銀行，合作建構具代表性的「泛化大金流模型」。此舉確保模型具備跨客群與業務的泛化能力，並能在市場層級展現準確性與有效性。

王俊權表示，該計畫將於年底前完成模型訓練，並提供給銀行公會，由財金資訊公司負責後續運轉，建立全金融體系共享的防詐偵測基礎，進一步實現「共創、共益、共享、共治」的合作目標。

AI 落地的關鍵：可信任 AI 的治理基礎

「判斷企業是否真正推動 AI 發展，要看它是否認真落實 AI 治理。」王俊權指出，可信任的核心，從建立共通的協定（Protocol）開始。共通協定指的是一套 所有參與者必須遵循的統一規範，讓不同系統、不同單位或不同國家在使用 AI 時，可以在同一個「遊戲規則」下協作與交換資料。若缺乏統一規範，不同系統之間將因知識深度與邏輯差異而難以互通，最終導致治理機制失效。當前各界雖積極發展 AI，卻普遍欠缺一致的約定。若協定未能先行確立，後續的授權與治理便無從談起。

當前各界雖積極投入 AI，但普遍對於共通協定的認知與導入尚有不足。應該如何著手？王俊權指出，協定必須優先被明確界定，例如：哪些資料需要回流、哪些應加以統一、哪些必須受到嚴格控管。唯有在這樣的基礎上，授權與治理才有落實的可能。

他同時也提醒，協定的設計必須兼顧外部環境與產業變動。若閉關自守，最終將被市場淘汰；若完全缺乏規範，則容易陷入混亂。因此，制定協定的過程必須秉持 「有所為、有所不為」 的原則。

以中信的內部推動經驗為例，即透過四種「選型」政策來界定應用範圍。雖然各家金融機構的業務需求不盡相同，難以完全一致，但王俊權的選型哲學在於 「不單押、能規模化」。所謂不單押是指避免依賴單一供應商，確保多平台並行的彈性；能規模化則是確保創新應用具備即時監控、防護與可擴散性，避免失控風險。

換言之，透過多元選擇與統一協定，才能確保 AI 治理在快速變動的環境下，仍具備彈性與韌性，並推動可信任 AI 的真正落地。

另一方面，隨著產業數位轉型進程加速，AI 治理的重要性日益浮現。王俊權進一步指出，AI 治理的重要性在於必須自上而下（top-down）清楚定義企業真正要解決的問題；更涉及基礎建設的策略，企業必須依據需求評估自身的運算架構是以「全面雲端」、「雲地混合」抑或「以地端為主」，並進一步思考如何最佳化投資，使算力與資源能符合長期的發展需求。

最後，協作機制是治理的另一個核心。無論是內部員工、外部廠商，或是不同的 AI 平台與模組，如何透過協定（Protocol）將多元的資源、使用者與應用情境整合起來，形成一個如同「交響樂團」般的協作系統，將決定治理能否真正落實。如同地圖系統能將龐大的地圖資訊與數據點有序串聯，AI 的治理同樣需要一套統一協定，確保不同來源的數據能共用、共識與共治。

從垂直產業模型到跨業協定

以中信目前推動的兩大跨業專案為例——「防詐金流履歷」與「金融專業大型語言模型」（FIN LLM）——可以觀察到金融業在 AI 發展中，正逐步邁向「垂直產業模型」的雛形。透過跨機構的共創與共用，不同銀行與金融單位開始嘗試將自身累積的數據與經驗，轉化為可共享的模型基礎。然而，這樣的模式能否真正落地，仍有賴於治理機制的健全，特別是是否能建立一套跨機構皆可遵循的協定。

王俊權指出，在防詐金流專案中，如同前面所提到的協定，首先會體現在金流 API 的固定化。透過統一規格，每一筆上傳的金流資料都能遵循一致標準，進而串聯甚或泛化成不同銀行的數據。藉由這樣的協定設計，AI 模型得以藉由數據驅動（data-driven）的跨行模擬與關聯分析，有效提升詐騙偵測的精準度。

另一方面，FIN LLM 的推動則展示了垂直產業模型的另一種可能。王俊權觀察到，全球的大型語言模型發展已呈現出一個明顯的趨勢，最初以通用型模型為主，例如 Llama 3 或 OpenAI ，隨著開源與公開資料逐漸被充分利用，通用模型的發展已經面臨限制，下一步便是朝向縱向產業資料的應用。

這也代表了使用場景正逐漸從「通用場景」走向「產業專用場景」，再深入到「場域專用」。可以看到全球主要的技術供應商（如 NVIDIA）也開始積極探索產業專屬的縱向模型。而金融業因擁有完整且不外流的金融數據，正是縱向模型的最佳試點。

他進一步說明，FIN LLM 的設計建構在三層協定之上：一是蒐集並授權大學層級的經濟學、統計學與金融教材，讓模型具備如「大學生等級」的基本理解力。二是透過金融研訓院提供的歷屆考題，將題庫蒸餾成標準化數據，使模型具備應試與決策能力。第三層則是納入金管會的裁罰案例與專用語彙，避免模型在專業情境中出現誤判或「幻覺」。

藉由這三層協定，FIN LLM 不僅強化了專業理解能力，也將金融產業的共通知識轉化為標準化規範，為可信任 AI 的落地提供了治理基礎。王俊權說，當協定建立形同「書同文、車同軌」的制度後，垂直產業模型才有機會成熟落地。然而，金融業之所以能先行，是因為其資料完整、制度成熟。然而，這套模式是否能複製到其他產業，仍是

一大挑戰，也是台灣 AI 發展能否跨域擴散的關鍵課題。

AI 在金融業的發展過程中，信任與韌性並非兩個平行的議題，而是相互依存的雙核心。信任決定了使用與採納的可能，韌性則確保了持續性與抗風險能力。

從防詐金流到 FIN LLM，金融業已展現一條可信任且具韌性的 AI 發展路徑。下一步的挑戰，是如何將這套模式推廣至更多產業，構築跨域共享的治理基礎，讓「可信任 AI 與韌性」成為台灣產業的競爭力核心。

行動方案

- 從數位轉型到 AI 落地，可信任與韌性是重中之重
- 風險預防 + 快速復原，AI 治理的三大核心要素
- 確保「無偏差」，完整風險治理須靠 AI 生態系建構

思科大中華區副總裁暨台灣區總經理
林岳田

“韌性網路已不再只是 IT 議題，而是攸關企業生存的核心基礎。”

因應未來 AI 的趨勢，最重要的就是強化數位韌性。不僅是數據中心、備援中心與防火牆的堆疊，而是在 AI 與雲端複雜的環境交織下，能否以更動態、更全面的方式整合。

行動方案

從數位轉型到 AI 落地，可信任與韌性是重中之重

隨著 AI 與雲端的快速普及，企業面臨的營運風險正不斷演變與升級。根據《2025 全球網路趨勢調查報告》指出，79% 的台灣企業曾因網路攻擊或設定錯誤導致重大中斷，95% 的 IT 領袖認為韌性網路架構對營運至關重要，98% 更指出網路是 AI、雲端與物聯網部署的關鍵基礎。

這些數據皆凸顯，除了算力，可信任（Trustworthy AI）與韌性（Resilience）更是企業 AI 落地佈署的基礎。

思科大中華區副總裁暨台灣區總經理林岳田指出，當前對可信任 AI 與企業韌性的討論，正如十年前的數位轉型浪潮。當時，唯有完善的流程設計與治理機制，數位化才能真正成為企業成長的助力。同樣地，AI 雖能提升效率並拓展應用，但要讓技術真正發揮作用，仍需建立以韌性為基礎的數位體質。

他進一步說明，所謂韌性，不僅包含實體層面的基礎建設與網路架構，還涉及人員的意識、思維與技能。唯有兼顧技術與組織能力，企業才能建立真正具備承載力的韌性基礎。打造可信任 AI 的首要關鍵也並非技術細節，而是治理原則的落實。可信任的前提是在決策與運作的各環節中，都必須將安全、隱私、人權與透明性納入考量。其後，才是進一步思考如何透過技術實作與部署來落地 AI。

四大實踐方向，確保 AI 符合治理要求

為協助企業在導入 AI 時兼顧創新與治理，林岳田以 Cisco 所提出的四大實踐方向為例，首先是建立負責任 AI（Responsible AI, RAI）框架。以安全、隱私、人權與透明性為原則，確保 AI 在設計與運作中符合治理要求。

其次則是主動識別新興 AI 風險，包括幻覺（Hallucination）、偏見（Bias）、不實內容等挑戰與傷害性輸出，並將其納入風險評估流程。

接著是導入防護機制與工具，Cisco 已與合作夥伴共同開發 SaaS 工具，以協助企業在應用 AI 代理人（Agent AI）時，檢驗資料來源的可信度，確保流程合規並滿足資安需求。

最後，必須和利害關係人共同制定規範，將 AI 的責任與邊界向內部與外部利害關係人全面揭露，當中涵蓋客戶、供應鏈與合作夥伴，建立透明且可追溯的機制，明確問責並確保責任分工。

林岳田提醒，企業在建立自己的 AI 系統時，需要特別注意兩個層面：一是面向客戶的應用系統，二是與安全相關的治理系統。這也意味著，企業必須謹慎評估並選擇數據來源，因為數據的品質與透明性直接影響 AI 的可信度。

以 Cisco 日前發佈的開源安全模型（Foundation-Sec-8B）為例，這個專為資安維運而建構的特定領域大語言模型（LLM），就納入了自家與合作夥伴的資安訓練資料，以及情資報告、漏洞資料庫、事件回應文件及資安標準等高品質資安訓練資料集，並持續進行預訓練以增強能力。企業可下載並結合自身資料，在 SOC（安全營運中心）中進行應用，例如輸入伺服器、網路設備或帳號行為的 log，以預判異常狀況並提供防禦建議。

他進一步強調，可信任 AI 的根基在於資料來源的安全、合規與透明。若缺乏這樣的基礎，再強大的算力與模型都難以支撐長期發展。

網路韌性將是 AI 時代的企業基石

許多人談到 AI，往往第一時間聯想到 GPU 算力，認為速度越快就能更有效地訓練模型。然而，林岳田強調，AI 真正要能落地，絕非僅靠算力。企業還需要具備完整的流程設計、清晰的治理機制、利害關係人的廣泛參與，以及能支撐整體運作的基礎建設。特別是高度韌性的網路與數位架構，更是關鍵支柱，這一點在過去多份研究與報告中皆被反覆驗證。

這樣的觀點，也從近期各項研究報告中獲得印證。根據「2025 全球網路趨勢調查報告」，79%的台灣企業曾因網路攻擊或設定錯誤導致重大中斷，78%預期未來兩年仍可能因資安事件受阻。這顯示，韌性網路已不再只是 IT 議題，而是攸關企業生存的核心基礎。

79%

台灣企業曾遭遇重大網路中斷事件

79% 台灣企業曾因網路攻擊、壅塞或設定錯誤導致重大中斷，顯示在數位轉型過程中，企業對強化網路安全與韌性的需求日益迫切。

同一份調查也指出，95% 的 IT 領袖認為韌性網路架構對營運至關重要，93% 的決策者計畫增加相關預算，98%則認為網路是 AI、雲端與物聯網部署的關鍵支撐。換言之，沒有穩固的網路，就無法承載未來龐大的算力需求。

「然而，企業往往低估了 AI 帶來的資安風險，」林岳田進一步提醒，因應未來 AI 的趨勢，最重要的就是強化數位韌性。不僅是數據中心、備援中心與防火牆的堆疊，而是在 AI 與雲端複雜的環境交織下，能否以更動態、更全面的方式整合。

企業在導入 AI 前，必須先釐清「為什麼要做 AI」，在此基礎上納入公平、隱私、透明等治理原則。不同產業有不同需求架構，例如，半導體的關鍵系統不上雲，金融業則受法規限制，資料保護有嚴格標準，因此不同產業在推動 AI 時，其架構形態必然有所不同。但不論選擇集中式或分散式、自建或上雲，核心都是確保基礎建設與網路具備韌性，能夠動態因應不同情境，並兼顧資安。這也是為何在高度複雜的環境中，統一的管理平台格外重要。

林岳田指出，Cisco 近年透過收購讓服務提供版圖更完整，打造能整合自家、合作夥伴乃至競爭對手系統的大型資料收集平台。這讓企業在 AI 導入過程中具備更高的可視性、可靠性與管理效率，並能進行跨平台的數據治理。

他同時也看到數據中心面臨轉型的轉捩點。過去金融機構慣用的備援中心模式，受限於頻寬與成本，難以全面推行。進入 AI 時代後，市場開始談論 「AI Ready Data Center」，其關鍵不僅在於算力，更在

於能否承受高耗電、提供足夠傳輸頻寬，並具備可靠的安全機制。

雲端的普及亦改變了規則。主管機關逐步允許企業將資料上雲，減輕自建機房壓力，但也帶來多雲環境的整合與治理挑戰。林岳田強調，從數據中心到傳輸層，企業必須以 端到端思維 強化基礎建設，確保「高可用性、可預測性與安全性」。唯有如此，才能在龐大算力需求與資安挑戰下，維持長期營運韌性。

Cisco 的 AI 生態系與顧問式服務

林岳田認為，AI 的生態系涵蓋算力、網路到顧問服務，版圖相當龐大。

Cisco 透過與 NVIDIA、AMD、Intel 等晶片大廠保持緊密合作，並確保其產品可與 Cisco 設備無縫整合，減輕客戶多平台管理的負擔。同時，Cisco 也自研高速網路與晶片，並在近兩年積極發展光纖技術（ROADM Optical Network），逐步構築從網路、晶片、光纖到資料平台的完整基礎建設。

除了產品與基礎建設，Cisco 也提供顧問式服務（Consulting Services）。內部的專家顧問會透過實地訪談與需求評估，協助企業釐清在 AI 導入過程中的目標、所需環境與系統組態，並提供完整建議，確保 AI 建置能與企業業務目標緊密對齊。

然而，Cisco 僅憑自身之力，難以服務數以百萬計的全球客戶，因此，林岳田提到，他們透過與各領域的 SI（系統整合商）合作，例如醫療、金融、製造等垂直產業，Cisco 能結合專業知識與在地經驗，協助客戶完成 AI 導入與後續維運，並確保落地成效。

他認為，Cisco 的價值在於一站式的整合能力：從 GPU 與算力、數據中心、網路基礎建設，到雲端管理、光纖傳輸、AI 顧問式服務，以及 AI 防禦（AI Defense）與應用控管，均可由 Cisco 與其生態系提供支援。Cisco 自我定位為企業在 AI 旅程中的 「Trusted Advisor」，提供多元選擇，並允許客戶依自身體質與需求靈活擴充。

林岳田強調，從建立負責任 AI 框架、識別新興風險、導入防護機制，到與利害關係人共同制定規範，這些治理原則的落實，以及數據來源的安全、合規與透明，都是確保 AI 可信賴的關鍵。同時，面對複雜的雲端與 AI 環境，企業必須以動態且全面的方式強化網路韌性，並透過整合性的管理平台，確保基礎建設的高可用性、可預測性與安全性。唯有如此，企業才能在數位轉型的浪潮中，穩健邁向 AI 時代，並維持長期營運韌性。

網達先進科技亞洲區總裁
李崇偉

若企業已建立資訊安全管理系統（ISMS），最佳做法是將 AI 治理納入其中，而非再建置一套平行的治理機制。

AI 帶來傳統資安未曾涵蓋的新挑戰，例如偏見、透明度與可解釋性。AI 治理應被視為「既有治理框架的延伸與補充」。此一做法不僅避免重複建置流程，也能確保 AI 與整體資安治理的一致性，為企業在面對新興風險時提供更完整的保護。

行動方案

風險預防 + 快速復原，AI 治理的三大核心要素

近年來，企業積極導入 AI 以推動創新，但實務上卻常遇到「想像與現實落差」。快速應用的渴望與路徑選擇的不確定，凸顯了數據治理、資安防護與標準化部署的挑戰。唯有建立在「可信任 AI」與「企業韌性」的基礎上，AI 才能真正成為持續創新的驅動力。

國際知名資訊通信科技（ICT）解決方案與託管服務整合商網達先進（Logicalis）亞洲區總裁李崇偉指出，AI 在企業環境中的責任遠大於消費者端，部署必須確保安全性與可靠性，同時避免錯誤、偏見與不當資訊的輸出。從許多實務案例可以看到，若模型輸出有不當內容，將對企業聲譽造成嚴重傷害；若遭攻擊者操縱、惡意誘導機器人回應，甚至可能出現以「一美元賣車」荒謬價格的情況。

因此，企業在導入大型語言模型（LLM）時，需建立前置防護與驗證流程，以降低生成風險；並持續關注如 prompt injection 等新型攻擊手法，確保系統環境不被入侵。換言之，可信任 AI 強調風險預防，AI 韌性則確保問題發生後能快速恢復。唯有兼顧兩者，AI 才能成為企業穩定營運與創新的可靠引擎。

李崇偉認為，企業自身的 AI 推動方法、訓練過程，數據來源及可靠度是判斷 AI 模型完整性的重要基準。而透過同一問題，測試企業內部 AI 的輸出，並同步比較其他 LLM 的回應，藉此檢視準確度是否一致或接近等橫向對照的方式，可大致評估企業 AI 部署的可靠性。

除此之外，企業本身的背景與商業模式亦是關鍵考量。例如企業擁有者是否具備良好聲譽與可信度？其次，其技術方法是否透明，特別是在模型訓練與數據管理上的作法。最後，還需審視該公司的願景、策略與商業模式。針對免費提供 AI 服務的產品，就必須進一步追問：它如何變現？是否依賴收集與轉售用戶資料？這些因素皆決定該服務能否被視為「可信任 AI」。

AI 治理三要素 須納入整體資安管理系統

企業若要同時兼顧「可信任 AI」與「韌性」，必須在多個面向上協調推進。李崇偉認為，以下三個核心要素尤為關鍵且缺一不可，分別是可信的數據來源、完善的內部治理機制，及資安防護能力。尤其是資安防護必須與治理緊密結合，而非被視為獨立環節。唯有透過端到端的安全設計，才能降低外部攻擊風險。

那麼，企業是否需將 AI 視為與一般資安同等級的治理對象，還是另行建立一套獨立機制？李崇偉認為，AI 治理應與既有資安與治理框架結合，而非完全獨立。因為 AI 並非孤立存在，而是依附於既有基礎架構與數據環境運作。例如，若企業已建立資訊安全管理系統（ISMS），最佳做法是將 AI 治理納入其中，而非再建置一套平行的治理機制。

然而，AI 也帶來傳統資安未曾涵蓋的新挑戰，例如偏見、透明度與可解釋性。這些議題無法僅依靠既有的 ISMS 完全解決，因此需要額外補強。換言之，AI 治理應被視為「既有治理框架的延伸與補充」。此一做法不僅避免重複建置流程，也能確保 AI 與整體資安治理的一致性，為企業在面對新興風險時提供更完整的保護。

情境、分類、監控與回朔三步驟逐步建立 AI 治理機制

在實際操作層面，李崇偉建議企業可依循三個步驟逐步建立 AI 治理機制：

第一：盤點 AI 使用情境

企業須先全面掌握內部各部門、各流程中 AI 的使用現況與潛在規劃，釐清 AI 的實際滲透範圍，作為後續治理的基礎。

第二：進行風險分類

不同應用場景涉及的風險程度不一。一般客服聊天機器人屬於低風險，而金融決策或醫療診斷等應用則屬高風險，必須導入更嚴謹的治理措施。透過風險分級，企業可合理配置資源並制定差異化治理策略。

第三，建立監控與回溯機制

治理不能只有在 AI 上線前審查，更需在運行過程中持續監測，並保留可回溯紀錄，以利在異常或爭議發生時快速追蹤來源與責任。

AI 必須透過長期監控與調整，才能確保可信任與韌性

他提醒，不少企業推動 AI 時存在「一次性部署」的迷思，忽略了持續治理的重要性。AI 與資安相似，必須透過長期監控與調整，才能確保持續的可信任性與韌性。

李崇偉認為，未來五年 AI 將會更加普及，不僅大型企業會廣泛使用，中小企業甚至政府單位也會大規模導入。而在可信任 AI 的面向上，將會聚焦於透明度與可解釋性。企業與使用者都將要求 AI 的決策過程能被追蹤、理解，而不再只是「黑箱運作」。

同時，對於偏見與公平性的要求也會更嚴格。不論在金融、醫療還是教育領域，一旦 AI 被發現存在系統性偏見，將會對企業聲譽與合規造成重大衝擊。

韌性的挑戰同樣趨於嚴峻，隨著 AI 深入基礎設施與關鍵產業，一旦系統故障或遭受攻擊，影響將不僅限於單一企業，而可能擴散至整個產業鏈。因此，未來的核心任務在於建立跨產業的韌性機制，確保即使 AI 系統出現問題，整個產業乃至社會都能快速復原。整體而言，可信任 AI 將更強調透明、公平與合規，而韌性則將走向跨產業、跨國協作。這將成為各國企業與政府必須提前規劃與準備的重要方向。

多數台灣企業對 AI 認知仍待澄清

將視角轉回台灣，網達先進事業開發部副總經理張晃崚指出，在談論「可信任 AI」時，重新檢討現有資安架構是關鍵的一步。多數企業雖已採購各類安全工具，但由於缺乏整合，往往各自為政，難以發揮整體效能。因此，Logicalis 的切入重點就在於協助企業將分散的工具整合為平台化架構，藉此提升監控效率，並在風險發生時快速應變，確保治理成效與企業韌性。

然而，許多企業在推動 AI 時仍存有迷思。例如擔心資料外洩至外部模型、誤以為傳統防毒或文字檢查工具即可保障 AI 安全，或單純將自家 AI 模型與 ChatGPT、NotebookLM 等服務比較，忽略治理層面的

要求。張晃崚強調，若這些觀念未被澄清，將直接影響治理與風險控管的效能。

他進一步指出，企業普遍期待能快速採用新技術，但往往受制於缺乏標準化流程，導致陷入 POC（概念驗證）反覆循環。真正的關鍵在於建立標準化架構，讓 AI 技術的導入與擴充如同搭積木般靈活。為此，以 Logicalis 所提出的平台化解決方案為例，透過將各項技術能力整合於同一平台；透過雲端即可快速啟用 Cisco HyperFabric AI 平台，協助企業建立標準模式，加速新技術的採用與應用擴展。

然而，在企業快速導入新技術的過程中，同樣也必須面對不斷演變的法規與合規挑戰。張晃崚認為，面對正在加速收緊的 AI 監管，企業需要建立完善的資料治理與紀錄機制，確保專案具備可追溯性，並轉化為可量化的 KPI，以因應未來潛在的監管要求。

私有化 AI 平台漸成主流，將衝擊原有架構

在落地實踐層面，私有化 AI 平台逐漸成為主流趨勢。許多企業傾向自行投資算力與基礎設施，但這同時帶來架構衝擊與設計挑戰。網達先進事業開發部資深 AI 架構師徐維中指出，整體環境可分為三個層次：第一層是使用者端的應用程式，企業必須具備足夠算力，能整合資料並選用適當模型，建構易於操作的平台；第二層是機房與網路基礎架構，需面對因 AI 運算帶來的流量暴增、可視性不足與 GPU 資源調度問題，並解決資料孤島挑戰，透過資料湖整合結構化與非結構化資料；第三層則是機房環境與能源管理，因 AI 運算伺服器能耗與散熱需求遠高於傳統伺服器，企業必須重新設計機櫃、能源系統，並同時滿足 ESG 要求。

然而，徐維中直言，傳統 IT 規劃多以「峰值設計」為基準，AI 導入後的運算需求變化更大，傳統設計方式難以應對。這意味著企業必須導入更精細的容量規劃與即時調整機制，以滿足不斷攀升的算力需求。

邁達特數位產品中心副總經理
林立棕

AI 的可信任性與韌性，已經不只是單一企業內部的挑戰，而是生態系合作的課題

基礎架構只是第一步。當硬體與軟體環境到位後，下一個挑戰是「應用」，這需要另一個生態系的支持。邁達特約從三年前開始經營並積極引入台灣本地的新創，主動尋找各行各業的雲端與 AI 應用，並逐步納入 AI 新創的解決方案。

確保「無偏差」，完整風險治理須靠 AI 生態系建構

在全球人工智慧浪潮下，AI 已不僅是效率提升與創新的工具，更成為影響企業治理與社會信任的關鍵議題。這幾年許多企業積極導入 AI，但在追求短期成效時，往往忽視了更根本的課題：如何確保 AI 在長期運作中，持續達到可信任（Trustworthiness）與韌性（Resilience）。

邁達特數位產品中心副總經理林立棕表示，可信任 AI 的核心在於「能否信任其回應與結果」。這看似簡單，卻蘊含龐大的治理挑戰與落實需求；而 AI 韌性正是其中不可或缺的基石。唯有確保系統具備韌性，才能避免創新應用最終轉化為潛在風險。

實際案例已顯示其重要性：加拿大某航空公司因客服 AI 誤報票價引發法律糾紛；台灣金融業則早已意識風險，導入 AI 時要求人工覆核，以降低誤導客戶的可能性。林立棕提醒，AI 治理的核心挑戰不僅在於技術精準度，更在於如何確保 AI 在無偏差的情況下，能持續滿足企業營運需求。

避免有害內容、慎防模型被操縱

目前企業很常見的 AI 應用多為客服應答機器人，協助同仁節省回覆時間。然而，若要真正建立信任，單靠工具部署並不足夠。企業必須設計相應的機制，確保 AI 回覆不僅正確，還能避免對品牌形象與客戶關係造成傷害。

林立棕認為，AI 應用的本質和過去 IT 應用的本質邏輯是一致，都屬於「應用」，過去我們稱之為 Web Application 或 ERP 系統，如今則是 AI Application。因此，業界普遍將焦點放在兩個核心環節，分別是「資料來源是否準備完善」及「避免 AI 產生有害內容」，前者確保輸入數據完整與正確，後者則是建立可信任輸出的必要前提。

其中，「避免有害內容」尤為關鍵，因為這直接關係到企業品牌聲譽與市場信任。林立棕指出，國際組織 OWASP 長期針對各類應用程式與系統提出指引，協助業界識別潛在風險，以確保系統安全與穩定運作。近年來，OWASP 更針對大型語言模型（LLM）提出「十大風險項目」，成為企業建構可信任 AI 的重要參考依據。

對於有意自行開發 AI 模型的企業，林立棕特別提醒，必須慎防模型被操縱。這在技術上相對容易發生，因為模型會依循自身邏輯生成回應。為此，企業至少需確保兩件事：一是輸入數據無誤；二是輸出內容安全且正確。避免相關風險的常見作法包括：一是限制模型接收的指令範圍，確保其僅能處理特定需求；或是透過資料限制，讓模型只能生成或回應特定內容；再者是結合 IT 技術檢測，輔以自動化工具，協助判斷輸出內容是否存在問題。

林立棕以 Cisco 近期積極推動的「AI Defense」概念為例，其中的 Secure AI Factory 架構就強調，若缺乏資安層面的防護，AI 架構絕不可能稱得上完整。因此，在協助企業導入治理時，會先釐清其 AI 應用需求，再指出需加強的面向，並透過 OWASP 架構，說明潛在風險與預防的必要性。

資料外洩並非唯一 台灣企業風險視角仍顯狹隘

然而，他也觀察到，目前台灣多數企業在討論 AI 風險時，仍主要聚焦於「資料外洩」。許多企業雖想導入 AI，卻未能真正理解後續可能帶來的治理挑戰。無論是「韌性」還是「可信任 AI」，核心都在於降低潛在風險，但多數企業尚未具體感受到其對營運的實際衝擊。

隨著國際討論熱度升高，越來越多台灣企業開始意識到相關議題的重要性，也逐漸反思若忽視這些問題可能帶來的後果。在此背景下，企業最迫切的需求就是如何快速建構出可落地的 AI 架構。但林立棕提醒，AI 真正落地時，牽涉到軟硬體的資訊基礎建設，複雜度極高。

過去幾年常認為企業可以自行評估、比較並組合不同方案，其實這種方式不僅耗時，也增加導入風險。相對而言，一體化的「AI Ready」解決方案已成為加速落地的選擇。透過軟硬體完整整合，企業不必再花時間進行繁瑣測試，而能直接將重心放在資料準備，以及 AI 回覆內容的正確性與品質優化上。

林立棕認為，當企業在導入 AI 並推向內部使用後，最核心的工作是持續檢查與改善 AI 回覆品質。唯有透過不斷優化，AI 才能真正成為可信任且具韌性的企業工具，而非短期導入的「炫技」應用。

AI 尚未成為台灣不可或缺基礎

林立棕指出，AI 的韌性可從兩個面向來理解：基礎架構（Infra）與應用場景。以目前多數企業的內部應用而言，AI 大多只是輔助效率的工具，如果暫時無法使用，頂多回到傳統的手動流程，雖然效率降低，但不至於全面停擺。對外服務亦然，例如客服 AI 雖日益成熟，但若 AI 無法運作，客戶仍可透過 0800 專線找到真人客服，影響有限。換句話說，AI 暫時還不是企業「不可或缺」的基礎。

雖然台灣目前尚未走到「AI 出問題即重創營運」的階段，但相較之下，IT 一旦出現故障，往往會立刻影響營收與營運，因此韌性才被視為關鍵課題。簡言之，AI 對台灣企業的角色仍以「效率工具」為主，尚未成為「不可或缺的基礎」。

然而，這樣的情況在國外已開始出現變化。以 Salesforce 為例，導入 AI 後立即裁撤 4000 名客服人員。此時若 AI 客服系統出現問題，便會對企業營運造成嚴重災難，因為服務模式已完全依賴 AI。相較之下，台灣尚未走到這一步，但隨著 AI 滲透更深、導入範圍擴大，未來「AI 韌性」勢必會成為企業治理中不可忽視的關鍵議題。

與韌性相關的是，企業需要進一步從「價值」角度來審視 AI 投資。有些客戶追求的並非立即的金錢回報或效率提升，而是品牌形象的強化；也有些客戶則會嚴格要求投資報酬率，將 AI 視為必須帶來具體效益的工具。林立棕說，對於已完成數位化的企業來說，未來推進 AI 化的可能性確實更高，但這並不意味著每個部門都需要自行「養一個 AI」。許多應用其實透過既有產品的「內建 AI 功能」即可滿足需求，尤其在資安領域，「Security for AI」與「AI for Security」的應用已相當普遍。

然而，他也提醒，當企業進一步考慮自建 AI 時，就不可避免要面對「資料主權」等議題：關鍵資料是否能放心交給外部？是否具備足夠的治理與安全保障？這些問題的答案，往往決定了企業的導入模式。

也正因如此，AI 的可信任性與韌性，已經不只是單一企業內部的挑戰，而是生態系合作的課題。從基礎架構供應商、軟硬體整合商，到在地新創與產業使用者，都必須在同一個框架下協作，才能讓 AI 真正具備「可落地、可持續、可信任」的特質。這也引出了下一個核心議題：如何打造一個支持可信任 AI 的完整生態系。

邊緣運算興起，凸顯新架構的重要性

林立棕指出，導入 AI 勢必伴隨成本。現階段 AI 的投入成本仍然偏高，但這樣的情況正在快速改變。實際上，從 2024 年開始已經出現成本大幅下降的跡象。

這不僅來自開源模型的興起，更關鍵的是整體運算成本的降低。過去 AI 運算幾乎完全依賴資料中心，如今隨著邊緣運算（Edge Computing）的發展，越來越多運算可以直接在邊緣裝置上完成。不僅工業的製造場域，甚至日常生活中的終端設備也逐步具備 AI 運算能力。他認為，未來 AI 應用架構將從集中化逐步走向分散化，更多運算功能會移動到使用者端。

所以，架構轉變帶來新的挑戰：如何確保這些分散式模型與資料不會被污染、操縱或感染？正好凸顯「韌性」與「資安」（Security）的緊密關聯。林立棕強調，安全性絕對是 AI 韌性不可或缺的一環，因為風險不僅影響自身，更可能波及他人，且影響範圍往往難以預測。

他進一步指出，這樣的挑戰同樣不是單一廠商能夠解決，而必須透過「AI 生態系」的力量來因應。例如 Cisco 一體機方案，本身就是一種生態系的展現，整合了不同技術廠商的解決方案，提供客戶一個「AI Ready」的完整基礎架構（Infra）。其中包含思科本身的高速網路設備，Pure Storage 提供的大量儲存能力，以及 Red Hat 的軟體支援，三者結合才能真正交付完整的 AI Ready Infra。邁達特做為與思科合作的代理商，也希望藉由這套方案，讓客戶清楚了解如何在這個基礎架構上發展 AI 應用。

不過，基礎架構只是第一步。當硬體與軟體環境到位後，下一個挑戰是「應用」，這需要另一個生態系的支持。林立棕說，邁達特約從三年前開始經營並積極引入台灣本地的新創，主動尋找各行各業的雲端與 AI 應用，並逐步納入 AI 新創的解決方案。

因為現實中，客戶不會直接說「我要買一台四顆 GPU 的伺服器」，而是提出具體需求：「我想解決某個問題，有沒有方案？」這時候，供應商就必須依照客戶需求，判斷並提供相應解決方案。林立棕直言，對企業客戶來說，最終需要的並非單一軟體，而是一整套「能直接落地使用」的完整解決方案。能將長期累積的基礎架構整合能力，結合本地 ICC（整合應用服務），最後交付給客戶一個能真正運作的生態系方案，正是邁達特的價值所在。

結論

新時代企業 AI 架構與治理基礎

結論

新時代企業 AI 架構與治理基礎

在人工智慧時代，企業面臨了前所未有的技術變革和風險挑戰，同時也促使新的架構與治理模式產生，以確保 AI 應用的可信任與企業的數位韌性。

在當代科技發展的脈絡中，監理與規範已跳脫產業阻力的管制思維與做法，而是以建構市場秩序與數位信任基礎為目標。若沒有 AI 規範，企業難以獲得產品與服務開發、跨境合作與長期投資等所需的穩定與可預期；但若規範僅著重於防範與限制，則會壓縮創新的空間並影響產業生態系的發展。因而，監理思維必須從單純的防弊，轉化為同時兼顧「創新促進」與「風險控管」的治理機制。

一、新的架構：從傳統資料中心到 支援 AI（AI-Ready）的基礎設施

隨著 AI 對算力、儲存與網路的需求急速攀升，基於以下因素，傳統的資料中心架構已難以負荷。企業必須朝可支援 AI（AI Ready）基礎設施演進，打造能支撐高強度 AI 應用的技術基礎建設。

強大的運算需求：AI 模型訓練與推論對基礎設施帶來三大極限要求：高 GPU 算力、高功耗和高頻寬。成千上萬的 GPU 需同時協作，產生巨大的電力消耗與散熱需求，並在節點間交換海量數據。這不僅是對伺服器本身功能的考驗，更是對機房供電、冷卻系統及網路架構的嚴峻挑戰。為此，業界正透過底層技術創新來應對，例如思科（Cisco）研發的 Silicon One 網路晶片，其設計目標便專注於提升 AI 驅動網路的能源效率，並提供業界最高頻寬的路由與交換能力，確保數據在 AI 叢集中暢行無阻。

混合雲整合：AI 應用的需求與功能需要極大的彈性與靈活度，可能在企業內部的地端（On-Premise）環境進行模型訓練以確保資料安全，再部署到公有雲上提供服務以利用其彈性擴展能力。這種跨越地端與雲端的混合雲模式成為常態，但也使得管理愈形複雜。企業迫切需要一個統一的平台來監控、管理和調度分散在各處的 AI 基礎設施。

標準化與彈性部署：AI 技術日新月異，一個長週期的 AI 專案很可能在完成前，其核心技術就已被市場淘汰。為了應對這種高速迭代，企業必須擺脫過去客製化、專案式的導入方法，轉向標準化、穩健、可靈活優化的整體部署架構。

資安優先架構：如前所述，AI 擴大了企業的資安攻擊面。因此，安全絕不能是事後添加的補丁，而必須是貫穿於基礎設施設計之初的核心原則，這意味著從網路、運算、儲存到 AI 軟體工具鏈的每一個環節，都必須內建安全機制。

容器技術：在台灣的討論中較少被提及，但對 AI 營運（MLOps）與安全至關重要的，是容器技術的應用。透過將 AI 應用及其所有依賴項打包在獨立的容器中，企業能夠實現環境的一致性、快速部署與彈性擴展。更重要的是，容器技術為安全管理提供了更精細的維度，能夠有效管理從底層 Linux 作業系統到容器層級的眾多網路協定與安全策略，是確保 AI 應用穩定、安全運行的關鍵技術。

上述這些複雜的技術要求，正推動著市場供應商朝向更深度的整合服務發展。

由於 AI 生態系涵蓋從算力、網路到顧問服務的龐大版圖，為此，市場上也可以看到一站式整合服務的出現。這類策略通常結合了硬體層面的合作與自研，例如既與晶片大廠（如 NVIDIA、Intel）合作以確保設備整合，也同時發展自家的網路晶片與光纖技術。在服務層面，部分廠商亦提供顧問式服務，協助企業評估需求並使 AI 目標與業務對齊；並透過與各垂直產業的系統整合商（SI）合作，結合在地經驗以協助導入。此類整合生態系的目標是為企業提供靈活的擴充選項，並透過統一的管理平台，協助落實 AI 治理，確保基礎設施的高可用性、可預測性與安全性。

二、新的治理模式：建構可信任 AI 框架

一個可信任的 AI 治理模式，旨在確保 AI 系統在其整個生命週期中，都能夠以理性、透明、公平、可解釋、安全和可靠的方式運行。

負責任的 AI 框架：建立 AI 治理的第一步，是確立一套全公司共同遵循的核心價值與原則。這些原則作為保護措施，旨在消除不準確、偏差、隱私和安全風險，以降低 AI 造成的損害風險。思科提出的「負責任的 AI 框架」（Responsible AI Framework, RAI）便是一個業界實例，其內含六大基石：

透明度（Transparency）：讓使用者了解自己正在與 AI 互動，並清楚其決策的依據。

公平性（Fairness）：主動識別並減輕 AI 系統的不公平偏見。

承擔責任（Accountability）：為 AI 系統的產出結果建立清晰的責任歸屬。

隱私（Privacy）：將資料隱私保護深植於 AI 系統的設計與運作中。

安全（Security）：保護 AI 系統免受內部及外部攻擊與濫用。

可靠性（Reliability）：確保 AI 系統在預期條件下能夠穩定、準確地運行。

資料治理與安全：AI 的信任始於資料的信任。企業必須對資料的整個生命週期，從收集、儲存、處理到銷毀，實施強而有力的隱私保護與安全策略。資料的安全性、完整性與公正性是治理的重中之重。一個值得信賴的 AI，其訓練資料來源必須是公開透明且無偏見的。

許多業界觀點均強調，可信任 AI 的根基在於資料來源的安全、合規與透明；若缺乏此基礎，再強大的算力與模型都難以支撐長期發展。

另一方面，治理的複雜性亦體現在「資料主權」（Data Sovereignty）上。例如，Cisco 在與全球五十多國政府合作的經驗中，深刻理解各國在資料主權（Data Sovereignty） 上的不同需求，包括資料必須完全在地化、區域性駐地，或跨境使用的限制。為了協助企業因應各國多樣化的法規要求，並在保障隱私的前提下，確保資料的安全與透明度。使得企業對於完整解決方案的需求提升。

智慧監管與業界標準的協力：面對 AI 的快速發展，傳統的立法模式顯得緩不濟急。僵化或過於寬泛的法規可能扼殺創新。因此，一種更

智慧的監管模式應運而生：著重於規範對法律和人權構成顯著影響的特定「高風險」使用案例。同時，應優先鼓勵並採納由業界主導的標準和指導方針，例如由安全 AI 聯盟（CoSAI）等組織所制訂的標準。這種模式賦予了產業足夠的靈活性，以適應不斷變革的技術，同時也為關鍵領域劃定了清晰的紅線。

主動防禦：新一代 AI 安全工具的應用：治理框架需要有力的工具來落地執行。針對 OWASP 提出的十大 LLM 應用程式風險，新一代的 AI Defense 產品提供了具體的防禦手段。這類工具能夠提供如盤點模型合法性、輸出內容合規性，並依使用者權限控管輸出結果、以 SaaS 模式支援企業落地等服務。

目前產業中的具體實例如思科（Cisco）所發佈的開源安全模型（Foundation-Sec-8B），該模型專為資安維運建構，據稱納入了情資報告、漏洞資料庫、事件回應文件等高品質資安訓練資料集。企業可下載並結合自身資料，應用於安全營運中心（SOC）中，例如透過分析日誌（log）來預判異常狀況並提供防禦建議。

三、IT 部門角色的轉變與人才培養

AI 不僅是技術議題，更是組織文化與人力資源的挑戰。

IT 部門的轉型：IT 不再只是基礎設施的維護者，而需成為跨部門的「橋樑」，推動標準化部署、即時監測與數據治理，協調業務需求與技術落地。

此一趨勢也反映在市場服務的轉變上。 業界已觀察到，IT 部門的角色正從維護者轉變為策略顧問，因此，供應商也開始提供相應的支援模式。例如，包括 Cisco 在內的企業皆提供顧問服務（Consulting Services），透過專家訪談與評估，協助企業釐清 AI 導入目標與所需環境。同時，這類供應商也積極與各垂直產業的系統整合商（SI）合作，藉此結合專業知識與在地經驗，形成新的服務生態以協助客戶導入。

提升 AI 素養：企業普遍存在的 AI 風險認知不足，以及員工對資安威脅的理解落差，是 AI 導入失敗的主要原因之一。因此，大規模地提升組織整體的 AI 素養變得至關重要。

目前已有企業透過如思科「全球網路學院」等計畫及專業 AI 認證，系統性地提升其員工的 AI 技能，培養出既懂技術又懂業務的跨領域人才，為 AI 的全面落地奠定基礎。

以新架構與治理模式奠定韌性基礎

AI 時代的核心不僅是運算力的提升，更在於能否建立可信任與具韌性的運作環境。新的基礎設施架構與治理模式，正是企業在 AI 浪潮中保持競爭力與信任的關鍵。

透過 AI Ready 基礎設施、負責任 AI 框架與人才再培訓的多重策略，企業不僅能降低風險，更能在數位轉型中建立長期韌性。強韌的架構為 AI 運行提供了穩定載體，而完善的治理則為 AI 的發展指明了正確方向。

面對複雜的雲端與 AI 環境，企業必須以動態且全面的方式強化網路韌性，並透過整合性的管理平台，確保基礎建設的「高可用性、可預測性與安全性」。唯有如此，企業才能在數位轉型的浪潮中，穩健邁向 AI 時代，並維持長期營運韌性。

附錄

- 附錄一、研究方法與設計
- 附錄二、參考資料

附錄一、研究方法與設計

1.1 研究目的與範疇

本研究旨在探索台灣產業在導入人工智慧（AI）時所面臨的挑戰與機會，並從全球趨勢、在地困境、專家觀點及政策現況等多重角度，為台灣建構一個整合「可信任（Trustworthiness）」與「韌性（Resilience）」兩大核心要素的 AI 實踐框架。研究範疇涵蓋 AI 策略、風險治理、資料管理、基礎設施、人才培育及法規政策等關鍵面向。

1.2 訪談對象選取標準

為確保研究觀點的廣度與深度，本次質性調查採用「立意抽樣（Purposive Sampling）」，邀請在台灣 AI 發展中具有關鍵影響力的專家。選取標準如下：

代表性： 涵蓋產（科技領袖、系統整合商）、官（數位發展部）、學研（頂尖研究機構、法人基金會）等不同領域的關鍵意見領袖。

專業性：在 AI 技術、資訊安全、數位政策、網路治理、產業應用（如醫療）等領域具有深厚的實務經驗與前瞻洞察。

多元性： 訪談對象的背景與觀點具備多元性，以確保研究視角的全面。

1.3 資料蒐集方式

本研究主要採用「半結構式深度訪談法」（Semi-structured In-depth Interview）。研究團隊擬定訪談大綱後，於 2025 年 7 至 10 月期間，透過線上及實體形式與多位專家進行平均 60 至 90 分鐘的深度訪談，並將過程完整記錄以供後續分析。

1.4 資料分析方法

研究團隊將所有訪談紀錄，採用「主題分析法」（Thematic Analysis）進行系統性歸納。透過反覆閱讀與比對，逐步提煉出共通的挑戰、核心觀點與策略建議，最終匯整為本白皮書核心篇章與其下的關鍵洞見。

附錄二、參考資料

1. Cisco. (2025). 2025 Global Networking Trends Report.
2. Cisco. (2024). 2024 Cisco Cybersecurity Readiness Index.
3. Cisco. (2025). Cisco AI Readiness Index.
4. Cisco. (n.d.). The Cisco Responsible AI Framework. Cisco Systems. https://www.cisco.com/c/dam/en_us/about/doing_business/trust-center/docs/cisco-responsible-artificial-intelligence-framework.pdf
5. 詹婷怡.(2025). 創新治理思維與規範是發展人工智慧競爭優勢必要基礎. 知勢. https://edge.aif.tw/innovative-governance-thinking-is-key-foundation/
6. 溫怡玲.(2025). AI 導入有「灰色地帶」，企業隱藏風險如何解決?. 知勢. https://edge.aif.tw/how-to-reduce-ai-risk/
7. U.S. Government. (2025). <America’s AI Action Plan.>
8. OWASP. (2025). <OWASP Top 10 for Large Language Model and Generative AI Applications.>
9. Artificial Intelligence Foundation. (2025). 2025 Taiwan Industry AI Adoption Survey and Implementation Guidelines.