ISACA

運用數位信任生態系框架實現值得信賴的人工智慧

Using the Digital Trust Ecosystem Framework to Achieve Trustworthy AI

治理

目 錄

摘 要

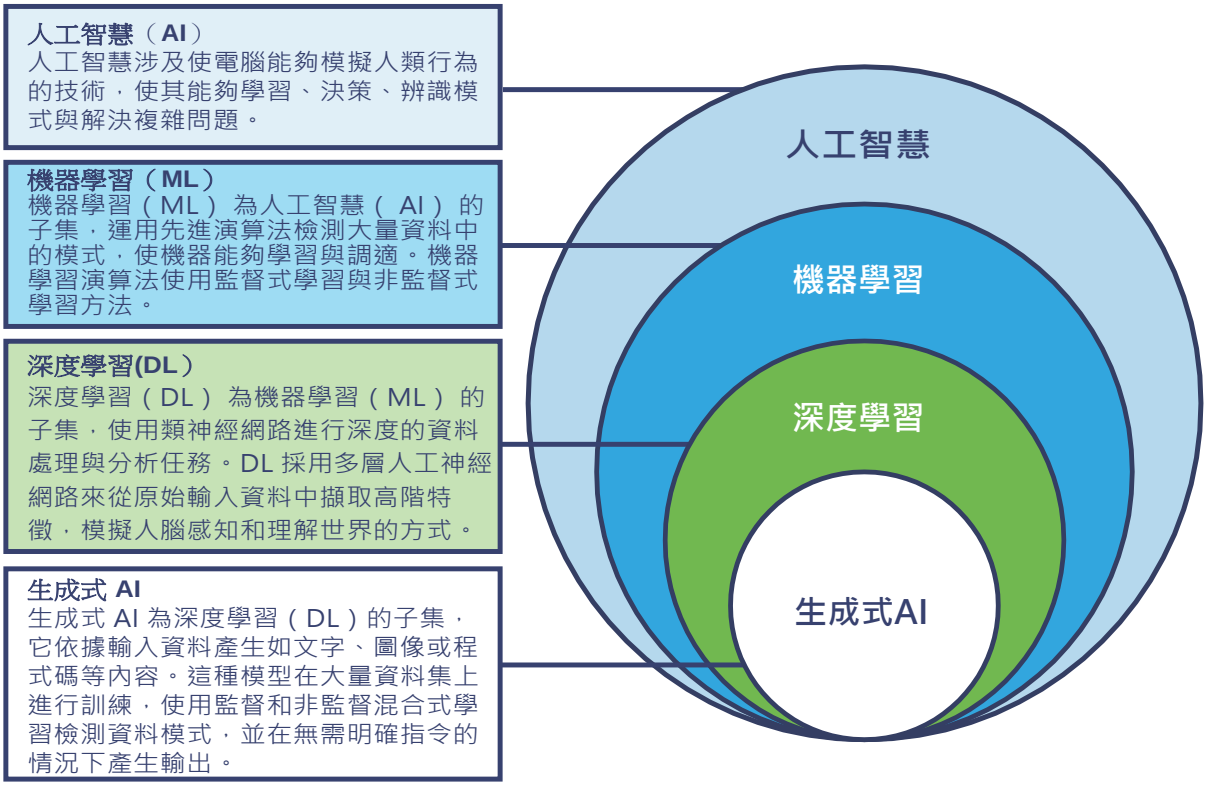
本白皮書探討了使用 ISACA 的數位信任生態系框架（DTEF）為採用人工智慧（AI）技術和服務的企業帶來的效益。內容提供了數位信任生態系框架（DTEF）如何協助評估新興科技風險與構建治理架構的指引，使組織在整個人工智慧（AI）生命週期中受益。本白皮書也強調了數位信任中的關鍵要素，這些要素是成功導入人工智慧（AI）技術與提供服務的基礎，並透過實際應用情境範例說明組織常見的挑戰與對應方式。

簡介

人工智慧（AI）應用已廣泛存在於社會的各個層面，常見範例包括聊天機器人、金融詐欺偵測及導航軟體。人工智慧 （AI）這個術語涵蓋機器學習（ML）、深度學習（DL）及生成式 AI（見圖一）；總體而言，它將持續顛覆各行各業，提供高效率與大規模的效率與效益。例如，在醫療領域，人工智慧 （AI）可實現個人化治療方案與預測性診斷管理；在金融服務業，人工智慧 （AI）運用數據驅動的洞察強化詐欺偵測與風險管理；而在能源產業，人工智慧 （AI）則協助優化電網管理與預測性維護，實現高效率與永續性的實際效益。

人工智慧 （AI）的影響力不限於企業明確導入的情境，其已被整合進多種的第三方應用中，例如常見辦公軟體與日常工作流程。此外，各部門如人力資源與行銷人員，也很可能已經在使用網頁型軟體進行職缺篩選或撰寫行銷內容。換言之，員工即使未意識到，往往也已在使用人工智慧 （AI）技術。事實上，越來越多尚未經企業正式授權使用的生成式 AI 工具，已逐漸演變為一種新型態的「影子 IT」。

圖1：人工智慧、機器學習、深度學習與生成式 AI 之比較視角



資料來源: Unraveling AI Complexity - A Comparative View of AI, Machine Learning, Deep Learning, and Generative AI, https://commons.wikimedia.org/wiki/File:Unraveling_AI_Complexity_A_Comparative_View_of_AI,_Machine_Learning,_Deep_Learning,_and _Generative_AI.jpg. This figure is available under the *Creative Commons Attribution-ShareAlike 4.0 International* license

儘管人工智慧（AI）能為企業帶來營運優化與效能提升，其潛在的風險亦不容忽視。如同歷來所有新興科技一樣，人工智慧很快便成為惡意行為者濫用的工具。到目前為止，生成式AI提高了商業電子郵件詐騙[1]（Business Email Compromise, BEC）的可信度，同時也降低了實施商業電子郵件詐騙（BEC）攻擊[2]所需的技術門檻。此外，人工智慧技術已被運用於地緣政治[3]、影像濫用[4]、政治宣傳[5]等領域，甚至曾在一樁成功的數百萬美元詐騙案件[6]中扮演關鍵角色。總體而言，人工智慧不僅在不同語境中被賦予不同的意義，其所引發的風險也依其應用類型而千差萬別。

人工智慧（AI）的普及程度無法低估。根據勤業眾信（Deloitte） 的觀察，全球正處於一場前所未有、普及程度極高的技術革命，而這場革命的核心，正是生成式 AI[7]。雖然早在 1950 年[8]便已有人工智慧（AI）的初步概念，但其技術發展的速度與影響範圍至今仍在持續擴展。與過往的技術變革類似，人工智慧的快速演進引發不少人的恐懼、懷疑與不信任。然而值得注意的是，生成式 AI 主要是用於釋放而不是取代人類潛力的關鍵工具。真正的挑戰，在於如何妥善分配與管理由人類與機器各自擅長的任務。使用人工智慧（AI）需要跨職能的領導，並且在傳統「人、流程、技術」三要素外，還需要審慎評估其對組織與社會所帶來的影響。以進入 ISACA 的「數位信任生態系框架」（Digital Trust Ecosystem Framework, DTEF）[9]。

人工智慧（AI）的普及程度無法低估。根據勤業眾信（Deloitte） 的觀察，全球正處於一場前所未有、普及程度極高的技術革命。

「數位信任生態系框架（DTEF）」的設計從組織所有利害關係人的角度出發，全面性地建立與維護數位信任。數位信任的涵義遠超過科技層面，其適用於整個組織內部與所有外部利益關係人。

選擇、建立、維護數位關係需要所有相關方的信心與透明度。供應商與消費者的需求、原則、價值觀與目標將影響信任程度。「數位信任生態系框架（DTEF）」提供了一套創新的數位信任和轉型的方法論，將「數位信任」視為優先核心，特別是在導入人工智慧技術時更應如此。

本白皮書探討了組織如何運用 ISACA 的「數位信任生態系框架（DTEF）」實現人工智慧技術與服務的數位信任解決方案。

[1] ISACA, "ISACA Glossary," https://www.isaca.org/resources/glossary
[2] Kelley, D.; "Wormgpt - the Generative AI Tool Cybercriminals Are Using to Launch Business Email Compromise Attacks," SlashNext, 13 July 2023, https://slashnext.com/blog/wormgpt-the-generative-ai-tool-cybercriminals-are-using-to-launch-business-email-compromise-attacks/; Shiebler, D.; "Generative AI Enables Threat Actors to Create More (and More Sophisticated) Email Attacks," AbnormalSecurity, 14 June 2023, https:// abnormalsecurity.com/blog/generative-ai-chatgpt-enables-threat-actors-more-attacks
[3] Klepper, D.; "Fake Babies, Real Horror: Deepfakes From the Gaza War Increase Fears About AI' s Power to Mislead," APNews, 28 November 2023, https://apnews.com/article/artificial-intelligence-hamas-israel-misinformation-ai-gaza-a1bb303b637ffbbb9cbc3aa1e000db47
[4] Saner, E.; "Inside the Taylor Swift Deepfake Scandal: 'It' s Men Telling a Powerful Woman to get Back in her Box' ," The Guardian, 31 January 2024, https://www.theguardian.com/technology/2024/jan/31/inside-the-taylor-swift-deepfake-scandal-its-men-telling-a-powerful-woman-to-get-back- in-her-box
[5] Hickey, M.; "Vallas Campaign Condemns Deepfake Video Posted to Twitter," CBS News, 27 February 2023, https://www.cbsnews.com/chicago/news/vallas-campaign-deepfake-video/; Harper, A.; Gehlen, B.; et al.; "AI Use in Political Campaigns Raising Red Flags into 2024 Election," ABC News, 8 November 2023, https://abcnews.go.com/Politics/ai-political-campaigns-raising-red-flags-2024-election/story?id=102480464; Ulmer, A.; Tong, A.; "Deepfaking it: America' s 2024 Election Collides with AI Boom," Reuters, 30 May 2023, https://www.reuters.com/world/us/deepfaking-it- americas-2024-election-collides-with-ai-boom-2023-05-30/
[6] Edwards, B.; "Deepfake Scammer Walks off With $25 Million in First-of-its-kind AI Heist," ARS Technica, 5 February 2024, https://arstechnica.com/information-technology/2024/02/deepfake-scammer-walks-off-with-25-million-in-first-of-its-kind-ai-heist/
[7] Deloitte, "Generative AI and the Future of Work," https://www2.deloitte.com/content/dam/Deloitte/us/Documents/consulting/us-ai-institute- generative-ai-and-the-future-of-work.pdf
[8] Turing, A.M.; *Computing Machinery and Intelligence*, Mind 49, 1950, https://redirect.cs.umbc.edu/courses/471/papers/turing.pdf
[9] ISACA, *Digital Trust Ecosystem Framework*, USA, 2022, www.isaca.org/dtef-ebook

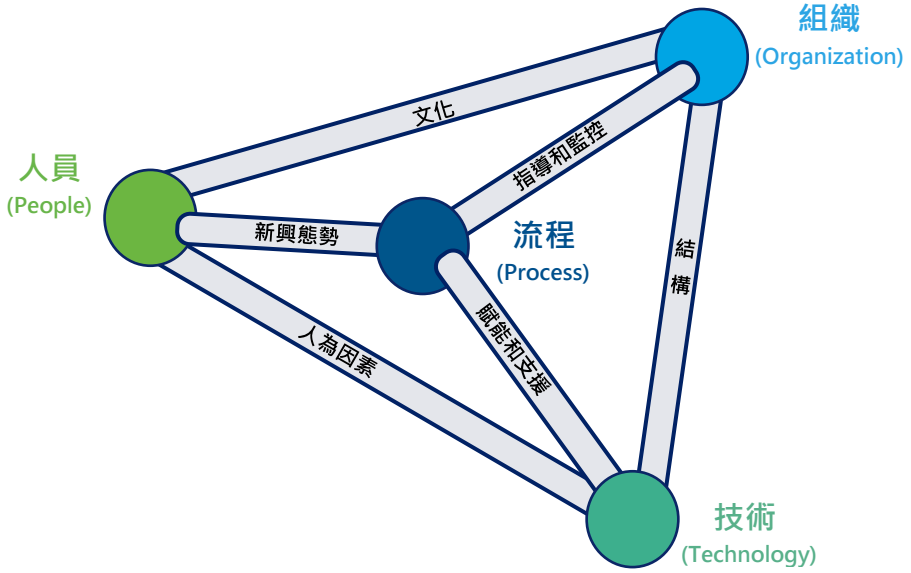
數位信任生態系框架（DTEF）概述

數位信任生態系框架（DTEF）定義了創建「數位信任生態系」的核心要素。此生態系須考量所有利害關係人，確保所有數位互動與交易皆為合法、可信賴，並涵蓋完整性、安全性、隱私、韌性、品質、可靠性及信任感[10]等關鍵要素。信任是一項企業在導入與部署人工智慧策略時，必須納入的核心原則。例如，能否信任人工智慧（AI）的輸出結果，取決於資料品質與資料保護的程度。人工智慧模型的開發必須建立於可信任的程式碼與透明性基礎上，並納入決策機制。最終使用者或人工智慧服務的消費者，必須能信任驅動決策的人工智慧演算法在準確性、隱私保護、安全性以及偏誤控制等方面皆經過驗證。此外，適當的人工智慧治理機制有助於企業高階管理者與主要利害關係人信任人工智慧（AI）所做出的決策是可解釋的，並確保其遵循公平性、倫理實務及具備法規與法律合規性。

數位信任生態系框架（DTEF）建構了一套知識體系，協助企業因應不斷變動的法律、法規與技術環境，符合現代商業需求，並處理外部與內部的影響因素及風險控管要素，使其能夠在具備數位信任的環境中持續營運。

數位信任生態系框架（DTEF）採用三維模型，建立於一項核心理論之上：在「人員」、「流程」、「技術」及「組織」的四個主要節點之間，存在多重相互依賴關係。這些節點之間透過多項動態互動進行連結，並構成此模型的最高層級架構。四個節點由六個領域連結，分別為：文化、新興態勢、賦能與支援、人為因素、指導和監督、以及架構。圖二呈現了節點與領域間的關聯性。

圖 2：數位信任生態系框架（DTEF）模型



資料來源：ISACA，數位信任生態系框架（DTEF），2022

[10] *同上。*

領域會影響一個或多個節點。例如，結構領域的變動必然會影響組織節點和技術節點。這些領域之間也以系統化的方式相互作用著。

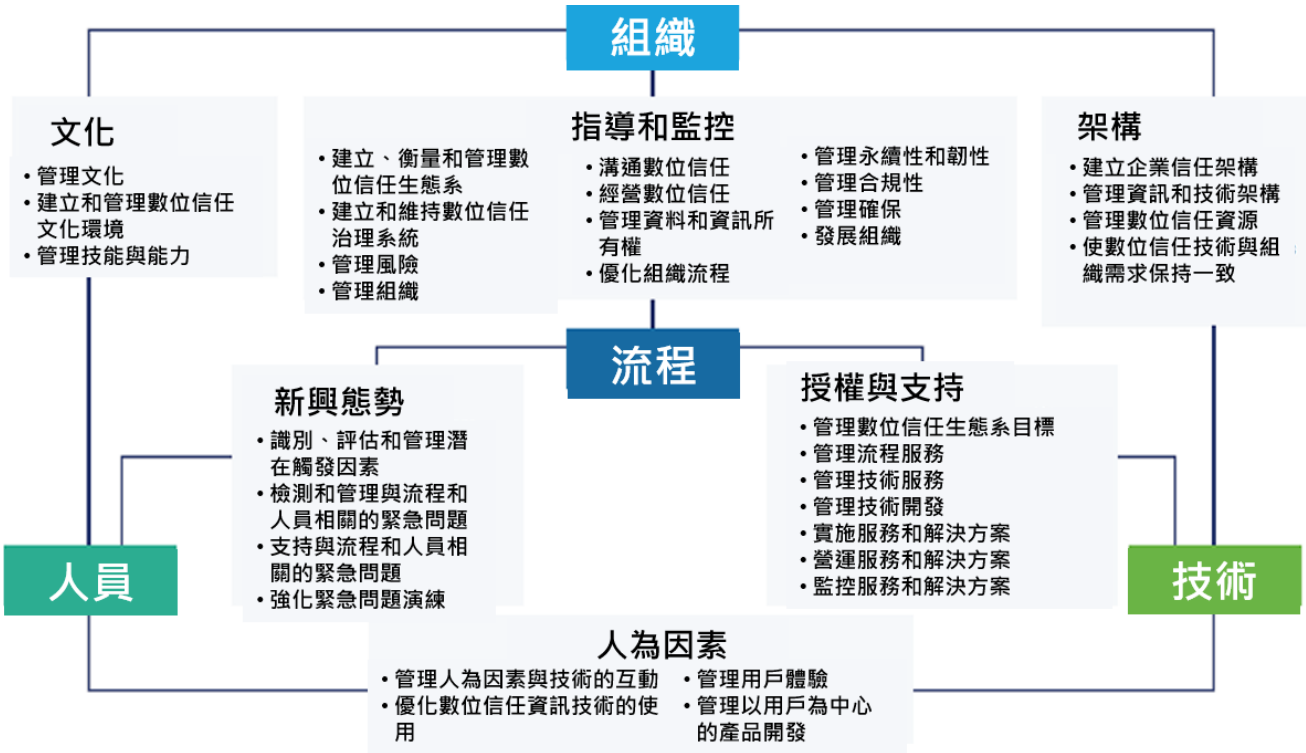
這些領域在組織中存在相互連結和複雜性並扮演著管理的關鍵角色，因為它們與不斷變動的法規、新興技術、新威脅、程序變更等密切相關。領域是由一組構成和結構的元素所組成。

數位信任生態系框架（DTEF）在每個領域中，透過信任要素來建立內容基礎。例如，架構領域被劃分為以下四項信任要素：

1. 建立企業信任架構。
2. 管理資訊與技術架構。
3. 管理數位信任資源。
4. 使數位信任技術與組織需求校準。

信任要素描述了維護數位信任所需的整體行動，並有助於避免或降低偏差。此框架的組成部分（參見圖 3）可用於確保在任何人工智慧的應用中，數位信任的原則皆能夠被實踐。

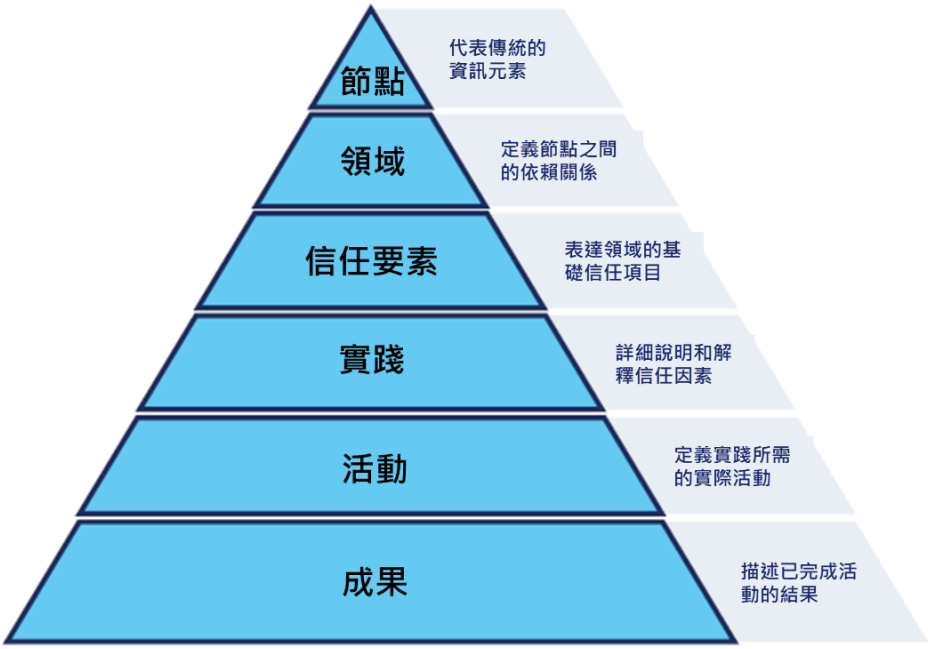
圖3：數位信任生態系框架（DTEF）組成要素



資料來源：ISACA，數位信任生態系框架，USA，2022

數位信任生態系框架（DTEF）提供了一個架構，用來在整個生態系中支持數位信任，同時並考慮到各種關係如何影響與消費者、客戶和使用者間的互動程度。圖4說明了數位信任生態系框架（DTEF）的主要結構要素。

圖4：數位信任生態系框架（DTEF）階層



數位信任[11]不僅僅關乎數位資訊與技術。數位信任更影響著企業的整體發展；因此，能夠展現數位信任的企業將獲得顯著的競爭優勢，並與消費者建立更良好的關係[12]。

人工智慧生命週期

未受控制的使用人工智慧產品，可能會帶來相當大的風險。由於其潛在影響可能波及整個企業（例如智慧財產權〔IP〕損失、品牌受損、法律訴訟），企業不僅需要了解現有已在使用的人工智慧工具，更需要理解員工試圖透過人工智慧（AI）解決的實際業務問題。在某種程度上，人工智慧（AI）只是一項技術，但因其細微差異，使其在生命週期各階段都需要進行嚴謹的治理與風險管理。人工智慧生命週期是一個從業務問題出發，再建立人工智慧解決方案的反覆迭代過程[13]。

如圖 5 所示，各步驟會在設計、開發與部署階段中不斷迭代。雖然讀者可能熟悉其他的人工智慧生命週期，其中可能包含更多細節。，但無論形式如何不同，人工智慧生命週期通常涵蓋以下四大組成成分：業務需求與目標、資料收集與準備、模型開發與評估、及營運部署與監控。

[11] ISACA defines digital trust as the confidence in the integrity of the relationships, interactions, and transactions among providers and consumers within an associated digital ecosystem. This includes the ability of people, organizations, processes, information, and technology to create and maintain a trustworthy digital world.

[12] Tower-Pierce, J.; "Wake up America: Digital Trust can Positively Impact Revenue," InCyber, 10 July 2023, https://incyber.org/en/article/wake-up- america-digital-trust-can-positively-impact-revenue/

[13] IT Modernization Centers of Excellence, "AI Guide for Government: A Living and Evolving Guide to the Application of Artificial Intelligence for the U.S.Federal Government," 26 March 2024, https://coe.gsa.gov/coe/ai-guide-for-government/print-all/index.html

圖5：人工智慧生命週期

設計

理解問題

人工智慧流程和系統擁有者、及其他相關利害關係人需明確辨識專案的關鍵目標與需求，才能有效定義預期的業務成果。組織必須明確定義他們希望人工智慧（AI）解決哪些業務問題。若未能清楚準確地掌握要解決的業務挑戰與預期結果，任何人工智慧解決方案都無法成功。

資料收集與探索

數據是所有人工智慧解決方案的基礎。本階段需蒐集並評估資料是否適合應用於目標人工智慧場景，包括辨識可用數據集、檢視數據品質問題、初步探索數據內容與擬定數據規劃。人工智慧模型僅應使用經過明確定義且符合需求的資料。

資料整理與準備

此階段包含將原始數據轉換成模型可用的格式。過程雖繁瑣且耗時，但對於建構能有效解決問題的人工智慧模型而言至關重要。

開發

建模

本階段聚焦於數據實驗，以確定最合適的人工智慧模型。在此階段，開發團隊通常需要進行反覆的對模型進行訓練、測試、評估與再訓練，才能找到最佳的模型與參數。人工智慧模型的訓練和選擇過程是互動的。沒有任何的人工智慧模型能夠在第一次訓練時就達到最佳表現。因此必須透過不斷迭代微調來完善。根據數據規模與類型不同，訓練過程可能需要極高的計算資源，甚至需使用專用硬體設備，因此它無法於普通的筆記型電腦上完成。

評估

一旦建立了一個或多個根據相關評估指標表現良好的人工智慧模型，就會使用新數據對它們進行測試，以確保它們具有良好的泛化能力並符合業務目標。

部署

移轉至正式環境

當開發的模型符合預期成果，並達到可在實際數據中運行的水準，即可部署至正式環境。在部署後，人工智慧模型將開始處理在訓練階段未曾見過的新資料。

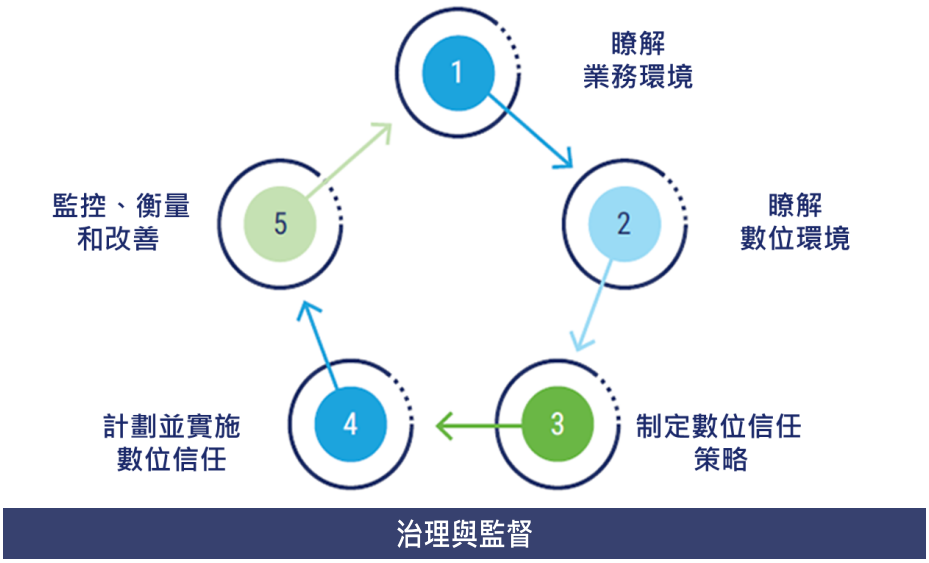
監控人工智慧模型輸出

於模型上線後，必須持續監控其輸出，確保能持續達到預期成果。此過程稱為泛化，即人工智慧模型能正確適應新的、從未見過的數據。在正式環境中，人工智慧模型可能出現「漂移」現象，這意味著性能會隨時間推移而導致效能下降。確實監控漂移是非常重要的，必要時也需要持續的更新 人工智慧模型。因此，人工智慧模型必須進行嚴謹且持續的監控與維護，以確保其長期穩定、可靠並持續解決業務問題。

以人工智慧為核心的數位信任生態系框架實施

數位信任生態系框架（DTEF）實施模型可協助組織以全面性的方式來規劃人工智慧專案。該模型包含五個主要階段，並以第六階段「治理與監督」作為前五項的基礎（見圖 6）。

圖6：數位信任生態系框架（DTEF）實施模型



資料來源：ISACA，數位信任生態系框架（DTEF）部署指引， 2024

圖 7 說明了 數位信任生態系框架（DTEF）實施模型、人工智慧生命週期與關鍵人工智慧特定活動之間的關係。

圖7：人工智慧生命週期與數位信任生態系框架（DTEF） 實施模型對應圖

瞭解企業環境

任何商業計畫都需要對現有的企業環境有清楚的瞭解，包含願景、使命、目標與策略；若缺乏這些，就可能導致方向出現偏差。對於人工智慧（AI）而言，因為其正值高度關注與寄予厚望的時期，因此越早與利害關係人互動是尤為重要。必須辨識企業風險與容忍度、業務痛點以及預期成果，因為並非所有事情都適合自動化。本階段的工作重點是界定範疇、假設前提與限制。與人工智慧（AI）相關的任務包括：

1. 制定企業的人工智慧願景、使命、目標和宗旨。
2. 瞭解與人工智慧相關的業務風險。
3. 識別使用人工智慧的產品和服務。

瞭解數位環境

與前一階段類似，本階段著重於數據收集，包括紀錄組織在產品與服務中的數位特性，並識別與數位互動的各種情境和使用案例。此階段與人工智慧（AI）相關的任務包括：

1.辨識企業目前已使用的人工智慧產品與服務。
2.辨識多元化的人工智慧利害關係人。
3.定義人工智慧的數位互動關係類型（例如：企業對消費者、企業對員工、政府對選民）。
4.定義人工智慧的數位互動媒介。

5.瞭解人工智慧數位供應鏈。
6.建立人工智慧數位互動使用案例。

制定數位信任策略

在此階段，企業需根據業務與數位環境，發展出數位信任策略。與人工智慧（AI）相關的任務包括：

1.紀錄人工智慧數位信任的策略性目標。
2.建立人工智慧系統與子系統的對應關係。
3.制定初步的人工智慧策略。
4.創建人工智慧商業案例。

規劃與實施數位信任

本階段是實際執行人工智慧（AI）導入的關鍵階段。為了成功完成，組織必須依據前三階段的成果，規劃人工智慧數位信任計畫並落實執行。與人工智慧（AI）相關的任務包括：

1.建立循環迭代的方法。
2.擬定人工智慧專案(計畫)方案。
3.為專案的每一步驟指派人工智慧負責人。
4.實際導入人工智慧。

監控、衡量與改善

最後，第五階段是建立持續運作模式的關鍵環節，涵蓋持續監控、衡量和改善，並與第一階段形成閉環，以確保持續迭代。在某些人工智慧應用（如機器學習）中，持續監控特別的重要，因為模型會隨著新資料的加入不斷改變輸出結果。此階段與人工智慧（AI）相關的任務包括：

1.確定人工智慧數位信任的關鍵衡量指標。
2.設定人工智慧衡量目標值。
3.蒐集、監控並回應衡量結果。
4.評估人工智慧的能力與成熟度。
5.持續改善人工智慧的數位信任環境。

治理和監督

貫穿所有階段的基礎就是治理與監督，以確保包含人工智慧（AI）在內的數位信任計畫能夠得到適當的範圍界定、執行和持續優化。與人工智慧（AI）相關的任務包括：

1.確立並採用人工智慧治理制度。
2.建立與管理人工智慧風險登記冊。
3.將人工智慧數位信任納入企業的治理、風險與法遵（GRC）架構中。

數位信任生態系框架（DTEF）的應用

與傳統軟體系統不同，人工智慧系統面臨獨特的挑戰，須要有專門的治理框架。隨著人工智慧系統以前所未有的速度持續發展，其對企業的影響也日益顯著。人工智慧（AI）使用量的快速增加，需要透過監督與結構化的方法來降低潛在傷害。由於企業目標與基礎設施的差異性極大，加上每個使用案例的獨特性，使得不可能涵蓋所有情境來說明數位信任生態系框架（DTEF）如何幫助企業在最大化組織價值的同時，降低人工智慧（AI）風險。

人工智慧使用量的快速增加，需要透過監督與結構化的方法來降低潛在傷害。

如今，人工智慧（AI）被廣泛採用，大量被整合進現有的軟體，或透過應用程式介面（API）來實現。理想情形下，凡是包含人工智慧功能的企業軟體更新，都應啟動既有的評估與核准流程，以進行商業檢視，包括使用條款審查，以及風險與報酬的衡量。值得注意的是，不同產品停用人工智慧功能的難易度差異極大，且操作方式並不總是直觀的[14]。

單一、孤立的供應商管理與採購流程，無法應對當今企業數位生態系的複雜性。產業資訊呈現一個嚴峻的現實：許多公司對第三方風險管理不足[15]，應用程式介面（API）攻擊事件卻不斷增加[16]。當中超過 40% 的公司甚至不清楚自己在使用哪些應用程式介面（API）[17]。再者，有超過一半的供應鏈資安事件被歸因於供應鏈本身的故障導致[18]，因此企業必須更重視應用程式介面（API）安全，以解決常見的漏洞問題[19]。本文其餘部分將探討一個常見的使用案例，以及一些數位信任生態系框架（DTEF）的代表性構成要素；但這些內容並非詳盡，而是示範性說明。

案例研究：客服聊天機器人

根據 Zendesk 的調查結果，生成式 AI 的興起，導致 70% 的企業顧客體驗（CX）主管重新評估顧客體驗功能[20]。該調查同時揭示了三大主題下的十項趨勢，而這些趨勢全部都與人工智慧有關，並需要各主要領域的數位信任專業人員必須具備認知、參與與監督能力。圖 8 展示了這些主題。

[14] Kaelin, M.; "How to Disable Windows 11 Copilot Through Registry File or Group Policy Editor," TechRepublic, 20 October 2023, https://www.techrepublic.com/article/how-to-disable-copilot-windows-11/; Salesforce, "Enable and Disable Data Cloud AI and Beta Features with Feature Manager," https://help.salesforce.com/s/articleView?id=release-notes.cdp_rn_2024_winter_feature_manager.htm&release=246&type=5

[15] Bolton, R.; "Third-Party Cybersecurity Risk Management — Updates for a Changing Risk Environment," Community Banking Connections, https://www.communitybankingconnections.org/Articles/2023/I2-I3/third-party-cybersecurity

[16] SALT Labs, "State of API Security Q1 2023," SALT, https://content.salt.security/rs/352-UXR-417/images/SaltSecurity-Report-State_of_API_Security.pdf; Matson, K.; "The Next big API Security Breach Looms: Here's how to Prepare," SC Media, 19 October 2023, https://www.scmagazine.com/perspective/the-next-big-api-security-breach-looms-heres-how-to-prepare

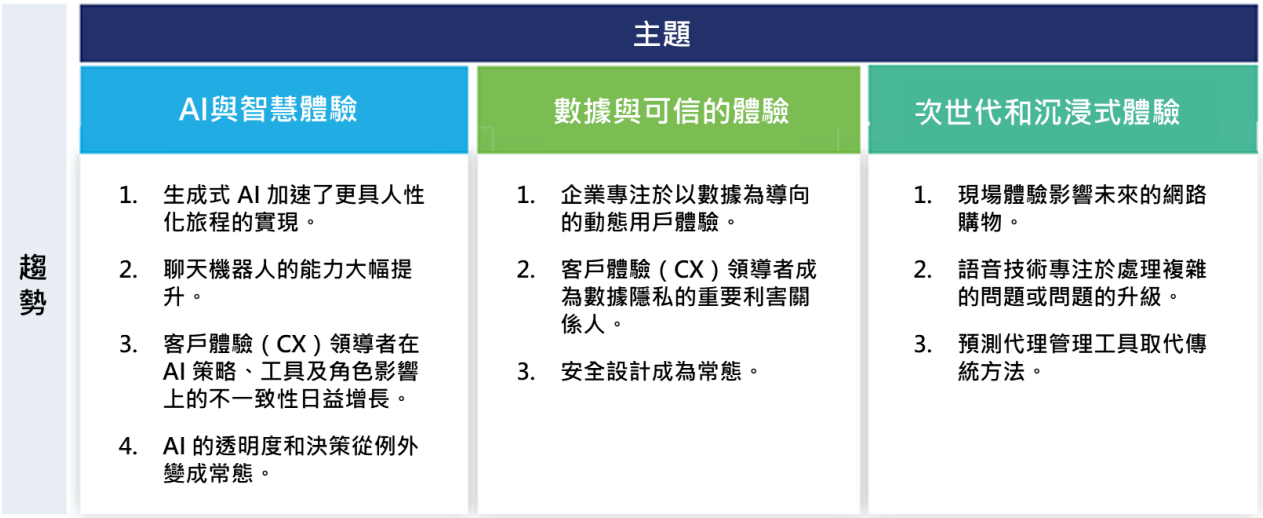
[17] Nagaraj, S.; "The State of API Security in 2023," InfoWorld, 2 November 2023, https://www.infoworld.com/article/3709450/the-state-of-api-security- in-2023.html

[18] *Op cit* Bolton, https://www.communitybankingconnections.org/Articles/2023/I2-I3/third-party-cybersecurity

[19] OWASP, "OWASP API Security Project," https://owasp.org/www-project-api-security/

[20] ZenDesk, "CX Trends 2024," https://cxtrends.zendesk.com/reports/cx-trends-report

圖8：2024年顧客體驗趨勢



資料來源：Adapted from Zendesk, "CX Trends 2024," https://cxtrends.zendesk.com/

在客服領域對生成式 AI 的期望值非常高，因此不難理解為什麼超過一半的顧客體驗主管正在探索人工智慧供應商。從事顧客體驗工作的人員對聊天機器人寄予厚望，認為它們將持續轉型為更強大的數位代理商。在此過渡期間，企業最好組建跨部門的團隊，以管理法遵問題，並最大限度地降低其與現有數位生態系整合所帶來的風險。

要從這些期望轉化為聊天機器人的實際應用，就必須了解它們的分類。聊天機器人通常被歸類為簡單型、智慧型或混合型。直到最近，聊天機器人主要是基於規則的，這提供了一致且可靠的體驗，但如今正逐漸透過自然語言處理（NLP）轉向人工智慧。除了基於規則方法外，聊天機器人也可能是基於關鍵字、選單、智慧型（具備情境理解）、混合型或語音互動[21]。總體而言，聊天機器人能夠提升客戶服務品質、克服語言障礙，並盡力消除因長時間等待電話而產生的挫折感。數位信任生態系框架（DTEF）已準備好應對策略、供應商選擇與管理、實施以及持續改進等方面的挑戰。

數位信任生態系框架（DTEF）如何提供協助？

無論是目前已經使用，或正在考慮使用人工智慧（AI）的類型為何，企業都必須制定一個與公司策略和成果一致的人工智慧策略。具體的策略與開發週期將會依人工智慧的類型而有很大差異。

以下列舉了部分在特定領域內適用的信任要素、實務與活動，這些可以對人工智慧聊天機器人的案例提供有用的指引。

[21] Shenoy, A.; "6 Types of Chatbots – How to Choose the Best for Your Business?," Yellow.ai, 9 January 2024, https://yellow.ai/blog/types-of-chatbots/

文化

人工智慧（AI）常常伴隨恐懼、不確定性與疑慮，特別是在它將如何影響個人或職業工作的程度上。建議企業應制定組織策略，將工作類型和人員的轉變那入考量。

組織發展負責人與招募經理之間的開放對話，可以幫助員工對人工智慧（AI）所帶來的風險與機會建立認同感。另一方面，企業與外部利害關係人之間，若能清楚說明人工智慧（AI）的目的、用途與治理方式，也將會產生助益。常見問題（FAQ）頁面就是一個範例。

文化領域中與人工智慧（AI）相關的考量包括：

- 定義人工部署與使用的目標文化：為何使用？如何使用？及其對內部和外部利害關係人的「文化足跡」是什麼？
- 解決任何阻礙人工智慧成功發展和整合的文化障礙。
- 從更廣泛的環境角度管理目標文化：例如接受度、採用率、知識和技能、常見的使用模式等。

當企業考慮採用人工智慧（AI）時，文化領域中的三個信任要素非常重要：

1. 管理文化（CU.01）：評估、調整與推廣有助於促進數位信任的組織文化和人類文化。
   實務編號 CU.01.02：重塑文化，包含傳達公司策略與價值觀的活動，並結合經驗教訓和外部要求與期望。此實務強調自上而下展現道德領導力的必要性。
   實務編號 CU.01.03：推廣文化，包含對員工的培訓與意識提升，使其理解並支持數位信任策略。
2. 建立和管理數位信任文化環境（CU.02）：定義並建立覆蓋整個生態系的數位信任管理體系。
   在「重塑文化」的實務中，活動 CU.02.02.6 為依不同外部利害關係人的需求，提供適當程度的資訊。
3. 管理技能與能力（CU.03）：這項信任因素著重於識別和維持所需人力資源的最佳技能、能力與能量。這些資訊對於管理人工智慧相關風險非常重要，當關鍵利害關係人未能理解其在確保與維護新興科技環境中數位信任的角色與責任，就可能導致風險。

人為因素

人為因素領域可以協助企業預測人力需求，也解決顧客體驗中需要考量與密切監控的部分。雖然科技相關的知識、技能與能力差距並非人工智慧（AI）所獨有，但人工智慧（AI）是最新且最具活力領域，因此在可預見的未來，將需要更有計畫的人才管理策略，以辨識並管理升級(再培訓)的需求，來確保整體人工智慧計畫和各別專案的成功。此外，人工智慧（AI）也需要持續監測，以確保模型能如預期運作。

人工智慧（AI）在人為因素領域的考量要素包括：

- 讓人工智慧的導入能被理解且易於使用，透過多元的溝通方式與不同的呈現格式來展示人工智慧驅動的結果。

- 定義支援使用者體驗並驗證人工智慧驅動結果的控制措施（例如：鑑別機制、專業上應有之注意等）。
- 定義並部署持續改善流程，引入回饋與前饋循環，確保人類與人工智慧的互動能持續優化。
- 設計分級與打分機制，讓人類參與者能夠對人工智慧實施的可信度給予不同層級的判斷。

在聊天機器人的案例中，從業人員可能需要關注：

- 實務 HF.01.04 及其所有相關活動：評估和管理科技相關人為因素的能力。
- 活動 HF.01.05.1：識別並修正與數位信任有關的錯誤與中斷來源。

結構

成功的人工智慧整合需要協調數據、人員、流程與技術[22]。要達成此目標，企業需要制定策略，一個自覺且深思熟慮的過程，包含選擇、妥協，以及在必要時說「不」，以滿足業務目標。依據開放組織（Open Group）的觀點，企業架構（EA）是一種策略工具，用來識別並縮小現況與未來狀態之間的差距。良好的企業架構能夠協助企業將策略轉化為實際執行[23]。

成功的人工智慧整合需要協調數據、人員、流程與技術

數位信任生態系框架（DTEF）架構領域涵蓋了定義、開發與管理整體企業架構的相關議題，內容包括企業架構模型中業務、資料、應用程式與技術層級的計畫、政策與標準等。

架構中與人工智慧（AI）相關的考量包括：

- 將人工智慧宇宙（AI Universe）映射到整體資訊通信技術（ICT）環境與面向業務的使用案例。
- 將人工智慧行為者（AI actors）嵌入到整體架構中，包含第三方和更遠的供應鏈。
- 定義並確保運行、控制與管理人工智慧行為體所需的資源需求。
- 考慮文化與新興態勢領域，持續採納新的或替代性的人工智慧使用案例，以強化架構。
- 管理並定期檢視人工智慧使用的商業與財務案例。

以聊天機器人案例，以下信任要素、實務與活動具有參考價值：

1. 管理數位信任資源（AR.03）

   此信任要素涉及識別、管理與控制確保數位信任所需的資源。這對於管理基礎設施的所有要素與元件至關重要。相關的實務與活動包括：

   - 實務 AR.03.02：管理數位信任應用程式。
     - 活動 AR.03.02.1：管理以客戶為導向的應用程式（如前端），以實現數位信任。
   - 實務 AR.03.05：管理數位信任營運。
     - 管理與控制數位信任及其相關架構的所有營運與正式環境面向。
     - 活動 AR.03.05.1：估算開發、取得或交付與數位信任相關營運計畫所需工作和資源的規模、工作量、期間與成本。

[22] Strickrodt, D.; "The Future of Enterprise Architecture and AI Integration. Embrace the Synergy.," Bizzdesign, 27 October 2023, https://bizzdesign.com/blog/the-future-of-enterprise-architecture-and-ai-integration/

[23] The Open Group, "A Practitioners' Approach to Developing Enterprise Architecture Following the TOGAF® ADM," https://pubs.opengroup.org/togaf-standard/adm-practitioners/adm-practitioners_3.html

2. 使數位信任技術與組織需求保持一致（AR.04）
   此信任要素涉及識別組織需求，並使技術與這些需求保持一致。

- 實務 AR.04.01：使技術對齊業務需求。
  - 活動 AR.04.01.1：識別相關的業務目標。

指導與監控

與人工智慧（AI）相關的潛在風險[24]，過去多半被視為未來的議題。然而，主動提前處理人工智慧風險，有助於確保企業安全且負責任的部署人工智慧技術。儘管認識到未來風險非常重要，但人工智慧技術的現況及當前的使用模式，已經帶來必須立即面對的風險。簡而言之，人工智慧（AI）可能會加劇既有的問題，例如缺乏品質控制或資料完整性不足，使系統更容易受到資安攻擊，甚至可能引發新的風險。

儘管認識到未來風險非常重要，但人工智慧技術的現況及當前的使用模式，已經帶來必須立即面對的風險。簡而言之，人工智慧可能會加劇既有的問題，例如缺乏品質控制或資料完整性不足，使系統更容易受到資安攻擊，甚至可能引發新的風險。

指導和監督領域包含建立、衡量、管理與治理數位信任生態系相關的主題，包括風險、溝通、資訊、永續與韌性、法遵、確信以及企業的整體發展。

指導和監督領域與人工智慧（AI）在相關的考量包括：

- 將人工智慧宇宙嵌入資訊（ IT）的大型治理、風險管理與合規（GRC）框架中。
- 在組織內外，持續套用治理、風險管理、法遵及確信等要素於人工智慧宇宙。
- 規劃人員與組織架構以引導和控制人工智慧行為者（包含其從導入到退役的生命週期）。
- 將人工智慧納入三道防線架構。

雖然於聊天機器人案例中，有許多相關的信任要素、實務與活動可以應用，但下列幾點值得考量：

1. 管理風險（DM.03）
   此信任要素涉及持續識別、評估並降低數位生態系相關的風險，使其處於企業高階管理階層設定的容忍範圍內。

- 實務 DM.03.01：指導與監控風險管理。
  - 活動 DM.03.01.1：識別角色與職責。
  - 活動 DM.03.01.2：建立風險偏好與風險容忍度。
- 實務 DM.03.02：識別數位生態系風險。
  - 活動 DM.03.02.2：識別風險擁有者。
  - 活動 DM.03.02.3：識別目前的風險控制措施(控制環境)。
  - 活動 DM.03.02.6：將數位生態系風險整合至更廣泛的企業風險管理（ERM）中。

[24] ISACA, "The Promise and Peril of the AI Revolution," 12 September 2023, https://www.isaca.org/resources/white-papers/2023/the-promise-and-peril- of-the-ai-revolution

2. 管理組織（DM.04）：此信任要素要求企業識別並組織結構，以支援數位信任生態系。
   - 實務 DM.04.01：管理組織架構。
     - 活動 DM.04.01.4：建立角色與職責，包括在適當情況下設立由高階主管、業務部門及資訊與技術（I&T）管理階層組成的信任指導委員會（或同等組織），以追蹤專案狀態、解決資源衝突，並監控服務水準與服務改進。
3. 管理數位信任（DM.06）：此信任要素要求組織維護適當的數位信任實務，特別是對於資訊的應用。
   - 實務 DM.06.01：盤點資訊資產。
     - 活動 DM.06.01.3：發現並盤點管理外部關係的合約及其他文件。

新興態勢

在當今快速變動且數位化的環境中，組織需要注意可能影響未來成功的外部環境因素。能夠考量這些因素的組織，將可以提升其敏捷性與韌性。從敏捷性的角度來看，組織能夠更快、更有彈性且更果斷地行動與調整。從韌性的角度來看，組織能夠預測、回應並適應變化或中斷。

在當今快速變動且數位化的環境中，組織需要注意可能影響未來成功的外部環境因素。

新興態勢領域專注於可能在流程與人員層面引發機會的事件與活動，包括內部變化、外部影響及人員層面的偏差。

與人工智慧（AI）相關的新興態勢考量包括：

- 透過生成式與非生成式 AI 來預測新興態勢（例如：界定人工智慧行為者的預期邊界）。
- 分析並驗證流程和結果模型（例如：人工智慧行為者應該達成什麼目標？可能的偏差是什麼？【需注意合理可預期的濫用】）。
- 控制人工智慧行為者的輸入變數、訓練及演化。
- 監控人工智慧行為者的突發行為，並在必要時進行調整，以維持可接受的信任水準。

與聊天機器人的使用案例，下列相關的信任要素與實務非常有用：

1. 識別、評估與管理潛在觸發因素（EM.01）：
   此信任要素涉及識別、評估並管理潛在觸發因素。
   - 實務 EM.01.01：識別並管理內部訊號，以及其相關活動。
2. 檢測和管理與流程和人員相關的緊急問題（EM.02）：
   此信任要素要求企業透過突發事件識別即將發生或已經發生的變化，並加以管理其結果。
   - 實務 EM.02.01：偵測內部突發事件及其所有的相關活動。

賦能與支援

賦能和支援領域是科技增能流程間的動態互連，科技驅動流程，而流程又支援科技的部署與運作。在此領域採納相關實務，有助於在聊天機器人部署至正式運作前，及早發現問題。

賦能和支援領域中與人工智慧（AI） 相關的考量因素包括：

- 將人工智慧宇宙嵌入服務價值鏈與服務管理。
- 將 人工智慧行為者進行定義並描述為流程、服務及整體服務組合的一部分。
- 引導並管控人工智慧的進一步發展（考慮供應商與使用者間的轉換）。
- 持續監控人工智慧的運作，並與新興態勢及人為因素領域建立連接介面。

與聊天機器人使用案例相關的信任要素、實務及活動如下：

1. 管理數位信任生態系目標 (ES.01)：
   此信任要素著重於確定流程與科技目標，並實現預期目標（包含品質考量）。
   - 實務 ES.01.01： 確定流程目標及其所有相關活動。
   - 實務 ES.01.03： 確定流程規格及其所有相關活動。
2. 實施服務和解決方案 (ES.05)：
   此信任要素要求組織規劃、協調並實施服務和解決方案。
   - 實務 ES.05.01： 規劃實施及其所有相關活動。
3. 監控服務和解決方案 (ES.07)：
   此信任要素著重於持續監控服務、解決方案及相關技術的運作，以確保數位信任。
   - 實務 ES.07.01： 監控流程運作。
     - 活動 ES.07.01.1： 管理流程運作的變更。
     - 活動 ES.07.01.2： 監控營運指標與控制措施。
     - 活動 ES.07.01.4： 監控流程的一致性。
     - 活動 ES.07.01.7： 辨識改進項目並納入品質管理流程。

結論

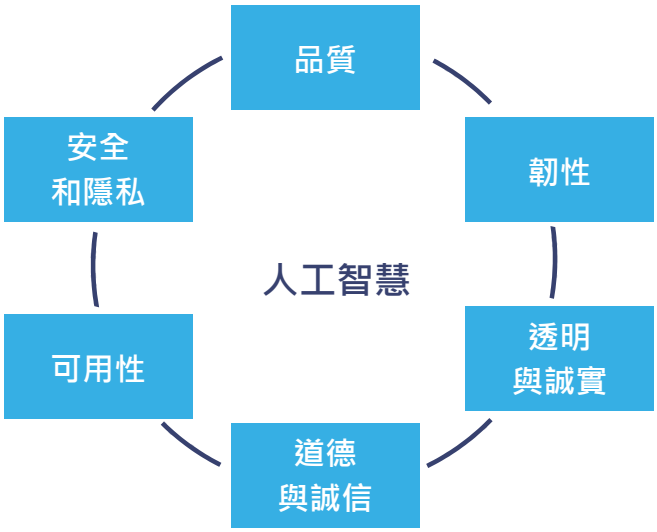
人工智慧並不是新事物，但近期的發展，特別是生成式 AI 的崛起讓市場急速發展。企業必須正視一個現實：至少有部分員工已經使用過某種形式的人工智慧（AI），並且可能讓智慧財產權暴露於風險之中。與此同時，各類型軟體也都逐步內建人工智慧（AI）功能，這要求企業在部署更新之前必須進行盡職調查。此外，隨著解決方案供應商爭相保持其競爭力，整體的競爭愈加激烈。

現實是，人工智慧（AI）已經到來，儘管存在風險，但鑒於其諸多優勢，無論行業或業務職能如何，要全面禁止是不切實際的。相關的立法與標準將成為建立護欄，以實現技術合乎道德和負責任使用的主要驅動力量。而立法和標準很可能會走上類似隱私法規的路徑，一個由複雜且不一致的法律網絡，這必然會為治理、風險管理與法遵（GRC）專業人士帶來困擾。

問題已不再是企業是否會採用以人工智慧（AI）為基礎的技術，而是採用的程度。即使企業不自行開發專屬模型，人工智慧（AI）也已經無處不在，且未經明確許可就使用熱門的生成式 AI 工具現象，已成為影子 IT 的另一種進化。企業需要制定與業務目標一致的人工智慧策略，並必須驗證使用的人工智慧（AI）是否真正解決業務問題，以及是否在可接受的風險容忍度之內。

為了達成這一目標，企業需要一個健全的框架，以確保其符合法規與人工智慧（AI）相關的法令法規要求、國際標準指引，以及客戶合約義務。雖然業界有許多不同領域的框架可供採用，但數位信任生態系框架（DTEF）能協助企業在人工智慧（AI）驅動的流程與系統中，實現品質、韌性、透明與誠實、道德與誠信、可用性、安全和隱私（見圖 9）。

圖9：值得信賴的人工智慧要素



數位信任生態系框架（DTEF） 同時也具備足夠的彈性，可以與其他框架結合使用。採用數位信任生態系框架（DTEF）來導入基於人工智慧（AI）的技術，意味著這些面向將在整個生命週期中得到考量，並且有助於打破組織中可能存在於流程、人員與技術間的孤島。當人工智慧（AI）在企業內被設計、開發與部署時，數位信任生態系框架（DTEF）的價值在於，如果能有效實施，它將幫助組織證明其符合法規、客戶要求、國際標準、內部政策與程序，最重要的是，它能確保解決方案以合於倫理且負責任的方式實現企業業務目標。

致謝

ISACA 謹此致謝：

開發領袖

Chetan Anand

CDPSE, CPISI, ICBIS, ICCP

AVP, Information Security and CISO, Profinch Solutions Pvt Ltd., India

校審專家

Joyce Chua

CISA, CISM, CDPSE, CAEG (Professional), (C)CISO, CFE, CIA, CIMP, CIPM, CIPP(A), CIPP(E), FIP, ITIL, MCP, PMP

United Overseas Bank, Singapore

J. Winston Hayden

CISA, CISM, CGEIT, CRISC, CDPSE

South Africa

Ed Moyle

CCSK, CISSP

Chief Information Security Officer for Drake Software, USA

Geetha Murugesan

CISA, CGEIT, CRISC, CDPSE, ISO 22301:2019, ISO 27001:2013, ISO 31000:2018, ISO 9000:2015

India

Rolf von Roessing

CISA, CISM, CGEIT, CDPSE, CISSP, FBCI

Partner, FORFA Consulting AG, Switzerland

ISACA 董事會

John De Santis, Chair

Former Chairman and Chief Executive Officer, HyTrust, Inc., USA

Brennan P. Baybeck, Vice-Chair

CISA, CISM, CRISC, CISSP

Senior Vice President and Chief Information Security Officer for Customer Services, Oracle Corporation, USA

Stephen Gilfus

Managing Director, Oversight Ventures LLC, Chairman, Gilfus Education Group and Founder, Blackboard Inc., USA

Niel Harper

CISA, CRISC, CDPSE, CISSP, NACD.DC

Former Chief Information Security Officer, United Nations Office for Project Services (UNOPS), USA

Gabriela Hernandez-Cardoso

NACD.DC

Independent Board Member, Mexico

Jason Lau

CISA, CISM, CGEIT, CRISC, CDPSE, CIPM, CIPP/E, CIPT, CISSP, FIP, HCISPP

Chief Information Security Officer, Crypto.com, Singapore

Massimo Migliuolo

Independent Director, Former Chief Executive Officer and Executive Director, VADS Berhad Telekom, Malaysia

Maureen O'Connell

NACD.DC

Board Chair, Acacia Research (NASDAQ), Former Chief Financial Officer and Chief Administration Officer, Scholastic, Inc., USA

Erik Prusch

Chief Executive Officer, ISACA, USA

Asaf Weisberg

CISA, CISM, CGEIT, CRISC, CDPSE, CSX-P

Chief Executive Officer, introSight Ltd., Israel

Pamela Nigro

ISACA Board Chair 2022-2023

CISA, CGEIT, CRISC, CDPSE, CRMA

Vice President, Security, Medecision, USA

Gregory Touhill

ISACA Board Chair 2021-2022

CISM, CISSP

Director of the CERT Division at Carnegie Mellon University's Software Engineering Institute, USA

Tracey Dedrick

ISACA Board Chair, 2020-2021

Former Chief Risk Officer, Hudson City Bancorp, USA

關於 ISACA

ISACA® (https://www.isaca.org) 是一個推動個人與組織追求數位信任的全球性社群。 50 多年來，ISACA 為個人與企業提供知識、證書、教育、培訓與社群，以促進其職涯發展、改造其組織並建立一個更值得信賴及合乎倫理道德的數位世界。 ISACA 是一個全球性專業協會與學習型組織，擴展其在資訊安全、治理、確信、風險、隱私及品質等數位信任領域工作的180,000名成員的專業技能。它已在全球 188 個國家共設立225 個分會。 ISACA 透過其 One In Tech 基金會，為資源不足及代表性不足的人們提供 IT 教育與職涯發展途徑之支援。

免責聲明

ISACA 設計並完成本創作「運用數位信任生態系框架實現值得信賴的人工智慧」（著作），主要作為專業人士的教育資源。ISACA 並不聲稱使用本著作任何內容將確保有成功的結果。 該著作不應被視為包含所有適當的資訊、程序及測試或排除合理指導獲得相同結果的其他資訊、程序與測試。 在確定任何特定資訊、程序或測試的適當性時，專業人員應將本身之專業判斷應用於特定系統或資訊科技環境所呈現的特定情況。

權利保留

ISACA

1700 E. Golf Road, Suite 400
Schaumburg, IL 60173, USA

電話： +1.847.660.5505

傳真：+1.847.253.1755

支援： support.isaca.org

網站： www.isaca.org

參加ISACA線上論壇：
https://engage.isaca.org/onlineforums

X: www.x.com/ISACANews

LinkedIn:
www.linkedin.com/company/isaca

Facebook:
www.facebook.com/ISACAGlobal

Instagram:
www.instagram.com/isacanews/

中文版致謝名單

ISACA 台灣分會高進光理事長

翻譯(按姓名筆劃): 陳政龍

校稿: 黃淙澤

編輯排版: 游恬欣