生成式AI導入指引

企業應具備的AI素養

財團法人資訊工業策進會
INSTITUTE FOR INFORMATION INDUSTRY

中華民國112年07月

目　錄

引水人的領航

人工智慧（Artificial Intelligence, AI）是指模擬人類認知能力的電腦計算技術，它已經改變了商業運作與人力需求方式，預期將更深層的影響產業的發展與競爭力移轉。而生成式 AI（Generative AI）隨著 AI 及運算技術成熟，也帶動 AI 生成內容（AI Generated Content, AIGC）熱潮。由於 ChatGPT 一夕之間成為大眾生活及工作的必備工具，生成式 AI 儼然即將成為下一世代經濟發展的通用技術（General Technology）之一，無法掌握發展趨勢者恐將失去市場競爭力。

放眼未來，大型語言模型（Large Language Model, LLM）將形成各式應用：1. 現有工具的使用方式將產生重大變化，並且將創造許多過去無法執行的專門任務。2. 未來 2-3 年內，幾乎所有白領（主管或基層等）工作都會受到某種程度的影響或重整。3. 惟各地政府法規或監管制度仍然會依國情、地域差異，限制或造成實際導入的時間差。因此，正視其帶來的機會與挑戰，將是企業發展必須掌握的關鍵。

生成式 AI 除了正面的影響，也帶來負面的衝擊，例如：為了讓 AI 學習，不小心洩漏組織的機敏資料；濫用智慧財產權；取代現有人力；道德限制能力弱，AI 可能混淆對於事實的正確認知等。這也是組織在應用 AI 時，必須先有的認知。

本生成式 AI 導入指引希望能幫助高階經理人與重要決策者：1. 了解 AI 對他們的組織影響，2. 如何建構組織 AI 能力。除了解發展 AI 時應考慮的事項，亦為其在應用範圍內的使用奠定基礎。

本指引透過資策會兩院四所及幕僚部門進行共創，就生成式 AI：技術和產業發展趨勢、導入與評估、風險管理及相關資源進行解析，期望能幫助一般企業及公協會組織，從基礎了解到布建運用時應注意事項，降低應用生成式 AI 進入門檻。

在構建此指引之時，我們也體會到產業變動快速，如果想要依循過去做事方法，寫一本完整並完美的指引手冊，並不切實際；故決定先以最快的速度，先提出基礎的指引，歡迎各界給予指教，團隊會持續修訂；後續並將擴增職務（如：產業分析、技術研發等）與領域（如：智慧醫材）等相關應用內容，進行 AI 信任治理案例分享，協助企業與相關組織發展並善用此一技術，最終達成可信任 AI（Trustable AI）境界。

第壹章 生成式 AI 發展趨勢

自 2022 年底 ChatGPT 問世後，短短兩個月內吸引全球超過 1 億活躍用戶數使用，帶動生成式 AI 自動創作內容的新型態生產方式。有鑑於生成式 AI 改變了工作流程與商業運作方式，預期將更深層的影響產業發展與競爭力移轉。生成式 AI 將成為下一世代經濟發展的通用技術之一，無法掌握趨勢者恐將失去市場競爭力。微軟創辦人比爾・蓋茲曾說：「像 ChatGPT 這樣的 AI 出現，就跟 PC 和網際網路誕生時一樣的重要」，足見其對人們的影響將是快速且全面性的。

以下將分別從生成式 AI 基本定義與技術潛力、產業應用機會，以及生成式 AI 導入案例，討論未來生成式 AI 技術與產業發展趨勢

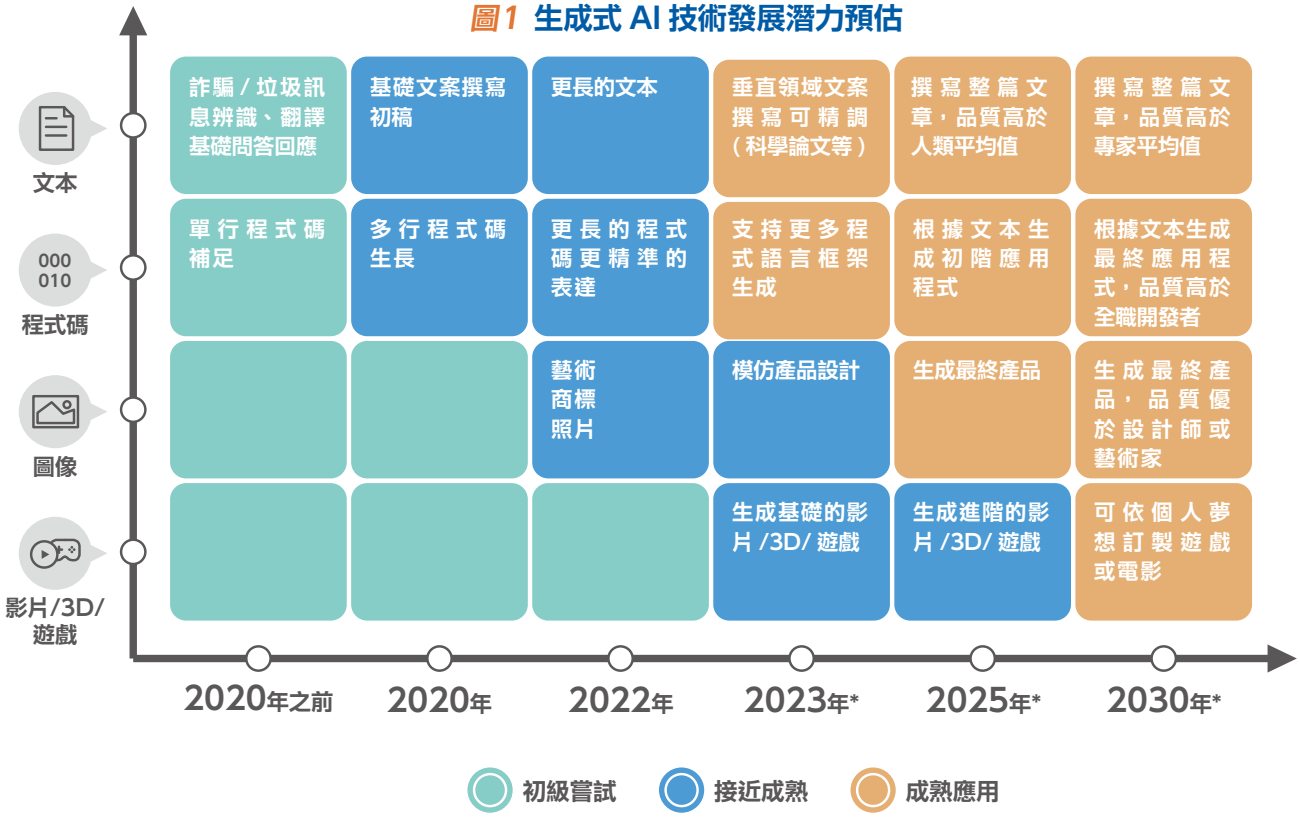
第一節 基本定義與技術潛力

生成式 AI（Generative AI）是 AI 技術的其中一項分支，讓機器學習模型研究歷史資料，除既有決策式 AI 的歸納與辨識功能外，生成式 AI 能進一步具備創造全新內容的能力，常見的技術包含生成對抗網路（Generative Adversarial Network, GAN）與 Transformer 模型等。

ChatGPT 即是基於 Transformer 模型所訓練出的大型語言模型（Large Language Model, LLM），OpenAI 使用大量文本資料預先訓練模型，使模型能夠學習自然語言的結構和規律，從而生成高品質、流暢的自然語言文本。再加上微調後，可進一步提高 ChatGPT 生成文本的準確度和流暢度。

生成式 AI 的應用領域隨技術發展推進逐漸廣泛，主要包含文本、程式碼、圖像、及影片 /3D/ 遊戲等內容領域。根據紅杉資本預測，生成式 AI 在文本與程式碼兩大領域之技術發展最為快速，預計至 2030 年將達到成熟應用，不論文章或程式的生成品質均高於專家平均值。

圖1 生成式 AI 技術發展潛力預估



資料來源：紅杉資本，資策會整理，2023 年

第二節 產業應用機會

雖然生成式 AI 仍在起步階段，相關應用仍不斷推陳出新。有關生成式 AI 帶來的產業應用機會，以下將從健康醫療、數位內容、生產製造、金融稅務與學術教育等產業領域舉例說明：

一、健康醫療

ChatGPT 可用於開發新的醫生筆記應用程式，減輕臨床文件撰寫負擔，包含讓醫生及護理師即時口述轉成文本，用於臨床診斷及醫囑紀錄，且能整合至電子病歷。生成式 AI 也能用於藥物開發，生成與現有蛋白質不同但具特定屬性（如：形狀、大小或功能）之蛋白質，有機會加速新藥開發。

二、數位內容

生成式 AI 可整合內外部資料並自動創造內容，在產品設計、藝術創作皆可提供創造協助。以動畫角色繪製流程為例，生成式 AI 可使用文字描述完善設計、調整數值優化，進行 AI 生成繪製，再輔以生成圖片與草稿進行拼貼，不斷重複步驟、多次生成結果，加速設計產品流程客製化。

三、生產製造

以韓國三星電子和 Naver 合作降低設計開發成本為例，運用 Naver 開發的大型語言模型及壓縮演算法，可加速開發專屬半導體解決方案。在人機對話方面，可增加機器人對環境的理解，例如：聽到「請幫我倒水」，機器人理解語義後，可完成移動、取杯、注水等一連串任務，未來有機會用於工廠搬運作業。

四、金融稅務

在投資理財方面可降低金融知識取用門檻，以生成式 AI 製作財務顧問對話機器人，輸入真實文件加以訓練，協助投資理財。在金流稅務方面，可運用生成式 AI 開發支付資訊整合功能，提供便捷的支付服務。在風險評估方面，能協助金融機構的資料科學家快速建立風險模型，減少開發過程的重複任務，並為利害關係人解釋結果。

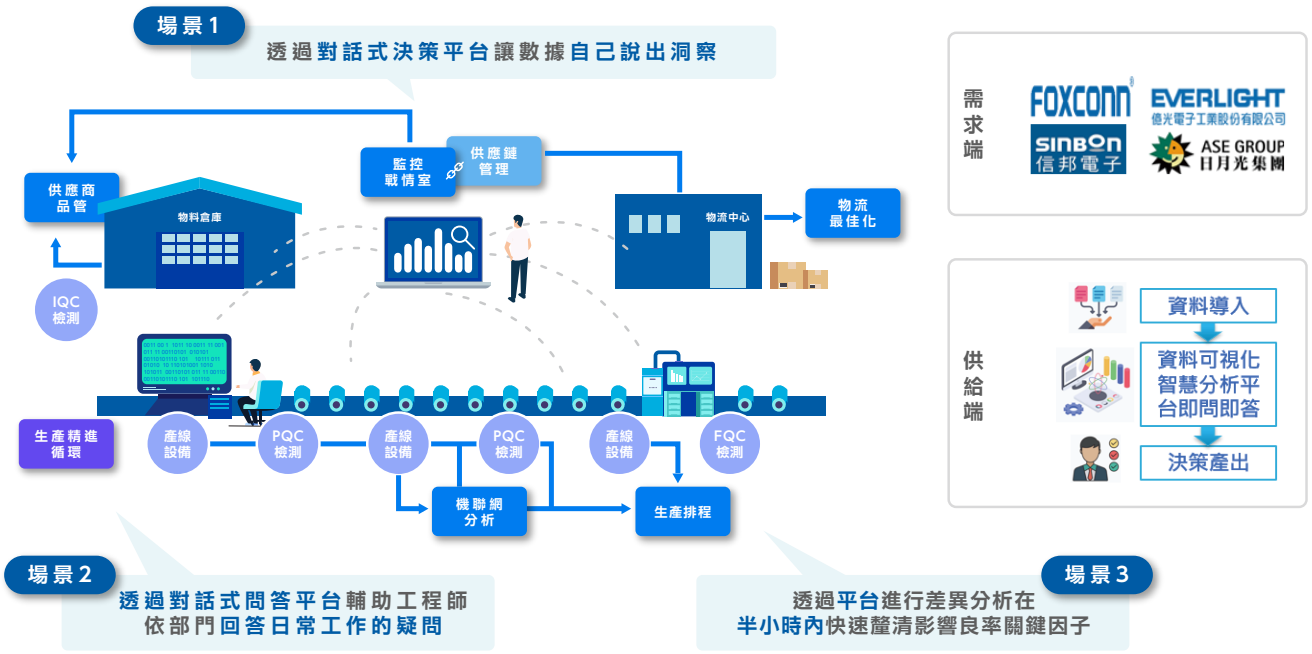
五、學術教育

在學術方面，為研究人員提供基於 GPT-3 的論文查詢助手，整合語義相似度與關鍵字對照的精確搜尋，並輔助摘要或擴大查詢。在語言學習方面，運用生成式 AI 結合語言學習平台，可解析練習題並產生角色扮演和對話，如：虛擬教師。

第三節 生成式 AI 導入案例

生成式 AI 新興應用案例以智慧工廠為例，導入生成式 AI 可實際應用於三大場景，（一）對話式決策平台：透過生成式 AI 分析機台數據並提出改善觀點；（二）對話式問答平台：輔助工程師依部門回答日常工作的疑問；（三）差異分析：在半小時內快速釐清影響良率的關鍵因子。上述三大場景都能透過生成式 AI 協助企業優化營運策略並提升產品品質，加速工廠數位轉型。

圖2 生成式 AI 智慧工廠導入案例

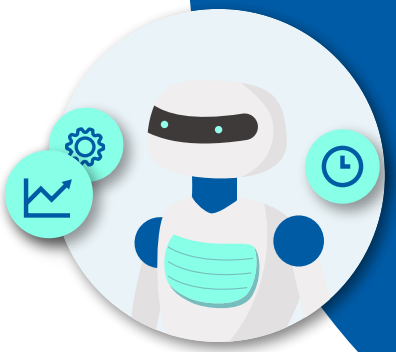


資料來源：各公司，資策會整理，2023 年

生成式 AI 不論對個人或企業都將帶來深遠的影響。於可見的未來，「AI 不會取代人類，但不會 AI 的人類會被取代」，唯有掌握生成式 AI 趨勢並能善用者，方能在這一波革命性的科技洪流中，保持領先與提升市場競爭力。

第貳章 生成式 AI 導入與評估

第一節 整體分析



導入生成式 AI 可望提升業務效能，例如為客戶服務中心提供智慧回答，或創作有深度與個人化的文章。然而，其引入不應忽視評估步驟。從效益角度，需要評估 AI 是否提升工作效率，是否能解決具體問題，或是否增進決策品質。同時，也需瞭解其可能帶來的風險，如資訊安全與隱私權問題。

此外，AI 道德與倫理亦是重要考量。AI 的決策透明度和可解釋性需要被仔細審視，避免偏見的發生。整體而言，生成式 AI 的導入與評估需要涵蓋多面向因素，包括技術效能、成本效益、風險管理，以及道德倫理等。在理解其強大能力的同時，我們也應謹慎對待其可能的衝擊。以下從導入策略、導入評估及自我查核三個面向進行說明。

第二節 策略思考

生成式 AI 為企業提供了創新和優化營運的絕佳機會。然而也帶來各方面的威脅與挑戰，如何降低導入這些技術的複雜度對於確保企業的未來發展至關重要。儘管面臨導入的高度挑戰，仍然可透過正確的專業知識和導入策略來克服，可以從以下幾個策略面向進行思考：

一、組織面

明確定義組織的策略目標、優先事項和生成式 AI 所需的資源分配。組織戰略應包括 AI 如何在組織內推動價值和創新的清晰願景，以及實現特定里程碑的發展路徑圖。定期審查和更新策略，以確保其與不斷變化的業務目標和技術進步保持一致。

在組織面，應強調道德考量、資料隱私以及遵守組織內不斷變化的法規的重要性，並制定 AI 開發和部署指南，確保所有利害關係人了解相關責任，鼓勵公開討論 AI 的道德影響，並促進問責和透明的文化。透過優先考慮最具影響力和投資回報的專案，平衡 AI 投資的成本和收益。探索合作夥伴關係或外部融資機會來支持生成式 AI 的發展計劃，並為生成式 AI 的開發、整合和持續維護分配資源做出明智的決策。

二、技術面

技術面，在選擇生成式 AI 的模型時，需謹慎考慮實際應用的需求，並針對 GPT、BERT、Transformer 等模型不同特性，評估其性能、可延展性和適用性。此外，數據量是否充足、資料品質是否可靠，均會影響到模型訓練的結果。在訓練時，需避免過度精緻和優化，保持生成能力與準確性的平衡。另外也需關注安全與隱私風險，確保符合道德法律標準。在部署時需考慮導入流程和使用過程的監控，並持續測試改進。最後，用戶反饋至關重要，應建立反饋機制並持續優化。綜合考慮上述這些技術構面，組織才能有效導入，實現更好的結果。

三、資料面

資料面，全面的資料治理框架對於確保資料隱私、安全性和合規性以及解決 AI 輸出中的偏見和公平問題至關重要。組織應該針對資料的生命週期進行管理，包含資料取得、儲存和使用的過程，並建立查核和監控的機制，以識別和減少潛在的偏見。

四、人才面

建立能夠推動創新和克服實施挑戰的熟練 AI 團隊。制定吸引頂尖人才的策略，例如提供具競爭力的薪酬方案、持續培訓的機會以及營造支持性的工作環境。透過促進職業發展並提供組織內的成長和晉昇機會來留住有價值的員工。鼓勵跨職能協作和資訊共享，以推動生成式 AI 的創新。為團隊成員建立有效溝通和合作的流程和平台，並通過認可和獎勵創造性解決問題和創新思維來促進持續改進的文化。

合作夥伴可以幫助組織隨時了解新興趨勢和監管變化，透過外部 AI 專家、技術提供者和相關機構建立關係將使組織獲得整合所需的知識、資源和支持，因此，與這些合作夥伴合作，透過最佳實務分享學習關於克服挑戰的各種經驗教訓。

五、管理面

在應用生成式 AI 技術時，管理為不可或缺的環節。管理面需要注意的是如何實施與監控這些策略及措施以確保應用生成式 AI 的效果或風險是在可控制的範圍中。包括對整體 AI 計畫的管理制度、里程碑、預期成果，以及應對可能問題的策略。內部及外部溝通，以及因應法規變遷和技術演進的彈性調整，並為了確保 AI 系統的安全性和正確性，定期的系統審核和風險評估皆不可或缺。

六、政策面

積極參與制定生成式 AI 政策，分享發展經驗和策略思維，並制定負責任的 AI 法規和行業標準，並跟監管機構、行業組織和其他利害關係人合作，幫助制定支持 AI 創新的政策，同時解決道德問題和對社會的影響。同時，對內外進行透明的溝通，例如與監管機構和其他利害關係人進行公開溝通，建立對使用 AI 的信任，並展示生成式 AI 技術的好處，分享 AI 計劃、成功案例和挑戰，展示致力於 AI 實踐的精神。

第三節 導入評估

企業內部導入生成式 AI 的大部分場景，來自於從既有資訊系統嵌入生成式 AI 的應用。以企業內部流程為例，電子郵件系統可提供撰寫資訊的素材。生產力應用程式可根據初稿進行加值與優化。財務軟體則將生成財務報告中的特徵進行描述。客戶關係管理系統則可據此提供與客戶互動的方式建議。這些功能都可提高企業知識工作者的生產力，此外，不同產業都可透過生成式 AI 來重塑組織內部的工作流程。

在導入時，可以根據所需資源的專案到資源密集型的專案依序切入。相關評估構面說明如下：

一、導入方式

依使用情境，選擇合適的導入方式，包含因應特定業務購買軟體工具或串接外部模型至內部系統，以及因應產業複雜性與專業性，可以選擇微調既有開源模型或自行訓練模型。

二、成本費用

依使用的軟體工具、串接外部模型、微調既有模型、自行訓練模型等面向進行思考。使用軟體工具需支付固定訂閱費用；串接外部模型需投入前期開發使用者介面費用與後期模型維護費用與人力；微調既有模型需人力清理、標註資料以及投入模型維護與運算的費用；自行訓練模型則更多人力、資金投入建設基礎架構與模型開發。

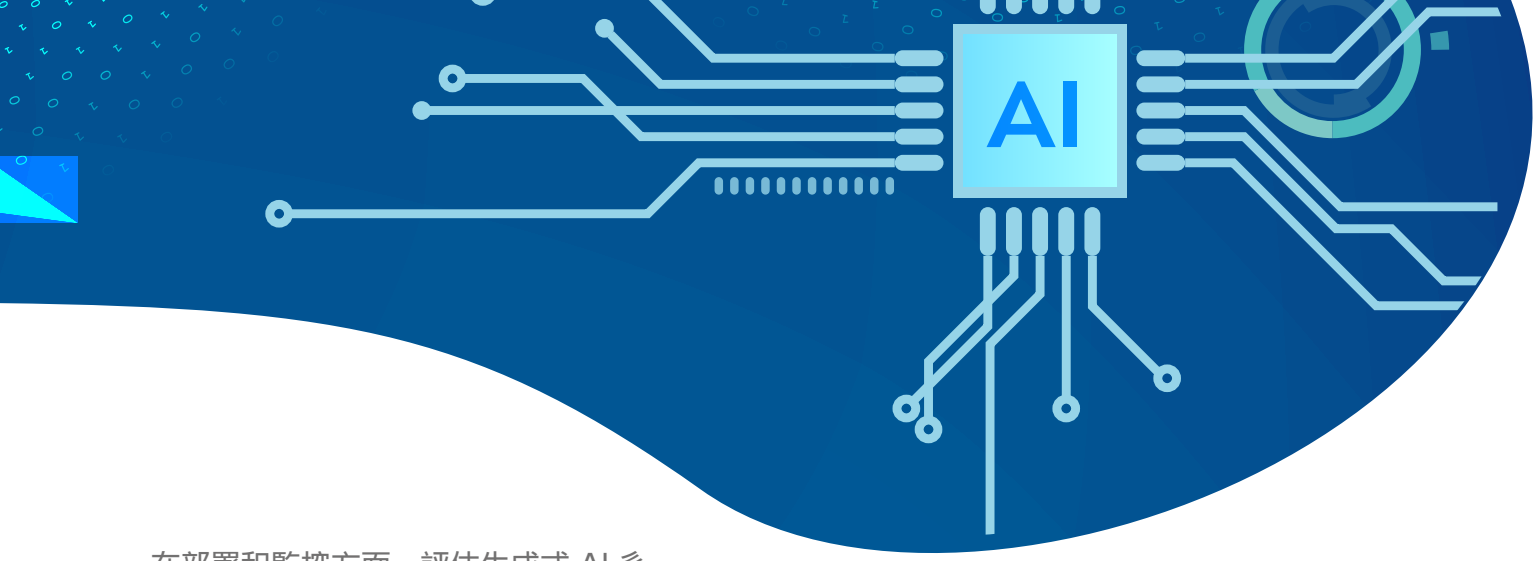
三、技術準備度

在模型選擇方面，選擇適合組織需求的生成式 AI 模型至關重要。不同的模型具有不同的能力和特性，如 GPT、BERT、Transformer 等。評估模型的性能、可擴展性和適用性，以確保其能夠滿足組織的需求。

此外，生成式 AI 需要大量的資料來學習和生成內容。組織是否擁有足夠的資料來支持訓練生成式 AI 的訓練。同時，評估資料的品質、多樣性和可靠性，以確保模型能夠生成高質量和可信的結果。

在訓練和優化方面，生成式 AI 的訓練是一個關鍵的過程。訓練、評估過程中的設置和參數選擇，確定最佳的模型性能。同時，需要注意防止過度精細和優化模型的生成能力和準確性。

在安全性和隱私保護方面，生成式 AI 可能會產生具有潛在風險的內容，如不實訊息、偏見或侵犯隱私等。評估生成式 AI 系統的安全性和隱私保護機制，確保生成的內容符合道德和法律標準。



在部署和監控方面，評估生成式 AI 系
統的流程和監控機制，確保生成模型能夠穩定運行並及
時檢測問題。進行系統性的測試和監控，以追蹤模型性能、錯誤並不斷改進系統。

最後是用戶反饋和改進，生成式 AI 應該具有反饋機制，使用者可以回報問題或提供改進建議。持續評估系統反應和改進，確保用戶的意見得到及時回應並用於優化系統。綜合考慮這些技術評估，可以幫助組織有效地導入生成式 AI，實現更好的結果。

四、資料準備度

使用既有模型或自行訓練模型時，需要準備專用資料。調整既有模型需要蒐集、清理、標註內部資料，建立內部資料集，用來客製化訓練，讓模型符合組織需求；自行訓練模型，除了準備內部資料外，也需要準備大量公開資料進行模型訓練。

五、監控方式

需要人員進行模型生成結果正確性與合適性的檢查；串接外部模型需要存取指令（Prompts）與結果並設定保護機制；調整既有模型需要分類問題並定期檢查模型安全性；自行訓練模型除了需要檢查正確性、安全性、合適性外，還需要特別注意使用公開資料的智慧財產權問題，避免違反法規規範。

六、監管要求

針對數據的隱私和安全問題，以及道德和倫理的問題，很多國家和地區，包括歐盟、美國、中國大陸和 OECD 等，都已經設定了 AI 倫理準則。雖然現在還沒有官方版本的 AI 倫理準則，但是
相關的法規已經開始出現，導入時應密切注意相關法規的最新發展。

第四節 自我查核表

表1 自我查核表

導入策略	準確度		
	(1) 未評估	(2) 評估中	(3) 已有方案
1、組織面	☐	☐	☐
2、技術面	☐	☐	☐
3、資料面	☐	☐	☐
4、人才面	☐	☐	☐
5、管理面	☐	☐	☐
6、政策面	☐	☐	☐

技術整備度	預算編列	技術	資料	流程盤查	監理規範
1、直接使用軟體工具	☐	☐	☐	☐	☐
2、透過API串接外部模型	☐	☐	☐	☐	☐
3、針對開源模型微調	☐	☐	☐	☐	☐
4、自行訓練企業的內部模型	☐	☐	☐	☐	☐

資料來源：資策會，2023 年

第參章 生成式 AI 風險管理

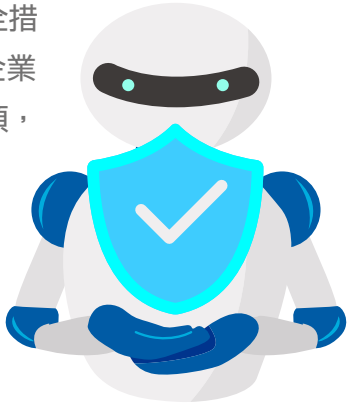
第一節 生成式 AI 風險管理項目

生成式 AI 擁有強大的應用性，然而，我們必須對這項技術所帶來的潛在風險保持警惕。舉例來說，生成式 AI 的濫用可能導致虛假資訊、惡意內容或冒充他人等問題出現。其次，生成式 AI 可能對個人隱私和安全構成威脅。此外，生成式 AI 的演算法也可能受到有意或無意的偏見影響，導致產生不公平和歧視性的結果。除了上述風險，以下是在應用生成式 AI 時常見需要注意和規避的風險管理項目。

一、公平性

AI 服務需要確保不因種族、性別、年齡、宗教、性取向等因素而產生不公平或偏見，必須建立公正的機制，以減少或消除偏見可能性。此外，在產業應用中，需要謹慎處理資料以確保公平性。例如：確保提供訓練資料給生成式 AI 的人在符合一定條件下能夠取得資料；或者平衡控制資料提供者的影響力與市場力量，確保資料的使用具有公平性。

二、隱私性



AI 服務需要確保使用者資料的隱私與安全，這意味著應該採取適當的安全措施，以保護使用者資料不被未經授權的人員存取、竊取或濫用。此外，企業在運用或發展生成式 AI 時，需在有效保護資料的前提下，明確以下事項，被應用於 AI 模型訓練的資料：確定哪些資料在被有效保護的情況下可以應用於 AI 模型的訓練。資料使用的授權範圍：明確員工的授權層級，限制能夠使用這些資料的人員，以確保資料的安全性和隱私保護。隱私提升技術（PETs）的應用：考慮運用隱私提升技術，如資料加密、匿名化和安全協議等，來增強資料的安全性和隱私保護。進一步探討和解決上述課題是十分重要的。

三、安全性

對於意圖不良者可能利用生成式AI來加速網路言論的攻擊。同時，AI系統也可能面臨防止駭客入侵、保護資料隱私、防範對抗式攻擊以及檢測和阻止惡意輸出等攻擊。因此，企業應該實施新的安全措施，包括身份驗證、加密、漏洞修補和監控，以確保 AI 系統的完整性、可靠性和可信度。定期進行安全測試、培訓人員意識、制定政策和遵守合規要求也是確保 AI 安全的重要環節。只有這樣，企業在引入 AI 應用時才能確保 AI 系統的基礎安全性。

四、智慧財產權（IP）確認和授權

利用生成式 AI 所產生的內容、圖片、影像、聲音，甚至是軟體程式，其智慧財產權的歸屬是一個待解的議題。若公司員工在使用生成式 AI 軟體時，將運算結果納入其工作成果，例如公開的報告或文章，是否有可能侵害他人的智慧財產權，進而導致公司捲入相關的法律訴訟？因此，企業在訓練生成式 AI 模型時使用的資料，以及模型輸出的成品，都可能帶來重大的智慧財產權風險，包括侵犯著作權、 商標、專利或其他受法律保障的素材。即使使用供應商所提供的生成式 AI 工具或應用，企業仍需充 分了解該公司對於訓練資料和生成的成品上的智慧財產權限制。

五、可解釋性

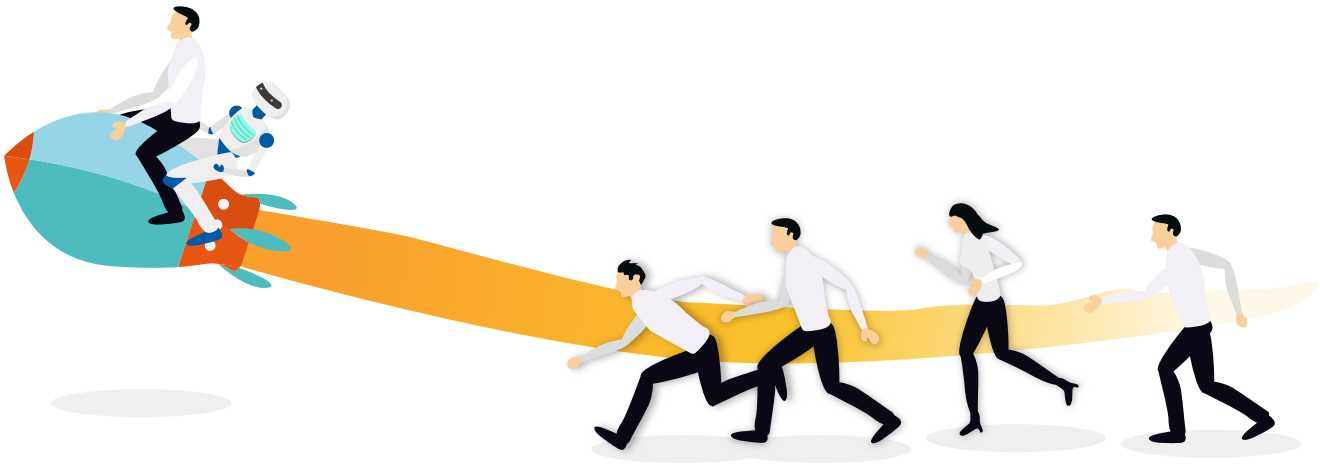
AI 服務需要提供透明的解釋，讓使用者了解 AI 如何作出決策，並且需要提供可解釋性，以讓使用者能夠理解 AI 生成的結果。在這波生成式 AI 趨勢下，生成式 AI 倚賴 LLM (Large Language Model) 進行內容生成和應用。然而，由於 LLM 具備數十億參數的神經網絡規模，解釋生成式 AI 提供的答案或內容成為引入生成式 AI 的重要挑戰。

因此，在企業應用 AI 時，可以借鑒安全 AI 開發框架和可解釋 AI（Explainable AI, XAI），以確保生成式 AI 的可解釋性，從而有效管理風險。此外，透過軟體物料清單（Software Bill of Materials, SBOM）等管理機制，可以獲得對 AI 模型的黑盒結構進行掌握，降低風險程度。藉助相關的可解釋 AI 軟體框架或使用 SBOM，企業可以在流程、資料和模型推論後的結果中找到可解釋性的依據，從而在所有情況下瞭解 AI 運行時的結果。

六、準確性與可靠性

生成式 AI 是一種能夠自我學習並生成新內容的 AI 技術。透過訓練和學習各種資料內容，它能夠產生新的、以前未見過的結果。然而，由於生成式 AI 本身的複雜性和創新能力，對於相同的指令或提示，可能產生不同的答案，這使得企業難以準確評估生成式 AI 的輸出結果的準確性和可靠性。

因此，AI服務需要確保生成的結果是準確且可靠的。為了確保這一點，在建立AI 服務時需要特別注意資料、資料偏見的問題，避免 AI 學習到不準確或不一致的回應。此外，需要設計並實施有效的監控機制， 以確保 AI 在運行時的可靠性。這包括定期審核 AI 的輸出，並對 AI 的錯誤進行調查和分析。通過監控和調整，我們能夠及時發現並修正 AI 的問題，從而提高其可靠性。



七、企業組織影響

基於 ESG 對整個產業的趨勢影響，在環境方面，生成式 AI 在運算時需要大量資源。因此，在應用生成式 AI 時，企業需要注意能源使用效率，並尋找更環保的運算方式，例如使用潔淨能源的資料中心。在社會方面，生成式 AI 的使用對現有工作人員的職能帶來新的挑戰。因此，組織需要重新檢視現有工作職能，重新探討職能的範疇和內容，並提供新的培訓和培育，以幫助員工適應生成式 AI 所帶來的新挑戰。

在治理方面，生成式 AI 能夠自動生成不同的內容和推論。因此，當組織應用生成式 AI 時，可以重新梳理現有工作流程，優化或重新設計工作流程，以充分利用生成式 AI 的能力提升企業效率。此外，更重要的是，企業應建立相應的政策和監管機制，以確保 AI 的使用符合法規和道德標準，並尊重資料隱私。

第肆章 生成式 AI 資源參考

第一節 生成式 AI 資源盤點

生成式 AI 需要大量資料與算力，促進雲端服務、數據資料庫、使用教學等產品發展，表 2 盤點生成式 AI 運算、數據、學習三方面資源供產業參考。

表2 資源盤點表

運算資源		
雲端服務	Amazon AWS	https://aws.amazon.com/tw/
	Google Cloud Platform	https://cloud.google.com/?hl=zh-tw
	Microsoft Azure	https://azure.microsoft.com/zh-tw

數據資源		
資料集平台	Kaggle	https://www.kaggle.com/
公開資料集	SQuAD:問答資料集	https://rajpurkar.github.io/SQuAD-explorer/
	ImageNet:視覺資料庫	https://www.image-net.org/
	Common Crawl:網頁資料集	https://commoncrawl.org/
	COCO Datasets:圖片資料集	https://cocodataset.org/#home
	Hugging Face Datasets	https://huggingface.co/docs/datasets/index

學習資源		
線上課程平台	Coursera	https://www.coursera.org/
	Edx	https://www.edx.org/
	DeepLearning.AI	https://www.deeplearning.ai/
	Tibame	https://www.tibame.com/
	udemy	https://www.udemy.com/zh-tw/
	Fast.ai	https://www.fast.ai/
廠商課程	Hugging Face NLP Course	https://huggingface.co/learn/nlp-course/chapter1/1
	Google Cloud Skills Boost	https://www.cloudskillsboost.google/course_templates/536
學術研究	arXiv	https://arxiv.org/

資料來源：資策會，2023 年

附錄

一、AI 模型之開發流程

AI 模型的開發流程，可大致分成以下幾個階段：

1. 定義問題：確認要解決的需求問題，如影像辨識、語音識別、文本辨識等。
2. 數據收集：收集大量有關問題的資料如影像、聲音、文本等。
3. 數據清洗：對資料進行處理和清洗，如去除噪音或不必要的干擾訊息，做訓練之準備。
4. 數據標記：為資料進行正確的標記，如對影像進行標記，使 AI 能識別影像中的物件物體正確學習。
5. 模型選擇：根據問題資料選擇適合的 AI 模型，如 DL、卷積神經網路等。
6. 模型設計：設計 AI 模型的架構和參數，例如神經元的數量、層數等。
7. 模型訓練：使用資料對模型進行訓練，調整模型參數以提高準確性和效能。
8. 模型評估：對模型進行評估，如進行準確性和泛用性等測試。
9. 模型優化：對模型進行優化，如調整參數、改進演算法等，以提高正確率及效率。
10. 模型部署：將訓練好的模型部署到實際場域或應用中，如將影像辨識模型應用到 CCTV 監視器中。

以上為基本 AI 模型開發流程，然實際運用情況仍需考慮軟硬體資源、導入現場需求及系統維運等落地使用之問題。

二、生成式 AI 對 AI 模型技術的衝擊

由模型技術的觀點來看，生成式 AI 帶來的影響可分成兩個面向：

1. 實務應用上，大型 LLM 的訓練成本太高且不符合產業經濟效益，故多數產業應用會朝向超小型 LLM 的趨勢邁進，也會逐漸走向本地端伺服器或邊緣端的方向走，如彭博（Bloomberg）近期推出的 BloombergGPT，即為專門針對商務金融資料訓練生成，量身打造的 LLM，目的為處多元化金融業相關的 NLP 任務問題。

2. 從 AI 的技術與模型發展方向來說，可由分辨式 AI（Discriminative AI）與生成式 AI 的不同論述起。
   (1) 分辨式 AI，類似於傳統的專家系統，透過一連串嚴謹的資料蒐集、標記，提供給 AI 學習以供「分辨」。此類 AI 系統通常具有特定目的，須針對此目的提供高品質具標籤的資料加以訓練，也因此分辨式 AI 產生的模型準確度較高，實務應用上也常用於判別資料正確與否、有明確結果輸出之應用場合，如車牌辨識、醫療影像辨識等系統。
   (2) 生成式 AI 運用 LLM 模型，則不需要人工提供資料或標記，而是運用海量的既有資料讓機器自主學習，產生資料與資料中的關聯或發掘隱藏的關係。少了數據需標註的限制後，訓練的資料倍增，搭配模型技術的演進與電腦算力，讓生成式 AI 得以實現。當然，過程中若予以標記資料或是強化學習（Reinforcement learning）等輔助訓練，將有助於加速 AI 模型收斂學習。

生成式 AI 對於技術面的衝擊，可歸納為：1. 中間任務的消失、2. 開發方式趨於統一。以自然語言來舉例，傳統 NLP 需分別進行諸如正確分詞、詞性標註、專名識別（Named Entity Recognition, NER）、句法分析、指代消解、語義 Parser 等眾多中間任務，拚湊起來再加上生成翻譯等才能完成最終 NLP 任務。LLM 的出現，打破了以上的限制，透過大量的預訓練，中間任務的學習已內化吸收至 Transformer 的模型內，可直接端對端解決最終任務，中間任務的功能就此消失。而 GPT 模型誕生後，技術開發方式也趨向統一，即不同子領域的特徵提取器逐漸由 LSTM/CNN 統一到 Transformer 上，也觀察到 Transformer 已逐步取代 CNN 等其他模型，有統一越來越多領域的趨勢。訓練與應用上，則朝向模型預訓練階段，加上視覺應用領域微調（Fine tune）或應用 Zero/Few Shot Prompt 模式來進行。在 GPT 中，所謂 Prompt 即指一個短語或一組單詞，可被用來引導模型生成特定類型的文字。

綜合以上趨勢可發現，現階段各大開源社群與技術產業，無不極力研發能於本地部署的 ChatGPT 供各應用所用。然承上分析可知，採用 ChatGPT 主要需要解決兩個問題：一為基礎的 LLM 的選擇（模型為何）、其次為 Instruction 訓練的操作，方法上主要仍採用微調（Finetune），差別在於數據集的取得；OpenAI 的 ChatGPT 採用 GPT-3+OpenAI 自己的數據集，以下簡列其他多樣方案供參考。目前開源社群使用最多為 LLaMA+Alpaca 這套方案，然僅有 Dolly 2.0 版可供商用，其餘各項皆不可商用，需特別注意。

各語言模型與訓練集整理

名稱	項目地址	基礎模型	訓練方法/資料集
Alpaca	https://github.com/tatsu-lab/stanford_alpaca	LLaMA	Alpaca
ChatGLM	https://github.com/THUDM/ChatGLM-6B	GLM	自定義資料集（1T）
Dolly	https://github.com/databrickslabs/dolly	GPT-J 6B	Alpaca
BELLE	https://github.com/LianjiaTech/BELLE	BLOOM	Alpaca轉中文+自定義數據集（0.5 ~ 2M）
OpenChatKit	https://github.com/togethercomputer/OpenChatKit	GPT-NEOX/Pythia	OIG-43M
FastChat/Vicuna	https://github.com/lm-sys/FastChat	LLaMA	shareGPT（70k）
gpt4all	https://github.com/nomic-ai/gpt4all	LLaMA	自定義資料集（800k）
lit-llama	https://github.com/Lightning-AI/lit-llama	Lit-LLaMA	Alpaca

資料來源：資策會，2023 年

三、軟體技術之發展建議

1. 關鍵軟硬體科研仍為基盤，需持續深耕

回首 AI 技術的發展歷程，自 1956 年 AI 元年開始，經歷了多次低谷與熱潮。近期 AI 挾著 ChatGPT 的爆紅而再度引起廣泛關注，細究其背後技術脈絡可略知一二。AI 自 1956 年受各界關注寄予厚望後，不久即因未達各界預期的效益而迎來了第一次的寒冬，直至 1980 年代的專家系統出現，透過專業領域知識、準則的導入，奠基 AI 在專門領域的運用基礎；然此時的 AI 系統，僅能就預先考慮過的問題予以解決，尚未具備自主學習及訓練能力，應用仍相當受限。

此後伴隨著網路 2.0 熱潮，AI 光芒略顯黯淡，進入了第二波的寒冬低谷。直到 2010 年代後期，得益於硬體運算能力提升與感測等大數據的可用性，AI 進一步往 ML 及 DL 方向發展。而後 Google 團隊於 2017 年論文《Attention Is All You Need》中提出的 Transformer 模型即是一種基於注意力機制的深度學習模型，廣泛用於自然語言處理任務，而大量的訓練資料與運算資源在 Transformer 技術中至關重要。以 Google 為例，其使用了龐大的多語言網絡資料集，包括翻譯文件、網頁內容、電子郵件等，同時利用其在線上平台所收集到大量用戶生成的翻譯數據，如網站翻譯和手機應用程式中的用戶翻譯，以擴展訓練資料；在硬體資源方面，Google 使用強大的分散式運算集群進行訓練。這些集群由數千台電腦組成，具有高度平行的運算能力。

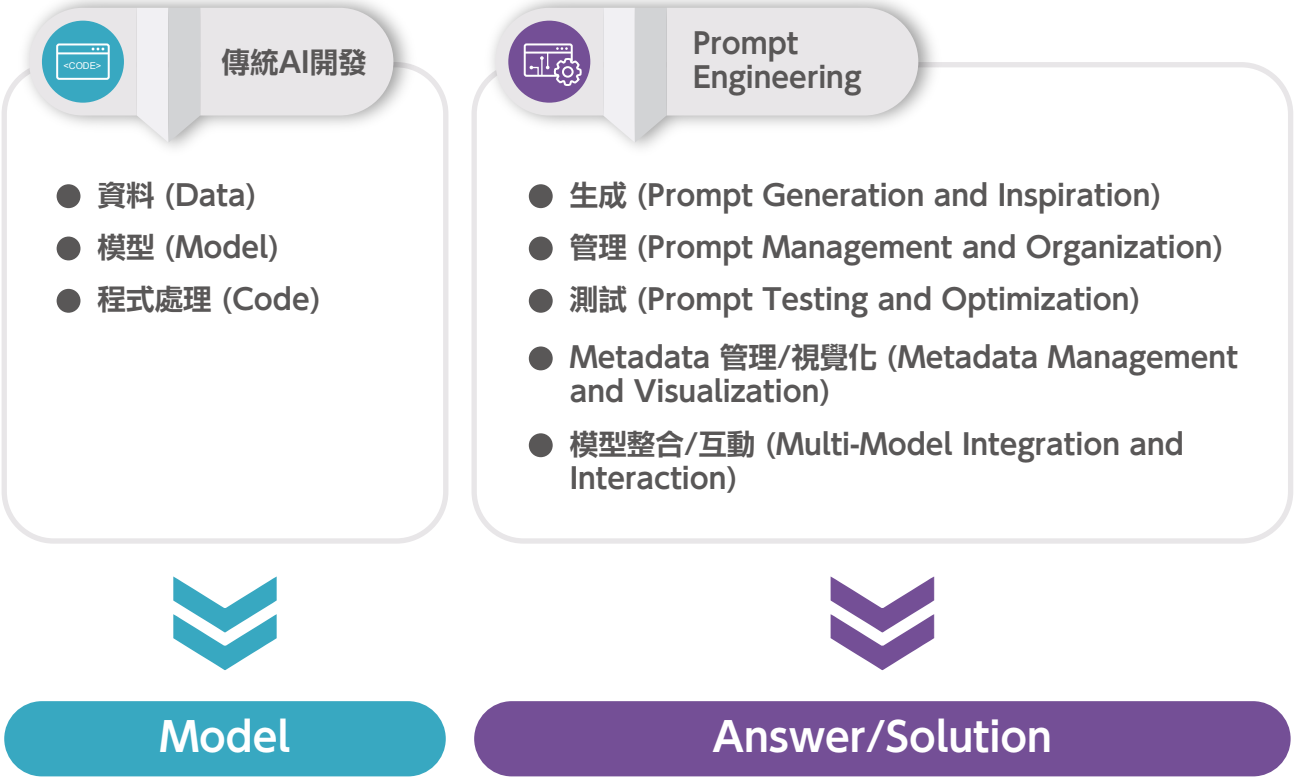
Google Transformer 技術的發展推動了大型語言模型的興起，這些大型模型基於 Transformer 架構，具有數十億或數百億個參數，能夠在多種自然語言處理任務中表現出色，這些模型的發展不僅驅動了自然語言處理的進步，還為人們帶來了更多的機會和挑戰，並在人工智慧領域產生重大影響。也就此，DL 導入至 NLP 範疇，各類研究如 Bert、GPT 等預訓練模型的提出，為 LLM 的研究及應用開啟了大門。

回顧過往歷程可知，基礎科研是一條漫長的累積長路，不論是軟體面的演算法、模型、訓練及效能的提升，乃至於硬體面的晶片及基礎算力發展，皆需要長時間的努力與累積，絕非一蹴可幾，若無過往的投入與耕耘，將難以達到目前的成果。AI 發展經歷過兩次的寒冬低谷，更讓人深刻體會到其中的艱辛與不易。

2. AI 開發方式改變

承前面章節所述，傳統 AI 開發需針對特定領域，由資料（Data）、模型（Model）、程式（Code）三大面向著手進行，才得以開發出適用之 AI 模型。然生成式 AI 的出現，將徹底顛覆 AI 的開發與運用方式，未來將不再仰賴研發團隊進行基礎模型訓練，轉而由 Prompt Engineering 工程方法取代之，其運作方式可參考下圖：

圖3 Prompt Engineering 所帶來的轉變



資料來源：資策會，2023 年

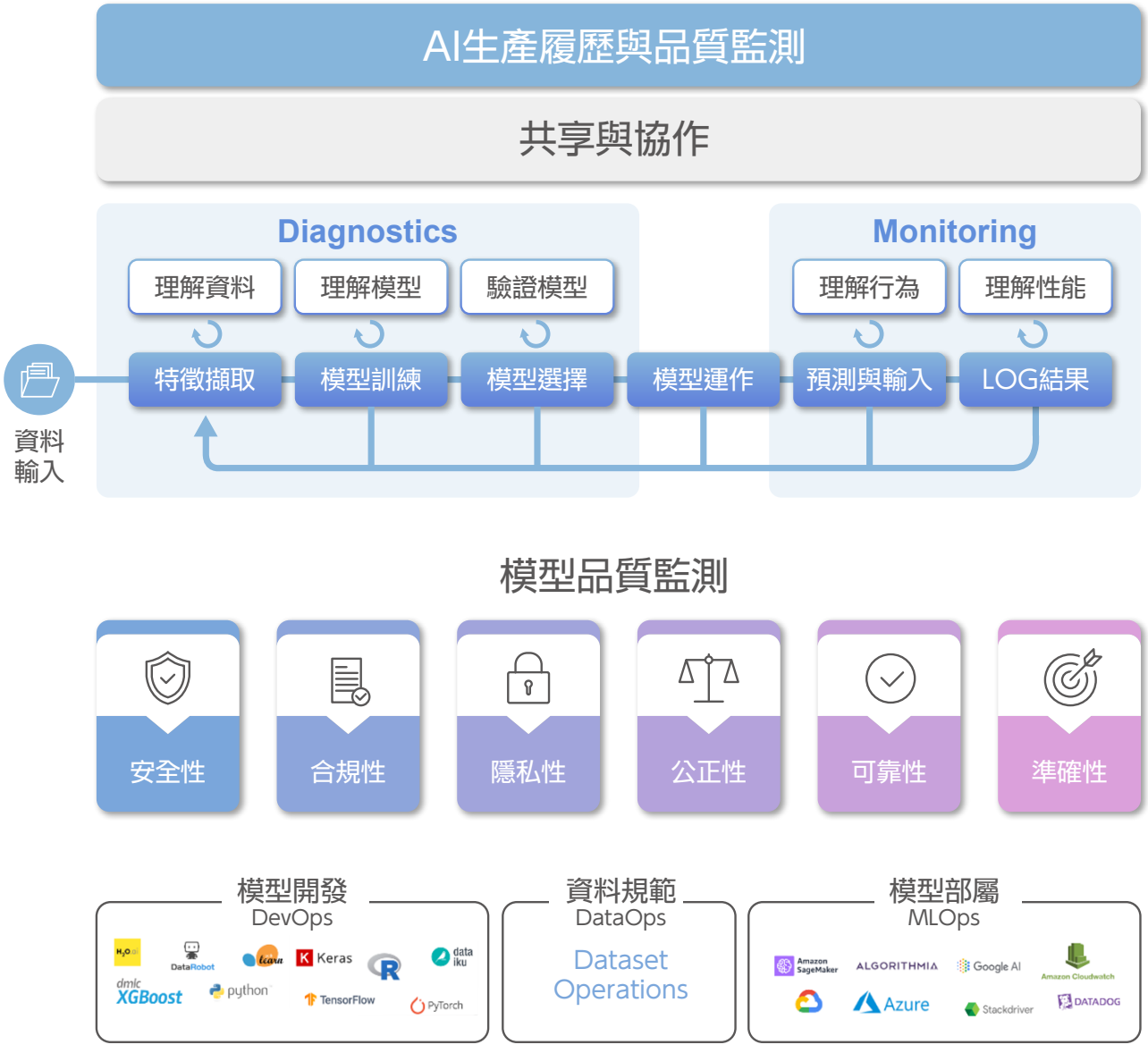
未來研發人員將透過生成式 AI，產出所需要的解答（answer）或解決方案（solution），並專注於查核確認 AI 的產出結果，成為審核者（reviewer）的角色。如何妥善有效地使用生成式 AI 進而管理組織生成式 AI，將會是未來研發所需關注與鑽研的方向。同時，開發過程中若涉及將使用者所輸入資料再訓練之情形，須確保使用者所提供的輸入資料符合相關法規和倫理要求，並採取必要的隱私保護措施，例如資料匿名化和加密，以避免個人身份和敏感訊息的洩露，以及限制資料的存儲和使用期限。開發者應透明地告知使用者所輸入資料的使用方式和目的，取得使用者的明確同意，並遵守隱私政策和相關規範，保護使用者的資料隱私權益。

3. 可信任 AI 重要性躍升，AI 生產履歷至關重要

隨著生成式 AI 的發展，AI 的可用性與取得性也越趨容易，相對的，AI 所衍生的惡意使用、隱私、資安、偏見、智財權等眾多風險議題隨之而來，也成為各方關注的焦點，各國政府及單位亦紛紛提出「可信任 AI」觀點與佈局。

進一步從技術觀點來探討可信任 AI，著重於「開發 AI 與使用 AI 的技術責任」面向，然由於 AI 的開發使用多發生於內部且易有盲點，故光靠研發團隊的自我認知與自律是不夠的，需取得團隊外部公正的品質憑證，進行「全方位的 AI 系統生命週期監管」，也就是「AI 生產履歷」的取得，在 AI 開發使用的前、中、後期，分別採納對應的工具與流程紀錄監管，建立有信任依據的 AI 產品或系統。

圖4 AI 生產履歷與品質監測架構



資料來源：資策會，2023 年

四、資訊安全之發展建議

1. AI 導入機會與風險並存，資安也要做到軟硬整合

AI 在過去 10 年帶來革命性突破，應用層面也從金融、醫療、交通、智慧製造等逐步向大眾擴散；從協助企業進行決策分析，逐步成為工作者的重要輔助工具，從而創造出龐大價值。

但在大量運用 AI 的同時，也因資料紀錄越詳細，資料外洩、隱私疑慮等風險隱憂也一一顯現，如何確保 AI 應用安全亦成為企業主的頭痛問題。

多數風險來自於不一致。有效的 AI 風險管理，可以從兩個構面切入，一是資料本身能不能信任；二是資料環境值不值得信任；前者需要落實企業的資料治理，後者則與企業整體資安部署密切相關，兩者並行才能在導入 AI 應用的過程中做到資訊安全的「軟硬整合」。

隨著 AI 應用發展將面臨四個主要資安挑戰趨勢：

(1) 仰賴雲端運算，延伸資安威脅

AI 應用大量依賴雲端、邊緣技術，使得與這些資料相關的保護和治理，從資料發現到分級分類、從存取控制到資料審計，都需要改變舊有的模式來應對快速生成的巨量資料帶來的變化，與串流間的資安防禦弱點。

(2) 攻擊手段多樣，傳統資安失靈

AI 面臨的攻擊手段也不斷翻新，像是對抗樣本攻擊（Adversarial Examples Attack）、模型竊聽（Model Eavesdropping）、模型詐騙（Model Spoofing）、訓練數據攻擊（Data Poisoning）、存取控制攻擊（Access Control Attack）等，不同攻擊手法與途徑增加防護難度，也導致傳統安全防禦手段如單點、靜態防護等特點，在應對大資料環境下的安全威脅時，防護效果已不敷效益或失效。

(3) 資料跨平台流轉，追蹤難掌控

各種巨量數據是 AI 運作的基礎，資料在同一平台或跨平台、跨公共雲間轉換，讓資料保護與威脅監控更具挑戰，也使得資料追蹤變得十分不可控，加上可信任的 AI 也需避免開發偏誤、濫用個資、倫理和道德議題等，讓 AI 治理更勢在必行。

(4) 存取方式演變，安控藏風險

當數據服務讓存取變得更方便時，也意味著便利性將成為安全控制的薄弱地帶。為了更方便、靈活運用各種 AI 模型數據，將透過 API（應用程式介面）來提供多元服務；然而從資安角度來看，開放的 API 對駭客來說是個入口，存在資安風險，嚴重時將導致機敏性資料被盜取，甚至造成營運中斷。

未來，企業在應用 AI 之際，本身也處於一個 IT（Information Technology）、OT（Operational Technology）、CT（Communication Technology）混合的新環境，如何建構資料安全與環境安全的「軟硬整合」治理策略，將攸關企業能否掌握商機，贏在數位信任的起跑點。

五、對 AI 素養之發展建議

隨著 AI 蓬勃發展，從各行各業研發與導入、公部門與縣市政府應用、到一般大眾也想多了解 AI 如何使用以及實質在生活應用方面能有所助益。近期由於生成式 AI 的發跡與使用門檻日趨便利，讓產業與社會大眾更想積極去了解、分析、使用與優化，但在過程中仍須有一些基本的素養與培訓課程，在使用生成式 AI 的同時，能保護自身在可信任的環境下去使用。

1. 公民素養課程

(1) 公民素養課程目標：讓大眾學習如何運用 AI 在日常生活中，再進一步介紹生成式 AI 之功能、案例以及使用上需要留意之處。

(2) 公民素養課程對象：一般民眾、軍公教人員、在學生。

(3) 公民素養課程規劃內容

1) AI 趨勢與概論：了解 AI 的基本概念、原理和應用領域，以及未來發展趨勢。

2) AI 生活應用：介紹 AI 在醫療、教育、金融、旅遊、製造等產業的應用案例，以及 AI 對個人生活的影響與挑戰。

3) 生成式 AI 介紹（如：ChatGPT）：介紹 ChatGPT 的功能、應用案例，以及 ChatGPT 使用上的限制與風險。

4) 倫理和社會責任：認識生成式 AI 技術所涉及的倫理和社會責任問題，如假新聞、網路暴力、隱私侵犯等，探討使用生成式 AI 需注意的社會議題以及應保護的公民權益。

2. 企業用戶課程

(1) 企業用戶 AI 素養課程目的：是讓企業培訓內部員工對 AI 技術和應用的理解和應用能力。再進一步讓企業學習何謂生成式 AI，如何運用生成式 AI 於現行營運或治理業務、並設計生成式 AI 於案場導入與驗證規劃、測試信度與效度、評估規模化協作契機。

(2) 企業用戶課程對象：各行各業單位、產業公協會、資服業者和新創。

(3) 企業用戶課程規劃內容

1) AI 產業應用案例：介紹生成式 AI 於各產業的應用案例，包括但不限於銀行業、零售業、製造業、醫療業、教育業、金融業等。

2) AI 信度及效度評判：理解 AI 信度評估的基礎概念、AI 模型訓練中的偏差與變異數，以及如何對 AI 產出進行評估和驗證。

3) AI 技術應用實作：介紹機器學習（ML）和 DL 常用演算法、如何篩選具最適合特徵的資料進行建模和優化，以及實作應用案例介紹。

4) AI 倫理與法律：說明企業在使用生成式 AI 所需肩負的責任與義務，包括道德和社會責任、資料安全保護以及法律合規性。

六、對產業資料生態系之發展建議

AI 產品倚賴大量資料進行訓練，故產業界與公部門能否共同協作資料生態系之建設，與 AI 能否蓬勃發展密不可分：

1. 促進資料共享與流通

(1) 解決產業資料釋出的不信任因素，包括資料權利定性與歸屬不明、缺乏經濟誘因、以及擔心資料被不當濫用等情況，藉由法規調適消弭前述問題，逐步推動產業資料共享與利用。

(2) 推動可受信任的資料中介服務：在資料主體與資料利用者之間建立橋接服務，設法於資料保護與資料應用中取得平衡，建立資料主體對於資料利用過程之信任，於符合個人資料保護規範前提下促進個人資料流通。

2. 形成產業 AI 應用之資料治理與管理規範

(1) 透過產業各界共同討論逐步形成初步規範，確保 AI 輸入與輸出的資料來源、品質（正確性、完整性）、偵測並降低資料選擇偏差。

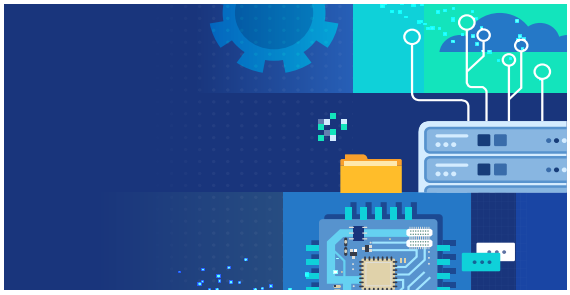
(2) 建立法遵與合規機制，檢視 AI 所利用資料之適法性：確保資料輸入與輸出人工智慧時無侵害第三方權益，如資料隱私 / 保護、商業機密、智慧財產權等。

(3) 建置資訊安全管理措施：AI 輸入與輸出資料整體過程中皆應設有適當安全防護，尤其涉及高風險資料。

3. 拓展 AI 資料生態系

(1) AI 資料生態系應注意①資料利用之公平性，確保需要資料訓練 AI 的人，在符合一定條件下也能取得資料；②平衡控制資料者的影響力與市場力量，確保資料近用公平；③確保貢獻 AI 資料發展之人也能獲得回饋。

(2) 透過法規調適、提供產業界發展指引與契約參考範本等作法，引導產業 AI 資料發展逐漸形成良性循環。



發行單位｜財團法人資訊工業策進會
發行人｜卓政宏
總編輯｜楊仁達
編輯委員｜何玲玲、林玉凡、洪春暉、張育誠、蒙以亨、蕭宏宜
（依姓氏筆畫排列）
編輯群｜王妍文、朱師右、何文楨、吳佩青、李珮瑄、林念潔、
（依姓氏筆畫排列）林逸均、林敬文、徐志浩、張乃文、陳宥蓁、黃世豪、
黃芳蘭、楊中傑、楊秉哲、楊淳安、楊凱婷、楊嘉栩、
葉宗翰、劉芸君、蕭淑玲、戴偉峻、鍾陳威、韓揚銘、
顏瑄、魏徹、顧振豪
地址｜10622 台北市大安區和平東路二段 106 號 11 樓
電話｜02-6631-8168
傳真｜02-2735-0655
網址｜https://www.iii.org.tw
Email｜contactus@iii.org.tw

財團法人資訊工業策進會
INSTITUTE FOR INFORMATION INDUSTRY