本所AI應用參考指引探討

運輸科技及資訊組 張益城 林詠純

研究期間114年2月至115年2月

摘要

生成式人工智慧之發展，對既有著作權法與行政法制提出結構性挑戰。本文以生成式人工智慧之生命週期為分析架構，區分模型訓練階段與實際上線運作階段，檢視各國法制對訓練資料使用、生成內容侵權風險及責任歸屬問題之回應。比較法觀察顯示，各國多透過合理使用、資料探勘例外或非享受性利用等既有理論，建構具條件限制之訓練合法性評價框架，惟訓練階段之容許並不當然延伸為生成結果之合法性，生成內容是否構成侵權，仍須回歸傳統侵權判斷架構。近年政策文件雖普遍將透明性與偏誤管理視為重要之前置治理工具，然其功能在於提升制度之可檢驗性與修正能力，而非確保人工智慧輸出本身即為正確或正當。在高度自動化與多方參與之運作模式下，既有以過失為核心之責任法制逐漸顯現侷限，補償導向治理雖可作為事後救濟之制度回應，仍無法取代決策形成過程本身之正當性要求。因此，本研究認為，公部門導入AI不應只是追求自動化效率，更應建立明確的「人為介入機制」，確保每一筆行政決定都能回歸法治國原則、保障權益的關鍵防線，藉以維繫可歸責之人類判斷、正當法律程序與行政責任體系。

關鍵詞：

生成式人工智慧、智慧財產權

一、緒論

生成式人工智慧（Generative AI）自 2022 年 ChatGPT 推出以來，迅速由學術與技術語境進入公部門、司法與行政決策等高風險領域，其引發的法律與規範課題已遠超過單一侵權或倫理問題，而對法治國（rule of law）核心原則形成結構性挑戰。一方面，生成式 AI 基於龐大資料訓練與概率性推論，能產出與人類創作高度相似的內容，卻在現有法律架構下，造成法律適用與責任歸屬判斷的模糊與分離，進一步影響法律權利義務的明確性及可預期性。這種效應不僅涉及著作權或合約等傳統法律領域，更深刻地牽動誰為決策負責、如何保持法律適用的一致性與可課責性(accountability)等核心法治價值。

其次，在公共部門與司法運作中，AI 工具若完全自動化或以最低人為監督介入，將可能侵蝕法律規範對行政行為的控制機制。例如，在法院系統或律師實務中，生成式 AI 已出現因提供虛構案例、錯誤引用而遭到法官批評或制裁的實際情況，顯示 AI 的不確定性與錯誤風險對司法正當程序與信賴保護帶來實質威脅。此類現象反映 AI 系統在高階法律判斷中難以符合現行法律對於理由形成、證據評價及法律適用精確性的要求。

在制度治理層面，許多國家與地區開始嘗試以立法或政策手段回應生成式 AI 的挑戰。例如在人工智慧基本法草案評估中，即明確提出 AI 的研發與應用必須在系統的整個生命週期中尊重法治、人權及民主價值觀，強調技術應在法律規範之下運作，而非取代法律本身的判斷功能。

因此，生成式 AI 對法治國原則的結構性衝擊不僅體現在侵權責任的模糊與法律界限重塑上，更涉及法律適用的機制設計、行政與司法決策的合法性以及人民信賴法律秩序的基礎。核心問題在於：當 AI 系統具備高度自動化與資料驅動的決策能力時，如何確保法律適用過程仍符合法治原則的可預見性、責任性與正當性？本研究即立基於此一結構性衝擊，探討生成式 AI 在法律訓練與生成階段可能衝擊傳統法治機制之處，並建構相應的法律與政策應對框架。

二、文獻回顧

針對人工智慧基本法、歐盟風險分類、金融業運用人工智慧指引、臺北市政府使用人工智慧作業指引、數位發展部公部門人工智慧應用參考手冊、行政院及所屬機關(構)使用生成式 AI 參考指引總說明及規定、ISO42001 國際標準、本所 ISMS 程序書回顧其內容。

(一)核心治理原則：可信任與負責任的AI

綜合各項規範，發展「可信任」且「負責任」的 AI 是共同的核心目標。我國《人工智慧基本法》與行政院指引均強調了七大基本原則：永續發展與福祉、人類自主、隱私保護與資料治理、資安與安全、透明與可解釋、公平與不歧視、以及問責。金融監督管理委員會（金管會）亦針對金融業提出了類似的六大核心原則，特別強調「建立治理及問責機制」與「重視公平性及以人為本」的價值。ISO42001 國際標準則為「人工智慧管理系統」（AIMS）提供了系統化的管理方法，要求組織從全景分析、領導作為到風險處理，確保 AI 的應用符合倫理與法規。

(二)風險基礎的管理機制（Risk-based Approach）

數位發展部公部門人工智慧應用參考手冊及臺北市政府使用人工智慧作業指引主張採取以風險為基礎的管理模式。歐盟法案與我國規範均將 AI 風險進行分類（如不可接受、高、中、低風險），並依風險程度採行對應的管制措施。

絕對禁止事項：包含未經人工檢視直接做成行政處分、侵害人民生命財產、或在非安全性環境處理機敏資料。

高度風險情境：涉及民眾權利義務、個人隱私或生物特徵辨識的應用，需嚴格監控與評估。

低風險情境：如機關內部行政作業、公開資料摘要等，可輔助使用但仍需人工核對。

(三) AI系統生命週期管理與實務引導

AI 的運用並非單一事件，而是涵蓋系統規劃、資料處理、模型建立至部署監控的完整生命週期。數位發展部指出，AI 專案與傳統資訊專案的差異在於其數據驅動的本質與效能的不確定性，因此建議採用「敏捷開發」與「小規模概念驗證（POC）」。在實務技術面，檢索增強生成（RAG）被視為解決生成式 AI 幻覺問題的重要技術手段，透過結合外部可靠知識庫提升精準度。此外，資料治理是 AI 成功的基礎，必須確保資料的完整性、準確性、真實性與即時性。

(四)人為參與與監督（Human-in-the-Loop）

AI 不得取代人類的自主思維與最終判斷。在數位發展部與金管會文件提及，決策中必須維持「人在指揮」（Human-in-command）或「人在迴圈內」（Human-in-the-loop）的機制，確保關鍵行為由人員檢核，並對 AI 產

出的結果承擔最終責任。對於生成式 AI 的產出，承辦人應秉持「不可完全信任」的態度，並在對外揭露時適當標註使用 AI 之情況。

(五)本所生成式AI管理程序書

綜合前述規範可知，AI 治理的關鍵不僅在於原則宣示，更在於制度化落實。為確保人為監督與風險控管機制能具體執行，本所依循相關指引精神，訂定生成式 AI 管理程序書，作為內部管理依據。本所生成式 AI 管理程序書核心目標在於確保技術應用符合國家安全、資安、隱私保護及法律規範，並有效地降低運作風險。該程序書適用於本所所有相關作業，要求全體同仁必須遵守其規範流程。

在「導入與評估機制」方面，程序書要求使用部門在研議導入技術前，必須先評估其必要性，例如是否能提升大量重複性任務的效率，或是有加速內容創建及提供創意靈感的需求。在正式運用前，必須填寫「生成式 AI 導入評估表」並經核准後方可執行，且因應 AI 技術發展迅速的特性，組織應定期審查評估結果以確保安全無虞。

針對「核心使用原則」，該程序書強調「輔助性原則」，即生成式 AI 僅作為輔助工具，不能取代承辦人的自主思維或人際互動，亦不得作為行政行為或公務決策的唯一依據，承辦人必須對內容進行專業的最終判斷。此外，程序書訂有嚴格的「限制性原則」，嚴禁將機敏資訊、營業秘密及個人隱私輸入 AI 系統；若需使用封閉式地端部署模型，則必須在確認系統環境安全後，依照資訊機密等級分級使用。

在「資通安全管理與風險控制」上，本所建立了一套完整的防護措施。在存取控制方面，僅授權特定人員使用並強制啟用「雙因素驗證（2FA）」，且須進行全程日誌記錄與異常行為監控。同時，程序書要求必須對 AI 產出的資訊進行審查，避免虛假內容，並定期評估涉及智財權、人權及隱私的風險。對於外部採購或合作事項，亦要求得標廠商須共同遵守此管理程序。

最後，為了達成「持續改善與教育訓練」，程序書規定本所應定期舉辦相關培訓，以提升同仁的風險管理能力。在追蹤機制上，使用部門至少每季需檢視導入評估表，且每半年必須填報「生成式 AI 使用檢核表」，以確保各項運用持續符合最新的法規與資安標準。

三、研究核心課題

綜合前述研究背景可知，生成式人工智慧所引發的法律爭議，並非

僅止於單一著作權侵權或技術風險，而係橫跨資料使用、內容生成、決策責任與法治原則之多層次問題。現行制度雖已嘗試於訓練資料使用、透明性、偏誤管理及責任分配等面向提出治理工具，然在人工智慧逐步介入甚至影響法律適用與行政決策之情境下，仍存在難以迴避的結構性疑問。

首先，在上線前（Ex Ante）階段，生成式人工智慧於模型訓練過程中大量使用既有著作與資料，已引發合理使用、非享受使用及文本與資料探勘等法律界線之爭議。儘管比較法上逐漸形成有限容許之趨勢，但此類合法性評價多集中於「是否允許使用資料」，而未必能回應資料使用結果對後續法律適用與責任結構所造成的影響。

其次，在上線後（Ex Post）階段，生成式人工智慧所產出的摘要、改寫、續集或其他高度結構化內容，已顯示出新型侵權樣態與責任缺口。由於生成內容並非單純重製原著作，卻可能在實質上重現其核心結構，使傳統以相似性與依賴性為核心的侵權判斷面臨舉證與責任歸屬上的困難。同時，既有過失責任與因果關係理論，亦難以充分回應高度自動化系統所造成的風險外溢問題。

更為關鍵者，在公部門使用人工智慧的情境下，問題已不僅指侵權或風險管理，而直接涉及行政行為之合法性、正當法律程序及法治原則的維持。依現行政策指引與治理框架，雖強調透明性、可解釋性與風險控管，然此類技術與程序性措施，仍無法完全回答以下核心疑問：當人工智慧實質參與法律適用或影響個案決定時，法律責任是否仍可明確歸屬於人類決策者？法治國所要求的個案判斷與規範演進能力，是否因此受到侵蝕？

本研究所欲處理之核心問題在於：

（一）生成式人工智慧在訓練與生成階段，如何對既有侵權判斷與責任體系造成結構性挑戰？

（二）現行以透明性、偏誤管理及風險控管為核心的治理工具，是否足以確保公部門使用人工智慧仍符合正當法律程序與法治原則？

（三）在人工智慧尚非法律主體之前提下，是否有必要以「人為介入」作為公部門正當使用人工智慧的規範性底線，以確保法律適用仍由人類承擔最終責任？

透過上述問題意識之釐清，本文將進一步檢視生成式人工智慧對法治國原則所形成的結構性衝擊，並嘗試建構一套以人為介入為核心、兼顧技術發展與法治要求之制度理解框架。

四、上線前階段（Ex Ante）：生成式 AI 訓練資料之法律界線

(一)訓練資料來源與使用合法性

（1）從訓練行為看著作權爭議：重製與侵權的兩難

在模型訓練的初始階段，通常需蒐集並分析大量既有文本、圖像或其他資料，以建立其語言理解與內容生成能力。此一訓練過程往往涉及對既有作品之重製與資料化處理，是否因此構成著作權侵權，乃生成式 AI 法律爭議中最早浮現，且對整體制度設計影響深遠之核心問題。從比較法觀察，各國法制並未就訓練資料之使用採取全面禁止或全面容許之立場，而係嘗試透過既有著作權理論加以調整與適用，逐步形塑出具有條件限制之合法性評價框架。

（2）比較法觀點：合理使用與立場對比

在美國法制下，上線前訓練資料之使用，主要透過合理使用（fair use）理論加以評價。其中，Google 圖書館（Google Books）計畫常被視為支持大規模資料處理合法性的代表性案例。該案中，Google 將大量書籍加以數位化並建立全文搜尋索引，雖涉及對作品之完整重製，惟法院認為其使用目的在於提供搜尋與資訊定位功能，而非供讀者取代原書閱讀，且其顯示內容僅限於高度受限之片段，進一步降低對原著作市場之替代可能性，未實質侵害原著作市場，因此認定具有高度轉化性（transformative use），構成合理使用。此一判斷模式，長期以來被視為資料分析與資訊技術發展得以在著作權法下取得空間的重要依據。

然而，近年生成式 AI 技術之發展，已使訓練資料使用的法律評價面臨新的挑戰。以 OpenAI 使用《紐約時報》新聞內容進行模型訓練之訴訟為例，原告即主張，該等訓練行為不僅涉及對新聞作品之大規模重製，且生成式 AI 系統後續所產生之內容，可能在資訊功能與市場效果上與原作品高度重疊，進而對新聞訂閱、授權與既有內容市場造成替代與侵蝕。與 Google Books 僅提供搜尋索引之情形不同，該案所涉及者，係以訓練成果支撐可直接生成完整敘述內容之商業化系統，其是否仍具備足以支撐合理使用之轉化性，成為爭議核心。

惟截至目前為止，前述 OpenAI 與《紐約時報》之訴訟仍處於審理階段，法院尚未就生成式 AI 訓練行為是否構成合理使用作出最終實體判決。整體而言，此一對照顯示，美國法制雖透過 Google Books 案確立資料導向使用之合法可能性，但在生成式 AI 得以直接產出具市場替代性內容之情

境下，既有合理使用框架是否仍可無條件延伸適用，尚未獲得明確答案，亦未形成明確之安全港規則。

相較之下，日本著作權法則透過明文規定建立訓練階段的合法性基礎。日本法引入「非以享受著作中表達之思想或情感為目的」之例外，允許在必要範圍內，以任何方式利用著作進行資料分析，包括為人工智慧訓練而對大量著作進行蒐集、比較、分類與分析。日本文化廳亦進一步說明，生成式 AI 開發商於模型訓練階段，原則上不以享受原著作內容為目的，因而可落入該例外範圍。同時亦指出，若訓練過程中出現為回應特定使用者需求而即時檢索原著作內容，或於微調與補充資料階段實質重現著作內容，則可能脫離「非享受」之範圍，顯示日本雖透過明文化例外事前開放訓練行為，仍藉由對利用態樣之實質審查，防止訓練行為轉化為變相之內容再利用。

歐盟透過立法方式回應生成式 AI 訓練對著作權體系所造成之結構性衝擊。具體而言，歐盟於《著作權單一市場指令》中引入文本與資料探勘（text and data mining, TDM）例外，分別就研究用途及一般用途之資料探勘行為設計不同規範架構，原則上允許對合法取得之著作進行分析，同時保留權利人以明確方式排除其作品被用於特定探勘行為之可能。此一制度安排顯示，生成式 AI 訓練並未被歐盟法一概排除於著作權規範之外，而係在既有權利體系內透過條件化例外加以調整。

（3）比較法綜合觀察：條件化容許與後續保留

綜合比較觀察可知，各國對生成式 AI 訓練資料之使用，雖採取不同法制技術，但呈現出若干共通趨勢：其一，訓練階段並未被一概視為侵權，而係在特定條件下予以容許；其二，合法性評價多圍繞於使用行為之性質與目的、其是否具有功能或效果上的轉化，以及是否對既有著作市場造成不當影響；其三，即便允許訓練資料使用，仍保留對後續生成行為及具體運作模式加以限制之空間。此亦顯示，訓練資料合法性問題並非終點，而是生成式 AI 法律治理之起點。

(二)透明性與偏誤作為前置治理工具

（1）透明性之制度目的：可檢驗性與可課責性

面對生成式人工智慧於訓練階段所可能引發的侵權與制度風險，近年各國法制與政策文件多將「透明性」與「偏誤管理」視為重要之前置治理工具。透過要求揭露訓練資料來源、模型設計假設與使用限制，並建立偏誤檢測與修正機制，期能降低人工智慧系統對權利與法治價值之潛在侵害。

然而，此類治理工具之功能與界限，仍有必要加以進一步釐清。

（2）透明性之界限：透明不等於可解釋

就透明性而言，其在生成式人工智慧治理中的主要目的，並非在於使所有利害關係人完全理解模型運作細節，而係確保系統之運作具備可檢驗性與可課責性。換言之，透明性作為一項制度要求，其核心在於提供事後審查、責任追究與制度修正之可能性，而非追求全面可理解的技術說明。此一理解亦反映於現行公部門人工智慧相關指引中(如：臺北市政府使用人工智慧作業指引、金管會金融業運用人工智慧(AI)指引、數位發展部公部門人工智慧應用參考手冊等)，透過要求紀錄訓練資料、模型版本、使用情境與限制條件，使人類使用者與監督者得以在必要時進行評估與介入。

（3）模型解釋技術之規範限度

惟「透明性」作為一項治理要求，若未進一步說明其應達成之程度與內容，容易在制度設計上產生期待落差，甚至誤將透明性等同於對人工智慧系統之全面揭露。然而，若將透明性理解為「只要揭露模型或資料，即可解決法治疑慮」，則可能高估其實際效用。必須強調的是，「看得到演算法」不代表「看得懂決策原因」。AI 的模型結構極其複雜，即便揭露了程式碼，也很難要求承辦人將其邏輯直接轉化為符合正當程序要求、可供法律評價之理由結構（reason-giving structure）。更進一步而言，透明性的限制並非人工智慧所獨有，而是決策體系本身即存在的結構性問題。即便在人類決策情境中，個人對其判斷形成過程亦未必能完整說明，其決定往往受限於經驗、直覺與制度背景，並非全然透明。

（4）透明性之層次化理解

在此脈絡下，近年實務上常透過「以另一模型解釋模型」之方式，試圖提升生成式人工智慧之可理解性，例如以簡化模型或後設模型（surrogate model）來輔助說明原 AI 模型的輸出結果。然而，此類作法所提供者，實際上仍僅為一種經技術轉換後之「理由近似物」，而非對原始決策過程之完整再現。此類解釋機制，固然有助於將模型輸出轉化為可供人類判斷之理由，從而回應實務理性層次之透明性要求，惟其並不當然等同於技術或因果層次之透明，亦不足以單獨承擔法律上對於正當性與責任歸屬之評價功能。若未清楚區分此一差異，反而可能使制度過度依賴形式上之可解釋性，而忽略其規範意義上的不足。

基於上述限制，有必要對生成式人工智慧治理中之透明性要求加以層次化理解，以避免透明性淪為空泛口號。綜合相關政策文件與學理討論，人工智慧之透明性要求，至少可區分為三個不同層次。

第一層，形式與程序層次的透明。

此一層次之透明，著重於是否揭露人工智慧之使用事實、適用範圍與基本功能定位，使利害關係人得以知悉其是否受到人工智慧系統之影響。此類透明要求主要服務於告知義務與程序正義之需求，亦為現行公部門人工智慧指引中最常見之規範形式。

第二層，實務理性（practical reason）層次的透明。

指的是決策過程是否能被一般人理解與接受，此一層次之透明，並不要求揭露人工智慧之完整技術細節，而係要求系統提供足以支撐實務判斷之理由，使人類決策者得以評估該建議是否適用於特定個案，並於必要時選擇採納或拒絕該建議。其重點在於確保人工智慧之輸出能轉化為可供人類判斷之理由結構，而非僅止於計算結果。

第三層，技術與模型層次的透明。

此一層次涉及模型結構、訓練資料細節與演算法設計等技術資訊，對於系統開發與特定高風險場景固然具有重要性，惟其內容高度專業化，未必能直接回應法律適用與個案判斷之需求，亦不宜作為一般治理情境下之最低要求。

（5）偏誤作為制度性現象

與透明性並列，偏誤管理亦被視為生成式人工智慧於訓練階段的重要治理工具。由於人工智慧系統透過既有資料進行學習，其訓練資料所反映之社會結構、歷史分布與制度選擇，極可能在模型輸出中被再製甚至放大。是以，近年政策文件與治理指引普遍要求於系統設計與運作過程中，進行偏誤辨識、測試與修正，以降低對特定群體或權利之不當影響。

然而，偏誤問題之本質，並非僅屬於人工智慧系統之技術缺陷，而是決策體系本身長期存在之現象。人類決策同樣深受認知捷思（heuristics）與社會慣性（social inertia）等因素所影響，而其判斷過程中所採取之評估標準與取捨方向，亦往往隱含特定價值取向與前設判斷，未必呈現為中立或一致之結果。若僅將偏誤視為人工智慧所獨有之問題，反而可能遮蔽人類決策長期累積之制度性偏差。從此角度觀之，人工智慧所顯現的偏誤，某種程度上係將既有社會與制度選擇「可視化」與「具體化」，使其更容易被檢驗與討論。

（6）效率與平等之價值張力

尤須注意者，在訓練階段進行偏誤修正時，實務上往往面臨效率與平等之結構性張力。為避免對特定性別或族群產生歧視性影響，系統設計者可能選擇移除或弱化相關變項，惟此舉亦可能導致模型對某些風險或結果

之預測準確性下降。此一現象顯示，偏誤治理並非單純之技術優化問題，而係涉及不同價值目標間之取捨。相關取捨本身，已屬規範判斷之範疇，難以由模型自行完成，亦無法僅憑技術指標加以正當化。

基於此，生成式人工智慧治理中對偏誤的要求，並非追求完全消除偏誤，而在於是否能夠被辨識、被評估，並在必要時加以修正。偏誤治理的重點，應置於制度是否具備持續監測、回饋與調整的能力，而非僅止於系統上線前的一次性檢測。此亦反映於公部門人工智慧相關指引中，例如要求機關於系統建置與運作過程中，紀錄訓練資料之來源與範圍、模型適用之使用假設，以及系統功能與決策權限之限制條件，使人工智慧所可能產生之偏誤，得以於事後被追蹤、檢驗並回溯其形成原因。

然而，與透明性相同，偏誤治理亦存在其制度界限。即便透過技術方法降低特定偏誤，人工智慧系統仍係依賴既有資料與模型假設進行推論，難以自行反思其適用結果是否符合特定個案之規範意義。換言之，偏誤管理固然有助於降低系統性風險，卻無法取代人類對於個案差異、價值衡量與責任歸屬之判斷。

（7）前置治理工具之制度定位與其界線

綜合而言，偏誤管理與透明性同屬生成式人工智慧治理之前置工具，其功能在於提升制度之可檢驗性與修正能力，而非確保人工智慧輸出本身即為正確或正當。此類前置治理工具，係作為後續責任評價與人為介入設計之前提，而非取代其法治功能。當生成式人工智慧系統完成訓練並實際投入運作後，其所產生之內容即可能直接影響權利義務關係之形成。此時，僅以前置之技術治理措施作為規範基礎，仍不足以回應法治國對於個案判斷、侵權風險控管與責任歸屬之要求，亦有必要進一步檢視生成內容於既有法律體系下之定位與法律效果。

五、上線後階段（Ex Post）：生成內容的法律性質與侵權風險

(一)生成內容是否構成侵權

（1）生成內容侵權問題之提出

相較於訓練階段圍繞資料使用合法性所展開之討論，生成式人工智慧於上線後所產出之內容，涉及多層次之著作權法問題，其中是否構成著作權侵權，尤牽涉直接之法律效果與權利衝突。此一問題的困難在於，生成內容多非逐字重製既有著作，而係透過改寫、摘要、延伸或重組方式呈現，

使傳統侵權判斷中以「重製」為核心之分析，在生成式人工智慧情境下呈現適用上的張力。

從比較法觀察，各國法制並未僅因生成內容由人工智慧產出，即當然排除其於特定情境下侵害他人著作權之可能性。相反地，多數實務仍回歸既有侵權判斷架構，檢視生成內容是否與既有著作構成實質近似(substantial similarity)，以及是否具備可將生成結果與特定既有著作之利用行為連結之判斷基礎。惟在生成式人工智慧之運作模式下，由於系統係基於大量既有資料進行學習，原告往往難以具體證明特定著作是否曾被用於訓練，或生成結果是否可追溯至單一作品之利用行為，致使侵權舉證在實務上面臨實質困難。

（2）著作權保護資格與侵權判斷之區分

我們必須將 AI 產出內容「是否受著作權相關規定保障」與「是否侵害他人」分別處理。即便 AI 生成的圖像因為缺乏人類原創性而無法取得著作權，這並不代表利用該圖像就不會侵害到現有的著作權市場。

（3）著作權的核心衝突：人類創作控制權的界定標準

美國法制下，美國著作權局（U.S.Copyright Office）於 Zarya of the Dawn 登記案件中，主要即係就前者問題表明立場，該等行政立場所處理者，僅為生成內容之著作權保護資格與登記範圍，並非直接就侵權成立與否作成判斷。美國著作權局於 Zarya of the Dawn 著作權登記案件中明確表示，著作權之核心保護標的為「人類作者之原創性表達」，而非純由人工智慧系統自動生成之內容。該局在 2023 年 2 月 21 日的官方信函中指出，作品中由 Midjourney 生成的圖像部分不屬於人類創作產品，因此不符合著作權法對於「原創性」與「人類著作」之要件，不能作為受保護之著作登記範圍；相對地，作品之文字部分及其對圖像的選擇、統整與編排，因為包含人類之創作判斷，仍屬受著作權保護的範圍。

與此相關，著作權局及其後續行政審查委員會（U.S. Copyright ReviewBoard）對於 A Recent Entrance to Paradise 類似申請也維持一致立場，即若著作之生成過程中缺乏人類對最終表達之實質創作控制（ultimate creative control），該生成物不具著作權保護資格。行政審查委員會在 2023 年的相關決定中屢次重申，著作權法要求著作首創性不可由機器獨自完成，而必須有足夠之人類精神創作參與，否則該作品不符合現行法上「作者」之定義。

上述立場亦反映於美國著作權局後續相關指引之中，即當申請人使用含有人工智慧生成素材之作品申請著作權登記時，登記申請將衡量使用者

對作品之最終表達所具有之實際創作控制程度；若人類之參與不足以產生原創性，該人工智慧生成素材本身將被排除於保護範圍之外。

（4）侵權認定的實務挑戰：如何面對「黑箱」下的舉證困境

日本法制則採取更為明確的整理方式。依日本文化廳對人工智慧與著作權之官方說明，生成式人工智慧於生成與利用階段，是否構成侵權，仍以傳統侵權判斷架構為準，亦即須檢視生成內容是否與既有著作具備「類似性」與「依據關係」。若二者同時成立，且不符合權利限制規定或未取得授權，則生成或後續利用行為即可能構成侵權；反之，若不具類似性或依據關係，則不構成侵權。

文化廳並進一步指出，即便無法確認特定既有著作是否實際被納入人工智慧之訓練資料，於生成結果與既有著作呈現高度相似，或使用者對該既有著作具有接觸可能性之情形下，依據關係仍可能被推定成立，以避免在生成式人工智慧高度自動化與資料不透明之情境中，因舉證困難而使權利保護落空。

日本文化廳亦特別強調，即便生成行為本身未必違反訓練或生成階段之相關規定，後續對生成內容之公開、散布或商業利用，仍須獨立檢視是否構成侵權。此一整理顯示，日本法制刻意將「訓練階段之容許」與「生成結果及其利用之合法性」加以切割，避免訓練例外或生成行為之合法性，被誤解為對生成內容後續利用行為之全面豁免。

（6）生成內容侵權判斷之具體展現。

在上述比較法架構下，生成式人工智慧侵權風險的關鍵，最終仍須回到法院如何於具體個案中，面對高度自動化與舉證困難的情境，實際操作「實質近似」與責任評價。中國大陸近期司法實務，即提供了生成內容侵權判斷之具體示範。

以廣州互聯網法院（2024）粵 0192 民初 113 號判決為例，該案係由網站廠商依使用者提示詞生成「迪迦奧特曼」相關圖像。法院於判斷時，並未著重於人工智慧是否曾接觸特定訓練資料，而係在舉證困難的前提下，直接檢視生成圖像在角色形象特徵、整體視覺表達及可識別性上，是否高度再現既有作品之創作性表達。法院據此認定，該生成圖像已構成對權利人作品權利之侵害，並判命被告承擔停止侵害及損害賠償責任。

當前各國司法實務（如上述之美、日、中案例）均反映出一項共同趨勢：法院並未因生成過程涉及人工智慧，而降低對「實質近似」的審查密度。綜合觀察可發現，儘管各國對創作保護的給予寬嚴不一，但「人為實質控制」始終是判定侵權與否的核心基準。此一發展不僅凸顯侵權風險已

從理論走入實務，更為本所未來利用 AI 生成圖表或輔助報告時，提供了一個明確的避險方向，即必須強化人為的最後審查與編輯。

(二)責任缺口與既有責任法制的侷限

（1）責任斷裂現象之出現

綜合前述比較法與實務發展可知，生成式人工智慧於上線後所產出之內容，雖可回歸既有侵權判斷架構進行評價，然在高度自動化與多方參與之運作模式下，既有責任法制逐漸浮現結構性侷限。此一侷限並非源自侵權要件本身，而係來自生成式人工智慧系統中「創作、生成與利用行為分離」所導致之責任歸屬上之斷裂現象。

（2）使用者責任之不確定性

首先，就使用者而言，其多僅透過提示詞輸入、參數選擇或生成結果取捨參與生成過程。此類行為是否已達可歸責之侵權程度，須視具體操作內容而定。若使用者明確以既有作品、角色或高度識別性元素作為生成目標，則其行為較容易被評價為具有侵權風險；反之，若僅進行一般性描述或抽象提示，則使用者對最終生成結果之控制程度有限，難以一概歸責。此種介於「創作參與」與「結果不可控」之間的狀態，使使用者責任呈現高度不確定性。

（3）平台注意義務之界限

其次，就平台或系統提供者而言，其雖設計、提供並控制生成式人工智慧之運作環境，惟其並非針對特定內容進行生成決定。既有責任法制中，平台責任多以「是否知情」、「是否可控制」及「是否即時移除」等標準進行評價，然在生成式人工智慧情境下，生成內容係即時產生，且其侵權風險往往須經事後比對方能辨識，使平台是否負有合理注意義務、以及注意義務範圍為何，成為難以明確界定之問題。

（4）非法律主體所造成之責任缺口

再者，生成式人工智慧本身並不具備法律主體地位，亦無法成為權利義務之歸屬對象。當生成結果同時包含系統訓練結果、使用者操作選擇與平台設計因素時，侵權責任難以單一歸屬，形成既有責任法制難以完全涵蓋之「責任缺口」。由於 AI 不是法律主體，不能成為賠償對象，一旦發生侵權爭議，最終的責任歸屬往往會回到輸入提示詞的使用者，或是疏於審查的平台端。這也說明了為什麼第七段結論與建議，在公部門應用中，人為介入不只是流程要求，更是防範法律風險的最後防線。

（5）中國實務之另一取徑：人類介入程度作為評價標準

中國大陸實務對此一問題，已呈現出不同切入角度。除前述廣州互聯網法院就生成結果侵權所作之判斷外，北京互聯網法院於（2023）京 0491 民初 11279 號判決中，則自生成內容之法律性質切入，強調使用者於生成過程中所體現之審美選擇與創作判斷。該案肯認，若使用者透過提示詞設計、參數調整及生成結果取捨，實質體現其個人創作判斷，則其對生成內容得成立著作權。該判決雖非直接處理侵權成立與否，亦不當然意味侵權責任即全數歸屬於使用者，然其所揭示之思維顯示，法院已意識到在生成式人工智慧運作架構下，必須透過「人類介入程度」作為重新配置法律評價與責任歸屬之關鍵。

（6）既有責任法制之制度極限

然而，即便透過區分使用者、平台與生成內容之法律性質，既有責任法制仍難以完全回應生成式人工智慧在法律適用層面所帶來的挑戰。特別是在公部門或高風險領域中，生成式人工智慧可能影響權利義務形成或法律效果判斷時，僅依賴事後侵權責任或損害賠償機制，無法充分確保法治國所要求之可預見性、可課責性與正當性。基於此，責任缺口的問題並非單純透過擴張既有侵權責任即可解決，而有必要進一步檢討制度設計，特別是透過人為介入機制，確保在關鍵決策或法律適用階段，仍由人類承擔最終判斷與責任，作為銜接生成式人工智慧與法治國原則之核心制度安排。上述責任斷裂與法制侷限，在公部門導入人工智慧於行政決策流程之情境下，將進一步轉化為正當法律程序是否仍能實質運作之法治風險。

六、公部門使用人工智慧之特殊法治風險

(一)行政決策自動化與正當法律程序

（1）行政決策自動化之程序風險

公部門導入人工智慧或演算法工具於行政決策流程中，最核心的法治風險之一，即在於是否仍能維持正當法律程序之實質功能。當決策過程高度自動化，甚至部分委由外部系統或私人機構執行時，既有程序結構即可能受到實質衝擊。

（2）案例觀察：形式決策與實質判斷之分離

愛爾蘭交通執法實務中由 GoSafe 公司操作自動測速設備之案例，即顯示行政決策自動化結合委外機制所引發之程序疑慮。該制度下，違規判斷高度仰賴技術設備與受委外公司之操作，受處分人往往難以理解測速與違規判斷之原理；加以罰單送達程序出現爭議，使當事人實際行使防禦與

救濟權利之可能性受到影響。此一案例顯示，縱使行政機關形式上仍為處分作成者，若實質判斷已高度依賴自動化系統，正當法律程序即可能淪為形式要求。

類似問題亦可見於美國賓州阿勒格尼郡所使用之家庭篩檢工具（Allegheny County Family Screening Tool, AFST）。該工具透過資料模型輔助社工評估兒童受虐或疏於照顧之風險，並未直接取代人類決定。然而，相關研究與實務討論指出，模型運作邏輯與權重配置不易向當事人說明，且其使用之歷史資料可能放大既有社會與制度性偏誤。即便保留人類決策，當行政實務高度依賴系統分數時，是否仍能確保個案差異被充分評估，成為正當法律程序上的疑問。

而紐西蘭的抉擇，則展現了不同的價值權衡：在評估類似預測性工具後，最終選擇停止實施。其考量並非技術效能不足，而係基於對歷史上不當拆散原住民族家庭之反省，認為該類系統可能對家庭權、程序正義與信賴關係造成過度衝擊。此一選擇顯示，行政決策自動化是否適合導入，並非單純效率問題，而涉及對正當法律程序與基本權保障之價值判斷。

（3）正當法律程序之結構性疑問

綜合上述案例可知，行政決策自動化所帶來的法治風險，並不取決於是否使用人工智慧或生成式技術，而在於決策過程是否仍保有可理解、可質疑、可課責的人類判斷結構。這類國際案例的啟示在於：自動化工具不能成為行政機關規避說理義務的藉口。如果行政處分的實質核心是由演算法決定，而承辦人無法解釋其邏輯，該處分在法治原則下將面臨極大的違法挑戰。

(二)責任歸屬模糊化與補償導向治理的興起

（1）責任歸屬模糊化之制度背景

隨著人工智慧與高度複雜科技系統的導入，行為與結果之間的因果關係日益模糊，既有以過失為核心的責任法制，逐漸顯現適用上的侷限。現代科技風險多源自多方協作與系統性運作，傷害可能在時間與空間上與設計、開發行為高度分離，使受害者難以指認具體可歸責之行為人。

（2）補償導向治理之興起

面對這種因果關係模糊的困境，若仍以傳統過失責任作為主要救濟途徑，將使受害者承擔過高之舉證負擔，實質救濟功能反而弱化。學理與實務因此逐漸轉向以補償為核心的治理思維，透過無過失補償機制（no-fault compensation schemes, NFCS），在不以證明過失為前提下，確保受害者能

即時獲得救濟。

此類補償導向機制已為多國（如以色列、紐西蘭、瑞典等）於醫療、交通或其他高風險科技領域之實務所採行，而非僅止於理論構想。其制度設計重點不在於放棄責任追究，而係改變責任實現之順序：由制度或保險機制先行補償受害者，再透過內部追償或風險分攤，使系統提供者承擔相應成本，以維持安全誘因。

前述補償導向治理，係針對高度不確定科技風險下，既有責任法制救濟功能不足所提出之制度回應。透過無過失補償機制，得以在責任歸屬難以釐清的情況下，確保受害者於損害發生後獲得即時救濟；惟此類機制所處理者，僅止於「結果如何回應」，而非行政決策形成過程本身之正當性。

（3）補償與人為介入之制度分工

因此，補償導向治理與人為介入並非替代關係，而係分別作用於不同制度層次之互補安排。前者處理責任法制在高度不確定風險下的救濟極限，後者則確保行政決策在導入人工智慧時，仍維持以人類作為最終判斷與責任承擔主體。雖然建立無過失補償機制可以降低受害者的舉證門檻，但為了維持公務運行的正當性，仍須透過「人為介入」確保每一筆決策都落在法律授權的軌道內。

七、結論與建議

(一)結論

生成式人工智慧的導入，並未使既有法律問題自然消解，而是在既有著作權法與行政法架構下，凸顯並放大若干原本即已存在的制度張力。本文透過區分生成式人工智慧之模型訓練階段與實際上線運作階段，並結合比較法與相關實務觀察，說明各國法制多係在既有理論基礎上調整其適用方式，而非建立一套全然脫離現行法體系之新制度。訓練階段之合法性評價，並不當然延伸為生成結果之合法性，生成內容是否構成侵權，仍須回歸既有侵權判斷架構加以檢視。

在治理層面，近年政策文件所強調之透明性與偏誤管理，確實有助於提升制度之可檢驗性與後續修正能力，惟其性質仍屬前置與輔助性工具，難以單獨承擔法治國所要求之正當性與責任評價功能。當生成式人工智慧實際投入運作，並可能影響權利義務關係之形成時，僅依賴技術治理措施或事後補償機制，於個案判斷、侵權風險控管及行政責任歸屬等面向，仍存在一定限制。

基於前述分析，本文認為，在公部門使用人工智慧之情境中，有必要進一步重視人為介入在制度設計中的角色。人為介入並非否定人工智慧於資訊處理或決策輔助上的功能，而是在影響人民權利義務之行政決策過程中，維持可歸責之人類判斷、可理解與可受審查之決策結構。透過此一方式，或可在導入人工智慧所帶來之治理效益與行政效率的同時，降低對正當法律程序與行政責任體系可能造成之衝擊。

整體而言，生成式人工智慧是否適合導入，並非單純的技術或效率問題，而涉及制度如何界定其使用範圍與作用方式。未來相關治理與法制發展，仍有賴在實務運作中持續觀察與調整，以在創新應用與法治要求之間，維持相對穩定的制度平衡。

(二)建議：人為介入作為公部門正當使用人工智慧之必要條件

在公部門導入人工智慧之治理架構中，人為介入並非僅屬技術操作上的選項，而應被理解為確保行政權行使符合法治國原則之制度核心。其意義並不在於否定人工智慧之效能，而在於使其運作能夠嵌入既有憲法秩序之責任結構與程序保障之中，成為強化行政品質之工具，而非取代人類判斷之主體。

首先，行政決策之正當性，根源於可歸責之人類判斷。依據法治國原則，行政機關對於影響人民權利義務之決定，必須存在可識別之責任主體。人工智慧系統固可提供高度精準之分析與預測，但其本身並非法律主體，無法承擔規範評價與責任歸屬功能。透過設計實質人為介入機制，使最終判斷仍由具備法律責任能力之行政人員作成，方能維持決策責任之人格化結構，並確保行政責任與司法審查制度得以有效運作。

其次，正當法律程序要求決策過程具備可理解性與可受審查性。人工智慧系統之輸出結果，往往建構於統計推論與模型權重之上，其理由形式未必直接符合規範論證之要求。人為介入之功能，即在於將技術輸出轉化為具法律意義之行政理由，使人民得以理解決策基礎並行使防禦權，亦使司法機關得以進行實質審查。此種轉譯與說理功能，乃人工智慧融入法治架構之必要橋梁。

再者，行政裁量與法律續造本質上涉及價值衡量與個案差異判斷。人工智慧雖可提供資料導向之建議，然其運作仍受既有資料與模型設定所限。透過人為介入機制，行政機關得以在系統建議之外，考量具體情境與規範目的，維持裁量彈性與規範發展空間，使人工智慧成為輔助決策之專業工

具，而非機械化適用規範之自動裝置。

最後，人為介入機制亦與補償導向治理形成制度上的互補關係。無過失補償制度固可於損害發生後提供救濟保障，然其處理的是結果層次之回應；人為介入則屬於決策形成階段之正當性保障。兩者結合，方能構成一套兼顧事前正當性與事後救濟之完整治理架構。如此設計，不僅能回應科技發展帶來之挑戰，更能在法治國原則之基礎上，建立公眾對行政運用人工智慧之制度信任。

總而言之，人工智慧之導入，不應被視為行政權行使方式之斷裂，而應透過制度設計，使其成為提升行政效率與決策品質之工具。在此過程中，人為介入並非對技術之不信任，而是對憲法秩序之承諾；其功能在於確保科技發展與法治原則得以協調並進，使公部門在創新與正當性之間取得動態平衡。

參考文獻

1. 美聯社：https://apnews.com/article/uk-courts-fake-ai-cases-46013a78d78dc869bdfd6b42579411cb
2. 立法院：https://www.ly.gov.tw/Pages/Detail.aspx?nodeid=46868&pid=255581
3. 台一國際智慧財產事務所：https://www.taie.com.tw/cloud-magazine.php?act=view&no=135
4. 中央研究院法律學研究所臺灣人工智慧行動網：https://ai.iias.sinica.edu.tw/nyt-suit-against-openai-and-microsoft/
5. 北美智權報： https://naipnews.naipo.com/34640/
6. 日本文化廳：https://www.bunka.go.jp/english/policy/copyright/pdf/94055801_01.pdf
7. 中央研究院法律學研究所資訊法中心：https://infolaw.iias.sinica.edu.tw/?p=672
8. AI・抄襲・智財權 / 楊智傑
9. 人工智慧與法學變遷 / 李建良
10. 人工智慧的法制基礎 破壞式創新的建構式法制/ 李建良等
11. 人工智慧來了！身為公民您該知道什麼？/ John Danaher等
12. 聯邦國際專利商標事務所：https://www.tsailee.com/News/Details?lc=en&News_id=1457
13. 資策會科技法律研究所：https://stli.iii.org.tw/article-

detail.aspx?no=64&tp=1&d=9092

14. 維基百科：https://en.wikipedia.org/wiki/Accident_Compensation_Corporation
15. 維基百科：https://en.wikipedia.org/wiki/The_Black_Box_Society
16. Lexology：https://www.lexology.com/library/detail.aspx?g=423ff4d2-69b0-468f-94e3-c39fdcc3cc6e&
17. 美國國家醫學圖書館：https://pmc.ncbi.nlm.nih.gov/articles/PMC2853784/
18. 國家圖書館：https://ndltd.ncl.edu.tw/cgi-bin/gs32/gsweb.cgi/login?o=dnclcdr&s=id=%22109NTU05194055%22.&searchmode=basic
19. 全球生成式AI著作權訴訟攻防戰 / 陳家駿