歐盟人工智慧法解說書

韓國企業實務應對指南

金炅鎭 律師

序幕

2026年8月2日，來自布魯塞爾的電話

2026年8月2日是個星期日。那天下午，我接到來自板橋一家人工智慧公司代表的電話。他的聲音聽起來與平時不同。他說收到了歐洲合作夥伴發來的一封電子郵件，內容指出他們的聊天機器人似乎從今天起違反了歐洲法律。他本以為不可能發生這種事，因為他們在歐洲連個辦公室都沒有，只不過是在服務中附帶了一個諮詢聊天機器人而已。

我問他，聊天機器人在與歐洲使用者對話時，是否會告知對方這是人工智慧。現場頓時陷入了一陣沉默。接著傳來的是反問：為什麼要告知這種事？而那句反問之中，正包含了撰寫本書的所有原因。

那天正是《歐盟人工智慧法案》第50條透明度義務全面施行之日。從那天起，與人類對話的人工智慧必須向使用者告知自身為人工智慧。板橋的那家公司甚至不知道自己屬於該法的適用對象，因為他們堅信自己在歐洲沒有法人，因此與歐洲法律無關。然而，該公司的聊天機器人當時正在與歐洲使用者對話，那一刻，公司的地址已不再重要，重要的是該人工智慧在歐洲正在從事何種行為。

我接到過許多類似的電話。有的是因為生成式人工智慧製作的廣告影片出了問題，有的是因為供應給歐洲銀行的審查系統觸法。這些公司都有一個共同點：全都不知道自己屬於監管對象，而當他們意識到時，相關法規早已施行。

市面上探討《歐盟人工智慧法案》的資料比比皆是。律師事務所的電子報、顧問公司的白皮書、摘要條文的手冊。大多數資料僅僅記錄條文規定了什麼並列出施行日期。即使讀過這些資料，板橋的那位代表恐怕依然不會意識到自己已經觸法。因為那些資料要麼是以歐洲企業為前提撰寫的，要麼就停留在單純整理條文的層面。

本書站在不同的立足點撰寫。我們從一開始就以總部位於首爾的企業為設想對象。相較於條文規定了什麼，本書更側重於執法機關會如何看待該條文。此外，在需要作出判斷之處，我們絕不迴避判斷。我會寫下諸如「我是這樣解讀該條文的」、「此處存在見解分歧」、「這部分目前尚不確定」等文句。哪怕作為參考資料的中立性稍微變薄弱，我也選擇留下這是一本「由人撰寫之書」的印象。

希望本書能發揮兩種作用。其一是作為判斷工具。期望考慮進軍歐洲的企業經營團隊，能花上30分鐘回答出「我們是否適用本法？若適用，屬於哪一個等級？」。其二是作為實務手冊。期望法規遵循負責人能列印出附錄的檢核表放在桌上隨時對照使用。因此，各章節的前半段為流暢易讀的文案，後半段則為可供查閱的參考資料。

有一點需要事先提醒讀者。本書的主題至今仍在持續變動。就在撰寫本文期間，施行日期發生了變更，指引也陸續發布。2026年6月29日，重新擬定多項日期的修正案獲得了最終批准。因此，本書中的日期與數據皆為獨立管理，也建議讀者在展開實際準備之前，務必再次確認該部分內容。土地至今仍在微微動搖。

那天晚上，我和板橋的那位代表進行了長時間的通話。我為他梳理了觸犯了什麼條款、應該先做些什麼，以及哪些事項尚有緩衝時間。通話結束時他說：「要是早點知道就好了。」本書正是為了那個「早點」而寫。希望這是一本讓人在8月2日接到電話之前，就能翻開閱讀的書。

目錄

第一部

為何韓國企業須研讀歐盟法規

第1章 域外適用的實體

就從釜山一家醫療器材公司的故事說起吧。這家公司開發了一款 AI 判讀輔助軟體，能讀取胸部 X 光影像並標註疑似結節部位。研發、訓練、伺服器運作全都在韓國國內進行。公司在歐盟既無分公司也無員工，更未曾簽署過任何合約。然而，一旦這款軟體以授權形式供應給德國某大學醫院的影像醫學科，並開始用於實際判讀輔助時，這家公司便成為歐盟人工智慧法（Regulation (EU) 2024/1689，以下簡稱 AI 法）的適用對象。即便伺服器位於釜山、僅有輸出結果傳至德國，依然如此；即便公司登記簿上沒有半行歐盟地址，依然如此。

許多韓國企業的法務人員常在此處摔跤。因為要接受「我們從未進入歐盟市場」與「我們屬於歐盟人工智慧法的規範對象」這兩句話能夠同時成立的事實，需要花費一些時間。我認為這一點正是貫穿整本書的第一道關卡。跨越國界的並非公司本身，而是產出物。而 AI 法正是以該產出物的到達地為基準來劃定管轄權。

1.1 首爾公司受歐盟法規範的三條路徑

AI 法第 2 條第 1 項是規定本法適用對象的條文。韓國企業涉入其中的路徑共有三條，依條文順序分別為 (a) 款、(b) 款與 (c) 款。

(a) 款路徑針對的是將 AI 系統或通用人工智慧模型（General-Purpose AI model，以下簡稱 GPAI 模型）投放於市場（placing on the market）或投入服務（putting into service）至歐盟市場的提供者（provider）。條文中特別加入「無論是否設立或位於歐盟境內」的字樣。倘若韓國企業在首爾總部開發 AI 系統或 GPAI 模型，並直接販售給歐盟客戶或透過 API 提供，即使在歐盟境內沒有任何實體分公司，亦須承擔提供者義務。在實務上常被忽視的一點是：投放市場可以透過函式庫形式、API 形式、直接下載，甚至是實體複製本來實現。這意味著僅憑一張 SaaS 合約，投放市場即告成立。

(b) 款路徑針對的是設立或位於歐盟境內的 AI 系統部署者（deployer）。首爾總部直接適用此路徑的情況較為少見。然而，若韓國企業在歐盟設立子公司或分公司，且該歐盟法人於業務上使用 AI 系統，則該歐盟法人即成為部署者。雖然在此架構下被規範的對象是當地法人而非總部，但總部亦不可掉以輕心，因為這正是需要建立金控/控股公司層級之合規（compliance）政策之處。

剩下的 (c) 款路徑才是本章真正的核心主題。當 AI 系統產生的產出物（output）於歐盟境內被使用時，無論其提供者與部署者是否部位於歐盟境外，均受本法規範。前述釜山醫療器材公司的案例正是屬於此路徑。我對該條文的解讀如下：歐盟立法者認定，在 AI 時代，以實體國界劃定管轄權的方式已不再有效。在服務透過雲端來往、僅憑一次 API 調用即可跨越國界的產業中，若以公司所在地為基準，規範將變成一張空網。因此，基準被轉移到了產出物的到達地 - - 亦即結果實際被使用的地點。

將這三條路徑總結成一句話即為：賣方、買方（使用方）、結果到達地點，只要這三個地點中有任何一個位於歐盟境內，法律便隨之而來。大多數韓國企業都是在第一條與第三條路徑中牽連其中。

1.2 提供者、部署者、進口商與經銷商之區分

AI 法第 3 條將供應鏈中的角色分為四種。即便是同一家公司，也會因交易結構的不同而對應至這四者之一，其義務負擔的輕重亦大相逕庭。

提供者（provider）定義於第 3 條第 2 款（AI 系統）與第 3 款（GPAI 模型）。指開發或指示他人開發 AI 系統或 GPAI 模型，並以自身名稱或商標將其於歐盟市場商業化的主體。在四個角色中，提供者的義務最為繁重。高風險 AI 系統提供者必須常態性運作風險管理系統、驗證訓練資料集的偏誤，並須從設計階段起即植入技術文件與自動記錄功能。GPAI 模型提供者自 2025 年 8 月 2 日起，須承擔四項義務：編製與維持技術文件、向下游整合業者提供資訊、建立歐盟著作權法合規政策，以及公開訓練內容摘要。這四項義務即使在《數位綜合法案》（Digital Omnibus）修訂中亦未經變動，意味著它們依然是有效適用的義務。

若 GPAI 模型的累計計算量超過 10 的 25 次方 FLOPs，而被推定具有系統性風險（systemic risk），則會增加一層義務：符合最新技術水準的模型評估、敵對性測試（即紅隊測試）、系統性風險的評估與減緩、向 AI 辦公室通報重大事故，以及確保適當水準的資安防護。一旦達到門檻，必須毫不延誤且最遲於 2 週內通報 AI 辦公室。若韓國國內的生成式 AI 新創公司正規劃自行開發大型模型並供應至海外，我認為應將此門檻計算放在商業計畫書的正文中，而非僅作為腳註。

部署者（deployer）定義於第 3 條第 4 款。指在其權限下基於業務目的使用 AI 系統的公務機關、機構、團體或其他法人，有時也被稱為 AI 系統的購買者或使用者。若是高風險 AI 系統的部署者，必須依據提供者給予的使用說明書運作系統、管理輸入資料的關聯性，並維持由人類掌握最終判斷的人類監督機制。如前所述，首爾總部直接成為部署者的情況極為罕見，常見的結構是由歐盟當地法人承擔部署者角色。

進口商（importer）定義於第 3 條第 6 款，指將歐盟境外生產的 AI 系統首次引進歐盟市場的主體。經銷商（distributor）定義於第 3 條第 7 款，指於歐盟市場販售該系統，但既非提供者亦非進口商之人。兩者皆扮演供應鏈中間人的角色。若韓國企業開發的 AI 系統由歐盟合作夥伴購入並首次投放於歐盟市場，則韓國企業為提供者，該合作夥伴為進口商。雖然責任有所劃分，但並不代表責任會隨之減輕。因為進口商與經銷商在將高風險 AI 系統投放於市場前，有義務確認是否已張貼 CE 標誌、是否附有歐盟符合性聲明

與使用說明書，以及提供者與進口商是否各自履行其義務。在合約中僅寫下「提供者為韓國企業，進口商為歐盟合作夥伴」即草草收尾的實務做法相當危險。若合作夥伴未能實際履行進口商側的義務，該不履行最終將演變成撼動韓國企業整條供應關係的意外事故。

補充說明一項條文：當 AI 系統經過實質性變更（substantial modification）或其預定目的發生改變而成為高風險系統時，承擔義務者並非原提供者，而是進行該變更的主體將作為新提供者承擔義務。若韓國企業製作的通用影像分析引擎被歐盟合作夥伴改造為醫療診斷用途使用，這意味著提供者地位可能會轉移至該合作夥伴身上。這一點的結果取決於合約中對於改造與再授權條款撰寫得有多精細。

1.3 產出物於歐盟境內被使用之情形：第 2 條第 1 項 (c) 款

關於此條文為何對韓國企業尤為重要，前文已有說明。在此我們將進一步深入探討。

第 2 條第 1 項 (c) 款明確規定：「無論是否設立或位於歐盟境內，當 AI 系統產生的產出物於歐盟境內被使用時，該提供者與部署者」均適用本法。實體位置與提供者的國籍皆非基準，最終產出物的到達地或產生影響的地點才是基準。我認為，這樣的設計是 AI 技術無國界的特性與保護歐盟公民健康、安全及基本權利之立法意旨相結合的結果。

在實務中，觸及此條文的路徑比想像中更加多樣。問題不單僅限於直接販售。提供 API、供應軟體函式庫、嵌入於行動應用程式（App）中的模型，這些全都會成為產出物流入歐盟的管道。若歐盟居民透過網頁或 App 使用韓國企業開發的 AI 翻譯服務，且該翻譯結果於歐盟境內被使用，單憑該服務即可成為適用對象。前述釜山醫療器材公司的判讀輔助軟體亦然。即使運算引擎在韓國國內資料中心運作、僅傳送輸出結果，只要該輸出結果實際影響了德國醫護人員的判斷，便立刻落入法律規範的效力範圍之內。

坦白說，「產出物於歐盟境內被使用」這一文句究竟確切涵蓋至何種程度，目前尚未完全定案。在如 AI 即服務（AI as a Service）這類經過多個階段傳遞結果的架構中，各階段主體的責任範圍極有可能依具體個案分別判定。我認為無須僅將此不確定性視為壞消息。在判例與執行案例皆尚未豐富的現階段，採取保守解釋並預先做好準備，反倒能在日後案例累積時立於有利地位。

關於執行層面再補充一些內容。AI 法的合規情況由 AI 辦公室（AI Office）進行監督，且 AI 辦公室對於 GPAI 模型提供者擁有獨占性管轄權。AI 辦公室會將是否遵守協調標準（harmonised standards）、是否簽署並實際遵守 AI 生成物透明度實務守則（Code of Practice）作為判定合規與否的依據。雖然簽署實務守則並非法律義務而屬自願性質，但我認為這比起單純的選項更接近於一種保險。因為日後在裁定罰鍰金額時，這有發揮減輕因素作用的空間。

我們也來精確梳理日期與金額。AI 法於 2024 年 8 月 1 日生效，自 2025 年 2 月 2 日起，針對操縱人類行為或社會信用評量等被禁止之行為（第 5 條）的規範全面施行。自 2025 年 8 月 2 日起，GPAI 模型提供者義務（第 51 條至第 55 條）開始適用，此部分在後文將說明之修訂中亦維持不變。截至目前為止皆為已過期的日期，無可爭議。

問題在於接下來原定於 2026 年 8 月 2 日施行的高風險 AI 系統（附錄 III，涵蓋招聘與人事管理、必要公共服務、生物識別、司法行政等 8 大領域）義務的全面施行日。在整理 NotebookLM 第一手資料時，資料尚將此日期記載為已確定的時間點，但經再次搜尋確認後，情況已有所變動。歐洲議會於 2026 年 6 月 16 日採納了《人工智慧數位綜合法案》（Digital Omnibus on AI）修正案，歐盟理事會於 6 月 29 日完成最終批准，並於 7 月 8 日簽署最終法案。透過此次修訂，附錄 III 高風險系統的單獨義務施行日已延後至 2027 年 12 月 2 日，而嵌入於已受歐盟產品安全法規規範之產品中的 AI 系統（附錄 I）的施行日則順延至 2028 年 8 月 2 日。以撰寫本文的 2026 年 7 月下旬為準，該修正案僅剩於官方公報（Official Journal）刊載一關，刊載三天後即可生效，預計最遲本月內將發生效力。然而，2026 年 8 月 2 日這個日期本身並非完全消失。第 50 條的大部分透明度義務以及 AI 辦公室對 GPAI 模型的罰鍰處分權，將在該日期照常生效啟用。這意味著僅有高風險規範被順延，透明度義務與 GPAI 的執行力仍將按原定計畫啟動。我認為此處正是分水嶺所在。絕不能僅憑新聞標題就認定「高風險規範整體延後，因此 2026 年可以高枕無憂」。被延後的僅是附錄 III 的一部分，其餘義務依然會按時向我們走來。

罰鍰金額分為三個層級。違反被禁止之行為（第 5 條）者，最高可處 3,500 萬歐元或全球年度營業額的 7%，以較高者為準。包含違反提供者、部署者、進口商、經銷商及 GPAI 模型提供者義務在內的大多數其他違規行為，最高可處 1,500 萬歐元或全球年度營業額的 3%，以較高者為準。向主管機關提供不正確、不完整或具誤導性資訊者，最高可處 750 萬歐元或全球年度營業額的 1%，以較高者為準。常會看到有資料將最後這筆金額誤寫為 75 萬歐元，但依據第 99 條條文原文，正確數字為 750 萬歐元。雖然只差一個位數，但這是要向董事會或管理階層報告的數字，務必精確掌握。中小型企業（SME）與新創公司在三個層級中皆享有優惠，適用定額與營業額比例中較低者。

1.4 歐盟境內代表：第 22 條與第 54 條

設立於歐盟境外的提供者若欲將高風險 AI 系統或 GPAI 模型投放於歐盟市場，必須透過書面委任於歐盟境內指定授權代表（Authorised Representative）。高風險 AI 系統部分依據為第 22 條，GPAI 模型部分則依據第 54 條。兩條文的共同點在於皆明確規定須透過書面委任進行指定，且兩者皆是強制為無實體存在之提供者建立法律接觸點的機制。

授權代表所做的事並不局限於保管文件。其必須驗證提供者的技術文件是否符合 AI 法的要求，並須將該紀錄保存 10 年。當歐盟市場監督主管機關或 AI 辦公室要求提供資訊時，授權代表須配合回應，並負有包含風險減緩措施在內的合作義務。若有必要，亦須代為辦理歐盟註冊程序。在此處，實務人員常會遺漏一個關鍵：一旦確認提供者未能遵守 AI 法，授權代表有義務主動終止委任並立即通報相關主管機關。也就是說，授權代表不僅是提供者的代言人，同時也是監視提供者的管道。在簽署代表合約時，若輕忽此條文的分量，日後將面臨困境。

亦存在例外情形。當 GPAI 模型以開源授權（Open Source License）發布時，只要未被歸類為具有系統性風險，指定代表的義務本身即可獲得豁免。然而，這並不意味著遵守著作權法政策（第 53 條第 1 項 (c) 款）與公開訓練內容摘要（同條 (d) 款）的義務也會一併豁免。這代表著即使以開源方式公開，也並不意味著能完全擺脫著作權問題。

再次整理準備期限：GPAI 模型提供者義務已自 2025 年 8 月 2 日起開始適用。對於在此之前已投放市場的模型，設有須於 2027 年 8 月 2 日前完成合規措施的過渡規定（即所謂的祖父條款，grandfathering clause），且此過渡規定將貫穿模型的整個生命週期。若總部設於首爾的公司規劃進入歐盟市場，我認為應將這些日期放在合約談判時程表中，而非僅放在事業路線圖的腳註裡。我曾屢次見過因匆忙尋找代表而在合約條件上陷於被動的情況。爭取到時間，就等於爭取到了談判籌碼。

此外，還須梳理與韓國國內法規的關係。韓國已於 2026 年 1 月 22 日施行《人工智慧發展與信任基礎塑造等相關基本法》（AI 基本法），並自同年 7 月 21 日起適用具體化委任事項的修正施行細則，新增了 AI 產品與服務確認制度等機制。韓國國內 AI 基本法的行政罰鍰約在 3,000 萬韓元以下，與最高可處全球營業額 7% 的歐盟 AI 法在量體上截然不同。儘管如此，亦不可輕視韓國國內規範。因為針對同一個 AI 系統，在韓國國內須遵守 AI 基本法，而對於流向歐盟的產出物則須遵守歐盟 AI 法，這種雙重規範架構已經形

成。還需注意的是，這兩部法律的風險分類標準與義務項目並非完全重疊。因此，無法直接將韓國國內法的應對檢核表複製移用到歐盟側。

1.5 自我診斷表

為使各位能於實務中即時確認先前閱讀的內容，以下將其整理為五個問題。若有任何一題回答「是」，請返回相對應的節次確認相關義務。

1) 貴公司是否以自身名稱或商標開發 AI 系統或 GPAI 模型，並直接販售給歐盟客戶，或透過 API、授權等形式提供？若是，貴公司即為第 3 條第 2 款及第 3 款規定的提供者。將適用第 2 條第 1 項 (a) 款，且須檢討依據第 22 條或第 54 條指定歐盟代表之義務。

2) 貴公司是否於歐盟境內設立子公司或分公司，且該法人於業務上使用 AI 系統？若是，該歐盟法人即為第 3 條第 4 款規定的部署者。若屬高風險系統，該法人層級須具備人類監督機制與輸入資料管理義務。

3) 貴公司開發的 AI 系統是否由歐盟合作夥伴購入並首次投放於歐盟市場？若是，合作夥伴即為第 3 條第 6 款規定的進口商，而貴公司為提供者。請重新檢視合約中是否載明確認進口商側履行 CE 標誌、符合性聲明及使用說明書義務之條款。

4) 即使貴公司於歐盟境內無實體存在，貴公司系統所產生的結果物或服務是否透過 API、函式庫、行動應用程式（App）傳達至歐盟境內使用者並實際被使用？若是，依據第 2 條第 1 項 (c) 款，須檢討完整的提供者義務。前述釜山醫療器材公司的案例即屬此種情形。

5) 貴公司系統是否涉及附錄 III 所列之招聘與人事管理、必要公共服務、生物識別、司法行政等領域？或是已內建於受歐盟產品安全法規規範之產品（醫療器材、機械、玩具、航空器等）中？若為前者，施行日已延後至 2027 年 12 月 2 日；若為後者，則為 2028 年 8 月 2 日。然而，切不可忽略於 2026 年 8 月 2 日啟動之第 50 條透明度義務以及 AI 辦公室對 GPAI 模型之罰鍰處分權，皆與此延期無關並將照常適用。

若釜山醫療器材公司的負責人重新面對這五個問題，大概會同時在第 1 題與第 4 題舉手。一家一直以為僅在韓國國內營運的公司，在歐盟法規條文上卻已同時承擔著兩項義務。究竟何時意識到此事實 - - 是在與醫院的合約已簽訂之後，還是在合約蓋章之前 - - 這當中的差別將決定從下一章開始展開的故事份量。

第2章 法律的骨幹

板橋的醫療保健新創公司代表上個月前來找我。他表示想把一項能在醫院急診室讀取患者表情與語氣並評估風險等級的服務，導入歐洲幾家醫院。他帶來的問題只有一個：「這個在歐洲可以直接使用嗎？」我在翻開條文前反問他：「該服務比較接近監控人類，還是比較接近協助人類？」代表停頓了一會兒。那段沉默便是本章的起點。歐盟《人工智慧法案》（Regulation (EU) 2024/1689，以下簡稱《AI法案》）正是以這個問題 - - 「該系統給人類帶來多大程度的風險？」 - - 建立起整體結構。

2.1 以「風險為基礎的方法」這種設計思想

《AI法案》於2024年8月1日生效。光是條文就有113條，若加上附錄，厚度足足有一整本書。然而貫穿這部法律的骨幹卻意外地簡潔，那就是「風險越大，義務越重；沒有風險，就沒有義務」的以風險為基礎的方法（Risk-Based Approach）。用美式邏輯來說，它不是在單一原則上不斷堆疊例外，而是先用名為「風險度」的尺來衡量系統，再根據刻度劃分管制強度。

這項設計在實務上的意義非常明確。即使是同一家公司開發的兩款 AI 產品，其中一款可能完全無需申報即可銷售，另一款則必須通過合格評定（conformity assessment）並取得 CE 標誌後才能販售。區分的標準並非技術的精密度，而是該技術對人類健康、安全及基本權利影響的大小。這就是為什麼板橋代表的情緒識別服務，會依據其究竟是醫院診療輔助工具還是勞工監控工具，而適用截然不同的法律規範。

《AI法案》將系統大致劃分為四個層級：絕對不得使用的被禁止行為、承擔嚴格事前義務的高風險系統、必須告知特定資訊的透明度義務對象，以及實際上可自由使用的最小風險系統。比起將這四個層級繪製成金字塔，我更喜歡將其比作設有四道門檻的走廊。您可以理解為，一個系統在這條走廊上前進時，依序接受是否能通過第一道門檻、第二道門檻的審查結構。

2.2 禁止、高風險、透明度與最小風險這四個層級

第一道門檻是根本無法通過的門。這屬於《AI法案》第5條所規定的「被禁止的 AI 行為」（prohibited AI practices）。代表性範例包括：在人類未察覺的情況下操縱其行為並造成傷害的系統；針對年齡或身心障礙等脆弱性以矇蔽其判斷力的系統；政府或企業對個人的社會行為進行評分並給予不利處分的分數機制（Social Scoring，社會信用評分）；以及在公共場所對不特定多數人進行即時人臉辨識的遠端生物特徵識別系統。該條款在法律生效6個月後的2025年2月2日起便已全面施行。板橋代表打算製作的情緒識別服務也必須先跨過這道門檻。如果是用於職場出勤管理，或是用於教育現場讀取學生情緒並反映於考評，除非符合特定例外（如安全或醫療目的），否則極可能觸犯本條規範。我建議在導入情緒識別功能時，務必將用途限定於「診療輔助」，並將其依據以書面文件留存。

通過第二道門檻後，即屬於高風險（High-Risk）AI 系統。此處有兩條進入路徑：其一是依據《AI法案》第6條第1項規定，受既有歐盟產品安全法規（附錄 I）規範的產品，例如基於 AI 的醫療器材或汽車零件中的安全組件；其二是列於《AI法案》第6條第2項及附錄 III 的 8 個應用領域，例如招募與人事管理、取得必要公共服務、生物特徵識別、執法、移民與庇護及國境管理、司法行政等領域。即便名稱列於附錄 III 中，若實際上未對人類的健康、安全或基本權利造成重大風險，仍有可排除於高風險之外的例外條款（第6條第3項）。關於如何判定此例外的委員會指南原定於2026年2月2日前出爐，但截至撰寫本文時，尚未能確認最終確定版本。在實務上，於該指南出爐之前採取保守態度 - - 即只要列名於附錄 III 就先假定為高風險並予以因應，會是比較安全的做法。

被歸類為高風險的提供者（provider），依據第16條將承擔沉重的負擔。這些負擔包括：在系統整個生命週期中識別並降低風險的風險管理系統、載明開發過程與評估結果的技術文件、包含自動生成日誌的紀錄保存、向部署者（deployer）提供充足資訊的透明度、允許人類最終翻轉結果的人類監督機制（human oversight）、精準度與穩健性及網路安全、品質管理系統、上市前的合格評定（conformity assessment）與 CE 標誌，以及上市後持續進行的上市後市場監督（post-market monitoring）。部署者（deployer）端同樣無法袖手旁觀，必須依據第26條規定，按照提供者給予的操作手冊運作系統、管理輸入資料，並切實維繫人類監督機制。

這第二道門檻的施行時間，是目前撰寫本書時變動最大的部分。原本針對附錄 III 高風險系統的義務，預計於法律生效 24 個月後的 2026 年 8 月 2 日起全面適用。然而，歐盟理事會於 2026 年 6 月 29 日對所謂的「數位萬用（Digital Omnibus）」法案包達成最終協議，並於 2026 年 7 月 8 日完成簽署，目前正等待刊登公報。根據該法案包，附錄 III 高風險系統的義務將延後 16 個月至 2027 年 12 月 2 日，而作為附錄 I 所屬既有產品安全法規之安全組件的 AI 系統則延至 2028 年 8 月 2 日。表面上的理由是正式會員國的協調機構與調和標準（harmonised standards）尚未準備就緒。我所參考的 NotebookLM 資料仍以 2026 年 8 月 2 日全面適用為前提進行整理，這一點便是資料與現實脫節之處。對於像板橋代表一樣目前正在設計產品的人來說，建議不要將此延期視為鬆一口氣的訊號，而是將其視為準備時間延長的訊號。監管機構給予更多時間，並非降低了要求標準，而是連他們自己都還沒準備好。

第三道門檻是透明度（Transparency）義務。第50條所規定的這項義務，即使是非高風險系統也同樣適用。如果開發了像聊天機器人一樣直接與人類對話的 AI 系統，除非依當時情況已顯而易見對方是 AI，否則必須告知該事實。對於 Deepfake（深偽）以及由 AI 生成或操縱的影像、語音、文字，必須予以標示，使其人工製作的事實得以彰顯。當部署者（deployer）使用情緒識別系統或依生物特徵對人類進行分類的系統時，也必須告知受測對象。這項義務無論是聊天機器人、影像生成器、語音合成工具還是寫作輔助工具，不論是否屬於高風險，皆廣泛適用。有趣的一點是，在本次數位萬用法案包中，第50條的核心骨幹依然於 2026 年 8 月 2 日施行。僅像浮水印（watermarking）這類以技術標示為 AI 生成物的具體執行手段，其適用期限延後了三個月至 2026 年 12 月 2 日。委員會於 2026 年 5 月 8 日發布了關於誰應如何履行聊天機器人揭露義務與深偽標示義務的指南草案，實踐守則（Code of Practice on Transparency of AI-Generated Content）亦同步籌備中。若板橋代表的服務會在醫院螢幕上顯示情緒分數，無論是否被認定為高風險，幾乎可以確定都會觸及這項第50條的告知義務。

第四個是最小風險（Minimal Risk）系統。既非禁止、非高風險，亦非透明度義務對象的大多數 AI 系統，例如垃圾郵件過濾器或推薦演算法等皆屬於此類。法律實際上幾乎沒有強制性義務，僅鼓勵自願遵守行為守則（codes of conduct）。不過，我並不建議因為屬於最小風險就置之不理。因為歐盟客戶在盡職調查過程中詢問是否自願遵守守則的情況日益增加，且未來服務範圍擴大而劃入附錄 III 領域的可能性也始終存在。

2.3 與 GDPR、產品責任指令及資料法案之關係

《AI法案》並非單獨存在的法律。首先必須審視其與歐盟《一般資料保護規則》（GDPR）的關係。《AI法案》第2條第7項明確規定，凡涉及個人資料處理者，包括 GDPR 在內的既有歐盟法律均直接適用。這意味著《AI法案》並非取代 GDPR，而是疊加在其上的結構。實務上重疊之處在於「影響評估」。若高風險 AI 系統處理個人資料，則必須同時進行依據 GDPR 第35條的資料保護影響評估（DPIA），以及依據《AI法案》第27條第4項的基本權利影響評估（Fundamental Rights Impact Assessment）；倘若已有執行完成的 DPIA，則可以補足基本權利影響評估的方式來減少重複作業。比起將兩份評估分開製作的公司，我認為在一份文件內將個人資料風險項目與基本權利風險項目並列陳述的公司，會在盡職調查中給人留下相當有利的印象。

亦須關注其與《產品責任指令》（Product Liability Directive）及既有產品安全法規的關係。若 AI 被應用於醫療器材、機械、玩具、航空器零件中，該產品所適用的既有安全認證便會與《AI法案》的高風險義務同時重疊。《AI法案》第74條第4項規定，若附錄 I 第 A 節所列之既有調和法規已設有提供同等保護水準的程序，則可適用該特定領域程序，以替代《AI法案》的特定程序（第79條至第83條）。這相當於在醫療器材認證體系等已有嚴密監管的領域中進行協調，避免雙重重複管制。然而，即便已經取得 CE 標誌，並不意味著可以豁免《AI法案》中的風險管理系統或技術文件義務。既有認證看重的是整個產品的安全，而《AI法案》則是分別審視其中嵌入的 AI 組件之風險。

《資料法案》（Data Act）的大部分條款已於 2025 年 9 月 12 日開始適用。擴大使用者與第三方對連網產品（connected products）所產生資料的存取權是該法案的核心，當其與《AI法案》中訓練資料摘要之公開義務（第53條第1項(d)款）相結合時，結果將朝著施壓要求更透明地揭露資料來源的方向發揮作用。《數位服務法》（DSA）與《網路韌性法案》（Cyber Resilience Act）亦是須與《AI法案》並列對照的相鄰法規。若線上平台使用 AI，就必須同時兼顧 DSA 與《AI法案》；若為嵌入 AI 的數位產品，則必須找出兩部法規在網路安全要求上的交集點。

我在向韓國企業說明這種關係時，總會使用相同的比喻：如果說 GDPR 是處理「個人資料」這種食材的法律，那麼《AI法案》就是處理用該食材做出的「料理」 - - 即系統本身的安全性與性能的法律。洗淨了食材並不保證料理絕對安全，料理做得好也不代表食材

來源合法正當。將這兩部法律分開檢討的公司，最終只會把相同的檔案製作兩次。

2.4 制裁力道：以全球營業額為基準裁處的罰鍰

《AI法案》第99條要求成員國建立有效、比例適當且具威嚇力的罰則體系。其所產生的罰鍰結構分為三個層級。最重的層級適用於違反第5條「被禁止行為」以及違反資料相關要求事項時。上限為 3,500 萬歐元或前一會計年度全球年總營業額的 7%，以較高者為準。這意味著，若板橋代表的情緒識別服務在觸犯第5條禁止領域的情況下於歐盟銷售，將可能以整家公司的全球營業額為基準，被處以最高達 7% 的罰鍰。我每次在諮詢場合都會再三強調：基準是「全球營業額」，而非僅限「歐盟營業額」。

中等層級適用於違反《AI法案》的其他義務（即除第5條外的大多數條款）。上限為 1,500 萬歐元或全球年總營業額的 3%，以較高者為準。通用人工智慧模型（General-Purpose AI model，GPAI）提供者的違規行為亦屬於此層級，《AI法案》第101條另行詳細規定了對 GPAI 模型提供者的罰鍰規範。GPAI 提供者的基本義務已自 2025 年 8 月 2 日起適用；若用於訓練的累積運算量超過 10 的 25 次方 FLOPs（每秒浮點運算次數），將被推定為具有系統性風險（systemic risk），進而須承擔第55條的附加義務，如模型評估與紅隊演練（adversarial testing）、保障網路安全、通報嚴重事故等。一旦超過此門檻，必須毫不延誤且最遲於 2 週內通報 AI 辦公室（AI Office），若錯過通報，該行為本身即會成為中等層級罰鍰的對象。

最低層級適用於向驗證機構或主管機關提供不正確或具誤導性資訊的情況。上限為 750 萬歐元或全球年總營業額的 1%，以較高者為準。中小企業（SME）在三個層級中皆享有例外規定：適用兩金額中「較低者」而非「較高者」。當歐盟機構或機關違規時，則採用與一般企業不同的標準，由歐洲資料保護監督官（EDPS）裁處罰鍰；若違反第5條，上限最高為 150 萬歐元，其他違規上限最高為 75 萬歐元，上限大幅降低。

AI 辦公室實際擁有裁處該罰鍰權限的時間點為 2026 年 8 月 2 日起。然而，正如前面 2.2 所述，在附錄 III 高風險義務本身已延至 2027 年 12 月 2 日的當下，高風險相關罰鍰實際發動的時間點自然也順延至其後。相反地，GPAI 義務與第50條透明度義務乃按原定時程推進，因此這兩個領域的罰鍰風險已處於發動狀態，或應視為自 2026 年 8 月起立即發動。由於日期散落在多個時間點，後文已整理成表格供參。

主管機關在審定罰鍰金額時考量的因素，亦列於第99條第7項。包括：違規的性質、嚴重程度與持續時間；受影響的人數與損害規模；屬於故意還是過失；違規企業的規模、

營業額與市場占有率；因違規獲得的財務利益或規避的損失；與主管機關的合作程度；企業所具備的技術與組織措施水準；是否自首通報；以及是否已在其他成員國因相同違規行為遭到裁罰。此處在實務上有張相當實用的牌：第101條第1項明定，遵循經 AI 辦公室與 AI 委員會評估為適當之實踐守則（Code of Practice）的履行承諾，可在計算罰鍰時作為減輕要素。我常解釋道，這與其說是扣減罰金的折價券，倒不如說更像是預先向主管機關說明「這家公司一直很合規」的聲明書。未遵循實踐守則雖不致於立刻處於劣勢，但只要有遵循紀錄，談判桌上就多了一張可用的籌碼。

亦必須提及「域外適用（Extraterritorial application）」。依據《AI法案》第2條第1項(a)款及(c)款，即便是位於歐盟境外的提供者，只要其系統產出結果（output）於歐盟境內被使用，即受本法管轄。即便板橋代表的公司僅在首爾設有辦公室，一旦將服務對接到歐盟的醫院，便會立即納入與歐盟境內競爭對手相同的罰鍰體系中。且該罰鍰並非依據在歐盟創造的營業額計算，而是以公司的「全球營業額」為基準。這意味著，即使在歐盟的營業額占不到全公司 1%，只要總公司營收規模龐大，便可能面臨高達全公司總營業額 7% 的風險。歐盟境外提供者必須以書面形式在歐盟境內指定授權代表（Authorised Representative），以作為與主管機關聯絡的管道；我曾多次觀察到，越是延後指定授權代表的公司，在真正發生事故時的應變越是遲鈍。

參考資料

生效日期表

2024 年 8 月 1 日，《AI法案》生效。 2025 年 2 月 2 日，第5條「被禁止的 AI 行為」全面施行。 2025 年 8 月 2 日，GPAI 模型提供者之基本義務（第53條）開始適用。 2026 年 8 月 2 日，第50條透明度義務施行、AI 辦公室裁處罰鍰權限生效，以及原定附錄 III 高風險義務全面適用日（因數位萬用包延至 2027 年 12 月 2 日）。 2026 年 12 月 2 日，第50條相關浮水印等技術標示手段之執行寬限期屆滿日。 2027 年 12 月 2 日，反映數位萬用包後附錄 III 高風險義務預定全面適用日。 2028 年 8 月 2 日，附錄 I 產品安全法規適用對象 AI 組件之義務預定適用日。

本表中 2026 年與 2027 年的日期，係反映數位萬用法案包雖已於 2026 年 7 月 8 日完成簽署、但截至撰寫本書時尚未確認刊登公報狀態之暫定數值。建議於刊登公報後，立即再次確認公報刊登日與最終條文編號。

檢核項目

我們的系統是否已先確認觸及第5條禁止清單。 若屬於附錄 III 的 8 個領域之一，是否已以書面文件檢討是否符合第6條第3項的例外要件。 若具備聊天機器人或生成式功能，是否已擬定第50條告知義務的執行方案（告知警語、浮水印）。 若訓練 GPAI 模型，是否已計算累積運算量是否接近 10 的 25 次方 FLOPs。 是否已將 GDPR 項下的 DPIA 與《AI 法案》項下的基本權利影響評估整合為單一文件體系。 若為歐盟境外提供者，是否已以書面指定授權代表。 是否妥善保存是否參與實踐守則及其相關紀錄。

板橋代表結束當天的諮詢離開時問道：「那麼，到底要準備到什麼時候為止呢？」我無法斬釘截鐵地指明一個日期，因為這部法律自己也正在把那個日期修改重寫了好幾遍。

第3章 2026年的時間表，重新繪製的地圖

板橋（Pangyo）一家新創公司的合規負責人重新打開了去年製作的應對時程表。有一半的儲存格被標示為紅色，這意味著日期錯了。去年此時該公司擬定的計畫，是將所有準備工作集中在2026年8月2日這單一時間點上。當時他們以為高風險人工智慧規定會在那天全面適用，研發團隊與法務團隊都只盯著那個日期全力衝刺。然而，該日期的相當一部分被往後延期了。確切來說，集中在8月2日的各項義務中，有些維持不變，有些則後退了十六個多月。

本章旨在將這份出現偏差的地圖交到各位手中。一旦錯過日期，本書其餘二十二章的內容都將隨之動搖。如果本應將資源投入在8月2日前急需具備的事項上，卻在現在死守著要到2027年才適用的義務，那就是順序倒置；反之，若誤以為延期而放手不管，結果卻觸犯了實際上維持原狀的義務，那將是更嚴重的事故。我在撰寫本章時，將每一個日期透過兩種管道進行了雙重比對確認：歐盟委員會與理事會發布的官方文件，以及撰寫本稿今日的最新報導。若兩者存在分歧，我會在文中明確指出。

讓我們先掌握大局。《歐盟人工智慧法》（Regulation (EU) 2024/1689）並非在同一天全面生效執行的法律。其架構是在生效後，各項義務依時間差逐步啟動。而2025年11月啟動的 Digital Omnibus 修訂案，則重新調整了部分時間差。這項修訂案於2026年6月29日獲得理事會最終批准，完成了立法程序，截至撰稿時的2026年7月24日，正處於等待刊登公報的階段。接下來您將看到的合規時間表，是將此修訂案疊加於原始法律條文之上、屬於此時此刻的地圖。

3.1 2024年8月1日生效，以及分階段適用之設計

《人工智慧法》於2024年8月1日生效。生效與適用（施行）是不同的概念。法律開始存在稱為生效，而各項具體義務開始實際產生約束力則稱為適用。本法並未在生效當天啟動所有義務，第113條針對各條文分別規定了不同的適用日期。

為何要這樣區分？因為規範對象的風險等級各有不同。對人類造成傷害風險較高的行為實踐必須儘早禁止；而備妥技術文件並通過合格評估等工作，則需要給予企業構建新系統的時間。因此，法律採取了「早期啟動禁止規定，延後啟動繁重準備義務」的時差設計。掌握這一設計原理，就能理解後續出現的日期為何以該順序排列 - - 即「緊急風險優先，需長時間準備的義務在後」的順序。

3.2 2025年2月2日：被禁止的實踐與人工智慧素養義務

最先啟動的是禁止規定。2025年2月2日，第5條所列舉之被禁止的人工智慧實踐管制全面適用。潛意識操縱或利用弱勢處境的系統、公私部門的社會信用評分、僅憑側寫（profiling）預測個人犯罪傾向的執法活動、無差別擷取網際網路或CCTV影像建立人臉辨識資料庫的行為、職場與教育機構中的情緒辨識、推論種族、政治傾向、工會成員身分、宗教或性傾向的生物辨識分類，以及執法目的之即時遠端生物辨識（存在有限例外），自該日起在歐洲既不能投放市場、不能作為服務提供，也不能實際使用。關於何種行為因何被禁，將在第15章按條文逐一探討。

同一天，第4條的人工智慧素養義務也隨之啟動。該義務要求提供者與部署者應確保其員工及相關操作人員具備負責任地處理人工智慧系統並評估風險的知識與技能。許多公司甚至忽略了這項義務已經適用的事實，因為這是一條容易被悄悄忽視的條款。第21章將重新審視實際上需要具備哪些條件。

違反禁止規定的罰則嚴厲程度在整部《人工智慧法》中高居榜首：最高可處3,500萬歐元或全球年度營業額的7%，以較高者為準。對於總部位於首爾的公司而言，比對評估在歐盟境內使用或投放歐盟市場的人工智慧系統是否符合此清單，是本書中不可拖延的首要工作。

3.3 2025年8月2日：通用人工智慧模型提供者的義務

自生效起整整過了一年的2025年8月2日，通用人工智慧模型（General-Purpose AI model, GPAI）提供者的義務正式啟動。第3條第63款所定義的通用模型，是指利用大量數據進行訓練，具有相當程度的通用性，並能廣泛執行各種任務的模型。語言模型是典型的例子，生成圖像、影片或語音的模型也可能包含在內。

第53條所賦予的義務要點如下：撰寫附錄XI（Annex XI）所要求的技術文件並在模型整個生命週期內保持最新狀態；向意圖引入該模型並整合至自身系統的下游提供者提供附錄XII所規定的資訊；制定尊重歐盟著作權指令（2019/790）第4條第3項權利保留聲明的政策；以及撰寫用於訓練之內容的公開摘要。此外，第55條則對具有系統性風險的模型提供者提出了額外要求，包括模型評估與紅隊演練（red teaming）、風險評估與減緩、重大事故通報，以及適當的資安防護措施。劃分是否具備系統性風險的門檻為累計訓練運算量達到 10 的 25 次方 FLOPs。一旦達到或預期將達到此門檻，最遲須在2週內向人工智慧辦公室（AI Office）通報。目前全球已知超越此門檻的模型大約在十二到十五個左右。這些義務的具體內容將在第三部第7章至第10章中逐一剖析。

同一天，治理規定也同步開始適用，人工智慧辦公室與各成員國主管機關的監督體系正式運作。然而，這裡隱藏著一個陷阱：義務啟動的日期與對違規行為開罰的日期並不相同。人工智慧辦公室設定了從2025年8月2日至2026年8月2日為期一年的所謂「無執法的適用」期間。這一年間義務雖然明確存在，但不會裁處罰鍰。切不可將這一段期間解讀為「沒有合規義務」。違規事實仍會持續記錄，並將從下一節所討論的日期起正式追究責任。

3.4 2026年8月2日：第50條透明度義務的全面施行與通用人工智慧執法權限

這是迫在眉睫的日期。2026年8月2日將同時發生兩件性質不同的事情。

其一是第50條透明度義務的全面施行。與人類直接互動的人工智慧系統，必須明確告知對方正與人工智慧對話；而生成合成音訊、圖像、影片及文本的系統，則必須以機器可讀的方式標示其產出，使其人工生成的特性可被偵測。使用情緒辨識或生物辨識分類系統的部署者，必須通知受影響的對象；對於深偽技術（Deepfake）與人工智慧生成的文本，亦須公開其經操縱或生成的事實。告知最遲應於首次互動或暴露時，以明確、可區分且可取得的方式進行。針對此項義務的討論將在第二部第4章至第6章持續展開。

其二是人工智慧辦公室對通用人工智慧模型執法權限的啟動。通用模型提供者的義務本身雖已於2025年8月2日啟動，但對違規行為實際裁處罰鍰的權限，則自一年後的這一天起正式行使。前述的「無執法的適用」期間即在此結束。罰鍰上限為全球年度營業額的3%或1,500萬歐元，以較高者為準。對於經營通用模型的企業而言，我將這一天解讀為寬限期結束的日子。

有一點需特別注意：在第50條標示義務中，僅對生成式系統設有一項例外。在2026年8月2日之前已經投入市場的生成式人工智慧系統，可獲得為期四個月的緩衝期，延至2026年12月2日前備齊標示義務即可。新推出的系統則無此寬限，自8月2日起立即適用。此一區分將在3.8節再次加以強調。

3.5 Digital Omnibus 修訂案：2026年6月 29日理事會最終批准帶來的改變

現在我們來看看讓這份地圖重新繪製的修訂案。歐盟委員會於2025年11月19日提出了名為 Digital Omnibus on AI 的法案組合，旨在減少《人工智慧法》施行過程中可能產生的延遲與不確定性，並釐清規則。其核心內容是建議將高風險人工智慧義務的適用日期從2026年8月2日往後延期。

該提案轉化為法律的過程如下：2026年5月7日，理事會、歐洲議會與歐盟委員會達成三方臨時協議；6月16日，歐洲議會全體會議批准該協議；6月29日，理事會完成最終批准。歐盟理事會表示，此項批准旨在釐清規則並使其推行更為順暢。截至撰寫本稿的今天（2026年7月24日）所確認的最新報導，最終法案已於2026年7月8日簽署，目前正等待刊登公報。有預測指出，為了避免與預定於8月2日施行的條文產生時差，最晚須在7月30日前刊登於公報。若依照刊登後第三天生效的原始規定計算，時間剛好吻合。然而，截至本稿截稿時，實際刊登日期尚未確定。建議讀者在閱讀本書時，再次確認是否已刊登於公報。

此修訂案所延後的是高風險人工智慧義務。然而，它並非僅僅進行延期，還另外新增了一項禁令：禁止在未獲得自由、具體且明確同意的情況下，擬真描繪實體人物私密身體部位或性行為而進行生成或操縱的人工智慧系統。這是一項針對所謂 nudifier 應用程式的條款，同時也涵蓋了兒童性剝削材料的相關風險。該禁令納入第5條，將於2026年12月2日生效。雖然備妥能可靠阻止此類產出的技術安全防護措施可透過避風港條款獲得免責，但適用對象不僅限於專為該用途設計的提供者，亦廣泛涵蓋可合理預見會產生此類產出的系統提供者。違規時處罰與禁止實踐同級，最高可處3,500萬歐元或全球營業額的7%。若忽略了這項延期修訂同時埋下了嚴厲新禁令的事實，就會犯下將此修訂案僅解讀為「鬆綁法規」的錯誤。

3.6 附錄III獨立型高風險人工智慧延期至2027年12月2日

高風險人工智慧主要分為兩大支柱。其一是附錄III所列舉的獨立型系統。這是指應用於招聘與人力資源管理、必要公共服務、生物辨識、移民與邊境管理、司法行政等八大領域，本身即具備完整功能的獨立人工智慧系統。第6條第2項為此分類的依據條文。

針對此類獨立型高風險系統，第16條（提供者）與第26條（部署者）的義務原本預定於生效24個月後的2026年8月2日起全面適用。Digital Omnibus 修訂案將此日期延後至2027年12月2日。從8月延至隔年12月，相當於推遲了16個月。有些報導寫17個月，但我親自計算了2026年8月與2027年12月之間的月份，確定是16個月無誤。

延期的背景在於尚未準備妥當的一方。企業需遵循的具體技術標準，即調和標準（Harmonised Standards，由 CEN-CENELEC JTC 21 制定之規格）尚未完全出爐，且各成員國成立監督本法之主管機關進度亦有所延誤。若在缺乏標準的情況下啟動義務，企業將面臨不知該符合何種標準卻可能遭受處罰的風險。將延期解讀為旨在降低此風險是合理且適當的。然而，務必記住裁處罰鍰的權限本身仍自2026年8月2日起生效。違反高風險義務的處罰上限為750萬歐元或全球營業額的1%，以較高者為準，而此上限本身是在2027年12月2日之後累積實際違規時才會開罰。如何將這一段空窗期轉化為準備期，將在第14章專章討論。

3.7 附錄I產品內建型高風險人工智慧延期至2028年8月2日

另一個支柱是歸類於附錄I、內建於既有產品中的高風險人工智慧。這適用於人工智慧作為安全元件嵌入至醫療器材、機械、玩具、航空器等已受歐盟產品安全法規管制的產品中，且該產品須接受第三方合格評估的情形。第6條第1項為此分類之依據。

此類產品內建型高風險人工智慧之法規程序完成時點，原本預定為生效36個月後的2027年8月2日。Digital Omnibus 修訂案將其再推遲12個月，移至2028年8月2日，比獨立型晚了八個月。這兩個日期有所差異的原因並非出於政治考量，而是與既有產品認證週期交織在一起的實務狀況。像醫療器材或汽車零組件這類原本就以數年為認證週期的產品群，需要更多時間將《人工智慧法》的要求整合至既有程序中。第8條第2項允許這種整合，使企業能在遵循既有安全法規的文件與程序上，疊加《人工智慧法》所要求的測試與報告，旨在減少重複作業。

釜山的醫療器材公司或韓國國內的汽車零組件公司即屬於此類。自家產品究竟屬於附錄I還是附錄III，將使準備期限相差八個月，且根據歸類不同，取得 CE 標誌（CE marking）程序的對口管道亦完全不同。我們將在第11章依具體案例深入剖析此一區別。

3.8 區分延期與未延期的界線

本章真正容易引發事故的地方就在這裡。許多人在聽到 Digital Omnibus 將多個日期往後延的訊息後，便誤以為整部《人工智慧法》都延期了。但在延期與未延期的事項之間，存在著一條明確的界線。

延期的是高風險人工智慧義務。附錄III獨立型延至2027年12月2日，附錄I產品內建型則延至2028年8月2日。

未延期的事項主要分為兩大類。第一類是第50條透明度義務的大部分內容。無論是聊天機器人的告知義務、合成內容標示，還是深偽技術的公開揭露，均維持在2026年8月2日不變。唯一的例外是針對在該日期前已經投入市場的生成式系統，給予截至12月2日為期四個月的寬限期。第二類是新設立的禁止生成未經同意之私密影像（NCII）及兒童性剝削材料之禁令。這並非被延期的條款，而是透過本次修訂首次啟動的義務，將於2026年12月2日生效。

因此，目前向歐洲輸出人工智慧系統或模型的韓國企業，在8月2日前急需調整的項目並非高風險規定，而是第50條。高風險方面爭取到了時間，但透明度方面時間緊迫。判斷的分水嶺就在這裡。我將這條界線作為決定準備順序的首要基準。我實際上見過好幾家公司因為反向解讀這條界線，將精力全數投入在應對高風險規定上，結果卻錯過了眼前的8月2日透明度標示要求。

施行時程總整理

2024年8月1日：《人工智慧法》生效。

2025年2月2日：第5條被禁止的實踐開始適用；第4條人工智慧素養義務開始適用。

2025年8月2日：通用人工智慧模型提供者義務（第53條、第55條）開始適用；治理規定開始適用；「無執法的適用」期間開始（至2026年8月2日）。

2026年6月29日：Digital Omnibus 修訂案獲理事會最終批准；等待刊登公報中（以截稿日為準）。

2026年8月2日：第50條透明度義務全面適用；人工智慧辦公室針對通用人工智慧模型違規行為之開罰權限啟動。

2026年12月2日：禁止生成 NCII 及兒童性剝削材料（第5條修訂案）生效；8月2日前投入市場之既有生成式系統標示義務寬限期結束。

2027年2月 2日：浮水印偵測可互操作性解決方案之具備期限（合成內容標示行為準則）。

2027年8月2日：2025年8月 2日前投入市場之既有通用人工智慧模型整備期限。

2027年12月 2日：附錄III獨立型高風險人工智慧義務（第16條、第26條）開始適用。

2028年8月2日：附錄I產品內建型高風險人工智慧法規程序完成時點。

這張表格是此時此刻的地圖。本書所涵蓋的主題仍在持續變化。在截稿之後，隨著公報刊登日期的確定、調和標準的陸續推出，以及新指引的發布，細部日期仍可能再次調整。我會持續將這些日期單獨列出並予以更新，也建議讀者在實際展開合規準備之前，再次核對這張表格。地圖雖然已經繪製完成，但地貌依然在微幅變動。

第二部

當前必須遵守的透明度義務

第4章 第 50 條，分為四個方向

幾年前，位於板橋的一家遊戲公司在客服管道導入了聊天機器人。起初這只是一個按幾個按鈕就會跳出預設答案的工具，但去年換成了對話型語言模型。如今，無論使用者問什麼，它都能像真人一樣回答。語句自然，甚至還能接梗開玩笑。當位於歐洲的使用者在凌晨詢問付款錯誤時，很難分辨螢幕另一端的存在究竟是真人還是機器。

這家公司一直以為自家的聊天機器人與監管無關，因為它並非高風險人工智慧。它既不決定錄用與否，也不進行信用評估，只是處理遊戲使用者的諮詢而已。然而，自 2026 年 8 月 2 日起，這個聊天機器人將受到歐盟人工智慧法案（Regulation (EU) 2024/1689，以下簡稱 AI 法案）第 50 條的規範。今天是 2026 年 7 月 24 日，剩下的時間只剩十幾天。

第 50 條所針對的不是危險的人工智慧，而是可能欺騙人類的人工智慧。這種區分是理解第 50 條的關鍵。高風險分類之所以成立，是因為該系統對人類的權利或安全有重大影響；第 50 條的透明度義務則不同，它的成立是因為該系統可能讓人將機器誤認為真人、將虛假誤認為真實。醫療器材判讀軟體屬於高風險，但不會欺騙人類；遊戲客服聊天機器人雖非高風險，卻有欺騙人類的可能。第 50 條管轄的就是後者。

因此，本章是許多韓國企業實際面臨的第一項義務。許多與高風險相關的規定，已在今年 5 月達成共識的所謂「數位綜合調整案」（Digital Omnibus）中延後實施。附錄 3 的高風險分類生效時間延至 2027 年，部分甚至延至 2028 年 8 月。除非是通用人工智慧模型提供者，否則那方面尚有寬限時間。然而，第 50 條在此次綜合調整案中並未被延後，仍於 2026 年 8 月 2 日照常適用。只有一個例外：第 50 條第 2 項的機器可讀標示，亦即浮水印相關義務，僅針對此前已上市的系統給予了為期四個月的寬限期，延長至 2026 年 12 月 2 日。其餘則全部自 8 月 2 日起生效。如果您的公司正在營運聊天機器人，或是運用人工智慧生成圖像、影片與文字輸往歐洲，第 50 條就是您最先撞上的牆。

第 50 條在內部又分為四個方向。依承擔義務的主體不同而區分，也依必須揭露的內容不同而區分。一家公司可能僅適用四者之一，也可能同時適用兩者或三者。若不清楚自己身處何種立場，就會在錯誤的地方浪費精力。

4.1 聊天機器人與對話型代理：表明非人類的方法

第 50 條第 1 項規範的是提供者（provider）的義務。提供者是指開發並推出旨在與人類直接互動的人工智慧系統的一方。聊天機器人、虛擬助理、自動語音應答皆屬於此類。條文內容簡短：必須告知該自然人，其正在與人工智慧系統進行互動。

條文看似簡短，實務卻非如此。問題出在例外條款。第 50 條第 1 項規定，若從合理通曉情況、謹慎且細心的自然人角度來看，屬於人工智慧已顯而易見者，得免予告知。這個「顯而易見」（obvious）正是陷阱所在。

企業很容易會有這種想法：我們的聊天機器人畫面上附有機器人圖示，名稱也取名為某某 Bot，任誰看都是人工智慧，因此符合例外條件。我不建議做此判斷。從執法的角度來看，主張例外的一方必須舉證其顯而易見性，一旦舉證失敗，即構成違反告知義務。歐盟執委會下設的人工智慧辦公室（AI Office）是以「非顯而易見」作為預設前提。要證明單憑一個圖示就能讓歐洲所有使用者意識到這是人工智慧，絕非易事。

最安全的做法是消除任何爭議空間。在對話開始的第一個畫面上，將「本諮詢由人工智慧進行」這句話顯眼地標示出來，關於是否顯而易見的爭辯便會蕩然無存。這項告知的代價僅為一句話；而主張例外卻敗訴的代價，則是依據第 101 條裁處的罰鍰。AI 法案第 101 條規定，對於違反第 50 條等義務者，最高可處以 1,500 萬歐元，或上一會計年度全球年營業額 3% 兩者中較高之金額。我認為這座天平顯然向其中一端傾斜。

告知的時間點也明確明定於條文中。第 50 條第 5 項規定，最遲應於首次互動時予以告知。在對話進行許久後才表明「其實我是人工智慧」，並無法滿足此一要件。告知方式亦有規定：必須明確且可資識別，並應遵循針對身心障礙使用者的無障礙需求。將文字隱藏在螢幕最下方的灰色小字中，極可能無法達到明確且可資識別的標準。人工智慧辦公室在審查此要件時，很有可能不會認為僅憑視覺訊號就已足夠。若為語音機器人，甚至可能要求同時具備聽覺訊號。

韓國企業經常忽略一個問題：聊天機器人並非自行開發，而是直接引進他人產品時的情況。如果是調用外國語言模型公司的應用程式介面（API）並搭載於自家服務中，第 50 條第 1 項的提供者是誰？開發系統並將其投放市場的一方即為提供者。若引進他人的模型並以自家名義打造聊天機器人服務投放市場，該聊天機器人系統的提供者就是該公司。

借用模型的事實並不能豁免其義務。

在此處，韓國企業還有一個容易忽略的情況。韓國國內自今年 1 月 22 日起實施的《人工智慧發展與信任基礎營造等相關基本法》（以下簡稱國內《AI基本法》），亦已包含類似的告知義務。以生成式人工智慧或高影響人工智慧推出服務的事業主，必須事先告知使用者該事實，並應於產出物上標示由人工智慧生成。單看結構，與第 50 條第 1 項、第 2 項如出一轍。因此我在進行韓國國內諮詢時，經常遇到客戶詢問：「我們已經完成國內法的因應，歐洲是不是照樣處理就可以了？」結構雖相似，份量卻截然不同。違反國內《AI基本法》的行政罰鍰上限為 3,000 萬韓元，且政府為了讓制度定錨，還設置了為期 1 年的宣導輔導期；而違反第 50 條的罰鍰上限為 1,500 萬歐元（折合韓元遠超過兩百億韓元）或全球營業額的 3%，並且完全沒有所謂的宣導輔導期。這正是絕不能將通過國內法合規的安心感直接套用到歐洲市場的原因。「雙重監管」這句話的意思並非義務產生了兩次，而是指必須用兩種不同的標準重新衡量同一件事。

4.2 合成內容機器可讀標示：浮水印發展到了何種地步

第 50 條第 2 項同樣是提供者的義務，但針對的對象不同。它是指生成合成語音、圖像、影片、文字的人工智慧系統提供者，其中也包含通用人工智慧模型。義務的內容是將其產出物以機器可讀的形式進行標示，使其人工生成或操弄的事實得以被偵測。

此處必須區分兩種標示。一種是人類肉眼可見的標示，例如「本影片由人工智慧製作」等字樣；另一種則是機器可讀的標示，這是嵌入檔案內部、肉眼無法看見但能透過驗證工具確認的標記。第 50 條第 2 項所要求的是後者，即機器可讀標示。肉眼可見的標示會在後文第 4 項「深度偽造公開」中再次登場。由於層級不同，提供生成式服務的公司往往需要兩者兼備。

條文以四個詞彙來描述機器可讀標示應具備的特性：有效、可相互操作、堅固且可靠。這四個詞在實務上的具體含義將於第 5 章結合浮水印技術詳細討論。在此僅強調：這種標示絕非裝飾。截圖一次、重新編碼一次就會被抹除的標記，無法達到「堅固」的標準。

過去幾個月來，這一領域的進展十分迅速。歐盟執委會於今年 6 月 10 日確定並發布了《人工智慧生成內容透明度實踐守則》（Code of Practice on Transparency of AI-Generated Content）。隨後於 7 月 8 日，執委會出具適當性評估意見，認定該守則足以支持第 50 條第 2 項、第 4 項及第 5 項義務的履行，並於次日（7 月 9 日）獲得人工智慧委員會（AI Board）的批准。欲簽署該守則的企業必須填寫簽署書並透過電子郵件提交至人工智慧辦公室，截止日期為 7 月 27 日晚間 6 點（中央歐洲時間）。撰寫本稿的今天是 7 月 24 日，對於正在考慮簽署的公司而言，只剩下三天時間。簽署者名單預計將在 8 月 2 日生效前公開。

實踐守則分為兩個層級。第 1 節探討推出生成式人工智慧系統之提供者的來源標示義務，第 2 節則探討引用該內容之部署者的標示義務。守則建議將浮水印、詮釋資料（Metadata）、指紋辨識等技術層層重疊使用，因為沒有任何單一技術能單獨達到完美。嵌入圖像中的浮水印可能因一張截圖而消失，詮釋資料則會在變更檔案格式的瞬間丟失。因此，守則將結合這三者的分層防禦法列為實務標準。

簽署雖屬自願，但我認為不簽署反而更加危險。人工智慧辦公室已預告將執法重點放在監督簽署者是否遵守守則上。未簽署的公司在缺乏守則這道安全網的情況下，必須僅憑

條文的抽象文字，自行證明其技術措施是否充分。而最終判定其適當性的一方，依然是人工智慧辦公室。

實踐守則具體要求重疊使用哪些技術，亦值得關注。其一是內容來源與真實性聯盟（C2PA, Coalition for Content Provenance and Authenticity）所創立的內容憑證（Content Credentials）方式。這是在檔案中嵌入經數位簽署的履歷資訊，記錄何時、由誰、使用何種工具製作，以及後續進行了何種修改。另一種則是肉眼不可見的浮水印。這是在像素或聲音訊號中刻下痕跡，使其即使經過壓縮或調整大小仍能一定程度地保留。問題在於這兩者各有缺點：內容憑證資訊會在變更檔案格式或截圖時整批遺失；肉眼不可見的浮水印雖能跨格式保留，但其本身無法說明究竟是由誰、於何時、用何種工具所製作。因此，守則建議將兩者疊加，並再加上可供比對原件的指紋辨識與存取紀錄，形成多層防禦結構。即便其中一層被突破，其餘層級仍能維護可偵測性。僅導入單一技術便自以為已盡到標示義務，據我判斷，並未達到該條文所要求的「堅固」程度。

第 2 項亦設有例外：僅提供標準編輯輔助功能的系統，以及為偵測與預防犯罪而經合法核准的系統。語法糾正工具即為前者的例子。若修飾語句的功能未實質改變內容，則無須將該產出物標示為人工智慧生成物。此項例外必須從嚴解釋。即使打著輔助編輯的名義，若實際上創造出了原本不存在的段落，便很難躲在「輔助功能」這面盾牌背後。

韓國企業極易觸法的領域是圖像與影片生成。利用生成式技術製作商品圖像或產出宣傳影片並投放至歐洲市場的行銷公司，或是將此類功能作為服務販售的新創企業，皆屬於此範疇。即便並非自行從頭訓練模型，而是引進他人的生成模型打造服務，只要將該服務的產出物投放至市場，便無法豁免於標示義務之外。

4.3 深度偽造公開義務與部署者的責任

在此處，義務的主體發生了變化。第 50 條第 4 項屬於部署者（deployer）的義務。部署者是指在其權限下實際使用人工智慧系統的一方，即非開發販售者，而是引進使用的一方。這種區分是第 50 條中劃分界線的重要標準。

第 4 項所針對的是深度偽造（Deepfake）。條文將深度偽造定義為：由人工智慧生成或操弄，模擬真實人物、物品、地點或其他實體或事件，使其看起來如同真實一般的圖像、語音或影片內容。部署者必須公開該內容係經人工生成或操弄的事實。這次要求的則是人類可識別的公開。若之前的第 2 項是檔案內部的機器標記，第 4 項則是向觀看者告知「這並非真實」的標示。

同為第 4 項，不僅適用於深度偽造，亦適用於為告知公眾利益事項而發布的人工智慧生成文字。若該文字經由人類進行實質審查並承擔出版編輯責任，則可免除此項公開義務。由人工智慧撰寫草稿、再由編輯核實事實並負責發布的新聞報導，即屬於此例外。若僅掛名編輯、實際上任由機器直接發布，該例外便不再成立。

我們將場景移至實際範例。首爾的一家廣告公司製作了一支瞄準歐洲市場的行銷宣傳影片。他們利用人工智慧合成了一位真實存在的知名演員臉孔，使其說出該演員未曾說過的話。這家公司並非製作該生成工具的提供者，工具是使用他人的產品；然而，他們是利用該工具製作深度偽造並將其公開發布的部署者。第 4 項的公開義務便落在了這家公司身上。

「部署者」這一身分對韓國企業而言格外重要的原因在於：許多韓國公司並非人工智慧的開發者，而是使用者。他們引進外國的生成模型製作內容，並將其發布給歐洲使用者。如此在第 50 條中填入該公司名稱的欄位便非提供者，而是部署者欄位。若僅將提供者義務視為他人之事，反而會忽略真正落在他自己身上的部署者義務。

先前介紹的實踐守則第 2 節，正是規範此項部署者義務。守則包含了如何標示深度偽造及公眾利益相關人工智慧生成文字的指南，涵蓋標示字樣與圖示的大小、位置及顯示時間等細節。附錄中亦載有三款可供選擇採用的歐盟通用圖示草案。與其從頭自行設計標示，不如直接引用這些圖示，更能減少日後因標示方式是否適當而引發爭議的情況。

第 4 項的例外將於第 5 節另行討論。藝術與諷刺的例外涵蓋於此，而該例外能延伸套用到何種程度，正是實務上的爭議焦點。

4.4 情緒識別與生物辨識分類系統的告知義務

第 50 條第 3 項再次規範部署者的義務。標的為情緒識別系統與生物辨識分類系統。營運此類系統的部署者，必須向暴露於該系統下的人員告知該事實。條文更明確規定，若涉及處理個人資料，應一併遵守《通用資料保護規則》（GDPR）及相關歐盟個人資料法規。

情緒識別系統是指試圖從人類的面部表情、聲音、肢體動作中解讀情緒狀態的系統。生物辨識分類系統則是指基於面部或指紋等生物特徵資料將人類劃分為特定群體的系統。客服中心在諮詢過程中透過顧客語調判斷其是否發怒的工具，或是門市攝影機估算訪客年齡層與性別以切換廣告的工具，皆可能涵蓋於此。

在此切勿混淆兩項規定。第 50 條第 3 項是「告知義務」，告知後即可使用；然而在情緒識別中，於職場與教育機構內使用則是「完全被禁止的行為」，觸及第 5 條。這並非「告知後使用」的問題，而是「根本不得使用」的問題。該禁令並非出自第 50 條，而是在後續章節中討論。韓國合法使用的員工情緒分析工具，在歐盟卻可能直接跨越禁止紅線，原因即在於此。目前只需區分第 50 條第 3 項的告知義務與第 5 條的禁止規定屬於不同條文即可。

第 3 項亦設有出於合法偵測犯罪目的之例外。然而該例外係以執法機關為前提，進軍歐洲的韓國一般企業幾乎無援引空間。此外，該例外還附帶一項條件：必須設置維護第三人權利與自由的保護機制，並遵守相關歐盟法規。即便主張例外，亦無法徹底免除個人資料保護義務。

這也是建議同時進行 DPIA（個人資料影響評估）與 FRIA（基本權利影響評估）的原因。滿足了告知義務，並不代表該系統自動遵守了個人資料處理原則。這兩項評估回答的是不同的問題：FRIA 詢問的是該系統對人類基本權利影響的程度，DPIA 則詢問在此過程中如何處理個人資料。越是像情緒識別與生物辨識分類這般將人類身心作為資料處理的系統，在實務上越適合將兩項評估放在單一程序中共同推進，而非分開處理。

第 50 條第 3 項並未另設寬限期。無法期待像第 2 項浮水印義務那般擁有四個月的緩衝期。該條項自 8 月 2 日起直接適用。若在歐洲營運含有情緒識別或生物辨識分類元素的服務，必須優先確認此一條款。

不少韓國公司試圖將在韓國開發的顧客諮詢分析工具直接移轉至歐洲法人的客服中心使用，其中用於客服人員培訓的聲音情緒分析功能即為典型代表。在韓國國內僅被視為諮詢品質管理工具的功能，一旦針對歐洲客服人員與歐洲顧客運作，便立即成為第 50 條第 3 項規定的部署者告知對象。若對客服人員使用，必須一併檢討是否觸及第 5 條「職場內情緒識別禁令」的風險；若對顧客使用，則必須滿足第 3 項的告知義務。即使是相同的功能，依據對象的不同，所適用的條文亦有所差異，這一點我在實務諮詢中格外頻繁地重複說明。

4.5 藝術、諷刺與安全目的之限制性例外能援引至何種程度

首次閱讀第 50 條的公司，最先尋找的往往是例外條款。以為「因為我們是藝術」、「因為我們是諷刺」，是不是就能得以豁免？我對這種期待抱持審慎態度。例外確實存在，然而條文所劃定的例外範圍遠比企業所企望的要來得狹窄。

先來看第 4 項關於深度偽造的例外。當內容屬於藝術、創作、諷刺或虛構作品的一部分時，公開義務並不會消失，只是公開方式得以放寬。只要以不干擾作品欣賞的適當方式予以揭露即可。例如在電影片尾名單中標註「本場景使用了人工智慧合成」，即屬於此類方式。這意味著無須在畫面中央持續顯示巨大警告字樣而導致無法觀賞電影。例外的本質在此顯現：這並非「無須揭露」的例外，而是「可以以較不干擾的方式揭露」的例外。

這項差異在實務上影響重大。若公司解讀為「我們是藝術，所以無須做任何標示」，那便大錯特錯。依然必須揭露，僅是在不干擾欣賞的前提下揭露即可。何者為藝術、何者為商業廣告，界線上必然會產生爭執。品牌宣傳影片即便帶有藝術元素，是否就能歸類為第 4 項所指的藝術作品，尚屬未知。我認為對於廣告目的明確的內容，與其依賴藝術例外，不如正式進行公開更為安全。前述實踐守則已將標籤的大小、位置乃至顯示時間進行了具體化，以該守則為基準會比主觀斷定更為妥當。

出於刑事犯罪偵測、預防、偵查與起訴目的之例外，亦散見於第 50 條各處。包含第 1 項聊天機器人告知、第 2 項機器可讀標示、第 3 項情緒識別與生物辨識分類告知、第 4 項深度偽造公開，這四項義務皆附帶經法律核准之執法目的例外。該例外係為維護公共安全這一明確公益所採取的措施。然而第 3 項亦明定在此情況下仍須嚴格遵守個人資料保護法規。就一般企業而言，認定幾乎無空間將此例外套用於自身業務，才是較為妥當的理解。

舉證責任問題貫穿了整項例外規定。欲適用例外的一方，必須自行舉證其系統符合該例外的要件。人工智慧辦公室一方面承認具有法律約束力的最終解釋權屬於歐洲法院，另一方面亦表明將針對個案單獨進行判斷。在執法案例尚未大量累積的當下，必須記住：例外的邊界越模糊，承擔該風險的一方就是主張例外的公司。

一目瞭然掌握第 50 條的對照表

將義務主體與內容如此區分後，便能輕鬆找出自家公司身處哪一個欄位。

第 50 條第 1 項。主體為提供者。與人類互動的人工智慧（聊天機器人等）須向使用者告知其為人工智慧。例外為屬於人工智慧顯而易見者，舉證責任由提供者承擔。適用日期為 2026 年 8 月 2 日，無寬限期。

第 50 條第 2 項。主體為提供者。合成語音、圖像、影片、文字之生成物須以機器可讀形式標示。例外為標準編輯輔助、合法犯罪預防。適用日期為 2026 年 8 月 2 日，既有系統寬限至 2026 年 12 月 2 日。

第 50 條第 3 項。主體為部署者。須向暴露於情緒識別、生物辨識分類系統下的人員告知，並一併遵守 GDPR 等個人資料法規。職場與教育機構內之情緒識別於第 5 條另行禁止。適用日期為 2026 年 8 月 2 日，無寬限期。

第 50 條第 4 項。主體為部署者。深度偽造須以人類可識別方式公開。公眾利益相關文字亦須公開，若由人類承擔編輯責任則例外。藝術與諷刺應以不干擾欣賞的方式公開。適用日期為 2026 年 8 月 2 日。

第 50 條第 5 項。告知時間點最遲為首次互動或暴露之時。方式應明確且可資識別，並遵守無障礙需求。

第 50 條第 7 項。人工智慧辦公室鼓勵制定實踐守則。《人工智慧生成內容透明度實踐守則》於 2026 年 6 月 10 日確定，7 月 9 日獲得人工智慧委員會批准，簽署截止時間為 7 月 27 日晚間 6 點（中央歐洲時間）。

自我診斷五大要點

貴公司是否正在營運與歐洲使用者對話的聊天機器人或語音機器人？若是，適用第 1 項。請確認第一個畫面是否有告知字樣。

貴公司是否利用人工智慧製作圖像、影片、文字並輸往歐洲？若屬於販售製作工具的一方，適用第 2 項機器可讀標示；若屬於利用該工具製作內容並輸出的一方，適用第 4 項深度偽造公開。兩者皆有可能適用。

貴公司是否在歐洲使用透過人類情緒或生物特徵資料進行判別的系統？若是，適用第 3 項告知；若用於職場或教育機構，請一併確認是否觸及第 5 條禁令。

貴公司製作的內容中，是否有以藝術、諷刺、創作為由而省略公開者？若有，請重新檢視該判斷是否留有書面紀錄，以及是否摻雜了廣告目的。

貴公司是否尚未決定是否簽署實踐守則？7 月 27 日晚間 6 點（中央歐洲時間）為截止時間。若決定不簽署，則必須另行準備採取何種技術措施以滿足第 2 項與第 4 項的要求。

最後以一個日期作為結尾。第 50 條將於 2026 年 8 月 2 日生效適用。對於在此日期之前已上市運作的生成式系統，其浮水印義務享有至 2026 年 12 月 2 日為止四個月的緩衝期；然而，對於該日期之後推出的新系統則沒有緩衝期，且第 1 項、第 3 項與第 4 項的告知及公開義務，從一開始就根本沒有所謂的寬限期。板橋的那家遊戲公司在上周的會議中，通過了在聊天機器人首頁畫面加入一行標示字樣的案子。開發團隊表示半天就能搞定。真正耗費時間的並非畫面設計，而是向管理階層說明我們為何屬於該條款適用對象的溝通會議。若您目前正向歐洲輸出任何內容，請算一算日曆上記至 8 月 2 日還剩下多少天。

第5章 水印的技術與法律

這是在板橋（Pangyo）一家圖像生成新創公司發生的真實案例。在針對投資人的 Demo 簡報場合上，團隊負責人輸入 Prompt 並生成了一張人物照片。畫面角落附帶了一個微小的 AI 標示圖示，若開啟檔案屬性，裡面還包含說明這張照片是在何時、由何種模型生成的數位簽署資訊。到這裡為止，一切都如預期般順利。問題出在接下來發生的事。投資人直接將該畫面截圖並上傳到自己的 Slack。當打開該截圖檔案時，畫面上看到的圖示雖然還在，但植入檔案內部的簽署資訊卻整組消失了。截圖這個極為平凡的操作，竟將精心附著的來源資訊一舉抹去。這個場景濃縮展現了第5章所要探討的核心問題：法律要求在 AI 生成的內容上留下標示，但該標示能否禁得起網路上內容常經歷的種種處理，完全是另一回事。

5.1 現有系統寬限至2026年12月2日，與新推出系統之無寬限期

歐盟《人工智慧法案》（Regulation (EU) 2024/1689，以下簡稱《AI法案》）於2024年8月1日生效，規範透明度義務的第50條則於生效24個月後的2026年8月2日起全面適用。第50條分為四款，各自針對 different 的對象。第1款要求如聊天機器人般與人類對話的 AI 系統，必須告知使用者正與 AI 進行互動。第2款要求生成合成語音、圖像、影片、文本之 AI 系統的提供者，須以機器可讀格式標示其產出物並使其可被偵測。第3款要求情緒識別或生物辨識分類系統的部署者，須向接觸該系統的人員告知該事實。第4款要求處理如 Deepfake 等經操縱之圖像、音訊、視訊的部署者，須公開其已被操縱的事實。

在此，直接影響實務運作的變化在於：經歐盟共同立法者達成共識的《數位綜合法案》（Digital Omnibus）修訂案中，僅針對第2款附帶了單獨的過渡措施。該修訂案於2026年5月7日在歐洲議會與理事會之間達成政治共識，並於2026年7月8日完成簽署。不過，以撰寫本稿的時間點為準，該案尚未刊登於《官方公報》（Official Journal）。但尚未刊登於公報並不意味著可以忽視；將立法者之間已達成共識並完成簽署的事項視為毫無依據，反而是更危險的判斷。該過渡措施的內容如下：若為2026年8月2日前已在歐盟市場推出或投入服務的生成式 AI 系統，只需在2026年12月2日前具備第50條第2款所要求的標示與偵測義務即可，等於獲得了4個月的緩衝寬限期。此外，在此寬限期之前已經問世的產出物，亦即在8月2日前製作的成果（包含 Deepfake 在內），並無須溯及既往進行標示。

相反地，新推出的系統則沒有這份餘裕。2026年8月2日之後首次在歐盟市場推出或投入服務的 AI 系統（包含生成式 AI 在內），自推出起即須具備第50條的所有要求。第2款的寬限僅適用於現有系統，且嚴格限於8月2日前已在市場上的系統，屬於狹隘的例外條款。第1款、第3款與第4款無論是現有或新推出系統，均無例外地自8月2日起適用。歐盟 AI 辦公室（EU AI Office）所闡明的原則亦同：凡屬第50條規範的所有 AI 系統，無論其最初於何時推出，只要自8月2日起在歐盟市場推出或投入服務，即必須遵守規範。在我看來，此條款結構實質上構成了對新進入者的懲罰 - - 給予已在市場立足的業者額外4個月時間，卻要求剛要進入市場的業者從第一天起就必須呈交完整合規的型態。

對總部位於首爾的企業而言，這種區分意味著準備負擔將因推出日期的設定而大相逕庭。若目前已有在歐盟營運中的生成式 AI 服務，只需在12月2日前完成水印與中繼資料（

Metadata）體系的建置；但若是8月2日之後才要在歐盟首次推出的新服務，就必須從設計階段起便完成標示功能後再行推出。若將兩者放在同一時程表上準備，新服務那一方勢必會延誤。根據第99條，違反規定時的行政罰鍰為上一財務年度全球年營業額的3%或1,500萬歐元（以較高者為準）。自8月2日起，各成員國的市場監督機關與 AI 辦公室將開始實質行使這項罰鍰處分權。與其因為獲得寬限而放鬆警惕，不如將寬限的這4個月視為用於實際技術落實的時間，這樣做更為妥當安全。

5.2 C2PA 標準與內容來源認證

《AI法案》第50條第2款本身隻字未提 C2PA 這個名稱。條文所要求的僅是以機器可讀格式標示產出物並使其可被偵測，並未指定該標示應透過何種技術實作。填補這一空白的是 AI 辦公室於2026年6月10日發布的《AI生成內容透明度實踐守則》（Code of Practice on Transparency of AI-Generated Content）。該守則將標示技術分為水印（Watermarking）、中繼資料（Metadata）與指紋識別（Fingerprinting）三種方式，並將條文要求「有效、具可互操作性、堅固頑強且可信」具體化為細部執行項目。細部措施 1.1.1 要求附帶數位簽署的中繼資料，1.1.2 則要求隱形水印。同時滿足這兩項條件的技術實務上僅有一種：即由內容來源與真實性聯盟（C2PA）所建構的「內容憑證」（Content Credentials）。

內容憑證實際執行的工作並不複雜。在內容生成的當下，系統會將使用何種工具、生成時間、經過何種編輯等資訊組成資料包，進行加密簽署後附著於檔案中。這個經簽署的資訊包被稱為「Manifest」（清冊）。內容每經過一次不同的工具處理，新的 Manifest 就會指向前一個 Manifest 並串接上去，結果使得單一檔案從生成到散布的歷程如鏈條般堆疊累積。若在這條鏈條中間夾雜了任何一個不驗證簽章的工具，歷程就會在此中斷，後續階段將處於無法證明來源的狀態。

法條中未寫明名稱，與在實務上事實上已凝固為單一標準，是兩件不同的事。實踐守則屬於自願性簽署的守則，並不具備法律約束力。然而，只要簽署該守則並按其內容執行，即成為證明符合第50條第2款規定的有力依據。結合第40條規定的「協調標準」（Harmonised Standards）制度來看，這種架構會更加清晰。遵循協調標準的高風險人工智慧系統或通用人工智慧模型，可享有被推定為符合相關要求的「符合性推定」（Presumption of Conformity）優惠。雖然 C2PA 尚未被正式指定為具備該地位的協調標準，但僅憑 AI 辦公室在實踐守則中事實上指明的技術路徑這一點，在監管機關面前拿什麼作為合規證據便已然確定。AI 辦公室也在同一天推出了標準圖示集。附著於由 AI 生成或編修之內容上的固定樣式 AI 標示是供人類觀看的一層，而 C2PA 內容憑證則是其下方由機器進行驗證的一層。

韓國企業在此需要掌握的實務重點有兩項。其一是雖然可以使用替代技術，但企業必須自行證明該技術具備與 C2PA 同等水準的可偵測性與開放存取性。守則未強制指定特定

技術，並不代表任何方式都能過關。其二是採用 C2PA 並非僅局限於圖像或影片生成管道（Pipeline）的問題。內容憑證的結構是在生成、編輯與散布等各階段串接數位簽章，因此若公司內部系統的任何一個階段斷鏈，最終產出物中的來源資訊就會整組遺失。我認為將此問題單純推給技術團隊處理極具風險；因為當法務部門需要向監管機關說明合規情況的當下，必須能夠從技術層面上指明簽章鏈究竟是在何處中斷。

5.3 堅固性要件：能否經受重新編碼與螢幕截圖？

第50條第2款規定標示技術必須具備有效性、可互操作性、堅固性（Robustness）與可信度。在這四個形容詞中，實務上尤以「堅固性」最難處理。因為這要求附著於內容上的標示，即使經過壓縮、變更解析度、重新上傳至其他平台，甚至是直接對螢幕進行截圖，都必須能夠留存下來。板橋新創公司發生的那一幕正好展現了這個問題。畫面上看到的圖示並非 directly 繪製於影像本身，而是檢視器讀取中繼資料後疊加顯示的標示；在截圖的瞬間，原始檔案的中繼資料便斷絕，僅留下畫面上擷取的像素。單靠中繼資料方式，標示在此類情況下便會消失。

與此相反，將訊號直接刻印於內容像素本身的隱形水印方式情況則有所不同。像 Google SynthID 這類技術，即使面對一般的重新編碼、輕微編輯，甚至像螢幕截圖這般中繼資料整組丟失的情況，留在影像本身的訊號依然能大幅度維持。這是因為訊號存在於內容內部而非中繼資料中。問題在於企圖有心抹去標示的對抗者。透過在影像中加入雜訊干擾訊號後再以生成式 AI 重新繪製的所謂「再生成攻擊」（Regeneration Attack），或是學習水印本身特徵進而將其剝離的「水印竊取攻擊」（Watermark Stealing Attack），都能在無肉眼可見畫質受損的情況下大幅降低偵測可信度。文字水印亦然，單憑重寫文句即可使訊號變得模糊。因此，在當前時間點，水印是一項提高「使 AI 生成內容變得不可追蹤」之成本的機制，而非能從根本上阻斷抹除的萬能鎖。

螢幕截圖在這場堅固性爭議中佔據了相當特殊的地位。截圖既非將檔案重新編碼，亦非降低畫質的編輯動作。這只是將螢幕上顯示的像素原樣重新拍攝的行為，因此附著於原始檔案上的中繼資料、簽章與 Manifest 都會剝落得乾乾淨淨。如內容憑證這般將資訊附著於檔案旁的方式在此刻會徹底失效。相反地，將訊號刻印於圖像像素陣列本身的 watermark 技術，即使被原樣重拍，該訊號也會隨像素一同轉移，因而具備存活的空間。我認為不應將這兩種技術視為競爭關係，而應視為互相補足漏洞的關係。中繼資料簽章能留下詳細的歷程，但經不起一次截圖；像素水印能抵禦截圖，卻無法告知是由誰在何時、製作了何物。唯有同時植入兩種技術，方能在其中一方被破譯或抹除時，另一方仍能留下最低限度的證據。

第50條第2款規定，在判斷堅固性時，應一併考量各類內容特點與限制、實作成本，以及當時的最新技術水準（State of the Art）。然而，究竟要抵禦重新編碼或螢幕截圖到

何種程度才能獲認定為具備堅固性，目前尚無定量標準。預計隨著 AI 辦公室的指引、協調標準以及各成員國執法案例的累積，該標準將逐漸明確。在當前時間點唯一能確定的是，「最新技術水準」這項標準具有隨時間推移而提高要求的特性。今天通過驗證的水印技術，無法保證在一年後仍被認定為具備堅固性。韓國企業若採取「在內容生成引擎中植入一次標示功能即大功告成」的態度，恐將面臨在下一個監管週期被評估為技術落後的風險。我建議應將此項要求視為需持續更新的研發（R&D）項目並編列於預算中，而非一次性實作。

5.4 禁止生成 NCII 與 CSAM：第5條修訂與2026年12月2日生效

第5條是直到最近才具備此一內容的條文。在此修訂前，第5條僅為禁止操縱性、剝削性及社會信用評分實踐的條款，並未明確提及非自願性性私密影像（NCII）或兒童性虐待素材（CSAM）。填補該空白的是歐洲議會與理事會於2026年5月7日達成政治共識的《AI數位綜合法案》（Digital Omnibus on AI）套案。該共識中包含將所謂的「去衣」（Nudifier）應用程式 - - 即未經同意脫去他人衣服或將其操縱為性化圖像、影片、音訊的 AI 系統，以及生成兒童性虐待素材的 AI 系統 - - 列入第5條禁止實踐中的內容。該套案已於2026年7月8日完成簽署，目前僅剩刊登於官方公報。

修訂後的第5條禁止在歐盟市場推出或使用此類系統的行為本身。適用範圍並不局限於最初即專為該用途設計的系統。若產生此類產出物屬於合理可預見的結果，即使提供者並無該意圖，亦可能落入監管範圍。不過，法案亦同時設有安全港（Safe Harbour）條款。若系統配備了能可靠阻絕此類產出物的有效技術安全防護機制，即可排除於此項禁止之外。因此，營運圖像生成服務的企業，必須能夠自行證明用於過濾系統防止生成該類內容的機制具備何等程度的可信度，方能脫離該條款的管轄射程。安全港所要求的可信度標準目前尚未透過判例或指引確定。因此，我建議比起事後編製過濾性能報告，不如從開發階段起即留存測試紀錄（Test Logs）以及誤判與漏判數據。因為在監管機關審查是否認定安全港資格時，能作為依據提出的資料歸根究底正是這些紀錄。施行日期與5.1節討論的第50條第2款寬限期結束日同為2026年12月2日。日期重疊並不代表兩者是同一件事。第50條第2款是要求現有生成式 AI 完成標示義務之寬限期的終結，而修訂後的第5條則是絕對禁止特定產出物本身之新規定的開始。一個是寬限結束的日子，另一個則是禁令開始的日子，只是偶然集中於同一日期而已。

違反法律的嚴重程度亦截然不同。如前所述，違反第50條依據第99條，罰鍰上限為營業額的3%或1,500萬歐元。相反地，違反第5條的禁止實踐，罰鍰上限則為營業額的7%或3,500萬歐元（以較高者為準），屬於《AI法案》所訂罰鍰體系中最高的層級。此外還疊加了歐盟成員國的刑事法。生成 NCII 或 CSAM 即使在沒有《AI法案》的情況下，在多個成員國早已屬於刑事處罰對象；《AI法案》第5條的修訂，更接近於將這項既有的禁令明確釘定於 AI 系統這一新工具上。若是在韓國開發圖像或影片生成服務並提供給歐盟使用者的公司，絕非將內容過濾列為事業後順位課題的時候。經營團隊必須親自認知到：一個突破過濾機制的產出物，可能直接衝擊公司整體營業額的7%。

參考資料

日期與義務對照表整理如下：2024年8月1日，《AI法案》生效。2025年2月2日，第5條禁止實踐全面施行。2026年8月2日，第50條全面適用日；自當日起，新推出的系統無一例外須遵守第50條第1款至第4款。現有系統的第1款、第3款與第4款亦自當日起遵守。2026年12月2日，兩項事件重疊：其一是2026年8月2日前推出的現有生成式 AI 系統第50條第2款寬限期結束；其二是依據修訂後第5條對去衣及 CSAM 生成系統之禁令開始施行。

判斷標準可歸納為三個問題：我們的系統於何時在歐盟市場推出或預計推出？以8月2日為基準，之前或之後決定了是否適用寬限。我們的標示技術是僅依賴中繼資料，抑或將訊號植入內容本身？抵禦截圖與重新編碼的程度由此分野。我們的生成系統是否配備過濾性私密影像或兒童相關內容的過濾機制，且能以文件證明其可信度？這決定了是否能適用第5條的安全港條款。

在此再次列出罰鍰區間：依據第99條的一般違規，為營業額的3%或1,500萬歐元。違反第5條的禁止實踐，為營業額的7%或3,500萬歐元。中小企業（SME）在兩個區間中均適用較低一方的上限。

在投資人截圖並上傳至 Slack 的那張照片上，至今仍只留下一個肉眼可見的微小圖示。原本位於其下的簽章已然消失，如今再也沒有人能夠證明該檔案出自哪個模型的哪條 Prompt。法律所要求的「堅固性」一詞，歸根究底，指的就是禁得起這一刻考驗的能力。

第6章 透明度實踐守則

設址於聖水洞的一家圖像與影片生成服務新創公司法務團隊，在2026年7月22日下午6點 - - 距離中歐時間截止期限僅剩一天時召開了會議。他們必須決定，究竟是現在就簽署在布魯塞爾確定的《透明度實踐守則》（Code of Practice），還是先觀望。由於該公司開發的影片生成服務也有歐洲使用者在使用，因此本就無法逃避歐盟《人工智慧法》（Regulation (EU) 2024/1689，以下簡稱《AI法》）第50條所規定的標示義務。只不過，究竟要以何種方式說明自己已履行該義務，以及說明材料要從何處取得，則是公司可以自行選擇的問題。

本章所探討的實踐守則（Code of Practice），雖然與第5章所見的通用人工智慧模型（GPAI model）實踐守則名稱相同，但兩者源頭不同。由《AI法》第56條所產生的GPAI實踐守則，針對的是少數大型模型提供者；相對地，第50條第7項預定的透明度實踐守則，則面向範圍廣泛得多的企業群體，從製作深偽（Deepfake）影片的新創公司、附設聊天機器人的客服中心，到批量生成合成圖像的廣告代理商皆涵蓋在內。歐盟執委會於2026年6月10日的閉幕全體會議（Closing Plenary）上，確定了這份透明度實踐守則的最終版本。在持續近九個月的多方協商過程中，來自產業界、學術界、公民社會、著作權人團體及會員國政府的參與者超過187人，並由AI辦公室（AI Office）委任的6位獨立專家領導草案起草工作。

有一點必須先予以釐清：第50條規定的透明度義務，並非僅靠這份實踐守則就能全部涵蓋。第50條第1項要求像聊天機器人這樣直接與人互動的人工智慧系統提供者，在設計上必須讓使用者能知悉自己正在與人工智慧對話；第50條第3項則要求使用情緒識別系統或生物辨識分類系統的部署者（deployer），必須向受影響的人告知該事實。第50條第7項將實踐守則列為對象的，僅有其中的偵測與標籤化義務，亦即第50條第2項與第50條第4項。因此，營運聊天機器人的公司或使用情緒識別功能的公司，即使簽署了本次實踐守則，這兩條條文的義務也不會自動獲得解決，這意味著依然需要各自進行額外的告知設計。

6.1 簽署的實質利益與舉證責任的轉移

透明度實踐守則涵蓋的範圍分為兩項：即《AI法》第50條第2項的標記與偵測義務，以及第50條第4項的標籤化義務。第50條第2項規定，生成合成語音、圖像、影片、文字的人工智慧系統提供者，有義務確保其產出物以機器可讀的方式予以標示，並能被偵測出係經人工生成。第50條第4項則要求製作或操縱深偽內容的系統，以及生成涉及公眾利益事項文字的系統部署者（deployer），向使用者告知該結果物係人工製作而成。

這兩項義務無論是否簽署，都將於2026年8月2日起直接適用。實踐守則所改變的並非義務的存在本身，而是「如何展現已遵守該義務」的方式。簽署企業無需從頭解釋自家公司為何認為當前措施已然足夠，而是可以將實踐守則所設定的項目 - - 例如採用了何種浮水印標準、附加了何種元資料（metadata）體系 - - 作為檢查清單（checklist）逐項證明即可。對會員國的市場監督機關而言，審查對象也變得十分明確。畢竟，直接認定「該企業遵循了歐盟經多方協商建立的標準框架」，比起每次都從頭剖析個別企業的自家方法論要容易得多。我認為這正是簽署實踐守則的真正實質利益所在。這並非產生了新條文，而是舉證的重心從企業轉移到了守則上，這也就是本章標題所指的「舉證責任的轉移」。

此外，在實務層面亦有收益。對於聲明遵守守則的企業，市場監督機關極有可能將監督力量集中於確認其是否遵循守則上。這意味著，原本可能投入未簽署企業的個別調查資源，在簽署企業身上會使用得相對較少。然而，有一點必須明確：與GPAI實踐守則不同，法律條文中並未明定「遵守透明度實踐守則」可作為違反《AI法》時減免罰鍰的減輕因素。《AI法》第101條規定的罰鍰減輕條款僅適用於GPAI模型提供者，而對違反第50條的處罰，則適用第99條規定的一般罰鍰體系，即全球總營業額的3%或1,500萬歐元中較高者。遵守實踐守則在該罰鍰計算中會占有多大權重，精確來說，是交由會員國監督機關的裁量權決定。

簽署程序本身也與GPAI實踐守則有所差異。GPAI實踐守則必須經過歐盟執委會進行適當性評估並予以批准的程序後，方能生效。而透明度實踐守則則是公布多方協商結果並徵求簽署的方式，因此批准程序的份量相對較輕。2026年7月22日的截止期限，是列名首批簽署者名單的窗口，在此之後極有可能繼續接受新簽署的企業。鑑於GPAI實踐守則在首次公開後亦有多次追加簽署者的先例，我預計本次也會採取類似走勢。不過，這項

預測也可能不準確，因此若有公司擬於事後補簽，應隨時確認AI辦公室的公告。

6.2 遵循與法律推定之關係

在此必須釐清一項關鍵區別：簽署並遵循透明度實踐守則，並不意味著會產生《AI法》所指的符合性推定（presumption of conformity）。符合性推定是僅賦予《AI法》第40條所規定的調和標準（harmonised standard）的優惠。若遵循由歐洲標準化組織（CEN、CENELEC）制定並刊登於《歐盟官方公報》的調和標準，該標準所涵蓋的要件即被推定為已遵守法律。雖然這是可反駁的推定，但舉證責任本身會轉移至監管機關，具有強大效果。透明度實踐守則並未能達到此一水準。

與GPAI實踐守則相比，這項差異更為顯著。GPAI實踐守則在《AI法》第53條第4項與第55條第2項條文本身中，即明定「遵循實踐守則」得作為履行義務的證明手段。這等同於條文直接指定可以將實踐守則用於此用途。反觀第50條第7項，僅規定AI辦公室應鼓勵並促進實踐守則的制定，並未在條文內建立「若遵守該實踐守則即視為已證明履行義務」的鏈結。因此，若要排列層級，順序如下：最頂層是調和標準的完整符合性推定；下一階是條文明確認可的GPAI實踐守則證明路徑；而透明度實踐守則則在此之下，缺乏條文依據，僅靠實務慣例發揮影響力。我認為絕不能混淆這一層級關係。如果將簽署透明度實踐守則誤以為是與GPAI實踐守則相同的法律盾牌，恐怕會有企業在日後會員國監督機關表示「僅憑簽署尚不足夠」時感到措手不及。

更精確地說，透明度實踐守則目前所具備的份量，更接近於展示「現有技術水準」（state of the art）的有力情況證據（circumstantial evidence）。由於這是由187位以上的利害關係人與6位獨立專家達成共識的產物，確實足以作為主張「此類措施已達業界通常水準」的依據。然而，其本身並非觸發法定推定效果的鑰匙。補充一點，截至撰寫本稿時為止，關於第50條所定義務的調和標準尚未公布。因此，就目前而言，透明度實踐守則實際上扮演著唯一的實務基準點角色；待調和標準出爐後，屆時的層級順序可能會再度產生變化。

執法窗口亦有所不同。GPAI實踐守則由AI辦公室自2026年8月2日起直接行使處以罰鍰的權限並進行監督。而與透明度實踐守則相關的第50條義務，則由會員國的市場監督機關執行。這意味著韓國企業所面對的對象並非布魯塞爾的AI辦公室，而是自家服務所流通的個別會員國監督機關；因此，實務應對窗口必須以該國語言、配合該國程序做好準備。

即便在會員國內部，窗口也並未統一歸口。若深偽影片涉及政治或選舉，廣播與媒體監管機構可能會一併介入審查；若屬於包含個人資料的合成內容，資料保護主管機關（DPA）則可能另行關注。我認為，企業若以為只需應對一家市場監督機關，可能會遇到同時收到多家機構質詢的情況。此外，用於闡明第50條範圍與適用對象的解釋指引（guidelines），在撰寫本文時仍停留在草案階段。這是實務工具（實踐守則）先出爐、解釋指引後跟進的反常順序。待指引公布後，對於實踐守則未涵蓋的空白之處 - - 例如對藝術性或諷刺性表達物的例外認定制於何種程度 - - 預計將能獲得更為明確的解答。

6.3 未簽署企業的替代路徑

選擇不簽署當然 collegiate 也是可能的選項，因為實踐守則本就出於自願。然而，第50條所規定的標記、偵測與標籤化義務本身，無論是否簽署都會原封不動地保留。未簽署企業必須自行說明並以書面文件記錄，為何其所採用的浮水印方式或部署者告知文案，達到了法律所要求的標準 - - 亦即有效（effective）、具可互操作性（interoperable）、堅固（robust）且可靠（reliable）的措施。在實踐守則已如業界標準般確立的情況下，這項文件化工作若完全不參考守則條文，反而更加困難。我認為，即便是選擇這條替代路徑的企業，實際上也大多會把實踐守則放在手邊，用來比對評估自家的措施。雖然這並非法律義務，但在實務上卻充當了指南針的角色。

那麼，未簽署企業實際上掌握的籌碼有哪些？我歸納出三項：第一，自行引進標記內容來源的元資料或浮水印技術，並以技術報告的形式記錄該選擇為何具有有效性與堅固性；第二，由外部審計員或技術專家評估自家措施的妥當性，以取得獨立的意見書；第三，在內部規定或使用條款中將標示程序予以明文化，藉此展示該措施並非出於偶然，而是在設計階段即已納入的結果。這三者比起簽署實踐守則並逐項對照檢查清單，都需要耗費更多工夫。然而，對於自身業務結構與實踐守則標準項目不太契合的公司 - - 例如製作特殊工業用圖像生成工具的公司 - - 這反而可能是更為坦誠的選擇。

時程規劃亦不可忽視。第50條義務原則上自2026年8月2日起全面適用；但根據2026年5月達成暫定協議的《AI綜合修正案》（AI Omnibus）之修訂，在此時間點前已上市的生成式AI系統，僅就機器可讀標記要件而言，可獲得延後至2026年12月2日的寬限期。至於可互操作性要件，則要到更晚的2027年2月2日才全面適用。若為必須對已有服務中的產品進行後續修改的公司，在該寬限期內便稍微留有些許完善措施的裕度。

另一點值得留意的是，目前正與實踐守則分別準備的第50條解釋指引。通常情況下，往往是解釋指引先出爐，實務工具的實踐守則隨後跟進；但本次順序卻完全顛倒。實踐守則已於2026年6月率先確定，而旨在闡明條文範圍與適用對象的指引則仍處於草案階段。這意味著未簽署企業在短時間內亦缺乏可替代守則所倚賴的官方解釋資料。決定不簽署的企業，最終只能密切關注AI辦公室未來的指引公布以及各會員國監督機關的初期執法案例，持續確認自家的措施是否並未落後。

站在韓國企業的立場來看，問題更加錯綜複雜。韓國國內的《AI基本法》中亦已納入生成式AI產出物的標示義務，從而產生了必須同時滿足歐盟規定與韓國國內規定的雙重合規負擔。若兩部法律的標示方式或文案要件未能完全重疊，則必須先研判能否以單一的浮水印與告知體系同時滿足雙方要求。若總部設於歐盟境外的企業，最好預先設妥能應對第50條相關詢問或調查的歐盟境內窗口（即擔任法定代理人角色的組織），以保安全。

聖水洞新創公司的法務團隊在當天的會議中傾嚮於簽署。畢竟其影片生成服務在歐洲使用者面前的曝光比重不低，且公司內部亦難以專門配置人力去與各監督機關個別交涉。不過，法務長最後留下了這樣一句話：簽署並不意味著結束，而是從簽署的那一刻起，內部檢視公司是否確實具備守則所要求的各個項目的新工作才正要開始。開發團隊要在何時之前嵌入浮水印模組、客服中心軟體劇本中要如何納入深偽告知文案，這些答案無論簽署與否，最終都得由各團隊自行填補。

參考資料

法律依據：《AI法》第50條第2項（標記與偵測）、第50條第4項（深偽與公益文字標籤化）、第50條第7項（實踐守則依據） 確定日期：2026年6月10日閉幕全體會議 首批簽署者名單截止時間：2026年7月22日18時（中歐時間），後續是否開放參與需確認AI辦公室公告 全面適用：2026年8月2日 既有系統標記要件寬限期：2026年12月2日（反映《AI綜合修正案》暫定協議時） 可互操作性要件適用：2027年2月2日 罰鍰依據：第99條（一般，營業額3%或1,500萬歐元中較高者），與GPAI專用減輕條款第101條分立 執法主體：會員國市場監督機關（與GPAI實踐守則由AI辦公室直接執行相對照） 法律性質比較：調和標準（第40條，完整符合性推定）居首；GPAI實踐守則（第53條第4項、第55條第2項，條文上證明路徑）次之；透明度實踐守則（第50條第7項，無條文依據僅作為實務慣例發揮作用）位居其下

第三部

給處理通用AI模型的企業

第7章 我們是通用人工智慧提供者嗎

來看一家位於大田的人工智慧新創公司。開發團隊從國外下載了一個公開的語言模型。公司用多年來累積的化學製程資料對該模型進行了重新訓練，如今其成果已達到能以文字說明化學設備異常徵兆的水準。公司開始將該模型付費提供給德國的兩家化學公司。執行長這樣認為：模型是別人做的，我們只是針對自己的業務需求進行了調整，因此就監管角度而言，我們更接近使用者。

這個判斷是否正確，正是本章探討的核心。如果這家公司被歸類為通用人工智慧模型（general-purpose AI model，以下簡稱 GPAI 模型）的提供者，那麼《人工智慧法案》第 53 條的一整套義務將全部適用。公司必須備妥技術文件、制定著作權政策，並公開用於訓練的內容摘要。僅是「進行了微調」的自我認知與「提供者」這一法律地位之間的差距，正是這家公司所忽視的風險。

先來解釋什麼是 GPAI 模型。《人工智慧法案》第 3 條第 63 款將 GPAI 模型定義為展現出相當程度的通用性（significant generality），且能勝任廣泛不同任務的 AI 模型。無論其是否採用大量資料透過大規模自監督學習方式進行訓練。只要模型無論以何種方式投放市場，能夠被整合至各種下游系統或應用程式中即可。語言模型就是常見的例子。單一模型既能翻譯、摘要，也能撰寫程式碼。這種廣泛的用途就是相當程度的通用性的意思。而開發並將其投放市場的一方就是提供者（provider）。

7.1 相當程度的通用性標準

「相當程度的通用性」這一詞彙並未在條文中被明確具體地逐一列舉。《人工智慧法案》前言第 97 條與第 99 條將大規模生成式 AI 模型列為 GPAI 的典型代表，理由在於其能靈活生成文字、音訊、影像、影片等各種形式的內容，並能輕鬆處理廣泛的任務。僅能進行人臉辨識的模型、僅能評定信用分數的模型則排除在外，因為它們被侷限於單一狹隘的工作中。相反地，無論問什麼都能給出像樣回答的大型語言模型，則被視為具備相當程度的通用性。

歐盟委員會自身也承認，無法將此判斷明確固定為一份精確的能力清單。取而代之的是，將運算能力（computational power）作為指標性標準（indicative criterion）。亦即測量模型訓練時所消耗的浮點運算次數（FLOPs）。從市面上的實務解讀來看，訓練運算量超過 10 的 23 次方 FLOPs 的模型，會被視為已跨過認定為 GPAI 模型的指標標準門檻。此外，即使將跨過該門檻的模型進行剪枝（pruning）或量化（quantization）以縮減體積所產生的衍生模型，只要原始模型超過該門檻，同樣無法擺脫被視為 GPAI 模型的認定。這意味著並非將模型縮小就能規避監管。10 的 23 次方這個數字與後面第 9 章將看到的系統性風險門檻（即 10 的 25 次方 FLOPs）處於不同的位置。前者是用於劃分是否屬於 GPAI 模型，後者則是在 GPAI 模型中對具備系統性風險者加重義務的門檻。這兩者絕不能混淆。

包含此指標標準的指南，歐盟委員會於 2025 年 7 月 18 日核准其內容並首次提出，並於 2025 年 11 月 19 日以通訊文件 C(2025) 7719 正式採納。該指南不具法律約束力。《人工智慧法案》條文以及日後歐盟法院的解釋優先於此指南。然而在實務中，歐盟委員會與 AI 辦公室實際上將以此標準審理案件，因此絕非可以憑藉「無約束力」為由加以忽視的文件。

對韓國企業而言，該標準在實務上令人棘手之處在於中間地帶。當開發出既不算超大規模、用途也不算極度狹隘的模型時，是否已跨過相當程度的通用性門檻便顯得模稜兩可。我認為在這種模糊地帶，將其視為「已跨過門檻」並預作準備才是安全做法。若自行判斷未跨過門檻，事後卻被認定為已跨過，就等於遺漏了第 53 條的整套義務。反之，若視為已跨過並著手準備，事後證實其實未跨過，也不過是準備工作稍微超前而已。兩者的風險比重明顯傾斜。此外，若簽署針對 GPAI 模型提供者的行為準則（Code of

Practice），在與 AI 辦公室的互動中證明自身合規的程序將會簡便許多，這一點也值得記住。

歐盟委員會在指南中特別強調，相當程度的通用性之判定必須始終經過個別案例評估（case-by-case assessment），並將根據技術、社會與市場的趨勢發展，以及歐盟法院的解釋，重新檢視並修正指南本身。這意味著目前劃定的界線在 3 年後並不保證仍在相同位置。因此，我建議不要將測量此門檻視為一次性工作，而應建立為每次重新訓練或進行重大調整時重新評估的標準流程。

7.2 生命週期的起點：大規模預訓練

要確定提供者義務從何時開始附隨，必須先釐清模型的生命週期從何處開始。歐盟委員會在指南中對模型及其生命週期採取了較寬廣的看法。模型的各個開發階段，例如小型實驗性訓練結束後的正式訓練，以及後續為了專業化或改進所進行的微調（fine-tuning），都被視為同一生命週期內發生的活動，而非獨立的模型。歐盟委員會將該生命週期的起點認定為大規模預訓練（large-scale pre-training）。亦即為了建立模型的一般能力，使用大量資料進行基本訓練的重大過程展開之時。

這個起點之所以重要，是因為部分義務從此時起就必須齊備。第 53 條第 1 項 (d) 款所要求的訓練內容公開摘要，即必須符合 AI 辦公室制定之標準格式的文件，若非從開始訓練的那一刻起就記錄投入了哪些資料，事後將無法追溯編製。著作權政策亦然。若要遵循歐盟著作權指令（Directive (EU) 2019/790）中有關文字與資料探勘（Text and Data Mining）的例外規定，在蒐集資料的那一刻，就必須已經確立能篩選出已聲明保留權利之著作的政策。技術文件編製與更新義務（第 53 條第 1 項 (a)、(b) 款）也要求在整個生命週期中維持最新狀態，因此若未從訓練初期建立文件化習慣，拖得越久債務便會堆積得越多。

針對系統性風險，則有更進一步的要求。歐盟委員會認為，在大規模預訓練開始之前，亦即從規劃階段起，提供者即可合理預測其模型是否會達到 10 的 25 次方 FLOPs 的系統性風險門檻。因此，《人工智慧法案》第 52 條第 1 項規定的通報（即「毫不延誤且最遲於 2 週內」之通報），不應在訓練結束後，而應在「預期達到門檻之時」即已完成。若等訓練結束後看結果才進行通報，就為時已晚。

因此，若韓國企業正在規劃大規模預訓練，在動工訓練之前必須先建立兩件事：一是記載何時投入了多少、何種資料的紀錄體系；二是決定不投入哪些資料的著作權政策。這兩者的具體要領將於第 8 章探討。《人工智慧法案》本身已於 2024 年 8 月 1 日生效，而 GPAI 模型提供者義務則自 2025 年 8 月 2 日起適用，因此若公司在此刻準備進行大規模預訓練，就代表已經進入義務運作的時間軸內。

7.3 微調與下游修改：何時會成為新的提供者

揭開本章序幕的大田新創公司的疑問就在這裡。當引進他人的模型並進行微調（fine-tuning），即利用自身資料重新訓練並修飾後，該公司是否會成為該修正後模型的「新提供者」？

歐盟委員會給出的答案非常明確。並非模型的任何修改都會使下游修改者（downstream modifier）成為新的提供者。這與歐盟《藍色指南》（Blue Guide）的原則一脈相承。《藍色指南》指出，若產品為了改變其原始性能、目的或類型而進行了重大變更或全面修改，可被視為新產品。GPAI 模型也適用相同的邏輯。下游修改者成為修改後 GPAI 模型之新提供者的時機，僅限於該修改對模型的通用性（generality）、能力（capabilities）或系統性風險（systemic risk）帶來重大變更（significant change）的情況。

這裡有一個必須精準掌握的要點。劃分是否成為新提供者的標準，就在於這三者：通用性、能力與系統性風險是否發生重大變更。微調後的成果是否具備「獨立預定用途」（independent intended purpose），或是否派生出與原始模型不同的獨立用途，並非此項判斷的標準。在實務上，人們往往容易混淆這兩者。很容易認為微調後的模型具備了化學設備診斷這一狹隘且具體的目的，成了獨立的產品，因此應該成為新的提供者。然而，歐盟委員會的標準看的並非該成果用於何處，而是該微調在多大程度上改變了原始模型。即使是針對狹隘目的進行微調，只要通用性、能力或風險性質未發生重大變更，就不是新提供者。反之，即使是針對廣泛目的進行的修飾，只要實際上大幅改變了模型的通用性或能力，就可能成為新的提供者。衡量標準在於變更的幅度，而非用途的寬窄。

在實務上該如何評估是否屬於重大變更呢？我建議提出三個問題來衡量：微調後的模型是否能夠執行原始模型無法做到的新任務，抑或是僅以更高精確度執行原始模型本就能做的事？基準測試（benchmark）分數是否大幅躍升，進而跨入與原始模型不同的效能等級，抑或是僅屬於誤差範圍內的改善？是否繞過了安全機制或產生了新的危險行為，抑或是完全繼承了原始模型的安全特性？若三個問題的答案皆為「是」，則傾向於認定為重大變更；反之若為「否」，則傾向於反面。化學設備診斷這一狹隘用途本身，並不屬於這三個問題中的任何一個。

在這個關鍵點上，韓國企業常常產生誤解，認為「既然原始模型不是我們做的，我們就沒有責任」。法律要找的不是「製作模型的人」，而是「將模型投放市場的人」。如果引進他人的模型進行實質變更後，以自身的名義提供給歐洲客戶，那麼對於「投放市場」這件事的大部分責任便轉移到了進行微調的公司身上。不過，成為新的提供者並不意味著要全盤承擔原始提供者的所有義務。《人工智慧法案》第 53 條第 1 項規定的義務 - - 即技術文件、著作權政策與訓練資料摘要 - - 僅限於該次修改或微調的部分。公司必須提供的文件與資訊僅適用於自身微調的過程以及該過程中所使用的資料，並無義務重新對原始模型本身的訓練過程進行文件化。然而，確實存在著應從原始提供者處獲取的資訊流，相關實務細節將於第 8 章與第 20 章探討。

若原始模型本身即屬於具系統性風險的 GPAI 模型，且下游修改符合成為新提供者的標準，則該修改後的模型亦會被推定為具系統性風險的 GPAI 模型。在此情況下，下游修改者必須毫不延誤且最遲於 2 週內通報 AI 辦公室，並須承擔包括模型評估、紅隊演練（adversarial testing）、風險緩和、重大事件通報及確保網路安全等加重義務。此外，非 EU 境內的下游修改者若以此方式成為新提供者，依據《人工智慧法案》第 54 條，必須在 EU 境內指定授權代表（authorized representative）。大田新創公司將模型提供給德國公司的瞬間，該項義務便立刻成為該公司的法律責任。

7.4 三分之一門檻討論與實務判斷

在微調相關的討論中，實務上常提及一個數字，即「三分之一」這個數值。歐盟委員會將「用於修改的訓練運算量超過原始模型訓練所用運算量的三分之一」列為判定是否屬於重大變更的指標標準。此門檻背後的邏輯相當簡單：若投入了如此規模的運算量進行微調，該微調極有高度蓋然性對模型的特性與性能帶來實質性的變化。

切勿誤解這三分之一的含義。它並非決定是否成為提供者的自動開關。歐盟委員會本身亦承認此標準具備前瞻性（forward-looking），其處理方式可能隨著技術與市場的發展而調整。即使微調消耗的運算量未超過三分之一，只要重大改變了通用性或能力，仍可被認定為提供者；縱使超過了三分之一，也未必自動成為系統性風險模型。這僅為一項指標，而非法律上死板固定的門檻。

實務上棘手之處另有所在。那就是下游修改者完全不知道原始模型的訓練運算量時。若原始提供者未分享該資訊，就根本沒有依據去計算三分之一這個標準。遇到這種情況時，AI 辦公室可要求下游修改者提供其他適當手段來證明合規。這意味著並非「沒有資訊就能免除義務」，反而可能轉向「舉證責任更加沈重」的方向。因此，若韓國企業欲引進使用他人的模型，最好從合約階段起就要求原始提供者提供訓練運算量資訊。事後喊「不知道」絕無法作為合法的抗辯理由。

坦白說，這三分之一與「重大變更」之間的精確關係，目前尚未經由實務案例沉澱成型。究竟何謂重大變更，以及三分之一這一數值在個案中將如何適用，仍待未來的執法案例累積後才會逐漸明確。在此刻很難做出絕對確定性的劃分。因此我建議，若公司以微調為業務內容，應自行記錄自身的微調對原始模型造成了多大程度的改變、運算量相較於原始模型達到何種比例，以及過程中使用了哪些資料。日後若針對「是否為提供者」產生爭議，該紀錄將成為判斷的依據。AI 辦公室的罰款處分權限將於 2026 年 8 月 2 日起全面啟動。閱讀本章的當下距離該日期已不遙遠，這一點同樣值得銘記。

檢視我們公司屬於哪一方的自我提問

我們開發的模型是用於多種任務的廣泛用途，還是僅執行單一狹隘的工作？若用途廣泛且訓練運算量在 10 的 23 次方 FLOPs 左右或以上，就必須重新審視相當程度的通用性門檻。

我們是否正在規劃大規模預訓練？若是，在動工訓練前必須建立資料紀錄體系與著作權政策。若預計會達到系統性風險門檻，也應提前準備通報。

我們是否將他人的模型進行微調後提供給歐洲？必須檢視微調是否重大改變了模型的通用性、能力或系統性風險。微調成果用於狹隘目的這一事實本身，並非判斷標準。

微調的運算量是否超過原始模型的 3 分之 1？若超過，則需一併檢視系統性風險相關標準。若不知道原始模型的運算量，現在就應在合約中加入條款要求提供該資訊。

相關日期彙整：《人工智慧法案》生效日：2024 年 8 月 1 日。GPAI 模型提供者義務適用日：2025 年 8 月 2 日。指南首次公佈日：2025 年 7 月 18 日，正式採納日：2025 年 11 月 19 日。2025 年 8 月 2 日前投放市場模型之合規緩衝期：至 2027 年 8 月 2 日止。AI 辦公室罰款處分權限全面啟動日：2026 年 8 月 2 日。

回到大田的那家新創公司。該公司表示自己只是微調了他人的模型。然而，這項微調實際上在多大程度上改變了模型的通用性與能力、運算量達到原始模型的幾分之幾，該公司至今尚未進行量化測量。在未經測量的情況下，該公司便已將模型提供給了兩家歐洲客戶。換作是我，我會告訴這家公司現在該做的事：在做出判斷之前，先量化出數字；只有在數字出來之後，才能論斷自己究竟是使用者還是提供者。

第8章 通用人工智慧提供者的基本義務

接到了板橋一家新創公司法務組長的電話。他說公司將自家的語言模型供應給德國製造業者的合約已接近收尾階段，但對方個法務團隊在合約中夾帶了一份附錄。標題是《AI法》第53條合規確認書。面對第一次看到的條文編號，他這樣問道：「我們開發的模型也沒有那麼大，這也必須遵守嗎？」

答案大致為「是的」。歐盟《人工智慧法》（Regulation (EU) 2024/1689，以下簡稱《AI法》）於2024年8月1日生效，而通用人工智慧模型（General-Purpose AI model，以下簡稱GPAI模型）提供者的義務則自2025年8月2日起開始適用。第53條是針對所有開發GPAI模型並投放於歐盟市場的提供者，無論是板橋的新創公司還是矽谷的巨型企業，無一例外皆適用的條款。與僅適用於具備系統性風險（systemic risk）大型模型的第55條不同，第53條自貼上GPAI模型標籤的那一刻起便隨之而來。

第53條第1項規定了四項義務：(a)款為依據附錄11（Annex XI）撰寫技術文件並保持最新狀態；(b)款為向意圖將模型整合至自身系統的下游提供者提供依據附錄12（Annex XII）的資訊；(c)款為制定遵守歐盟著作權法的政策；(d)款則是依據AI辦公室（AI Office）製作的格式撰寫訓練內容公開摘要並予以公開。前兩者是交付給主管機關或合約相對人的文件，後兩者則是對外公布的政策與摘要本。板橋法務組長收到的確認書，正屬於對方基於其中(b)款、為了履行自身義務所要求的檔案。

8.1 技術文件，附錄11的要求

附錄11規定了模型在投放市場前應準備之檔案的骨幹。從名稱、版本、推出日期、發布方式、授權條件等一般資訊開始，到架構與參數規模、模態（文字、影像、語音等）、使用何種資料集進行了多少訓練、資料如何收集、清理與過濾、評估是由內部進行或委託外部機構、已知的侷限性以及訓練所消耗的能源量等，皆須載明。技術文件這個名稱看似樸實，實質上卻接近於從誕生那一刻起的成長記錄。

這份記錄的起點在實務上經常令人困惑。AI辦公室認為GPAI模型的生命週期（lifecycle）始於大規模預訓練（large pre-training run），即為了建立模型的一般能力而使用大量資料進行的基礎訓練開始之時。此後進行的微調 or 後續改進工作，並不被視為建立新模型，而是被視為同一個模型生命週期內的下一個階段。訓練計算量（training computation）不僅包含預訓練，連同生成合成資料與微調所使用的計算量也會累計計算。並非僅對海外開源模型進行微調就能擺脫義務約束。

技術文件必須在整個生命週期內保持最新狀態，主管機關要求時須能立即提出，且在投放市場後須保存10年。根據《AI法》第78條，接收文件的一方負有包含智慧財產權與營業秘密在內的保密義務，並須具備資安防護措施，但也不應過度信任此條文。資訊一旦落入主管機關手中，外洩的管道便多了一條，因此詳細記載至何種程度，法務團隊必須自行劃定界線。

8.2 提供給下游提供者的資訊，附錄12

第53條第1項(b)款指向稍有不同的方向。若附錄11的技術文件是面向監管機關，附錄12的資訊則是面向打算引進模型並搭載於自身產品的下游系統提供者。板橋新創公司的相對人 - - 德國製造業者正處於此位置。該公司預計將板橋新創公司的模型整合至自家的工業用AI系統中，唯有取得關於模型的能力與侷限、可預測性與可靠性、可再現性與安全性等資訊，他們自身才能履行《AI法》規定的義務。若無法理解所整合的模型，下游提供者便根本無法履行其義務，因此GPAI提供者的資訊提供是支撐整個價值鏈的核心支柱。

內容雖與附錄11重疊，但焦點不同。如何整合至自身系統、需要何種硬體與軟體需求、輸入與輸出的格式為何等直接關係到整合實務的資訊被置於優先位置。該資訊同樣受第78條保密義務的約束。下游提供者若認為GPAI提供者提供的資訊不實或根本未提供，可依第89條第2項享有提出申訴的權利。板橋法務組長收到的確認書，正屬於對方試圖事先杜絕提出申訴可能性的檔案，而在空白表單上僅僅簽名與具備實際內容後簽名，完全是兩回事。

對韓國企業而言，該項義務與域外適用緊密相連。總部設於歐盟境外的提供者若將GPAI模型投放於歐盟市場，或其產出物於歐盟境內使用，依據《AI法》第2條第1項(c)款均適用本法，並須依第54條或第22條於歐盟境內以書面指定授權代理人（Authorised Representative）。由於代理人常兼任資訊提供管道，因此資訊套件最好與指定代理人一併規劃擬定。

8.3 歐盟著作權法合規政策，《指令》第4條第3項的權利保留

第53條第1項(c)款要求GPAI提供者制定遵守歐盟著作權法及相關法規的政策。核心在於《數位單一市場著作權指令》（Directive (EU) 2019/790，以下簡稱《DSM指令》）第4條。第4條第1項承認文字與資料探勘（Text and Data Mining, TDM）的例外情況。這意味著大量抓取著作物用於模型訓練的行為原則上不構成著作權侵害。然而第4條第3項附帶了條件：若權利人已以適當方式明確聲明保留（reservation of rights），則TDM例外並不適用於該著作物。

實務便是在此產生分歧。若為公開於網路上的內容，權利保留得以機器可讀手段（machine-readable means）標示，例如 robots.txt 檔案或元資料（metadata），而GPAI提供者則必須動用先進技術（state-of-the-art technologies）識別並尊重此類標示。規避付費牆（paywall）限制存取的行為、突破《DSM指令》第6條第3項技術保護措施的行為、以及抓取未向公眾公開內容之網站的行為皆在禁止之列。網路爬蟲若無法透過日誌證明確實讀取並過濾了該標示，政策便只不過是紙上宣言。

該政策必須適用於相關模型各自的整個生命週期，但政策本身只需制定一份並適用於所有模型即可。若為引進GPAI模型進行微調 or 修改而成為新提供者的情形，著作權政策義務僅限於該修改與微調所使用的資料，並不意味著責任須追溯至原始模型的訓練資料。

實務守則（Code of Practice）便是在此一環節登場。依據《AI法》第56條由AI辦公室擬定的GPAI實務守則於2025年7月10日公布最終版本，由透明度、著作權、安全與保障（safety and security）三個章節組成。包含亞馬遜、Anthropic、Google、IBM、微軟、OpenAI在內的26個機構簽署了全部章節，xAI僅部分簽署了安全與保障章節，而Meta則以法律不確定性為由，成為唯一全面拒絕簽署的公司。簽署後可獲得證明合規的便捷途徑，並減輕AI辦公室的監管負擔。然而，AI辦公室本身已明確聲明遵守實務守則並不影響著作權及相關法規的適用與執行，因此該文件並非能在會員國著作權法面前充當盾牌的檔案。

8.4 訓練內容公開摘要，格式與營業秘密的界線

第53條第1項(d)款要求針對GPAI模型訓練所使用的內容撰寫並公開足夠詳細的公開摘要（Public Summary of Training Content）。歐盟委員會（European Commission）於2025年7月24日發布了說明通知（Explanatory Notice）與標準格式。格式分為識別模型與提供者的一般資訊、訓練所用資料的種類與來源、資料處理過程三個部分。資料來源項目則細分為大規模公開資料集、透過網路爬蟲收集的資料、自第三方取得或授權的資料、使用者生成的資料或合成資料。

該義務的目的在於平衡兩種利益關係。著作權人與監管機關有權知悉模型是吸收何種養分成長，而提供者則有權保護營業秘密。《AI法》第78條雖賦予保護智慧財產權、機密商業資訊與營業秘密的義務，但並不意味著摘要本本身得以機密填滿。欲作為營業秘密保護的細節，不應放進公開摘要，而應作為僅向AI辦公室或會員國主管機關提交的獨立文件交付。公開至何種程度、隱藏自何處開始的界線法律並未替我們劃定，因此我認為各公司應事先透過內部指引劃清此界線。

若訓練資料資訊未留存，或尋找資訊本身會造成不成比例的負擔（disproportionate burden），則被認定為例外情況。若為2025年8月2日之前已投放市場的模型，雖無須重新訓練或撤銷過去的工作，但仍須在著作權政策與訓練內容摘要本中敘明並證明「無資訊」或「尋找過程負擔過重」的事實本身。因不知情而沉默，與主動說明不知情，在法律上處於完全不同的立足點。

執行與時間表

AI辦公室對GPAI提供者處以實質罰款的權限將自2026年8月2日起生效。在此之前，將著重於行使資訊索取權（第91條），以掌握提供者的合規計畫並引導合作。一旦開始實際處以罰款，違反包含第53條在內的GPAI義務時，依據《AI法》第101條，得處以上一會計年度全球年總營業額的3%或1,500萬歐元（以較高者為準）作為行政罰鍰。

對於在2025年8月2日前已投放市場的GPAI模型提供者，將給予至2027年8月2日為止共24個月的寬限期。AI辦公室於2025年11月19日正式採納GPAI義務範圍解釋指引，明確指出並非每次對微小修改的衍生模型都視為獨立提供者，僅有進行重大修改的一方才會被轉移提供者義務。雖然執行幾行程式碼進行微調不會立刻使人成為GPAI提供者，但其

界線仍須帶入具體案例中方能確認。

參考資料

時間表如下：2024年8月1日生效；2025年7月10日實務守則最終版本；2025年7月24日訓練內容摘要格式；2025年8月2日GPAI義務開始適用；2025年11月19日採納義務範圍解釋指引；2026年8月2日處罰權限生效；2027年8月2日現有模型寬限期結束。

檢查清單如下：確認自家模型是否屬於GPAI，以及大規模預訓練於何時開始；備妥達到附錄11水準的技術文件草案，並建立10年保存體系；於合約之外單獨準備欲提供給下游提供者的附錄12資訊套件；檢查網路爬蟲管道（pipeline）是否確實讀取機器人排除標準（robots.txt）與權利保留元資料；按格式撰寫訓練內容公開摘要草案，並將劃分公開與非公開範圍的內部標準留存為文件；確認是否已於歐盟境內指定授權代理人，並在法務團隊內部共享「簽署實務守則無法替代遵守著作權法」的共識。

板橋的法務組長最終在確認書上簽了字。只不過在簽字前，他多花了三天時間。那是他與開發團隊面對面坐下來，逐一核對是否確實具備附錄12所要求資訊的三天。若沒有那三天，簽名也不過就是一張紙罷了。

第9章 系統性風險模型

9.1 10的25次方 FLOP 門檻與計算量估算

我曾接到過來自韓國板橋一家 AI 新創公司的電話。他們當時正在訓練下一代語言模型，計算 GPU 叢集的使用時間後發現，在訓練結束之際，公司似乎會跨越一條意料之外的界線。那條界線正是 10 的 25 次方 FLOP，也就是浮點運算 1 京的 1000 兆倍。這個數字可能一時難以體會，但若說明 GPT-4 等級模型的訓練計算量大約就在這個數值附近或已超越此門檻，您應該就能大致掌握其規模了。

《人工智慧法案》（Regulation (EU) 2024/1689）第 51 條第 2 項規定，通用人工智慧模型（General-Purpose AI model，簡稱 GPAI 模型）用於訓練的累計計算量若超過 10 的 25 次方 FLOP，即推定其具有「高影響力能力」（high impact capabilities）。第 3 條第 65 款所定義的「系統性風險」（systemic risk），是指源自當代最高水準 GPAI 模型能力之風險，且該風險可能對歐盟市場產生重大影響並蔓延至整個價值鏈。超越此門檻的模型將被推定為具有系統性風險的模型，其提供者須承擔第 52 條與第 55 條規定的加重義務。

在此，韓國板橋這家新創公司的代表存在一個誤解。依規定，若達到或預計將達到該門檻，必須毫不延誤地在最遲 2 週內發出通知，而他以為通知的受文者是「人工智慧辦公室」（AI Office）。但我對此條文有不同的解讀。第 52 條規定的通知受文者是「歐盟委員會」（European Commission），而非人工智慧辦公室。人工智慧辦公室雖然是歐盟委員會下設的新組織並扮演處理實務的窗口角色，但條文文字本身明確規定受文者為歐盟委員會。由於許多實務資料將兩者混用而導致混淆普遍存在，但書面文件應寄往何處，在實務中是與罰款金額同等重要的關鍵問題。

通知書中必須以兩位有效數字標示導致達到門檻的計算量，並附上如何估算該計算量的說明。雖然即便符合此門檻，提供者仍有機會提出據以主張實際上並未引發系統性風險的理由來對分類提出異議，但歐盟委員會已明確表示舉證責任在於提供者。這絕非一項輕鬆的程序。

讓我們再看一個日期。GPAI 模型提供者的義務已自 2025 年 8 月 2 日起正式適用。而自撰寫本文時起 9 天後的 2026 年 8 月 2 日起，歐盟委員會的罰款裁處權將正式生效。違規時的罰鍰為上一財務年度全球總營業額的 3%，或是 1500 萬歐元，兩者取其高。目前尚無公司收到罰款通知書，但從下週開始，情況將有所不同。

10 的 25 次方這個數字本身並非固定不變。第 51 條第 3 項規定，授權歐盟委員會得根據技術發展，透過授權法案（delegated acts）調整該門檻。這意味著以目前標準看似安全的公司，也可能在下次修訂時改變其地位。

對總部位於首爾的公司而言，只要自家模型的產出物在歐盟境內被使用（《人工智慧法案》第 2 條第 1 項 c 款），這一切討論便不再與己無關。公司原本應在 2025 年 8 月 2 日之前就建立好計算量估算體系，若至今尚未建立，現在開始為時未晚，但必須抓緊時間。

9.2 基於硬體的估算與基於架構的估算

讓我們回到前面韓國板橋新創公司的案例。該公司的工程團隊主管曾因不知該 how 估算計算量而感到困惑。人工智慧辦公室的指南提出了兩種切入方法。

第一種是「基於硬體的方法」。將所使用的 GPU 或其他硬體單元數量（N）、總使用時間（L，單位為秒）、硬體單元的理論最高性能（H，每秒 FLOP）、GPU 平均利用率（U）這四個數值相乘，透過公式 C(FLOP) = N x L x H x U 計算出計算量。若混合使用了不同型號的 GPU，則必須以加權平均值來設定 H。韓國板橋那位團隊主管面臨的難題就在於此：他們在訓練初期使用了 A100，後期則混合使用了 H100，由於這兩種設備的理論性能相差數倍，如何設定加權平均值便成了關鍵所在。

我們用具體數字來計算看看。假設使用 8000 張 H100 運作 30 天，平均利用率為 40%，則 N 為 8000，L 為將 30 天換算為秒數的約 259 萬秒，H 為單張 H100 的理論最高性能（每秒約 1000 TFLOPS），U 為 0.4。將這四個數值相乘，大約會得到 8.3 乘以 10 的 24 次方 FLOP。這雖然比門檻值 10 的 25 次方低了一個位數，看似可以放心，但只要將利用率提高估算至 50%，計算結果就會突破 1.0 乘以 10 的 25 次方。這意味著 Excel 表格中的一個利用率欄位，就決定了是否負有通知義務。

另一種則是「基於架構的方法」。這種方式是將神經網路在訓練期間進行的前向傳播（forward pass）與反向傳播（backward pass）次數，乘以單次完整傳播所進行的運算量；在實務上，常使用 6 x P x D 這個近似公式。其中 P 為參數數量，D 為訓練 Token 數量。雖然該近似公式計算簡便，但若採用特定的最佳化技術，可能會與實際計算量產生偏差。

從執管當局的角度來看，無論使用這兩種方法中的哪一種皆可，但核心關鍵在於必須在通知書中載明採用了何種方法。若隱瞞或含糊帶過方法論，其本身就會成為違規的線索。人工智慧辦公室於 2025 年 7 月 18 日核可了該指南內容，並於同年 11 月 19 日正式採納（C(2025) 7719）。雖然這是不具法律約束力的解釋性指引，但卻是預先告知歐盟委員會將如何實際解讀條文的實質指南。指南中所列的估算方法本身也已明定具有範例性質，因此在特定情況下當局會認同哪一方，目前尚無先例。

兩種方法都基於同一個前提：不僅是預訓練（pre-training），連同合成資料生成、微調（fine-tuning）等用於提升模型能力的所有運算，都會計入累計計算量（前言第 111 條）。實務上的確有公司忽略了微調階段的計算量會將總量推升突破門檻的情況。我曾聽聞有案例僅計算預訓練階段的計算量便掉以輕心，最後加總微調運算量後才發現已超越門檻。

在實務上，我給予的建議只有一個：切勿僅擇一使用，而是從模型開發初期起，就養成同時以兩種方法分別計算並互相比對的習慣。若兩個數字存在較大偏差，該偏差本身就會成為需要在通知書中加以說明的內容。當首爾總公司建立計算量記錄系統時，在整個開發生命週期中留存日誌（logs），未來無論是要提出異議還是進行通知，都能作為依據。

9.3 模型評估與敵對性測試

超越計算量門檻後所面臨的後續要求更為沉重。《人工智慧法案》第 55 條第 1 項 a 款要求具有系統性風險的 GPAI 模型提供者採取包括模型評估在內的措施，特別是包含「敵對性測試」（adversarial testing，通常稱為紅隊演練/red-teaming）的程序。您可以將其理解為聘請倫理駭客來對自家模型進行實際攻擊測試的程序。這項要求旨在確保於模型的整個生命週期中，持續審視其是否會產出有害結果或存在被濫用的空間。

自 2025 年 8 月 2 日起，此項義務已正式適用。不過，對於在此日期前已在歐盟市場推出的模型提供者，則給予至 2027 年 8 月 2 日止、為期 24 個月的緩衝期。切不可忽視新推出的模型與既有模型之間在適用時間點上的差異。

從執管當局的角度來看，該條文之所以棘手，原因在於對於「何謂充分的評估」目前尚無成熟的標準。人工智慧辦公室自身也承認，外部模型評估者的生態系仍處於發展階段，其所採用的方法論在初期可能會良莠不齊。即便如此，當局仍會以「最新技術水準」（state of the art）作為判斷基準。我對該條文的解讀如下：缺乏標準並不能作為免責事由；反而正因為缺乏標準，將自己設定的基準及其依據作成書面文件留存，在日後會有利得多。

在實務中常被問到的問題還包括：究竟該在內部建立紅隊能力，還是委託外部專業機構？雖然外包給外部機構眼下看起來較為方便，但既然人工智慧辦公室已公開提及外部評估者生態系尚不成熟，若選擇的機構不當，日後其可信度可能會遭到質疑。若是我的話，會建議將核心的紅隊能力保留在內部，僅針對特定領域（如生物安全、網路攻擊模擬等）引進外部專家予以補充。

若資訊不充分或收到科學小組（Scientific Panel）的警告，人工智慧辦公室亦有權親自展開評估。這意味著若內部的紅隊報告不夠完善，當局就會親自主動介入審查，因此報告的品質直接決定了監管的強度。

首爾總公司的應對順序如下：首先，在 2025 年 8 月 2 日之前（由於目前已過該日期，應立即執行），確認自家模型是否為滿足相當通用性標準的 GPAI 模型，以及計算量是否超過門檻。接著，從設計階段起便將涵蓋整個模型生命週期的評估與敵對性測試體系融入其中。這種被稱為「合規源於設計」（compliance by design）的做法，比起事後

修改，從一開始就納入的成本要低得多。

9.4 重大事故通報

如果說前一節的紅隊演練是事前預防，本節探討的則是事故實際發生後的情形。第 55 條第 1 項 c 款要求具有系統性風險的 GPAI 模型提供者追蹤、記錄並通報與「重大事故」（serious incident）相關的資訊，但這次的受文者有所不同 - - 為人工智慧辦公室，以及在必要時通報相關成員國的主管機關。切勿將此與 9.1 中提到的向歐盟委員會進行的通知相混淆。計算量門檻通知是發給歐盟委員會，而重大事故通報則是發給人工智慧辦公室。即使處理的是同一套法規，書面文件的遞送對象亦有所區分。

人工智慧辦公室所指的重大事故分為兩種。其一是對模型或其實體基礎設施的嚴重網路安全侵害，包含模型參數外洩或網路攻擊等；其二是因 GPAI 模型的異常運作或事故，直接或間接導致第 3 條第 49 款 a 目至 d 目所規定的 AI 系統相關重大事故。然而，在我所取得的資料中，尚無法確認第 49 款各目的具體內容。不確定的部分最好明確標註為不確定，這部分有待後續修訂版本中，待歐盟委員會發布進一步指引後再行補齊。

2025 年 11 月，歐盟委員會正式發布了具系統性風險 GPAI 模型的重大事故通報表格。該表格包含 Word 與 PDF 兩種格式，已成為展示履行《GPAI 行為準則》（Code of Practice）第九項承諾（Commitment 9）的文件，並透過 EU SEND 平台進行提交。同時，亦附帶了安全與風險管理活動記錄必須至少保存 10 年的要求。表格的頒布意味著，如今已難以再以「尚未準備好」作為藉口。

人工智慧辦公室已明確表示，期待業者能進行事故發生前的「主動通報」（proactive reporting）。這可能是《GPAI 行為準則》中的承諾，或是替代性合規證明方式的一環，並獨立於第 55 條第 1 項 c 款規定的重大事故通報義務之外。在此處我的研判有所分歧：若劃分法律義務與實務建議，法律義務是在事故發生後進行通報，但考量與主管機關的關係，在出現事故徵兆的階段主動聯繫，更有利於降低日後的調查強度。與監管機關的關係與其說是扣分關係，倒不如說是建立信任的關係。

總部位於歐盟境外的韓國企業，在將 GPAI 模型推出至歐盟市場之前，必須以書面指定設立於歐盟境內的授權代表（Authorised Representative，《人工智慧法案》第 54 條）。該授權代表須作為與人工智慧辦公室溝通的窗口，並具備代為通報重大事故的能力。相關成員國的主管機關通常是以授權代表設立所在國為核心來劃定，因此將授權代表設立於哪個國家，便決定了日後將與哪個國家的主管機關對話。建議您現在就確認授權代

表合約中是否已明確載明事故通報業務。

9.5 確保網路安全與 2026 年 7 月《網路安全與人工智慧行動計畫》

第 55 條第 1 項 d 款要求在整個生命週期中，確保對模型及其實體基礎設施提供適當水準的網路安全保護。前一節討論的重大事故中，嚴重的網路安全侵害便與此項義務直接相關。在實務中常被忽視的一點是，受保護的對象不僅限於模型本身，還涵蓋支援該模型的伺服器與網路。

在撰寫本文的當下，一項不可忽視的事態已經發生。2026 年 7 月 7 日，歐盟委員會發布了《網路安全與人工智慧行動計畫》（Communication COM(2026) 577 final）。該聲明發布之際，正好是我記錄 9.1 中韓國板橋新創公司案例前後，時間點巧妙得令人吃驚。這項計畫並非新法律，亦未對企業增加新的義務；準確地說，它是疊加在已實施的《人工智慧法案》、NIS2 及《網路韌性法案》（Cyber Resilience Act）之上，提供資金與市場機會的協調框架。

該行動計畫的核心可歸納為三點：第一，在歐盟層級提升第三方在 AI 模型進入歐盟市場前評估其能力與風險的能力，此為支援人工智慧辦公室監管功能的措施。第二，開啟開源軟體韌性運動（Open Source Software Resilience Campaign），由歐盟網路安全局（ENISA）、歐盟委員會、成員國及開源社群共同支援關鍵開源專案的維護與安全。第三為資金支援：透過「歐洲地平線」（Horizon Europe）與「數位歐洲計畫」（Digital Europe Programme），截至目前多年度財務框架（MFF）結束前已投入約 2 億歐元，並設定目標在 2026 年底前透過歐洲創新委員會基金（EIC Fund）促成對網路安全與 AI 新創公司達 1 億歐元規模的投資。

從執管當局的角度解讀這項行動計畫，其意圖十分明確。歐盟委員會自 2026 年 8 月 2 日起將開始對 GPAI 模型行使監督權，並已明確強調將鎖定引發系統性網路安全風險的模型。罰款裁處權生效的日期與行動計畫鎖定目標的時間點精準契合，這看似並非偶然，而是經過精確設計的步驟。

首爾總公司在這項行動計畫中應留意的要點有二：一是參與提升第三方評估能力公開徵求的管道已然敞開；二是越是大量採用開源元件的模型，越需要密切關注開源韌性運動的動向。《GPAI 行為準則》已於 2026 年 6 月獨立確立最終版本，簽署該準則具有被推定為合規的優勢。然而，準則中具體的技術標準與調和標準（harmonised standards）仍有部分尚未完整刊載於歐盟官方公報（OJEU）。

再過 10 天，歐盟委員會的罰款倒數時鐘就將啟動。建議將這份行動計畫的事實清單（factsheet）放在首爾合規負責人的桌上。8 月 2 日早晨，或許就得再次翻開那份事實清單了。

參考資料

重要日期：2024 年 8 月 1 日《人工智慧法案》生效。2025 年 8 月 2 日 GPAI 模型提供者義務開始適用。2025 年 7 月 18 日 GPAI 指南草案獲核可，同年 11 月 19 日正式採納（C(2025) 7719）。2025 年 11 月公布重大事故通報表格。2026 年 6 月定稿《GPAI 行為準則》最終版本。2026 年 7 月 7 日發布《網路安全與人工智慧行動計畫》（COM(2026) 577 final）。2026 年 8 月 2 日歐盟委員會罰款裁處權生效，既有模型緩衝期之起算點。2027 年 8 月 2 日 2025 年 8 月 2 日前推出之模型的緩衝期屆滿。

通知受文者判斷基準：計算量達到門檻之通知（第 52 條）為歐盟委員會。重大事故通報（第 55 條第 1 項 c 款）為人工智慧辦公室及必要時之相關成員國主管機關。應於公司內部規章中明確明定這兩份書面文件的受文者有所不同。

檢查項目：確認自家模型是否為 GPAI 模型，以及是否超過 10 的 25 次方 FLOP 門檻。分別透過「基於硬體」與「基於架構」兩種方法進行計算量雙重估算。制定預計達到門檻時於 2 週內向歐盟委員會通知之程序。決定內部紅隊能力與使用外部評估機構之範圍。建立重大事故定義與通報表格（EU SEND）提交體系，相關記錄保存至少 10 年。檢視模型與實體基礎設施之網路安全體系。確認歐盟授權代表（第 54 條）合約中是否包含代辦通知及通報條款。另行檢視是否與韓國國內《AI 基本法》產生雙重監管衝突。

第10章 致擬採用開放原始碼例外條款的企業

韓國板橋的一家新創公司開發團隊負責人曾前來找我。他們打算將自家訓練的韓語語言模型上傳至 Hugging Face，公司法務團隊卻為了其中一項授權條款文字纏鬥了數週。他們希望免費釋出，但保留公司名稱；同時又擔心日後大企業直接拿去謀利會太委屈，因此想限制商業用途。然而，當他聽到光是這項條件，就會導致該模型在歐盟人工智慧法案（Regulation (EU) 2024/1689，以下簡稱《AI法案》）面前無法被認定為開放原始碼時，臉色頓時凝重了起來。本章所要探討的內容，正是該神情背後的原因。

《AI法案》已於 2024 年 8 月 1 日生效，針對通用人工智慧模型（General-Purpose AI model，以下簡稱 GPAI 模型）提供者所規定的義務，主要集中在第五章。這些義務自 2025 年 8 月 2 日起開始適用，而 AI 辦公室（AI Office）的罰款裁量權則將於 2026 年 8 月 2 日起正式啟動。也就是說，當您閱讀本書時，相關義務早已生效，僅剩罰款這把利刃尚未拔刀出鞘罷了。若是在 2025 年 8 月 2 日前已進入市場的 GPAI 模型，可享有截至 2027 年 8 月 2 日的合規緩衝期；然而，在此之後新推出的模型則會立即成為適用對象。

《AI法案》在這堆沉重的義務之上，開了一扇喘息的窗口。亦即針對以免費開放原始碼（free and open-source）授權發布的模型，給予部分義務的豁免。其依據條文為第 53 條第 2 項與第 54 條第 6 項。歐盟委員會認為，此類模型能對研究與創新做出貢獻，並為歐盟經濟提供成長契機，若參數、架構乃至使用資訊皆予以公開而具備高度透明度，便值得減輕其義務負擔。不過，這附有相應條件；只要遺漏了其中任何一項條件，無論公司多麼真心認定並發布為開放原始碼模型，在《AI法案》面前都將被視為與付費封閉型模型無異。

10.1 免費開放原始碼授權的四項自由

《AI法案》所要求的四項權利分別為：存取、使用、修改與散布。這四者之中只要缺少任何一項，該授權條款在《AI法案》下便無法被認定為免費開放原始碼。

存取（Access）意味著任何利害關係人，皆應能在無需支付費用或受其他限制的情況下自由取得模型。只要不做國籍等不合理的歧視，實施使用者驗證等合理的安全與安保措施是被允許的。換言之，要求會員註冊或電子郵件驗證是可以接受的。

使用（Usage）是指原始提供者不得以智慧財產權為武器阻止他人使用或收取使用費的權利。不過，要求衍生物的標示（attribution）與散布須採用與原本相同或相似條件的限制，是允許的。亦即所謂的著佐權（copyleft）條件可行，但直接閹割用途本身的限制則不可行。

修改（Modification）是指任何人皆可在無需付費或不受限制的情況下修改模型的權利。無論是要進行微調還是剪裁權重，完全取決於使用者。

散布（Distribution）是指存取、使用並修改模型的人，能在限定條件下再次分發該成果的權利。要求保留創作者標示或著作權聲明等可作為限制條件，甚至無法阻止接收者以封閉式專有授權發布衍生物。我認為其核心在於敞開再散布的大門，而非試圖去控制再散布成果的授權型態。

這四項自由與 Linux 基金會等開源社群數十年來所雕琢的定義極為相似。唯一的差別在於《AI法案》將其以明文法條訂死。一旦授權文字觸及這四個條件中的任何一項，即便原始碼已公開在 GitHub 上，在法律層面上依然不屬於開放原始碼。

在此有必要釐清一項常見混淆。在開發者社群中，常將開放原始碼與開放權重（open weight）混用。只要能下載權重檔案就稱為開源的慣例十分常見，但《AI法案》的認定並非如此寬鬆。參數、模型架構乃至使用資訊都必須可公開存取，且在此之上，前述的四項自由必須在授權條款文字中獲得保障。僅丟出權重卻將使用條件維持模糊的常見做法，不足以獲得本章所論述的豁免。

10.2 未符合要件的典型案例

實務中經常踩雷的條款顯而易見。韓國板橋開發團隊負責人所希望的非商業性使用限制即為代表。限制模型僅能用於非商業、非營利研究目的的條款，會被認定為嚴重削弱了使用權利。因為《AI法案》認為授權條款必須開放給任何目的使用。

第二種是禁止散布。雖可使用模型但不得再散布的條件，侵犯了散布權利本身。第三種是使用者規模門檻。亦即當月活躍使用者（MAU）超過特定數值時須取得額外授權的條件，最具代表性的便是 Meta 的 Llama 授權系列因此類結構而引發爭議。第四種則是對特定使用案例要求額外商業授權。研究用免費但商業散布需另行簽約的常見雙重結構，便屬於此類。

這四個案例看似名稱不同，歸根究底皆源於同一個根源：提供者既不想完全放棄控制權，又覺得被稱為封閉型心生負擔，因而試圖打造中間地帶的心理。我認為將此類授權單獨稱為原始碼公開（source available）授權更為精確。這屬於看得見程式碼卻非開放原始碼的狀態，而《AI法案》在此點上極為冷酷。其立場在於：看得見與自由完全是兩回事。

這四種典型限制最終都會觸及 10.1 所提到的四項自由中的使用或散布。因此在審閱授權條款時，將條文逐一帶入這四項自由進行比對是最快的方法。在實務上，我認為僅憑這一步比對就能過濾掉一半以上的問題。

有一點需要特別說明。開放原始碼例外條款所豁免的義務，為第 53 條第 1 項 (a) 款與 (b) 款，亦即撰寫技術文件與向下游系統提供者提供資訊的義務，以及第 54 條指派歐盟境內代表的義務。實際上，若歐盟境外提供者確實符合真正的開放原始碼條件且不存在系統性風險，便無需設置代表。本書其他章節曾說明韓國企業無一例外皆須指定歐盟代表，那是基於封閉型模型或具系統性風險模型為前提的說明；在此特別更正：純粹適用開放原始碼豁免對象的模型，可能根本不受代表義務拘束。雖然條文如此規定，但在實務上，授權條款是否真正符合要件始終存在爭議空間，因此我認為預先設立代表仍是較為安全的做法。

10.3 變現禁止要件的實際適用

即便在條文上完全符合四項自由，仍可能錯失豁免，原因在於變現（monetization）禁止要件。《AI法案》前言（recital）第 103 條明確指出，模型的存取、使用、修改、散布皆不得要求財務補償。歐盟委員會對此概念的解釋相當寬廣。

歐盟委員會視為變現的代表性形式即為雙重授權（dual licensing）。亦即以單一免費開放原始碼授權釋出，但對商業使用要求另行支付費用的授權結構。非商業研究目的限制、禁止再散布、使用者規模門檻，以及對特定使用案例要求商業授權，全都被歸入同一範疇。這與 10.2 所探討的典型未符合要件案例重疊絕非巧合，因為這些限制歸根究底都是為了收取費用的機制。

在此實務工作者常有一個誤解，以為只要模型本身免費發布就不算變現。然而，歐盟委員會表示，即便是諮詢服務、支援服務、雲端託管費用等看似與模型無關的收益模式，也會視個案具體情況予以審視。「判定始終基於個案進行」是歐盟委員會的用語，而我每次閱讀這句話時，都會將其視為灰色地帶極廣的訊號。模型本身免費，但僅對 API 存取收費的常見開源商業模式，亦難以斷定處於安全地帶。

若要給出一個衡量基準，大致如下：是否針對模型本身（即權重與程式碼）收費是一次性基準；至於對輔助使用者便利使用該模型的附加服務收費，大致上仍屬安全。然而，若該附加服務實際上扮演了存取模型的唯一管道，例如理論上公開了權重，但實際上只能透過公司營運的付費 API 使用，這便等同於用金錢阻斷了存取權利本身。我建議在此邊界線上，規模越大越應採取保守判斷。

在此必須堅定強調一點：即便授權條款符合所有四項自由且無任何變現行為，若屬於具有系統性風險（systemic risk）的 GPAI 模型，將完全不適用開放原始碼豁免。第 53 條第 2 項即如此規定。被推定為具系統性風險模型的基準，為訓練時累積運算量超過 10 的 25 次方浮點運算（10^25 FLOPs）。一旦跨過此門檻，即便完整公開原始碼且不賺取半毛商業利益也無濟於事。必須無例外地履行第 52 條與第 55 條規定的所有加重義務。訓練運算量接近此門檻的公司，無論是否開源，都應參閱其他專章，該部分已在第九章進行探討。

10.4 豁免後仍殘留的兩大義務

即便符合四項自由、毫無變現行為且無系統性風險，從而完整獲得豁免認定，仍有兩項義務會貫徹始終：建立著作權政策的義務，以及公開訓練內容摘要的義務。

第 53 條第 1 項 (c) 款要求建立遵守歐盟著作權法及相關法規的政策。其中包含識別並尊重《數位單一市場著作權指令》（Directive (EU) 2019/790，DSM 指令）第 4 條第 3 項權利保留（reservation of rights）的程序。通俗來說，若創作者以機器可讀的方式標示不允許將其著作用於 AI 訓練，則必須具備能識別並過濾該標示的新技術。不得規避付費牆等技術性保護措施，且該政策必須適用於公司所開發之各個模型的整個生命週期。無須為每個模型重新撰寫政策，單一政策可跨多個模型通用。

第 53 條第 1 項 (d) 款規定，須按照 AI 辦公室所訂定的格式，針對訓練所使用的內容製作並公開足夠詳細的摘要。AI 辦公室於 2025 年 7 月 24 日釋出了解釋性通知（Explanatory Notice）與標準格式，該格式訂定了公開摘要應包含的最低標準。雖然官方表明其宗旨在於在保護營業祕密與著作權人及監管機構的知情權之間取得平衡，但實際上要公開至何種程度、何處止步，各家公司的判斷必然有所分歧。在此處，我建議公司內部應預先針對非公開範圍制定獨立基準。雖然可以第 78 條的保密義務為依據，但若過度縮減公開範圍，AI 辦公室可能會對摘要本身的真實充實度提出質疑。

這兩項義務未被列入豁免範圍的原因相當明確。因為以免費開放原始碼授權發布模型，並不代表該模型是以何種資料訓練、如何處理著作權等問題會自動變得透明。即便原始碼公開，訓練資料的來源卻往往一片漆黑。歐盟委員會正是為了補足這一漏洞，才不論是否開源都保留了這兩項義務。

換言之，本章的豁免屬於部分豁免而非全面豁免。或許無須撰寫技術文件或設立代表，但著作權政策與訓練內容摘要這兩份文件，無論是開源還是封閉型都必須同樣準備。我偶爾會看到韓國企業實務工作者誤以為走開源路線就能從文件作業中解脫，這話只對了一半。

執法當局的視角與罰款

AI 辦公室自 2025 年 8 月 2 日起監督 GPAI 模型提供者的義務履行，並自 2026 年 8 月 2 日起行使對違規行為處以罰款的權限。歐盟委員會於 2025 年 7 月 18 日率先公開了

GPAI 模型義務範圍的指南草案，而包含相同內容的正式 Communication C(2025)7719 final 則於 2025 年 11 月 19 日獲得採行。草案與正式版本之間有四個月的間隔，期間實務運作皆以草案為基準。我認為該指南扮演了填補條文空白的角色，但最終的法律確定性仍需仰賴歐盟法院裁判先例的累積才能定案。

AI 辦公室已明確表示將對變現概念進行廣義解釋。不僅對雙重授權或因使用限制而產生的額外授權要求進行審查，甚至連間接的獲利活動也在調查範圍之列。其宗旨旨在防止開放原始碼豁免遭到濫用，因此僅打磨授權條款文字卻保留原有收益結構的做法，極可能行不通。一旦確認違規，依據第 101 條，最高可處以上一會計年度全球年總營業額的 3% 或 1,500 萬歐元（以較高者為準）的罰鍰。

坦白說，本章所探討的許多判斷基準尚未經由裁判先例予以固化。指南僅提供方向，並不會對個別授權文字逐一裁決。運算量估算方法論無論是以硬體為基礎還是以架構為基礎，皆承認 30% 的誤差範圍，因此處於 10^25 FLOPs 門檻附近的公司，光是計算方式本身就可能引發爭議。在歐盟法院裁判先例累積之前，我認為採取保守解釋與行動才是安全之舉。

韓國企業當前應做之事

總部位於首爾的公司若開發 GPAI 模型並投放至歐盟市場，或者其產出物於歐盟境內被使用，依據第 2 條第 1 項 (c) 款，本章內容將完整適用。實務工作者當前應確認的事項（不分先後順序）如下：

首先須確認自家模型是否屬於 GPAI 模型，以及訓練運算量是否接近 10^25 FLOPs 門檻。若超過或可能超過該門檻，無論用開源包裝得有多漂亮，都無豁免可言。

應對照授權條款原文，逐一帶入存取、使用、修改、散布四項自由進行比對。確認是否隱藏了非商業條件、禁止散布、MAU 門檻，或對特定用途要求商業授權等條款。

即便授權條款文字無瑕，也須重新檢視實際收益結構中是否存在可被解讀為變現的空間。支援服務或託管費用等周邊收益模式亦屬於檢查對象。

應向組織內部明確宣導：即便在完整獲得開放原始碼豁免認定的前提下，建立著作權政策與撰寫訓練內容摘要這兩項作業依然存在。這兩者不僅是法務團隊的事，而是從資料

收集階段起就必須共同設計的工作。

須牢記指派代表與否取決於自家模型是否真正符合開放原始碼豁免要件。不過，若對是否符合要件存在爭議空間，我認為預先設立代表才是避免日後陷入困境的做法。

參考資料

四項自由判斷基準 存取：禁止要求付費，禁止國籍等不公平歧視，允許合理的驗證程序 使用：禁止利用智慧財產權請求使用費，允許要求標示・散布條件的著佐權 修改：允許無付費或限制的變更 散布：允許要求保留創作者標示・著作權聲明條件，阻止再散布為封閉型衍生物則屬違規

摧毀豁免的典型條款 模型僅限用於非商業、非營利研究目的 可以使用模型，但不得散布 月活躍使用者數超過一定數值時需要額外授權 特定商業使用案例須另行簽訂合約

不可豁免的義務 第 53 條第 1 項 (c) 款 建立遵守歐盟著作權法政策，識別 DSM 指令第 4 條第 3 項權利保留 第 53 條第 1 項 (d) 款 撰寫訓練內容公開摘要，遵守 AI 辦公室標準格式

時間表 2024 年 8 月 1 日：《AI法案》生效 2025 年 7 月 18 日：GPAI 指南草案公開 2025 年 7 月 24 日：訓練內容摘要標準格式公開 2025 年 8 月 2 日：GPAI 義務開始適用 2025 年 11 月 19 日：採行指南正式版本 Communication C(2025) 7719 final 2026 年 8 月 2 日：AI 辦公室罰款裁量權啟動 2027 年 8 月 2 日：2025 年 8 月 2 日前推出之模型的合規緩衝期結束

對於那位韓國板橋的開發團隊負責人，我最終是這樣告訴他的：若想加入非商業條件，那是公司的自由，但那一刻起就必須放棄本章所說的豁免。取而代之的是，必須二選一：要麼選擇如實準備技術文件並設立代表的封閉型路線；要麼徹底開放條件，並承擔著作權政策與訓練內容摘要這兩項作業的開源路線。那天，他把玩著名片，許久沒有說話。

第四部

高風險AI：邁向2027年的準備

第11章 我們的產品屬於高風險嗎

從仁川南洞工業區的一家電池管理系統公司說起吧。為方便起見，我們簡稱這家公司為「新興」。新興生產電動車用電池管理系統（即 BMS），作為零組件供應給歐洲整車製造商。當電芯溫度異常飆升或電壓失衡時，立即切斷電路的安全邏輯是該公司產品的核心。同天下午，松島一家人力資源管理軟體新創公司的負責人也撥電話給法務團隊。我們稱這家公司為「韓國人才」（Korea Talent）。韓國人才開發了一款人工智慧招募工具，能讀取應徵者的履歷並對面試候選人進行排名，並將其賣給了德國與荷蘭的幾家中型企業。兩家公司都帶著相同的問題前來找我：我們的產品屬於高風險嗎？

這個問題看似簡單，但兩家公司得到的答案卻走上了完全不同的道路。新興的 BMS 是嵌在「汽車」這一既存產品內部的安全組件，而韓國人才的招募工具本身就是一套完整的服務。人工智慧法對這兩者的規範依據的是不同的條文、不同的時間表以及不同的舉證方式。一個是附錄 I，另一個則是附錄 III。本章將從這個分水嶺出發，將分為八大領域的附錄 III 與嵌入既存產品內部的附錄 I 並列剖析，探討夾在中間的篩選條款所帶來的舉證負擔，最後再將韓國的六大主力產業一一帶入比對。

11.1 附錄 III：獨立型系統涵蓋的八大領域

人工智慧法第6條第2項規定，列於附錄 III 的人工智慧系統均被視為高風險。這裡的分野標準並非取決於系統搭載於何種硬體，而是取決於該系統的「預期目的」（intended purpose）指向何處。像韓國人才的招募工具這類獨立提供服務的軟體，正是該附錄的主要範疇，實務上分為八大領域。

第一是生物辨識。遠端生物特徵識別系統，以及推論敏感屬性的生物特徵分類系統均屬於此範疇。不過，像無差別收集臉部影像、在職場與教育機構進行的情緒識別，或是推論種族、政治傾向、性取向的生物特徵分類等危害較大者，則根本不屬於附錄 III，而是直接劃入第5條的禁止清單。留在附錄 III 的生物辨識雖然相對輕微，但依然屬於規範沉重的一環。

第二是用於關鍵數位基礎設施、道路交通，以及水、瓦斯、供暖、電力供應的安全管理與營運中的安全組件。在檢視電池產業時，這個領域將變得尤為重要。後面會再詳細說明。

第三是教育與職業訓練。包含入學審查、學習成果評估、考試作弊偵測、教育程度分發等。

第四是招募與僱用、勞務管理、取得自營職業機會。包含招募公告分析、應徵者篩選與排名、升遷與解僱相關決定、工作分配與績效評估等均屬於此領域。韓國人才的產品正屬於這個領域。

第五是取得必要的民間與公共服務。這個領域又分為四個分支：(a) 公共機關的社會保障給付資格評估、(b) 信用度評估與信用評分、(c) 人壽與健康保險的風險評估與保費計算、(d) 緊急救護優先順序分類。金融產業正好首當其衝。

第六是執法。包含個人風險評估、測謊、證據可信度評估、用於犯罪預測的個人輪廓分析等。

第七是移民、庇護與邊境管理。包含測謊、風險評估、申請審查、人員識別等。

第八是司法行政與民主程序。包含輔助法律研究與事實關係解釋，或對選舉結果及投票行為產生影響的系統。

在這八大領域中，韓國企業在實務上頻繁遇到的主要是招募、信用評估與保險、教育、必要服務、生物辨識這五個分支。至於執法、移民與邊境管理、司法行政，民間企業很少直接成為提供者，因此本章將不做過多討論。

有一點必須從一開始就明確定調：系統只要對自然人進行個人輪廓分析，無論屬於附錄 III 的哪一個領域，一律被視為高風險。所謂個人輪廓分析，是指自動處理並評估或預測個人的工作績效、經濟狀況、健康、偏好、信譽、行為、位置或移動軌跡等屬性。韓國人才的招募排名工具處理每位應徵者的履歷與回答以預測其招募適合度，完全符合這一定義。這意味著後文將討論的篩選條款，對韓國人才而言根本是一扇從未開啟的大門。

如第3章所述，關於附錄 III 獨立型系統的第16條（提供者義務）與第26條（部署者義務）全面實施日期，原定為2026年8月2日，後因《數位綜合法案》（Digital Omnibus）修訂而延至2027年12月2日。這意味著韓國人才擁有了更充裕的時間，並不代表義務本身消失了。我認為這段寬限期不應作為延後準備的藉口，而是充份做好準備的餘裕。

11.2 附錄 I：嵌入既存產品內部的人工智慧

現在讓我們轉到新興的故事。第6條第1項提出了與附錄 III 完全不同的兩項條件。第一，人工智慧系統被用作列於附錄 I 的歐盟既存協調法規所適用產品的安全組件，或者該人工智慧系統本身即為此類產品。第二，該產品依據該協調法規，在投放市場或投入服務前必須接受第三方符合性評估。必須同時滿足這兩項條件，才會被歸類為高風險。

列於附錄 I 的協調法規涵蓋《醫療器材法規》（2017/745）、《體外診斷醫療器材法規》（2017/746）、《機械指令/法規》、《玩具安全指令》、乘用車《型式認證法規》（2018/858）、《農林車輛法規》（167/2013）、《二輪及三輪車輛法規》（168/2013）、《船舶設備指令》（2014/90/EU）以及航空相關法規等。新興的 BMS 是適用乘用車型式認證法規的汽車安全組件，而該汽車原本就是必須通過型式認證這項第三方符合性評估的產品。兩項條件均已滿足，因此屬於高風險。

附錄 I 方面值得關注的焦點在於誰是提供者的問題。當人工智慧系統嵌入附錄 I 的產品中時，製造該產品的製造商會被視為人工智慧法上的提供者。如果新興以整車製造商的名義或商標將 BMS 隨汽車一同推向市場，那麼將由該製造商承擔第16條規定的義務。若新興僅供應零組件，且未在最終產品上標示自家商標，則新興本身可能並非人工智慧法規定的提供者。然而，在與整車製造商簽訂的供應合約中，多數情況下都會承擔配合遵守人工智慧法所需的資訊提供義務，因此在實務上，零組件廠商也承擔著實質上相當於提供者的責任。我建議務必逐條確認合約條款中的這一部分。

第8條第2項允許針對附錄 I 的產品，將人工智慧法要求的測試、報告與資訊，整合至既存安全法規的文件與程序中。例如在依據《醫療器材法規》編製的技術文件中，疊加人工智慧法所要求的風險管理與訓練資料偏誤驗證。此舉旨在減少重複，對於像新興這樣已經具備汽車零組件認證體系的公司而言，是實務上相當受歡迎的條款。

附錄 I 產品內嵌型高風險人工智慧的管制程序完成期限，原定為2027年8月2日，經《數位綜合法案》修訂後延後1年至2028年8月2日。這比附錄 III 晚了八個月。我認為這並非出於政治考量，而是與既存產品認證週期相扣合的實務考量。因為像汽車或醫療器材這類原本就以數年為單位進行認證週期的產品群，將人工智慧法的要求嵌入既存程序中需要花費更多時間。

11.3 第6條第3項的篩選機制，以及舉證負擔的陷阱

名列附錄 III 並不意味著無條件成為高風險。第6條第3項規定，即使是附錄 III 所列的人工智慧系統，若不會對自然人的健康、安全或基本權利帶來重大危害風險，即不被視為高風險。這就是所謂的篩選條款。這裡有一點必須特別釐清：該篩選條款僅對附錄 III 開放。附錄 I 產品內嵌型系統並無此類例外。新興的 BMS 在滿足汽車安全組件條件的那一刻起即屬於高風險，完全沒有接受額外例外審查的空間。有無篩選條款的差異，正是劃分附錄 I 與附錄 III 的第二條重要分界線。

篩選條款實際適用的條件分為四種：系統僅執行狹隘且程序性的任務；改善人類已完成的活動結果；僅偵測決策模式或偏離該模式的偏差，而不取代人類已作出的判斷或在缺乏適當人工審查的情況下主導判斷；僅執行與附錄 III 用途評估相關的預備性任務。只要滿足這四項條件中的任何一項，即可主張適用篩選條款。

然而，這四項條件會因為一項但書條款而瓦解 - - 對自然人進行個人輪廓分析的系統，完全不適用該篩選條款。讓我們再次回想韓國人才的招募排名工具：處理每位應徵者的資料以預測其適合度本身就是個人輪廓分析，在那個瞬間，通往篩選條款的大門便已關閉。無論將「狹隘的程序性任務」或「預備性任務」的論述構建得多麼精細，都無濟於事。在實務諮詢中，我經常遇到企業自我說服、將自家系統稱為「輔助性推薦工具」的情況。然而，只要具備自動評估應徵者個人屬性並進行排名的功能，這種說詞便無法成立。

主張適用篩選條款的提供者，必須承擔全部的舉證負擔。第6條第4項要求認為自身屬於附錄 III 但非高風險的提供者，在投放市場或投入服務前，須將該判斷依據作成文件紀錄。該文件係國家主管機關要求時須立即提交的資料，且依據第49條第2項規定，亦須登錄於歐盟資料庫中。若有公司選擇不登錄而私下主張適用篩選條款，這並非合規，而是坐以待斃等待被揭發。第80條賦予市場監管機關重新評估提供者劃分為非高風險系統的權限。若主管機關判定實質上屬於高風險，提供者將收到要求毫不延誤地符合人工智慧法規範的糾正命令。篩選條款並非供人自行判斷後默默溜走後門，更像是以隨時準備證明其判斷正當性為前提的前門。

記載該篩選條款具體適用案例的第6條指引，原定於2026年2月2日前出爐。該期限現已屆滿。歐盟執委會直到2026年5月19日才提出草案，利害關係人意見徵詢則於6月23日

截止。截至撰寫本稿的2026年7月24日，最終版本仍未公布。草案揭示的方向與迄今所述內容並無二致：從嚴解釋篩選條款，且不適用於個人輪廓分析。不過，一旦最終指引公布，具體案例預計會再增加，因此我認為現階段的自我評估文件宜保留為暫定版本，待最終本出爐後再行修訂較為安全。

11.4 套用至韓國六大主力產業

現在將上述架構逐一套用至韓國最具代表性的六大產業。這完全是我基於迄今處理的案例與公開資訊所作出的實務判斷，個別產品的實際分類仍須詳細剖析該產品的具體預期目的方能確定。

半導體為低風險。半導體晶片本身並非人工智慧系統，因此非人工智慧法的直接管制對象。然而，若是在半導體生產線上防止機器人誤動作或控制有毒氣體外洩的基於人工智慧之安全系統，情況則有所不同。若該控制系統嵌入適用機械法規的設備作為安全組件，即有可能屬於附錄 I。我認為半導體產業絕大多數均處於此狹隘例外之外。

汽車為高風險。正如新興的案例所示，先進駕駛輔助系統（ADAS）、緊急煞車、駕駛人監控、電池管理系統的安全相關功能，大多屬於附錄 I 的安全組件。由於 directly 攸關人身安全，反而很難規避高風險的判定。雖然準備期限已延至2028年8月2日而較為充裕，但考慮到汽車型式認證這項既存認證程序本身即以數年為單位運行，我認為必須從現在開始將人工智慧法的要求嵌入其中，才能趕上該期限。

家電為低風險。智慧電視的內容推薦、智慧冰箱的食材管理、無防碰撞功能的掃地機器人之自主運行，均屬於篩選條款所針對的典型輔助功能。這意味著它們是對決策結果不產生實質影響的便利功能。然而仍有例外：與兒童安全相關的玩具、整合特定醫療器材功能的健康照護家電、以生物辨識控制出入的智慧門鎖，均可能落入附錄 I 或 III。建議家電企業切勿將每項產品籠統劃為低風險，而應先個別過濾是否附帶安全相關的附加功能。

電池為中風險。該產業是劃分為附錄 I 與 III 的代表性案例。用於電動車的電池管理系統屬於汽車安全組件，故歸為附錄 I。另一方面，大型儲能系統（ESS）的過熱預防與火災切斷系統，有空間被視為參與電力供應網營運的安全組件，從而可能編入附錄 III 的第二個領域 - - 關鍵基礎設施。這意味著在同一個電池產業內部，車用與大型 ESS 用途會適用不同的附錄，因此必須按產品線分別進行分類。

醫療器材為高風險。基於人工智慧的醫療影像診斷輔助軟體、手術機器人、病患監控與治療計畫系統，均屬於適用《醫療器材法規》（2017/745）或《體外診斷醫療器材法規》（2017/746）的產品。在該法規中被歸類為 IIa 級以上的軟體，原本就是必須接受第

三方符合性評估的對象，因此幾乎自動滿足人工智慧法附錄 I 的兩項條件。韓國醫療器材企業若已取得 CE 認證，在確認該認證等級的瞬間，其在人工智慧法上的高風險與否事實上也已基本確定。

金融為高風險。信用度評估與信用評分屬於附錄 III 5(b)，人壽與健康保險的風險評估與保費計算屬於 5(c)。像是授信審查、保險承保、投資顧問聊天機器人等對個人財產與機會產生實質影響的人工智慧，大多落入此領域。透過篩選條款脫身的道路實質上也已被堵死。因為信用評分或保費計算本身即為自動處理並評估個人信用、財產與風險傾向的操作，難以脫離個人輪廓分析的定義。若韓國金融機構針對歐盟客戶提供信用評估或保險承保人工智慧服務，我建議從一開始就以高風險為前提進行準備。

將六大產業綜合歸納：汽車、醫療器材與金融為高風險，電池為中風險，半導體與家電為低風險。然而，「低風險」的判定僅基於目前銷售的產品配置，並不保證會延續至下一版本。一旦在家電上新增生物辨識出入功能，或在半導體設備中加入與人身安全直接相關的自動判斷功能，級別便可能隨之改變。建議將此評估視為對當前產品配置的快照。

再補充一個韓國企業所處的位置。上述六大產業中，只要有任何一項產品投放至歐盟市場，或者其輸出成果在歐盟內部被使用，即會受到人工智慧法第2條第1項(c)款規定的域外效力管轄。總部設於歐盟境外的提供者必須以書面指定歐盟境內的授權代表；若被歸類為高風險，亦將與韓國國內正在準備的《人工智慧基本法》要求互相扣合。由於兩部法律所要求的文件紀錄與風險管理體系有相當大的重疊，我建議在製作因應歐盟的文件時，即將國內法的因應納入同一架構下進行設計。分開製作兩套反而造成更大的資源浪費。

參考資料

判定步驟四階段。第一階段：確認我們的產品是否為適用歐盟既存產品安全法規（醫療器材、機械、汽車、玩具等）之產品的安全組件，或是該產品本身。若是，則檢視該產品是否屬於第三方符合性評估對象。若兩者皆符合，則屬於附錄 I、高風險，且無篩選條款適用。第二階段：若不屬於附錄 I，則檢視是否屬於附錄 III 的八大領域（生物辨識、關鍵基礎設施、教育、招募、必要服務、執法、移民與邊境管理、司法行政）之一。第三階段：若屬於上述領域，則確認是否對自然人進行個人輪廓分析。若屬個人輪廓分析，則立即判定為高風險，且不適用篩選條款。第四階段：若不屬於個人輪廓分析，則審

查是否滿足「狹隘的程序性任務」、「改善人類已完成活動」、「僅限於模式偵測的輔助」、「預備性任務」這四項條件之一；若滿足，則將該判斷依據作成文件紀錄，並登錄於歐盟資料庫中。

韓國六大主力產業摘要。半導體為低風險，例外為安全控制設備。汽車為高風險，屬於附錄 I，施行期限為2028年8月2日。家電為低風險，例外為玩具與生物辨識出入。電池為中風險，車用 BMS 屬於附錄 I，大型 ESS 屬附錄 III 檢討對象。醫療器材為高風險，適用 MDR/IVDR IIa 級以上。金融為高風險，屬於附錄 III 5(b)・5(c)，因個人輪廓分析而排除篩選條款。

日期表。2026年2月2日：第6條指引原定期限（已屆滿）。2026年5月19日：歐盟執委會公布高風險分類指引草案。2026年6月23日：草案意見徵詢截止。2026年8月2日：第50條透明度義務全面實施，以及人工智慧辦公室啟動罰款處分權限（高風險義務本身尚未實施）。2027年12月2日：附錄 III 獨立型高風險人工智慧義務（第16與26條）開始實施。2028年8月2日：附錄 I 產品內嵌型高風險人工智慧管制程序完成時點。

新興的 BMS 負責人與韓國人才的招募工具負責人，在當天下午各自帶著不同的結論走出會議室。新興甚至無需評估篩選條款，從一開始就著手進行高風險的合規準備；而韓國人才光是承認自家產品正進行個人輪廓分析這項事實，就花了整整兩週。兩家公司的差異並非技術上的差距，而是取決於能否坦誠審視自家產品究竟對人類產生何種影響。

第12章 高風險系統提供者的義務體系

讓我們以一家位於板橋（Pangyo）的招募解決方案新創公司為例。這家公司開發了一項透過分析履歷與面試影片來為招募人員提供應徵者（候選人）排名的服務，目前正打算將此服務導入德國與荷蘭的客戶端。用於招募審查的 AI 屬於歐盟人工智慧法（Regulation (EU) 2024/1689，以下簡稱 AI 法）附錄 III 第 4 款規定的高風險（High-Risk）AI 系統。從這一刻起，這家公司不再只是販售單一服務的業者，而是成為必須完整承擔 AI 法第三章所要求之六大義務的提供者（provider）。本章所要探討的正是這六大義務：風險管理系統、資料治理、技術文件與自動記錄及向部署者提供資訊、人類監督、準確性、穩健性與網路安全，以及符合性評估、CE 標誌與歐盟資料庫登記。

首先，讓我們整理一下日期問題。本書資料調研所使用的 NotebookLM 筆記中，將高風險義務的全盤適用日期記為：附錄 III 為 2026 年 8 月 2 日，附錄 I 為 2027 年 8 月 2 日。這是依據 AI 法於 2024 年 8 月 1 日生效後計算 24 個月與 36 個月的原始時程，因此計算本身並沒有錯。然而，由於理事會於 2026 年 6 月 29 日最終批准了 AI 數位綜合法案（Digital Omnibus on AI），該時程已被延後。隨著協調標準（harmonised standards）的制定有所延遲，加上成員國通知機構（Notified Body）的指定工作進度未能跟上，附錄 III 獨立型高風險系統的義務適用日已被延至 2027 年 12 月 2 日，附錄 I 產品內嵌型系統則延至 2028 年 8 月 2 日。因此，本章所指的適用日並非筆記中的舊日期，而是這個新日期。之所以特地指出資料與搜尋結果不一致之處，是因為像板橋新創公司這樣的企業，可能會因輕信舊日期而倉促準備導致漏洞百出，或是反過來聽到延期的消息就完全放手不管。有趣的是，作為罰款依據條款的第 99 條與第 101 條本身，仍按原定時程於 2026 年 8 月 2 日起生效。違反如第 5 條禁止行為或第 50 條透明度義務等已施行條款者，自該日起即成為處罰對象。不過，就本章所探討的第 9 條至第 15 條，以及第 43 條、第 48 條、第 49 條而言，由於高風險義務本身尚未全面適用，因此我認為實際執行的重心將移至 2027 年 12 月與 2028 年 8 月之後。

12.1 風險管理系統：第9條所要求的循環結構

板橋新創公司的開發團隊負責人最先遇到的條款 beer 便是第 9 條。該條文並非製作一份文件就大功告成的形式主義，而是要求在高風險系統的整個生命週期（lifecycle）中持續運行的程序。必須預先識別並評估風險，透過變更設計或使用者指引等手段加以降低或消除，單獨檢視該系統對未滿 18 歲的未成年人或脆弱群體會產生何種影響，並經過確認合規性與效能的測試。核心在於這四個階段並非一次性結束，而是每當修改系統時都必須重新運行。若是招募 AI，就意味著每當更新應徵者排名演算法時，都必須重新走一遍這個循環。

第 9 條第 10 項允許若提供者已根據其他歐盟法律運作內部風險管理程序，得將第 9 條第 1 項至第 9 項的要素疊加或整合至該現有程序中。若是屬於醫療器材或機械類等附錄 I 的產品，得根據第 8 條第 2 項，將 AI 法的要求嵌入現有協調法規的文件與程序中。像板橋新創公司這樣開發附錄 III 獨立型系統的企業，由於缺乏此類現有架構，必須從頭重新設計第 9 條的風險管理系統；而我認為在這種情況下，反而是將其與品質管理系統（第 17 條）合為一體進行設計更為妥當。因為完全沒有必要各自建立兩套將風險評估結果即時反映至系統修正過程的管理程序。

執法機關 AI 辦公室（AI Office）在評估此風險管理系統時，會以系統的預期目的（intended purpose）為基準。若套用了協調標準，亦可享有符合性推定（Presumption of Conformity）的益處。然而如前所述，由於協調標準本身尚未健全完備，在目前時間點，與其依賴標準，不如準用國際標準（如 ISO/IEC 23894 等 AI 風險管理指南）並將自家程序保留為書面文件會更為安全。不確定的部分當然必須註明不確定。風險管理系統的具體實施方式與風險評估細部標準，依系統種類的不同很大程度上存在彈性空間，而這項空白預計將由原本預定於 2026 年 2 月 2 日前出臺的第 6 條指引以及未來的協調標準來補足。

12.2 資料治理與訓練資料品質：第10條與偏誤問題

招募 AI 真正的引信就在這裡。若使用過去的招募資料來訓練模型，資料中原已存在的性別或畢業學校偏誤就會被直接繼承下來。第 10 條規定用於高風險系統的訓練（training）、驗證（validation）與測試（testing）資料集必須具備相關性（relevant）、代表性（representative），並盡可能無錯誤（free of errors）且完整（complete）。第 10 條第 2 項更進一步明確要求，必須評估資料收集的設計選擇、收集過程、包含標註（annotation）、標籤化（labeling）、清洗（cleaning）與增強（augmentation）在內的準備過程、治理體系、負責人的角色與責任、用於偏誤檢測與緩和的評估，乃至於特定情況下資料集可能不適當的可能性。對於板橋新創公司而言，這意味著必須從審視過去 5 年招募資料中哪些職類、哪些年齡層被低估代表（underrepresented）的統計屬性開始著手。

此處出現了與個人資料問題相互交織的交集。第 10 條第 5 項允許提供者僅在為了偏誤檢測與矯正而嚴格有必要的情況下，例外處理諸如種族、健康資訊等特殊類別個人資料（special categories of personal data）。然而，這僅能在完全遵守 GDPR（Regulation (EU) 2016/679）及相關指引所要求之保護措施的前提下始得為之。我對此條文的解讀如下：若為了消除偏誤而隨意收集性別或種族資訊，其行為本身即構成違反 GDPR，因此 AI 法第 10 條所給予的例外只不過是把門縫稍微打開，並非意味著可以隨意進出。在實務上，透過匿名化或使用第三方代理變數（proxy variables）的方式通過這道窄門會較為安全。

這項要求亦適用於未直接用於模型訓練的資料，且資料集還必須反映 AI 系統實際應用的地理、語境、行為與功能環境，這些都是容易被忽略的部分。若將以韓國資料訓練出的招募 AI 直接導入德國市場，可能會被指責資料未反映德國的勞動市場結構或履歷習慣。AI 辦公室在判斷資料的相關性、代表性、完整性及是否存在偏誤時，同樣會以預期目的為基準，並參考是否套用了協調標準或共通規範（common specifications）。關於如何定量測量相關性、代表性與完整性的具體技術標準，目前尚未透過協調標準全部釐清，該部分的解釋可能因系統種類而異。

12.3 技術文件、自動記錄以及應提供給部署者的資訊

第 11 條要求在投放市場前建立技術文件（technical documentation），並在此後持續更新至最新狀態。應包含於文件中的項目列於附錄 IV（Annex IV），範圍相當廣泛，包括系統的預期目的與設計邏輯、架構、資料要求與訓練方法論、驗證與測試程序及結果、風險管理系統運作記錄，乃至於變更履歷。由於中小企業與新創公司獲准採用簡化格式，以板橋新創公司的規模而言，妥善利用此放寬的架構較為有利。

第 12 條要求具備自動記錄，即日誌（log）功能。高風險系統在運作期間必須能夠自動記錄事件，該記錄將用於識別系統風險、追蹤實質性變更，並支援整個生命週期的監控。若為招募 AI，這意味著事後必須能夠重構輸入了何種履歷、產出了何種分數，以及參與該判斷的模型版本為何。若缺乏此日誌，當發生事故時，無論是提供者或部署者都無法解釋究竟發生了什麼事。

第 13 條要求將相關資訊交付給部署者，使其能夠實際操作該日誌與系統本身。說明書中必須包含為進行人類監督所應採取的措施、運作系統所需的運算資源與硬體、系統的預期壽命、維護方法，以及收集、儲存與解讀第 12 條日誌的方法。若將這三條條文串成一句話，便是：第 11 條要求提供者建立能自行解釋的狀態，第 12 條要求留存佐證該解釋的證據，而第 13 條則要求將該證據與解釋交付給部署者。本章的這一節比其他章節短，是因為分配到的 NLM 資料本身僅留有導入部分，缺少對條文本文的詳細敘述。這三條條文的具體內容已在本節中直接依據已公開的 AI 法原文及附錄 IV 進行補齊。

12.4 人類監督：第14條描繪的控制權

第 14 條規定高風險系統的設計必須包含人機介面，以便自然人（natural persons）能對其進行有效監督。目的十分明確：即使遵守了所有其他要求，對於殘餘的風險，人依然必須能夠在最後關頭及時攔截。條文將系統應提供給部署者的能力分為五種：理解系統能力與局限並能監控異常徵兆的能力；意識到人類無批判性地依賴 AI 輸出之「自動化偏誤」（automation bias）的能力；正確解讀輸出的能力；在特定情況下拒絕使用系統或忽視並翻轉輸出的能力；以及透過如停止按鈕等程序使系統中斷運作的能力。

套用至招募 AI 則如下：招募人員不能直接盲信 AI 給出的排名，而是當發現排名異常時，能夠追問為何得出此結果，並在必要時忽略該排名。板橋新創公司在此常犯一個錯誤：在介面上僅顯示排名與分數，卻隱藏了依據。如此一來，人類監督便有名無實。規定部署者側義務的第 26 條亦要求部署者依據提供者的說明書運作系統、管理輸入資料的相關性，並維持由人掌握最終判斷的人類監督體系；因此第 14 條並非由提供者單獨承擔的條款，而是與部署者相輔相成的條款。

條文各處頻繁出現「適當且比例相稱」的能力這一表述，而這究竟是指何種程度，必然會因系統種類與使用環境而異。因為急診室診斷輔助 AI 與招募排名 AI 所要求的人類監督強度絕不可能相同。AI 辦公室在判斷此適當性時，同樣會以預期目的為基準，並參考協調標準或共通規範。

12.5 準確性、穩健性與網路安全：第15條要求的技術最低標準

第 15 條要求高風險系統具備適當水準的準確性（accuracy）、穩健性（robustness）與網路安全（cybersecurity），並在整個生命週期中持續維持該水準。在準確性方面，必須於說明書中載明系統的準確度水準及相關指標（metrics）。若為招募 AI，意味著必須在文件中寫明以何種指標測量排名預測的準確度，以及該數值為何。穩健性則是指面對錯誤、不一致或缺陷時仍不致崩潰的彈性。亦即即使輸入值稍微異常，系統也不應產出荒謬的結果。

網路安全項目在實務上相當棘手。必須針對資料與模型污染（data and model poisoning）、對抗性範例（adversarial examples）、違反保密性等威脅，根據系統的風險與語境配備技術性措施。若有人惡意篡改履歷資料，實施讓特定候選人自動淘汰的攻擊，這正是資料污染攻擊。若缺乏防止此類威脅的措施，即構成違反第 15 條。值得注意的是，根據網路安全驗證制度 Regulation (EU) 2019/881 取得驗證或做出符合性聲明的系統，被推定為符合第 15 條的網路安全要求。衡量準確性、穩健性與網路安全「適當水準」的定量標準，目前同樣尚未透過協調標準完全釐清。儘管歐盟執委會表示將透過利害關係人與計量機構發展基準（benchmark）與測量方法論，但在現階段，板橋新創公司所能依賴的最終只有準用 ISO/IEC 24029 或 NIST AI RMF 等國際指南來建立自有的指標。

12.6 符合性評估、CE 標誌與歐盟資料庫登記

若截至目前探討的第 9 條至第 15 條屬於系統內部的設計要求，則第 43 條、第 48 條與第 49 條 fishing 便是對外證明該結果的程序。第 43 條提出了兩條分岔路：若符合協調標準，提供者得採用自行評估的內部控制（Annex VI）模式；若缺乏協調標準，或屬於附錄 III 第 1 款規定的生物識別（biometrics）系統，則必須走向經由經授權通知機構（Notified Body）審查的第三方驗證（Annex VII）模式。招募 AI 屬於附錄 III 第 4 款，原則上採內部控制。然而，若摻雜了哪怕些許生物識別要素，例如joined了分析面試影片中表情或聲音的功能，在未完整套用協調標準或未使用共通規範的情況下，那一刻起便會跨入附錄 III 第 1 款，從而可能被強制要求進行通知機構審查。這也是我經常向企業強調的一點，開發團隊往往會忽略：一旦偷偷嵌入生物識別要素，符合性評估的沉重程度將截然不同。

若 AI 內嵌於醫療器材或機械類等附錄 I 產品中，則遵循現有產品安全法規的符合性評估程序，但必須將 AI 法第三章的要求嵌入該程序中。第 8 條第 2 項即為允許此項整合的依據。若系統或預期目的發生實質性變更（substantially modified），則必須重新接受評估；但在持續學習的系統中，預先確定並形成書面文件的變更視為例外。此處關於實質性變更的界線何在再次變得模糊，究竟在徹底更換招募 AI 的排名演算法與調整數個參數之間的何處會觸發重新評估義務，僅憑條文無法一刀切定。

通過符合性評估後，應加貼 CE 標誌（第 48 條）。必須加貼於明顯、清晰且不易磨滅之處；若系統難以進行物理加貼，得替換為加貼於包裝或隨附文件上，若為僅以數位方式提供之系統，則得替換為顯示於介面或機器可讀程式碼中。若通知機構參與了審查，則必須在 CE 標誌旁一併標明該機構的識別碼。此外，亦須依據第 47 條撰寫歐盟符合性聲明（EU Declaration of Conformity）並保存 10 年；若產品涉及多項歐盟法律，得整合為單一聲明。

最後是歐盟資料庫登記（第 49 條）。提供者，或若為歐盟境外企業則為其授權代表（Authorised Representative），必須在投放市場前登記自身與系統。經常被忽視的一點是，不僅限於附錄 III 所列的高風險系統，依據第 6 條第 3 項篩選條款自行判斷不屬於高風險的系統亦為登記對象。在這種情況下，必須將為何判斷不屬於高風險的自我評估依據作成書面文件留存。這意味著即使板橋新創公司日後得出自家招募 AI「非高風險」

的結論，也必須登記並保存該判斷依據。用於執法、移民、庇護、邊境管理等領域的特定高風險系統將登記於資料庫的非公開區段，部分與關鍵基礎設施相關的系統則會於國家層級另行登記。

對於總部設於歐盟境外的企業，亦即像板橋新創公司這樣的韓國公司而言，在投入歐盟市場前，還有一項重疊於所有程序之上的義務：必須以書面形式指定設立於歐盟境內的授權代表。該代表負有保存技術文件與符合性聲明副本 10 年並於主管機關要求時提交的義務。值得一提的是，協調標準的制定已超越 CEN/CENELEC 原本設定的 2025 年 8 月期限，目前仍持續進行中。這項空白亦是導致數位綜合法案延期的背景，且在未來標準完成之前，符合性評估的具體標準難免保持流動狀態。

參考資料

適用日期整理。AI 法生效：2024 年 8 月 1 日。罰款依據條款（第 99 條、第 101 條）生效：2026 年 8 月 2 日。附錄 III 獨立型高風險系統義務全面適用：2027 年 12 月 2 日（因數位綜合法案自 2026 年 8 月 2 日延期）。附錄 I 產品內嵌型高風險系統義務適用完成：2028 年 8 月 2 日（因數位綜合法案自 2027 年 8 月 2 日延期）。通知機構於 AI 法下的指定申請期限：約 2028 年 2 月 2 日。

罰款區間。違反附錄 III 第 4 款招募 AI 等高風險義務（第 16 條等第三章），依據第 99 條第 2 項處以全球年營業額 3% 或 1,500 萬歐元中較高者。違反第 5 條禁止行為則處以 7% 或 3,500 萬歐元中較高者，遠重於本章之義務。

自我檢查清單。風險管理系統是否與品質管理系統連結合一？訓練資料的統計屬性與偏誤評估記錄是否保存？若處理了特殊類別個人資料，是否已單獨將 GDPR 項下的保護措施作成書面文件？技術文件是否已填滿附錄 IV 的所有項目？日誌功能是否確實達到足以重構事件的水準？供部署者使用的說明書中是否包含人類監督方法與日誌解讀方法？介面上是否顯露排名或判定的依據，使部署者得以實際忽視或翻轉？準確度指標是否以數字載明於說明書中？是否有針對資料污染或對抗性攻擊的防禦措施？是否已再次確認混入哪怕任何一項生物識別要素？是否已指定歐盟授權代表？是否已妥善落實歐盟資料庫登記（包含判斷為非高風險的情況）？

在這十二個項目中，能夠自信回答「是」的項目有多少，以及在 2027 年 12 月與 2028 年 8 月到來之前還能增加多少個肯定答案，如今將留給板橋新創公司的開發團隊負責人

去面對。

第13章 部署者的義務

盆唐的一家醫療保健新創企業決定在其德國子公司導入招募審查人工智慧。從經營公司的角度來看，這個案例並沒有什麼特別之處。軟體已經開發完成，公司只需要購買並使用即可。然而，即便是購買並使用的角色，也同樣受到法律約束。《人工智慧法案》將這一角色稱為「部署者」（deployer），並在第26條與第27條中規定了部署者必須遵守的事項。以為「既然不是自己開發的就沒有責任」的想法，在讀完本章之後最好打消。

13.1 遵守使用說明、配置人類監督與日誌保存六個月

部署者義務的起點，在於完全遵循提供者所編寫的使用說明書。提供者會在使用說明中載明該招募審查人工智慧應在何種條件下、輸入何種資料，以及如何解讀其結果。部署者不得在該範圍之外運行該系統。如果該人工智慧僅通過履歷評估用途的驗證，而公司卻將其用於在職員工的升遷審查，那麼在那一刻，公司就已經偏離了指引，一旦發生問題，責任便會轉移到公司身上。

接下來是人類監督的配置。正如在第12章中所見，提供者必須設計系統以允許人類介入。部署者則必須安排實際執行該設計的人員。絕不能隨便指派任何人。該人員必須理解系統憑藉何種依據篩選應徵者，並擁有在發現異常時推翻結果的權限與能力。如果把這項工作交給人資團隊最資淺的員工，讓其僅看著螢幕按下批准按鈕，這不是監督，而只是裝模作樣的監督。執法當局在事故發生後展開調查時，最先質詢的就是這一點：監督者實際上能看到什麼，以及擁有何種權限。

隨之而來的是監控與記錄的義務。在系統運行期間必須觀察異常徵兆，若發現嚴重的誤動作或事故，必須停止使用並通報提供者及相關主管機關。此外，還必須保存系統自動留下的日誌。保存期限至少為六個月。「六個月」這個數字並無談判空間。日後若針對招募結果提出訴訟，公司必須能夠出示當時人工智慧是根據什麼做出何種判斷，才能進行辯護。刪除日誌的公司，等於根本沒有任何可用於辯護的資料。

13.2 對勞工導入人工智慧，事前通知

這正是韓國企業經常忽略的一點。當針對勞工導入高風險人工智慧時，公司必須預先將該事實告知勞工。順序並非在導入後才進行公告，而是在導入前進行通知。

這與韓國國內的慣例截然不同。在韓國國內，引進新的績效考核工具或出勤管理系統時，雖然有時會為員工舉辦事前說明會，但法律並未對此做出強制規定。歐洲在此一點上則有所不同。如果是影響勞工處境的高風險人工智慧（如招募、調派、績效考核、解僱等），導入前的通知即為法律義務。若前述盆唐的醫療保健公司在其德國子公司使用招募審查人工智慧，除了該人工智慧將處理的應徵者之外，有時也必須預先通知現有的勞工代表或勞工委員會。我認為這項通知義務應該由法務團隊而非人資團隊來負責處理。人資團隊容易將通知視為一般的宣傳或公告，但這是法律規定的程序，一旦遺漏本身即構成違法。

13.3 基本權影響評估，應由誰執行

部署者義務中最為繁重的一項是基本權影響評估。這由第27條所規定，通常簡稱為FRIA（Fundamental Rights Impact Assessment）。這是在使用高風險人工智慧之前，預先審視其可能對人的基本權利產生何種影響的程序。

這項義務並非適用於所有部署者。精確劃分適用對象正是本節的核心。身為公務機關的部署者，以及提供公共服務的民間部署者，在選用附錄 III 的高風險系統時，必須進行此項評估。單純的民間企業將高風險人工智慧用於內部業務，並不意味著會自動承擔此項義務。

然而存在兩項例外，這兩項例外在實務上更為重要。附錄 III 第5項 (b)款（用於信用評等的人工智慧）以及 (c)款（用於人壽保險及健康保險之風險評估與定價的人工智慧），即使是單純的民間部署者，也同樣屬於基本權影響評估的適用對象。若一家總部設於首爾的金融科技公司以部署者身分在歐洲使用用於貸款審查的信用評等人工智慧，該公司雖非公務機關亦未提供公共服務，但仍無法規避此項評估。保險公司亦然。若認為自己與公共性質無關、屬於純粹民間企業而可跳過 FRIA，這種判斷在信用評等與保險這兩個領域中是錯誤的。

評估應包含的內容已經明確規定：該系統的使用目的、時間與方式；受影響人員的範圍；與該使用相關的風險；人類將如何進行監督；以及風險實際顯現時將採取何種減緩措施。這五項內容必須在首次使用前整理完成，若使用方式或情境發生重大變更，則須重新更新。完成評估後，須將結果通知市場監督機關。規範規定須使用人工智慧辦公室（AI Office）所編製的問卷表單，但截至我進行檢視時，該表單的最終確定版本尚未公開，實務工作者之間正流通著自行編撰的草案。

與個人資料保護影響評估（即 GDPR 的 DPIA）之間的關係也是需要釐清的部分。第27條第4項規定基本權影響評估旨在補充 DPIA，是「補充」而非「替代」。若公司已經完成 DPIA，可以將該資料引用至基本權影響評估中。如何蒐集與處理應徵者的個人資料已包含於 DPIA 中，因此沒有必要重寫該部分。然而，DPIA 是著重於個人資料保護的評估，而基本權影響評估則涵蓋歧視、受公平審判權等超越個人資料範疇的基本權利。在兩項評估重疊的部分可以共享資料，而在不重疊的部分則須另行填補。如果有公司認為做了 DPIA 就等於完成了 FRIA，那就是誤讀了該條款。

在此處需要釐清一個日期問題。本稿參考的 NotebookLM 資料記載，針對附錄 III 獨立式高風險系統的第26條與第27條義務將於2026年8月2日起全面適用。然而在2026年5月7日，歐洲議會、理事會與執委會就稱為「數位綜合法案」（Digital Omnibus）的修正案達成暫定共識，其中包含將附錄 III 獨立式高風險系統的義務適用時間延後16個月至2027年12月2日的內容。截至我撰寫本稿時，該共識尚未經過歐洲議會與理事會的正式採納及公報刊登，仍處於暫定狀態。也就是說，在8月2日到來之前，有很大可能被正式確定；若果真如此，原先知曉的8月2日適用即告作廢，並移至2027年12月。我建議韓國企業：沒有理由放緩原先為了趕在8月2日而緊鑼密鼓進行的準備。因為若未獲正式採納而8月2日逕自過去，原本規定的日期就會原封不動地維持效力。不過，請務必瞭解該日期確實存在延後的可能性，並請在8月初之前親自確認該修正案是否刊登於歐盟公報。

以下列出本章須確認的事項：

遵守使用說明：不偏離提供者規定的條件與用途。

人類監督：實際配置具有理解與介入能力及權限的人員。

監控與日誌：因應異常徵兆，且日誌至少保存六個月。

勞工通知：針對勞工導入高風險人工智慧須於導入前告知。由法務團隊而非人資團隊負責。

基本權影響評估：公務機關及提供公共服務的民間部署者，在選用附錄 III 系統時為義務對象。即便是純粹的民間企業，信用評等（第5項 b款）與人壽・健康保險風險評估・定價（第5項 c款）亦毫無例外屬於適用對象。得與 GDPR DPIA 共享資料，但不得以其代之。

確認施行日期：附錄 III 獨立式高風險系統的第26條與第27條義務原本預計於2026年8月2日生效，但因《數位綜合法案》共識，有可能延後至2027年12月2日。請親自至公報確認是否已獲正式採納。

讓我們回到盆唐的那家醫療保健公司。在其德國子公司導入招募審查人工智慧之前，公司首先應該問的不是「這項人工智慧有多精密」，而是「我們作為部署者，究竟處於何種立場」，這才是必須先確認的事。

第14章 如何利用延期期間

板橋的一位新創公司合規負責人上週寄了封電子郵件給我。CEN 與 CENELEC 的調和標準都還沒出爐，2026 年 8 月 2 日的高風險系統義務卻將照常生效，他想知道究竟該以什麼為依據來證明符合性。這個問題沒有標準答案。只不過，如何調配目前手中現有的工具，將決定 8 月 2 日早晨的神情。

14.1 調和標準制定延遲的背景

歐盟《人工智慧法》（Regulation (EU) 2024/1689）於 2024 年 8 月 1 日生效，並對高風險 AI 系統提供者賦予了第三章規定的義務。這項義務將自 2026 年 8 月 2 日起全面適用，同日 AI 辦公室裁罰罰款的權限也將同步啟動。違反規定的罰款金額，為上一會計年度全球總營業額的 3% 與 1,500 萬歐元兩者中較高者。等於是一個日期同時開啟了義務與處罰。

在技術上落實這項義務的機制就是調和標準。歐盟執委會於 2023 年 5 月 22 日以決定 C(2023) 3215 向歐洲標準化委員會（CEN）與歐洲電工標準化委員會（CENELEC）發出標準化請求 M/593。若依據該標準開發並驗證系統，便能享有被視為符合本法相關規定的符合性推定（Presumption of Conformity，第 40 條）效益。由於不必逐條對照法條，對提供者而言是最為確實的捷徑。

然而，CEN 與 CENELEC 超出了受託的 2025 年 8 月期限。截至 2026 年 7 月目前為止，標準草案的編撰工作仍在持續進行中。機構內部為了彌補延誤，決議採取了幾項應急措施。凡在審議（Enquiry）表決中獲得通過者，無須經過另外的正式表決即可直接公布草案，並另行組建了小型撰寫小組來處理進度最為落後的六份草案。其目標是在 2026 年第四季前推出第一批標準。其中對應第 17 條品質管理系統（QMS）的 prEN 18286 已推進至審議階段。

在此我想提出的判斷如下：即使標準草案能在 2026 年內出爐，也不意味著能立即帶來符合性推定的效益。執委會在歐盟公報刊登引用文號的審查程序仍未完成，而這項審查相當耗時。因此，認為到了 8 月 2 日拂曉之際，實際上並沒有韓國企業能夠以調和標準為依據主張符合性推定，這樣想會更為穩妥。標準頒布的消息，與能憑藉該標準獲得法律效益的事實，兩者屬於不同的時間點。

執委會並非不知道這段落差。法律已保留彈性，若標準未及時或未以令人滿意的方式開發完成，執委會得透過執行法規採行共通規範（Common Specifications，第 41 條）。實務守則（Code of Practice，第 56 條）亦被明定為在標準獲批前所使用的臨時工具。AI 辦公室表態將把遵守實務守則的紀錄視為計算罰款時的減輕事由，執委會甚至正在評估透過「數位綜合案」（Digital Omnibus）提案為高風險規定的適用提供額外寬限期。這展現出承認延誤並多管齊下準備替代方案以填補空白的姿態。

對韓國企業而言，這項背景所代表的含義非常明確。不必寄望在 8 月 2 日能拿到「標準」這張完美的答案卷。相反地，從現在起就必須準備好在沒有答案卷的情況下，要呈現什麼給閱卷者看。

14.2 缺乏標準時的替代方案：ISO/IEC 42001 與 NIST AI RMF

填補缺乏標準之空白的第一個替代方案是實務守則。依據第 56 條規定的實務守則，被認可為證明符合第 53 條第 1 項（所有通用人工智慧模型提供者之義務）與第 55 條第 1 項（具系統性風險模型提供者之義務）的簡易方法。反之，若不遵循實務守則，AI 辦公室可能會要求提供更詳盡的資訊，因此漠視實務守則反而會增加文書作業負擔。

第二個則是國際標準。ISO/IEC 42001 是用以認證組織層級 AI 管理系統（AIMS）的標準，進入 2026 年以來，獲證案例迅速增加。然而，此處有一關鍵點必須釐清：取得 ISO 42001 證書並不代表會自動享有《人工智慧法》的符合性推定效益。因為符合性推定是僅授予給在歐盟公報上刊登引用文號之調和標準的特權，而 ISO 42001 是排除在歐盟調和程序之外的國際標準。僅持有 ISO 42001 的提供者，在發生爭議時仍須自行承擔舉證責任。我常遇到實務人員在此點產生誤解，將 ISO 證書誤認為法律盾牌。精確地說，應將其視為製作盾牌的材料，而非盾牌本身。

在同屬 42000 系列的標準中，相當值得關注的是 prEN 18286。該標準是針對第 17 條品質管理系統、適用於個別系統層級的調和標準候選案，且已推進至審議階段。若說 ISO 42001 是認證整體組織治理的工具，prEN 18286 則是一旦完成便能取代該位置、範圍更窄且更直接的工具。切勿將這兩項標準混為一談，我認為目前先以 ISO 42001 建立骨架，待 prEN 18286 登載於公報後再將個別系統文件疊加其上的兩階段策略，會是更為穩妥的做法。

第三個位置則是 NIST AI RMF。不過坦白說，《人工智慧法》條文中完全沒有任何提及 NIST AI RMF 之處。該框架係由美國國家標準技術研究所（NIST）於 2023 年 1 月推出的自願性指南，截至 2026 年當前，版本 1.0 仍為現行版本。它既無認證程序，亦無處罰規定。其架構由治理（Govern）、對照（Map）、測量（Measure）、管理（Manage）四個功能組成，而 NIST 本身公開了將這些子項目與 ISO 42001 條文進行對照的對應表，證實兩者重疊之處相當多。Govern 對應 ISO 42001 的第 5、6 節，Map 對應影響評估，Measure 對應監測，Manage 則對應營運管制。

因此，在實務中切實可行的組合如下：將 NIST 的四大功能作為劃分內部管制項目的骨架，並將《人工智慧法》條文作為該骨架所須填補的義務清單。若韓國企業因美國方面的業務而已使用 NIST RMF 架構內部管制，完全不必拋棄該架構重頭來過。只需在一張

對照表上，將《人工智慧法》條文、ISO 42001 條款及 NIST RMF 子類別並列標明於同一個管制項目中即可。與個別運作三個框架並在每次稽核時拿出不同文件相比，單憑這張對照表就能以極少的人力維持運作。

總結而言，本節的判斷如下：目前沒有任何事物能替代調和標準的法律地位。然而，若重疊運用 ISO 42001、NIST AI RMF 以及實務守則，即使缺乏標準，依然能在監管機關面前確保備妥充份的文件厚度。儘管文件厚度並不等同於法律防禦力，但可作為已盡力準備的客觀情狀證據。

14.3 延期所造成的資料治理空白與自願應對

在三節之中，我認為本節最為緊迫。原因非常簡單：資料是日後無法倒退修改的材料。

《人工智慧法》第 10 條要求用於高風險 AI 系統訓練、驗證及測試的資料集必須具備相關性、代表性、無錯誤性與完整性。該條文要求考量資料收集與準備、偏見檢測與緩減，乃至系統實際應用之地理、情境與行為環境。這項義務同樣自 2026 年 8 月 2 日起全面適用，但由於缺乏調和標準，目前尚無關於應以何種數據證明相關性或代表性的官方公認標準。告訴大家具備何種條件才算充分的評分標準表，目前處於空白狀態。

此處的問題不僅僅在於耗費時間。現在所收集的資料將決定未來系統的性能與極限。若將資料治理押後、苦等標準出爐，這段期間累積的資料集本身可能早已在缺乏相關性或代表性的狀態下固化。對法條條文的解讀日後還能追趕補救，但使用具偏見之資料完成訓練的模型，在標準公布後便再也無法逆轉。因此，我將這段空白解讀為「技術上的不可逆性」問題，而非「法律上的不確定性」。這意味著完全沒有等待的理由。

能自願填補之處主要有兩點。第一是通用人工智慧模型實務守則。其中包含撰寫訓練內容公開摘要的義務，這與資料治理直接相關。遵循實務守則會被認可為符合第 53 條與第 55 條的簡易方法，亦可在 2026 年 8 月 2 日之後計算罰款時作為減輕事由。第二則是涵蓋第 9 條風險管理、第 11 條技術文件、第 12 條自動記錄、第 14 條人類監督、第 15 條準確性、強健性與資安防護的自我文件化體系。不能因為缺乏標準就推遲文件化本身。相反地，越是缺乏標準，就越要縝密記錄建立自我標準的依據及依據該標準所作的驗證紀錄，日後監管機關詢問依據時，才能拿得出憑據。

AI 辦公室得向未遵循實務守則之提供者要求更詳細資訊的條文，也是值得深思的一點。該條文可理解為：主動建立標準的企業只需展示其標準即可；反之，完全未建立任何標準的企業，則必須依據辦公室的要求從零開始製作資料。歸根究底，先動起來的一方日後受到的折磨會比較少。

執委會預定於 2026 年 2 月 2 日前推出第 6 條指引，同時也在評估共通規範與數位綜合案寬限方案。雖然必須持續關注此一趨勢，但絕不能顛倒「關注趨勢」與「現即建立資料治理體系」的先後順序。我給板橋那位負責人的回覆也是如此：不要等待標準，而是從現在累積的資料開始，建立自我檢測相關性與代表性的程序。標準只是日後驗證該程

序的工具，並無法替你建立程序本身。

參考資料

日期 2024 年 8 月 1 日，《人工智慧法》生效 2025 年 8 月 2 日，通用人工智慧模型（GPAI）義務開始適用 2026 年 2 月 2 日，預定公布第 6 條指引 2026 年 8 月 2 日，高風險 AI 系統（附錄 III）義務全面適用，AI 辦公室裁罰罰款權限啟動 2027 年 8 月 2 日，附錄 I 產品內建型高風險 AI 系統全面適用，2025 年 8 月 2 日前推出之 GPAI 模型合規寬限期結束

判斷基準 在調和標準刊登於歐盟公報前，不得主張符合性推定。 ISO/IEC 42001 認證係展現組織治理之依據，其本身並不給予法律推定效益。 遵循實務守則既為簡便證明方法亦為罰款減輕事由，未遵循則將導致被要求提供額外資訊。 資料治理非苦等標準公布之對象，而是應自當前所累積資料起即刻實施之對象。

自我核對項目 是否已再次確認自家系統是否屬於附錄 III 高風險類別或 GPAI 類別 是否已檢視 GPAI 實務守則中關於訓練資料摘要之條款並主動加以採用 除建置 ISO/IEC 42001 AIMS 外，是否已將 NIST AI RMF 四大功能整理為管制項目對照表 是否就第 9 條、第 10 條、第 11 條、第 12 條、第 14 條、第 15 條分別備妥內部文件與驗證紀錄 是否已以書面指定歐盟境內授權代表

第五部

被禁止的AI行為

第15章 不得跨越的界線

板橋的一家 HR Tech 新創企業會議室。會中提出了一項在招募視訊面試工具中加入表情分析功能的提案。該功能可讀取應徵者的眼球運動與聲音顫抖，從而對緊張度與誠實度評分。在韓國國內，類似工具已被用於部分大企業的招募流程中，且並沒有直接禁止此舉的法律條文。問題在於，這家公司恰好準備在德國分公司的招募中採用相同的工具。在負責律師確認條文後，會議進入了完全不同的局面。因為在韓國國內合法的工具，在跨越國境的瞬間，變成了可能面臨刑事處罰等級裁罰金的對象。

《歐盟人工智慧法案》（Regulation (EU) 2024/1689，以下簡稱 AI 法案）第 5 條正是引發此類狀況的條文。這部於 2024 年 8 月 1 日生效的法案，以「基於風險的方法」（risk-based approach）為骨幹，將 AI 系統分為四個等級，而第 5 條所涵蓋的，正是其中最高層級、完全不被允許的「禁止之實踐」（prohibited practices）。高風險人工智慧或有限風險系統只要遵守義務即可上市，但若屬於第 5 條所列情形，無論準備多少文件都無濟於事。既不能製造、也不能銷售或使用。因此，實務工作者將此條文稱為 AI 法案的第一道也是最堅固的合規關卡（first hard compliance gate）。其施行日期也最早，自生效 6 個月後的 2025 年 2 月 2 日起便已全面施行；若有違反，依據第 99 條第 3 項，最高可處以 3,500 萬歐元或全球年營業額 7% 中金額較高者之裁罰金。這是整部 AI 法案中最沉重的罰款區間。

15.1 潛意識操縱與利用脆弱性

第 5 條第 1 項 (a) 款旨在禁止使用超出人類意識範圍的潛意識技術，或故意散布操縱性、欺騙性技術，實質扭曲人類行為並顯著損害其判斷能力，進而造成重大損害的情況。(b) 款則禁止利用因年齡、身心障礙或社會經濟狀況而產生的脆弱性，從而導致相同結果的情況。這兩款共同要求的最終是兩項要件：行為的實質扭曲，以及重大損害（significant harm）。必須同時滿足這兩者，違規行為才成立。資料來源指出，輕微的不便或微小的經濟損失並不包含在此，但我認為該界線在實際運作中可能會劃得相當低。從重複出現「可能對健康、安全或基本權利產生嚴重影響的水準即視為重大損害」的敘述來看，執法機關寬鬆放寬此門檻的可能性似乎不高。

這裡有一個容易忽略的要點。條文所關注的並非開發者的意圖，而是系統產生的實際效果。即使板橋新創公司在開發招募工具時完全沒有操縱的意圖，但若該工具實際上模糊了應徵者的判斷力並造成不利影響，該條文仍將直接適用。這是實務中最常被誤解的地方。大家常辯解稱自己沒有惡意，但 AI 辦公室（AI Office）所問的問題截然不同。除了預定用途（intended purpose）外，他們還會一併評估合理可預期的誤用（reasonably foreseeable misuse）。若在電商 App UI 中加入的微誘導（nudge）引導了兒童或高齡使用者進行非自願的付款，即使設計者並未期望此種結果，也難以避免被認定違規。

釜山一家醫療器材公司的經歷也十分相似。該公司在開發針對老年患者推薦保健食品的聊天機器人時，將對話流程設計為重複曝光與緊急性文案。在韓國國內，這或許能被視為一種行銷技巧，但因其瞄準高齡者的脆弱性並扭曲購買行為，在歐盟體系下便構成了違反 (b) 款的疑慮。

15.2 社會評分

第 5 條第 1 項 (c) 款禁止公家機關或民間企業以個人的社會行為或性格特徵為依據進行評分，並因該評分導致個人在與原本情境無關的地方遭受不利待遇，或是遭受與該行為相比過於不利的待遇。雖然人們容易聯想到國家監控並給國民分級的情景，但該條文對民間企業同樣適用。若徵信機構抓取 SNS 貼文或線上行為數據建立綜合信任度分數，並將該分數用於貸款審查或保險核保，便會正面觸犯此條款。

資料來源中附帶了「在符合法定條件的情況下（where the legal conditions are met）」這一前提，但該條件究竟為何尚未確定。我認為這一點必須等待歐盟委員會後續的指引說明。不過實務判斷基準相當明確：評估標的行為與使用該分數的情境是否相互脫節？其結果相較於原始行為是否過於嚴苛？若這兩個問題的答案均為「是」，就必須視為危險訊號。這意味著韓國金融科技（FinTech）業界一直在運作的許多替代信用評估模型，有必要以該基準重新審視。

執法機關在審視此條款時，核心關注的同樣不是開發者的說明，而是系統實際產生的效果。即使設計信用評估模型的一方聲稱僅打算用於貸款審查，但若該分數流向招募或租賃審查等與原始情境截然不同的領域，我認為違規判定依然成立。由於裁罰權自 2025 年 2 月 2 日起即已開放，一旦確認違規，AI 辦公室預計將毫不猶豫地採取制裁措施。

15.3 即時遠端生物辨識：原則禁止與例外

第 5 條第 1 項 (h) 款原則上禁止在公眾得進入之空間（publicly accessible spaces）出於執法目的使用即時遠端生物識別系統。在廣場、地鐵站、體育場等任何人皆可出入的地方，透過 CCTV 即時掃描面部並比對身分的系統，即為典型的規範對象。

例外範圍並不寬廣。必須滿足以下三項要件之一，且還必須跨越「嚴格必要（strictly necessary）」這一門檻：尋找綁架、人口販運、性剝削受害者或失蹤者的定向搜尋；防止對生命或身體安全構成特定、實質且迫在眉睫的威脅或恐怖攻擊；以及搜尋涉嫌觸犯附錄 II 所規定之犯罪且在該會員國可判處 4 年以上監禁刑罰的犯罪嫌疑人。我認為「嚴格必要」這一門檻將被狹義解釋。若能透過其他侵害較小的手段達成目的，則不予認定例外。

若韓國安控業者或智慧城市解決方案企業企圖向歐盟地方政府供應即時人臉辨識 CCTV，此條款將是首要關卡。雖然非出於執法目的、用於商業設施防竊或門市行銷的人臉辨識並不屬於第 5 條 (h) 款本身的適用對象，但仍須另行檢視其設計是否採用事後比對而非即時比對，以及是否觸及「公眾得進入之空間」的定義。由於附錄 II 具體列出了哪些犯罪無法僅憑資料來源確認，因此在簽約前務必對照原文。

從執法機關的角度來看，我認為在觸及隱私與行動自由的條款中，公共場所即時生物辨識屬於份量最重的一環。因此，儘管機關會在公共安全與基本權利之間進行審慎衡量，但對於稍微偏離例外要件的使用行為，預計將施以嚴格的衡量標準。站在與地方政府簽約的韓國企業立場，確認發包單位如何將例外要件文件化會是較為安全的做法，這是為了避免因發包單位判斷錯誤而導致供應商一併承擔責任的局面。

15.4 情緒辨識：在職場與教育機構完全禁止

第 5 條第 1 項 (f) 款禁止在職場與教育機構中投放、提供服務或使用推論自然人情緒的 AI 系統。韓國實務工作者在此經常產生混淆。AI 法案中有兩個處理情緒辨識的條文，且性質截然不同。第 50 條是透明度義務條款，要求在自使用情緒辨識系統時告知使用者。若客服中心使用通話中分析客戶情緒的工具，在第 50 條規範下只需告知即可。相反地，在適用第 5 條 (f) 款的職場與教育機構中，並非告知就能解決。從一開始，使用行為本身就被禁止。這意味著相同的情緒辨識技術，僅因場所不同，就會從申報對象轉變為刑事處罰等級的禁止對象。

例外情況僅以醫療或安全原因（medical or safety reasons）狹隘地開放。監控飛行員或管制員疲勞度的系統常被舉為例外案例。相反地，我認為以測量員工工作投入度或將學生上課專注度分數化為目的者，並不屬於例外。即使包裝成效率或生產力，條文所審視的依然是實際使用情境。板橋新創公司檢討的面試表情分析工具，鑑於招募是進入職場的關卡，將其解釋為屬於該條款的職場範疇較為安全。

執法機關預計不會寬鬆接受醫療或安全目的這項例外。合約或產品說明書中一旦出現「提升效率」或「增加生產力」等字眼，反而可能被視為該系統不屬於例外對象的證據。若是在人力資源管理、教育科技（EdTech）、客戶應對領域銷售服務的韓國企業，首要任務是檢視是否將情緒推論功能設計為可選式開關，以及該選項是否採取了絕不對職場與教育機構客戶開放的結構。

15.5 在韓國可行但在歐盟不可行的領域

在進行諮詢時，我常收到整理「韓國法律未禁止但在歐盟遭禁項目」的要求。在此優先列出三項。

第一是職場內的情緒分析。韓國國內沒有直接禁止此舉的法律。雖然可透過《勞動基準法》或《個人資料保護法》進行迂迴規制，但並沒有禁止情緒辨識本身的條文。歐盟則截然相反。如 15.4 所述，由第 5 條 (f) 款正面予以阻絕。

第二是民間綜合評分。韓國金融科技與電商業界已營運多個替代信用評估模型，且將線上行為數據反映於信用分數的作法在監理沙盒內普遍獲得允許。在歐盟，禁止社會評分的條款不分公私，因此若原封不動導入相同模型，違反第 5 條 (c) 款的疑慮將大幅增加。

第三是面部影像的無差別收集。韓國企業在建立人臉辨識資料庫時，從網際網路或 CCTV 影像中大量抓取影像的方式，只要想辦法符合《個人資料保護法》規定的同意要件，業務本身就不至於被阻斷。然而 AI 法案第 5 條 (e) 款將透過此類無差別網路刮取（scraping）建立或擴充人臉辨識資料庫的行為本身，直接規定為禁止之實踐。

除此三項外，第 5 條亦在 (d) 款禁止僅憑側寫（profiling）或性格特徵評估犯罪可能性系統，並在 (g) 款禁止基於生物識別數據推論種族、政治傾向、工會成員身分、宗教或性傾向的分類系統。基於犯罪相關事實輔助人類判斷的工具仍保留為 (d) 款例外，但僅憑側寫計算風險值的模型則排除在例外之外。我認為韓國評分業者跨越這條界線的案例並不少見。

在此必須深入探討域外適用與授權代表的問題。AI 法案第 2 條第 1 項 (c) 款規定，即使總部設於歐盟境外，只要其輸出成果（output）在歐盟境內被使用，即受本法管轄。即便首爾總部僅在韓國伺服器上運作模型，一旦該成果傳送給德國分公司或法國客戶，便會直接觸犯第 5 條。歐盟境外企業在產品上市前，必須以書面指定設立於歐盟境內的官方授權代表（Authorised Representative），且在韓國國內亦須另外適用《AI 基本法》，因而陷入必須在兩個監管架構下同時管理同項服務的情境。由於韓國《AI 基本法》尚未設立如第 5 條這般的絕對禁止清單，填補兩法間隙的責任最終落到了企業內部合規組織的肩上。

此外還需補上近期變更的一處。歐洲議會與理事會於 2026 年 5 月 7 日達成暫定協議、並於 6 月 29 日獲理事會最終批准的《數位綜合法案》（Digital Omnibus on AI），在第 5 條中新增了兩個項目：生成非同意親密影像（NCII），包括逼真合成或操縱特定人身體部位的所謂「脫衣」（nudify）工具，以及生成兒少性剝削物（CSAM）。凡未取得當事人自由、具體且在充分知情下給予之明確同意即合成其身體的 AI 系統，現已納入第 5 條禁止對象。該規定將自 2026 年 12 月 2 日起適用，違規裁罰金設定為 1,500 萬歐元或全球營業額 7% 中金額較高者，低於其他第 5 條違規行為 3,500 萬歐元的上限。營運帶有深偽（Deepfake）合成功能之影像編輯 App 或角色生成服務的韓國企業，務必在今年內審視此條文。

在第 5 條面前有效的應對方式並非事後補救，而是設計階段的阻絕。預先將禁止事項植入系統架構、演算法、資料收集方式與使用者介面中，其成本遠低於上市後再剔除功能。鑑於條文本身不少地方讀起來較為抽象，建議在初期企劃階段加入諮詢精通 AI 法案與 AI 倫理之律師或顧問的程序，一併檢視貴公司系統的預定用途與可預期的誤用。

參考資料

日期 2024 年 8 月 1 日，AI 法案生效。 2025 年 2 月 2 日，第 5 條全面施行，裁罰權開始。 2026 年 5 月 7 日，數位綜合法案達成暫定協議。 2026 年 6 月 29 日，理事會最終批准。 2026 年 12 月 2 日，NCII 與 CSAM 生成禁止條款施行。

罰款 第 5 條一般違規：3,500 萬歐元或全球年營業額 7% 中金額較高者（第 99 條第 3 項）。 NCII 與 CSAM 生成違規：1,500 萬歐元或全球年營業額 7% 中金額較高者。

禁止項目一覽 (a) 潛意識操縱。 (b) 利用脆弱性。 (c) 社會評分，涵蓋公家與民間。 (d) 僅憑側寫評估犯罪風險；基於客觀事實輔助人類評估者屬例外。 (e) 透過無差別網路刮取人臉影像建立人臉辨識資料庫。 (f) 職場與教育機構內的情緒辨識；醫療或安全目的屬例外。 (g) 基於生物識別數據推論種族、政治傾向、工會成員身分、宗教、性傾向之分類。 (h) 公眾得進入之空間內的即時遠端生物辨識（執法目的），僅具三項狹隘例外。 NCII 與 CSAM 生成，2026 年 12 月新增。

貴公司系統自我檢視清單 貴公司 AI 功能之設計是否朝向模糊使用者判斷力的方向？ 是否有針對兒童、身心障礙者、高齡者等脆弱群體的功能？ 是否有從員工或學生的行為、表情、聲音中推論情緒或狀態的功能？ 是否綜合個人的線上行為進行信任度或風險度評

分，並將該結果用於其他情境？ 是否從網際網路或 CCTV 大量收集人臉影像以建立資料庫？ 合成或編輯身體影像的功能是否在未經同意確認程序的情況下運作？ 是否已在歐盟境內指定官方授權代表？是否已一併整理與韓國國內《AI 基本法》的雙重義務？

第 5 條並非如其他章節所處理的高風險系統義務那般，只要備妥文件並踐行程序即可解決的條文。高風險系統只要辦理登記、接受評估並標示警語，便能開闢留存於市場的道路。然而一旦觸及第 5 條，根本沒有這條路可走。一旦跨越，一切便告結束。板橋會議室裡的表情分析功能，最終決定在德國分公司的招募中剔除，但在韓國國內招募中仍繼續使用。相同的程式碼在跨越國境時於合法與違法之間搖擺不定，這樣的現狀我們還能負擔多久，則是另一個問題了。

第六部

與韓國法律的交會

第16章 AI基本法與EU人工智慧法的雙重管制

就從板橋（Pangyo）一家招募解決方案新創公司的故事說起吧。這家公司開發了一款透過分析履歷與面試影片來評定應徵者排名的AI。他們在供應給韓國國內大企業人資團隊的同時，也正好打算將相同的產品引進位在匈牙利的汽車零件子公司。該公司代表這樣問道：既然在韓國已經通過了《AI基本法》的審查，是否直接在歐洲販售就可以了呢？我當場回答說不是這樣。這兩部法律雖然都將「招募」這項相同的活動列為管制對象，但其要求的文件、程序，甚至是判斷主體均不相同。這家公司目前正處於必須同時滿足兩部法律的處境。第16章討論的就是這個立場，也就是韓國企業實際上處於國內法與歐盟法之間的立場。

16.1 2026年施行修訂版《AI基本法》的兩個層面

韓國的《人工智慧發展與建立信任基礎等相關基本法》（以下簡稱《AI基本法》）本法於2026年1月22日正式施行。此時施行的內容是管制的骨架。高影響人工智慧業者的安全性確保措施義務、生成式AI標示義務、國內代理人指定義務等，均在此時生效。然而，隨著施行細則於2026年7月21日修訂施行，又疊加了另一個層面。本次施行細則修訂的重心並非管制，而是振興。《AI產品與服務確認制度》以及政府採購優惠措施，正是透過這份施行細則予以具體化。

這項架構我整理如下：本法於2026年1月建構了管制的骨架，半年後的施行細則則在該骨架上添補了推動產業發展的血肉。這意味著同一部法律中同時包含了取締的手與推動的手，而這種雙重性正是理解《AI基本法》的首要關鍵。歐盟《人工智慧法》較接近密實編織了處罰與市場准入要件的強制性法規；而韓國《AI基本法》則更偏向於先鋪設最低限度的安全網，再透過採購與認證來引導企業。閱讀條文時若忽略了這種溫差，便會在研讀韓國國內規定時誤以為會具備歐盟式的嚴格度，從而做出錯誤判斷。

韓國政府表示，施行初期將實施至少1年以上的宣導期。這意味著與其立即處以罰鍰，不如先給予糾正建議與改善空間。歐盟實際上並無此類宣導期的概念。自2026年8月2日起，歐盟AI辦公室將針對違反高風險AI與通用人工智慧模型（GPAI）義務的行為直接行使罰款裁量權。兩部法律在同一年並行進入全面適用階段，一邊以宣導開始，另一邊則以罰款開始。這種溫差造成了韓國企業的錯覺。我認為，若將在韓國國內順利過關的慣例直接套用到歐盟，可能會發生一開始就收到罰鍰通知書的情況。

16.2 高影響AI與高風險AI：概念相似但網目不同

兩部法律在採用基於風險的方法（Risk-based approach）這一點上使用相同的語言，但撒網的方式卻有所不同。

《AI基本法》第2條第4款將高影響人工智慧定義為對人的生命、身體安全、基本權利產生重大影響或恐將引發風險的人工智慧，並將其範圍限定為用於總統令規定的11個領域。這些領域包含：能源供應、飲用水生產工序、醫療保健提供及相關系統運作、醫療器材與數位醫療產品的開發與利用、核能設施管理與運作、犯罪偵查與逮捕、對權利義務產生重大影響的決定（如招募與貸款審查）、交通工具與設施及系統的運作與管理、公共服務、教育中的學生評估，以及總統令規定的其他領域。由於列舉的項目有十項，最後一項為總統令授權條款，因此通常稱為11個領域。

歐盟《人工智慧法》則以高風險這一單一概念開啟了兩扇門。一扇門是用於附錄I列出的現有產品安全法規（如醫療器材、汽車、玩具等）之安全元件的AI；另一扇門則是附錄III明確指定特定使用情境的AI，亦即生物識別、關鍵基礎設施、教育、就業、必要民營與公共服務、執法、移民・庇護・邊境管理、司法行政與民主程序。此處有個顯著的要點：歐盟附錄III不僅包含執法，還明確納入了移民・庇護・邊境管理、司法行政與民主程序。韓國國內的11個領域中雖有犯罪偵查，但並沒有直接針對移民、司法行政或選舉相關系統的項目。這代表網目的大小本身就有所不同。若是開發移民審查AI或裁判輔助AI的韓國國內企業，即使在韓國國內法上未劃入高影響領域，向歐盟出口時仍須另行確認是否有可能被歸類為高風險系統。

歐盟法中還有一個韓國國內法所沒有的機制，即第6條第3項的過濾條款。即使是名列附錄III的AI，若不存在對健康、安全或基本權利造成重大危害的風險，或者對決策結果未產生實質影響，該條款便為其開闢了排除於高風險之外的途徑。這項過濾機制對企業而言是喘息空間，但對執法當局而言則是判斷產生分歧之處。我對該條文的解讀如下：並非只要列入附錄III就會自動成為高風險，而是意味著還有一扇必須由企業自行證明無實質影響力的門。韓國國內法中沒有這種例外過濾機制。只要屬於11個領域，即直接判定為高影響人工智慧。這意味著雖然沒有避開的出口，但判斷標準明確；而歐盟雖然門敞開著，卻必須承擔自行證明才能通過該門戶的負擔。

義務的沉重程度也不同。歐盟高風險AI提供者必須具備風險管理系統（第9條）、資料治理（第10條）、技術文件（第11條）、自動記錄（第12條）、人為監督設計（第14條）、正確性・堅固性・資安（第15條），並完成第三方合格評定、CE標誌及歐盟資料庫註冊後，方可進入市場。這是一種強制的事前認證體系。韓國《AI基本法》雖然要求高影響AI業者履行安全性與可信度確保措施義務，但其方式帶有較濃厚的自主風險管理、文件化、使用者通知等努力義務性質；即使違反，也是經過糾正建議與改善命令後再處以罰鍰的階段性結構。這意味著比起事前認證，韓國更接近事後確認與改善要求。在我看來，如果說歐盟是在門口檢查文件的方式，那麼韓國則是先允許進入、待出現問題時再要求修改的方式。哪一種方式正確屬於政策判斷問題，但從實務工作者的角度來看，結論是必須分別針對歐盟產品制定事前認證時程，並針對韓國國內產品制定事後因應體系。

附錄III高風險AI系統的義務自2026年8月2日起全面適用，附錄I產品內建型則於2027年8月2日完成適用。韓國國內高影響AI義務則自2026年1月22日起已生效。僅就時間點而言，韓國領先了半年以上，但我再次強調，領先並不代表更為嚴格。

16.3 深偽技術：讓肉眼識別還是讓機器讀取

兩部法律雖然都正面處理深偽技術，但要求的標示方式著重點不同。

《AI基本法》第31條將透明度義務構成為三項：高影響或生成式AI正在運作的事實之事前告知、身為AI生成物事實之標示，以及對深偽成果物之個別告知與標示。其核心在於該標示必須能透過使用者的感官予以確認。法律要求必須是能透過視覺或聽覺、或使用軟體輕鬆確認的方法，且為考量主要使用者的年齡、身體或社會條件而能明確識別的方法。在我看來，在螢幕一角加上浮水印，或是在播放開始處以字幕呈現告知文字的方式即屬於此類。這是一種讓人透過眼睛看或耳朵聽來識別的方式，即所謂的「肉眼識別型標示」。

歐盟第50條的脈絡則不同。第50條第2項要求製作合成語音、影像、影片、文本之系統的提供者，須以機器可讀格式（machine-readable marks）標示其產出物。這指的不是肉眼可見的字幕，而是偵測軟體能在檔案內部讀取出的浮水印或元資料簽章等技術。此標示甚至附帶了必須具備有效性、可跨平台互操作性、堅固性與可信度的標準。第50條第4項更另外要求製作深偽技術的散布者，必須公開該圖像・語音・影片為人為生成或操縱的事實。總結來說，歐盟等於是雙重要求機器讀取的標示與向人類告知的通知，而韓國則偏重於人類可識別的單一標示。

這種差異分化了技術投資的方向。若僅看韓國國內市場，透過螢幕字幕或語音提示即足以履行義務。但若要將相同的內容推向歐盟市場，就必須另外加上浮水印技術或內容來源標準（可聯想到如C2PA這類體系）。以板橋新創公司製作的AI虛擬化身廣告影片為例，韓國國內用只需在螢幕下方加上一行AI生成標示字幕即可；但若要將同一部影片上傳至歐洲法人頻道，則意味著必須在該影片檔案本身嵌入機器可讀取的簽章。這不再是多加一行字幕的問題，而是要在內容製作流程中另外嵌入獨立模組的問題。

適用時間點亦不一致。韓國國內第31條義務自2026年1月22日起已然生效。歐盟第50條自2026年8月2日起適用，在此之前已於市場推出的生成式AI系統之標示與偵測義務獲得展延至2026年12月2日，且在此之前製作的內容本身並無追溯標示的義務。在當前這個時間點，韓國企業可以整理為：國內標示義務應已在履行中，而歐盟標示義務則處於準備階段。

16.4 韓國國內代理人與歐盟授權代表：有門檻與無門檻的差異

讓境外業者在境內設立代理人的概念，兩部法律是一致的。然而，在誰須負擔該義務上則出現了決定性的分歧。

《AI基本法》第36條及施行細則第29條規定，在韓國國內無住所或營業所的人工智慧業者中，僅有達到一定規模以上者才須指定韓國國內代理人。其規模標準為：前一年度營業額達1兆韓元以上；或人工智慧服務部門前一年度營業額達100億韓元以上；或截至前一年度底直前3個月內韓國國內使用者人數每日平均達100萬人以上。只要跨過這三者之一就必須設立代理人，但反過來說，低於此門檻的海外業者則豁免國內代理人義務。國內代理人的職責亦相當明確：代理提交大規模人工智慧安全性確保措施結果、申請高影響人工智慧確認，以及協助履行高影響人工智慧安全性與可信度確保措施。

歐盟方面則完全沒有這種營業額或使用者人數的門檻。第22條要求高風險AI系統提供者、第54條要求通用人工智慧模型（GPAI）提供者，在進入歐盟市場前，均須以書面委任狀指定歐盟境內的授權代表（Authorised Representative）。無論規模大小、營業額多少，只要將高風險AI或通用AI模型推向歐盟市場，無一例外皆須設立代表。唯一的例外是不帶有系統性風險的開源通用AI模型。代表負責確認合規文件並保存10年，配合市場監督當局與AI辦公室的要求，並處理歐盟資料庫註冊等程序。若提供者違法，代表甚至擁有終止委任並立即通報當局的權限。

從板橋新創公司的角度來看這種差異，情況如下：若這家公司打算將招募AI作為高風險系統推向歐洲市場，即使是營業額尚未達100億韓元的早期新創公司，也必須設立歐盟代表才能進入市場。反之，若歐洲某家AI企業進入韓國販售招募AI，只要該公司在韓國國內的服務營業額或使用者人數未達施行細則的門檻，即可在不設立國內代理人的情況下開展業務。在我看來，在以規模為基準撒網的韓國，與僅憑活動性質撒網的歐盟之間，越是規模較小的韓國企業，承受來自歐盟方面的負擔就相對越發沉重。

請一併關注罰款結構。歐盟規定，違反禁止行為時，處以全球年營業額7%或3,500萬歐元中較高者；違反其他高風險及通用AI模型義務時，處以3%或1,500萬歐元中較高者。該裁罰權針對禁止行為部分自2025年2月2日起生效，其餘義務部分則自2026年8月2日起生效。若沒有代表，將根本缺乏因應此罰款風險的對口，其結果實際上與被驅逐出歐盟市場無異。

16.5 AI產品與服務確認制度與採購優惠

本節討論的不是管制，而是誘因機制。2026年7月21日施行的修訂版施行細則具體化了「AI產品與服務確認制度」。其架構為：企業向韓國智慧資訊社會振興院轄下的韓國軟體產業協會（KOSA）申請確認後，由韓國資訊通信技術協會（TTA）對人工智慧系統的實際使用水準進行技術審查，並由KOSA向通過審查的產品與服務頒發確認書。獲得確認書的AI產品與服務將自2026年8月起在政府採購市場獲得優惠。

歐盟《人工智慧法》中並沒有這種採購優惠的概念。歐盟選擇透過合格評定與CE標誌鎖定市場准入門檻，跨過該門檻的產品則在無額外優惠的情況下於市場競爭。相反地，韓國只要滿足最低限度的安全要件，市場准入本身相對開放，並透過政府採購這項明確的誘因，引導企業自發性地完成確認程序。「以鞭子鎖門的歐盟」與「以紅蘿蔔引導排隊的韓國」，兩者的對比在此顯得格外清晰。

我向實務工作者提出如下建議：若是針對韓國國內公共機構進行業務開發的AI企業，請勿將取得此確認書視為規避管制的措施，而應作為業務策略來應對。另一方面，若是針對歐盟市場的產品線，這份確認書並無法替代歐盟的CE標誌或合格評定。韓國國內確認書是僅在韓國國內採購市場中具有意義的標記，在歐盟市場則必須另行從頭開始走附錄III的程序。我曾親眼見過有人將這兩者混為一談，竟拿著韓國國內確認書提示給歐盟買家的失誤。

參考資料

雙重管制對照表

項目 / 韓國《AI基本法》 / 歐盟《人工智慧法》 管制性質 / 本法 2026-01-22 管制骨架，施行細則 2026-07-21 振興 / 強制性事前認證，違反時立即裁罰 核心概念 / 高影響人工智慧（第2條第4款，總統令11個領域） / 高風險AI（第6條，附錄I・III） 例外過濾 / 無，符合11個領域即確定 / 第6條第3項，若無實質影響可予排除 領域範圍 / 能源・飲用水・醫療・核能・犯罪偵查・招募貸款・交通・公共服務・教育等 / 除上述項目外，尚包含移民・庇護・邊境管理、司法行政・民主程序 深偽標示 / 第31條，肉眼・聽覺可識別之告知・標示 / 第50條，機器可讀取之浮水印 + 個別告知 代理人指定標準 / 第36條・施行細則第29條，營業額1兆・AI營業額100億・使用者100萬三者擇一 / 第22條・第54條

，無門檻，高風險・通用AI模型全體 確認・採購制度 / AI產品・服務確認制度（KOSA核發，TTA審查），2026-08 採購優惠 / 無此制度，僅存在CE標誌與市場競爭

檢核表：即將同時在歐盟與韓國國內推出的團隊

一、是否已分別標明自家產品是否屬於韓國國內11個高影響領域及歐盟附錄III項目。 二、是否已檢討是否有機會透過歐盟第6條第3項過濾條款排除於高風險之外，並將該判斷依據留存為文件。 三、是否已分別在生成的內容中植入韓國國內用的肉眼標示與歐盟用的機器可讀取浮水印。 四、是否已計算營業額與使用者人數門檻，以確認是否須指定韓國國內代理人。 五、若欲進入歐盟，無論營業額規模為何，是否已透過委任狀完成歐盟代表的指定。 六、是否已將韓國國內確認書與歐盟合格評定視為不同程序進行分別管理。

板橋的那位代表最終在牆上並排貼上了兩份合規日曆。一份是韓國國內高影響確認與採購優惠申請時程，另一份則是歐盟代表委任與附錄III合格評定時程。相同的產品、相同的公司，日曆卻有兩份。這家公司在下一季會先填滿哪一份日曆，我也仍在持續關注中。

第17章 與個人資料保護的交會

我接到一家位於板橋的醫療影像 AI 公司的聯絡。他們打算向德國幾家醫院提供輔助診斷肺癌的軟體，但當地代理人寄來了一份問卷。內容詢問：訓練所使用的胸部 X 光影像從何處取得？是否獲得患者同意？若未獲得同意，是依據什麼法律依據進行處理？公司代表問我：我們以為只要準備好因應 AI 法案即可，為什麼一直提到個人資料？我這樣回答：AI 法案與個人資料管制從一開始就不是能切割開來的兩件事。從用鏡頭拍攝人像、收集診療紀錄、訓練對話的那一刻起，就已經是個人資料問題，而 AI 法案只是在此之上再疊加了一層規範。

17.1 GDPR 與個人資料保護法的同時適用

AI 法案（Regulation (EU) 2024/1689）於 2024 年 8 月 1 日生效。該法第 2 條第 7 項明確規定：關於個人資料、私生活及通訊保密保護的歐盟法律，對於因 AI 法案所規範之權利及義務而處理的個人資料同樣適用，且 AI 法案不影響 GDPR（Regulation (EU) 2016/679）或個人資料保護指令（2016/680）。簡單來說，AI 法案並未取代 GDPR，而是疊加在 GDPR 之上運行。

兩部法律的側重點有所不同。GDPR 關注個人資料如何被收集、處理與傳輸，以及當事人的權利是否受到保障；AI 法案則關注系統本身的安全性、效能與風險管理。然而，兩者皆強調透明度、問責性、資料品質與人工監督，因而存在許多重疊之處。最典型的例子就是自動化決策。當 AI 系統自動做出與個人相關的決策時，會觸及 GDPR 第 22 條；若該系統被歸類為高風險，則會同時觸發 AI 法案的額外義務。這意味著單一行為將同時受到兩部法律的規範。

這裏正是韓國企業經常忽略的盲點。進軍歐盟的韓國公司實際上面臨著三層防護網的限制。第一層是 GDPR。由於第 3 條的域外適用條款，即使在歐盟沒有據點，只要向位於歐盟境內的人員提供商品或服務，或監控其行為，GDPR 就會直接適用。第二層是韓國的《個人資料保護法》。作為韓國公司，自然必須遵守國內法，而諸如跨境傳輸規範等條款同樣適用於流向歐盟的資料。第三層則是 AI 法案。依據第 2 條第 1 項 (c) 款，只要 AI 系統的產出物在歐盟境內被使用，即納入適用範圍，且該 AI 法案內部又嵌有個人資料相關條款，最終形成必須通過三道個人資料管制關卡的架構。這也是我給那位板橋公司代表的答覆：個人資料問題並非獨立於 AI 法案之外，而是早已包含在 AI 法案內部。

罰款機制同樣呈現雙重結構。AI 法案採分階段實施：自 2025 年 2 月 2 日起，若違反第 5 條所禁止的 AI 實務做法，最高可處以 3,500 萬歐元或全球總營業額 7%（取其高者）的罰款。自 2026 年 8 月 2 日起，針對違反高風險人工智慧系統與通用人工智慧模型義務者，最高可處以 1,500 萬歐元或營業額 3% 的罰款，屆時 AI 辦公室的罰款處分權限亦將正式發動。若同時重疊違反 GDPR，則會另行加處最高 2,000 萬歐元或營業額 4% 的罰款。這代表若分別違反兩部法律，就必須各自繳納罰款，一次資料處理的失誤可能在帳簿上記下兩筆罰金。

實務上也有一處令人樂見之處。AI 法案第 27 條要求部署高風險人工智慧系統的一方進行基本權影響評估（Fundamental Rights Impact Assessment, FRIA），並明定該評估可補充 GDPR 第 35 條的個人資料保護影響評估（Data Protection Impact Assessment, DPIA）。這意味著已在執行 DPIA 的公司，只需在其架構上疊加 AI 法案所要求的項目，即可整合為單一評估。不必準備兩套文件，對合規負責人而言無疑是一絲喘息之機。

17.2 跨境傳輸與適當性認定

AI 法案並未針對資料跨境傳輸設立新條款。當個人資料透過 AI 系統傳出歐盟境外時，GDPR 第 5 章的跨境傳輸規定將完全、原封不動地適用。因此，韓國企業只需思考一個問題：在我們目前的資料流向中，是否有符合 GDPR 第 5 章的合規途徑？

最便捷的途徑是適當性認定。若歐盟委員會認定特定國家的個人資料保護水準與歐盟相當，則前往該國的資料傳輸無需標準合約條款或額外保護措施即可獲準。韓國於 2021 年 12 月 17 日獲得了這項適當性認定。然而，此處有一點必須精確釐清：市面上常有資料宣稱韓國已獲得歐盟全面批准個人資料傳輸，但這並非準確的說法。該認定僅針對韓國個人資料保護委員會所監督的領域。受金融委員會監督的個人信用資訊 - - 即依據《信用資訊法》處理的金融業資料 - - 並未納入此適當性認定的範圍內。銀行、信用卡公司或金融科技公司若欲將歐盟客戶的信用資訊傳輸至韓國，至今仍須備妥標準合約條款（Standard Contractual Clauses, SCC）等額外保護措施。若如醫療影像 AI 公司般非屬金融領域，固然可直接享有適當性認定的效益；但若是開發金融科技或信用評等 AI 的公司，則務必確認此一例外規定。

在 AI 法案中亦可間接看到跨境傳輸的痕跡。第 2 條第 4 項規定，若第三國的公共機關或國際組織在執法或司法合作的架構下使用 AI 系統，且該機構就個人基本權保護提供適當的保護措施（adequate safeguards），則可免除 AI 法案的適用。此處直接引用了 GDPR 所使用的概念用語。這意味著兩部法律看待跨境傳輸的用語源於相同根源，在實務上也代表直接套用 GDPR 第 5 章的判斷標準亦不致偏離過遠。

資料從韓國再次傳輸至第三國的情形也極易被忽略。得益於適當性認定，資料從歐盟傳至韓國的路徑十分順暢。然而，若韓國公司將該資料再次轉傳至越南的開發團隊或美國的雲端服務，則這第二次傳輸與韓國獲得的適當性認定無關。必須重新檢視該國是否獲得歐盟單獨的適當性認定，或是否需要採取如 SCC 等保護措施。由於 AI 開發過程中資料跨越數國伺服器與人員實屬常態，將這第二次傳輸路徑繪製成表留存較為安全。

17.3 訓練資料中所含個人資料的處理依據

AI 法案第 10 條針對高風險人工智慧系統的開發者課予資料治理義務。要求用於訓練、驗證與測試的資料集必須符合目的、具代表性、低錯誤率且具完整性。然而，這項要求絕不意味著當資料包含個人資料時，可免除 GDPR 所要求的合法處理依據。資料品質要求與處理依據要求是並存的獨立標準。無論資料多麼乾淨、多具代表性，若最初缺乏收集該個人資訊的法律依據，違反 GDPR 的事實依然存在。

GDPR 第 6 條第 1 項列出了處理個人資料的六項依據，而在 AI 模型開發實務中最常使用的依據為 (f) 款「正當利益」（legitimate interest）。歐洲資料保護委員會（EDPB）於 2024 年 12 月 17 日採納的第 28/2024 號意見書中，整理了如何將此依據應用於 AI 模型的開發與部署階段。這份應愛爾蘭資料保護委員會（DPC）要求提出的意見書，雖然承認正當利益可作為 AI 模型開發的法律依據，但要求分三個階段進行測試與衡量：追求的利益是否合法、明確且真實存在（非投機性質）；該處理對於實現該利益是否切實必要；以及該利益是否足以權衡並勝過資料主體的權利與自由。EDPB 以協助使用者的對話型代理人或改善資安為例，指出此類服務雖可以正當利益為依據，但公司必須證明該處理的必要性，以及確實進行了權利衡量。

韓國的情況則截然不同。《個人資料保護法》將「同意」視為處理依據的核心。不過，第 28 條之 2 規定，若出於統計製作、科學研究、公益性記錄保存等目的，則允許在未取得資料主體同意的情況下處理假名資料。韓國個人資料保護委員會於 2025 年 8 月發布了生成式人工智慧相關個人資料處理指引，開始明確規範如何將此假名資料條款應用於 AI 訓練資料。因此，歐盟是以正當利益的利益衡量邏輯，而韓國則是以同意與假名化處理這兩道門，分別嘗試為 AI 訓練資料中的個人資料處理提供合法化基礎。由於這兩種依據的要求並非完全重疊，因此在歐盟以正當利益為依據收集的資料，是否能直接依據韓國標準重複使用，仍需各公司單獨進行檢視。

這個領域至今仍在持續變動。EDPB 意見書發布僅約半年，個人資料保護委員會的指引發布時間則更短。對於如何將利益衡量邏輯應用於透過網路爬蟲收集的大規模訓練資料，以及假名資料條款是否涵蓋至 AI 微調（fine-tuning）階段等問題，目前尚無確定的答案。在現階段，我的看法如下：大方向上歐盟偏向利益衡量，韓國則劃分為同意與假名化處理，但詳細標準未來仍有數次調整空間。這並非建議拖延決策，而是指應將目前

建立的依據作成書面紀錄留存，並在每次新指南出爐時重新對齊檢視。

AI 法案第 59 條額外提供了一個例外途徑。若是在監管沙盒（Regulatory Sandbox）內，為了公共安全或公共衛生等重大公共利益而開發 AI 系統，則可將原本出於其他目的收集的個人資料重新用於訓練。然而，其條件極為嚴格：必須無法透過去識別化或合成資料達到目的、資料必須存放在隔離環境中、不得分享至沙盒外部，且在參與結束後必須刪除。條件如此繁複本身即是一個訊號，顯示該例外難以作為日常營運的商業依據。

參考資料

日期 2021 年 12 月 17 日：歐盟採納對韓國之 GDPR 適當性認定（不含金融委員會監督之個人信用資訊） 2024 年 8 月 1 日：AI 法案生效 2024 年 12 月 17 日：EDPB 採納第 28/2024 號意見書 2025 年 2 月 2 日：AI 法案第 5 條禁止事項規範生效（最高 3,500 萬歐元或營業額 7%） 2025 年 8 月：個人資料保護委員會公布生成式 AI 個人資料處理指引 2025 年 8 月 2 日：GPAI 提供者義務生效 2026 年 8 月 2 日：高風險人工智慧系統義務全面生效，AI 辦公室罰款處分權限正式發動（最高 1,500 萬歐元或營業額 3%）

檢核項目 訓練資料中是否包含個人資料？若有，是否已將 GDPR 第 6 條之依據做成書面紀錄留存？ 是否混雜了特種個人資料（第 9 條規範對象）？ 是否已分別單獨確認從歐盟至韓國、以及從韓國至第三國的傳輸路徑？ 若屬金融領域資料，是否係採用 SCC 而非適當性認定？ 是否有將 DPIA 與 FRIA 整合為單一評估的空間？ 歐盟代理人是否同時承擔個人資料相關文件保管及查詢回應之職責？

第七部

執行：建構合規計畫

第18章 90 天啟動計畫

在板橋的一家人力資源管理軟體公司，合規團隊主管被召進了執行長辦公室。一家歐洲客戶在簽約前前來詢問該公司對《人工智慧法案》的因應現況，執行長問主管：「90 天內我們該從哪裡、準備什麼？」主管手裡拿著包含上百個條文的法規集與附錄，一時之間毫無頭緒，不知該從哪一頁翻起。

對於這個問題，答案在於「順序」。《人工智慧法案》不會一股腦地全盤壓向企業。每個條文的施行日期不同，罰鍰的重量也 different。90 天的時間雖然不足以完全遵守法案的所有條文，但用來建立處理順序卻綽綽有餘。本章將該順序分為四個階段：建立盤點清單、進行分類、衡量差距，以及設定優先順序。

18.1 建立 AI 盤點清單：連看不見的 AI 也要按部門清查

先明確說明一點：《人工智慧法案》中沒有任何一個條文要求建立盤點清單。然而，這項工作之所以放在 90 天計畫的首位，是因為無論是第 6 條的高風險分類、第 11 條的技術文件，還是第 52 條的辦公室通知，若不知道自己公司有哪些人工智慧在何處運作，就根本無法著手。這並非條文直接規定，而是為了遵守條文所必須經歷的實務步驟。我稱之為「實質上的義務」。

在盤點調查中，企業特別容易遺漏的並非肉眼可見的 AI 產品，而是隱形的 AI。公司宣傳銷售的人工智慧服務大家都知道，問題在於旁邊的附屬功能：人資團隊用來篩選履歷的商業軟體自動推薦功能、客服中心聊天機器人附帶的回答推薦引擎、行銷團隊使用的廣告最佳化工具目標定位演算法、財務團隊的異常交易偵測功能。這些東西既非公司親自開發，平時也不被稱為人工智慧，但大多符合《人工智慧法案》的定義。因此，調查不能僅止於開發團隊，還必須按部門全盤清查，涵蓋人資、行銷、客戶支援、財務乃至法務。

實際詢問時，各部門的反應各不相同。若問行銷團隊「有沒有使用人工智慧？」，通常會先得到「我們沒在用」的回應。但如果追問「使用哪種廣告平台？」，情況就截然不同了。自動決定廣告曝光順序的功能、將偏好相似的顧客歸類成群的功能，往往早已內建於該平台中。行銷團隊不稱之為人工智慧，只當成一般的廣告工具。因此，必須將提問從「是否有使用人工智慧？」改為「是否有自動排名、推薦或篩選的功能？」，才能獲得正確答案。法務團隊也有需要另外確認的事項。合約中偶爾會寫明對方提供的服務包含人工智慧功能，但該條文往往只留在法務團隊的抽屜裡，實際使用該服務的事務部門甚至不知道其中含有人工智慧。

調查方法無須過於繁複。可以從要求各部門主管寫下「業務中所使用的工具裡，凡具有自動判斷、推薦或篩選功能者請全部列出」開始。除了公司內部開發的系統外，外部採購的 SaaS 工具及其隨附的人工智慧功能也必須納入。引進時並不知情、後來才發現是以人工智慧為基礎的案例，在實務中屢見不鮮。

每個項目需要記錄的資訊有五項：用途為何、處理何種資料、屬於內部開發還是外部引進、與歐洲市場或來自歐洲的資料是否有接觸點，以及是否掌握訓練所需的運算規模。最後一項雖僅適用於自行訓練模型的企業，但若未事先記錄，日後在判定是否屬於通用

人工智慧模型時，就必須從頭重新調查。

這份盤點清單在實務上扮演兩個角色。其一是作為下一節「分類與角色認定」的素材；其二是作為法律所要求的外部登記基礎，例如第 49 條針對高風險系統要求的歐盟資料庫登記。登記是向公司外部提出的盤點清單，而目前製作的則可視為內部版本。

18.2 依系統進行分類與角色認定

當所有項目都列入盤點清單後，便將各項目對照第 2 章設定的四個層級：屬於被禁止的行為、高風險、須遵守透明度義務，還是最小風險。評定應徵者優先順序的工具極可能落在附錄 III 的僱用領域；廣告目標定位工具大抵屬於最小風險；而製作深度偽造（Deepfake）或假裝與真人對話的聊天機器人，則屬於須遵守透明度義務的範疇，歸第 50 條管轄。

若為自行訓練通用人工智慧模型的企業，則需額外進行一項確認：該模型是否屬於透過大量資料進行自監督學習並執行廣泛任務的通用人工智慧模型（GPAI），以及累積訓練運算量是否超過 10 的 25 次方次浮點運算（FLOPs），進而被歸類為具系統性風險的模型。該基準的測量方法論目前尚未完全定案。由於尚存不確定性，營運自有模型的企業若遇到是否超過此數值的模糊情況，建議採取保守態度，視為接近具系統性風險並進行因應，會較為安全。

評定層級後，再將第 1 章提到的角色表套用到各項目上：在這個系統中，我們是提供者、部署者，還是兩者皆是？若僅將外購的招聘工具用於自身業務，則屬於部署者；若將自行開發的模型提供給歐洲市場的下游業者，則成為該模型的提供者。此處有一個實務人員經常混淆的盲點：如果將通用模型拿來進行微調（Fine-tuning）或修改結構，根據其變更幅度，原本屬於部署者的公司可能會變成修改後新模型的「新提供者」。以為只是區區幾行微調的程式碼，在法律上卻等於確立了新提供者的身份。我看過許多公司遺漏了這一點。只要動過模型，就必須先確認該修改是否會導致提供者身份轉移。

來看一個角色判定容易混淆的實際場景。位於汝矣島的一家金融科技公司，引進海外公開的開源語言模型用來潤飾信用評等文案。若直接使用原始模型，該公司屬於部署者，而出產模型的海外公司才是提供者。然而，若該公司利用韓國國內的貸款審查資料重新訓練該模型，使其甚至能給出審查等級，情況就截然不同了。若變更幅度巨大，這家金融科技公司本身就會變成該修改後模型的提供者，並須一併承擔屬於附錄 III 領域「信用評等」的高風險系統提供者義務。僅憑使用開源模型並不能免除義務，關鍵在於修改了什麼以及修改幅度有多大。

本節產出的成果是一張表格：這張表格為盤點清單上的每個系統與模型標註了一個層級與一個角色。這張表格將作為下一節「差距分析」的輸入資料。

補充一點：歐盟委員會原本預計於 2026 年 2 月前提出第 6 條高風險分類基準的指南，但實際草案延至 2026 年 5 月 19 日才釋出，公眾意見徵詢則於 6 月 23 日結束。這比預定時程延後了近三個月。這意味著分類基準的詳細解釋尚未最終定案，因此建議對處於模糊邊界的系統採取保守判斷，直到該指南的最終版本出爐為止。

18.3 差距分析

備齊盤點清單與分類表後，便將該表與條文並列，衡量缺乏了什麼。這被稱為「差距分析」（Gap Analysis）。這個用語同樣未明文寫在法律條文中。第 56 條針對通用模型提供者訂定了《實踐守則》（Code of Practice），並規定不遵循該守則的提供者須透過其他方式證明合規，而「其他方式」的實體，正是將自己目前的措施與守則要求的措施進行對比的工作。遵循實踐守則的一方在舉證負擔上會輕鬆許多。

若為被歸類為高風險的系統，需要對照的條文清單相當長：是否具備第 9 條的風險管理體系？是否落實第 10 條的資料治理（即檢視訓練資料的品質與偏見）？是否已撰寫包含系統結構與開發過程的第 11 條技術文件？從第 12 條的自動記錄功能、第 14 條可由人工干預監督的設計，到第 15 條的正確性、穩健性與網路安全保障，逐一對照。凡是缺乏的部分，即為「差距」。

若為通用模型提供者，則需檢視是否備齊以下四項：技術文件、提供給使用該模型的下游業者的資訊、著作權法遵循政策，以及訓練所用內容的公開摘要。若被歸類為具系統性風險的模型，則還需再加上模型評估、事故通報與網路安全。

差距分析的產出物為「差距清單」以及用以填補該差距的「路線圖」（Roadmap）。內容應區分列出：需要技術改善的項目、需要建立新流程的項目、只需整理文件的項目，以及需要招募新人員或進行培訓的項目。實際執行實踐守則中所載措施的紀錄，在依據第 101 條計算罰鍰時可被視為減輕事由。僅找出差距後置之不理，與記錄下尋找並填補差距的過程，兩者在日後面對罰鍰時會產生相當大的差異。

讓我們回到前面提到的招募排名工具。該工具已從盤點中篩選出，被歸類為附錄 III 僱用領域，且公司已被認定為部署者。現在將此工具對照第 9 條至第 15 條：沒有風險管理體系；沒有檢視履歷資料偏見的紀錄；技術文件僅掌握在銷售工具的供應商（Vendor）手中，公司手頭並無此文件；雖有可由人工覆核最終決定的流程，但未予以文件化。這四行便直接成為差距清單的前四行。部署者的義務雖然比提供者輕微，但並非全然沒有，因此應從向供應商索取技術文件副本並取得偏見檢視結果開始做起。

18.4 設定優先順序：依日期與裁罰重量

整理至此，需要一個標準來將收集到的盤點清單、分類與差距清單排成一列。我採用兩個軸線：何時開始施行、違規時處以多少罰鍰。將這兩個軸線交叉比對，順序自然就會浮現。

在建立此順序前，有一點必須先釐清。開始調查時參考的 NotebookLM 資料，曾假設高風險人工智慧系統義務將於 2026 年 8 月 2 日起全面適用。以該法案原始訂定的時程來看確實如此。然而，透過搜尋確認的最新狀況有所變化。經歐洲議會於 6 月 16 日、理事會於 6 月 29 日通過，並於 7月 8 日簽署的《數位綜合法案》（Digital Omnibus），將附錄 III 獨立型高風險系統的義務施行日期從 2026 年 8 月 2 日延後十六個月至 2027 年 12 月 2 日。內置於產品中的附錄 I 高風險系統義務則從 2027 年 8 月 2 日順延一年至 2028 年 8 月 2 日。該修訂案將於刊登於官方公報三天後生效，但在撰寫本書時仍處於等待刊登狀態。雖然預計最晚在 7 月內刊登，但萬一刊登延期，原本預定的 8 月 2 日全面適用時程便會重新復活。由於此處尚存不確定性，建議在 8 月初再次確認官方公報。

此一延期大幅改變了優先順序。順序如下：

第一順位為第 5 條的「被禁止的行為」。該條文已於 2025 年 2 月 2 日起施行，罰鍰高達 3,500 萬歐元或全球總營業額的 7%，為本法案中最重的處罰。此外，《數位綜合法案》還在禁止清單中新增了如未經同意合成生成性別/性愛影像等項目。由於該條文已經生效，必須立即在 90 天計畫的第一週內，確認自家人工智慧中是否有任何觸及此紅線的情形。

第二順位為第 50 條的「透明度義務」，包含表明為聊天機器人的告知，以及深度偽造與合成內容的標示。此項義務不受《綜合法案》影響，仍按原定時程自 2026 年 8 月 2 日起施行。同時，AI 辦公室的普遍罰鍰處分權限亦於同日生效。可以說是義務與執法權限在同一天同步登場。從撰寫本書的現在到施行日期所剩時間不多，因此在 90 天計畫中，實際工作量應最大程度地分配給這個項目。

第三順位為「通用人工智慧模型義務」。該義務已於 2025 年 8 月 2 日起施行，若尚未著手準備，實際上已經處於落後狀態。這正是必須在 90 天內完成檢視的原因。

第四順位為「高風險附錄 III 義務」。按原定計畫，此順序本應排在第二順位之前，但得益於《綜合法案》延至 2027 年 12 月，爭取到了緩衝時間。然而，爭取到時間並不意味著獲得豁免。差距分析產出的路線圖仍應持續推進，只是有了調節進度速度的空間。

第五順位為「附錄 I 產品內置型高風險義務」，到 2028 年 8 月為止時間最為充裕。

在排定這五個順序時，我不會僅參考日期。當施行日期臨近但罰鍰較輕的項目，與施行日期尚遠 knowledge 但罰鍰重的項目發生衝突時，我會賦予罰鍰更大的比重。這正是為何附錄 III 義務即便延至 2027 年 12 月，也沒有降至第五順位以下的原因。雖然爭取到了時間，但罰鍰依然處於 1,500 萬歐元或 3% 的區間，該區間與透明度義務具有同等重量。在路線圖中維持第四順位的位置，但在 90 天內投入的實際工作時間集中在第二與第三順位，這是我建議的比重分配。

在此順序之上，還需疊加「指定歐盟境內授權代表」。若作為第三國提供者向歐洲推出高風險系統或通用模型，依據第 22 條與第 54 條必須設立代表。一俟備妥盤點清單與路線圖，即應將實際資訊提供給該代表，如此主管機關前來詢問時，代表才不致於兩手空空。

將 90 天拆劃為三個月，如下所示：

第一個月（第 1 天至第 30 天）：跨開發、人資、行銷、客戶支援、財務、法務等所有部門，同時展開按部門進行的盤點調查。同週內優先確認是否符合第 5 條的被禁止行為。由於該條文已經施行，最為急迫。

第二個月（第 31 天至第 60 天）：將盤點清單上的項目對照四個層級與角色表，確定分類。本月實際進行第 50 條透明度義務的履行、套用聊天機器人告知文案與深度偽造（Deepfake）標籤技術。被歸類為高風險或通用模型的項目則進入差距分析階段。

第三個月（第 61 天至第 90 天）：完成差距分析並確定差距清單。向高層管理階層報告依日期與罰鍰重量建立的優先順序路線圖，並取得預算與人力核可。若已指定歐盟境內授權代表，則交付盤點清單與路線圖。於 2026 年 8 月初，再次確認《數位綜合法案》是否已實際刊登於官方公報，以及時程係維持原案還是採用新時程。

整理查閱參考資料如下：

施行日期：被禁止行為為 2025 年 2 月 2 日。通用模型義務為 2025 年 8 月 2 日，2025 年 8 月 2 日前發布的模型寬限至 2027 年 8 月 2 日。透明度義務（第 50 條）與 AI 辦公室罰鍰處分權限為 2026 年 8 月 2 日。高風險附錄 III 義務原定為 2026 年 8 月 2 日，經《數位綜合法案》變更為 2027 年 12 月 2 日（需確認是否刊登於官方公報）。高風險附錄 I（產品內置型）義務原定為 2027 年 8 月 2 日，變更為 2028 年 8 月 2 日。

罰鍰：違反被禁止行為處 3,500 萬歐元或全球總營業額 7%；違反其餘大部分義務（高風險、透明度、通用模型一般事項）處 1,500 萬歐元或 3%；提供不準確資訊處 750 萬歐元或 1%。以全球營業額為基準，取兩者中較高者。

各部門盤點檢查：開發團隊的自研模型與 API 串接；人資團隊的招募與評估工具；行銷團隊的目標定位與推薦引擎；客戶支援團隊的聊天機器人與回答推薦；財務團隊的異常交易偵測；法務團隊知悉的合約 AI 條文。

90 天後，板橋的那位主管拿著一張盤點表格走進執行長辦公室。共有四十二個系統，其中三個為高風險候選、五個為透明度對象，其餘皆為最小風險。表格旁並列著五行優先順序。執行長所詢問的並非完美無瑕的合規，而是「該從哪裡開始」，而這張表正是答案。要完全遵守法律，90 天或許不足；但要明白該從何守起，90 天已然足夠。

第19章 文件化體系

位於韓國板橋的一家新創公司決定將其高風險人工智慧系統推向歐洲市場。他們做了風險管理，具備了資料治理，也經過了測試。然而，負責稽核的律師在會議室裡問道：「證明這些事項的文件在哪裡？」負責人員面面相覷。工作是做了，卻沒有留下任何記錄。

這家公司在人工智慧法案面前所處的境地，正是本章的主題。監管當局無法進入公司的腦海中一探究竟。當局唯一能掌握的只有文件。無論風險管理得多麼細緻，如果沒有留下相應過程，在當局面前就等同於完全沒有管理。文件化並非附帶的事務，而是證明其法律存在狀況的唯一途徑。

19.1 保留什麼、以何種形式、保留多久

要保留哪些內容，取決於自身扮演的角色。如果是高風險系統提供者，則須附上第 11 條規定的技術文件與第 17 條規定的品質管理系統（Quality Management System）記錄。第 17 條要求建立一套涵蓋人工智慧開發治理、品質管理、測試與驗證、風險管理、資料治理、技術文件規格、上市後監測以及重大事故通報的文件化體系。如果是通用人工智慧（General-Purpose AI, GPAI）模型提供者，則需保留第 53 條規定的技術文件與訓練內容公開摘要。如果是部署者，則應備妥第 26 條規定的監督記錄與自動生成日誌。

在形式方面，分為有指定格式與無指定格式兩種情況。關於訓練內容公開摘要，人工智慧辦公室（AI Office）已於 2025 年 7 月 24 日發布標準表單，直接使用該表單即可。至於技術文件等由附錄 XI、附錄 IV 規定項目的內容，只需逐一填補該等專屬項目。在有既定格式之處，遵循該格式較為安全。若使用自行製作的表單，日後連自己都很難發現漏掉了什麼。

保存期限因條文而異，這是實務上最常被忽略的地方。依據第 12 條，高風險系統的自動生成日誌至少須保存六個月；若個人資料相關法律要求更長的期限，則從其規定。依據第 18 條，提供者必須在系統投放市場或投入服務後，將技術文件、品質管理系統記錄、驗證機構的決定以及歐盟適合性聲明副本保存 10 年。代表總部位於第三國之提供者的境內授權代表（Authorised Representative），同樣必須將該技術文件保存於隨時可提供給當局調閱的狀態達 10 年。六個月與 10 年，若不將此差異依文件種類整理成表格，每當負責人員更換時，標準就會隨之搖擺。

19.2 因應稽核：如何證明文件與實際運作一致

備妥文件與通過稽核是兩回事。在稽核中真正面臨考驗的，是文件記載的內容與系統實際運行的模式究竟有多吻合。

常見的不一致情況如下：技術文件上記載系統依循此流程做出判斷，然而實際系統經歷了數次更新後，正以另一種方式運作。文件依然原封不動地留在抽屜裡，只有系統獨自向前邁進。當主管機關將文件與日誌、文件與實際輸出進行對比時，這種偏差很快就會暴露無遺。而且，一份不一致的文件比完全沒有文件更糟糕，因為這會留下公司向當局提交虛偽不實資料的印象。第 17 條所要求的品質管理系統記錄正是發揮價值之處。透過展示在設計、開發、部署、運作與變更全過程中實際採用的流程與結果，成為證明文件與實際運作相符的佐證資料。

因此，我建議不要將文件化視為單一次的工作，而是將其視為維護的流程。也就是直接將更新文件的步驟嵌入修改系統的開發流程之中。只要在部署核准檢查清單中加入一項確認相關技術文件是否更新的項目，就能防止文件與實際情況產生差距。雖然這需要多費一些心思，但在稽核現場，文件與系統互相吻合的公司與並非如此的公司，將會受到截然不同的對待。

19.3 回應歐盟當局資訊調閱請求的程序

人工智慧法案賦予主管機關廣泛的資訊調閱權限。高風險系統提供者應依據國家主管機關（competent authority）的合理請求，提出證明符合第 2 節要求事項的所有資訊與文件（第 21 條第 1 項）。這些文件必須以當局指定之歐盟官方語言中當局易於理解的語言提出。若日誌處於提供者的控制之下，亦須開放存取權限（第 21 條第 2 項）。市場監管當局更可進一步要求透過 API 或其他遠端技術手段，存取訓練、驗證與測試資料集本身（第 74 條第 12 項）。監督基本權利保護的國家機關針對附錄 III 的高風險系統，亦有權要求提供以可存取的語言及形式呈現之文件（第 77 條第 1 項）。

通用人工智慧模型方面的對接管道則有所不同。人工智慧辦公室得請求提供履行義務所需的一切資訊，包含依據第 53 條第 1 項 (a) 款規定的技術文件（第 91 條），且該文件須具備附錄 XI 所規定的項目。必須在人工智慧辦公室規定的期限內以最新形式提出，必要時甚至可能被要求提供對模型本身的存取權限（第 92 條）。運算量超過 10 的 25 次方 FLOPs 之具系統性風險模型的提供者，應毫不延誤地、最遲於 2 週內先行通報人工智慧辦公室（第 52 條第 1 項）。對於重大事故或資安侵害事件，同樣負有向人工智慧辦公室及相關當局通報的義務（第 55 條第 1 項 (c) 款）。在所有這些程序中，營業秘密、智慧財產權與機密資訊均受保護（第 78 條）。若在回應調閱請求而提交的文件中含有敏感內容，應明確請求給予機密對待，並在文件上記明機密標示。

若未在公司內部事先指定此對接管道，當請求真正降臨時，法務、開發與業務部門將互相推託從而錯過期限。平時就必須規劃好請求會經由何種管道進入、由誰接收並轉交至哪個部門，以及由誰彙整文件並以何種語言和形式提出。若為第三國提供者，境內授權代表將成為與當局溝通的第一線管道，因此必須預先將所需文件交付給代表，並告知應對要領。若代表手中沒有文件，便無從回應請求。人工智慧辦公室自身已表明將採取合作、循序漸進且具比例原則的方法，並建議若在遵行上遇到困難，可主動聯繫尋求支援。然而在我看來，這種合作態度僅適用於誠實回應的公司。對資訊調閱請求回應疏漏或不誠實，其本身就會被視為獨立的違規行為。

相關日期也必須一併留意。人工智慧法案已於 2024 年 8 月 1 日生效，通用人工智慧模型提供者的義務自 2025 年 8 月 2 日起適用，人工智慧辦公室自此時起可作出相關決定。附錄 III 獨立型高風險系統的義務將於 2026 年 8 月 2 日起全面適用，人工智慧辦公室裁

處罰鍰的權限亦於同日全面生效。罰鍰依違規類型而異：違反一般義務者，處上一會計年度全球總營業額 3% 或 1,500 萬歐元兩者中較高者；違反禁止行為者，處 7% 或 3,500 萬歐元兩者中較高者。此罰鍰計算基準亦包含未回應資訊調閱請求或提供虛偽不實資訊的情形。

有一點目前仍有討論空間。「合理請求」、「易於理解的語言」、「所需的一切資訊」等表述具體指涉何種程度，在各個案例中可能會有不同的解釋空間。歐盟執委會有義務依據第 96 條發布指引，並已於 2025 年 11 月 19 日發布關於通用人工智慧模型提供者義務範圍的指引（C(2025) 7719 final）。然而該指引同樣附帶了必須考量個別案例特性的前提。歸根究柢，只能根據自身狀況進行解釋，並留下該解釋的依據。

參考資料

文件種類與保存期限。高風險系統自動生成日誌至少 6 個月（第 12 條）。技術文件、品質管理系統記錄、驗證機構決定、歐盟適合性聲明於投放市場後 10 年（第 18 條）。第三國提供者之境內授權代表保存的技術文件亦為 10 年。

資訊調閱請求之法律依據。高風險系統為第 21 條（資訊與文件）、第 74 條第 12 項（資料集存取）、第 77 條第 1 項（基本權利主管機關存取）。通用人工智慧模型為第 91 條（資訊調閱請求）、第 92 條（模型評估存取）、第 53 條第 1 項 (a) 款（技術文件）、第 52 條第 1 項（系統性風險通報，2 週內）、第 55 條第 1 項 (c) 款（重大事故通報）。機密保護為第 78 條。

關鍵日期。生效日為 2024 年 8 月 1 日。通用人工智慧模型義務適用日為 2025 年 8 月 2 日。高風險系統全面適用及罰鍰權限生效日為 2026 年 8 月 2 日。罰鍰為 3% 或 1,500 萬歐元，禁止行為則為 7% 或 3,500 萬歐元，兩者皆取較高者。

內部準備事項。製作依文件種類區分的保存期限表。在系統更新流程中納入文件更新步驟。預先指定接收資訊調閱請求的窗口與負責部門。預先將文件交付給境內授權代表。機密資訊於提交前標記完畢。

稽核從不會毫無預告地降臨。從調閱通知書飛來的那一刻起，時鐘就開始運轉。在那一刻翻箱倒櫃的公司，與打開早已整理妥當資料夾的公司，兩者之間的差距將決定最終的命運。

第20章 合約與供應鏈

假設板橋的一家新創公司要開發一款諮詢用聊天機器人。由於沒有資金訓練自家模型，因此透過 API 引進歐洲大型科技公司的通用人工智慧模型（GPAI model），並加上自家邏輯，作為一套 AI 系統推出。合約通常是該大型科技公司預先擬定的標準使用條款，是一整包點擊一次即表示同意的條款。單憑這份標準條款，是否就等於履行了歐盟人工智慧法案（Regulation (EU) 2024/1689，以下簡稱 AI 法案）的合規義務？我認為並非如此。AI 法案已在條文中明確規定上游模型提供者與下游系統提供者之間應在何時、以何種形式交付何種內容，而該條文的位階高於標準條款。應該根據該條文重新擬定合約才對。

第 19 章探討了主管機關要求提供資訊時，企業應 how 回應。第 20 章則著眼於前一階段，即在供應鏈中企業之間應在合約中納入哪些條款。合約是法規遵循的工具，而非替代法規的產物。

20.1 應納入 AI 供應商合約的條款

AI 法案第 25 條第 4 項規定，高風險 AI 系統提供者與供應整合至該系統之 AI 系統、工具、服務、元件、程序之第三方應訂立書面合約。合約中應根據最新技術水準（state of the art）明確規範必要的資訊、能力、技術存取權及其他支援，以使下游提供者能完整履行 AI 法案規定的義務。形式必須為書面，僅憑一封電子郵件是不夠的。不過，以免費開源授權釋出的工具、服務及元件免除此項義務。釜山一家醫療器材公司的法務團隊就曾忽略過：通用 AI 模型本身即使是開源的，也不適用此豁免條款。

我們將合約中應納入的項目分為三個部分。

資訊提供條款。具體載明上游供應商應針對其所供應之元件交付何種內容。高風險 AI 系統提供者負有風險管理系統（第 9 條）、資料治理（第 10 條）、技術文件（第 11 條）、自動記錄（第 12 條）、人工監督（第 14 條）、準確性・強健性・網路安全（第 15 條）等義務，若要證明這些義務，就必須瞭解上游元件的資料來源、局限性及日誌記錄方式。必須針對這六個條文逐一列明應於何時之前交付何種文件。若僅含糊地寫上「提供必要之支援」，在發生爭議時，上游供應商便有空間主張其已提供最低限度的支援。

協助條款。規定當接獲稽核、事故調查或主管機關的資訊要求時，上游供應商應於多久時間內回應。正如第 19.3 節所述，主管機關的資訊要求必須在指定期限內回應，但若該資訊有相當大一部分掌握在上游供應商手中，下游提供者僅憑一己之力將無法遵守期限。最好將協助期限明確寫為若干個工作日，甚至載明違約時的處置措施。這是合約中最常被遺漏的條款。

變更通知條款。此條款規定當上游模型更新或安全措施變更時，必須事前通知下游提供者。AI 法案規定 GPAI 模型提供者有義務在模型的整個生命週期內更新技術文件（第 53 條第 1 項），但若更新事實未自動傳達給下游客戶，這項義務就等同於僅存在於紙面上。我曾見過因單一模型變更而導致下游系統必須重寫整份風險管理計畫的案例。若能在合約中事先規定變更通知的期限與通知內容的具體程度，就能爭取到應對的時間。

2025 年 3 月歐盟公共採購社群所公開的 AI 模範合約條款（MCC-AI）雖然是用於公共採購，但其資訊提供與協助條款的骨幹仍極具參考價值。然而，該條款已聲明其本身並不處理智慧財產權、個人資料保護與損害賠償問題，因此絕非直接複製貼上即可了事。

20.2 應從上游模型提供者取得的資訊

反過來說，若韓國企業處於下游系統提供者的位置，究竟需要從上游模型提供者取得什麼，合約才算完整？

首當其衝的核心是第 53 條第 1 項 (b) 款及附錄 XII。GPAI 模型提供者必須編製資訊與文件並維持最新狀態交付給下游提供者，以便下游提供者能將模型整合至其系統中。附錄 XII 所要求的內容包含模型的概述、預期任務、與其他系統的整合方式、上市日期與分發方式，以及與硬體・軟體的交互作用。若缺少這些資訊，下游提供者將無法填寫附錄 IV 的技術文件，而一旦無法填寫，符合性評估（Conformity Assessment）本身便會停擺。

除此之外，還必須額外注意三點。著作權法遵循政策：依第 53 條第 1 項 (c) 款規範，遵守指令 2019/790 之文字與資料探勘（Text and Data Mining, TDM）例外規定之政策及權利保留標示。訓練內容公開摘要：第 53 條第 1 項 (d) 款要求之此份摘要，將依據歐盟執委會於 2025 年 7 月 24 日公開的說明書與標準格式撰寫並予以公開。整合所需之支援：第 25 條第 4 項所要求之資訊、能力與技術存取權。在 20.1 節中作為下游提供者應要求之項目的該條文，在此處則反轉呈現為應接收之項目。

若模型伴隨系統性風險（systemic risk），應取得的資訊層面將會再增加一層。第 51 條將以超過 10 的 25 次方浮點運算數（10^25 FLOPs）之計算資源所訓練的模型歸類為具系統性風險之模型；第 55 條則要求該提供者進行模型評估、通報重大事故與網路安全侵害事件，並於達到臨界值時在 2 週內通知 AI 辦公室（AI Office）。下游提供者必須透過合約取得該評估結果與事故紀錄，方能反映於自家的風險管理計畫中。在上游模型已被默默歸類為具系統性風險之模型，而唯獨下游提供者毫不知情的情況下，我認為這是最危險的盲區。

開源模型則存在例外情況。免費且開源授權的模型在特定條件下可免除技術文件與供下游提供者使用之資訊義務，但前提是必須公開模型參數（包含權重）、架構資訊及使用資訊。這只不過是將公開方式從合約轉變為公開管道而已。

自 2025 年 8 月 2 日起，GPAI 模型提供者的義務已開始適用，而在此之前上市的模型則享有直至 2027 年 8 月 2 日的寬限期。若是準備訂立新合約的公司，首要任務便是確認

對方的模型是否處於寬限期內。

20.3 應提供給下游客戶的資訊

這次我們假設韓國企業是開發並銷售 GPAI 模型的上游提供者。20.2 節中的項目此時便反轉為應主動提供的一方之義務。企業必須準備好將附錄 XII 的資訊、著作權政策及訓練內容公開摘要交付給下游客戶。

問題便由此產生。附錄 XII 所要求的資訊中包含了模型是以何種資料、如何進行訓練的內容，而這同時也是公司的核心技術資產與營業秘密。AI 法案第 78 條第 1 項雖規定主管機關應保護營業秘密、智慧財產權及資料保密性，但該條文是針對主管機關設限，並非免除向下游客戶提供資訊的免死金牌。附錄 XII 所規定的最低限度項目，並非可以營業秘密為由予以刪減的性質。

因此在實務上，資訊會分為兩個層次。第一層是附錄 XII 所明定、下游客戶履行其義務所絕對不可或缺的資訊。這部分不得以營業秘密為由扣除。第二層則是支撐該項目的詳細實作、訓練演算法的具體程式碼或內部超參數（hyperparameters）。這部分只要抽象化至說明「是以何種類型的資料進行如何訓練」的程度來交付，即可滿足附錄 XII 的要求。我建議將此區分製作成文件範本並附於合約之後。

亦可採取附加保密協定（NDA）的方式。若必須提供更詳細的資訊，可將該資訊指定為機密資訊並限制其使用目的。然而，NDA 是用來限制所交付資訊之用途的工具，而非縮減應交付資訊範圍的工具。此處還有另一個槓桿：AI 法案第 89 條第 2 項承認下游提供者有權以上游 GPAI 模型提供者違規為由向 AI 辦公室提出申訴。若懈怠於提供資訊，上游提供者面臨的可能不再是合約爭議，而是被召至監管機關面前的窘境。

20.4 責任分配與免責條款的局限

無論合約擬定得再怎麼嚴密，依然存在無法逾越的界線。AI 法案於第 3 條中定義了提供者、部署者（deployer）、進口商及分銷商的角色，並對每個角色賦予不同的義務。高風險 AI 系統提供者所承擔的風險管理、資料治理、技術文件及符合性評估義務，因與保障公共安全及基本權利之目的緊密相連，絕非憑合約中的一行文字便能轉嫁他人。我稱之為不可委任（non-delegable）義務。透過合約約定損害賠償的分攤比例或許可行，但無法透過合約改變監管機關將對誰實施裁罰。

AI 辦公室與國家主管機關擁有直接對違規當事人處以罰鍰並下令採取改正措施的權限。無論下游提供者取得多麼優渥的免責條款，AI 辦公室都不會審視該條款，而是直接針對在 AI 法案下應承擔責任的提供者開罰。免責條款僅在當事人事後進行內部結算時有用，無法成為阻擋主管機關親自執法的盾牌。

從罰鍰的規模便能明顯看出此一局限。違反禁止實踐行為（第 5 條）最高可處 3,500 萬歐元或全球營業額的 7%（以較高者為準）；違反其他義務最高可處 1,500 萬歐元或全球營業額的 3%（以較高者為準）。若涉及個人資料，還會疊加 GDPR 的違規罰鍰，最高可達 2,000 萬歐元或營業額的 4%。商業合約的免責上限通常僅為合約金額的數倍，而這些按營業額百分比計算的罰鍰輕易就能超越該上限。至於強制退出市場或聲譽受損等非財務性損失，根本就不是免責條款當初所能涵蓋的範疇。

法律體系的版圖亦有所改變。人工智慧民事責任指令（AI Liability Directive）草案已於 2025 年 2 月 11 日由歐盟執委會宣布撤回意向，並於 2025 年 10 月 6 日刊登於公報而正式廢止。表面理由是利益相關者之間未能達成共識。取而代之的是，經修訂的產品責任指令（(EU) 2024/2853）將適用於 2026 年 12 月 9 日之後投放市場的產品，該指令明文將軟體與 AI 系統視為產品。自 2026 年底起，若因 AI 系統缺陷造成損害，製造者將承擔無須證明過失的無過失責任（Strict Liability / 嚴格責任）。即使是因為未及時進行更新或安全性補丁而導致缺陷持續存在，亦將被認定為缺陷。AI 法案下的行政罰鍰與產品責任指令下的民事賠償責任，可能針對同一起事故同時併發。

當韓國企業以首爾為總部向歐盟市場推出 AI 系統，或其輸出結果於歐盟境內被使用時（AI 法案第 2 條第 1 項 (c) 款），第 20 章從 20.1 到 20.4 的這四小節在同一份合約中是相互牽連的。若 20.1 的條款不夠完善，20.2 中本應取得的資訊一開始就無法流入；而若

無法取得該資訊，20.3 中要交付給下游客戶的資訊也會變得漏洞百出。正如 20.4 所言，無論這些條款擬定得有多好，在監管機關面前，個別企業的最終責任依然存在。合約是地圖而非盾牌。只不過，這張地圖並無法抹去 AI 辦公室所劃定的執法邊界。

建議您不妨翻開目前手中的供應商合約確認一下。資訊提供條款中是否載明了條文編號？協助期限是否明確規定為工作日？變更通知條款究竟是否存在？如果都沒有，那麼該份合約便依然停留於 AI 法案誕生之前的世界。

參考資料

合約檢查要點 1. 是否已載明對應第 9 條、第 10 條、第 11 條、第 12 條、第 14 條、第 15 條各條文之資訊提供項目及條文編號。 2. 協助條款中是否已以工作日為單位，明確規範因應稽核、事故調查及主管機關資訊要求之期限。 3. 變更通知條款中是否已規定通知時點及內容之具體程度。 4. 是否設有從上游 GPAI 模型提供者處取得附錄 XII 項目、著作權政策及訓練內容公開摘要連結之程序。 5. 是否設有當模型被歸類為具系統性風險之模型（超過 10^25 FLOPs）時接獲通知之條款。 6. 是否備有劃分應提供予下游客戶之資訊與作為營業秘密保護之資訊的基準文件。 7. 免責條款上限金額之設計是否已考量 AI 法案罰鍰（營業額之 3% 至 7%）與 GDPR 罰鍰（營業額之 4%）。 8. 是否已重新審視瑕疵及無過失責任條款，以因應 2026 年 12 月 9 日產品責任指令適用後之情勢。

關鍵日期 2024 年 8 月 1 日，AI 法案生效。 2025 年 2 月 2 日，禁止之實踐行為相關規範全面施行。 2025 年 3 月，歐盟公共採購社群公開 AI 模範合約條款（MCC-AI）。 2025 年 7 月 24 日，公開訓練內容公開摘要說明書與標準格式。 2025 年 8 月 2 日，GPAI 模型提供者義務開始適用，AI 辦公室之 GPAI 決定權限啟用。 2025 年 10 月 6 日，人工智慧民事責任指令撤回案刊登於公報。 2026 年 8 月 2 日，高風險 AI 系統義務全面適用，AI 辦公室裁罰權限全面啟動。 2026 年 12 月 9 日，修訂版產品責任指令（(EU) 2024/2853）開始適用。 2027 年 8 月 2 日，2025 年 8 月 2 日前上市之 GPAI 模型合規寬限期結束。

第21章 組織與人員

這是在板橋一家軟體公司真實發生的會議場景。在決定向歐洲子公司引進招聘審查人工智慧時，法務長表示「這是開發團隊該寫技術文件的活」，開發長則回應「我們只管寫程式碼，員工培訓與向主管機關申報是法務的事」。事業部總監整場會議只是靜靜聽著，最後問了一句：「所以這件事，我們三個到底誰要負責？」這項提問沒有既定答案，正是本章的出發點。人工智慧法並未指定特定的部門或職務。然而，如何佈局人員與組織，卻成了決定罰款金額的關鍵變數。

21.1 AI 素養義務，第4條與第26條是不同的條文

人工智慧法第4條要求提供者與部署者，必須確保員工以及代表其處理人工智慧的人員具備足夠水準的 AI 素養（AI literacy）。該條文已於2025年2月2日起正式施行。無論公司使用的是何種風險等級的人工智慧，即便不是高風險，哪怕只是引進了一款聊天機器人，都會觸及第4條。

這裡常出現一個誤解。在第4條草案階段，討論原本朝向對企業施加具體義務的方向進行，但歐盟執委會隨後將重心轉移至會員國與執委會自身的政策課題上。因此，第4條本身對企業的要求較接近宣示性文字，真正具體可落實的義務則出自第26條。第26條規定使用高風險人工智慧的部署者，有義務培訓員工進行人類監督（human oversight）。關於這項培訓義務，依附錄 III 獨立型高風險系統的標準，原定於2026年8月2日全面適用；然而正如第13章所指出的，隨著2026年5月《數位綜合法案》達成暫行共識，適用時間有可能延後至2027年12月2日。在我看來，這一點絕不能混淆。高風險培訓義務（第26條）固然可能延後，但適用於所有提供者與部署者的通用素養義務（第4條）已經施行，且並非綜合法案討論的對象。如果板橋那家公司因為只使用聊天機器人，就以為看到8月2日延期的消息便能鬆一口氣，那便是將第4條與第26條混為一談了。

法律並未以清單形式規定培訓的具體內容。不過，考量到人工智慧辦公室（AI Office）累積的實務案例，以及未來即將出爐的第8條至第15條、第25條指引，預計將與人類監督及風險管理結合處理，由此可以推知，這項素養培訓絕非拿一張形式上的結業證書就能了事。建議各位制定包含以下四項內容的培訓計畫，並留下完成紀錄：系統運作原理、極限與潛在偏見、如何閱讀性能指標，以及識別需要人工介入的時機。若將這些紀錄作為第17條品質管理體系或第11條技術文件的一部分進行管理，日後面臨稽核時便無需另行搜尋。至於罰款部分則較為模糊。第99條列出的罰款區間中，並未單獨提及第4條。因此，認為並不存在單純針對違反第4條的固定罰款標準，才是精準的理解。在實務上，會員國的市場監管機關極有可能將其納入一般義務違反區間（最高1500萬歐元或全球營業額的3%）一併裁量，或者在發生事故時，將其作為人類監督失敗的背景事由加以引用。

21.2 設立組織圖上不存在的職位

人工智慧法並未明文要求任何公司都必須成立人工智慧治理團隊。然而，幾乎沒有一家公司能由單一工作人員兼任處理風險管理體系（第9條）、資料治理（第10條）、技術文件（第11條）、品質管理體系（第17條）與人類監督（第14條）。條文本身雖未強制要求特定組織結構，但將所有條文串聯起來後，事實上若無專責組織便無法運作。我將此視為法律留下的空白：形式上給予自由，內容上卻實施了強制。

實務上有一套名為「90天啟動計畫」的推動程序。該程序分為三個階段：建立人工智慧系統盤點清冊、釐清各系統屬於提供者還是部署者的角色，以及進行找出漏洞的差距分析。首先必須決定由誰來主導這三個階段。我建議設置直屬於執行長的人工智慧治理負責人 - - 無論名稱是首席人工智慧官還是人工智慧合規負責人皆可 - - 並由法務、開發、事業三個部門的代表在其麾下定期召開會議。若由這三個部門中的任何一方獨攬大局，另外兩方就會淪為配角，而變成配角的部門往往會敷衍塞責地填寫自己負責的文件。我曾多次目睹此類事故發生。

人工智慧辦公室的關注焦點也在於此。人工智慧辦公室已明確表示，將深入審查風險管理體系與品質管理體系是僅停留在紙面上，還是在組織內部切實運作。若公司製作了完美的政策文件，卻未分配執行人力、預算與權限，該文件在稽核中反而會成為不利的證據。因為有政策卻無負責人這件事本身，就會被解讀為風險管理的漏洞。

21.3 法務懂法律，開發懂技術，事業懂市場，以及三方共同判斷的場合

法務部門的職責相當明確：梳理人工智慧法與國內人工智慧基本法重疊之處、在供應商與客戶合約中加入人工智慧法相關條款、指定並管理 EU 境內授權代表，以及回應人工智慧辦公室或會員國主管機關的資訊調閱請求。若公司涉及通用人工智慧（GPAI）模型，確認是否符合著作權政策亦屬法務職責。

開發部門則負責從設計階段起將法律融入程式碼中。這包括將風險管理體系與資料治理落實於實際管道中、撰寫與更新技術文件、加裝自動記錄功能，以及在系統中植入第50條所要求的 AI 生成物機器可讀標示。認為開發部門無需了解法律條文的想法極具風險。在不了解條文情況下寫出的程式碼，事後法務再怎麼修飾也無法彌補。

事業部門則聚焦於市場。事業部門須先確定本公司在此供應鏈中究竟屬於提供者、部署者還是輸入商，識別系統的風險等級，並管理上市時程與策略。向下游部署者說明系統用途與限制事項，必要時提供培訓，亦屬事業部門的職責。若屬具有系統性風險的 GPAI 模型，必須在兩週內通報人工智慧辦公室，而能掌控此緊迫時限的，終究是第一時間掌握市場狀況的事業部門。

即使進行了上述分工，仍會剩下無法單純歸類於三者之一的交叉判斷事項。例如：引進 GPAI 模型進行微調時，是否會產生新的「提供者」地位？我們的系統是否會落入附錄 III 的信用評等或保險風險評估範圍？這類判斷無法單憑法務、開發或事業任何一方單獨作出。這必須將法律層面的解讀、技術層面的變革，以及市場層面的意義擺在同一個檯面上互相對齊，才能得出結論。我認為將這三方召集至同一個會議體系中，絕非形式上的安排，而是實質上的必要。會議絕不能僅止於留下紀錄後便散場，還必須明確釘牢會中決議由誰執行、由誰確認，會議才具有真正的意義。

在此列出本章應確認的核心要點：

第4條 AI 素養：自2025年2月2日起對所有提供者與部署者施行，與風險等級無關。

第26條 人類監督培訓：針對高風險部署者。原定2026年8月2日適用，因《數位綜合法案》有可能延後至2027年12月2日。切勿與第4條混淆。

罰款區間：第99條中無針對第4條的專屬條文。可能被納入一般義務違反區間（最高1500萬歐元或全球營業額3%）裁罰，或作為事故調查時的背景事由。

治理組織：法律雖未強制規定形式，但建議設立直屬於執行長聯絡負責人，並由法務、開發、事業三部門代表組成定期會議體系。

各部門分工：法務負責法規、合約、代理人及主管機關溝通；開發負責設計、文件、日誌與標示技術；事業負責角色確定、等級識別、上市策略、下游培訓及主管機關通報。

需交叉判斷事項：提供者地位變更、是否屬於附錄 III 等事項，應由三部門共同判斷，並指定結論的執行者與確認者。

讓我們再次回到板橋的那間會議室。對於事業部總監提出的問題 - - 「我們三個到底誰要負責？」 - - 答案並非只有一個。然而，若會議結束時仍無人能回答這個問題，那麼該公司恐怕已經忽略了第4條或第26條中的某項規範。

第22章 監管沙盒與事前諮詢

板橋的一家 AI 醫療保健新創企業代表曾問我：他們想把一套能讀取視網膜影像並篩檢糖尿病視網膜病變的軟體賣給德國醫院，但在獲得正式認證之前，有沒有辦法用真實患者的資料進行驗證？我這樣回答：自2026年8月2日起，將會有一個解答，那就是各會員國營運的 AI 監管沙盒。

22.1 會員國沙盒使用條件

《人工智慧法》（Regulation (EU) 2024/1689）第57條明確規定，會員國至少須設立一個國家級 AI 監管沙盒，並於2026年8月2日前正式運作。這項義務的主體是會員國，這一點至關重要。這意味著它不是企業要主動申請的對象，而是政府必須先建置好的設施。此義務可透過與其他會員國主管機關共同設立來履行，亦可透過參與其他已營運之會員國沙盒的方式予以替代。然而，唯有在該項參與能保障本國國民享有同等程度的存取權時，方獲認可。此外，並不禁止在區域或地方層級設立額外沙盒；歐洲資料保護主管機關（EDPS）亦可為歐盟機構設立獨立沙盒，並以與國家主管機關相同的地位運作。執委會得提供技術支援、諮詢與工具，但無法代替會員國承擔設立義務本身。

沙盒內進行的事宜均遵循提供者與主管機關事先合意之沙盒計畫。其核心在於依據計畫於指定期間內進行系統開發、訓練、測試與驗證，亦得於主管機關監督下進行真實環境測試（real world conditions）。主管機關擔任協助符合法規規範的監督者與顧問角色。在風險管理中，特別是針對基本權利、健康與安全相關風險的辨識與緩和過程中，提供指導與支援；若提供者提出請求，主管機關將開立證明其於沙盒內所作所為的書面證明與結束報告。這兩份文件日後在符合性評估或市場監督階段具有相當的比重。因為法律直接要求市場監督機關與驗證機構（notified bodies）應積極採納該報告，並於合理範圍內將程序提前辦理。

最受矚目的條款是罰鍰豁免。沙盒參與者若於實驗過程中對第三人造成損害，仍無法免除其民事責任。這一點理所當然。然而，只要在遵守計畫與參與條件、且出於善意遵循主管機關指導的前提下，對於違反《人工智慧法》的行政罰鍰將不予裁罰。針對違反其他歐盟法律或國內法的事項，若屬相關機關積極監督並提供指導的情形，亦同。我將此條款解讀為監管機關宣示不會從一開始就將企業視為敵對對象。然而，若出現對健康、安全或基本權利的重大風險，主管機關得隨時暫停或永久終止沙盒活動，該決定亦將通報予 AI 辦公室。這意味著沙盒絕非逃避監管的庇護所。

國家主管機關設立沙盒後，必須將該事實通報 AI 辦公室與 AI 委員會（AI Board）；AI 辦公室則會在網站上公開各國沙盒計畫與清單並保持最新狀態。具體營運方式，亦即資格與選拔標準、申請程序、參與條件等共通原則，將由執委會依據第58條以施行細則訂定。此施行細則旨在防止全歐盟以迥異的標準營運沙盒，並於其中定明中小企業與新創

企業原則上享有免費存取的權利。

從韓國企業的角度來看，考量相對明確。依據 AI 法第2條第1項 (c) 款，只要產出物於歐盟境內被使用即適用域外管轄的公司，在將高風險系統推向市場之前，先透過沙盒進行驗證會更加安全。考量到違規時最高可處3,500萬歐元或全球營業額 7% 的罰鍰結構，僅憑「出於善意遵循指導」的事實即可開啟免除全部罰鍰的途徑，這絕非微小的誘因。一旦取得結束報告與書面證明，通往 CE 標誌認證的時間也會縮短。然而，歐盟境內授權代表在此過程中須兼任與主管機關的聯絡窗口，且針對實驗期間可能發生的第三人損害，亦須另行投保責任保險。

亦有尚不確定的部分。由於施行細則尚未確定，資格標準與申請程序的具體型態在各會員國之間可能呈現差異。此外，對於「善意遵循指導」這項要件主管機關於事後會有多嚴格的審查，以及非歐盟企業實際上能在多大程度上獲得比照歐盟境內中小企業的優先權，皆須等待各國累積初期案例後方能判斷。我個人抱持樂觀態度，但比起成為第一批申請者，更傾向建議先觀察數個月後的先例再行申請。

22.2 中小企業優惠措施

《人工智慧法》第62條要求會員國與 AI 辦公室在法規設計階段，即須特別考量中小企業（特別是新創企業）的處境。釜山一家醫療器材新創企業代表向我提出的問題，正可在此條文找到解答。他問的是：是否必須繳納與大型企業相同的符合性評估費用？答案是「否」。第62條第1項規定，高風險系統符合性評估費用應依企業規模、市場規模等相關指標按比例調降，並應依 AI 委員會之請求定期重新審查該基準。

在沙盒存取權方面，中小企業（SME）與新創企業亦排在優先順位。只要符合資格要件與選拔標準，即享有免費優先存取的權利，僅例外費用會另行收取。此外，會員國必須在全歐盟建置並維護標準化的申請模板與單一資訊平台，舉辦針對 SME 的意識提升活動與培訓，並承擔評估與推廣 AI 相關公共採購最佳實務的角色。第63條更進一步允許微型企業（指無合夥或關聯企業，員工未滿10人且營業額或資產總額在200萬歐元以下者），得透過降低文件負擔的方式滿足部分品質管理系統要件。

此外亦出現名為「小型中型企業（Small Midcaps）」的中間層級。指員工未滿750人且年營業額在1.5億歐元以下或年度資產負債表總額在1.29億歐元以下的公司，並將過去僅限微型與中小企業享有的部分減輕文件負擔優惠擴大至此等規模的公司。其技術文件要件有所調降，違規時裁處的罰鍰亦改以 3% 或 1,500萬歐元中「較低者」為基準，而非較高者。然而，此處有一項重大例外：若從事第5條所禁止的操作性 AI 慣例或社會信用評分等行為，即便是小型中型企業，仍將直接適用最高3,500萬歐元或全球營業額 7% 的一般上限。我在進行諮詢時特別強調這一點，因為規模小所寬待的是文件與程序的負擔，而非禁止行為本身的嚴重性。

韓國企業應留意的實務事項可整理為三點。第一是先確認自家公司屬於哪一個範疇。須根據員工人數、年營業額、資產負債表總額，核實屬於 SME、新創企業、微型企業或小型中型企業，方能決定向會員國主管機關提交何種文件。第二是切勿放過費用減免與優先存取的申請。即使有規定，若未提出申請亦無法適用。第三是若屬於微型企業或小型中型企業，得依據第63條的文件簡化條款為由，預先縮減應對文件的範圍進行協商。然而，「歐盟境內」這一字眼是否直接適用於非歐盟企業，目前經案例確認者尚不多，因此我建議先透過歐盟授權代表向各會員國主管機關以書面方式進行確認。

22.3 與 AI 辦公室的溝通管道

若沙盒與費用減免是制度的骨架，那麼實際踏入該制度的管道便是溝通管道。執委會依據第58條，應開發供利害關係人與主管機關接觸並尋求非拘束性指導的單一專用介面。除此之外，AI 辦公室於2025年10月正式啟用「AI 法服務台（AI Act Service Desk）」與「單一資訊平台（Single Information Platform）」。其結構為透過線上表單提交問題後，由與 AI 辦公室緊密合作的專家團隊予以解答，且各國主管機關資訊及各會員國沙盒現況亦可在該平台上查詢。2025年11月19日，進一步說明與限縮通用人工智慧（GPAI）模型提供者義務範圍的指南（C(2025) 7719 final）獲得批准並一併公開。

主管機關或 AI 辦公室主動聯繫的情形亦有法律依據。高風險系統提供者依據第21條，須應主管機關之合理請求，提供證明符合規範之文件，且該文件應以主管機關易於理解之歐盟官方語言撰寫。自動生成日誌之存取（第21條第2項）或市場監督機關對訓練、驗證、測試資料集之存取（第74條第12項）亦屬請求對象。GPAI 模型提供者之情形則略有不同。AI 辦公室直接作為窗口，得要求提供依據第91條之資訊請求與依據第92條之模型評估存取；具備系統性風險之模型（即訓練算力超過 10^{25} FLOPs 的模型）提供者，最晚須於兩週內將該事實通報 AI 辦公室。若發生嚴重事故或資安侵害，亦伴隨迅速通報之義務。無論透過何種途徑往來，營業秘密、智慧財產權與機密商業資訊均依據第78條受到保護。

AI 辦公室將自身定位為具合作性、階段性且符合比例原則的執法者。其旨意在於若在合規上有困難，切勿先擔心罰鍰，而應主動聯繫；反之，針對未遵循實踐守則或合規情況不明的提供者，則明確表示得要求提供更多資訊，甚至請求模型評估之存取。我在這一點上總是給予韓國企業相同的建議：預先組建由法務、開發、業務部門共同聯動的內部應對團隊，並將技術文件、QMS 紀錄、日誌、GPAI 訓練內容摘要整理至隨時皆可調閱的狀態。資訊請求通常給予的期限十分緊迫，若到了那一刻才開始尋找文件，為時已晚。

參考資料

核心期限 2026年8月2日 會員國 AI 監管沙盒正式運作截止、高風險系統義務全面適用、AI 辦公室裁罰權限全面啟動 2025年8月2日 GPAI 模型提供者義務開始適用 2025年2月2日 第5條禁止慣例全面施行 2025年10月 AI 法服務台與單一資訊平台上線 2025年11月19日 GPAI 指南 C(2025) 7719 final 獲得批准

檢查項目 1. 是否已依據員工人數、年營業額、資產負債表基準，確認貴公司屬於 SME、新創企業、微型企業或小型中型企業何者？ 2. 是否已於 AI 法服務台確認目標會員國的沙盒計畫與申請窗口？ 3. 歐盟境內授權代表是否已準備好擔任與主管機關的聯絡窗口？ 4. 是否已評估針對沙盒參與期間可能發生之第三人損害的責任保險？ 5. 是否已常態性整理技術文件、QMS 紀錄與日誌，以便能即時回應資訊請求？

法律所開啟的大門已敞開。究竟是要率先敲響那扇門，抑或是觀察其他公司先行進入後再隨之跟進，最終取決於各家企業的風險偏好。

後記 將管制視為成本，還是當作進入障礙？

讀到本書這裡的公司負責人員，想必已經感到疲憊。需要遵守的事項實在太多。進行標示、備妥文件、指定授權代表、接受評估，並完成登記。若想在歐洲銷售產品，就必須通過這一切。因此，將管制視為成本是極為自然的事。這不是一項能增加營收的工作，卻又耗費金錢與時間，帶來沉重負擔。許多公司都是這樣看待歐洲《人工智慧法》的。

我想在這最後一章提出另一個觀點。那就是不將管制視為成本，而是將其視為進入障礙的觀點。確切來說，是尋找一個「對他人而言是障礙，對自己而言卻是門戶」之位置的觀點。

管制沉重，意味著對所有人來說都很沉重。瞄準歐洲市場的其他國家競爭對手，也同樣站在這道牆前。比別人先一步越過這道牆的公司，就已經先進入了那些仍在牆前猶豫不決的競爭對手無法進入的市場。提早完成管制因應，這本身就成為了一種領先優勢。歐洲對人工智慧施加如此沉重的管制，反過來說，也代表只有通過該管制的人工智慧才能留在歐洲市場。擠進通過管制的少數行列之中，這就是競爭。

我認為這個觀點對韓國企業而言尤具意義。韓國企業長期以來在通過嚴苛管制的過程中成長起來。透過將產品輸往安全管制嚴格的市場來提升品質，而這份品質又轉化為競爭力。人工智慧管制也可以放在這樣的位置來看待。在符合歐洲要件的過程中建立起來的資料治理與風險管理體系，不僅僅用於歐洲，更會讓整間公司的人工智慧整體變得堅實。因應管制，本身就成了公司的實力。

當然，這並不意味著管制變得輕微。正如整本書所強調的，義務極為沉重，截止日期日益臨近，裁罰力道也十分巨大。主張將管制用作進入障礙，並非要否定這份重量，而是要先一步承擔這份重量。當其他人拖延推遲時，率先做好準備的公司，將會站在管制所造就的那扇窄門之內。

合上這本書，我重新回到了最初的那個問題。前言中提到的板橋公司，當時並不知道自己會受規範涵蓋。讀過本書的公司，如今已經清楚明白。明白自己會透過何種途徑受規範、以何種身份受規範，以及該在何時之前完成何事。擁有這種認知的公司與缺乏認知的公司，在相同的管制面前將站在完全不同的位置。一方會在 8 月 2 日接到電話，而另一方則在該日期之前便已準備妥當。

將管制視為成本，還是當作進入障礙？這個問題的答案並不在條文之中，而存在於公司決定何時、以何種方式面對該條文的選擇之中。牆已經立起，剩下的就是決定何時越過

這道牆。期待能在率先越過的那端與各位相見。

附錄

附錄 1 施行時程整合年表

基準日 2026年7月23日。本年表之日期為動態變動。實際展開準備前，請務必再次確認最新狀態。

歐盟人工智慧法案

2024年8月1日 人工智慧法案生效。

2025年2月2日 第5條禁止行為開始適用。第4條人工智慧素養義務開始適用。

2025年8月2日 通用人工智慧模型提供者義務（第53、55條）開始適用。治理規定開始適用。第99條裁罰依據生效。

2026年6月29日 Digital Omnibus 修訂案歐盟理事會最終批准。於官方公報刊登後第3日生效。

2026年8月2日 第50條透明度義務全面適用。針對通用人工智慧模型之執法權限啟動。

2026年12月2日 第5條 NCII 與 CSAM 生成禁令生效。2026年8月2日前已投放市場之既存生成式系統，其標示義務寬限期結束。

2027年2月2日 具備通用人工智慧浮水印檢測可互操作性解決方案之期限。

2027年8月2日 2025年8月2日前已投放市場之既存通用模型之合規調整期限。

2027年12月2日 附錄 III 獨立型高風險人工智慧義務開始適用。

2028年8月2日 附錄 I 產品內建型高風險人工智慧義務開始適用。

韓國 AI 基本法

2025年1月21日 AI 基本法制定並公布。

2026年1月22日 本法全面施行。透明度義務（第31條）、安全性確保義務（第32條）、高影響力事業營運者職責（第34條）、國內代理人（第36條）等。

2026年 7月 21日 修訂版施行細則施行。AI 產品與服務確認制度、政府採購優惠、使用成本補助等產業振興措施。

已延期與未延期事項

已延期 附錄 III 獨立型高風險（2027年12月2日）、附錄 I 產品內建型高風險（2028年8月2日）。

未延期 第50條透明度義務（維持於 2026年8月2日）。NCII 與 CSAM 禁令並非延期，而是新增規定（2026年12月2日）。

附錄 2 系統分類判斷流程圖

依序回答問題，即可找出我們系統的定位。各問題旁的章節編號可參閱詳細說明。

第1階段 境外適用與否（第1章）

問題：我們的人工智慧是否投放於歐洲市場、是否有歐洲使用者使用，或其輸出結果是否用於歐洲？
三者滿足其一 → 適用對象。進入第2階段。
三者皆否 → 目前排除適用。惟須再次確認是否確實毫無接觸點。

第2階段 角色確定（第1章）

問題：在此系統中我們扮演何種角色？
開發並推出 → 提供者。
引進使用 → 部署者。
進口 → 進口商。 / 於供應鏈中提供 → 分銷商。
單一系統可能重疊具備多重角色。

第3階段 禁止事項確認（第15章）

問題：此系統是否觸及第5條八項禁止行為之一？（職場與教育機構的情緒識別、社會信用評分、剝削弱勢、無差別收集臉部影像、生成NCII/CSAM等）
是 → 禁止。不得於歐洲使用。在此終止。
否 → 進入第4階段。

第4階段 風險分級分類（第2章）

問題 A：是否與人類互動（聊天機器人）或生成合成內容、深偽技術（Deepfake）？
是 → 透明度義務對象。第50條義務（第4、5、6章）。2026年8月2日生效。

問題 B：是否開發或進行實質修改後推出通用人工智慧模型？（第7章）
是 → 通用人工智慧模型提供者。第53條義務（第8章）。若超過 10^25 FLOPs 則附加第55條義務（第9章）。

問題 C：是否作為附件一產品（醫療器材、汽車、機械類等）的安全元件，且該產品屬於第三方適合性評估對象？（第11章）
是 → 附件一高風險。無過濾機制。2028年8月2日生效。適用第12、13章義務。

問題 D：是否屬於附件三八大領域（生物辨識、關鍵基礎設施、教育、就業、必要服務、執法、移民、司法）？（第11章）
是 → 附件三高風險候選對象。進入第5階段過濾機制。2027年12月2日生效。

以上皆未觸及 → 最低風險。無特別義務。

第5階段 附件三過濾機制（第11章）

問題：是否符合第6條第3項四項例外情形之一（狹義程序性任務、改善人類決策結果、模式偵測、準備工作）？
是，且未對自然人進行個人化分析（Profiling） → 可排除為高風險。惟須將依據作成書面文件並登記於歐盟資料庫。
進行個人化分析（Profiling） → 排除例外適用。確定為高風險。
否 → 確定為附件三高風險。適用第12、13章義務。

第6階段 韓國法平行確認（第16、17章）

即使完成歐盟分類，亦須一併對照韓國《人工智慧基本法》（高影響第2條、標示第31條、國內代理人第36條）與個人資料法規（GDPR、PIPA）。遵守一方並不免除另一方之義務。

附錄 3 自我檢測檢核表

印出來放在桌上使用的用途。請對各項目回答「是」或「否」，並在對應章節中確認回答為「否」的項目。

A. 適用範圍與角色（第1章）

是否已確認我們的人工智慧是否接觸到歐盟市場、使用者或產出物其中之一？
是否已針對各個系統確定我們屬於提供者、部署者、進口商或經銷商中的哪一種角色？
作為第三國提供者，若將高風險人工智慧或通用模型投放至歐洲市場，是否已指定歐盟境內授權代表（第22條、第54條）？
是否已實際將技術文件與相關資訊移交給授權代表？

B. 禁止事項（第15章）

是否已將國內使用的人工智慧對照第5條的八大禁止事項逐一檢核？
是否未在職場或教育機構中使用情緒識別工具？
是否未依綜合評分對個人進行評斷，並在其不相關的領域給予不利待遇？
生成式模型是否具備防止產出非自願性親密影像（NCII）及兒童性虐待材質（CSAM）的技術防護措施，且能證明其有效運作？

C. 透明度（第4、5、6章）

與歐洲使用者對話的聊天機器人或語音機器人，首頁畫面是否設有人工智慧告知事項？
生成式產出物是否帶有機器可讀標示，且該標示能否承受截圖與重新上傳？
製作並發布深偽內容（Deepfake）時，是否進行能讓人識別的公開揭露？
標示機制是否採用多層架構（C2PA 中繼資料 + 隱形浮水印）？
是否已計畫在 2027 年 2 月 2 日前具備浮水印偵測的跨平台互通性？

D. 通用模型（第7、8、9、10章）

是否已評估我們的模型是否具備相當程度的通用性，並已跨越通用人工智慧模型的門檻？
是否已備妥技術文件（附錄 XI）、下游資訊（附錄 XII）、著作權政策及訓練內容摘要？
是否已透過兩種計算方式，確認累計訓練計算量是否超過 10^25 FLOPs？
若欲適用開放原始碼豁免條款，是否已符合四大自由與免費條件，並落實留存的兩項義務（著作權與摘要）？

E. 高風險（第11、12、13、14章）

是否已評估我們的產品是否依據附錄 I 或附錄 III 被歸類為高風險人工智慧？
若欲採行篩選條件豁免，是否已將該判斷依據做成書面文件備查？
是否已具備風險管理（第9條）、資料治理（第10條）、技術文件與日誌紀錄（第11條、第12條）、人工監督（第14條），以及正確性、強健性與安全性（第15條）？
是否已有符合性評估、CE 標誌及登錄歐盟資料庫（第43條、第48條、第49條）的計畫？
若作為部署者，是否已確認是否需向勞工發出通知及進行基本權影響評估（第27條）？
在寬限期內，是否已於現階段啟動資料治理？

F. 韓國法律與個人資料（第16、17章）

是否已確認是否適用韓國《AI基本法》的高影響人工智慧定義及標示義務（第31條）？
若為進軍韓國的外國業者，是否已確認韓國國內代理人（第36條）要件；若為我們，是否已確認歐盟授權代表要件？

是否已確認歐洲使用者的個人資料會同時受到 GDPR 與韓國《個人資料保護法》（PIPA）的管轄？
是否已分別在歐洲與韓國取得訓練資料個人資料處理的合法依據？

G. 執行體系（第18、19、20、21章）

是否已按部門建立人工智慧清冊（包含影子 AI / 隱藏的人工智慧）？
文件內容是否與實際系統運作一致，且系統更新時是否同步綁定文件更新？
供應商合約中是否已納入資訊提供、協助與變更通知條款？
是否已根據相應角色落實人工智慧素養（第4條）並保存紀錄？
是否已預先確定回應主管機關資訊調閱請求的窗口與程序？

附錄 4 第50條告知文案範例（韓文・英文）

這並非直接套用的最終完成文，而是起點。實際文案應根據服務情境與法律諮詢進行調整。告知最晚應於首次互動或曝光時，以明確且可識別的方式，並符合無障礙（可存取性）要求予以呈現。

1. 聊天機器人・對話型代理（第50條第1項）

韓文
本諮詢由人工智慧進行。若您希望與真人客服人員連線，請於下方提出請求。

英文
You are chatting with an AI system. If you would like to speak with a human agent, please request it below.

簡短型
韓文 本服務為人工智慧客服。
英文 This is an AI assistant.

2. 語音代理（第50條第1項）

韓文
您好。這是人工智慧語音服務。接下來的回應將由人工智慧進行。

英文
Hello. This is an automated AI voice service. Your call will be handled by an AI system.

3. 生成式內容標示（第50條第2項，供人類辨識之標示）

韓文
本內容係由人工智慧生成。

英文
This content was generated by AI.

圖像・影片用圖說
韓文 人工智慧生成圖片
英文 AI-generated image

注意 除此肉眼可見之標示外，檔案中尚需一併包含機器可讀標示（如 C2PA 等），始符合第50條第2項之規定。請參閱第5章。

4. 深度偽造揭露（第50條第4項）

韓文
本影片包含利用人工智慧生成或操縱之畫面。

英文
This video contains images or audio that have been artificially generated or manipulated using AI.

藝術・諷刺用途之減輕標示（不影響欣賞體驗之方式）
韓文 本作品使用了人工智慧合成技術。（標示於片尾名單等處）
英文 This work uses AI-synthesised elements. (in end credits or equivalent)

5. 公益性文本（第50條第4項）

韓文
本文章由人工智慧生成，並經編輯審查後發布。

英文
This text was generated by AI and published under human editorial review.

注意 若缺乏真人實質審查與編輯責任，則不得主張適用此例外規定。請參閱第4章。

6. 情緒識別・生物辨識分類告知（第50條第3項）

韓文
本服務利用人工智慧分析您的情緒狀態。（或分類您的生物特徵資訊。）

英文
This service uses AI to analyse your emotional state. (or to categorise biometric data.)

注意 於職場與教育機構中使用情緒識別屬禁止行為（第5條），而非僅需履行告知義務。請參閱第15章。

附錄 5 技術文件目錄範本

高風險系統為附錄 IV、通用模型為附錄 XI 所要求項目的目錄範例。實際撰寫時請對照該附錄原文。文件必須與實際系統一致，更新時應一併修正（第19章）。

A. 高風險人工智慧系統技術文件（基於附錄 IV）

1. 系統一般資訊
1.1 目的、提供者、版本
1.2 系統互動之其他系統與硬體
1.3 使用者介面說明
1.4 供部署者使用之指引

2. 系統之架構與開發
2.1 開發方法與階段
2.2 設計規格、系統架構
2.3 資料要求（訓練、驗證與測試資料之來源、範圍及特性）
2.4 人類監督措施及其設計
2.5 預先確定之變更與持續維持符合性之方法

3. 系統之效能與限制
3.1 準確性、強健性與網路安全指標及測試結果
3.2 可預見之故障與風險
3.3 對特定群體產生效能差異之可能性

4. 風險管理
4.1 風險管理系統說明（第9條）
4.2 已識別之風險與減緩措施

5. 變更履歷與維護
5.1 各版本變更紀錄
5.2 上市後監測計畫（第72條）

6. 符合性
6.1 所採用之調和標準與共通規範
6.2 歐盟符合性聲明副本（第47條）

B. 通用人工智慧模型技術文件（基於附錄 XI）

1. 模型一般資訊
1.1 模型名稱、提供者、發布日期、版本
1.2 架構與參數數量
1.3 輸入與輸出格式
1.4 授權條款

2. 開發資訊
2.1 訓練方法與過程
2.2 訓練與測試資料之類型及來源、資料處理方法
2.3 計算資源與訓練計算量（FLOPs 估算與依據）
2.4 已知能力與限制

3. 供下游提供者使用之資訊（附錄 XII）

3.1 模型整合所需之技術資訊

3.2 能力、限制與使用條件

4. 著作權與訓練內容

4.1 歐盟著作權法遵循政策（Directive 2019/790 第4條第3項）

4.2 訓練內容公開摘要（第53條第1項第d款，辦公室標準格式）

5. 系統性風險（適用時，第55條）

5.1 模型評估與對抗性測試結果

5.2 系統性風險評估與減緩措施

5.3 重大事故應變機制

5.4 網路安全措施

附錄 6 人工智慧供應商契約條款範例

本篇為草案。實際契約應經過法律諮詢後確定。各條款之背景請參閱第 20 章。

1. 定義條款

本契約所稱人工智慧法，係指 Regulation (EU) 2024/1689 及其修正案與下位規範。提供者、部署者、高風險、通用人工智慧模型等用語，依人工智慧法所定之定義。

2. 角色與地位確認條款

當事人就本契約所供應之人工智慧系統或模型，確認各自於人工智慧法上之地位（提供者・部署者・進口商・分銷商）如下。（以表格明示。）

供應商應以書面告知該系統是否屬於高風險或系統性風險，及其依據。

3. 資訊提供條款

供應商應於契約開始履行前及受請求時，無遲延地提供需求方履行其於人工智慧法上義務所需之資訊（包含附錄 XII 及第 13 條資訊）。

屬通用模型者，供應商應提供技術文件、能力與限制、整合條件、著作權合規及訓練內容摘要相關之資訊。

4. 變更通知條款

供應商於所供應之系統或模型之性能・結構・訓練資料有重大變更時，應事前通知需求方。因未經通知之變更致需求方違反義務者，其責任由供應商承擔。

5. 協助條款

主管機關為資訊請求、調查、稽核或重大事故因應時，當事人應相互誠實協助，並於合理期間內提供必要資料。

6. 保密條款

依本契約所提供之資訊中屬於營業秘密者，未經對方事前同意，不得用於目的外或向第三人公開。但依法令或主管機關之合法要求所為之提供，不在此限。

7. 境內代理人條款（適用時）

供應商為位於第三國之提供者時，供應商應依人工智慧法第 22 條或第 54 條選任歐盟境內代理人，並將其資訊通知需求方，且應提供技術文件與資訊以利代理人執行任務。

8. 責任與免責界限條款

當事人間損害之分擔依本契約所定。但確認人工智慧法直接賦予各當事人之管制上義務，不問本契約之責任分配或免責約定，各自仍應承擔。

9. 準據法與管轄

（視個案而定。應檢討與歐洲市場相關爭議之準據法與管轄選擇對實際執行之影響。）

附錄 7 用語對照表（韓文・英文・法條）

歐盟人工智慧法 Regulation (EU) 2024/1689，韓國 AI 基本法 2026 年 7 月 21 日施行現行本基準。條文確認 2026 年 7 月 23 日。

歐盟人工智慧法

提供者 / provider / 第 3 條第 3 款
部署者 / deployer / 第 3 條第 4 款
境內授權代表 / authorised representative / 第 3 條第 5 款（高風險第 22 條、通用模型第 54 條）
進口商 / importer / 第 3 條第 6 款
分銷商 / distributor / 第 3 條第 7 款
營運者 / operator / 第 3 條第 8 款
投放市場 / placing on the market / 第 3 條第 9 款
於市場提供 / making available on the market / 第 3 條第 10 款
啟用 / putting into service / 第 3 條第 11 款
人工智慧素養 / AI literacy / 第 4 條
禁止之人工智慧實踐 / prohibited AI practices / 第 5 條
高風險人工智慧系統 / high-risk AI system / 第 6 條（附錄 I・III）
風險管理系統 / risk management system / 第 9 條
資料與資料治理 / data and data governance / 第 10 條
技術文件 / technical documentation / 第 11 條（附錄 IV）
紀錄保存與日誌 / record-keeping, logs / 第 12 條
透明度與資訊提供 / transparency and provision of information / 第 13 條
人類監督 / human oversight / 第 14 條
準確性、強健性與網路安全 / accuracy, robustness and cybersecurity / 第 15 條
高風險提供者之義務 / obligations of providers / 第 16 條
品質管理系統 / quality management system / 第 17 條
價值鏈上之責任 / responsibilities along the AI value chain / 第 25 條
部署者之義務 / obligations of deployers / 第 26 條
基本權影響評估 / fundamental rights impact assessment (FRIA) / 第 27 條
調和標準 / harmonised standards / 第 40 條
共通規範 / common specifications / 第 41 條
符合性評估 / conformity assessment / 第 43 條
歐盟符合性聲明 / EU declaration of conformity / 第 47 條
CE 標誌 / CE marking / 第 48 條
登記 / registration / 第 49 條
特定人工智慧系統之透明度義務 / transparency obligations for certain AI systems / 第 50 條
深層偽造 / deepfake / 第 50 條第 4 項
內容標示實踐守則 / codes of practice on labelling / 第 50 條第 7 項
通用人工智慧模型 / general-purpose AI model / 第 3 條第 63 款（分類第 51 條）
系統性風險 / systemic risk / 第 3 條第 65 款（第 51 條）
具系統性風險之通用人工智慧模型 / GPAI model with systemic risk / 第 51 條
通用模型提供者之義務 / obligations for providers of GPAI models / 第 53 條
訓練內容公開摘要 / summary of content used for training / 第 53 條第 1 項 d 款
通用模型境內授權代表 / authorised representatives of GPAI providers / 第 54 條

具系統性風險之通用模型提供者義務 / obligations for GPAI with systemic risk / 第 55 條
通用人工智慧實踐守則 / codes of practice / 第 56 條
人工智慧監理沙盒 / AI regulatory sandbox / 第 57 條
真實條件測試 / testing in real world conditions / 第 60 條
人工智慧辦公室 / AI Office / 第 3 條第 47 款
歐洲人工智慧委員會 / European Artificial Intelligence Board / 第 65 條
國家主管機關 / national competent authorities / 第 70 條
高風險歐盟資料庫 / EU database for high-risk AI systems / 第 71 條
上市後監測 / post-market monitoring / 第 72 條
重大事故通報 / reporting of serious incidents / 第 73 條
市場監督 / market surveillance / 第 74 條
提出申訴權 / right to lodge a complaint / 第 85 條
個人決策說明請求權 / right to explanation / 第 86 條
罰鍰 / penalties / 第 99 條
通用模型提供者罰鍰 / fines for GPAI providers / 第 101 條
生效與施行 / entry into force and application / 第 113 條

韓國 AI 基本法

人工智慧 / artificial intelligence / 第 2 條第 1 款
人工智慧系統 / AI system / 第 2 條第 2 款
人工智慧技術 / AI technology / 第 2 條第 3 款
高影響人工智慧 / high-impact AI / 第 2 條第 4 款
生成式人工智慧 / generative AI / 第 2 條第 5 款
人工智慧事業 / AI business operator / 第 2 條第 7 款
人工智慧開發事業 / AI development business operator / 第 2 條第 7 款 a 目
人工智慧利用事業 / AI use business operator / 第 2 條第 7 款 b 目
使用者 / user / 第 2 條第 8 款
受影響者 / affected person / 第 2 條第 9 款
訓練資料 / training data / 第 2 條第 12 款
人工智慧倫理原則 / AI ethics principles / 第 27 條
透明度確保義務 / transparency obligation / 第 31 條
安全性確保義務 / safety obligation / 第 32 條
高影響人工智慧之確認 / confirmation of high-impact AI / 第 33 條
高影響事業之責任義務 / obligations of high-impact AI operators / 第 34 條
高影響影響評估 / high-impact AI impact assessment / 第 35 條
指定國內代表人 / designation of domestic representative / 第 36 條
事實調查 / fact-finding investigation / 第 40 條
罰則 / criminal penalties / 第 42 條
罰鍰 / administrative fines / 第 43 條

附錄 8 EU 主管機關與韓國國內支援機構

聯絡方式變動頻繁，故不列出具體電話號碼，改為記載官方機構及其對外窗口。實際查詢前，請先至各機構官方網站確認最新窗口。

EU 方面

歐盟執委會人工智慧辦公室 (European Commission, AI Office)
通用人工智慧模型監督，以及指引、實踐守則、表格發布的核心機構。
官方入口網站 digital-strategy.ec.europa.eu (Shaping Europe's digital future)

人工智慧法服務台 (AI Act Service Desk)
提供條文、附錄、問答集的執委會官方諮詢窗口。
官方入口網站 ai-act-service-desk.ec.europa.eu

歐洲人工智慧委員會 (European Artificial Intelligence Board)
成員國協調機構。透過執委會入口網站公開其活動。

成員國國家主管機關 (national competent authorities)
由各成員國指定的市場監督與通報機關。必須個別確認目標進駐成員國的主管機關。各成員國規定有所不同，相關清單會在執委會入口網站更新。

EU 官方法規入口網站 (EUR-Lex)
確認 Regulation (EU) 2024/1689 及修訂案之官方原文與公報刊載內容。
官方入口網站 eur-lex.europa.eu

韓國方面

科學技術情報通信部
《AI基本法》主管部會。為高影響確認、國內代理人申報、安全性執行結果提交之窗口。
官方入口網站 msit.go.kr

個人資料保護委員會 (PIPC)
主管《個人資料保護法》，發布人工智慧相關個人資料指引。韓-歐盟資料適足性認定相關窗口。
官方入口網站 pipc.go.kr

韓國網際網路振興院 (KISA)
個人資料與資安實務支援。

大韓貿易投資振興公社 (KOTRA)
支援拓展歐洲市場之企業，提供當地法規資訊。

中小企業暨新創企業部及其附屬機構
支援中小企業因應海外法規與認證。

實務建議
應先確認目標進駐成員國的國家主管機關以及監理沙盒（第22章）營運與否。EU 方面的第一手資料務必以上述官方入口網站的原文進行比對。第二手解析（律師事務所電子報等）可用於掌握方向，但日期與條文必須以第一手資料為準進行確認。

附錄 9 主要條文要旨

如需法律引用，請務必對照 EUR-Lex 的 Regulation (EU) 2024/1689 官方原文（及 2026 年修訂版）。以下為實務中常查閱條文之要旨整理，非條文原文本身。

歐盟人工智慧法

第 2 條（適用範圍）
適用於在歐盟投放人工智慧之提供者、歐盟境內之部署者，以及產出物於歐盟境內使用之第三國提供者與部署者。此為域外適用之依據。（第 1 章）

第 3 條（定義）
提供者、部署者、進口商、分銷商、歐盟授權代表、投放市場、投入服務、通用人工智慧模型、系統性風險等定義。（第 1、7 章）

第 4 條（人工智慧素養）
確保處理人工智慧之人員具備相關素養之義務。於 2025 年 2 月 2 日起適用。（第 21 章）

第 5 條（禁止之行為）
禁止潛意識操縱、利用脆弱性惡意剝削、社會評分、無差別收集臉部影像、工作場所及教育機構之情緒識別、推論敏感資訊之生物特徵分類、即時遠距生物特徵識別等。2026 年修訂版新增禁止生成 NCII（非自願性色情影像）及 CSAM（兒童性虐待物）（2026 年 12 月 2 日）。（第 15、5 章）

第 6 條（高風險分類）
將高風險分為附錄 I 之產品內建型與附錄 III 之獨立型。第 3 項為附錄 III 之篩選條件（若無重大風險則排除，但進行輪廓塑造時排除例外適用）。（第 11 章）

第 9 條至第 15 條（高風險提供者要件）
風險管理、資料治理、技術文件、紀錄保存、透明度與資訊提供、人工監督、正確性、強健性及資安防護。（第 12 章）

第 22 條・第 54 條（歐盟授權代表）
第三國提供者於將高風險（第 22 條）或通用模型（第 54 條）投放至歐盟前，須指定歐盟授權代表。代表負責保存技術文件（通用模型為 10 年）並與主管機關合作。（第 1 章）

第 26 條・第 27 條（部署者義務與基本權影響評估）
遵循使用說明、配置人工監督、保存日誌（至少 6 個月）、通知勞工。公務機關及提供公共服務之部署者，須針對附錄 III 系統進行基本權影響評估（FRIA）。（第 13 章）

第 43 條・第 48 條・第 49 條（符合性評估、CE 標誌與登記）
附錄 III 第一領域得由公告機構介入，其餘則實施內部控制。須標示 CE 標誌並於歐盟資料庫完成登記。（第 12 章）

第 50 條（透明度）
第 1 項聊天機器人告知（提供者）、第 2 項合成內容之機器可讀標示（提供者）、第 3 項情緒識別與生物特徵分類告知（部署者）、第 4 項深偽與涉及公益之文字揭露（部署者）、第 5 項時機與方式。第 7 項實踐守則。於 2026 年 8 月 2 日起適用。（第 4、6 章）

第 51 條・第 53 條・第 55 條（通用模型）
第 51 條系統性風險推定（10^25 FLOPs）。第 53 條基本義務（技術文件、下游資訊、著作權政策、訓練內容摘要）。第 55 條系統性風險加重義務（評估、紅隊演練、事故通報、資安防護）。（第 7、8、9 章）

第 99 條（罰鍰）
違反禁止規定處 3,500 萬歐元或全球營業額 7%；違反其他義務處 1,500 萬歐元或 3%；提供不實資訊處 750 萬歐元或 1%。兩者取其高者。（第 2 章）

第 113 條（生效與適用）
分階段適用日期之依據條文。（第 3 章）

相關歐盟法律

Directive (EU) 2019/790 第 4 條第 3 項
著作權人對文字與資料探勘之權利保留。通用模型著作權政策之依據。（第 8 章）

韓國 AI 基本法

第 2 條第 4 款（高影響人工智慧）
對生命、身體安全及基本權利產生重大影響，並應用於列舉之 11 個領域中之人工智慧。（第 16 章）

第 31 條（確保透明度義務）
高影響與生成式 AI 使用前告知、標示生成式產出物、深偽明確識別告知。（第 16 章）

第 32 條（確保安全性義務）・第 34 條（高影響業者責任）
達運算量標準以上系統之風險管理；高影響業者之風險管理、說明及文件保存。（第 16 章）

第 36 條（國內代理人）
超過營業額與使用者人數標準之境外業者指定與申報國內代理人。（第 16 章）

歐盟人工智慧法解說書

ISBN |

作者 | 金炅鎭

發行人 | 金炅鎭

出版 | 金炅鎭律師出版社

出版登記 | 2025. 3. 10. (第2025-000015號)

地址 | 韓國首爾特別市東大門區典農路91，白日大廈304號

電話 | +82-2-6338-1905

電子郵件 | kimkj008@gmail.com

附註）本書部分內容與編輯過程借助了人工智慧工具。

支持這本 AI 圖書館書籍

若本書對您有所助益，歡迎透過 PayPal 支持
未來的翻譯、公開版本與開放 AI 知識計畫。

PayPal

paypal.me/kimkjj

掃描 QR 碼或開啟上方連結。