

# 歐盟資料探勘限制規定於生成式 AI 訓練階段之適用性初探

陳俞安

## 壹、前言

## 貳、歐盟關於生成式 AI 訓練階段利用著作之法制架構

- 一、數位單一市場著作權指令
- 二、人工智慧法與著作權義務的具體規範
- 三、數位單一市場著作權指令與人工智慧法之交互關係

## 參、資料探勘限制規定之界線與生成式 AI 訓練之適用

- 一、生成式 AI 訓練與資料探勘的本質區別
- 二、資料探勘限制規定的法律困境
- 三、歐盟司法實務之觀察

## 肆、歐盟生成式 AI 制度檢討與未來展望

- 一、現行生成式 AI 制度檢討
- 二、歐盟未來展望與發展

## 伍、結語

---

作者現為經濟部智慧財產局國際及法律事務室科長。  
本文相關論述僅為一般研究探討，不代表任職單位之意見。

## 摘要

隨著生成式人工智慧（Artificial Intelligence, AI）的快速發展，其於訓練過程大量利用他人著作之情形，引發廣泛爭議。針對此類利用，歐盟現行著作權之法律基礎主要為 2019 年通過之「數位單一市場著作權指令（下稱 CDSM 指令）」有關資料探勘（Text and Data Mining）之權利限制規定，以及 2024 年發布之「人工智慧法」要求 AI 提供者遵守 CDSM 指令之退出機制，並履行公開訓練資料詳細摘要等透明化義務。

然而，多數會員國對於資料探勘限制規定及退出機制是否適用於 AI 訓練仍存疑義。歐洲議會 2025 年發布之研究報告亦指出，資料探勘旨在進行資料分析與知識提取，難以涵蓋生成式 AI 訓練內化資料並重組內容的特性。本文藉由梳理現行歐盟法制架構，以歐洲議會研究報告及修法動向為核心，探討生成式 AI 訓練與資料探勘在本質上之差異，與資料探勘限制規定於生成式 AI 訓練適用所面臨之困境，並進一步分析未來制度可能改善的方向，供未來研究參考。

關鍵字：生成式 AI、AI 與著作權、人工智慧法、資料探勘、訓練資料、通用型 AI 實務準則

Generative Artificial Intelligence (Generative AI)、AI and Copyright、AI Act、Text and Data Mining、Training Data、General-Purpose AI Code of Practice

## 壹、前言

隨著生成式人工智慧（Artificial Intelligence, AI）的快速發展，其在模型訓練過程需要大量利用他人著作，且可能生成與原著作競爭的內容，此種重製他人著作之行為是否構成合理使用引發高度爭議，因其利用方式雖與傳統資料探勘（Text and Data Mining）概念相似，卻又遠超過其範圍。歐盟於 2019 年通過之「數位單一市場著作權指令（Directive on Copyright in the Digital Single Market，下稱 CDSM 指令）」為分析大量資料等創新技術需求，訂有資料探勘權利限制與例外規定（下稱資料探勘限制規定）。2024 年頒布之「人工智慧法（AI Act）」（下稱 AI 法）進一步要求 AI 提供者遵守 CDSM 指令中有關資料探勘限制規定賦予權利人之退出權，試圖以現有著作權法制架構回應生成式 AI 帶來的挑戰。

然而，從歐盟執委會於 2024 年對會員國進行之政策問卷總結<sup>1</sup>顯示，多數會員國對於資料探勘限制規定是否適用 AI 訓練仍存疑義。歐洲議會於 2025 年 7 月發布「生成式 AI 與著作權—訓練、創作、規範」之研究報告（下稱生成式 AI 研究報告）<sup>2</sup>更指出，現行 CDSM 指令有關資料探勘限制規定係為進行資料分析與知識提取，難以涵蓋生成式 AI 訓練內化資料並重組內容的特性。因此本文旨在探討歐盟資料探勘限制規定是否可適用於生成式 AI 之訓練？

本文將依序爬梳現行歐盟關於 AI 與著作權之法制架構，釐清生成式 AI 訓練與資料探勘的本質差異，進而探討資料探勘限制規定在生成式 AI 訓練上的適用困境，最後提出未來制度可能改善的方向，期望為後續研究提供參考。

<sup>1</sup> Policy questionnaire on the relationship between generative Artificial Intelligence and copyright and related rights – Revised Presidency summary of the Member States contributions. The Council of the EU and the European Council, <https://www.consilium.europa.eu/en/documents/public-register/public-register-search/?AllLanguagesSearch=false&OnlyPublicDocuments=false&DocumentNumber=16710%2F24&DocumentLanguage=EN> (last visited Jan. 1, 2026).

<sup>2</sup> Lucchi N. Generative AI and Copyright: Training, Creation, Regulation. Policy Department for Justice, Civil Liberties and Institutional Affairs, European Parliament, June 2025.

## 貳、歐盟關於生成式 AI 訓練階段利用著作之法制架構

### 一、數位單一市場著作權指令

CDSM 指令於 2019 年 3 月通過，係為解決數位時代下傳統著作權法面臨之挑戰，其中一項重要改革即是訂定資料探勘限制規定，以促進數據分析和 AI 技術的發展<sup>3</sup>。

依 CDSM 指令第 3 條規定，僅限「研究組織或文化遺產機構」為「科學研究之目的」進行資料探勘，在適用主體上強調非營利性及公益性，並限於「合法接觸 (lawful access)」的著作，由於適用範圍相當限縮，因此禁止權利人以契約條款排除之<sup>4</sup>；而第 4 條則不限適用主體及目的，同樣限於就「合法接觸」的著作進行資料探勘，但賦予權利人以適當方式明示保留作品被使用於資料探勘的權利（例如以機器可讀之方法標示保留），即「退出機制 (Opt-out)」，透過原則開放使用之制度設計，簡化資料利用程序及成本，同時維持著作權人之控制權，試圖兼顧技術創新與著作權保護。

然而所謂「合法接觸」之界定<sup>5</sup>與「退出機制」之執行，各會員國間的實踐方式有很大差異，例如「網路上可自由接觸之內容 (freely available online)」是否即屬於「合法接觸<sup>6</sup>」、退出方式至今尚無統一標準，使得資料探勘限制規定在實務運作上充滿不確定性。

### 二、人工智慧法與著作權義務的具體規範

2024 年 7 月發布之 AI 法<sup>7</sup>，作為歐盟全面性監管 AI 之法律，目的在於確保 AI 模型能尊重基本權利、安全和道德原則以維護公共利益，本身並未創設新的著

<sup>3</sup> 2019 年指令發布時之新聞稿，歐洲議會網站，<https://www.europarl.europa.eu/legislative-train/package-better-access-to-digital-goods-services/file-jd-directive-on-copyright-in-the-digital-single-market> (last visited Jan. 1, 2026).

<sup>4</sup> 歐盟 CDSM 指令第 7 條。

<sup>5</sup> CDSM 指令序言第 14 點指出依據開放接觸 (open access) 政策、簽訂授權契約等合法取得方式 (如付費訂閱期刊、資料庫)，以及網路上可自由接觸之內容皆屬之。

<sup>6</sup> Lucchi N., *supra* note 2, at 29, 232-233. 斯洛維尼亞將「網路上可自由接觸內容」排除在合法接觸要件之外。

<sup>7</sup> 2024 年 8 月 1 日正式生效，其中關於通用 AI 模型相關義務及罰則，於 2025 年 8 月開始實施。

作權規定，但其於序言第 105 點<sup>8</sup>明確指出，CDSM 指令之資料探勘限制規定亦適用生成式 AI 的開發與訓練，惟當權利人依 CDSM 指令第 4 條明示退出時，通用 AI 模型提供者（General-Purpose AI，下稱 GPAI 開發商）<sup>9</sup>即須取得權利人授權始能利用。此外，AI 法第 53 條要求 GPAI 開發商遵守歐盟著作權法，並應依據 AI 辦公室提供的範本公開訓練資料詳細摘要。

為促進 GPAI 開發商落實前述義務，確保 CDSM 指令既有之著作權保護機制得以落實，歐盟 AI 辦公室自 2024 年底陸續與利害關係人磋商，並於 2025 年 8 月發布最終版本之通用型 AI 訓練內容公開摘要範本（Template for the Public Summary of Training Content for General-Purpose AI Models，下稱 AI 訓練摘要範本）及通用型 AI 實務準則（General-Purpose AI Code of Practice，下稱 AI 準則）。

首先，AI 訓練摘要範本將訓練資料來源區分為公開資料、已獲授權的私人資料、網路爬取資料、用戶資料、合成資料等，並依資料性質訂有不同程度之揭露要求。例如對於已取得授權之資料集，僅需確認是否簽訂授權協議及說明資料形式（如文字或音訊）；相較之下，若屬網路爬取資料，則須提供較多來源說明。然而，該範本仍遭批評僅停留於概括性摘要，難以使權利人判斷其作品是否被納入訓練<sup>10</sup>，且即使 GPAI 開發商未履行揭露義務可能面臨高額行政罰鍰<sup>11</sup>，著作權人仍無從因而獲得直接救濟或利益<sup>12</sup>。

<sup>8</sup> AI 法序言第 105 點：通用模型，尤其是大型生成模型，……這類模型的開發和訓練需要存取大量文本、圖像、影片和其他資料。在這種情況下，文字和資料探勘技術可能被廣泛用於檢索和分析此類可能受著作權和相關權利保護的內容。除非適用相關著作權例外和限制，否則任何受著作權保護內容的使用都需要獲得相關權利人的授權。(EU)2019/790 指令引入了例外和限制，允許在特定條件下為了文字和資料探勘的目的重製和提取作品或其他題材。根據這些規則，權利人可以選擇保留其對作品或其他題材的權利，以防止資料探勘，除非這是為了科學研究目的而進行的。在（權利人）以適當方式明確保留選擇退出之權利的情況下，通用 AI 模型提供者如果要對這類作品進行資料探勘，需要獲得權利人的授權。

<sup>9</sup> AI 法序言第 97、99 點：通用 AI 模型係指透過無監督學習或強化學習等方式訓練大量資料，以勝任執行各種不同任務的能力之模型。大型生成式 AI 模型即通用 AI 模型的典型例子。

<sup>10</sup> Barry Scannell & Leo Moore, EU Releases AI Training Data Template, WILLIAM FRY, <https://www.williamfry.com/knowledge/eu-releases-ai-training-data-template> (last visited Jan. 9, 2026).

<sup>11</sup> AI 法第 101 條規定，GPAI 開發商違反法定義務，得處以最高 1,500 萬歐元或前一年全球年營業額 3% 之罰鍰。

<sup>12</sup> Quintais, João Pedro. Generative AI, copyright and the AI Act. 56 Computer Law & Security Review: The International Journal of Technology Law and Practice, 2025.

其次，AI 準則從「透明度、著作權及安全性」三大面向提供 GPAI 開發商具體執行建議，其中透明度和著作權章節適用於所有 GPAI 模型，而安全性章節僅適用於存在系統性風險者：

- (一) 在透明度部分，須向 AI 辦公室及下游開發者提交技術文件（模型文件表單），紀錄模型訓練過程、內部運作細節及模型使用限制等，作為符合法規要求之證據與審查依據。
- (二) 在著作權部分，1、僅能使用合法接觸的資料訓練 AI 模型，不得規避科技保護措施，亦不得爬取「持續且反覆侵權」的盜版網站資料；2、應採取適當防護措施以避免模型輸出侵權內容；3、遵守機器人排除協議（Robot Exclusion Protocol, robots.txt），以確保權利保留聲明得以發揮保護效果。

然而，該準則作為自願性工具，主要著重於提供 GPAI 開發商操作指引，權利人團體普遍認為該準則欠缺有效監管機制，反而使 GPAI 開發商可藉由自願簽署以減輕行政負擔<sup>13</sup>，與 AI 法第 53 條保障著作人權益之原意背道而馳<sup>14</sup>。

### 三、數位單一市場著作權指令與人工智慧法之交互關係

從前述規範可觀察到，AI 法訂定遵守著作權法義務，促使 GPAI 開發商於設計階段即納入遵守著作權法考量，間接強化了 CDSM 指令資料探勘限制規定執行的效果。此外，在生成式 AI 發展的脈絡下，資料探勘限制規定已不再被視為單純鼓勵技術創新的工具，而須與 AI 法所訂義務共同解讀，例如，「合法接觸」的範圍，基於 CDSM 指令第 3 條為科學研究目的之公益性，原有觀點認為只要未規避技術保護措施或付費機制，即使接觸到未經授權公開於網路上的內容（即非法內容）也不因而阻礙其適用<sup>15</sup>。然而，第 4 條為商業目的是否能比照適用，歐

<sup>13</sup> 歐盟執委會 2025 年 7 月 10 日新聞稿，General-Purpose AI Code of Practice now available，<https://digital-strategy.ec.europa.eu/en/news/general-purpose-ai-code-practice-now-available> (last visited Jan. 9, 2026).

<sup>14</sup> CISAC, Joint statement by a broad coalition of rightsholders active across the EU's cultural and creative sectors regarding the AI Act implementation measures adopted by the European Commission, <https://www.cisac.org/Newsroom/articles/joint-statement-broad-coalition-rightsholders-active-across-eus-cultural-and> (last visited Jan. 9, 2026).

<sup>15</sup> 若「合法接觸」被解釋為須確保資料的「合法來源」，研究機構將面臨高昂的審查成本與法律不確定性，且一旦經通知須移除非合法資料導致需重新訓練模型，成本極高，恐使 CDSM 指令第 3 條形同虛設，不利科學研究政策目的。資料來源：Margoni, Thomas, TDM and Generative AI: Lawful Access and opt-outs. Forthcoming in *Auterus & Media*, 15-16 (2024).

盟會員國間對此看法不一，但 AI 準則現已明確將「持續且反覆侵權的盜版網站」排除在 GPAI 開發商可「合法接觸」的範圍之外，嘗試在保障著作人權益和推動技術發展間重新取得平衡。

## 參、資料探勘限制規定之界線與生成式 AI 訓練之適用

歐洲議會在參考各會員國應對生成式 AI 的著作權法律實踐、利害關係者意見後所發布之生成式 AI 研究報告<sup>16</sup>指出，資料探勘限制規定適用於生成式 AI 時，可能產生若干制度性風險與法律適用上的困境。本文遂依據該報告所提出之核心觀點，結合相關評論進行綜整與分析如下。

### 一、生成式 AI 訓練與資料探勘的本質區別

根據 CDSM 指令第 2 條第 2 款之定義，資料探勘係指「任何旨在分析數位形式文本與資料之自動分析技術，以便生成包括但不限於模型、趨勢、相關性等內的資訊」。指令訂定時之 AI 發展著重於醫療保健、安全等領域之應用<sup>17</sup>，主要目的在萃取知識、發現新趨勢<sup>18</sup>，且生成之資訊並不涉及原始著作的「表達」。

反觀生成式 AI 模型透過編碼與序列化（tokenization）處理資料，並不斷調整內部參數（權重）以重組、模仿原始資料的風格與內容，甚至涉及「記憶」（memorization）原始著作，故生成內容可能重現原始著作<sup>19</sup>。以 Getty Images 指控 Stability AI 未經授權使用其圖片訓練模型案為例，部分生成內容甚至出現

<sup>16</sup> 同註 2。該報告雖無法律效力，但其綜整歐盟各會員國政策調查回應意見，反映出實務上許多懸而未決的問題，仍具有相當價值。

<sup>17</sup> 2017 年 CDSM 指令期中審查報告，歐盟官網，<https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex:52017DC0228> (last visited Jan. 8, 2026).

<sup>18</sup> CDSM 指令序言第 8 點：新技術使得對數位形式的資訊（例如文字、聲音、圖像或資料）進行自動化計算分析成為可能，這通常被稱為文字和資料探勘。文字和資料探勘使得處理大量資訊以獲取新知識和發現新趨勢成為可能。

<sup>19</sup> Adam Buick, Copyright and AI training data—transparency to the rescue?, *Journal of Intellectual Property Law & Practice*, Volume 20, Issue 3, 182–192(2025).

Getty 的浮水印<sup>20</sup>，顯示生成式 AI 之應用已非單純的資料探勘。因此，直接適用資料探勘限制規定似乎忽略了本質不同而產生的風險。

## 二、資料探勘限制規定的法律困境

### （一）難以通過「三步測試（Three-step test）」<sup>21</sup>

依據歐盟資訊社會著作權指令第 5 條第 5 項規定<sup>22</sup>，CDSM 指令中關於資料探勘之限制規定須通過三步測試，以免過度侵害著作權人權益，要件包括 1、僅適用於某些特殊情況；2、不得與著作的正常利用相衝突；3、沒有不合理地損害作者的合法利益。

然而生成式 AI 訓練多屬大規模商業應用，難以被視為「特殊情況」；其次，生成內容可能與原著作市場競爭，甚至取代其市場，進而削弱原著的「正常利用」功能；最後，AI 訓練重製原始著作後，通常會為了後續模型評估、調整等需求而長期保存於資料庫，其利用範圍遠超出單純資料分析之必要限度<sup>23</sup>，且生成內容可能再現原著作<sup>24</sup>，損害權利人利益甚鉅。由此可見，生成式 AI 之訓練行為，難以通過三步測試之檢驗。

<sup>20</sup> Getty Images (US), Inc. v. Stability AI, Inc., No. 1:23-cv-00135, 4, (D. Del. filed Feb. 3, 2023).

<sup>21</sup> 「三步測試」（Three-Step Test）最早規定在伯恩公約，後來擴展至「世界智慧財產權組織著作權條約」（WIPO Copyright Treaty, WCT），係用以檢視著作權權利限制與例外規定是否過度侵害權利人之準則。

<sup>22</sup> 歐盟資訊社會著作權指令第 5 條第 5 項規定原文：The exceptions and limitations provided for in paragraphs 1, 2, 3 and 4 shall only be applied in certain special cases which do not conflict with a normal exploitation of the work or other subject-matter and do not unreasonably prejudice the legitimate interests of the rightholder.

<sup>23</sup> CDSM 指令第 4 條第 2 項規定，根據前項所進行的重製，可保留至為資料探勘目的所需的時間內（as long as is necessary for the purposes of text and data mining），即僅得於資料探勘處理過程或必要的時間內保存重製物，一旦處理過程結束，即須將其刪除。

資料來源：Members' Research Service, AI and copyright: The training of general-purpose AI, Think Tank pages of the European Parliament, <https://epthinktank.eu/2025/04/28/ai-and-copyright-the-training-of-general%E2%80%91purpose-ai/> (last visited Jan. 8, 2026).

<sup>24</sup> 如 The New York Times v. OpenAI 案，ChatGPT 生成內容再現紐約時報原文片段。

## （二）退出機制難以執行，影響授權市場發展

- 1、缺乏統一且「機器可讀」的技術或協議：不同平台或業別方法不一（如 robots.txt、TDMRep<sup>25</sup>），會員國間退出機制規範亦不相同<sup>26</sup>，德國地方法院判決<sup>27</sup>甚至將以自然語言撰寫的服務條款視為足夠之退出方式，增加退出機制之模糊性，無法降低 AI 開發商之風險。而權利人無法有效「退出」，等同無法透過與 AI 開發商簽訂授權協議獲得應有報酬。
- 2、著作權人難以有效控制其著作，對個別創作者負擔過重：一旦作品被轉換為其他格式（如影片或圖檔），或由第三方重新上傳至其他平台，權利人便無法確保對作品的持續控制。此外，創作者必須為每件作品逐一聲明「退出」，並持續監測作品是否被利用，相當耗費心力。

AI 法雖訂有透明度義務以回應此問題，然對訓練資料的揭露僅止於概括性摘要，難以使權利人確認其作品是否已被爬取或成功退出，且高度仰賴 GPAI 開發商的善意遵循。

## （三）缺乏對創作者的公平報酬機制

有論者主張 AI 學習資料與人類閱讀資料內化之行為相同<sup>28</sup>，故無須支付創作者費用，但此論點過於簡化現實，因 AI 學習在規模和深度上超越了人類閱讀的限制<sup>29</sup>，且即使是人類學習仍須支付費用、無法毫無限制的複製整個圖書館的書籍，因此為訓練 AI 而重製他人著作仍須支付報酬，

<sup>25</sup> TDMRep 是為配合 CDSM 指令第 4 條的規定而開發的，它提供了一個簡單的布林值（Boolean-based flag）供發行商標記是否保留權利，該協議的採用率在不同國家差異很大。

<sup>26</sup> 比利時、捷克明訂須採用技術協定（如 robots.txt、metadata）；法國依服務條款主張退出，缺乏機器可讀標準；德國及荷蘭明確要求須機器可讀。資料來源：Lucchi N., *supra* note 2, at 73-74.

<sup>27</sup> Kneschke v. LAION e.V., Landgericht Hamburg, Sept. 27, 2024.

<sup>28</sup> 除了 Meta 等 AI 開發商於訴訟中主張外，近期美國 Bartz v. Anthropic 案之法官也認為 AI 使用書籍訓練，就像一個「立志成為作家的讀者（a reader aspiring to be a writer）」，不是單純為了複製，而是經由閱讀後轉化，生成與原著不同的作品。

資料來源：Bartz et al. v. Anthropic PBC, No. 3:24-cv-05417, p13 (N.D. Cal. Jun 23, 2025).

<sup>29</sup> Robert Brauneis, Copyright and the Training of Human Authors and Generative Machines, 48 COLUM. J. L. & ARTS 1 (2024).

以免削弱權利人創作之動力。雖然 CDSM 指令序言第 73 點提及創作者就其內容有權獲得「適當且相稱的報酬」，但資料探勘限制規定中卻無具體補償機制，導致 GPAI 開發商得以利用大量著作訓練，獲取龐大商業利益，但創作者卻無從獲取合理報酬，受到廣泛批評，亦是創意產業與 AI 開發商分歧日益加深之主因。

### 三、歐盟司法實務之觀察

前述理論上的爭議，隨著相關案件陸續進入法院審理，正透過歐洲的司法判決逐步獲得釐清。首先，在德國攝影師 Kneschke 訴 LAION 案中，爭點在於被告作為非營利組織，能否於建置 Laion-5B 資料集時重製原告上傳至圖庫網站之照片。德國漢堡地方法院 2024 年 9 月判決認為，被告行為符合德國著作權法第 60d 條（依據 CDSM 指令第 3 條所增訂）所規定之「為科學研究目的」資料探勘限制規定，即使後續該資料集被商業公司用於 AI 模型訓練，仍不影響其原本之非商業性質。法院同時補充說明，被告重製行為亦符合德國著作權法第 44b 條一般資料探勘之目的（同 CDSM 指令第 4 條），僅因權利人已依該條規定聲明退出，故無法適用該條規定。值得注意的是，本案僅涉及建置訓練資料集之行為，屬於模型訓練前的準備階段，並未涉及生成式 AI 的實際訓練，不過判決凸顯了 CDSM 指令第 3 條與第 4 條在適用主體及限制之差異，未來可能被 AI 開發商策略運用，產生資料洗白（Data Laundering）風險<sup>30</sup>。

德國慕尼黑法院 2025 年 11 月就德國音樂著作權集體管理團體 GEMA 訴 OpenAI 案所作判決，更直接回應生成式 AI 訓練的核心爭議。該案中，原告主張 OpenAI 在訓練 ChatGPT 模型時將多首受保護的歌詞轉化為參數，使其被「記憶」於模型中，並可依使用者提示生成原始歌詞。法院認為，訓練階段將原告歌詞「記憶」於模型，或是後續輸出階段重現原始歌詞，均構成著作權法上的重製行為。

法院指出，資料探勘限制規定的立法原意，旨在規範建置訓練資料集所必要之重製行為（例如將作品轉製為數位格式或暫存於記憶體），由於該行為僅供後

<sup>30</sup> 即研究機構依 CDSM 指令第 3 條較寬鬆之規定建置資料集後，再轉交給商業 AI 開發商訓練模型之風險，美國著作權局對此也同樣表示憂慮，認為研究機構建置之資料集須與 AI 開發商利用行為脫鉤，並留意實務上彼此間之分工與關聯性。資料來源：U.S. Copyright Office Report, *Copyright and Artificial Intelligence Part 3: Generative AI Training*.

續系統自動化分析以「萃取資訊」，並未涉及著作之利用，故並未要求向權利人支付報酬。然而，儘管 AI 模型訓練原則上屬資料探勘範疇，本案被告於訓練模型過程中不僅「萃取資訊」，而是將作品完整記憶並重現著作已對著作權人權益造成實質侵害，明顯超出資料探勘限制規定之適用範圍<sup>31</sup>。此判決是歐洲首例法院認定 AI 模型訓練過程本身即可能構成侵害著作權的案例，對釐清該限制規定之適用界線具有指標性意義。

綜上所述，不論是從本質分析、司法實務可知，現行資料探勘限制規定在生成式 AI 之訓練的正當性與合法性，似有待重新檢視，學界評論<sup>32</sup>也多持相同看法。雖然 AI 法將該限制規定視為歐盟推動 AI 創新發展的重要工具，但在欠缺配套措施的情況下，不論在應對 AI 開發商的需求與潛在風險，或在促進授權機制與保障創作者報酬權等方面，皆顯示現行規範仍有不足。

## 肆、歐盟生成式 AI 制度檢討與未來展望

歐洲議會基於前述生成式 AI 研究報告，認為生成式 AI 實務運作與傳統資料探勘存在本質上的差異，因此在法規適用上面臨許多困境，並認為生成式 AI 訓練階段大規模重製著作、模型記憶、內化資料及商業應用方式，已超出 2019 年 CDSM 指令制定時所預設之資料探勘情境，因此 CDSM 指令的主要推動者 Axel Voss 議員於 2025 年 9 月進一步領銜提出法律修正案<sup>33</sup>，試圖檢討現行制度之不足。以下即以該研究報告與修正案改革建議為基礎，並結合相關案例與評論進行分析。

<sup>31</sup> GEMA v. OpenAI, 42-43, L.G. München I, Nov. 2025.

<sup>32</sup> Tim Dornis, The Training of Generative AI Is Not Text and Data Mining, *European Intellectual Property Review*, volume 47, 65-78 (2025).

Eduardo Alonso, Nicola Lucchi, AI And Copyright "Hallucinations": Does the Text and Data Mining Exception Really Support Generative AI Training?, *European Intellectual Property Review*, volume 47, 515-526 (2025).

Schack, Haimo Auslesen von Webseiten zu KI Trainingszwecken als Urheberrechtsverletzung de lege lata et ferenda 77 NJW - Neue Juristische Wochenschrift 113-118 (2024).

Matthew Sag and Peter K. Yu, The Globalization of Copyright Exceptions for AI Training, 74 *Emory Law Journal* (2025).

<sup>33</sup> 「著作權與生成式 AI：機會與挑戰」(2025/2058(INI)) 修正案概要 (AMENDMENTS 1 – 370)。

## 一、現行生成式 AI 制度檢討

### (一) 從「退出機制」轉向「事先授權」(Opt-in) 的可能

歐洲議會指出，僅依賴現行資料探勘限制規定之退出機制，已不足以保障創作者權益，甚至將其適用於生成式 AI 訓練批評為「嚴重法律錯誤適用」。因此報告主張應回歸著作權法一般原則，採取「事前授權(opt-in)」，以確保權利人對其作品使用具有實質控制權。

雖有論者憂心此舉將使 AI 開發商取得資料成本增加，進而阻礙技術創新，但從近年市場實務發展觀之，AI 開發商為降低侵權風險，以及對高品質資料的需求，促使其積極與權利人協商、簽訂授權協議。2025 年瑞典集體管理團體 STIM 主動推出 AI 音樂授權制度，允許新創公司 Songfox 使用者生成歌曲與翻唱；Getty Images 授權 Perplexity 於搜尋結果中使用其圖像<sup>34</sup>；環球音樂集團與音樂創作平台 Udio<sup>35</sup> 和 Stability AI<sup>36</sup> 公司分別達成和解與策略合作，共同開發音樂創作平台和工具；迪士尼公司授權 Sora 超過 200 個動畫與角色的資料庫生成影片<sup>37</sup> 等。從此些授權趨勢可看出權利人逐漸在訴訟以外，也傾向建立可持續的授權市場，甚至形成新的商業合作模式，並未因此阻礙 AI 開發商之技術發展。

<sup>34</sup> Getty Images 授權 Perplexity 在其 AI 搜尋服務中，於使用者查詢結果階段，即時顯示 Getty 的圖像內容，並附帶來源與權利標示，使用戶能夠存取具有適當署名的優質內容。Getty Images 另授權 Defined.ai 公司使用其圖像建置高品質資料集以訓練 AI 模型。

資料來源：<https://www.investing.com/news/stock-market-news/getty-perplexity-sign-multiyear-licensing-deal-to-boost-ai-powered-search-visuals-4323700?> (last visited Jan. 8, 2026).

<https://defined.ai/press-room/strategic-engagement-with-getty-images> (last visited Jan. 8, 2026).

<sup>35</sup> Universal says struck first licensing deal for AI music, AFP, <https://economictimes.indiatimes.com/tech/artificial-intelligence/universal-says-struck-first-licensing-deal-for-ai-music/articleshow/124952768.cms?from=mdr> (last visited Jan. 8, 2026).

<sup>36</sup> UMG strikes strategic alliance with Stability AI to develop 'next-generation' AI music-making tools, musicbusinessworldwide, <https://www.musicbusinessworldwide.com/umg-strikes-strategic-alliance-with-stability-ai-to-develop-next-generation-ai-music-making-tools> (last visited Jan. 8, 2026).

<sup>37</sup> The Walt Disney Company and OpenAI Reach Landmark Agreement to Bring Beloved Characters from Across Disney's Brands to Sora, The Walt Disney Company, <https://thewaltdisneycompany.com/news/disney-openai-sora-agreement/> (last visited Jan. 8, 2026).

## （二）增訂報酬請求權，輔以集體管理機制分配報酬

就公平報酬問題而言，相較於掌握較多資源的大型企業或內容平台，個別創作者往往缺乏與 AI 開發商談判的能力，在歐洲議會所提修正案中，即建議在資料探勘限制規定中增訂報酬請求權，或至少訂有談判與分配機制，以確保著作被用於 AI 訓練時，作者能獲得公平且適當的報酬<sup>38</sup>。

然而，鑒於 AI 訓練涉及大規模使用著作以及透明度有限等問題，報酬請求權仍須搭配可行的集體授權制度，始能協助個別創作者和小型業者獲得適當報酬，並有效率的授權大量作品予 AI 開發者<sup>39</sup>。例如英國在資料探勘立法遭到創意產業強烈反彈後，主要著作權授權機構（出版業為主）已規劃推動「生成式 AI 訓練集體授權制度」<sup>40</sup>；德國音樂著作權集體管理團體 GEMA 也推出 AI 授權模式，要求 AI 開發商就訓練用途及後續生成內容支付報酬<sup>41</sup>。此種收益分配方式不僅能降低 AI 開發商交易成本，也有助增加創作者收益來源。

## （三）標準化退出機制並強化透明度義務

現行退出機制因無統一標準與技術規範，導致權利人難以有效退出。而僅公開訓練資料摘要，無法讓創作者驗證自身作品是否被用於 AI 訓練，導致退出機制形同虛設。因此，至少應建立標準化且機器可讀的退出機制，或設置可供創作者登記的資料庫，使 AI 開發商能透過自動化爬蟲辨識並排除已標示退出的著作<sup>42</sup>，以增進透明度，同時降低權利人的負擔。

<sup>38</sup> 同註 33。

<sup>39</sup> Authors Guild, Authors Guild Policy Proposals Regarding the Development and Use of Generative AI, <https://authorsguild.org/app/uploads/2023/05/Authors-Guild-Policy-Proposals-Regarding-the-Development-and-Use-of-Generative-AI.pdf> (last visited Jan. 8, 2026).

<sup>40</sup> Lindsay Clark, AI training license will allow LLM builders to pay for content they consume, The Register, [https://www.theregister.com/2025/04/24/uk\\_publishing\\_body\\_launches\\_ai](https://www.theregister.com/2025/04/24/uk_publishing_body_launches_ai) (last visited Jan. 8, 2026).

<sup>41</sup> 收取 GenAI 服務提供者淨收入的 30%。資料來源：[https://www.digitalmusicnews.com/2024/10/25/gema-ai-licensing-model-details/?utm\\_source=chatgpt.com](https://www.digitalmusicnews.com/2024/10/25/gema-ai-licensing-model-details/?utm_source=chatgpt.com) (last visited Jan. 8, 2026).

<sup>42</sup> 同註 33。

然而，逐筆公開訓練資料，實務上因資料量龐大而不具可行性，為兼顧透明度與可操作性，修正案則建議 AI 開發商應保存爬蟲紀錄，並在權利人申請時提供檢視<sup>43</sup>，或是建立未履行透明義務的賠償規範<sup>44</sup>。唯有強化透明度義務，降低資訊不對等情形，權利人才能確實掌控作品的使用情形，落實退出機制功能，進而促成新的收益模式<sup>45</sup>與健全的授權市場。

## 二、歐盟未來展望與發展

綜合前述分析可知，歐盟現行 CDSM 指令資料探勘限制規定適用於生成式 AI 訓練仍有侷限，且面臨 AI 開發商與創意產業利益失衡的問題。從整體國際趨勢觀之，各國雖採不同法律路徑，但對於將生成式 AI 全面適用資料探勘限制規定，普遍持審慎態度。例如：澳洲<sup>46</sup>、韓國<sup>47</sup>及英國<sup>48</sup>因創作者與 AI 產業間仍存歧見，而暫緩推動相關立法；美國訴訟實務亦在涉及使用非法資料來源訓練 AI

<sup>43</sup> 同註 33。

<sup>44</sup> 同註 12。現行違反 AI 法透明度義務僅有行政罰鍰，權利人無從因而獲得直接救濟或利益。

<sup>45</sup> 美國圖庫公司 Shutterstock 與主要的 AI 開發者（如 Meta、Apple、Amazon、OpenAI）簽訂了授權協議，僅在 2023 年就帶來了約 1.04 億美元的年收入。資料來源：<https://petapixel.com/2024/06/04/shutterstock-made-104-million-licensing-assets-to-ai-devs-last-year/> (last visited Jan. 8, 2026).

<sup>46</sup> Jonathan Feder, John McLellan, Laura McFadzean, K&L Gates LLP - IP Law Watch. Copyright Act in Australia Won't Permit Free use of Copyright Works in AI. The National Law Review, <https://natlawreview.com/article/copyright-act-australia-wont-permit-free-use-copyright-works-ai> (last visited Dec. 30, 2025).

<sup>47</sup> 韓國文化體育觀光部 2025 年 6 月 13 日新聞稿。[https://www.mcst.go.kr/attachFiles/viewer/skin/doc.html?fn=20250613080135490963094493\\_PRESS20250613080135660336.hwp&rs=%2FattachFiles%2Fviewer%2Fresult%2Fpress%2F&](https://www.mcst.go.kr/attachFiles/viewer/skin/doc.html?fn=20250613080135490963094493_PRESS20250613080135660336.hwp&rs=%2FattachFiles%2Fviewer%2Fresult%2Fpress%2F&) (last visited Dec. 30, 2025).

<sup>48</sup> 英國政府 2024 年至 2025 年間針對 AI 訓練使用著作權作品之公眾諮詢結果，88% 的回應者（主要為創意產業）表示應以授權為前提，7% 回應者表示維持現有著作權法規定，僅有 3% 回應者支持引入允許權利人保留權利的資料探勘限制規定。資料來源：Copyright and artificial intelligence: progress report, <https://www.gov.uk/government/publications/copyright-and-artificial-intelligence-progress-report/copyright-and-artificial-intelligence-statement-of-progress-under-section-137-data-use-and-access-act> (last visited Dec. 30, 2025).

模型<sup>49</sup>、生成與原著作相似或競爭性內容<sup>50</sup>時，傾向認定可能損害原作者既有市場與潛在收益，因而難以成立合理使用，此一方向與德國 GEMA 訴 OpenAI 案相互呼應。至於日本雖訂有相對寬鬆之資料探勘規定，但近期 Perplexity 公司遭訴一案<sup>51</sup>，顯示即使以訓練 AI 模型為由，仍不得無限制利用著作。顯見如何在技術創新的同時衡平創作者利益，是各國皆面臨的關鍵課題。

為回應此挑戰，歐盟於 2025 年 11 月成立「著作權知識中心」<sup>52</sup>，為創作者提供管理與保護著作之指引，並於 2026 年 2 月啟動對 CDSM 指令的檢視<sup>53</sup>。此外，歐洲議會亦於 2026 年 3 月通過修正案<sup>54</sup>，其核心建議—包括強化訓練資料透明度、建立權利人報酬請求權，以及推動標準化退出機制等，將成為歐盟後續調整著作權與生成式 AI 政策，以及修訂 CDSM 指令的重要依據。

<sup>49</sup> Bartz v. Anthropic PBC, No. 3:24-cv-05417 (N.D. Cal. June 23, 2025). 法院認為下載盜版書籍於訓練資料庫將破壞出版市場的正常運作，不得與後續訓練模型之行為視為整體而免除法律責任。另認為掃描（重製）合法購買之書籍於訓練資料庫，以及後續訓練模型之行為皆具有高度轉化性而構成合理使用，但強調如果被告生成與原告著作高度相似的內容，則可能無法構成合理使用。

<sup>50</sup> Concord Music Group, Inc. v. Anthropic PBC, No. 3:24-cv-03811 (N.D. Cal.). 法院於駁回原告動議之初步意見表示，被告 Anthropic PBC 公司利用受保護歌詞訓練其生成式 AI 模型 Claude 且受有利益，其導致輸出原始歌詞給用戶之行為需負擔侵權賠償責任。

Kadrey v. Meta Platforms, Inc., No. 3:23-cv-03417 (N.D. Cal. June 25, 2025). 法院雖認為 Meta 公司利用他人著作訓練模型具有高度轉化性而構成合理使用，但強調係原告未能提出足夠證據，如能舉證 AI 生成之競爭性內容將造成市場稀釋，對原告市場造成損害，是原告致勝關鍵。

<sup>51</sup> 繼續閱讀新聞之後，朝日新聞和日經新聞也對美國 Perplexity 公司提起聯合訴訟，指控其未經授權使用文章，索賠總額達 44 億日元。ITmedia AI+, <https://www.itmedia.co.jp/aipplus/articles/2508/26/news100.html> (last visited Jan. 8, 2026).

<sup>52</sup> 透過一站式平台匯集著作權相關資訊、法規和判例，協助各界降低 AI 時代的侵權風險。歐盟著作權知識中心網頁，<https://www.euipo.europa.eu/en/copyright-knowledge-centre> (last visited Feb. 10, 2026).

<sup>53</sup> EUobserver, MEPs call for new copyright rules when AI trains on protected works, <https://euobserver.com/206337/meps-call-for-new-copyright-rules-when-ai-trains-on-protected-works/> (last visited Mar. 12, 2026).

<sup>54</sup> 歐洲議會 2026 年 3 月 10 日以 460 票贊成、71 票反對的表決結果，通過「著作權與生成式 AI」修正案之建議，後續將作為歐盟執委會修法之參考。<https://www.europarl.europa.eu/news/sv/press-room/20260306IPR37511/protecting-copyrighted-work-and-the-eu-s-creative-sector-in-the-age-of-ai> (last visited Mar. 12, 2026).

### 伍、結語

綜合前述歐盟法制檢討與各國因應方向，可作為我國未來 AI 政策之重要參考，首先，不宜將資料探勘限制規定視為 AI 發展的唯一解方，仍應回歸「事前授權」原則，以避免創作者權益於科技發展中被弱化。其次，鑒於個別創作者欠缺與 AI 開發商談判的籌碼，可協助創意產業推動集體授權機制，促進雙方協商可持續的授權模式，從而確保創作者能分享 AI 發展之經濟利益。最後，輔以軟性治理工具，如訂定 AI 著作權指南或加強對使用者宣導，以降低生成侵權內容之風險等。透過法制、產業與治理多面向的配合，方能在技術創新與創作者權益間取得合理的平衡。