大型語言模型在訓練階段與使用檢索增強生成的著作權風險挑戰及案例研究

廖先志 *、陳鍾誠 **

* 作者現為法務部國際及兩岸法律司調部辦事檢察官。
** 作者現為國立金門大學資訊工程系助理教授。
本文相關論述僅為一般研究探討，不代表本局及任職單位之意見。

摘要

本文探討大型語言模型在訓練階段與使用檢索增強生成技術時所面臨的著作權風險，本文首先分析預訓練、微調和檢索增強生成技術的原理，並根據我國著作權法規範之合理使用原則，評估三者於利用目的及性質、著作性質、利用質量及比例，以及利用結果對著作潛在市場影響等面向的侵權風險。本文認為，預訓練的風險最低，微調風險其次，而檢索增強生成的風險最高。接著，本文以 ISO 31000 風險管理框架為基礎，設計一個以跨國移交受刑人法為主題的智慧問答系統。為降低著作權侵權風險，系統僅使用法律條文、施行細則、法院判決和取得合法授權的文獻作為資料來源。系統採用微調後的 ChatGPT 3.5 Turbo 模型，並結合檢索增強生成技術，以 Google Colab 平臺運行。本文最後總結大型語言模型之訓練及其在法律領域應用所面臨的著作權挑戰，並提出未來研究方向。

關鍵字：人工智慧、大型語言模型、訓練模型、檢索增強生成、著作權
AI、Large Language Model（LLM）、Model Training、Retrieval-Augmented Generation（RAG）、Copyright

壹、前言

大型語言模型（Large Language Model, LLM）作為生成式AI（Generative AI）的核心技術[1]，廣泛應用於文本生成、推理及專業領域的知識探索等領域，實用價值極高。然而，LLM在設計與運行過程中，高度依賴各個領域的訓練資料與動態檢索資料：訓練又可細分為預訓練（Pre-training）與微調（Fine-tuning）等階段（技術細節均詳見下參、一），訓練LLM的資料來源不僅包括學術論文、專業期刊、新聞報導、法律條文等各式文本[2]，還涵蓋圖像、音頻、視頻等多模態數據，為模型提供穩定且專業的基礎知識；而動態資料則透過使用檢索增強生成（Retrieval-Augmented Generation, RAG），能及時更新特定領域的知識並回應最新問題[3]。RAG並可在相當程度上減少LLM產生「幻覺」（hallucination）現象，避免LLM生成錯誤或不完整的資訊，導致使用者被誤導[4]。在法律應用場景中，這樣的挑戰尤為重要，因為法律文本的生成與法律問題的解答需要高度準確性與權威性，這要求LLM能精準學習或正確擷取相關的法律知識;但隨之而來的，無論是學習靜態資料還是擷取動態資料，其操作過程是否符合著作權規範，都是設計AI應用系統時必須審慎應對的核心問題。

基於上述討論，以下本文首先探討我國與美國著作權法中關於合理使用的規範，並分析LLM在預訓練、微調與RAG技術應用場景中的潛在侵權風險。在此基礎上，本文引入ISO 31000風險管理框架，並應用這個架構來設計一個跨國移交受刑人法AI智慧問答系統，這個系統是使用OpenAI的ChatGPT 3.5 Turbo模型。本文通過這個系統的設計與應用，深入探討如何通過技術手段與法律策略的協作，平衡AI技術的發展與著作權保護之間的矛盾。最後，本文提出未來在設計類似系統時使用RAG技術的可能改進方向，期望本研究能為LLM在專業領域的應用提供具體參考。

---

1 Ashish Vaswani et al., Attention Is All You Need, 30 ADVANCES IN NEURAL INFO. PROCESSING SYS. 5998-6008 (2017).

2 Kevin D. Ashley, Artificial Intelligence and Legal Analytics: New Tools for Law Practice in the Digital Age, Cambridge University Press (2017).

3 Patrick Lewis et al., Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks, 33 ADVANCES IN NEURAL INFORMATION PROCESSING SYSTEMS, 9459-9474 (2020).

4 Varun Magesh et al., Hallucination-Free? Assessing the Reliability of Leading AI Legal Research Tools, ARXIV, abs/2405.20362, https://arxiv.org/abs/2405.20362 (last visited Dec. 12, 2024).

貳、LLM 訓練階段與 RAG 有關之我國及美國著作權法相關規定及合理使用

一、我國著作權法

在 LLM 的訓練與 RAG 應用過程中，資料來源的合法性是評估 LLM 法律風險的核心要素之一。這些資料可能是不受著作權法保護的著作，例如法律條文、政府公報、法院判決、已超過著作權保護期間的著作，以及不具原創性的內容[5]（如事實描述）或僅屬構想（idea）而非具體表達（expression）的部分[6]；同時，資料也可能包含受著作權保護的著作，例如新聞報導、學術論文、判決評釋及判決等彙編。在訓練 LLM 及應用 RAG 的過程中，其牽涉到的行為通常有對資料的重製行為（reproduction）[7]，即將文本內容解析並內化於模型參數中；此外，也可能因為對文本進行摘要、篩選或重組，而涉及對資料的改作（adaptation）[8]。同樣，如果將資料在多伺服器間傳輸或分享，則可能涉及公開傳輸（public transmission）行為。對於受著作權保護的文本，這些操作都可能構成侵權，除非符合合理使用（fair use）或法律例外條款[9]。

而根據我國著作權法第 44 條至第 63 條的合理使用規定，對於著作的利用是否構成合理使用，需綜合審酌「一切情狀」，並重點考量以下基準：（1）利用的目的及性質，包括是否屬商業或非營利教育用途；（2）著作的性質；（3）利用部分在整個著作中所占的比例與質量；（4）利用對著作潛在市場及現在價值的影響[10]。此外，我國著作權法第 65 條第 2 項進一步將例示之 4 項判斷基準以外事實之「一切情狀」，例如利用人之善意與否、公共利益、行為妥適性及社會福利等因素納入考量，認為合理使用的判斷不應僅局限於單一基準，而需以人類智識文化資產之公共利益為核心，進行綜合判斷[11]。

---

5 蔡惠如，著作權合理使用概括規定之回顧與前瞻，智慧財產權月刊第 209 期，頁 4-25，2016 年 5 月。

6 我國著作權法第 10 條之 1 規定：「依本法取得之著作權，其保護僅及於該著作之表達，而不及於其所表達之思想、程式、制程、系統、操作方法、概念、原理、發現。」

7 Daniel J. Gervais et al., The Heart of the Matter: Copyright, AI Training, and LLMs, SSRN, Sept. 21, 2024, https://ssrn.com/abstract=4963711 (last visited Dec. 12, 2024).

8 章忠信，人工智慧訓練與著作之合法利用，智慧財產權月刊第 304 期，頁 5-26，2024 年 4 月。

9 我國著作權法第 65 條第 1 項。

10 我國著作權法第 65 條第 2 項。

11 智慧財產法院 102 年度民著上字第 1 號民事判決要旨參照。

二、美國著作權法中之轉化與合理使用

美國著作權法是我國著作權法合理使用規定的重要參考來源，其中「轉化性使用」（transformative use）的概念在合理使用的判斷中占據關鍵地位，也對我國司法實務產生了深遠影響[12]。轉化性使用是指在利用他人著作時，是否對原著作進行了充分的實質改變，使其目的或性質與原作不同，並創造具有新價值的作品。這種改變可能表現在資訊重組、美學創新或新視角應用等方面，其核心在於創作出與原著作具有不同功能或表達的新作品，而非單純的重製行為。根據美國法院的實務見解，當使用行為具高度轉化性時，即便具有商業目的，也更傾向於被認定為合理使用。這是因為轉化性使用通常降低了對原作市場的替代效應，同時有助於知識創新與文化多樣性，實現著作權保護的平衡目標[13]。雖然我國著作權法第 65 條關於合理使用的基準中未明確提到轉化性，但我國司法實務在解釋第 65 條的各個基準時，特別是在「利用的目的及性質」與「利用結果對市場潛在影響」兩項基準的分析中，已隱含參考了美國轉化性使用的概念，例如，當使用行為有助於社會知識或文化多樣性，且對原作市場替代效應不高時，通常更容易被認定為合理使用[14]。

參、訓練階段與 RAG 的著作權風險分析

一、訓練 LLM 與 RAG 之技術簡述

（一）訓練階段—預訓練與微調

預訓練和微調是 LLM 訓練過程中的兩個核心階段，二者在目標、資料利用和對模型結構的影響上存在顯著差異。預訓練的主要目標是學習大規模通用語言模式，通過多樣化的資料來源捕捉語法結構、語義理解

12 蔡嘉裕，著作權「轉化性使用」之我國本土案例分析，智慧財產權月刊第 271 期，頁 47-77，2021 年 7 月。

13 Jiarui Liu, An Empirical Study of Transformative Use in Copyright Law, 22 Stanford Technology Law Review 163-241 (2019).

14 章忠信，同註 8。

及上下文關係，從而構建一個具備廣泛泛化能力的基礎模型。在此階段，模型的各層參數均接受訓練，實現分層次的功能：例如，低層參數專注於詞法和句法結構的學習，中層參數側重於語義處理與上下文依賴，高層參數則負責捕捉長距離依賴關係並增強生成能力[15]。相比之下，微調專注於針對特定任務或領域資料進行優化，主要調整高層參數以適應特定場景需求。微調階段使用的小型、高度專業化資料集，通常具備高商業價值與高原創性，是該領域任務特徵的集中代表。通過精細的參數調整，微調能顯著提升模型在特定應用中的表現[16]，尤其是在專業領域（如法律）中的應用，微調能針對高複雜性和專有性文本進行優化，使 LLM 更能準確地處理專業領域中的細節與邏輯挑戰[17]。

（二）RAG

RAG 是結合 LLM 與外部資料檢索的技術，用於生成準確性更高的回應。RAG 的運作原理是通過檢索模組在產生回應前，先從知識庫中提取相關文檔送給 LLM，LLM 在檢索結果的基礎上，生成答案以回應使用者的問題[18]。這種讓 LLM 可以先檢索閱讀後再回答的「開書考」（Open Book）方式，避免了模型僅依賴其內部訓練過程中的統計學習來回答問題，讓生成內容可以被驗證，從而提升了準確性和用戶的信任感[19]。尤其在法律領域，RAG 可以引用例如法條或判例等權威性資料，使回應更具專業性和可靠性[20]。然而，應注意的是，RAG 仍無法完全消除 AI 的幻覺，研究顯示，某些 AI 法律系統即使採用了 RAG 技術，仍存在幻覺，例如

15 Alec Radford & Karthik Narasimhan, Improving Language Understanding by Generative Pre-Training, 2018, https://cdn.openai.com/research-covers/language-unsupervised/language_understanding_paper.pdf (last visited Dec. 12, 2024).

16 Jeremy Howard and Sebastian Ruder, Universal Language Model Fine-tuning for Text Classification, PROC. OF THE 56TH ANN. MEETING OF THE ASS'N FOR COMPUTATIONAL LINGUISTICS, 328-339 (2018).

17 Mike Lewis, Yinhan Liu, Naman Goyal et al., BART: Denoising Sequence-to-Sequence Pre-training for Natural Language Generation, Translation, and Comprehension, PROC. OF THE 58TH ANN. MEETING OF THE ASS'N FOR COMPUTATIONAL LINGUISTICS, 2020.

18 Patrick Lewis et al., *supra* note 3.

19 Varun Magesh et al., *supra* note 4.

20 *Id*.

錯誤地陳述了判例結果或對法條的錯誤引用[21]。但無論如何，採用 RAG 仍是減少幻覺的一個相當有效的技術方式[22]。

二、以我國著作權法合理使用規定分析預訓練、微調、RAG 三者之侵權風險

若在未經授權的情況下，使用如新聞報導或學術論文等受著作權保護的作品進行 LLM 的訓練，無論是預訓練還是微調，皆可能涉及重製、改作甚至公開傳輸行為（我國著作權法第 3 條第 2 款），而有侵害著作權的風險[23]，國內外雖有不少文獻討論到訓練 LLM 時的著作權侵權問題，但多數文獻均將訓練階段視為一體，並未詳細區分訓練階段討論其不同的著作權侵權風險[24]，但實際上，從前述討論可知，LLM 訓練階段中之預訓練與微調其技術本質並不相同，且 RAG 屬外部動態資料來源，並未變更 LLM 本身，因此三者之侵害著作權風險本不相同，本文以下將根據上述我國著作權法規範之合理使用，考量如利用的目的及性質、著作的性質、所利用的質量及比例、利用結果對著作潛在市場影響等四項判斷基準，對 LLM 之預訓練、微調與 RAG 下的著作權風險進行分析與比較：

（一）利用的目的及性質

預訓練的目的是構建通用語言模型，多數屬於非營利性基礎研究，公益性較高，主張合理使用的可能性較高；微調階段多用於特定任務的優化，其主張合理使用的可能性相對較弱；而 RAG 因直接針對特定用戶的需求，用途通常非常明確，如果用途不具公益性，RAG 主張合理使用的可能性最低。

---

21 Matthew Dahl, Varun Magesh, Mirac Suzgun and Daniel E. Ho, Large Legal Fictions: Profiling Legal Hallucinations in Large Language Models, ARXIV, abs/2401.01301 (2024), https://arxiv.org/abs/2401.01301 (last visited Dec. 12, 2024).

22 Varun Magesh et al, *supra* note 4.

23 章忠信，同註 8。

24 如章忠信，同註 8 及 Daniel J. Gervais et al., *supra* note 7.

（二）著作的性質

此款所稱「著作的性質」是指「被利用著作之性質」[25]，預訓練使用的資料來源多樣，可能包含有具著作權保護或公共領域等不同種類的來源，侵權風險不一；而微調階段因為常針對某一特定知識領域，所以除了使用公共領域的資料外，微調通常還會依賴如學術論文與法律評論等高原創性與高商業價值的作品，在使用這類作品來微調時，侵害著作權的風險就會顯著增加；至於 RAG，因為檢索資料的性質同樣可能涉及高原創性資料，且生成內容直接使用檢索資料，故本文以為，RAG 的侵害著作權風險至少與微調相當，甚至更高。

（三）所利用的質量及比例

預訓練通常涉及大規模資料集，從網路上取得大量的文本之後，讓 LLM 學習如何透過接龍的方式進行寫作，這些用來訓練的文章被用作 LLM 的學習語料，讓 LLM 能透過大量語料的訓練學會寫作。雖然預訓練是將這些文章完全加以參數化，但由於資料規模龐大，每篇文章，甚至是每一本書，在整個龐大的訓練語料當中，占的比例通常不到萬分之一，單一作品在整體資料中的比例極低。在這個階段，模型學習的內容主要是綜合資料的抽象模式，而非單一作品的具體表達方式，因此，單一作品對 LLM 整體的影響通常有限，且 LLM 經預訓練後所生成內容通常較難直接與原作建立明顯關聯[26]。

然而，應注意的是，在海量的預訓練資料集中對單一作品的完整使用，與我國著作權法第 65 條第 3 款中傳統「質量與比例」基準並不完全相同，因為該基準通常指的是單一作品內部使用部分的比例，而非單一作品在資料集中的比例，儘管如此，考量到單一作品在訓練資料集中所占比例甚低、預訓練生成內容的高度抽象性及其對單一作品核心內容的依賴程度較低等因素，本文以為，在實質上，預訓練對單一作品的使用與傳統「質

[25] 蕭雄淋，著作權法論，頁 214，五南圖書出版有限公司，2021 年第 9 版。
[26] Jeremy Howard and Sebastian Ruder, *supra* note 16.

量與比例」原則具有相近之合理性基礎，因此，預訓練階段或可類推適用我國著作權法第 65 條第 3 款規定，而得以主張合理使用[27]；惟此主張仍有待未來我國或其他國家司法實務能進一步釐清完整重製行為與部分引用間之界線。然而即使如此，應該要注意的是，即使單一作品占比低，預訓練過程中仍可能因為刻意強化特定作品的風格或特徵，而導致模型產生過度擬合（overfitting）現象，進而增加侵權風險[28]。

而微調過程通常使用小型、高度專業化的資料集，單一作品的比例顯著提高，且生成內容可能直接反應出原作特定表達方式，導致侵權風險升高。例如，若微調資料集中包含某作者全集，微調後之 LLM 生成內容中高度再現原作語句或段落，都可能被認定為無法主張合理使用。同時，與預訓練相較，微調階段更容易因為資料來源過於集中或相似度太高，而出現過度擬合特定作者風格的情況，這也進一步增加侵權風險。至於 RAG，因其技術的特性使其生成內容直接依賴檢索到的資料，尤其是在小型專業資料庫中，單一作品的比例更高，且相比於預訓練或微調，RAG 在生成答案時可能直接包含具體引用的語句或段落，致使侵權風險增加。

（四）利用結果對著作潛在市場影響

預訓練生成的內容通常是通用語言模式，與特定原作的市場功能無直接競爭。因此，預訓練的生成內容通常對市場的替代效應較低，對市場價值的影響也相對有限。相比之下，微調針對特定領域的專業資料集進行優化，生成內容往往包含高專業性或高原創性的表達，例如法律意見、判決摘要或專業評論等。這些生成內容可能與原作在市場功能上形成直接競爭，導致用戶對原作需求的減少，進而對其市場價值造成實質影響。RAG 技術的市場影響更為顯著，特別是當生成內容直接取自外部資料庫或文獻，並與其市場功能重疊，可能會削弱這些服務的市場需求，進一步威脅其經濟價值，其侵權風險也因此顯著提高。

---

27 類推適用係針對法律漏洞，可參見：王澤鑑，法律思維與民法實例，頁 305-306，2003 年。

28 Daniel J. Gervais et al., *supra* note 7.

綜合比較，三者在侵權風險上呈現遞增趨勢：預訓練風險最低，因為其使用大規模、多樣化資料集，單一作品比例低，生成內容抽象，市場影響有限，主張合理使用的可能性較高；微調侵權風險較高，因資料集集中於特定領域，涉及高原創性作品，生成內容與原作相似度高，市場影響顯著，故主張合理使用的可能性降低；而 RAG 因為是從直接透過檢索從資料庫或文獻中提取資料生成答案，生成之內容與檢索資料高度相關，市場功能重疊，替代效應明顯，主張合理使用可能性最低，所以侵害著作權的可能性最高。

肆、以 ISO 31000 風險管理理念下設計之跨國移交受刑人法智慧問答系統

一、跨國移交受刑人法簡介

跨國移交受刑人法制定於 102 年，該法對與其他國家（包括中國大陸及港澳地區）的受刑人移交程序進行了詳細規範，所謂移交受刑人，是基於人道考量，將受刑人送回其國籍國去服刑，移交受刑人屬於廣義的刑事司法互助之一[29]。在我國，刑事司法互助研究者本來就不多，跨國移交受刑人的專門研究者更少。儘管如此，我國與其他國家間近年仍有移交受刑人的個案持續發生[30]，顯示該法在實務中具有一定的應用價值。但如果每碰到一個個案，都要從頭開始研究相關法律及程序，效率甚低，且難免發生錯誤，所以一個能夠回答與這部法律及程序等相關問題的人工智慧問答系統，有其價值。本智慧問答系統旨在為一般使用者與專業法律人士提供一個專注於跨國移交受刑人法的智慧問答系統（下稱移交受刑人問答系統），其目標是正確回答使用者對於法條適用、移交程序及相關司法實務見解等常見問題。

[29] 廖先志，跨國移交受刑人法簡介，法學叢刊第 58 卷第 1 期，頁 131-162，2013 年 1 月。
[30] 法務部，跨國移交受刑人制度簡介，https://reurl.cc/xpleR5（最後瀏覽日：2024/12/12）。

二、以 ISO 31000 標準對移交受刑人問答系統之風險管理與分析

ISO 31000 是國際通用的風險管理標準，旨在為組織提供應對不確定性的架構，支持決策制定並確保既定目標的實現。該架構包含風險識別、風險分析、風險評估、風險處理和風險監控五個核心步驟，通過系統化的流程協助組織有效管理各類風險[31]。在法律 AI 系統設計中，特別是處理高度專業化的法律文本時，ISO 31000 提供了一套能有效應用於技術與法律交叉領域的分析工具，以下即應用 ISO 31000 的架構來討論應如何設計移交受刑人問答系統。

（一）風險識別

移交受刑人問答系統旨在解答使用者關於跨國移交受刑人法的法律問題，故可能需整合法律條文、施行細則、法院判決、學術文章、新聞報導及其他資料作為知識來源。然而，不同資料來源的法律屬性和使用方式各異，導致潛在的著作權侵權風險。

（二）風險分析

在資料來源層面，本系統需使用的文本包含多種性質。法律條文、施行細則與法院判決屬於公共領域或公文書[32]，法律上不具著作權保護風險，而學術文章、新聞報導及部分網頁內容則可能具高度原創性，若未經授權即使用，可能構成著作權侵權糾紛。在技術面，微調涉及將特定資料內化進模型參數，RAG 則在生成階段直接引用外部資料，兩者皆有可能再現原始著作內容，產生潛在侵權問題。

（三）風險評估

從前述風險來源分析可知，法律條文、施行細則與法院判決因其公共領域屬性，風險極低，是安全且穩定的知識來源，且能顯著提升系統

[31] International Organization for Standardization (ISO), ISO 31000:2018 Risk Management – Guidelines, https://www.iso.org/standard/65694.html (last visited Dec. 12, 2024).

[32] 我國著作權法第 9 條。

的專業性和信任感。相比之下，學術文章與新聞報導如未獲授權，且直接被用於生成內容時，可能構成替代市場，風險程度較高。

從技術面來看，侵權風險則呈明顯排序：RAG 因回應內容直接依賴檢索結果，且常包含原始語句，風險最高；微調雖為內化處理，但資料集中度高也仍有再現風險，風險次之；預訓練階段主張合理使用的可能性最高，故風險最低。

（四）風險處理

1、技術選擇

基於風險與效能的平衡考量，本文以為，應優先選用侵權風險較低且能滿足專業需求的技術方案，避免採用高成本且缺乏針對性的訓練方式。在法律領域實務中，也常見以既有語言模型進行微調，而非自建預訓練模型[33]。另可搭配如 RAG 等動態更新技術，以提升系統回應的準確性與權威性。

2、資料選擇

鑒於本系統採取的微調與 RAG 技術對資料來源的侵權風險相對較高，應採取更為保守的資料策略，包括僅採用公共領域或合法授權資料，排除高風險來源，並確保引用透明化，以提升系統回應的可信度與可追溯性。

（五）風險監控與持續改進

未來，系統將動態監控新增資料的合法性，並結合使用者的回饋以優化模型性能。同時，基於技術選擇的局限性，繼續探索如何在保持答案權威性的基礎上，進一步降低潛在侵權風險。

[33] Paul D. Callister, Generative AI Large Language Models and Researching the Law, SSRN, Aug. 12, 2024, https://ssrn.com/abstract=4927675 (last visited Dec. 12, 2024).

三、基於風險管理架構下的移交受刑人問答系統實作

基於上述的 ISO 31000 風險架構及系統效能的考量，以下，分為技術實作與資料來源兩部分來說明移交受刑人問答系統的架構：

（一）技術實作

移交受刑人問答系統選擇使用 ChatGPT 3.5 Turbo，這是一款專為生成式 AI 應用設計的高效語言模型，能有效處理文本生成與語言理解任務[34]。在訓練方法上，本文結合微調與 RAG 技術，以兼顧系統的專業性與動態性。系統運行環境為 Google Colab，其內建 GPU（圖形處理單元）加速可顯著提升嵌入生成與資料檢索的運算效率，確保能快速處理大規模資料[35]，確保系統能在短時間內完成大規模資料的處理。此外，系統介面採用 Gradio[36]，設計簡潔、使用直觀，用戶可直接輸入法律問題並獲得回應。

1、微調 LLM 的技術實踐

微調的目的是內化與移交受刑人相關的專業知識，使模型能準確且一致地回答涉及該法的問題[37]。本系統採用 OpenAI 提供的 Playground 工具[38]進行微調，該工具支援用戶通過 JSONL（JSON Lines）格式資料進行模型微調，提供了一個簡便的接口，適合用於對現有模型（如 ChatGPT 3.5 Turbo、ChatGPT 4o-mini）進行特定領域的優化，本文即以 ChatGPT 3.5 Turbo 為基礎完成模型微調，並通過 OpenAI 所提供的 API 接口使用微調後之模型來完成即時查詢[39]。

---

34 OpenAI, GPT-3.5 Turbo, https://platform.openai.com/docs/models/gpt#gpt-3-5-turbo (last visited Dec. 12, 2024).

35 Google Colab, Colaboratory - A Google Research Project, https://colab.research.google.com/ (last visited Dec. 12, 2024).

36 Gradio, Gradio Documentation, https://gradio.app/docs/ (last visited Dec. 12, 2024).

37 Jeremy Howard and Sebastian Ruder, *supra* note 16.

38 OpenAI, OpenAI Platform, https://platform.openai.com/ (last visited Dec. 12, 2024).

39 API（Application Programming Interface）是一種軟體接口，允許開發者通過程式設計與工具或服務進行交互。使用 OpenAI 的 API，用戶可以將 ChatGPT 等模型整合到自己的應用中，用於回答問題、生成文本或執行其他自然語言處理任務，OpenAI, OpenAI API Documentation, https://platform.openai.com/docs (last visited Dec. 12, 2024).

2、RAG 的技術實踐

微調雖然有其功效，但微調常見的缺點在於如果知識更新速度過快，要即時更新，就必須重新微調，導致成本較高；且微調未必能完全避免模型產生不精確的回答[40]。為解決這一問題，本系統除微調外，再以微調後的模型為基礎，同時使用 RAG 技術，結合動態檢索模組與生成模組，實現即時解答最新法律問題的能力。為達成此目標，本系統首先通過一種名為 FAISS（Facebook AI Similarity Search）的工具[41]，將微調階段使用的資料整理成適合快速檢索的格式。具體而言，這些資料被分割成約 1,000 字的小段，並設置 120 字的重疊區間，確保在分段後仍能保留文本的上下文關係。隨後，本系統使用 OpenAI 的嵌入模型，將這些文字轉化為一種數學向量形式存入資料庫中[42]。當使用者輸入法律問題時，系統會根據問題的語意，自動從資料庫中挑選最相關的小段文字，並將檢索到的內容作為生成答案的基礎。此外，本系統不但在生成過程中會參考檢索到的資料進行回應；將答案提供給使用者時，還會附加相關資料的引用來源，以增強答案的專業性與可信度。

（二）資料來源

基於風險及效能之平衡考量，移交受刑人問答系統的資料主要來源包括三類：首先是跨國移交受刑人法的條文與施行細則，這些法條為系統提供了明確且權威的法律依據；其次是整理後的法院判決摘要，這些摘要選取了相關案例的核心要旨，有助於模型學習具體案例的法律適用邏輯；最後是使用作者撰寫與整理的學術文章，這些文章對相關的法律與實務有深度的探討。

---

40 Zorik Gekhman et al., Does Fine-Tuning LLMs on New Knowledge Encourage Hallucinations?, ARXIV, abs/2405.05904 (2024), https://arxiv.org/abs/2405.05904 (last visited Dec. 12, 2024).

41 J. Johnson, M. Douze & H. Jégou, Billion-Scale Similarity Search with GPUs, IEEE TRANSACTIONS ON BIG DATA (2017), https://arxiv.org/abs/1702.08734 (last visited Dec. 12, 2024).

42 OpenAI, Embeddings Documentation, https://platform.openai.com/docs/guides/embeddings (last visited Dec. 12, 2024).

所有資料經過嚴格整理後以 JSONL 格式儲存，JSONL 是一種輕量級的資料格式，用於存儲結構化資料。每條紀錄作為獨立的 JSON 對象，存放在文本文件的單獨一行中，便於處理大量資料或進行批量操作。相比傳統的 JSON 格式，JSONL 格式在可讀性與效能上表現更佳，尤其適合用於模型的訓練。檔案中，每筆紀錄包含用戶提問（prompt）與模型回答（completion）所組成的問答對（Question-Answer Pair），以確保符合 OpenAI 微調格式的要求。

伍、結論與未來研究方向

在閉源（closed-source）的 LLM 中，由於模型建立者不公開訓練資料的來源，著作權擁有者可能無法確定其作品是否被使用或侵權；即使是 Llama 等開源（open-source）模型，其訓練方式與資料集來源等細節也未必完全公開[43]。2024 年剛通過的歐盟 AI 法案（AI Act）雖然要求 AI 模型提供者公開訓練資料來源，但提供者仍可能僅提供資料集或來源的概括性描述，並不一定會提供含具體作品來源清單[44]。在這種狀況下，多數時候著作權利人只能透過特定提示或輸入，觀察模型是否生成與其作品相似的內容來「推測」LLM 是否使用了相關資料[45]。而對於主要目的在於僅需要 LLM「學會」專門領域知識的一般系統設計者來說，受限於經費、技術等因素，通常只能微調 LLM，且常使用到 RAG 技術，因此，理解與分析 LLM 訓練及應用等各階段的著作權風險顯得尤為重要。本文從法律與技術雙重角度，深入探討了 LLM 在處理法律文本時的著作權挑戰，並在預訓練、微調與 RAG 三個階段的侵權風險上進行了詳細比較：預訓練因使用大規模資料且生成內容抽象，侵權風險相對最低；微調因依賴小型專業化資料集，侵權風險因而提高；RAG 則因直接檢索外部資料進行生成，侵權風險最高。這些發現補充

43 Rishi Bommasani et al., Foundation Model Transparency Reports, 7 PROC. AAAI/ACM CONF. AI ETHICS & SOC'Y 181, 181-195 (2024).

44 Tech Policy Press, How the EU AI Act Can Increase Transparency Around AI Training Data, https://www.techpolicy.press/how-the-eu-ai-act-can-increase-transparency-around-ai-training-data/ (last visited Dec. 12, 2024).

45 葉奇鑫、許斌，AI 大語言模型訓練與著作權合理使用之思考－以紐約時報對 OpenAI 訴訟案為中心，全國律師第 28 卷第 6 期，頁 5-19，2024 年 6 月。

了現有文獻對這些應用情境的細化研究[46]，並結合我國著作權法與美國合理使用中的轉化性概念，探討了技術應用與合法性之間的平衡。

此外，本文作者還設計了以跨國移交受刑人法為核心的智慧問答系統，通過微調和 RAG 技術結合，實現了法律應用的專業性與動態性。這個系統初步測試結果顯示，能在數秒內提供高準確度的回答，展示了生成式 AI 在法律領域的應用潛力。未來可以再著重於評估系統效能，如生成準確性、引用透明性及用戶滿意度；同時探索資料庫分層策略，嘗試將受保護與無保護資料分離，並設計相應的生成策略，以設計更合規的生成策略，進一步平衡 AI 的實用性與合法性。

[46] Agency for Cultural Affairs, Government of Japan, General Understanding on AI and Copyright in Japan, https://www.bunka.go.jp/english/policy/copyright/pdf/94055801_01.pdf (last visited Dec. 12, 2024).