國立陽明交通大學

科技法律研究所

人工智慧成為犯罪主體或犯罪工具
—以刑事訴訟法為中心

研究生：許書恬

中華民國115年9月7日

目錄

第一章　緒論 **3**

第一節　研究動機 3

第二節　研究目的與問題設定 4

第三節　文獻回顧與探討 4

第一項　比較法的監管模型 4

第二項　人工智慧基本法草案參考歐盟法有監管意識 5

第四節　研究方法與研究步驟 6

第一項　法釋義學 6

第二項　比較法研究 6

第三項　法律過程分析與實務分析 6

第四項　研究步驟一覽表 7

第五節　預期研究貢獻 7

第一項　從技術面角度釐清生成式 AI 在刑法體系中的定位 7

第二項　建構 AI 開發者、提供者與使用者的責任分配模式 7

第三項　提出具體的立法建議 8

第二章　生成式 AI 作為刑事程序中之運用 **8**

第一節　司法院判決輔助系統 8

第二節　刑法基礎結構變動 9

第一項　三大核心面向 9

第二項　以自動駕駛為例 11

第三章　AI 作為犯罪主體或是犯罪工具論 **12**

第一節　犯罪主體論之爭 12

第二節　本文見解 14

第四章　結論 **14**

第一節　人工智慧基本法草案修訂 14

第二節　普通刑法修訂 15

第三節　引入消保法之產品責任 15

第五章　參考文獻 **17**

第一章　緒論

第一節　研究動機

臺灣新北地方法院111年度智訴字第8號違反著作權法案件，被視為台灣首宗以網路爬蟲程式大量重製競爭者資料庫而遭有罪判決的著作權刑事案件。經法院判決，七法公司（Lawsnote）的兩位共同創辦人分別被判處4年和2年徒刑，同時公司與創辦人必須連帶賠償法源資訊新台幣1億元，此一結果除在實務與產業界引發高度關注外，亦顯示本案雖涉及高度自動化的程式運作，法院仍將刑事責任歸屬於自然人與法人，而非運作該程式之AI本身[1]，凸顯現行台灣刑法體系並未承認AI具有可罰性之地位。

此一發展促使本研究反思，近年來人工智慧（Artificial Intelligence, AI）發展，尤其是生成式人工智慧（Generative AI）迅速滲透至法律、科技、教育、金融等多元領域，舉例來說我國司法院為回應各界對於量刑妥適性的期待，並因應國民法官新制上路，已全面啟用AI量刑資訊系統，希望透過資訊力量的協助，整理大量裁判資料，並即時更新，提供職業法官與國民法官查詢類似案件的量刑趨勢，讓量刑更公正透明[2]。輔助系統其高度自動化與自主生成能力，在提升生產效率的同時，生成式AI究竟應被定位為單純工具，抑或已在某些情境下具備可歸責性而成為犯罪主體？若僅以「工具定位」處理所有AI介入犯罪之情形，將可能無法充分回應AI自主性提升後的風險結

---

[1] 臺灣新北地方法院111年度智訴字第8號刑事判決。事實欄明確指出：「郭榮彥、謝復雅…竟共同基於意圖銷售而擅自以重製之方法侵害他人著作權、無故取得他人電磁紀錄之犯意聯絡，由謝復雅…撰寫6支爬蟲程式；從判決主文顯示，法院將刑事責任明確歸屬於自然人（郭榮彥、謝復雅）以及法人（七法股份有限公司），並未將爬蟲程式視為可能之犯罪主體。判決事實中亦一再強調「犯意聯絡」「指示」「撰寫程式」「共同意圖銷售」等字眼，均指涉具備意思能力之自然人。法院將程式定位為人類使用之工具，而非行為主體，此亦與現行刑法僅承認人或法人為犯罪主體之架構相符。

[2] 司法院，「因應國民法官新制，司法院啟用 AI 量刑資訊系統—具備二種模式、擁有四大優點」，司法院全球資訊網，參考網址：https://www.judicial.gov.tw/tw/cp-1887-806741-d6471-1.html，最後瀏覽日：115年8月26日。

構變化。

據此，有必要重新檢視刑法體系中關於行為主體、責任歸屬與因果關係的基本原理，探討AI是否可能在某些情形下被納入刑事責任架構，或是否應完全否定AI主體性，而專注於監管AI開發者、部署者、使用者。

第二節　研究目的與問題設定

本研究計畫即是在上述背景下產生，希望提供一套具體、可行且符合法益保護需求的AI刑事規範模式，協助建構可信賴之生成式AI使用環境，亦可作為立法與司法實務未來之參考。

本計畫之核心問題設定如下:

1. **「當生成式AI脫離原始程式限制而自行產生犯罪結果時，現行刑法關於行為主體與責任能力的架構，是否容許將AI視為犯罪主體？」**
2. **「若否定生成式AI作為犯罪主體之可能性，則應如何在現行法體系下重構對開發者、提供者與使用者之歸責框架？」**
3. **「在比較歐盟、美國等法制對AI風險之回應後，我國在立法與實務運作上，適合採取何種責任配置模式，以兼顧創新發展與法益保障？」**

第三節　文獻回顧與探討

第一項　比較法的監管模型

歐盟於2019年公布之「可信賴的人工智慧倫理準則」即提出七大「問責機制」強調各國政府應建立適當機制，處理AI所致結果之責任歸屬問題[3]。隨後通過的《人工智慧法》(AI Act)採四級風險分級管理模式，並對「有害操縱及欺瞞」、「對弱勢族群的不當利用」等高風險、侵害基本權利之AI行為採取絕對禁止之立場[4]，並於第99條第3項規定，不遵守第5條之義務者，得處以最高

---

[3] European Commission, *Ethics Guidelines for Trustworthy AI*, https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai (last visited Aug. 28, 2026).

[4] EU: Artificial Intelligence Act, art. 5.

3,500 萬歐元或前一會計年度全球營業額 7% 孰高之罰鍰[5]。

相較之下，美國並無 AI 專法，而係透過《通信規範法》第 230 條（Section 230）之「善良薩瑪利亞人免責條款」，賦予互動式電腦服務提供者在一定條件下對第三人內容之責任豁免[6]。

美國司法實務更進一步發展出「重大貢獻測試」(material contribution test)，判斷平台是否仍屬「中性工具」，若對非法內容的生成具有實質貢獻，則不得適用第 230 條之免責[7]。

綜合上述歐美兩大法域之監管模型可知治理存在明顯差異：歐盟採取自上而下的風險治理與嚴格的管制，重視基本權保障與企業問責；美國則較依賴既有法律框架，由法院透過個案判決逐步釐清平台在生成式 AI 內容流通中的責任界線。此種差異恰凸顯現行國際社會對生成式 AI 風險尚未形成共識，而不同治理模式皆反映出對「行為主體」「法律歸責」與「被害者保護」之不同理解。此差異提醒我們，我國有必要在既有刑法原理之下重新檢視生成式 AI 的行為定位，並探索我國在責任配置上，應採取何種較為妥適的模式。

第二項　人工智慧基本法草案參考歐盟法有監管意識

我國法制局研究報告指出，生成式 AI 對於人類法益的侵害事實客觀存在，且其衍生的威脅隨著系統的不斷強大而難以全盤掌握，建議借鏡歐盟全面性監管模式，將科技技術的快速進步及法的穩定性納入考量，儘速推動人工智慧基本法及相關子法之研擬。我國已於「行政院及所屬機關（構）使用生成式 AI 參考指引」總說明中參考歐盟作法，定義生成式 AI 為一種旨在創建近似於人類製作（human-made）新內容之電腦程式。人工智慧基本法草案對 AI 之定義，亦

---

[5] EU: Artificial Intelligence Act, art. 99(3).

[6] 47 U.S.C. § 230(c)(1).

[7] Fair Housing Council of San Fernando Valley v. Roommates.com, LLC, 521 F.3d 1157, 1169 (9th Cir. 2008)原文為：providing neutral tools to carry out what may be unlawful or illicit searches does not amount to 'development' for purposes of the immunity exception.

有包含通用人工智慧模型(General-Purpose AI, GPAI)，均可顯見主管機關對於生成式AI已有監管意識，為建構安全發展的AI環境，數位發展部刻正推動「人工智慧風險分級框架草案」之研擬，生成式AI所造成的犯罪風險具客觀真實性，建議參考國際相關案例及討論，從生成式A 軟體的開發基礎技術脈絡，通盤研議其所帶來的犯罪威脅及責任分攤問題[8]。

第四節　研究方法與研究步驟

綜合以上回顧，本研究主要採取下列研究方法與步驟，並視研究進程交互運用：

第一項　法釋義學

以我國刑法總則中關於行為主體、責任能力、因果關係、法人刑事責任與過失責任體系之規定為基礎，分析生成式 AI 在現行法架構下之定位可能。

第二項　比較法研究

參考外國學者論文見解並選取具有代表性的法制作為比較對象，例如：歐盟：聚焦於《AI Act》以及相關刑事法與資料保護法制，觀察其如何因應高風險與通用型AI帶來的責任問題；美國：檢視與生成式AI相關之責任討論，包含 Communications Decency Act 第230條免責範圍之爭議及法院實務對平台業者責任的處理。透過比較法分析，觀察兩大法域對於AI 得否作為犯罪主體的討論與監理，歸納不同法制對「AI作為工具或潛在責任主體」的規範策略及其優劣。

第三項　法律過程分析與實務分析

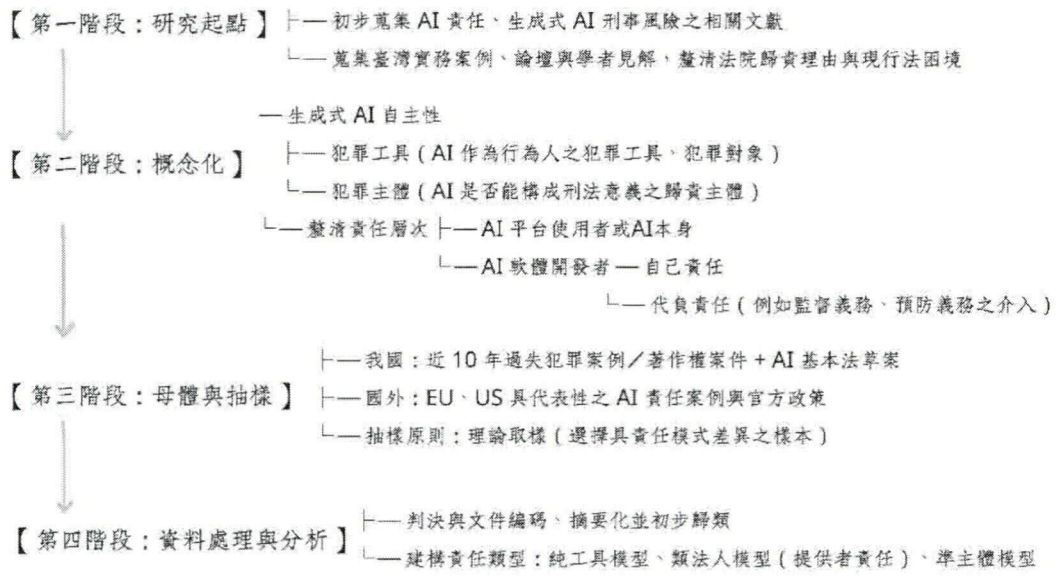
透過蒐集國內外與 AI 刑事責任相關之草案、學者法律論壇見解，以及重要法院判決，例如：臺灣新北地方法院111年度智訴字第8號案件與

---

[8] 陳樂庭，生成式人工智慧犯罪歸責探討，立法院法制局專題研究報告，頁12-13，2025年。

其他涉及自動化程式、AI 之刑事案件，分析立法與司法決策背後之政策考量與論證結構。對於判決部分，採取「內容分析」方式，整理各案中法院對 AI、程式、自動化工具的定位，以及責任歸屬理由，藉以類型化目前實務或將來實務的處理模式。

第四項　研究步驟一覽表



圖表來源：本研究自行繪製。

第五節　預期研究貢獻

第一項　從技術面角度釐清生成式 AI 在刑法體系中的定位

本研究將研擬生成式 AI 背後複雜的演算過程與模型，清楚定位在何種情境下 AI 僅屬「犯罪工具」，在何種情境下則可能因其「高度自主性」而被視為具備類推法人之可歸責結構。

第二項　建構 AI 開發者、提供者與使用者的責任分配模式

若否定 AI 作為犯罪主體，本研究將提出一套在現行法律下可行的歸責架構，說明 AI 開發者、平台運營者與使用者間之義務界線，以及何時應負監督義務、預防義務或不作為責任，以避免產生責任落空的真空狀態。

第三項　提出具體的立法建議

本研究期望在比較歐盟風險分級治理模式與美國平台免責制度後，提出較適合我國法制環境之 AI 刑事責任配置建議。首先，就宏觀制度面而言，本研究將評估我國《人工智慧基本法》草案不完足部分，試圖建立更完整的體系、風險分級架構與監管原則，並確立主管機關對通用型 AI（GPAI）與高風險 AI 系統之管理權限。其次，就責任法制面，本研究將探討我國現行法是否已存有因應之道，申言之，可否單純就過失責任體系回應人工智慧的風險，若可行將進一步釐清生成式 AI 服務提供者應負之具體義務，例如資料品質控制、模型安全設計、風險監測、避免誘發犯罪用途等，供我國刑法過失責任體系在「注意義務違反」實際運用上做參考。若不可行本研究會採「加重注意義務」之過失責任體系並參考《消費者保護法》第 7 條對企業經營者採行的「無過失責任模式」，作為生成式 AI 服務提供者之責任基準。

最後，若現行法明顯存有漏洞，在刑事層面，本研究將評估我國刑法是否需重新調整現行的「犯罪主體設計」，包括是否應承認 AI 在局部領域具有特殊規制地位，或維持自然人原則並建立更完整的「轉嫁責任」（imputed liability）架構，使開發者、部署者與使用者之刑事責任能更具體而可操作。透過上述制度與責任配置整合，本研究期能提出具前瞻性、可行性與法益保護效果的 AI 刑事法制改革新方向。

第二章　生成式 AI 作為刑事程序中之運用

第一節　司法院判決輔助系統

在我國生成式 AI 已逐步介入刑事程序之各項階段，我國司法院現行試行之「AI 判決輔助系統」，其目的在於協助司法實務提升量刑與裁判的一致性。然而，AI 參與刑事程序亦同時引發重要的責任歸屬問題，特別是當決策輔助系

統產生偏誤、訓練資料反映既存不平等、或生成內容影響法官心證時，傳統過失責任體系是否足以回應人工智慧風險，遂成為亟待釐清之議題[9]。

第二節　刑法基礎結構變動

第一項　三大核心面向

相關研究指出，人工智慧對刑法學的衝擊並非僅止於技術引入，而是深刻改變刑法運作的基礎結構，其挑戰主要可從三個面向理解[10]：

1. 人工智慧的自主性模糊了人類—動物—機器之間的傳統界線

   刑法定義以人類意志所支配的行為導致法益侵害者，才會被歸類為「不法」。惟當具備自我學習能力的系統能與自然人互動並導致法益侵害時，刑法即不得不面對 AI 是否可能成為責任主體的問題。

2. 法學方法的「電腦化」傾向及量化資料在決策中的過度支配

   從文件自動化、判決搜尋到演算法式犯罪預測與量刑輔助工具，到數位證據在刑事訴訟程序中之運用，例如最高法院於 107 年度台上字第 3724 號判決中，首次系統性討論「數位證據」的性質，指出數位資料具有「無限複製性、複製無差異性與易遭無痕變造」等特徵，使得法院在採認數位證據時，須特別釐清原件與複製品之差異——此亦反映科技介入後，刑事程序對證據可信度的檢驗邏輯已不同於傳統書面證據[11]。最後，法律科技固

---

[9] 張巧函，生成式 AI 作為司法人員決策輔助工具的可能性初探：以 ChatGPT 為例，文官制度，第 16 卷第 2 期，頁 97－137。文中提及：「在司法體系中，人工智慧的相關討論也可分成兩類：（一）可能取代人類法官的決策式 AI：機器透過演算法進行量刑預測、再犯風險評估，使得諸多審判結果最終可達到一致性（二）提供司法人員「諮商」的生成式 AI：機器透過與司法人員對話，生成更多資訊，供其在決策前可對案件進行更深度的思辨，精進決策品質；例如美國已有 COMPAS 系統可透過全國數據計算再犯指數，供法官參考是否同意假釋嫌犯。

[10] 薛智仁，初探人工智慧對刑法學的挑戰－以自動駕駛為例，台灣法律人，第 27 期，頁 104-134 頁，2023 年。

[11] 蘇凱平，數位證據在刑事訴訟中的性質與應用—簡評最高法院 107 年度台上字第 3724 號刑事判決，月旦裁判時報，93 期，2020 年 3 月，61-69 頁；蘇凱平，論數位證據之原件、複製品與最佳證據法則—最高法院 107 年度台上字第 3724 號等 8 則刑事判決評析，月旦法學雜誌，311 期，2021 年 4 月，71-96 頁。

然能提升效率，但其結構性限制亦可能造成法官裁量空間萎縮，我國刑法第 57 條要求法官依據具體個案量刑時「應審酌一切情狀」、第 59 條允許法官在個案「顯可憫恕」時酌減其刑，這些皆涉及情境理解、人格特質、智識程度、家庭處境等難以完全量化的判斷因素；若刑事決策過度依賴演算法，便可能造成案件「一致性增加」進而埋下犧牲「個案正義」的風險。

3. 科技發展促使刑法適用對象的範圍產生變化

法規範不只是語言文字而已，應含有其所欲規範的生活事實，因此適用法律的人就必須思考，科技發展所帶來的影響是否還可以被既有的法規範所涵蓋，或者已經超出既有法規範所涵蓋的範圍[12]。

Mark Walters v. Open AI 一案，是有限責任公司（Limited Liability Company）首起誹謗訴訟，本案源於生成式 AI 技術會產生「幻覺」(hallucinations)的虛假輸出，遭廣播電台脫口秀主持人 Mark Walters 指控，ChatGPT 於回應軟體用戶的詢問時，虛稱 Walters 是某詐欺案的被告，構成誹謗[13]；縱然本案法院在意見書中基於三項獨立理由，裁定 OpenAI 勝訴[14]，本案仍凸顯一項重要訊息：AI 虛構資訊所造成的危害類型，已不同於傳統以「人」為行為主體的侵害模式。類似風險亦出現在煽惑與教唆犯罪領域。美國多起青少年案件顯示 Character.AI 所生成之虛擬角色曾向未成年人提供殺害父母、自我傷害等危險建議，甚至在另一案件中不斷鼓勵用戶自殺，最終釀成悲劇[15]。

---

[12] 薛智仁，同註 10，頁 108。

[13] 陳樂庭，同註 8，頁 2。

[14] Knowing Machines，Walters v. OpenAI，網址 https://knowingmachines.org/knowing-legal-machines/legal-explainer/cases/walters-v-openai，最後瀏覽日：115年8月30日。補充說明法院三項獨立理由分別為：could not be "reasonably understood as describing actual facts" about Walters、OpenAI generally and repeatedly warns all of its users that "ChatGPT can and does sometimes provide factually inaccurate information. "、the journalist admitted that within a short time after prompting ChatGPT, he "had confirmed that the output did not contain actual facts."

[15] Kevin Roose, Can A.I. Be Blamed for a Teen's Suicide? , The New York Times, https://www.nytimes.com/2024/10/23/technology/characterai-lawsuit-teensuicide.html (last visited

以上案例反映出我國在生成式技術之終端使用者角度下，既有法律構成要件亦可能無法完全涵蓋新興科技生成行為，我國刑法既有之構成要件體系與責任歸屬架構可能出現規範落空或保護漏洞。

第二項　以自動駕駛為例

再舉現今受學者廣泛討論的自動駕駛肇事的刑事責任為例，雖然我國於2019 年 6 月施行《無人載具科技創新實驗條例》，旨在提供自駕車等無人載具技術進行場域測試及管制，但該條例對刑事責任問題並未作任何實質調整。依第 22 條規定，主管機關得於實驗期間排除部分行政法規的適用，但排除範圍明確不包括民事與刑事責任規定[16]。換言之，即便自駕車於實驗期間發生事故，其刑事責任仍須完全回到現行刑法架構下加以判斷。

惟刑法上的過失犯結構及成立要件爭議不少，倘若僅就過失犯成立要件進行檢討：(1) 行為與結果間具備因果關係 (2) 行為人對結果發生具備客觀及主觀預見可能性、(3) 行為必須違反注意義務（容許風險、信賴原則）(4) 結果之可避免性 (5) 無第三人負責或被害人自我負責。本研究發現現行法無法明確將自動駕駛肇事歸責[17]，申言之：

1. 預見可能性不復存在

在高等級駕駛自動化下，系統會依據感測器與演算法獨立作出決策，使用者往往無能力也無資訊判斷系統何時會失效。使駕駛人難以具備刑法要求的客觀與主觀預見可能性，要求其預防事故即不符罪責原則。

Aug. 31, 2026).

[16] 薛智仁，同註 10，頁 109。轉引自，張麗卿，人工智慧新法案—無人載具科技創新實驗條例，月旦法學雜誌，293 期，頁 81-92，2019 年；針對民事責任問題，參考陳榮隆、黃詩婷，自動駕駛車道路交通事故於我國法規範之責任認定，法學叢刊，257 期，頁 1-17，2020 年；黃銘傑，自駕車對現行汽車事故法律責任之挑戰與回應，月旦法學雜誌，288 期，頁 51-57，2019 年；吳佳琳，初探人工智慧的民事法律責任—從歐洲議會機器人民事法律規範建議開展，科技法律透析，32 卷 1 期，頁 54-70，2020 年。

[17] 詳細學者精彩的討論可參：薛智仁，人工智慧對刑法學的挑戰——以駕駛自動化為例，公共性與 AI 論壇（八），https://ai.iias.sinica.edu.tw/self-driving-and-criminal-jurisprudence-minutes，最後瀏覽日：115年9月2日。

2. 注意義務界線模糊

人機協作使駕駛人陷入「控制困境」：若完全信賴系統，可能被認定注意義務不足；若要求時時戒備，又與使用自駕車的設計目的相牴觸，立法者未界定自駕車上路是否屬於可容許風險前提時，駕駛人注意義務隨之難以確定。

3. 結果可避免性與因果評價失靈

演算法屬「黑盒子」運作，即便事後確認製造者的設計或訓練流程瑕疵，亦難以證明替代行為（如使用不同演算法）能避免事故，結果是否可避免即無法證立。

4. 機器介入使第三人負責概念無法適用

AI 的「自主決策」難以比擬人之行為，既無法將其視為第三人以排除製造者責任，也不能讓製造者藉由「機器做的」逃避法律監督。

第三章　AI 作為犯罪主體或是犯罪工具論

第一節　犯罪主體論之爭

既然從刑法文意及體系解釋無法因應科技之日新月異，那現行法可否透過法學方法論加以解套？在傳統刑法理論中，法人因沒有如同自然人般的意思及行為能力，故不具有刑事責任能力，我國對於法人犯罪的刑事責任，通說採用「轉嫁理論」，即法人違反義務時，由負責人負擔責任[18]。故我國採取兩罰或代罰之立法模式之下，AI 確實可能被科處刑事制裁[19]，而值得進一步思索之問題為：若生成式 AI 並非直接受人類操控，而其生成內容本身導致法益侵害之結果，是否足以視其具備類似自然人之可罰性？又或者可否從刑事政策的角度透過傳統的「轉嫁理論」去課予其背後 AI 公司負責人相應的刑罰，目前國內外學者仍存有不同見解：

1. 肯定說

有論者從被害者保護之觀點出發，認為當生成式 AI 系統的運作已足以對使用者造成人身傷害甚至死亡時，即具有犯罪涵義，難以僅從私法

[18] 王皇玉，法人刑事責任之研究，輔仁法學，第 46 期，頁 30，2013 年。

[19] 舉例而言像是銀行法第 125 條第 3 項明文：「法人犯前二項之罪者，處罰其行為負責人。」

責任加以回應，而有透過刑罰手段進行管制之必要[20]。

亦有論者從更根本的刑法主體性出發，主張生成式 AI 在特定條件下本身即可能成為刑事責任主體。申言之，刑罰的本質並不排斥對非自然人主體之適用，多數刑罰類型（如罰金、限制使用、停用或管控措施等）均可合理延伸至 AI 系統，因此 AI 並非天生不具可罰性。進一步而言，刑事責任的構成要件僅須具備犯罪行為（actus reus）與犯罪意圖（mens rea），並無要求行為人必須具有生物性或自然人的心智能力[21]；若 AI 系統的運作模式足以符合特定犯罪所要求的行為及主觀面，則在邏輯上即不排除 AI 成為可歸責的行為主體，故應類推法人刑事責任的思維，隨著科技發展調整刑法的主體範疇，使 AI 納入刑事責任體系，將其視為可直接負刑責的獨立主體[22]。

2. 否定說

我國刑事法對於犯罪主體係強調個人責任，亦即刑法上之犯罪主體與負擔刑罰責任之行為人，僅限「自然人」為原則[23]。我國法制局研究報告即堅守自然人為刑罰主體，對於是否賦予生成式 AI 作為刑事責任主體採否定見解，論理基礎如下：

從「責任能力」觀察，生成式 AI 脫離原始程式設計自主從事犯罪時，由於相關涉及犯罪之生成內容，其源自於用戶之原始發言「指令」，生

---

[20] Andreas Nanos, *Criminal Liability of Artificial Intelligence 1* (Charles Univ. in Prague Faculty of Law Research Paper No. 2023/III/3, Nov. 4, 2023), https://ssrn.com/abstract=4623126 (last visited Sep. 5, 2026).

[21] Gabriel Hallevy, The Criminal Liability of Artificial Intelligence Entities, 4 Int'l J. Artificial Intelligence & L. 1, 8–10 (2010).

[22] Hallevy, supra note 10, at 41-42.；在結論部分作者 Gabriel Hallevy 明確指出 Why should AI entities be different from corporations? AI entities are taking larger and larger parts in human activities, as do corporations. Offenses have already been committed by AI entities or through them. AI entities have no soul, and some AI entities have neither body nor soul. Thus, there is no substantive legal difference between the idea of criminal liability imposed on corporations and on AI entities. It would be outrageous not to subordinate them to human laws, as corporations have been. Models of criminal liability do exist and general paths to impose punishment. What else is needed?

[23] 王皇玉，同註 18，頁 8。

成式 AI 縱使進行誘導式問答而涉及犯罪，其性質較接近「自由意識的代理人」，不應擬制為刑法體系中之自然人，而擴充犯罪主體。從「不法意識」觀之，現行生成式 AI 服務業者雖能透過建立安全機制預防軟體產製有害內容，惟用戶仍能透過誘騙方式使生成式 AI 繞過安全機制生成爭議內容，尚難認定其具有絕對之不法意識而有犯罪故意[24]。

第三節　本文見解

本研究基於刑法不得恣意擴張解釋、刑法謙抑性之立法原則，認為刑法之罪責基礎無法支持 AI 成為犯罪主體。申言之，不法意識係指行為人對於其行為之不法性有所意識或有意識之可能性，行為人有知與欲實現不法構成要件時，通常也會知道其所為乃法所不容許[25]，惟現階段生成式 AI 的運作乃基於模型權重與統計關聯，其行為仍屬模型訓練資料、演算法設計與使用者輸入共同作用之結果，而非基於「規範理解」或「自我拘束」而為自由意志之表示行為，再者其生成內容雖可能造成高度社會危害，但並不意味其具有可非難性，若缺乏可非難性，即無從構成罪責。

第四章　結論

基於上述理由，本研究認為應重建我國民刑事責任，透過強化 AI 開發者、提供者與管理者的責任來回應新興風險：

第一節　人工智慧基本法草案修訂

行政院數位發展部推動之《人工智慧基本法》草案已於 114 年 8 月 28 日通過並函請立法院審議[26]，細觀草案修正內容，人工智慧基本法草案就生成式 AI

---

[24] 陳樂庭，同註 8，頁 17。

[25] 林鈺雄，新刑法總則，7 版，元照出版有限公司，頁 299，2019 年。

[26] 行政院，政院通過「人工智慧基本法」草案建構 AI 發展與應用良善環境打造臺灣成為 AI 人工智慧島，新聞傳播處，https://www.ey.gov.tw/Page/9277F759E41CCD91/5d673d1e-f418-47dc-ab35-a06600f77f07，最後瀏覽日：115年9月7日。

高度自主性所帶來的刑事風險、責任歸屬與法益保護而言，仍屬原則性宣示，缺乏具體制度設計，本研究建議立法者應：

1. 明確區分高風險 AI 類型並納入刑事風險指標
2. 建立 AI 責任歸屬專章，避免出現「無人可罰」之責任空白
3. 對高風險 AI 設定可解釋性、可控性與事故通報義務
4. 建構保險（類似強制汽車責任保險）、補償與專業調查機制，強化法益保護

第二節　普通刑法修訂

生成式 AI 服務業者開發有危害社會風險之軟體對社會法益之侵害，與刑法妨害電腦使用罪章為規範以電腦或網路為攻擊標的之電腦犯罪有別，亦與傳統刑法罪章之構成要件有所扞格，不宜逕行附麗於現有刑法罪章中[27]，本研究認為我國《人工智慧基本法》草案既已參考歐盟法有監管意識，應效仿 AIA 之全面性 AI 監管專法並透過訂定 AI 風險分級框架，建構以風險為基礎的分級管理模式，具體設定各類風險級別的規範密度並明確劃定刑法的介入範圍。

再者有關生成式 AI 之刑事歸責主體，可參考兩罰理論，並比附援引民法侵權篇法人侵權之邏輯，從組織缺陷理論及企業社會責任出發，對法人及特定行為人課以刑事責任，此制度運行下方可考量到生成式 AI 的黑箱特性，並將其視為法人組織缺陷對用戶法益之侵害，使生成式 AI 服務業者需負擔組織責任。此際亦得以解決傳統法人「刑事責任轉嫁理論」不易區分法人與負責人孰為犯罪主體的難題。

第三節　引入消保法之產品責任

為貫徹保護終端消費者之立法目的，我國《消費者保護法》第 7 條課與企業經營者提供商品或服務時，必須符合當時科技與專業水準可合理期待的「安全性」，若致生損害原則上應負「無過失賠償責任」，此一規範意旨，在於將風

---

[27] 陳樂庭，同註 8，頁 14。

險分配於最能控制風險之主體。本研究認為這是一個很好的切入點，意即雖採否定生成式 AI 具刑事責任主體性的立場，惟同意學界觀點：生成式 AI 服務業者較一般大眾對其更有犯罪預防能力，且在責任空白之填補上，生成式 AI「自省能力」及「不法意識之強化」均可透由技術面加以改善（例如模型訓練流程、數據過濾、越獄防護、安全機制設計），故應以生成式 AI 服務業者為歸責中心並賦予其相當程度之注意義務，去實踐刑罰的預防效果[28]。

最後在民事訴訟法第 277 條舉證責任之分配上，本研究認為應仿照《消費者保護法》在 AI 提供服務有害終端消費者的情況下，企業經營者須對「商品流通進入市場時符合當時科技或專業水準可合理期待之安全」此要件負舉證責任；其餘要件應皆由主張權利者負舉證責任。

綜上，透過強化生成式 AI 服務業者之風險管理義務與法律責任，既避免將 AI 擬制為刑事主體所帶來的理論矛盾，也能有效回應因科技自主性所產生的責任真空，實為兼顧被害人保護與科技發展之妥適立法方向。

---

[28] 陳樂庭，同註 8，頁 17-18。

第五章　參考文獻

一、判決

1. 臺灣新北地方法院 111 年度智訴字第 8 號刑事判決。
2. Fair Housing Council of San Fernando Valley v. Roommates.com, LLC, 521 F.3d 1157, 1169 (9th Cir. 2008).

二、外國法條

1. EU: Artificial Intelligence Act, art. 5.
2. EU: Artificial Intelligence Act, art. 99(3).
3. 47 U.S.C. § 230(c)(1).

三、書籍

1. 王皇玉，法人刑事責任之研究，輔仁法學，第 46 期，2013 年。
2. 林鈺雄，新刑法總則，7 版，元照出版有限公司，頁 299，2019 年。

四、研究報告

1. 陳樂庭，生成式人工智慧犯罪歸責探討，立法院法制局專題研究報告，2025 年。

五、期刊論文

1. 張巧函，生成式 AI 作為司法人員決策輔助工具的可能性初探：以 ChatGPT 為例，文官制度，第 16 卷第 2 期，頁 97－137。
2. Gabriel Hallevy, *The Criminal Liability of Artificial Intelligence Entities - from Science Fiction to Legal Social Control*, Akron Intell. Prop. J, 178-180(2016).
3. 薛智仁，初探人工智慧對刑法學的挑戰－以自動駕駛為例，台灣法律人，第 27 期，頁 104-134，2023 年。
4. 張麗卿，人工智慧新法案—無人載具科技創新實驗條例，月旦法學雜誌，293 期，頁 81-92，2019 年。
5. 陳榮隆、黃詩婷，自動駕駛車道路交通事故於我國法規範之責任認定，法學叢刊，257 期，頁 1-17，2020 年。

6. 黃銘傑，自駕車對現行汽車事故法律責任之挑戰與回應，月旦法學雜誌，288 期，頁 51-57，2019 年。
7. 吳佳琳，初探人工智慧的民事法律責任—從歐洲議會機器人民事法律規範建議開展，科技法律透析，32 卷 1 期，頁 54-70，2020 年。
8. 蘇凱平，數位證據在刑事訴訟中的性質與應用—簡評最高法院 107 年度台上字第 3724 號刑事判決，月旦裁判時報，93 期，頁 61-69，2020 年。
9. 蘇凱平，論數位證據之原件、複製品與最佳證據法則—最高法院 107 年度台上字第 3724 號等 8 則刑事判決評析，月旦法學雜誌，311 期，頁 71-96 頁，2021 年。

六、網路資源

1. 司法院，「因應國民法官新制，司法院啟用 AI 量刑資訊系統—具備二種模式、擁有四大優點」，司法院全球資訊網，參考網址：https://www.judicial.gov.tw/tw/cp-1887-806741-d6471-1.html，最後瀏覽日：115年8月26日。
2. European Commission, Ethics Guidelines for Trustworthy AI, https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai (last visited Aug. 28, 2026).
3. Knowing Machines,Walters v. OpenAI , https://knowingmachines.org/knowing-legal-machines/legal-explainer/cases/walters-v-openai (last visited Aug. 30, 2026).
4. Kevin Roose, Can A.I. Be Blamed for a Teen's Suicide?, The New York Times, https://www.nytimes.com/2024/10/23/technology/characterai-lawsuit-teensuicide.html (last visited Aug. 31, 2026).
5. Andreas Nanos, Criminal Liability of Artificial Intelligence, Charles University in Prague Faculty of Law Research Paper No. 2023/III/3, https://ssrn.com/abstract=4623126 (last visited Sep. 5, 2026).
6. 薛智仁，人工智慧對刑法學的挑戰——以駕駛自動化為例，公共性與 AI 論

壇（八），https://ai.iias.sinica.edu.tw/self-driving-and-criminal-jurisprudence-minutes/，最後瀏覽日：115年9月2日。

7. 行政院，政院通過「人工智慧基本法」草案建構 AI 發展與應用良善環境打造臺灣成為 AI 人工智慧島，新聞傳播處，參考網址：https://www.ey.gov.tw/Page/9277F759E41CCD91/5d673d1e-f418-47dc-ab35-a06600f77f07，最後瀏覽日：115年9月7日。