JUDICIAL WEEKLY 中華民國115年5月29日 第2310期
司法周刊 【司法文選別冊】

AI 在法院運用與界限

文／廖建瑜 *

摘要

AI 技術經歷符號主義、機器學習、深度學習到大型語言模型（LLM）與生成式 AI 的發展，核心能力體現在數據分析與預測（如量刑預測）、自然語言處理（如摘要與草擬文書）及流程自動化（如 OCR 與自動分案）。這些能力被視為解決訴訟爆炸、提升司法效率及運用大數據的關鍵驅動力。

各國法院積極導入 AI，例如新加坡推行 AI 翻譯、交通事故索賠系統及 GPT-Legal 等輔助工具，並與美國 HarveyAI 合作。中國與巴西則運用 AI 於卷證分析、量刑建議及上訴審查自動化。然而，美國使用 COMPAS 系統評估再犯風險引發種族偏見與正當法律程序的爭議。

AI 應用伴隨著顯著風險，包括「AI 幻覺」（虛構判例）、演算法偏差（種族或性別歧視）、錨定效應影響法官心證，以及 Deepfake 證據的真實性挑戰。此外，將卷證上傳至商用雲端 AI 亦引發嚴重的個資外洩與機密保護疑慮。

各國針對法院 AI 運用制定了不同強度的規範。美國採分散式管理，透過個別命令與律師倫理規範約束；歐盟強調倫理與基本權保障，堅持人類保有最終決定權；南美如阿根廷與巴西則採嚴格的行政許可與風險分級制。台灣雖已推動語音辨識、網頁聊天機器人等應用，並制定人工智慧基本法

* 臺灣高等法院刑事庭審判長兼法官、國立成功大學法律學博士

摘要

及發布行政院及司法院參考指引，但仍缺乏針對律師外部使用者的規範、明確的證據法則與風險分級審查制度。

本文建議台灣司法應建立 AI 風險矩陣，將應用場景依影響程度分為低、中、高風險區進行分級控管。同時，應設計詳細的 AI 使用紀錄表，以確保數位履歷的可追溯性，並對外部律師與機關內部所屬人員建立雙軌的揭露義務與違規處置機制，另就 AI 生成之資料如何具有證據資格應參考美國聯邦證據規則第 707 條草案之精神，以及法院欠缺自建封閉式地端部署之生成式 AI 模型下，法院卷證分析如何運用 AI 之具體指引。最終目標是建立風險分級與人機協作為核心的司法 AI 模式，確保科技輔助不損害審判獨立與程序正義。

關鍵詞：生成式 AI（Generative AI）、人機協作（human-in-the-loop, HITL）、AI 幻覺（hallucination）、AI 偏見（Bias）、深偽（deepfake）、風險矩陣（Risk Matrix）

目　次

壹、AI 的發展與在法律領域適用之核心能力 2

一、AI 的發展階段 2

二、AI 在法律領域可以運用之核心能力 4

三、AI 進入法院的驅動力 6

貳、AI 在法院運用的態樣與衍生問題 7

一、AI 在法院運用態樣 7

二、AI 在法院運用衍生之問題 11

參、AI 在各國法院運用之規範 18

一、美國法院運用 AI 規範與實務趨勢 18

二、歐洲法院運用 AI 規範與倫理原則 27

三、英國、紐西蘭、澳洲、加拿大、香港等英國法系法院 AI 規範 34

四、亞洲國家法院運用 AI 規範 42

五、南美司法系統之 AI 應用規範 48

六、小結：各國法院 AI 規範的共通原則與差異 51

肆、我國現況及未來展望 ── 結語 53

一、盤點目前我國司法使用 AI 現況 53

二、我國司法關於 AI 應用之規範 55

三、台灣與世界各國運用 AI 規範之主要差異 60

四、我國現行 AI 運用規範之漏洞 61

五、未來展望 62

AI 在法院運用與界限

壹、AI 的發展與在法律領域適用之核心能力

人工智慧（artificial intelligence, AI）一詞，首見於 1955 年 8 月 31 日約翰麥卡錫（John McCarthy，後被尊稱為 AI 之父）等人為在達特茅斯學院召開暑期研究於提案中所產生之名詞[1]，經過 70 年的發展後，AI 已成為資本市場競逐的標的[2]，也是莘莘學子爭相學習的職場敲門磚[3]，並成為職場上必備的工作技能[4]。

一、AI 的發展階段

AI 並非橫空出現，而係隨著技術演進而有不同階段的呈現：

（一）AI 符號主義（Symbolicism）與規則系統階段（1950s–1980s）

大約在 1950 至 1980 年代，早期 AI 核心是「如果… 則…」(If-Then) 的邏輯規則。專家先將所有的判斷規則輸入系統，人工智慧可以通過處理符號和符號運算來實現。此階段係由人先決定規則強調專家系統，AI 結果產出均係透過可解釋的邏輯推理，常見體現在一般資料庫之查詢[5]。

（二）AI 機器學習 (Machine Learning, ML) 階段（1990s–2010）

人工智慧的機器學習階段是指不依賴專家手動輸入所謂規則，而是餵給電腦海量的數據資料，讓它自己從數據中利用數學方式找出參數之學習過程，找到最佳預測模

1 J. McCarthy, M. L. Minsky, N. Rochester, C. E. Shannon (1955), A PROPOSAL FOR THE DARTMOUTH SUMMER RESEARCH PROJECT ON ARTIFICIAL INTELLIGENCE, http://jmc.stanford.edu/articles/dartmouth.html, last visited on 2025/10/14.

2 RexAA，AI 投資佔比超 50%：1,927 億美元背後的資本狂歡與困境，鉅亨號，2025/10/04，https://hao.cnyes.com/post/198471, last visited on 2025/10/14.

3 許維寧、張博瑞，資工電機大熱門！台大、師大碩士班報考創紀錄考場爆滿還需多借，聯合新聞網，2025/02/13，https://udn.com/news/story/6928/8547019, last visited on 2025/10/14.

4 莊雨潔，AI 世代求職焦慮調查：月花 5000 元投資進修，1111 人力銀行產經新聞網，2025/07/04，https://www.1111.com.tw/news/jobns/161433, last visited on 2025/10/14.

5 Ran He, Jie Cao, Tieniu Tan (2025), Generative artificial intelligence: a historical perspective, National Science Review, 12(5), nwaf050, https://doi.org/10.1093/nsr/nwaf050.

式。在此階段 AI 擅長預測與分類而無法創造新的內容，常見體現在一般風險評估之預測[6]。

（三）AI 深度學習（deep learning）（2012–2016）與語言理解（2017–2020）

2016 年 3 月 9 日標榜深度學習的 AlphaGo 打敗韓國圍棋棋士李世乭，對於人工智慧模擬人類思維進而模仿人類行為能力一鳴驚人，此種透過層層非線性函數組成的神經網路，取徑不同架構如卷積神經網路（Convolutional Neural Network, CNN）、循環神經網路（Recurrent Neural Network, RNN）和生成式對抗網路（Generative Adversarial Network, GAN），分別擅長不同資料類型[7]，使得 AI 在影像、語音與文字處理能力提昇[8]。自然語言處理（Natural Language Processing, NLP）是一種可讓電腦解釋、操縱及理解人類使用的語言之技術，並進而回應，更重要的是語言資料上下文關聯性能理解，例如代名詞是指誰[9]。常見體現在語意檢索或關鍵句萃取之功能。

（四）大型語言模型與生成式 AI（2020– 迄今）

大型語言模型（Large Language Model, LLM）利用大量資料進行預訓練的超大型深度學習語言模型，為大量參數類神經網路組成，負責理解和生成自然語言文本。而所謂生成式 AI（generative AI）是使用自然語言做為輸入，並由大型語言模型（LLM）提供技術支援而產生文本、圖像、聲音、動畫、3D 模型及影片等之 AI 應用程式，因此除與人聊天互動外，AI 首次具備了寫作的能力，顯而易見發揮在文書處理與論理建構上[10]。

---

6 Iqbal H Sarker (2021), Machine Learning: Algorithms, Real-World Applications and Research Directions, SN Computer Science, 2(3), Article 160, https://doi.org/10.1007/s42979-021-00592-x.

7 林婷嫻，【深度學習】如果電腦有神經，可以教它做什麼？中央研究院研之有物，2017/08/17，https://research.sinica.edu.tw/deep-learning-2017-ai-month/, last visited on 2025/10/14.

8 Olufisayo S. Ekundayo & Absalom E. Ezugwu (2025) Deep learning: Historical overview from inception to actualization, models, applications and future trends, Applied Soft Computing, 159, Article 111482, https://doi.org/10.1016/j.asoc.2025.113378.

9 Sajun, A. (2024). A historical survey of advances in transformer architectures. Applied Sciences, 14(10), Article 4316, https://doi.org/10.3390/app14104316.

10 Ansari, A., et al. (2025). A survey of transformers and large language models for ECG diagnosis: Advances, challenges and future directions. Artificial Intelligence Review, 58, 261, https://doi.org/10.1007/s10462-025-11259-x.

2022 年 7 月 21 日 Midjourney 專長為圖像生成之 AI 應用程式公開，隔月 Jason Allen 即使用創作名為「Théâtre d'Opéra Spatial」作品在科羅拉多州博覽會數位藝術類競賽拿下首獎[11]；同年 11 月 30 日 OpenAI 發表 ChatGPT-3.5，不到一週用戶即突破百萬，2025 年 2 月每週活躍用戶（WAUs）已達 4 億人，同年 10 月已上升至 8 億人[12]，迄今模型已進展至 GPT-5.1[13]，生成式 AI 之使用已變成日常生活的一部分。

二、AI 在法律領域可以運用之核心能力

基於上述的技術演進，AI 在法律領域的核心能力可歸納為三大類：

(一) 數據分析與預測

這主要依賴機器學習，特別藉由監督式學習與特徵工程（Feature Engineering），所謂監督式學習，就像是給 AI 一位導師，拿已知的考題和標準答案去訓練它。所謂考題即特徵工程的重點，必須把一份非結構化的判決書，翻譯成 AI 看得懂的結構化數據。例如，在一個竊盜罪的量刑預測模型中，特徵可能包括：犯罪次數：（3）（數值）；是否和解：（1）（1 ＝是，0 ＝否）；竊盜金額：（55000）（數值）；犯後態度：（2）（1 ＝良好，2 ＝普通，3 ＝惡劣）；答案（Label ／標籤）：這就是該案法官「最終判處的刑期」（例如：6 個月）。

AI 的學習過程或者稱模型訓練透過下列途徑：

1. 餵養資料：例如將過去司法院裁判書公開查詢系統數十萬份「竊盜案」的「特徵」（考題）和「刑期」（答案）一起丟給 AI。
2. 找出關聯：AI 會使用複雜的統計模型（例如「隨機森林 Random Forest」或「類神經網路 Neural Network」）去計算每一個「特徵」與「答案」之間的數學關聯性（權重）。

---

11 Kevin Roose (2022), An A.I.-Generated Picture Won an Art Prize. Artists Aren't Happy., The New York Times, 2022/09/02, last visited on 2025/10/14.

12 林羿君，ChatGPT 週估活躍用戶突破 8 億大關，八個月暴增一倍，鉅亨網，2025/10/09，https://m.cnyes.com/news/print/6183064, last visited on 2025/10/14.

13 OpenAI 2025 年 11 月 12 日發表最新版本，https://openai.com/zh-Hant/index/gpt-5-1/；6 日後 Google 亦發表最新模型 Gemini 3, https://blog.google/intl/zh-tw/products/explore-get-answers/gemini-3/, last visited on 2025/12/05.

3. 學習成果：AI可能學習到「是否和解」這個特徵對於刑期的影響權重非常高；而「竊盜金額」在超過某個閾值後，其影響力才會顯著增加。

4. 產生預測：當一份新案件的特徵（例如：犯罪次數（1），是否和解（0），金額（3000））被輸入時，模型就會基於它學到的權重，計算出一個機率最高的預測刑期區間（例如：2 至 4 個月）[14]。

(二) 自然語言處理

AI 的讀寫能力，也是生成式 AI 的核心技術，透過 Transformer 模型與大型語言模型的技術演進，從關鍵字比對進步到理解上下文，並學會了人類語言中複雜的邏輯、指代、轉折與因果關係。透過抽象式摘要（Abstractive Summarization），AI 可以在閱讀完數百頁的文本後，非以「複製貼上」關鍵句子，而是基於它學到的語言模型，用自己的話重新撰寫出一份涵蓋所有核心事實的摘要 [15]。

甚至具有文件草擬（Drafting）的能力，LLM 本質上是一個超級文字接龍的機率模型，當我們給它一個提示（Prompt），例如：「撰寫一份酒駕判決書的事實段落，被告酒精濃度 0.28 毫克…」，AI 就會根據它學習過的數十萬份不能安全駕駛之判決書範本，去預測「下一個最有可能出現的字」。因此，它會一個字一個字地生成出「被告…於… 時許，駕駛… 沿… 路… 行駛，經警攔查，測得吐氣所含酒精濃度達每公升 0.28 毫克…」。因為在它的模型中，這串文字是機率最高的組合 [16]。

(三) 流程自動化

生成式 AI 透過光學字元辨識（Optical Character Recognition, OCR）、資訊擷取

14 Kleinberg, J., Mullainathan, S., & Raghavan, M. (2017). “Inherent Trade-Offs in the Fair Determination of Risk Scores. In C. H. Papadimitriou (Ed.), 8th Innovations in Theoretical Computer Science Conference (ITCS 2017) (pp. 43:1-43:23). Schloss Dagstuhl--Leibniz-Zentrum fuer Informatik, https://doi.org/10.4230/LIPIcs.ITCS.2017.43.

15 Rush, A. M., Chopra, S., & Weston, J. (2015). A neural attention model for abstractive sentence summarization. Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing (EMNLP) (pp. 379-389). Association for Computational Linguistics, https://doi.org/10.18653/v1/D15-1044.

16 Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., ... & Amodei, D. (2020). Language models are few-shot learners. Advances in Neural Information Processing Systems 33 (NeurIPS 2020) (pp. 1877-1901). Curran Associates, Inc. DOI:10.48550/arXiv.2005.14165.

（Information Extraction, IE）與規則引擎（Rules Engine），運用串接，將規律日常工作主動進行而非被動聽取指令。自動分案的功能而言，其運作流程：

1. 文件數位化（OCR）

當事人提交一份紙本起訴狀（若未採取數位線上起訴），AI 透過 OCR 會掃描這份文件，將其從電子卷證的圖片檔（Image）轉換為可搜尋的文字檔（Text）。

2. 資訊擷取（IE）

透過特殊自然語言技術 AI 會自動閱讀這份文字檔，並像填表格一樣抓取關鍵資訊：例如：原告：（王小明）、被告：（甲建設公司）訴訟標的金額：（5,500,000 元）、關鍵字：（房屋漏水、瑕疵擔保、解約）[17]。

3. 分類與分流（Classification & Routing）

資訊擷取後 AI 可以自動判斷這份起訴狀的案件類型與訴訟標的金額，透過規則引擎（Rules Engine），系統接著會觸發預先設定好的「如果… 則…」邏輯，例如：IF 案件類型＝“ 營建工程 ” AND 訴訟標的金額 1 億元以上＝“ 重大 ”；THEN 分案至 “ 民事 - 營建專股 ” AND 自動將案件設定為重大案件 [18]。其他諸如自動檢查當事人提交的書狀是否符合程序要件、自動生成開庭通知與排程等。

三、AI 進入法院的驅動力

AI 普及並非一蹴即成，但目前成為全球法院共同討論運用的顯學，這背後有三個主要的驅動力：

(一) 驅動力一：訴訟爆炸

全球司法體系共同面對的「癌症」即案件量負擔超載，特別是 covid-19 疫情過後 [19]，在台灣加上詐欺橫行，不論新增（民事）或未結（刑事）案件激增約四分之

17 M. Saravanan, B. Ravindran & S. Raman (2009). Improving legal information retrieval using an ontological framework, Artificial Intelligence and Law 17 (2):101-124, https://doi.org/10.1007/s10506-009-9075-y.

18 Ashley, K. D. (2017). Artificial intelligence and legal analytics: New tools for law practice in the digital age. Cambridge University Press, p56-69.

19 Barry Godfrey, Jane C. Richardson and Sandra Walklate (2022) The Crisis in the Courts: Before and

一[20]，但司法人力（法官、法助、書記官）養成無法速成，且編制上又是受限中央政府機關總員額法人數管制天花板，在案多人少情況下導致法官過勞、案件遲延或審判品質下降，最終侵蝕了人民對司法的信任。 AI 被視為解決此一困境的關鍵解方。它可能是一個 24 小時待命的助理，能將司法人員從繁瑣的文書工作中解放出來，讓法官專注於更核心的法律判斷，進而提升整個法院的案件處理量。

(二) 驅動力二：對司法效率與品質的追求

人民不僅要求正義，還要求即時的正義。 AI 在流程自動化上的能力，能大幅縮短案件等待時間。同時，AI 在數據分析上的能力（如量刑建議），有助於減少量刑歧異，提升司法的客觀性與一致性。

(三) 驅動力三：數據與技術的成熟

AI 的燃料是大數據及資料庫，隨著法院長年推動電子卷證和判決書公開，積累了足夠龐大、乾淨的本土司法數據，足以餵養和訓練出強大的 AI 模型。伴隨技術的突破，尤其是生成式 AI 在自然語言理解與文字生成上的驚人能力，使得 AI 不再是數位玩具，而是真正能投入第一線實戰的生產力工具。

貳、AI 在法院運用的態樣與衍生問題

一、AI 在法院運用態樣

筆者於 2025 年 8 月奉派前往亞洲數位司法最先進之新加坡參訪，本節主要觀察目前新加坡法院 AI 使用情況以及世界其他國家具體展示 AI 如何被運用在法院的日常工作中。

(一) 新加坡

在新加坡 AI 落地在法院運用可分為以下部分：

Beyond Covid, The British Journal of Criminology, Volume 62, Issue 4, July 2022, p 1036–1053, https://doi.org/10.1093/bjc/azab110.

20 劉宜妗，近年法院及檢察機關司法人力配置及相關經費運用之探討，立法院預算中心，112 年 8 月，https://www.ly.gov.tw/Pages/Detail.aspx?nodeid=46365&pid=230909, last visited on 2025/10/16.

1. 對外提升民眾服務

⑴ AI 線上即時翻譯

在多語社會中，語言常常是司法可近性的最大障礙，透過網頁 AI 智慧上下文之沉浸式翻譯功能，立即獲取翻譯，或者在法庭 AI 雙向即時通譯，不必等待人工翻譯，減省開庭時間，亦能清楚知道翻譯正確性，讓語言不再成為司法參與的門檻，而是變成平等參與的基礎保障。

⑵ 交通事故線上請求系統（Motor Accident Claims Online, MACO）

交通事故糾紛，是各國司法系統最常見的案件類型之一，新加坡對於一般民眾提出 MACO 系統，運用 AI 技術如下：

① 決策樹邏輯（Decision Tree）

使用者需回答一系列關於事故情境的問題，例如：車輛相對位置與行駛方向；是否有違反交通規則（闖紅燈、超速）；是否有行人或第三方涉入。系統依據答案，逐層套用既有案例規則，最終生成一個責任比例建議，例如「原告 70%，被告 30%」。

② 量化參考

系統內建過往判決與和解數據，能根據傷害類型、醫療費用、車輛損失等，生成「賠償金額區間」。例如：一名頸部損傷，系統可能顯示過往案例的賠償區間為 3,000-5,000 新幣。

⑶ 離婚財產分配估算器（Divorce Assets Informative Division Estimator, AIDE）

新加坡法律援助局（Legal Aid Bureau）於 2020 年推出，可依使用者提供的婚姻期間、財產種類與貢獻（經濟或照護）等資料，推估夫妻財產分配比例。

上開 AI 生成之資訊讓一般民眾在考慮是否提告前，可先使用 MACO 及 AIDE 系統，評估勝算與風險。若系統顯示自己責任比例偏高或分配金額過低，當事人可能選擇和解而非訴訟，另外亦可作為調解基礎，以系統顯示合理賠償金額據此引導回歸理性範圍。

⑷ 線上爭議解決（Online Dispute Resolution, ODR）

新加坡在家事案件中，AI 可能被用於分流，當民眾提出符合規則之電子訴狀後，AI 系統可以分析案件的複雜性、爭議點（例如子女監護權或財產分配），然後將案件

引導至最適合的解決途徑（例如調解、諮商或法庭審理），案件自動分流以最適途徑解決紛爭，減省民眾訴訟時間成本及訟源。

2. 提供專業律師與法院內部人員（法官及其他人員）使用

⑴ GPT-Legal

新加坡 2024 年 9 月由資訊通信媒體發展局（IMDA）與新加坡法律學會（SAL）共同開發，其關鍵特點是使用本地龐大的法律內容庫 LawNet 進行訓練，該模型的首個具體應用是為超過 15,000 份法院判決書生成摘要，這些摘要仿照「新加坡法律報告」（Singapore Law Reports）的官方判決要旨格式，包含主題詞、重要事實、爭點與法院見解。透過此功能，使用者可於幾分鐘內了解長篇判決的核心內容，大幅節省研究時間。 LawNet 法律檢索平台，特別是其「Intelligent Research」（智慧研究）功能，早已超越了關鍵字比對，它允許法律專業人士使用自然語言提問，例如：「房東在何種情況下可以合法沒收押金？」，系統能理解語意並回傳最相關的法條與判例。

⑵ 智能案例檢索系統（Intelligent Case Retrieval System, ICRS）

智能案例檢索系統是一個案例資料庫，專注於人身傷害案件，並建構出「三層傷害本體（ontology）」以結構化各類傷害類型，使檢索更精確。其定位是「專業端參考工具」，主要供律師、法官與保險公司使用。律師輸入「傷害類型」或「案情描述」，系統即提供相似案例。例如：輸入「脊椎骨折」，系統會顯示過去十年所有相關判決，標示賠償金額區間，並將傷害類型分為：第一層：大類（頭部、脊椎、四肢、內臟）。第二層：細分（脊椎骨折、椎間盤突出）。第三層：更細緻（第七頸椎骨折）。這樣的層級結構，避免檢索結果過於分散。

⑶ Harvey AI

2023 年 8 月新加坡司法機構與美國法律 AI 新創公司 Harvey 簽訂了合作備忘錄，在小額索償庭（Small Claims Tribunals, SCT）試點 AI 應用。2025 年 11 月推出 AI 文件摘要工具，借用 AI 整理眾多文件及證據而生成簡單的摘要，讓兩造清楚對方主張，也讓法官節省時間在審理前了解案情。

⑷ 證據審查（Evidence Review）

生成式 AI 亦被嘗試應用於證據審查與摘要。法院案件常涉及龐大文件量（契約、

電子郵件、證言、報告等），因此新加坡法院正在測試 AI 工具，以自動分類、聚集並擷取關鍵資料。這些系統正於程序前會議、證據開示及卷宗準備等階段試行。

⑸ 新加坡大型律師事務所已率先導入 AI，用於訴訟文件審查、併購盡職調查及合約分析。主要應用包括：

① 文件審查與電子證據開示（e-Discovery）：利用預測編碼與文件分類演算法，自動標記關鍵與機密文件，縮短人工審閱時間。

② 合約盡職調查：AI 可自動擷取重要條款（如控制權變更、賠償條款），協助律師在短時間內完成大批合約審閱。

③ 生成式 AI 文件起草：如與 Harvey 合作，讓律師輸入提示即可生成保密協議、僱傭契約草稿，供後續審閱修訂。

④ 意見摘要化：利用 LLM 工具快速生成內部備忘錄或客戶簡報摘要[21]。

(二) 美國

AI 技術領先的美國，在法院運用 AI 係由各州法院主導自行發展，除有如新加坡應用模式外，其最著名也最具爭議的應用集中在風險評估，所謂替代性制裁犯罪矯正管理剖析軟體（Correctional Offender Management Profiling for Alternative Sanctions, COMPAS）系統。 AI 會根據被告的多項數據（如犯罪史、年齡、社會關係等）來預測其再犯風險。法官在決定是否允許保釋、判處緩刑或決定刑期長短時，會將此風險評分作為重要參考[22]。

(三) 中華人民共和國

在中國的「智慧法院」（Smart Court）系統是全球在 AI 應用上最全面且部署規模最大的，由上海市高級人民法院協作，科大訊飛（iFlytek）公司開發「206 System」，於 2019 年 1 月 23 日首次適用於庭審，該系統除智慧檢索，提供更精準的法條與語意之判例檢索，亦包括卷證分析與爭點整理，包括自動摘要、標記證據、整理雙方爭

21 Nydia Remolina (2025), AI in the judiciary: The Singapore Case, Singapore Management University School of Law Research Paper Forthcoming, pp.8-22.

22 Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016). Machine bias: There's software used across the country to predict future criminals. And it's biased against Blacks. ProPublica. https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing, last visited on 2025/10/18.

點；具有量刑及賠償建議系統，係根據歷史數據提供量刑或賠償金額的參考區間。最後，連針對事實單純或重複性高的案件（如酒駕、簡易民事）具有自動生成判決書初稿生成之功能[23]。甚至檢察官以 AI 模型量化社會危險性，如山西運城人民檢察院開發的風險評估系統，作為羈押與保釋決定參考[24]。

(四) 巴西

在巴西 VICTOR AI 系統透過識別僅屬一般性影響無具有重大法律意義的案件，自動化審查最高法院的上訴案件，法院書記需花費 44 分鐘評估上訴是否符合程序進行，而 VICTOR AI 只需幾秒鐘。巴西勞工法院最高行政管理機關巴西勞工司法最高委員會推出了 Chat-JT，一款生成式 AI 工具，協助法官、法院工作人員及實習生自動化任務（如法律研究、文件分析、標準化摘要的擬定）。該平台於 2025 年 2 月推出，允許用戶根據司法工作流程中的特定需求打造個人化助理。透過與其他內部系統整合，Chat-JT 提供跨內部資料庫的先進搜尋功能，旨在簡化日常運作並改善勞工法院內的決策流程[25]。

二、AI 在法院運用衍生之問題

(一) 資訊正確性風險 ── AI 幻想生成虛構判決或錯誤解釋

目前法院運用 AI 最常見的問題，是源於所謂 AI 幻想（hallucination）或妄想（delusion）在法院文件出現虛構判決案例、法條、法律期刊論文或雖引用真實案例、期刊論文但判決內容或作者結論卻完全錯誤或被嚴重扭曲錯誤引用，常見於律師提交之請求或答辯狀，亦有出現在法院裁決中[26]，依據史丹佛大學人工智慧研究院的最新研究，法律 AI 模型在處理查詢時，每 6 次就有 1 次或更多出現幻覺現象，若使用通用型

23 Changqing Shi, Tania Sourdin, Bin Li (2021) The Smart Court – A New Pathway to Justice in China? International Journal for Court Administration 12(1), Art 4, pp.1-29.

24 Zhiyuan Guo, Jiajia Yang (2025) The Application of Artificial Intelligence in China's Criminal Justice System, Legal Issues in the Digital Age. Vol. 6. No. 1., at pp.88-89.

25 OECD. (2025). Governing with Artificial Intelligence: The State of Play and Way Forward in Core Government Functions. OECD Publishing, https://doi.org/10.1787/795de142-en, at p.266.

26 Aaron Mak, Judges say staff's use of AI led to errors, PliticoPro, 2025/10/23, https://subscriber.politicopro.com/article/2025/10/judges-say-staffs-use-of-ai-led-to-errors-00620307, last visited on 2025/10/25.

聊天機器人在法律查詢中的幻覺率高達 58% 至 82%[27]，根據學者 Damien Charlotin 維護的 AI 幻覺案例數據庫統計[28]，全球有 541 件，美國已有 366 件案例，而隨著 AI 盛行在 2025 年案例數爆發性成長約 6.5 倍；最常犯錯主體是一般訴訟當事人，其次是律師；而文件犯錯類型為虛構判決或來自現有案例的虛假引文，其次是錯誤表述的判例法或法律規範[29]。 AI 法律幻想之產生係因 AI 生成是預測接下來最看似合理的單詞，但缺乏驗證事實準確性的功能，在法律領域更容易因下列因素發生：

1. 檢索錯誤

檢索增強生成（Retrieval-Augmented Generation, RAG）需先從龐大的法律資料庫中檢索相關文本，但法律問題往往沒有單一明確答案，由於判例間存在時效與層級差異，若檢索到語義相似但法律上無關或過時的判決，模型仍可能引用，導致似是而非的錯誤。

2. 生成錯誤

即使檢索到正確文件，模型在生成回覆時仍可能誤解判決見解或層級關係（如將地院見解誤認為最高法院見解）；混淆角色（把辯護人主張誤寫成法院見解）；過度歸納或推論未載明的規則。這類幻覺較危險，因為形式上引用真實案例，但結論錯誤，難以被使用者察覺。

3. 資料與語境錯配

RAG 系統雖連結真實資料，但若引用的來源與模型生成內容語意無關或誤解上下文，即為根據錯誤基礎之幻覺。例如：模型引用真實判決來支持不相干的主張，表面

27 Faiz Surani、Daniel E. Ho et al., AI on Trial: Legal Models Hallucinate in 1 out of 6 (or More) Benchmarking Queries, Stanford University: Human-Centered Artificial Intelligence (HAI), 2024/5/23, https://hai.stanford.edu/news/ai-trial-legal-models-hallucinate-1-out-6-or-more-benchmarking-queries, last visited on 2025/10/25.

28 Damien Charlotin, AI Hallucination Cases, https://www.damiencharlotin.com/hallucinations/, last visited on 2025/11/14，統計至 2025 年 11 月 10 日。

29 Damien Charlotin 統計 2023 至 2024 年美國僅 43 件，2025 年已有 323 件；前揭註網頁亦有針對犯錯主體、犯錯類型作出統計。

無幻覺，實際上是錯誤對應[30]。

縱使明白出錯原因，上開研究發現即使是客製化的法律人工智慧工具仍然有 17% 至 33% 左右出現幻覺[31]。因此，AI 在法院運用如何避免幻想成為第一要務。

（二）AI 演算法偏差（algorithmic bias）生成風險評估或量刑失準

所謂演算法偏差指的是人工智慧系統在做決策或預測時，其結果系統性地、不公平地偏向或損害了某個特定群體（例如基於種族、性別、年齡、地區等）。而這個偏差不是程式的錯誤（Bug），而是它忠實地反映了某些不公平的模式所導致的結果。而造成此一偏差的成因可能有如下因素：

1. 標籤偏差（Label Bias）

指 AI 訓練資料中的結果變數（label）有系統性偏差。例如以「是否被再次逮捕」作為再犯指標，而非「是否再次犯罪」。若警察對黑人社區更頻繁巡邏，即使犯罪行為相同，黑人被「再次逮捕」的機率更高。因此模型表面上準確，但實際上只是學習了種族化的逮捕模式，而非犯罪風險。

2. 特徵偏差（Feature Bias）

指輸入變數本身有種族偏差，例如「前科次數」、「先前逮捕紀錄」。若這些特徵在黑人族群中更容易被累積（非因犯罪而是因執法密度高），演算法便學到這種差異，結果使黑人被評為再犯高風險。

3. 樣本偏差（Sample Bias）

若 AI 風險評估訓練資料只包含「被逮捕過的人」，就排除了未被逮捕但有犯罪行為者。由於逮捕本身受種族偏見影響，樣本即不具代表性。因此演算法對不同族群的準確度會顯著不同[32]。

---

30 Magesh, V., Surani, F., Dahl, M., Suzgun, M., Manning, C. D., & Ho, D. E. (2025). Hallucination-free? Assessing the reliability of leading AI legal research tools. Journal of Empirical Legal Studies, 22(2), at. pp. 223-225.

31 Varun Magesh、Faiz Surani et al. at p. 216.

32 Neil, R., & Zanger-Tishler, M. (2025). Algorithmic bias in criminal risk assessment: The consequences of racial differences in arrest as a measure of crime. Annual Review of Criminology, 8, pp.97–119, https://doi.

如前所述，COMPAS 在美國是司法系統用以預測被告再犯風險的工具，早於 2016 年即遭披露對黑人不公平，黑人被錯誤標示為「高再犯風險」的比例幾乎是白人的兩倍 [33]。著名案例 State v. Loomis[34]，被告 Eric L. Loomis 因涉及駕車掃射事件（drive-by shooting），被控多項重罪。 Loomis 之後與檢方達成認罪協議，就部分罪名認罪，法院於量刑時採用了 COMPAS 風險評估工具，依據被告的年齡、性別、犯罪前科、社交背景等多項因素，預測其再犯風險分數。該分數被列入量刑報告中，顯示 Loomis 具有「高再犯風險」。法官於量刑理由中援引該分數，判處被告六年監禁。 Loomis 不服，主張法院依賴 COMPAS 風險評估結果量刑，違反正當法律程序權（Due Process），認該系統演算法不透明，且有性別及種族偏見，威斯康辛州最高法院認為一審法院在量刑時得將 COMPAS 風險評估作為參考，而非得以該評估結果作為唯一或決定性依據（rely on），並要求法院仍須就其他事證與量刑因素加以說明，是以，並未違反被告之正當法律程序權 [35]。

（三）AI 生成結果產生錨定效應（Anchoring Effect）

所謂錨定效應是指：決策者在被要求估計某個數值時，傾向於將其估計值建立在他們遇到「第一個」或「最顯著」的數值（錨點）上的，例如：檢察官的求刑數字，受錨定效應影響會成為心理錨，導致判斷偏移，從學者研究發現法律領域中的數值決策（例如判處的刑期或賠償金額）確實容易受到決策情境中顯著數字影響。這種錨定效應可能帶來不良後果，使法院的裁決產生偏誤或不穩定 [36]。 AI 若以大量案例資料為基礎運算，理論上能降低此類認知偏誤與情緒干擾；但若訓練資料本身有偏差，也可能放大錨定偏誤，或者不瞭解 AI 演算過程而盲從。因此研究者認為，AI 生成之建議

---

org/10.1146/annurev-criminol-022422-125019.

33 Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016, May 23). Machine bias: There's software used across the country to predict future criminals. And it's biased against Blacks. ProPublica, https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing, last visited on 2025/11/14.

34 State v. Loomis, 881 N.W.2d 749 (Wis. 2016).

35 State v. Loomis, 881 N.W.2d 749, 764–65 (Wis. 2016) (Roggensack, C.J., concurring)；另參見 Harvard Law Review, Note, State v. Loomis, 130 Harv. L. Rev. 1530, 1532–1534 (2017).

36 Bystranowski, P., Janik, B., Próchnicki, M., & Skórska, P. (2021). Anchoring effect in legal decision-making: A meta-analysis. Law and Human Behavior, 45(1), .1–23, https://doi.org/10.1037/lhb0000438.

或預測結果同時具有減少與再製錨定效應的雙面性[37]。在採取一般國民之陪審團、裁判員或國民法官制度下，就 AI 預測結果產生判斷上的影響與職業法官之差別，有待時間發展累積研究方有所得。

（四）AI 生成或處理之數位證據真實性與可靠性

AI 生成或處理的證據（AI-generated or processed evidence），通常泛指透過人工智慧系統（特別是具備機器學習過程）所產出或處理的資料[38]，諸如語音增大或影像辨識圖像增強、專家意見分析整理、乃至統計預測，作用是增強或澄清現有內容，這些資料可能被用於證明刑事或民事案件中的待證事實。然而，利用 AI 技術產生逼真但完全捏造的內容，用於虛假地牽連個人犯罪、製造虛假供詞、篡改歷史記錄或捏造新聞事件，所謂深度偽造（deepfake）的語音影片及圖像[39]，被當事人主張作為證據進入法庭時，由於 AI 系統產生內容係不透明的過程，甚至專家也難以驗證特定輸出是如何產生的。具備機器學習過程的 AI 技術，其內部的獨立運算過程是未知的猶如黑盒子（Black Box），缺乏可解釋性，使得 AI 生成證據關於真實性（authenticity）與可靠性（reliability），究竟是深偽或只是強化證據效果，將使法院在審查其輸出結果時面臨困難[40]。

2024 年 3 月 29 日美國華盛頓州金郡高等法院[41]審理被告 Puloka 被控謀殺案件中，被告抗辯係自我防衛，辯護人提供一段由證人用 iPhone 拍攝的影片作為證據，但有解析度低，且包含嚴重的「動態模糊」之問題，於是提交了使用「Topaz Labs AI」程式中的 AI 影片編輯工具增加了清晰度，並進一步在 Adobe Premier Pro 中處理過之增強版

37 Watamura, E., Liu, Y., & Ioku, T. (2025). Judges versus artificial intelligence in juror decision-making in criminal trials: Evidence from two pre-registered experiments. PLOS ONE, 20(1), e0318486, p2, https://doi.org/10.1371/journal.pone.0318486.

38 曾小榕、金孟華（2025），AI 在刑事證據法之研究：以美國聯邦證據規則為核心，歐美研究，第 55 卷第 1 期，頁 3。

39 Rebecca A. Delfino, Deepfakes on Trial 2.0: A Revised Proposal for a New Federal Rule of Evidence to Mitigate Deepfake Deceptions in Court, Loyola L. Sch., L.A., Legal Studies Research Paper No. 2025-10 (2025), https://ssrn.com/abstract=5188767, last visited on 2025/11/14.

40 Gesser, A., Pastore, J., Kelly, M., Dorward, J., & Kohan, G. A. (2025, March 20). Federal Judicial Conference to revise rules of evidence to address AI risks. Debevoise & Plimpton LLP. https://www.debevoise.com/insights/publications/2025/03/federal-judicial-conference-to-revise-rules-of, last visited on 2025/11/15.

41 Washington v. Puloka, No. 21-1-04851-2 (Super. Ct. King Co. Wash. March 29, 2024).

本，使物體增加了銳利度、清晰度和更平滑的邊緣，而來源影片則包含帶有「塊狀」邊緣圖案的模糊影像。檢方的專家證人 Grant Fredericks 是一名經過認證的司法鑑識影片分析師，認為執行司法鑑識影片分析時的重點是影像完整性，而不是為用戶創造更流暢、更具吸引力的產品，本案使用的 AI 工具，與原始影像中的像素數量相比，增強每個影格大約 16 倍的像素，並且使用了任何司法鑑識影片專家都未知曉或審查過的演算法和增強方法。此外，證明了該 AI 方法創造了虛假的影像細節，且該過程不為司法鑑識影片專家社群所接受，因為它會改變影片部分內容的含義，沒有保持來源影片中的物體原始形狀和顏色。一般公認的影像放大技術所創建的影片產品，可以在一般公認的影片處理程式中重現。而 Topaz Labs AI 模型則使用一種稱為機器學習的不透明過程來放大和增強影片。

法院認為在刑事審判中引入使用 AI 工具增強的影片是一種新穎技術，使用這些 AI 工具的證據主張者必須證明，任何專家的意見或理論都是基於相關群體所接受的方法論。採納使用新穎科學理論或原理的證據是以弗萊標準（Frye test）[42] 來衡量，以是否已在相關科學群體中獲得普遍接受為準。為了確定是否已達成科學意見的共識，法院會審查主張普遍接受的證詞、文章和出版物、在該群體中的廣泛使用。本案用以判定 AI 增強影片可採性的相關科學社群是「司法鑑識影片分析群體」，Topaz Labs AI 增強工具使用機器學習演算法，尚未經過同行評審，目前無法被該社群重現，也未在該社群中被普遍接受。辯方未提供任何州或聯邦上訴法院的判決，無論來自哪個司法管轄區曾審查過，更不用說批准過在刑事或民事審判中引入 AI 增強影片。辯方未提供任何文章、出版物或其他法律來源，有批准在刑事或民事審判中引入 AI 增強影片之證據。

Topaz Labs AI 增強模型產生的影片證據不符合關聯性要件（美國聯邦證據規則 Federal rule of evidence, FRE 401），因為其產生的影片並未完整地顯示實際發生的情況，而是使用不透明的方法來呈現 AI 模型「認為」應該顯示的內容。此外，採納此 AI 增強證據將導致問題混淆和目擊者證詞混亂，並可能導致一場關於 AI 模型所使用的、無法經同行評審的過程的「審判中的審判」（Trial within a Trial），因此其任何相關性，均被 FRE 第 403 條所述的不公平偏見的危險大幅超越。不公平偏見的危險在本案中尤

42 Frye v. United States, 293 F.1013, 1014 (D.C. Cir. 1923).

其令人擔憂，因為陪審團可能會專注於 AI 增強影片中所描繪事件期間個人和物體的大小與位置。因此，本案最終排除 AI 生成增強影片之證據。

除上開案件外，陸續發生美國法院 Matter of Weber as Trustee of the Michael S. Weber Trust 案[43]，紐約州遺囑法庭針對信託與財務爭議案件。當事人之一主張損害賠償專家提交財務計算分析，用於估算信託或財務損害，產生爭議的核心事實專家在進行財務計算時使用 Microsoft Copilot（大型語言模型生成式 AI）進行數學與財務推算，但無法向法院說明 Copilot 的資料來源、如何得出結果、模型邏輯與推算原理。法官進行「實驗式檢驗」輸入相同問題，Copilot 每次生成不同結果。法官詢問 Copilot 自身的可靠性時，Copilot 回答：「應由人類專家再次檢查」，法院認為無法達到專家證據所需的可靠性。法院拒絕採信該 AI 輔助之財務計算（證據被排除）。

（五）AI 卷證分析之個資保護

AI 文字識別及檢索整理生成文字能力增強後，法院內部人（含法官在內）對於使用 AI 文書處理能力舉凡案情摘要（summary）、時序整理（timeline）、供述內容比對（compare statements）、找出矛盾與爭點（conflict identification），最後甚至生成判決草稿，以節省人力時間，抱以高度期待。然而，法院卷證內含個資包括原本法定閱卷時即應對外保密之性侵被害人、未成年人資訊外，也包括個人資料保護法中一般個資：如當事人（原告、被告）、告訴人、被害人及訴訟參與人、證人之出生年月日、國民身分證統一編號、護照號碼、特徵、指紋、婚姻、家庭、教育、職業等，刑事警詢筆錄上基本欄位，被告犯罪前科也是法院刑事卷證必備資料，甚至特種個資中病歷、醫療、性生活、健康檢查也是法院卷宗內常見資料之一。而法院卷證原本蒐集之目的係為審判，然而將卷證內含個資資料傳輸至外部伺服器（雲端）落入第三方（OpenAI、Google、Microsoft）可能作為訓練大型語言模型之素材，構成另一次未經個資當事人同意之利用。並且於 2025 年 7 月、8 月 ChatGPT、Grok 陸續被傳出用戶與 AI 之對話在 google 搜尋可以完整查看外[44]，早從 2023 年 3 月韓國三星電子（Samsung Electronics）

43 Matter of Weber as Trustee of the Michael S. Weber Trust, 220 N.Y.S.3d 620 (N.Y. Sur. Ct. 2024).

44 Liv McMahon, Hundreds of thousands of Grok chats exposed in Google results, BBC, 2025/08/21, https://www.bbc.com/news/articles/cdrkmk00jy0o, last visited on 2025/11/15.

半導體部門授權部分員工使用 ChatGPT 協助工作，隨即發生了三起資料外洩事件：

1. 員工將半導體設備量測相關的機密原始碼（Source Code）複製到 ChatGPT 要求除錯。
2. 程式碼優化：員工上傳了與良率計算或設備識別相關的程式碼，請求 AI 進行優化。
3. 會議記錄摘要：員工將內部會議錄音轉成的文字檔全文上傳，要求 ChatGPT 製作會議紀要。

這些資料被傳送至 OpenAI 的伺服器，三星公司意識到這些資料無法被撤回，且可能成為 AI 模型訓練的一部分，進而有洩漏給其他使用者的風險，並最終決定在公司內部網路與設備上全面禁止使用生成式 AI 工具 [45]。

故從上開分析，欲利用公開商用之生成式 AI 應用程序，不論 AI 公司對外標榜隱私權之保護 [46] 或上傳資料不做為模型之訓練 [47]，直接將法院卷證上傳 AI 分析，顯不可行，因此如何在追求審判效益與個資保護取得平衡即成難題。

參、AI 在各國法院運用之規範

本文主要介紹美國、歐洲、英美法系其他國家，以及亞洲與南美等 15 個司法管轄區之法院 AI 使用規範。為避免僅是逐國羅列規定，難以看出制度差異與發展方向，本文採取橫向比較的方式，從對律師及外部使用者之規範、對法院內部人員與機密資訊之規範、對 AI 生成或處理之證據的證據法則及 AI 系統本身之風險分級與倫理框架等四個面向，觀察各國法院規範的設計。

一、美國法院運用 AI 規範與實務趨勢

美國法院體系在應對 AI 使用方面，採取了個別分散化和即時反應的策略，其特徵

45 Mark Gurman (2023), Samsung Bans Staff's AI Use After Spotting ChatGPT Data Leak, Bloomberg, 05/02/2023, https://www.bloomberg.com/news/articles/2023-05-02/samsung-bans-ChatGPT-and-other-generative-ai-use-by-staff-after-leak?embedded-checkout=true, last visited on 2025/11/15.

46 以 ChatGPT 或 Gemini 為例，均通過美國註冊會計師協會（American Institute of Certified Public Accountants, AICPA）認證符合安全性、可用性、機密性、處理完整性及隱私（確保所有資料處理均依法規收集、保留、揭露和處置）SOC2 五大關鍵原則，並且合規達 12 個月屬 Type2 等級。

47 例如 ChatGPT 企業版或商業版於使用約定條款中宣傳不會使用包括輸入或輸出內容來訓練或強化模型。

是聯邦規則、聯邦法官的個別命令與州律師公會的道德指引併行。

(一) 律師使用 AI 提出訴狀產生虛構判例事件及責任

如前所述生成式 AI 最大風險是產生幻覺，即編造出看似合理但事實虛假或不存在的資訊，特別是虛構法律判例和引用。

1. 指標性虛構判例事件 ── Mata v. Avianca, Inc. 案 [48]

本案被視為法律界 AI 倫理的里程碑事件。原告律師 Steven A. Schwartz 在人身傷害損害賠償之民事訴訟中，使用 ChatGPT 進行法律研究，提交了一份引用了至少六個完全虛構判例的法律文件以反駁被告答辯，對造律師找不到相關判決後向法院反映原告律師似乎捏造判決。 Schwartz 提出宣誓書並附上判決的 PDF 檔，當法院檢視了這些所謂的判決後，注意到其外觀、格式、內容雖模擬了法院判決的寫作方式，但存在許多與真正法院文件不符的地方，包括語句、引用方式、邏輯連貫性等。而於 2023 年 4 月 11 日及 4 月 12 日發出命令，要求原告律師提供進一步資訊，並解釋這些無法查證的案例來源，Schwartz 提出聲明中承認使用了 ChatGPT 來尋找判決，而 ChatGPT 產生了這些「案例」，其不知道 ChatGPT 會生成虛構內容，有請 ChatGPT 再確認判決存在，但未自己確認即直接引用在法律文件上等語。原告律師 Schwartz 及代理律師 LoDuca 在被告通知案例不存在後，仍未立即撤回文件，亦未向法院主動說明問題，而是在法院多次下令後才提供完整說明。紐約南區聯邦地方法院的法官 P. Kevin Castel 認為律師在向法院提交文件時，必須在合理查證後確認主張、抗辯及法律見解是有法律依據的，或至少是非濫用性的、具誠信的法律主張。本案原告律師並非是單純疏失，而是刻意避免發現真相（conscious avoidance），違反了聯邦民事訴訟規則（Federal Rule of Civil Procedure, FRCP）第 11 條（Rule 11）的合理查證要求。 Rule 11(c)(4) 允許多種制裁方式，包括：非金錢命令、罰金、研習課程及轉介紀律委員會。法院最終對 Schwartz 處以 5,000 美元罰款，並要求他們向客戶和被虛假引用的法官發出道歉信。此案確立了核心原則：律師對其提交文件的準確性負有不可推卸的責任。

48 Mata v. Avianca inc. 678 F. Supp. 3d 443 (s.d.n.y. 2023).

2. Park v. Kim 案 [49]

聯邦第二巡迴上訴法院在處理車禍侵權案件中，原告律師 Jae S. Lee 在上訴狀中引用了由 ChatGPT 生成的虛假案例。法院認為律師有義務閱讀並驗證她所引用的判決是否存在，因此認為違反 Rule 11 規定，將該律師轉介至其紀律委員會（Grievance Panel）進行懲戒，並明確指出聯邦程序規則已經提供足以處理生成式 AI 使用所引發疑慮的完整框架，律師不能將 Rule 11 責任轉嫁給 AI 技術 [50]。

3. Mid Central Operating Engineers Health and Welfare Fund v. HoosierVac, LLC 案 [51]

在請求雇主給付案件中，被告律師 Rafael Ramirez 承認他依賴生成式 AI 工具，結果產生了虛構的案件引用（fictitious case citations），在三個不同的場合向法院和對造律師提交了包含「不存在案件」（non-existent cases），法院駁回了律師對 AI 產生幻覺一無所知的辯解，以違反 Rule 11 規定判處 6,000 美元罰款。

此外，在科羅拉多州，一名律師也因使用 ChatGPT 提供的虛假判例而被停職案例 [52]。法院一致認為，律師使用了 AI 絕不能成為提交不準確、誤導性或完全虛假資訊的藉口。律師必須對 AI 生成的所有內容進行獨立核實。

（二）AI 生成證據之證據法則

1. 聯邦證據規則

美國聯邦司法會議證據規則諮詢委員會（Advisory Committee on Evidence Rules）積極應對 AI 證據帶來的挑戰，2024 年 11 月同意制定正式提案，並於 2025 年 5 月 2 日以 8 比 1 投票通過 [53]，隨後於 2025 年 8 月釋出進行公眾意見徵詢（Public Comment），預計持續至 2026 年 2 月 16 日，同年 5 月 7 日委員會進行表決。委員會提出了證據規則修訂方向：

---

49 Park v. Kim, No. 22-2057 (2d Cir. Jan. 30, 2024).

50 判決原文 “The responsibilities imposed by Rule 11 cannot be outsourced to A.I.”.

51 Mid Central Operating Engineers Health and Welfare Fund v. HoosierVac, LLC. No. 2:24-cv-00326-JPH-MJD, 2025 WL 1511211 (S.D. Ind. May 28, 2025).

52 People v. Zachariah C. Crabill, No. 23PDJ067 (Colo. Presiding Disciplinary Judge Nov. 22, 2023).

53 陳郁潔，美國法院擬修正《聯邦證據規則》以規範人工智慧生成內容之證據能力，資訊工業策進會科技法律研究所，https://stli.iii.org.tw/article-detail.aspx?no=64&tp=1&d=9399, last visited on 2025/11/15.

⑴ 增訂聯邦證據規則（Federal Rules of Evidence, FRE）第 707 條（機器生成證據）

正式條文內容：「當機器生成證據（Machine-Generated Evidence），在沒有專家證人的情況下被提出，且如果由證人作證將受第 702 條規則約束時，法院只有在該證據符合第 702 條 (a)-(d) 款的要件時才可採納該證據。本規則不適用於簡單科學儀器的輸出。」。

本條增訂意謂 AI 輸出之證據內容，如果它是出自人類之口，該人類必須是個專家，因此必須符合專家證詞的標準，即它必須具有協助事實認定、基於充分的事實或數據、是可靠原則和方法的產物以及反映原則和方法在事實上的可靠應用。如果是簡單科學工具輸出，例如溫度計、秤和其他常用設備，則不受該規則的約束。根據委員會的註釋說明，FRE 第 707 條規則的目的是防止機器生成證據的提出者，透過直接提供機器輸出來規避 FRE 第 702 條規則的可靠性要求，而在這些 AI 輸出若由人類專家作為意見提出時，本應受 FRE 第 702 條規定約束的情況下。因此 FRE 第 707 條規定旨在將第 702 條專家證人之標準延伸到 AI 生成輸出的情境中。根據擬議的 FRE 第 707 條規定，法院分析的重點將是 AI 輸入提示詞及其他資訊的充分性、AI 平台的內部流程以及所得輸出的有效性，並要求證據開示（Discovery）[54]。

⑵ 修訂 FRE 第 901 條以解決 Deepfake 問題

委員會於 2024 年 4 月因應 Deepfake 審議 Delfino 教授 [55] 與 Grimm、Grossman 教授 [56] 的版本關於 Rule 901 的修正提案。二版本都主張將 Deepfake 證據真實性的最終決定權從陪審團重新分配給法官，由法官根據 FRE 第 104 條 (a) 決定其可採性，另外都要求證據提出的一方提高證明為真實的負擔。這是為了緩解陪審團容易被 Deepfake 欺騙的風險，且實證研究支持法官比陪審團更適合處理 Deepfake 偵測。最後委員會並未

54 Committee on Rules of Practice and Procedure of the Judicial Conference of the United States. (2025, August). Preliminary draft of proposed amendments to the Federal Rules of Appellate, Bankruptcy, Civil, and Criminal Procedure, and the Federal Rules of Evidence. Administrative Office of the U.S. Courts, https://www.uscourts.gov/rules-policies/proposed-amendments-published-public-comment, last visited on 2025/11/15.

55 Delfino, R. A. (2023). Deepfakes on trial: A Call To Expand the Trial Judge's Gatekeeping Role To Protect Legal Proceedings from Technological Fakery, Hastings Law Journal, 74(2), 293-348.

56 Grimm, P. W., Grossman, M. R., & Cormack, G. V. (2021). Artificial intelligence as evidence. Northwestern Journal of Technology and Intellectual Property, 19(1), 9-106, https://scholarlycommons.law.northwestern.edu/njtip/vol19/iss1/2.

通過修正提案而留待後續觀察[57]。

2. 州層級的因應措施

除了聯邦層級的修訂外，一些州也開始制定法規來應對 AI 證據：

⑴ 路易斯安那州：於 2025 年 8 月 1 日生效的 Act No. 250 要求律師在提供證據前，須「合理盡責地驗證證據的真實性」。如果律師知曉或應知曉證據是虛假或經人工操縱，但未披露此事實，將可能面臨藐視法庭的處罰[58]。

⑵ 紐約州：2023 年 11 月提出了修正《刑事訴訟法》和《民事訴訟法和規則》的法案（A 8110），要求人工智慧生成之證據，除非有獨立且可採納的證據實質支持，且該證據的提出者證明人工智慧在創造或處理證據時的特定使用是可靠性與準確性的，否則將被禁止[59]。

（三）關於律師使用 AI 聯邦法官之個別命令（Standing Orders）模式

由於美國缺乏統一的國家法律來規範在法院 AI 之運用，聯邦法院的應對方式主要是由個別法官發布常設命令（Standing Orders），針對其法庭上的案件設定個別要求，有下列情形：

1. 強制揭露核實模式

這是最普遍採用的模式[60]。例如，德州北區聯邦地方法院的 Brantley Starr 法官發布了最具代表性的命令，要求律師和自訴人必須提交證明，聲明文件未經 AI 起草，或所

57 Advisory Committee on Evidence Rules. (2024, November). Minutes of the meeting of April 19, 2024. In Advisory Committee on Evidence Rules November 2024 agenda book (pp. 15–35). Administrative Office of the U.S. Courts, https://www.uscourts.gov/sites/default/files/2024-11_evidence_rules_committee_meeting_agenda_book_final_10-24.pdf.

58 Quinn Emanuel Urquhart & Sullivan, LLP. (2025, November 6). Adapting the rules of evidence for the age of AI. https://www.quinnemanuel.com/the-firm/publications/adapting-the-rules-of-evidence-for-the-age-of-ai/, last visited on 2025/11/15.

59 Law Journal Editorial Board, Address the Issue of AI-Generated Evidence, 2023/11/12, https://www.law.com/njlawjournal/2023/11/12/address-the-issue-of-ai-generated-evidence/, last visited on 2025/11/15.

60 de Seingalt, B. (2025, Summer). The buck stops with you: Professional responsibility in the age of generative AI. The Federal Lawyer, 72(2), at p.34-35.

有 AI 生成的文字和引用已由人類使用傳統法律資料庫進行準確性檢查[61]。此命令的理由是這些平台容易產生幻覺和偏見，且對法律沒有忠誠。夏威夷聯邦地方法院也要求當事人披露所使用的 AI 工具，並證明 AI 生成部分的準確性已由律師或自訴人核實[62]。伊利諾州北區聯邦地院的 Gabriel Fuentes 法官也要求披露 AI 的使用情況及使用方式，並重申單純依賴 AI 不滿足 Rule 11 的合理查詢要求[63]。

2. 機密資料禁止輸入模式

這類命令特別關注 AI 系統可能導致客戶機密資訊洩露的風險。美國國際貿易法院的三位法官（Stephen Vaden、Jane Restani 和 Claire Kelly）要求律師除了披露 AI 使用外，還必須證明使用該程序未導致任何機密或商業專有資訊洩露給未經授權的第三方[64]。加州司法委員會所提出指引亦明定，法院職員不應將機密、個人識別或其他非公開資訊輸入到公共生成式 AI 系統中[65]。

3. 嚴格禁止模式

雖然相對少見，但有些法官完全禁止在提交給法庭的文件中使用 AI。例如，俄亥俄州南區聯邦地院的 Michael Newman 法官和北區聯邦地院的 Christopher Boyko 法官均禁止使用 AI 在任何準備提交給法院的文件上。北卡羅來納州西區聯邦地院也發布了命令，禁止使用 AI 平台準備訴狀或備忘錄[66]。

4. 小結：差異化與法院自治現象

由於缺乏統一規範，律師在不同法庭可能面臨截然不同的合規要求。截至 2025 年 4 月，有超過 62 位聯邦地方法官發布了與 AI 使用相關的命令或規則。雖然大多數命令

61 Grossman, M. R., Grimm, P. W., & Brown, D. G. (2023). Is disclosure and certification of the use of generative AI really necessary? Judicature, 107(2), at p.70.

62 Thomson Reuters. (2025.July). 50 state statutory surveys: Artificial intelligence (AI): Use of artificial intelligence in the practice of law - State. Westlaw, last visited on 2025/11/16.

63 Bucher, A. (2025). Navigating the power of artificial intelligence in the legal field. Houston Law Review, 62(4), at pp.838.

64 de Seingalt, B. supra 60, at p.35.

65 Judicial Council of California. (2025, June 16). Report to the Judicial Council: Judicial branch administration: Rule and standard for use of generative artificial intelligence in court-related work (Item No. 25-109).

66 de Seingalt, B. supra 60, at p.36.

都要求披露和核實[67]，但具體細節（如披露的範圍、是否涵蓋非生成式 AI）存在顯著差異，這增加了跨轄區執業的複雜性。

（四）州法院關於法院內部人員使用 AI 之規範

相較於聯邦法官的個別命令，一些州級最高法院或司法委員會正在推動適用於全州法院系統內部使用 AI 的統一政策。例如：

1. 猶他州 Utah（保密最嚴格）

猶他州司法委員會（2023 年 10 月）針對法官和法院員工頒布臨時規則，其核心在於資訊保密。規則嚴格禁止將任何可能識別案件或個人的機密資訊輸入生成式 AI 工具中。此外，規則允許 AI 用於法律研究和起草行政命令或初稿，但不得僅依賴 AI 內容，且法官必須對其進行徹底審核[68]。

2. 紐澤西州 New Jersey（禁止 AI 取代法官裁量）

紐澤西州最高法院的原則聲明（2024 年 1 月）強調 AI 僅為輔助工具，不得取代法官的自主決策功能，但可作為支持和增強司法職能的工具。該聲明針對法官及其工作人員僅可將 AI 用於特定目的，將 AI 的適當使用範圍限定於初步資料收集、文件整理或總結等任務，目的在於維護司法獨立性和公正性[69]。

3. 阿肯色州 Arkansas（禁止將法院資料輸入公共 AI）

阿肯色州最高法院的第 25 號行政命令（Order 25）提醒所有參與法院系統的人員（包括公眾、律師、法院工作人員），當將客戶或法院資料輸入生成式 AI 系統時，需注意這些工具可能會保留數據用於訓練其 LLM。輸入機密或封存資料前應確認系統是否會保留數據，且不當揭露可能違反相關法院規則（如 Administrative Order No.19）、法規及職業行為準則。明確禁止法院行政辦公室員工、各級法院書記官及其員工，以及任何可存取內部 CourtConnect 系統的人員，故意將州法院的內部數據暴露給生成式

67 de Seingalt, B. supra 60, at p.34.

68 Browning, J. G. (2025). The dawn of the “AI Judge”? Generative artificial intelligence and its impact on appellate courts. The Journal of Appellate Practice and Process, 25(2), at pp.355-356.

69 Id., p.358.

AI。受禁止規範的人員若為法院利益進行 AI 相關研究，可請求最高法院自動化委員會批准。法院資訊系統部門則被允許為法院利益進行相關研究 [70]。

4. 伊利諾州 Illinois（強調透明性但不強制揭露）

伊利諾州最高法院的政策（2025 年 1 月生效）採取了相對寬鬆的立場，鼓勵並不阻止在遵守法律和道德標準下使用 AI。該政策要求法官對最終決定負責，並禁止 AI 應用程式危及敏感資訊（如機密通信、個人識別資訊 personally identifiable information, PII）。值得注意的是，伊利諾州政策不要求律師在法庭文件中強制披露 AI 的使用情況 [71]。

5. 加州 California（禁止將個資輸入公共 AI）

加州司法委員會（Judicial Council of California）。2025 年 6 月 16 日發布法院使用 AI 指引，於 2025 年 9 月 1 日生效，規範法律從業人員及司法人員（包括法官），禁止行為包括不得將機密、個人識別資訊（PII）或非公開資訊輸入至「公共生成式 AI 系統」（即資料可被外部存取或用於訓練模型的系統）、不得使用 AI 對受保護群體（如種族、性別、身心障礙等）造成非法歧視或差別影響。若提供給公眾的書面、視覺或音訊作品的最終版本完全由生成式 AI 產出，必須進行揭露，但法官在審判工作中保有是否揭露的裁量權 [72]。

（五）律師公會關於律師使用 AI 之倫理規範

1. 美國律師協會（American Bar Association, ABA）意見書

律師公會的倫理指引，特別是意見書，為律師在 AI 時代的執業提供了權威性的框架。 ABA 於 2024 年 7 月發布了第 512 號正式意見書，為全國各州的倫理指引提供了藍圖。主要有下列四項核心：

---

70 In re Creation of Arkansas Supreme Court Administrative Order No. 25, 2025 Ark. 117 (2025).

71 Browning, J. G.supra 68. at pp.368-369.

72 Judicial Council of California. (2025). Report to the Judicial Council: Judicial branch administration: Rule and standard for use of generative artificial intelligence in court-related work (Report No. 25-109). https://business.cch.com/srd/20250722-25-109-20250718-25-109-updated20250616.pdf, last visited on 2025/11/17.

⑴ 能力義務（Competence, Rule 1.1）：律師有責任理解與其執業相關的新技術（包括 AI）的利益與風險。律師必須對 AI 的功能、限制（如幻覺）和數據處理方式有基本的了解。

⑵ 保密義務（Confidentiality, Rule 1.6）：這是最受關注的義務之一。律師在使用 AI 工具處理客戶資訊前，必須採取合理措施保護客戶機密，不能將機密資訊輸入到缺乏適當安全和保密機制的公共 AI 模型中。

⑶ 監督義務（Supervision, Rules 5.1 & 5.3）：主管律師有責任制定明確的 AI 使用政策，並對下屬律師及非律師助理進行充分的培訓和監督，以確保他們的使用符合道德規範。

⑷ 對法庭的坦誠（Candor toward tribunals, Model Rule 3.3）：律師對法庭負有坦誠義務。當律師使用 AI 進行法律研究或起草文件時，最大的風險是 AI 生成的錯誤或虛假內容，ABA 意見書明確指出，律師有義務確保提交給法庭的所有內容都是準確且有事實或法律依據的。如果律師未能對 AI 產生的法律引文進行核實，並因此向法庭提交了不存在的案例，這將構成對法庭坦誠義務的嚴重違反。

此外，因為律師使用 AI 提高了效率，這對律師如何向客戶收費產生了影響。律師不能以人工方式所需的時間來向客戶計費。收費必須反映 AI 帶來的效率收益[73]。律師必須以透明的方式向客戶說明其費用的計算方式，適當地「納入效率收益」（account for efficiency gains），以確保費用在 AI 時代保持合理[74]。

2. Florida、West Virginia 等州的差異化要求：

⑴ 佛羅里達州（Florida）：佛羅里達州律師協會的倫理諮詢意見似乎鼓勵使用人工智慧。該意見指出：「律師可以在法律實務中使用生成式人工智慧，但必須保護客戶資訊的機密性，提供準確且稱職的服務，避免不當的收費行為，並遵守適用的律師廣告限制。」，雖然佛羅里達州的做法並未要求披露人工智慧的使用情況，但它仍然強調律師履行其倫理義務，包括讓客戶了解這項技術的使用情況並保護客戶的利

73 Fees and Expenses, Rule 1.5.

74 de Seingalt, B. supra 60, at p.37.

益[75]。

⑵ 西維吉尼亞州（West Virginia）：西維吉尼亞州律師倫理意見要求律師在使用生成式 AI 前，應取得客戶的知情同意，且須書面確認。該意見還提出拒絕使用能提高效率和準確性的技術，可能被視為未能提供稱職的法律代表[76]。

二、歐洲法院運用 AI 規範與倫理原則

歐洲各國的 AI 規範通常體現了更強烈的中央統籌和倫理驅動特徵，著重於基本權利、透明度、可解釋性，並與歐盟 AI 法案（Artificial Intelligence Act，下稱 AI Act）緊密銜接。

(一) 歐盟

1. 對於法院內部與外部人員共同規範 ── 司法系統 AI 使用倫理憲章

歐洲委員會（Council of Europe）的歐洲司法效能委員會（European Commission for the Efficiency of Justice, CEPEJ）於 2018 年 12 月 4 日發布的司法系統 AI 使用倫理憲章，為 AI 使用原則性之管制框架。該憲章的規範對象非常廣泛，涵蓋了所有參與 AI 工具設計、部署和使用的公共及私人利害關係人，包括使用者如法官、檢察官、律師等法律專業人士。主要有五項原則上指示：

⑴ 基本權保障原則：憲章要求在司法系統內導入和使用 AI 時，必須保障基本人權，例如歐洲人權公約和個人數據保護公約。

⑵ 非歧視與偏誤防止原則：強調應預防 AI 系統產生歧視或偏見，確保 AI 工具的設計和使用不會對基本權利產生不利影響。

⑶ 可解釋性與透明性：要求 AI 決策過程和輸出必須具備可解釋性與透明度。

⑷ 人機協作（Human-in-the-loop, HITL）保有決策權：AI 系統僅能輔助人類在法律案件中的決策。人類應對最終決定保持控制權，並有權拒絕純粹由機器生成的決策。

---

75 Kies, J. M. (2025, January). How courts should deal with generative AI: A judge's perspective. Wisconsin Lawyer, 98, at p.46.

76 West Virginia Lawyer Disciplinary Board. (2024). AI and the rules of professional conduct: Advisory ethics opinion 24-01. West Virginia State Bar.

⑸ 品質與安全性原則：使用的司法大數據必須來自確定的法律來源。 AI 工具必須在安全的環境下運行，且處理後的模型需具備高度的準確性與安全性。憲章並要求持續評估與 AI 工具認證機制的實施可行性 [77]。

2. 歐盟人工智慧法案（AI Act）— AI 系統風險分級管控

2024 年 8 月 1 日歐盟作為全球首部針對人工智慧的監管專法生效，共 113 條及 13 個附件，規範對象為 AI 系統與通用人工智慧模型（General Purpose Artificial Intelligence Model, GPAI），雖非直接規範使用者，然而從序言雖稱 AI 系統運用有助司法效益，但更強調保護基本權包括獲得有效補救和公平審判的權利、辯護權和無罪推定，揭櫫用於司法行政的 AI 系統分類為高風險，因為它們可能對民主、法治、個人自由以及獲得有效補救和公平審判的權利產生重大影響 [78]，而於第 6 條和第 8 條用於協助研究和解釋事實和法律，以及將法律應用於具體事實的 AI 系統、或用於評估犯罪受害者風險、測謊儀、評估證據可靠性、評估累犯風險等或者用於替代性爭議解決機構（alternative dispute resolution bodies）的 AI 系統，且其結果對當事人產生法律效力，均歸屬高風險 AI 系統範疇，除非係純粹的附帶行政活動，例如司法判決、文件或資料的匿名化或假名化，或人員之間的通訊等，不影響個案實際司法行政的 AI 系統，則不屬於高風險。

AI Act 要求對高風險 AI 系統實施持續、迭代的風險管理系統 [79]，供應商必須建立和記錄上市後監控系統（post-market monitoring system），系統性地收集、記錄和分析 AI 系統在整個生命週期內的性能數據，以確保持續符合要求，並且必須在技術上允許自動記錄事件（日誌），以便追溯系統功能，並用於上市後監控。此外，供應商必須建立系統，向相關主管機關報告因使用其 AI 系統而導致的任何嚴重事件（如死亡、嚴重

77 European Commission for the Efficiency of Justice. (2018). European ethical charter on the use of artificial intelligence in judicial systems and their environment. Council of Europe.，本文完稿後 CEPEJ 於 2025 年 12 月 19 日有發布最新指引，可參見 European Commission for the Efficiency of Justice (CEPEJ). (2025). Guidelines on the use of generative artificial intelligence in courts (CEPEJ(2025)18Final). Council of Europe, https://rm.coe.int/cepej-2025-18-final-guidelines-gen-ai-courts/1680b332b4.

78 EU Artificial Intelligence Act Recital (4), (48), (61).

79 EU Artificial Intelligence Act Article 9.

健康損害、關鍵基礎設施中斷或侵犯基本權利[80]）。

深偽內容的生成者或操控者必須揭露內容為人工生成或操控；旨在向公眾提供公共利益資訊的 AI 生成文本，必須揭露其人工生成或操控性（除非經過人工審查），此為本法第 4 條所明定特別揭露義務。

本法核心原則為 AI 工具可以支持法官的決策權或司法獨立性，但不應取代它；最終的決策制定必須保持由人驅動的活動[81]。禁止僅基於個人特徵或個性評估，來評估或預測個人犯罪風險的 AI 系統；禁止在公共可及空間使用「即時遠端生物識別系統」（Real-time remote biometric identification systems），除非有嚴格的例外情況（例如針對特定嚴重犯罪嫌犯的定位或識別[82]）。

(二) 德國

德國聯邦和各州司法部長會議（Justizministerinnen und Justizminister von Bund und Ländern）於 2025 年 6 月 5 日在第六屆聯邦 ── 各州數位峰會（Bund-Länder-Digitalgipfel）針對 AI 在司法系統中應用所達成的共同聲明（Gemeinsame Erklärung zum Einsatz von Künstlicher Intelligenz (KI) in der Justiz），其副標題為「利用機會，承擔責任，確保透明度」（Chancen nutzen, Verantwortung übernehmen, Transparenz gewährleisten），說明使用 AI 態度。該共同聲明的主要內容係強調在司法系統中採用 AI 工具，必須在提升效率的同時，嚴格遵守法治國家的核心原則，並將人類的角色置於中心。

1. 核心原則與目標

聲明確認在司法領域使用 AI 具有巨大的潛力，可以顯著提高效率、優化流程，並能使尋求法律幫助的人更容易與法院和檢察機關聯繫。承諾將推動使用負責任、公正、可說明追溯（nachvollziehbar）、可靠的 AI 系統。

2. 關於人類監督與最終責任（核心要求）

80 EU Artificial Intelligence Act Article 12、72、73.

81 EU Artificial Intelligence Act Recital (61).

82 EU Artificial Intelligence Act Recital (41)、第 5 條 (H).

聲明明確指出：最終的決策仍然是一項由人類掌控的行為（Die endgültige Entscheidungsfindung bleibt eine von Menschen gesteuerte Tätigkeit.）。強調最終的裁判或決定仍為人所為，不可由 AI 自主代替司法裁量。

3. 司法人員職能與責任的不可或缺性

不僅法官，而是所有司法領域人員皆是不可或缺的，在 AI 運用中意謂人類司法工作者的職能與責任為不可或缺，即不能由 AI 取代或完全自動化，也確保工作權。

4. 建立監督與評估機制

包括對 AI 系統的安全、可靠性、數據品質、風險管理機制的制度化。

5. 促進司法人員的知能建設與變革管理

確保司法部門具備操作、評估、解釋 AI 系統的能力。

6. 透明公開

向民眾、當事人說明 AI 系統使用情形、算法邏輯、決策流程，並確保可理解、可說明及可追溯性。

7. 德國與歐盟 AI Act 的銜接

德國作為歐盟成員國，其司法領域的 AI 應用也必須符合歐盟的 AI Act 規定，特別是在涉及高風險 AI 系統的應用上，必須確保其準確性、穩健性，並避免歧視和偏見[83]。

(三) 法國

法國最高法院 AI 工作小組（Cour de cassation）於 2025 年 4 月發布了專題報告，針對下列三個核心提出意見：

1. 法院對 AI 的需求盤點

工作小組對法院內不同部門需求進行全面調查，需求包括：

83 Justizministerinnen und Justizminister des Bundes und der Länder. (2025, 5. Juni). Gemeinsame Erklärung zum Einsatz von Künstlicher Intelligenz in der Justiz, https://www.enorm.bund.de/SharedDocs/Pressemitteilungen/DE/2025/0605_Digitalgipfel.html, last visited on 2025/11/18.

⑴ 大量司法資料檢索與彙整工具。

⑵ 法條、判決交叉比對與自動關聯分析。

⑶ 相似判決搜尋工具。

⑷ 程序管理、案件分流、排程協助工具。

⑸ 文件摘要、草稿生成（判決書起草輔助）。

⑹ 自動匿名化。

⑺ 自然語言查詢。

法國最高法院已把 AI 視為司法流程中「節省時間、提升品質」的重要工具，但排除自動化裁判。

2. AI 評估方法

法國最高法院報告提出司法 AI 必須進行三重評估：

⑴ 倫理評估

檢查是否符合：基本人權、公平性、反歧視、人類可控性、避免濫用、不得削弱司法獨立及當事人權利保障（例如上訴權）。

⑵ 法律評估

審查是否符合歐盟一般資料保護原則（General Data Protection Regulation, GDPR）和歐盟 AI Act，特別是針對敏感司法資料的處理，必須有嚴格的保護措施。演算法在法律上可否作為基礎、程序是否必須由法官主導、AI 是否會影響審判中立性以及演算法產生之結果是否可作為證據。

⑶ 技術評估

主要考量模型可解釋性、透明度、錯誤風險、資料品質與偏誤及資訊安全。

3. 法官如何使用 AI 的具體規範

報告提出具體的司法使用方式，區分為：

⑴ 可使用（但需監督）：包括大量資料之搜尋、判決、法條之交叉比對、案例相似度輔助、自動摘要、判決草稿輔助、文書校對、自動匿名等。

⑵ 不可使用（禁止區域）：包括 AI 不得作為裁判者或決定者、不得以 AI 運算結果直接作為判決理由（不可引用演算法作為判決之法律基礎）、不得交由 AI 自動做出裁

判結論以及不得使用黑箱、不可解釋模型作為司法決策依據。

報告強調最終決定必須由人類法官作出，司法使用 AI 的合法性來自於人類掌控、透明、可審查、保護基本權利[84]。

（四）波蘭（Poland）

波蘭 2024 年發布的《關於在司法和執法中使用 AI 的建議》（Recommendations on AI in justice and law enforcement）建議採用 AI 來實現記錄數位化、自動化筆錄、起草命令和決策、使用聊天機器人以及搜尋判例，並推出數位司法助理（Digital Judicial Assistant, DJA），旨在協助法官撰寫法庭裁決背後的理由，預計將於 2026 年中旬開始在選定的法院使用。目前，該工具在波茲南地區法院（District Court in Poznań）進行測試，但僅限於協助行政秘書工作。雖然波蘭沒有針對含法官在內法院內部人員卷證使用 AI 的專門禁令，但作為歐盟成員國必須遵守 GDPR 規範[85]。

就律師運用 AI 的專業指引，業據波蘭全國律師委員會（National Council of Attorneys-at-Law）於 2025 年發布《關於律師應如何使用 AI 工具的建議[86]》（RECOMMENDATIONS ON HOW ATTORNEYS-AT-LAW SHOULD USE AI-BASED TOOLS），分成三階段規範：

第一階段：準備使用 AI 工具（Recommendations for preparing to use AI tools）

在此階段，律師需在選擇工具前進行技術與法律評估：

1. 確定用途：明確律師使用 AI 的具體目的（如通用支援或特定訴訟用途）。
2. 了解工具運作與數據處理：評估 AI 如何處理輸入數據、是否使用網路數據（Grounding）、數據存儲位置（如雲端）以及過濾機制。
3. 確認合約版本：確定適用於該 AI 工具的具體合約及條款（包含授權條款）。
4. 保密性評估：審查合約中關於輸入與輸出內容的保密條款，特別是專業保密

84 Cour de cassation. (2025). Cour de cassation et intelligence artificielle: préparer la Cour de demain, https://www.courdecassation.fr/files/files/Publications/IA%20-%20Rapport%202025/Rapport_IA_2025_Web.pdf.

85 Szeliga, A., Kalinowski, M., Gil, M., Macura-Targosz, S., & Tujakowska, A. (2025). Poland: Law and practice. In Artificial intelligence 2025 (pp. 1-31). Chambers and Partners.

86 National Council of Attorneys-at-Law. (2025). AI in the work of an attorney-at-law: Recommendations on how attorneys-at-law should use AI-based tools (1st ed.). Krajowa Izba Radców Prawnych.

（Professional Secrecy）的保護，確保數據不被用於訓練模型。

5. 著作權評估：確認律師對輸入內容擁有權利，並釐清供應商對輸出內容的權利主張。
6. 侵權防禦：確認供應商是否承諾在發生第三方智慧財產權索賠時為用戶辯護。
7. GDPR 合法評估：評估個資處理角色（控制者 / 處理者）、簽署數據處理協議（DPA）、評估數據跨境傳輸及是否需進行隱私衝擊評估（DPIA）。
8. 安全措施配置：確定需實施的技術與組織安全措施（如加密、身份驗證）。
9. 評估 AI 代理（Agents）：評估是否使用 AI Agents 來自動化組織流程。
10. 技術整合評估：評估 AI 工具與現有律師系統（如 CRM、文件管理系統）整合的技術可行性與安全性。
11. 微調模型（Fine-tuning）評估：評估是否需要使用內部數據對模型進行微調訓練。

第二階段：實施 AI 工具（Recommendations for implementing AI tools）

在此階段，律師需處理客戶關係與內部規範：

1. 審查客戶合約：檢查現有客戶合約中是否有禁止或限制使用 AI 的條款。
2. 通知客戶：識別哪些情況下必須通知客戶正在使用 AI（例如涉及重大戰略諮詢時），一般標準工具使用可能無需通知。
3. 取得客戶同意：確定哪些情況需取得客戶明確同意（例如使用客戶數據訓練模型、或是客戶指定特定律師親自處理的事項）。
4. 費用調整：評估使用 AI 是否影響律師費計費方式（如按時計費或固定費用），以及是否需轉嫁 AI 工具成本。
5. 制定內部規範：制定內部使用 AI 的建議書或指導原則（包含禁止事項、提示詞撰寫規則、驗證規則等）。
6. 進行內部培訓：在使用前對員工進行 AI 工具操作培訓。

第三階段：使用 AI 工具（Recommendations for using AI tools）

在此階段，律師需進行持續的監控與管理：

1. 定期評估：定期評估所使用的 AI 工具及市場上針對律師的新工具。
2. 更新內部規範：隨著 AI 技術發展，定期審查並更新內部的 AI 使用建議書。

3. 定期培訓：舉辦定期的內部培訓課程，分享經驗並介紹新功能。

4. 建立知識庫：建立並管理內部知識庫（包含經過驗證的提示詞、合約範本等）。

5. 事件管理：建立 AI 事件管理機制，處理如數據洩露等安全事件，並採取預防措施。

（五）小結：歐洲國家的共同特徵

1. 將 AI 視為輔助工具非決策者

歐洲各國，無論是通過 CEPEJ 倫理憲章還是德國的共同聲明或法國最高法院報告，都堅定地將 AI 定位為輔助工具，嚴禁取代法官的裁量權或最終決定權。

2. 透明性要求普遍高於美國

歐盟 AI Act 對高風險 AI 系統（例如，用於輔助評估犯罪風險或影響基本權利的系統）有嚴格的透明度、可解釋性和合格評鑑要求。

3. 注意資料治理與偏誤

歐洲規範高度關注無歧視原則、演算法偏見，以及個資保護（與 GDPR 銜接），強調 AI 系統的設計和使用必須確保司法程序的公平性與中立性。

4. 多為「司法行政」主導而非個別法官主導

歐洲多數國家的指引或策略（如德國司法部長會議、法國最高法院報告）是由中央司法行政部門或最高法院等體系性機構推動，旨在制定統一的治理框架，而非美國常見的由個別法官發布的常設命令。

三、英國、紐西蘭、澳洲、加拿大、香港等英國法系法院 AI 規範

（一）英國（英格蘭及威爾斯）：針對司法機構 AI 使用指南

該指南（2023 年 12 月 12 日第一版，目前 2025 年 10 月 31 日更新第三版）適用於英格蘭及威爾斯各級法官、司法助理、法律顧問、官員及其他支援人員。

1. 機密性與隱私：不得將非公開資訊（如私人或機密資訊）輸入公開的人工智慧聊天機器人。應將任何輸入至公開聊天機器人的資訊視為已公開。

2. 問責制與準確性：在使用或依賴人工智慧工具提供的任何資訊之前，必須檢查其準確性。人工智慧工具可能產生虛構的案例、引文或引用，提供不正確或誤導性的法律資訊，或出現事實錯誤。
3. 人類監督：維護人類對人工智慧系統的掌控能力。
4. 律師的責任：應告知訴訟參與者，他們對提交給法院的人工智慧生成資訊負有責任，如同對待任何其他類型的證據一樣。
5. 使用提醒：若發現當事人提交的書面材料似乎使用了人工智慧聊天機器人所生成，法官可詢問相關的準確性檢查工作。 AI 可用於摘要大量材料、撰寫簡報、電郵等行政任務，較適合用於對已知資訊進行非決定性的確認。但不建議法官以公開聊天機器人進行法律研究或法律分析，原因是擔心其可驗證性不足和可能產生虛構判例。
6. 個人責任：

⑴ 最終責任：司法職位擔任者需對以其名義產出的內容負個人責任。

⑵ 不可取代親自審閱：法官必須閱讀基礎文件，AI 不能取代法官與證據的直接接觸。

⑶ 設備使用：應使用法院工作設備而非個人設備來存取 AI 工具以避免機密外洩[87]。

(二) 紐西蘭：針對法院內外部人員使用生成式 AI 指南

2023 年 12 月 7 日紐西蘭法院人工智慧諮詢小組提出經由「法院首長委員會（Heads of Bench Committee）」、「紐西蘭審裁處執行委員會（Tribunals Aotearoa Executive）」及各審裁處及管理局主席批准，分別針對法官、律師及非律師關於在法院使用生成式 AI 的指引。

1. 針對司法人員部分

⑴ 指引規範對象：包括法官、司法官員、審裁處成員及司法支援人員（包括助理、私人秘書、書記官及法律研究顧問）。

⑵ 指引內容重點

① 理解生成式 AI 及其限制

87 Courts and Tribunals Judiciary. (2025, October 31). Artificial intelligence (AI): Guidance for judicial office holders, https://www.judiciary.uk/wp-content/uploads/2025/10/Artificial-Intelligence-AI-Guidance-for-Judicial-Office-Holders-2.pdf, last visited on 2025/11/19.

使用者應認知生成式 AI 是基於機率預測文字組合，而非從權威資料庫檢索答案，並非搜索引擎。另外，生成式 AI 現有模型對紐西蘭法律或程序要求的訓練數據有限。此外，生成式 AI 產出內容可能看似具說服力，但事實上不正確、不完整、具誤導性或有偏見。

② 維護保密、禁制令與隱私

使用者不得輸入非公開、機密、受禁制令保護或具法律特權的資訊。輸入的資訊可能被 AI 記憶並用於回答他人，導致洩密。應使用公務設備及電子郵件，並在可能的情況下關閉聊天紀錄（Chat History）功能。

③ 確保問責與準確性

使用者使用或依賴生成式 AI 提供的資訊前，必須獨立查核其準確性。 AI 可能捏造案例、引註或法條。

④ 意識道德問題

AI 反映訓練數據中的偏見，且通常缺乏紐西蘭的文化脈絡（特別是毛利與太平洋島國文化價值）。

⑤ 揭露使用情況

使用者不需要揭露是否使用了生成式 AI，支援人員需與督導法官討論其使用方式及風險緩解措施。

⑥ 留意他人的使用

法官需警惕律師或無律師代表的當事人（Lay Litigants）在書狀中使用 AI。若文件中有特別警示訊號：例如使用美式拼寫、引用陌生的海外案例、內容看似完美但有明顯實質錯誤。

⑦ 建議與不建議的用途

建議使用在摘要資訊、改變語氣（如改寫為淺顯英語）、演講撰稿大綱、行政任務（需謹慎個資）。需額外小心在法律研究使用 AI，應僅限於對已知正確資訊的高層次解釋，不可替代查閱權威來源。不建議使用在法律分析，因為 AI 缺乏理解內容或人類推理的能力，無法進行批判性檢視，盲目依賴將損害司法正直性 [88]。

88 Courts of New Zealand. (2023, December 7). Guidelines for use of generative artificial intelligence in courts

2. 針對律師部分 [89]

除適用針對法官指引①～④部分，針對律師特別強調律師有責任確保提供給法庭的所有資訊皆為準確。在使用生成式 AI 提供的資訊（含法律引註）前，必須進行查核。此外，律師不需要依慣例主動揭露是否使用了生成式 AI，除非法庭或審裁處有此要求 [90]。

3. 針對非律師部分

針對代表自己或他人出庭的「非律師人員（Non-lawyers），指引特別強調生成式 AI 不能取代合格律師，無法提供量身定做的法律建議；生成式 AI 可能協助識別相關法律原則（需查證）、整理事實結構、起草基本文件、或提供語法和語氣的建議；使用者有責任確保提供給法庭的所有資訊皆為準確；除非法庭或審裁處詢問，否則不需要預設主動揭露是否使用了 AI，但法庭有權要求您揭露是否使用了 AI [91]。

(三) 加拿大

2024 年 5 月 7 日加拿大聯邦法院（Federal Court）針對參與加拿大聯邦法院法律程序的律師、當事人及訴訟參與人，適用於提交給法院且為訴訟目的而準備的文件所發布的指引 [92]，其內容重點如下：

1. AI 生成內容的聲明

若提交的文件中包含由 AI 直接創建或生成的內容，必須通知法院及其他當事人。也適用於 AI 扮演類似「共同作者（co-author）」角色的情況。上開情形必須在文件的

---

and tribunals, https://www.courtsofnz.govt..nz, last visited on 2025/11/18.

89 本文審稿後，紐西蘭法律協會 (NZLS) 於 2026 年 2 月發布了更詳盡的《律師生成式 AI 使用指南》，參見 New Zealand Law Society. (2026, February). Generative AI guidance for lawyers, https://www.lawsociety.org.nz/professional-practice/rules-and-maintaining-professional-standards/generative-ai-guidance-for-lawyers/, last visited on 2026/05/06.

90 Courts of New Zealand. (2023, December 7). Guidelines for use of generative artificial intelligence in courts and tribunals: Lawyers, https://www.courtsofnz.govt.nz, last visited on 2025/11/18.

91 Courts of New Zealand. (2023, December 7). Guidelines for use of generative artificial intelligence in courts and tribunals: Non-lawyers, https://www.courtsofnz.govt.nz, last visited on 2025/11/18.

92 Federal Court. (2024, May 7). Notice to the parties and the profession: The use of artificial intelligence in court proceedings, https://www.fct-cf.gc.ca, last visited on 2025/11/20.

第一段明確寫出聲明，指出整份文件或特定段落使用了 AI，而法院對於做出 AI 使用聲明，不會據此做出不利推斷。若 AI 僅用於建議修改、提供推薦、或對人類已創作的內容進行批評（且由人類手動實施修改），則無需聲明。亦不適用於僅遵循預設指令的系統自動化、語音識別或文件編輯工具[93]。

2. AI 使用原則

使用 AI 生成的法律參考資料或分析時應格外謹慎，必須使用公認且可靠的來源（如官方法院網站 CanLII）來查核判例、法規或評論，以防範幻覺或深偽。並且為確保準確性與可信度，必須由人類核實所有 AI 生成的文件與內容，這符合法律專業的一般標準。法院在未經公眾諮詢前，不會使用 AI 或自動決策工具來做出裁決或判決。

此外，加拿大安大略省法律委員會（Law Commission of Ontario, LCO）於 2025 年 4 月針對 AI 在刑事司法系統（執法、保釋、量刑、審判、監管）中的應用、風險與法律改革建議發布的報告，其中關於審判部分[94]，建議針對 AI 證據的特定可採性規則，例如要求在採納前必須通過類似專家證人測試的可靠性聽證會。另外，制定規則強制揭露 AI 系統的相關資訊（如錯誤率、訓練數據、演算法限制），而不受商業機密特權的絕對保護。再者，建議設立類似「司法與科學委員會」的獨立機構，專門評估法庭所用科學技術（含 AI）的可靠性。此外，加強法官和律師對 AI 技術的理解，以避免盲目信任技術。在量刑部分，建議 AI 風險評估分數僅能作為減輕刑罰的因素，不應作為加重刑罰的唯一依據。高風險評分不應單獨呈現給法官，必須有其他外部證據支持[95]。

---

93 加拿大於 2025 年 12 月發布《2024-2025 年聯邦法院規則全面審查最終報告》，報告說明在諮詢過程中，反覆出現建議是規則應明確規範 AI 在法院文件及專家證據中的使用，原指引（2024）及修正指引（2025 年）對於違反 AI 聲明之法律效果：不利費用裁定（adverse cost award）命令說明為何不應被認定藐視法庭（order to show cause），已提供足夠機制。委員會建議目前不修訂規則以規範 AI 使用，但此議題應隨技術及其在法律程序中的運用持續演進而繼續檢視，參見 Federal Court of Appeal and Federal Court Rules Committee, 2024-2025 Global Review of the Federal Courts Rules Final Report of the Sub-Committee December 2025, https://www.fct-cf.ca/en/pages/law-and-practice/amendments-to-the-rules/global-review-of-the-federal-courts-rules-final-report-december-2025.

94 Thompson, P., & Neubauer, E. (2025, April). AI at trial and on appeal: Paper 4 in the LCO AI in criminal justice project. Law Commission of Ontario, https://www.lco-cdo.org/wp-content/uploads/2025/11/LCO-AI-in-Criminal-Justice-Paper-4-AI-at-Trial.pdf, last visited on 2025/11/20.

95 D'Andrea, A., & Christian, G. (2025, April). AI and the assessment of risk in bail, sentencing and recidivism: Paper 3 in the LCO AI in criminal justice project. Law Commission of Ontario, https://www.lco-cdo.org/wp-

3. AI 生成資料作為證據的問題

LCO 於 2025 年 11 月整理發布的專家討論紀要[96]，該報告的核心在於探討當 AI 生成的分析結果（例如：複雜的金融犯罪模式分析、臉部辨識比對結果或預測警政數據）進入法庭成為「證據」時，現有的證據法規則（如鑑定證人標準、傳聞證據排除原則）是否足以應對。這份總結報告將專家意見歸納為以下四個關鍵衝突點。

⑴ 黑盒子與對質權的衝突

技術透明度：專家指出許多警用 AI 系統受「商業秘密」保護，律師無法檢視其演算法代碼。這可能侵害被告在憲法（加拿大權利與自由憲章）下的對質權，因為辯方無法有效交叉詢問一個「不可解釋」的演算法。專家建議建立「法庭指定專家」制度，由具備技術能力的第三方在保護商業秘密的前提下，對演算法進行審核。

⑵ 鑑定證據的門檻（The Daubert / Mohan Standard）

可靠性審查：討論了法院如何應用 R. v. Mohan[97] 標準（加拿大鑑定證據標準）來審查 AI。報告指出傳統的「普遍接受度」在快速迭代的 AI 領域難以界定。專家建議法院不應只審查 AI 的設計，更應審查該 AI 在「特定案件數據集」上的表現（即「特定情境下的可靠性」）之動態驗證。

⑶ 披露義務（Disclosure Obligations）

Stinchcombe 原則的延伸：在加拿大，檢方有義務披露所有相關資訊。報告探討了檢方是否必須披露 AI 的訓練數據（Training Data）及其可能存在的偏誤評估報告。專家強調若無法提供數據來源的完整監管鏈（Chain of Custody），AI 生成的報告應被視為不具證據能力。

⑷ 司法認知與陪審團偏見

專家擔心法官或陪審團可能對「機器產出的結果」產生過度信賴之自動化偏誤，將其誤認為是客觀事實而非推論。建議應強制要求在 AI 證據旁附帶「不確定性評分」

---

content/uploads/2025/11/LCO-AI-in-Criminal-Justice-Paper-3-AI-and-Risk-Assessment.pdf, last visited on 2025/11/20.

96 Law Commission of Ontario. (2025, November). Summary of proceedings: Roundtable on AI evidence in criminal proceedings, https://www.lco-cdo.org/wp-content/uploads/25-12-18-LCO-Toronto-Roundtable-Report.pdf.

97 R. v. Mohan, [1994] 2 S.C.R. 9.

或「誤差率」。

(四) 澳洲[98]

新南威爾斯州最高法院（Supreme Court of New South Wales, NSW）於 2024 年 11 月 21 日發布法院使用者及給法官的 AI 指引，適用對象包括在 2025 年 2 月 3 日起，所有在 NSW 最高法院進行的訴訟程序之法律實務工作者（包括法官、律師、訴訟代理人等）、未受律師協助之當事人及專家證人。適用範圍排除使用 AI 拼字、文法修正、逐字稿、翻譯工具、排版、搜索引擎或使用 AI 強化專門法律檢索系統都不視為使用生成式 AI。再者，應適當告知說明使用 AI 風險，如幻覺、偏見等。對於律師及訴訟當事人主要指引內容如下：

1. 核心禁令與限制

⑴ 機密資料輸入禁止：受法院禁制令、隱私保護或法定義務保護的機密資訊，嚴禁輸入至任何生成式 AI 程式中。

⑵ 宣誓書與證人陳述書（Affidavits & Witness Statements）：

① 禁止使用 AI 生成證人的證據內容或意見。

② 禁止使用 AI 來修飾、潤色、強化或淡化證人的證詞。

③ 必須揭露：證人必須在文件中聲明未使用生成式 AI 來產生內容。例外：若需使用 AI 製作附件（Annexure）或證物，必須先取得法院許可（Leave）。

⑶ 專家報告（Expert Reports）：未經法院事先許可，不得使用生成式 AI 起草專家報告。若需使用（例如用於分析大數據、聲音或影像），必須說明使用的軟體版本、是否為封閉式系統、以及使用的效益。若獲准使用，必須在報告中詳細揭露使用了哪些 AI 工具、輸入了什麼提示詞等細節。

⑷ 書狀與陳詞：若使用 AI 協助準備書狀，使用者必須在文件中驗證並聲明所有引用的判例、法條與學術資料皆為真實、準確且與案件相關。驗證過程本身不得使用 AI 進行。

98 除本文介紹外，尚有維多利亞最高法院 2024 年 5 月亦有發布 AI 指引，差異僅在是否完全禁止 AI 生成或處理之證據，Supreme Court of Victoria. (2024). Guidelines for litigants: Responsible use of artificial intelligence in litigation. https://www.supremecourt.vic.gov.au/, last visited on 2025/11/20.

2. 對法官的規範

⑴ 判決書撰寫禁止：法官不得使用生成式 AI 來撰寫判決理由、評估證據，亦不得將判決書草稿輸入 AI 進行編輯或校對。

⑵ 研究與助理管理：法官若使用 AI 進行輔助研究，必須自行核實結果，不可替代傳統研究方法。法官應要求其助理若有使用 AI 必須揭露並進行驗證。

⑶ 監管當事人：法官應敏銳地察覺當事人是否未經揭露即使用 AI，並留意 AI 生成內容的「危險訊號」（Red Flags），例如：引用不存在的判例或法條；使用非本地（如美國式）的拼寫或法律術語；內容重複或分析邏輯明顯有誤[99]。

（五）香港

香港司法機構政務處（Judiciary Administration）於 2024 年 7 月針對法官、司法人員及司法機構支援人員，執行司法和行政職務時使用生成式 AI 的原則，主要指導原則如下：

1. 禁止授權司法職能：法官必須親自且獨立地做出司法決定，嚴禁讓生成式 AI 接管司法職能，AI 僅能作為輔助工具。
2. 理解限制並確保準確性：使用者需了解 AI 可能產生不準確、誤導或有偏見的內容（如幻覺捏造案例），在使用 AI 生成的資訊前，必須進行查核與驗證。
3. 維護資訊保安與保密：僅使用司法機構提供的設備，避免使用個人設備或公共 WiFi，勿將私人、機密或敏感資訊輸入至開放式 AI 聊天機器人，因為輸入的資料可能會被系統保留並公開。若可行，應關閉聊天記錄功能，並拒絕 AI 平台不必要的權限要求。
4. 防止侵犯著作權：避免上傳受著作權保護的資料進行摘要或分析，並注意 AI 生成內容可能引發的著作權問題。
5. 注意偏見：AI 的回應基於訓練數據，可能包含文化、倫理或地理偏見，使用者需加以修正。

---

99 Supreme Court of New South Wales. (2024). Supreme Court practice note SC Gen 23: Use of generative artificial intelligence (Gen AI) and guidelines for New South Wales judges, https://supremecourt.nsw.gov.au/documents/Practice-and-Procedure/Practice-Notes/general/current/PN_SC_Gen_23.pdf, last visited on 2025/11/20.

6. 承擔責任：法官與職員需對以其名義產出的材料負個人責任，即使該材料部分源自 AI。
7. 留意法院使用者的 AI 使用：若發現訴訟當事人（特別是無律師代表的訴訟人）可能使用 AI 準備文件，法官應詢問其是否已核實準確性；律師亦有責任確保呈交內容的準確性。
8. 潛在用途與限制：

⑴ 適用：摘要資訊（需人工核對）、演講 / 簡報撰寫大綱、翻譯輔助、行政工作（如起草郵件，但需注意個資）。

⑵ 不建議：不建議用於法律分析，因為 AI 缺乏對語境的理解及批判性思考能力，容易得出不準確結論[100]。

（六）小結：英國法系模式之比較

1. 對引用真實性最為重視（與美國相近）：英國、紐西蘭、澳洲、加拿大、香港的指引都高度關注幻覺風險，要求律師或訴訟參與人對引用的真實性負最終查證責任。加拿大甚至必須事先聲明使用 AI 生成之部分。
2. 對機密保護要求接近歐盟：這些普通法司法管轄區對機密資料的保護要求極高，特別是針對公開 AI 模型，嚴禁輸入敏感資訊。
3. 不鼓勵將 AI 用於實質法律研究：英國、紐西蘭和香港的法官指引都明確建議不鼓勵或不建議將公開 AI 用於法律研究或法律分析，而是應限於整理、摘要或行政性任務。

四、亞洲國家法院運用 AI 規範

亞洲地區法院對於 AI 使用的規範，特別是對 AI 生成證據性材料，採取了比歐美更為具體和嚴謹的態度。

100 Judiciary Administration. (2024, July). Guidelines on the use of generative artificial intelligence for judges and judicial officers and support staff of the Hong Kong Judiciary. The Government of the Hong Kong Special Administrative Region, https://www.judiciary.hk/doc/en/court_services_facilities/guidelines_on_the_use_of_generative_ai.pdf, last visited on 2025/11/21.

(一) 新加坡：法院使用者生成式 AI 指引

新加坡最高法院 2024 年 9 月 23 日發布法院使用者生成式 AI 指引[101]，並於 2024 年 10 月 1 日生效，適用對象包括檢察官、律師、當事人、證人等所有法院使用者，適用案件包括新加坡最高法院（Supreme Court）、國家法院（State Courts，包含小額索償、僱傭索償與社區糾紛解決法庭）以及家庭司法法院（Family Justice Courts）。指引核心內容如下：

1. 一般原則與責任歸屬

法院不禁止使用生成式 AI 工具準備法庭文件，但使用者必須承擔生成內容的全部責任。除非法院特別詢問，否則使用者不需要預先聲明其使用了生成式 AI。律師與自我代表當事人仍有義務確保提交給法院的所有資料皆經過獨立驗證、準確、真實且適當。

2. 禁止事項與限制

禁止使用生成式 AI 來產生、偽造、修飾或竄改任何證據（例如：不能要求 AI 憑空捏造一份宣誓書的草稿內容作為證據）。指引強調生成式 AI 不是搜尋引擎，它可能會產生看似具說服力但實際上不準確或虛構的內容。若法院懷疑文件來源不當，可要求：說明是否使用 AI、宣示遵守指引、必要時提交宣誓書（affidavit）。

3. 準確性核實義務

使用者必須對 AI 生成的內容進行事實查核與校對，而且必須確認 AI 引用的判例、法條或文章是真實存在的，且內容正確。使用者不能使用另一個生成式 AI 工具來驗證原本 AI 生成的內容是否正確，必須使用可靠的來源（如 eLitigation 官方判決查詢系統 GD Viewer[102] 或新加坡法規網站）進行查證。

---

101 Singapore Supreme Court (2024). Registrar's Circular No. 1 2024. Guide on the use of generative artificial intelligence tools by court users, https://www.judiciary.gov.sg/news-and-resources/registrar's-circulars/circular-details/registrar's-circular-no.-1-2024-supreme-court, last visited on 2025/11/21.

102 新加坡官方網站 https://www.elitigation.sg/gd/, last visited on 2025/11/21.

4. 保密與智慧財產權

使用者不得在 AI 工具中輸入機密或敏感資訊，因為這些資訊可能會被公開或被 AI 工具儲存。使用 AI 生成內容時，若有引用他人著作，必須適當標註原始來源。

5. 違規後果

若未遵守本指引，法院可採取以下措施：

⑴命令使用者支付訴訟費用（若為律師，可能需個人承擔）。

⑵忽視該文件或降低其證據權重。

⑶對律師採取紀律處分。

⑷可能涉及藐視法庭、個資違法等法律後果。

新加坡法務部（Ministry of Law）於 2025 年 9 月 1 日發布法律界使用生成式 AI 指引（Guide for Using Generative AI in the Legal Sector）公告為期一個月之公眾諮詢草案[103]，適用對象比上開指引更廣，適用於任何在新加坡處理法律工作的人員，有三大核心原則：

1. 專業倫理：無論是否使用 AI，法律專業人員必須對所有工作成果負全責，且必須審查、分析並驗證所有 AI 生成的內容，特別是針對幻覺風險。在自己不熟悉的專業領域使用 GenAI 時應更加謹慎，需額外查核權威來源。
2. 保密性：使用生成式 AI 工具時必須確保客戶機密不外洩，並應區分「免費工具」與「企業級工具」。使用免費工具時不得輸入客戶機密或敏感資訊，且輸入前需將資料匿名化。使用企業級工具時，需確認服務條款明確禁止將輸入輸出資料用於訓練 AI 模型。需考量客戶資料是否被儲存在新加坡境外及其合規性。
3. 透明度（Transparency）：若使用生成式 AI 會實質影響客戶利益（如費用、服務交付方式）或涉及資料處理風險，應考慮向客戶揭露。並應解釋 AI 如何實施、強調律師仍負最終責任、說明資料保密措施，並提供客戶選擇退出（opt-out）的選項[104]。

---

103 新加坡律政部 2026 年 3 月 6 日正式公告，參見 Ministry of Law, Singapore. (2026, March 6). Guide for using generative AI in the legal sector, https://www.mlaw.gov.sg/files/Guide_for_using_Generative_, last visited on 2025/11/21.

104 Ministry of Law, Singapore. (2025). Guide for using generative AI in the legal sector (Draft for public consultation), https://www.mlaw.gov.sg/public-consultation-on-guide-for-using-generative-artificial-

(二) 韓國

韓國最高法院人工智慧研究會（법원 인공지능연구회）於 2025 年 2 月發布司法領域人工智慧使用指引（사법에서의 인공지능 활용에 관한 가이드라인），針對司法機關（法院整體）、法官及訴訟當事人提供使用 AI 時之建議，核心規範內容如下：

1. 司法機關開發與引進 AI 的六大原則

⑴ 基本權保障原則：不得侵害國民基本權，應優先考慮能提升社會弱勢司法接近權的系統。

⑵ 信賴性原則：必須確保準確性，並努力透過技術措施排除數據偏見。

⑶ 合法性原則：需與整體法律秩序調和，遵守個資法與智慧財產權法規。

⑷ 責任性原則：AI 僅為輔助，系統設計須讓法官能驗證結果準確性，並由法官承擔最終責任。

⑸ 透明性原則：應將數據來源、訓練過程與演算法文件化，並考慮向大眾公開。

⑹ 未來導向原則：系統設計應保持開放性以適應模型變更，並持續確保預算與制度依據。

2. 法官使用 AI 的規範

⑴ 憲法價值維護：法官不能將事實認定與法律解釋的工作完全委託給 AI。

⑵ 人機協作與責任：法官需認知到 AI 會有幻覺與偏見，必須對 AI 產出的結果進行批判性評估與驗證，並對最終判決負全責。

⑶ 機密與個資保護：禁止將個資、案件具體內容、營業秘密、判決書草稿輸入至商用 AI 工具中。避免使用法院官方信箱註冊商用 AI 帳號或在帳號中標示法官身分。若需使用，應確認該 AI 工具提供「不使用輸入數據進行模型訓練」的選項。

3. 訴訟當事人使用 AI 的規範

⑴ 揭露義務：若懷疑或經對造指出當事人提交的資料（書狀、影音證據）是由 AI 製作，法院可命提出者揭露是否使用了 AI。

---

intelligence-in-the-legal-sector/, last visited on 2025/11/21。正式版本已於 2026 年 3 月 6 日發布。

⑵ 訴訟指揮：法院可要求當事人具體說明使用了何種 AI 工具、輸入了什麼提示詞，以及採取了何種驗證措施。

⑶ 證據能力與深偽：針對 AI 生成的證據（如 Deepfake 疑慮），需嚴格審查其證據能力。若當事人使用 AI 撰寫書狀，法院可警告其須對內容錯誤（如引用不存在的判例）自行承擔法律後果[105]。

(三) 中華人民共和國

中國最高人民法院於 2022 年 12 月 8 日發布了《關於規範和加強人工智能司法應用的意見》[106]，提出了具體原則：

1. 核心定位：輔助審判原則（最關鍵的紅線）

⑴ AI 不得取代法官：文件明確規定，無論技術發展到何種水平，人工智能都不得代替法官裁判。

⑵ 僅供參考：AI 的輔助結果僅可作為審判工作或審判監督管理的參考。

⑶ 責任歸屬：司法裁判始終由審判人員作出，裁判職權由審判組織行使，司法責任最終由裁判者承擔。

⑷ 自主選擇權：各類用戶（包括法官與當事人）有權選擇是否利用 AI 提供的輔助，並有權隨時退出與 AI 的互動。

2. 基本運作原則

⑴ 安全合法：禁止使用不符合法律法規的 AI 技術和產品，且不得損害國家安全，不得侵犯合法權益。必須確保國家秘密、網絡安全、數據安全和個人資訊不受侵害。

⑵ 公平公正：保證 AI 產品無歧視、無偏見，且確保不因技術介入、數據或模型偏差而影響審判過程和結果的公正。

⑶ 透明可信：AI 系統的司法數據採集、法律語義認知、輔助裁判邏輯等環節，必須是可解釋、可測試、可驗證的。系統投入應用時，必須說明並標識其功能、性能與局

105 법원 인공지능연구회 . (2025). 사법에서의 인공지능 활용에 관한 가이드라인 , https://www.lexology.com/library/detail.aspx?g=879a6b11-ce47-40ae-ad28-d87e076addcc, last visited on 2025/11/22.

106 中國最高人民法院《關於規範和加強人工智能司法應用的意見》，https://www.court.gov.cn/fabu/xiangqing/382461.html, last visited on 2025/11/22.

限，確保結果可預期、可追溯。

⑷ 公序良俗：AI 應用不得違背公序良俗，不能損害社會公共利益和秩序，不能違背社會公共道德和倫理。並且需防範倫理道德風險，建立風險管控與責任查究機制。

3. 具體應用範圍（允許與鼓勵的方向）

⑴ 全流程輔助辦案：支持證據指引與審查、法律法規推送、類案推送、裁判輔助、法律文書輔助生成與審查等。

⑵ 輔助事務性工作：支持電子卷宗自動分類、案件信息自動回填、筆錄自動生成、送達地址推薦等，以減輕工作負擔。

⑶ 輔助司法管理：支持案件裁判偏離度預警、不規範司法行為自動巡查、廉潔司法風險防控等。值得注意這裡旨在利用 AI 技術自動偵測與巡查司法過程中不符合規範的行為，目的是「提升司法管理質效」並「保障廉潔司法」，顯示其核心作用在於透過自動化技術強化對司法權運行的內部監督與風險控制。

⑷ 服務社會治理：支持訴訟諮詢問答、訴訟預期輔助評估、社會治理風險預警等。

4. 安全保障與風險防範

⑴ 知識產權保護：支持司法 AI 的創新與專利、著作權申報，並切實保護相關知識產權。

⑵ 審查機制：通過司法人工智能倫理委員會等機制，採用倫理審核、合規審查、安全評估等方式，防範化解應用過程中的安全風險。

中國對於律師運用 AI 並無專屬規範，但 AI 使用必須遵守個人信息保護法（Personal information Protection Law, PIPL）、演算法推薦管理規則及科技倫理規範等一般性規範。

（四）小結：亞洲模式觀察

1. 法院從上而下 AI 管制：新加坡、韓國及中國均由法院體系中最高法院層級發布法院 AI 使用指引。
2. 最嚴格 AI 使用控管：新加坡在不強制披露的情況下，卻對 AI 生成之證據一律封殺，對證據性材料有明確的紅線，反映出對 AI 破壞程序正義和證據本質的擔憂。韓國則

是對於案件具體內容、甚至連法官身分都不得對商用 AI 透露。中國模式則賦予 AI 對司法工作流程一定程度的監督與預警功用。

五、南美司法系統之 AI 應用規範

(一) 阿根廷

阿根廷里奧內格羅省（Poder Judicial, Provincia de Río Negro）的高等法院於 2024 年 10 月 1 日發布司法部門使用生成式人工智慧良好實務議定書（Protocolo de buenas prácticas para el uso de inteligencia artificial generativa en el Poder Judicial）適用於該省包括法官在內司法部門的所有人員，該議定書旨在推動負責任、合乎道德且審慎地使用生成式 AI，並強調 AI 僅是輔助工具，不能取代人類能力。具體指引如下：

1. 行政管理與授權機制

人員若要在工作中使用生成式 AI，必須填寫申請表，並經由司法管理資訊化委員會主席批准。若有涉及改變工作模式的 AI 專案，需經高等法院事先授權。

2. 四大核心目標與具體使用規範

⑴ 風險緩解

① 強制驗證：AI 經常會產生看似合理但錯誤的內容（幻覺），使用者必須批判性地審查並透過可靠來源驗證所有 AI 生成的內容。

② 非搜尋引擎：不應將生成式 AI 當作搜尋引擎使用，因為它們是基於機率預測文字，而非檢索事實。

③ 避免過度依賴：使用者應接受培訓以了解風險，避免因 AI 回答的自信語氣而喪失批判能力。

⑵ 人類控制與禁止授權決策

① 禁止委派決策：使用者不得將決策權完全交給 AI，司法判決的責任完全在於人類官員。

② 全程監控：若將 AI 整合至系統中，必須確保人類能介入從生成到最終審查的每個階段。

⑶ 資料保護與隱私

① 禁止輸入個資：若 AI 工具的隱私政策未保證不將數據用於模型訓練或商業用途，則禁止輸入個人數據或機密資訊。

② 匿名化處理：若必須使用，應先採取技術手段將數據匿名化或假名化。

⑷ 持續培訓

建立持續的培訓計畫，涵蓋技術、法律與道德層面，並推動跨學科合作，包括提示詞的精進[107]。

(二) 巴西

巴西國家司法委員會（Conselho Nacional de Justiça, CNJ）2020 年 8 月 21 日第一次針對人工智慧應用所發布巴西人工智慧司法指引（Resolução n° 332/2020）[108]，2025 年 3 月 11 日發布第二版指引，特別針對生成式 AI 與大型語言模型，並於 120 日後生效，本指引適用對象包括法官在內、訴訟當事人及鑑定人，重要指引內容如下：

1. 生成式 AI 的使用規範

⑴ 輔助性質：生成式 AI 僅能作為輔助或支持決策的工具，禁止將其作為自動決策工具。法官必須對 AI 生成的內容進行指導、解釋、驗證與修訂，並對最終決策負全責。

⑵ 數據輸入禁令：禁止將受司法保密（segredo de justiça）或敏感數據輸入至私人或外部的生成式 AI 系統中，除非數據已在源頭完全匿名化。

⑶ 禁止高風險用途：禁止使用私人或外部生成式 AI 處理本指引定義為高風險或過度風險的任務。

⑷ 強制培訓：使用者必須先接受關於 AI 限制、風險及道德使用的培訓。

---

107 Superior Tribunal de Justicia de la Provincia de Río Negro. (2024). Acordada Nº 15/2024: Protocolo de buenas prácticas para el uso de inteligencia artificial generativa en el Poder Judicial, https://www.pensamientopenal.com.ar/legislacion/91487-rio-negro-protocolo-buenas-practicas-uso-inteligencia-artificial-generativa-poder, last visited on 2025/11/24.

108 National Council of Justice. (2020). Resolução nº 332, de 21 de agosto de 2020: Ética, transparência e governança na produção e uso de inteligência artificial no Poder Judiciário. Brasília, Brazil, https://atos.cnj.jus.br/atos/detalhar/3429, last visited on 2025/11/24.

⑸ 判決書註記：若法官使用生成式 AI 輔助撰寫判決，可自行決定是否在判決書中提及，但系統後台必須有自動紀錄以供審計。

2. 風險分級制度

將 AI 應用分為高風險與低風險：

⑴ 高風險：包括識別自然人特徵與行為模式（側寫）、評估證據價值、犯罪事實認定、預測判決結果、以及情感識別等生物特徵分析。

⑵ 低風險：包括行政管理自動化（分類、摘要）、內部判例檢索輔助（不替代人類評估）、生成草稿文本（需法官審閱）、轉錄影音、文件匿名化處理等。

3. 登錄透明政策

所有司法機關開發或使用的 AI 解決方案（計算模型）必須登錄國家司法委員會（Conselho Nacional de Justiça, CNJ）所管理 Sinapses（全國司法 AI 平台），並進行分類標明屬於高風險或低風險，以便使用者和 CNJ 知道如何負責任地使用和監督。各法院需公開 AI 名稱、版本、風險等級，並由 CNJ 定期公布全國 AI 清冊。

4. 禁止事項

司法機關禁止開發或使用以下 AI 解決方案：

⑴ 無法人工審查或導致用戶絕對依賴而無法修改結果的系統。

⑵ 犯罪預測：利用個人特徵評估犯罪可能性或再犯率來作為判決依據。

⑶ 社會評分：基於行為或社會地位對人進行分類或排名。

⑷ 情感識別：利用生物特徵識別情緒[109]。

(三) 小結：南美國家管制特色

1. 高度的「行政管控」與「事先授權」機制

阿根廷對於 AI 之運用係採事先許可制，須先取得批准或事先授權才能使用，在 AI 運用上罕見嚴格。

109 Conselho Nacional de Justiça. (2025). Resolução Nº 615, de 11 de março de 2025: Estabelece diretrizes para o desenvolvimento, utilização e governança de soluções desenvolvidas com recursos de inteligência artificial no Poder Judiciário, https://atos.cnj.jus.br/files/original155530202503146 7d4517244566.pdf, last visited on 2025/11/24.

2. 明確的「風險分級」與「禁止事項」清單

巴西的規範在風險管理上尤為細緻，不僅僅是原則性的指導，而是直接在法規中定義了具體的風險等級與絕對禁止的技術類型，這在其他國家的法院指引中較為少見。

3. 強調提示詞工程的具體教學

阿根廷的指引不僅規範不能做什麼，還花費篇幅教導法官如何做得好，在議定書中包含了關於提示詞的詳細教學章節，具體建議法官如何設定角色、語氣、背景與格式，以獲得更好的生成結果。

4. 對「外部私有模型」的嚴格數據隔離

南美國家對於將數據輸入至商業公司（如 OpenAI、Google）擁有的模型中，採取極度保守的態度。

5. 強制性的教育培訓要求

不同於部分國家僅將培訓列為建議，南美國家的指引將培訓列為強制性義務。

六、小結：各國法院 AI 規範的共通原則與差異

本文介紹 15 個國家或司法管轄區的法院運用 AI 規範情況總結如下：

(一) 各國共同一致的規範

儘管各國法制不同，但對於法院使用 AI 規範有以下高度共識的五項核心原則：

1. 人類保有最終決定權

AI 僅被定位為輔助工具，無論在何種情況下，AI 系統嚴禁取代法官的司法裁量權或最終決定權，法官必須親自做出司法決定，不可授權給 AI。

2. 使用者需負查核與驗證責任

由於生成式 AI 具有產生幻覺的風險，律師或使用者必須對 AI 生成的所有內容（特別是引用的判例、法條或學術資料）進行獨立核實，使用者不能以不知道 AI 會產生虛構內容作為提交不準確資訊的藉口。

3. 嚴格的保密與個人資料保護

各國普遍要求禁止將非公開、機密、個人識別資訊（PII）或受保護的案件資料輸入至公共或商用的生成式 AI 模型中，以防資料洩露或被用於訓練模型。

4. 禁止偽造與誤導法庭

律師對法庭負有坦誠義務，不得提交由 AI 虛構的判例或證據，這種行為會受到司法機構最嚴厲的處罰，包括停職、罰款或職業紀律處分。

5. 避免偏見與確保公平

各國普遍要求 AI 應用不得造成非法歧視、影響基本權利或產生演算法偏見。在使用 AI 系統時，需要評估訓練數據及演算法是否存在偏誤，以確保司法程序的公正性與中立性。

(二) 各國法院 AI 規範的相異之處

1. 對律師或外部使用者之規範

包括律師、訴訟代理人及自我辯護當事人使用生成式 AI 的許可範圍與義務，例如：是否必須揭露使用 AI、對 AI 生成內容是否負查核與驗證責任、違反時可能遭受的程序制裁或職業懲戒。美國透過聯邦民事訴訟程序法第 11 條、各級法院的 Standing Orders 以及州律師公會意見，波蘭、英國、紐西蘭、新加坡等國，則分別由律師公會或法院發布專門指引，形成高度專業自律的規範脈絡。

2. 對法院內部人員使用與機密資訊之規範

此一面向聚焦於法官、司法行政人員及法院 IT 部門等內部使用者，如何在工作流程中導入 AI，同時避免將卷證資料、個資或國家機密輸入商用雲端模型。歐洲 CEPEJ 倫理憲章、德國聯邦與各州司法部長共同聲明，以及英國、紐西蘭、澳洲等法院指引，皆強調：AI 僅能作為輔助工具，禁止以公開聊天機器人處理未公開資料，並要求司法機關建置封閉或受控環境的 AI 系統。

3. 對 AI 生成或處理之證據的證據法則

此部分關係到 AI 產生或加工的內容能否作為證據進入法庭，以及應適用何種證據

法標準。美國試圖透過修正聯邦證據規則第 707 條，將 AI 生成內容置於類似專家證人之審查架構；州法院亦於實務上對經 AI 強化或生成之影片、圖像採取嚴格證據能力審查。新加坡、澳洲等國，則在法院使用者指引中，直接禁止以 AI 捏造或修飾證人陳述與宣誓書內容，並要求證據真實性須以獨立方式加以驗證。

4. 對 AI 系統本身之風險分級與倫理框架

若將法院視為「高風險 AI 使用場域」，如何設計一套兼顧基本權保障與技術創新的倫理與監管架構，即成各國規範重要差異所在。歐洲透過 CEPEJ 倫理憲章與德國聯邦與各州司法部長會議共同聲明，提出基本權、非歧視、可解釋性與人類最終決定權等原則，並與 EU AI Act 高風險系統規範銜接；南美國家如巴西、阿根廷的法院與立法機關，則強調事先授權、風險分級與禁止性清單。中國大陸則在「智能法院建設指導意見」與相關技術規範中，將安全合法、公平公正、透明可信、公序良俗等要求具體化為 AI 系統開發與部署時必須遵守的運作原則。

肆、我國現況及未來展望 ── 結語

一、盤點目前我國司法使用 AI 現況

依司法院 112 年 8 月揭櫫司法數位 2.0[110]，盤點目前法院運用 AI 之技術，計有以下部分：

(一) 電子卷證智慧化分析

利用機器學習技術，針對毒品、竊盜、詐欺、傷害及妨害名譽等 5 種刑事案件，提供電子卷證書籤自動生成功能，協助法官快速獲取資訊，並針對全案電子卷證，提供多種個人資料的自動遮隱機制，藉以減輕書記官在提供閱覽前的處理負擔。惟因處理速度過慢以及遮隱成效有限，仍須人工介入，導致使用率不高。

(二) 智慧客服小幫手

導入自然語言處理技術，於司法院官網提供「智慧客服小幫手」。民眾可透過對話

110 司法院數位政策 2.0. 2023/08/09，https://www.judicial.gov.tw/tw/cp-247-919057-c3448-1.html, last visited on 2025/11/28.

方式進行諮詢，系統亦支援「庭期查詢」及「開庭進度查詢」功能，減輕法院訴訟輔導的人力負擔。

(三) 法庭語音辨識系統

配合國民法官新制，已於全國 21 所地方法院的 28 間國民法官法庭建置「法庭中文語音辨識系統」，利用 AI 語音辨識技術協助書記官製作筆錄，精確辨識開庭語音內容，提升開庭效率並減少筆錄修改需求。一般法庭雖亦部分建制語辨系統，但因辨識正確率未臻理想，實際開庭使用頻率不高。

(四) 裁判書草稿自動生成

司法院採用 TMT5 語言模型架構，在裁判書生成任務上進行監督式訓練的微調（fine-tuning），模型訓練階段以起訴書及配對的裁判書資料為主。草稿生成階段主要以對應的「起訴書」為生成素材，系統會即時分析犯罪事實、證據等段落資料，以提高草稿的準確度、可讀性及可解釋性。並採「主機完全落地」架構建置，相關資料在本地端進行訓練與生成，以確保個人隱私與公務機密。適用案件以不能安全駕駛（酒駕等）及幫助詐欺，預計試辦毒品罪，民事案件則規劃民事交通損害賠償、消債更生、清算事件。另外強調 AI 僅作為輔助工具，撰寫的裁判草稿僅供法官製作裁判時參考。審判的核心權力仍保留在法官手中，AI 不會取代法官認定事實、適用法律、決定量刑 [111]。此舉旨在減輕法官的工作負荷，提升審理效率並減少撰寫疏誤。然而，後續由於有下述之爭議，故最後僅有規劃而無進展。

(五) 量刑智慧分析系統

配合國民法官制度施行，於 112 年 2 月 6 日建置 AI 量刑資訊系統，藉由 AI 自然語言處理技術，在判決原始文字中標註相關量刑事由，並以自動擷取關鍵字的方式，搜尋及判讀加重、減輕法條的適用，且在標註過程中，作為機器學習訓練資料，取代人工標註方式，達到即時更新資料庫的效果，分成事實型模式與評價型模式，前者係從判決認定之犯罪事實揀選量刑因子進行歸納整理，使用者可以個案的事實內容進行

111 司法院審慎發展生成式 AI 應用新聞稿 2023/8/27，https://www.judicial.gov.tw/tw/cp-1887-929494-8a9fb-1.html, last visited on 2025/11/28.

查詢，目前計有殺人、搶奪暨強盜、傷害、不能安全駕駛、肇事逃逸、詐欺、竊盜、毒品、槍砲及妨害性自主案件，提供法官、律師及一般民眾查詢[112]；後者則從判決說明的量刑審酌事由，歸納係說明刑法第 57 條哪一款的量刑審酌事由，並研判係屬對被告有利、對被告不利的量刑因子，進行歸納整理，使用者可以個案量刑情形進行查詢，目前並未對律師或一般民眾開放使用。

二、我國司法關於 AI 應用之規範

(一) 沿革

行政院於 2023 年 8 月 31 日通過由國科會草擬行政院及所屬機關（構）使用生成式 AI 參考指引，而司法院公告數位 2.0 政策在前，且其中內容概述第 4 點又涉及 AI 生成判決之爭議問題，故引起同年 9 月 26 日全國律師聯合會、民間司法改革基金會、台北律師公會等民間相關司法及人權團體，聯合召開記者會，發表「AI 草擬判決的三大疑問與三大風險」聲明，要求召開公聽會，廣納各方意見並參考現行國際的標準，擬定使用規範後，再開放使用[113]。司法院翌日即以新聞稿分別就系統建置、訓練方式部分及避免生成內容錯誤之風險部分等 AI 技術層面加以說明，認為認定事實部分完全由法官自行決定並無使用生成式 AI 技術，系統亦未進行分析卷證，由法官認定有罪後，適用法律部分的例稿式生成內容則提供給法官參考。而系統的訓練資料均已去識別化，不含當事人個資，訓練資料包含起訴書及裁判書等資料逾 10 萬筆，應不至於有訓練資料不足的問題。另在使用生成式 AI 內容部分，為避免 AI 可能有生成內容錯誤的問題，在生成之後再結合規則法（Rule-based），以提高生成內容的準確性，應不會出現 AI 幻想之問題。此外，並承諾制定司法院使用生成式 AI 參考指引，而系統持續調整優化尚未正式上路等情[114]。

立法院於 114 年 12 月 23 日三讀通過人工智慧基本法，總統於 115 年 1 月 14 日公

112 司法院事實型量刑資訊系統，https://intellisen.judicial.gov.tw/, last visited on 2025/11/28.

113 民間司法改革基金會記者會 ── AI 草擬判決的三大疑問與三大風險要減輕負擔，也要控制 AI 風險，2023/09/26，https://www.jrf.org.tw/articles/2550, last visited on 2025/11/28.

114 司法院就民間團體 112 年 9 月 26 日召開「AI 草擬判決的三大疑問與三大風險」記者會所提建議之回應新聞稿 2023/9/27，https://www.judicial.gov.tw/tw/cp-1887-951341-9add3-1.html, last visited on 2025/11/28.

布，全法共計 20 條。就法律性質而言，本法定位為框架性、原則性之基本法，即軟法並無強制性規定，著重建立橫向治理架構而非針對特定產業或領域設計之專法，其第 1 條明定立法目的係「促進以人為本之人工智慧研發與產業發展，建構人工智慧安全應用環境…… 保障人民基本權利」。就規範效力層次而言，本法第 1 條後段，確立本法就 AI 事項相對於其他法律之優先適用地位。就核心規範精神而言，本法第 4 條揭示七大基本原則，作為政府推動 AI 研發與應用之指導準則，其一，永續發展與福祉原則，明定應兼顧社會公平及環境永續，並提供適當教育及培訓，降低數位落差，使國民適應 AI 帶來之變革，此為整體 AI 治理之社會基礎；其二，人類自主原則，明定應允許人類監督，落實以人為本並尊重法治及民主價值觀；其三，隱私保護與資料治理原則，明定應採「資料最小化原則」，妥善保護個人資料隱私，避免資料外洩風險；其四，資安與安全原則，明定 AI 研發與應用過程應建立資安防護措施，防範安全威脅及攻擊，確保系統之穩健性與安全性；其五，透明與可解釋原則，要求 AI 之產出應做適當資訊揭露或標記，以利評估可能風險，並了解對相關權益之影響；其六，公平與不歧視原則，明定應盡可能避免演算法產生偏差及歧視等風險，不應對特定群體造成歧視之結果；其七，問責原則，確保承擔相應之內部治理責任及外部社會責任。

本法在風險管理制度方面，雖第 2 條第 1 項規定中央主管機關為國家科學及技術委員會，但同條第 2 項又賦予各目的事業主管機關就所涉事項辦理之權，並且在第 16 條明定由數位發展部推動與國際介接之 AI 風險分類框架，各目的事業主管機關應循此框架訂定以風險為基礎之管理規範；第 17 條則就高風險 AI 應用明定責任歸屬及歸責條件，並應建立救濟、補償或保險機制。第 19 條更明定「政府使用人工智慧執行業務或提供服務，應進行風險評估，規劃風險因應措施」，並應依業務性質訂定「使用規範或內控管理機制」，此條明文適用於司法機關等所有政府機關，為司法院訂定 AI 使用規範提供明確的授權依據。此外，第 18 條課予政府應於本法施行後 2 年內完成不符本法規定之法規制（訂）定、修正或廢止，意味相關主管機關（含司法院）負有主動調適既有規範之義務。

然而，人工智慧基本法全文並無針對司法領域或司法機關就 AI 應用設置專節或專條規定，本法射程以行政推動、產業促進與基本原則宣示為主，法院運用 AI 所涉及之審判獨立、正當法律程序及證據法則等核心課題，性質上難以由此框架法單獨處理，

有賴司法院依職權另行補充或透過立法途徑建構司法專項規範。

(二) 指引架構

我國在人工智慧基本法生效前後，均係由各機關在其職權或所掌事務範圍內，為實現一定之行政目的，以不具法律上強制力之方法，促請特定人在 AI 事項為一定作為或不作為之行政指導，通常是以指引方式呈現。

1. 2025 年 3 月 7 日司法院及所屬各機關發展人工智慧參考指引

本次司法院的指引框架主要著重於「發展」人工智慧（AI）系統，並涵蓋了廣泛的 AI 應用，包括人工智慧、人工智慧系統及模型的定義。

⑴ 適用範圍及核心目標

針對司法院及所屬各機關發展人工智慧應用，應以資料整理及偵錯校對等輔助作業為主，並且人工智慧不得取代法官判斷證據價值、認定事實、適用法律或量刑等核心工作。

⑵ 人權、倫理與自主性原則

發展 AI 系統應落實以人為本，平等尊重所有人的基本人權與人性尊嚴。因此，必須避免訓練資料和模型存在種族、性別、宗教、階級、黨派等的偏見或歧視，以免影響人為判斷或決策的公平性，並且人工智慧系統應由使用者自主選擇是否使用其輔助，並在過程中得隨時終止使用。此外，也應重視風險管理，定期審查及更新。同時，應定期追蹤模型成效及更新訓練資料，以持續優化並避免產出過舊或不合時宜的內容。

⑶ 透明度與可解釋性要求

AI 系統開發者應適當揭露訓練及產出內容所依據的相關資訊，以確保使用者及受影響者能得知系統決策的要素，並且應致力於使產出內容具有可解釋、可測試、可驗證的性質，以減少人工智慧幻覺，並提供說明文件。因此應對資料收集、資料標籤以及使用情形進行適當記錄，並建立相關紀錄保存制度。

⑷ 安全與個人資料保護

AI 系統發展應依據《資通安全管理法》規定辦理，確保系統、軟體、演算法等技術運作環境的安全性。用於訓練模型及產出內容所依據的資料應受到適當保護，避免

未經授權之竄改、存取及使用。個人資料的蒐集、處理及利用，應遵守智慧財產權並符合相關法令規範，並建立適當的管理措施，以維護資料當事人權益。此後，AI 系統發展情形應提報司法院科技司法推動小組[115]核備後並公布會議紀要。

2. 司法院 115 年 1 月 27 日運用人工智慧工具應遵守公務機密維護相關規定函示

司法院於 115 年 1 月 27 日以院台廳司二字第 1150500116 號函，基於供不特定公眾免費或付費使用之商業性 AI 模型，縱宣稱或標示不將輸入資料作為模型訓練之用，但其於資料處理、傳輸、雲端儲存及運算過程，仍可能有資安、個資保護及機密外洩之風險，對於司法院所屬各機關人員於運用 AI 工具時，依法官法第 18 條、法官倫理規範第 16 條及公務員服務法第 5 條第 1 項等規定，負有守密義務[116]。基此，衡酌個案卷證資料具高度機密性，須避免機敏資料外流，使用 AI 工具時應審慎為之，不得將含有個人資料、機密資訊或未對公眾公開之卷證內容（如未公開之起訴書、書狀、筆錄、調查訪視紀錄、裁判草稿等），上傳或輸入至前揭商業性 AI 模型或相關平台。

3. 行政院及所屬機關（構）使用生成式 AI 參考指引

行政院的指引框架則聚焦於「使用」生成式 AI（Generative AI），著重於公務人員在執行業務時，如何管理這類技術帶來的特定風險。

⑴ 適用對象及核心目標

適用對象除行政院及所屬機關（構）人員外，尚包括向外採購得標之法人、團體或個人。另公營事業機構、公立學校、行政法人及政府捐助之財團法人，得準用本指引。此外，其他機關要訂定使用生成式 AI 規範時，亦可參照此指引內容。

使用生成式 AI 主要目的在於提升行政效率，同時要避免可能帶來的國家安全、資訊安全、人權、隱私、倫理及法律等風險。 AI 產出類似於人類製作的新內容，但其資料可能涉及智慧財產權、人權或業務機密之侵害。人員使用時應秉持負責任及可信賴之態度，掌握自主權與控制權，並秉持安全性、隱私性與資料治理、問責等原則。

115 原指引係提報司法 e 化推動委員會，司法院 101 年 12 月 12 日召開第一次司法 E 化推動委員會，見司法院數位政策（2018.07），頁 28。嗣於 115 年 3 月 30 日以司法院院台資一字第 1151200599 號函修正本指引第 14 點改提報科技司法推動小組。

116 司法院及各級法院就保密事項，訂有「司法院公務機密維護注意事項」、「司法院暨所屬機關主管應保密事項」、「臺灣高等法院公務機密維護作業要點」、「法院辦理涉及國家機密案件保密作業辦法」等規定。

⑵ 使用限制與機密保護（嚴格限制）

生成式 AI 產出資訊須由業務承辦人進行客觀且專業之最終判斷，不得取代業務承辦人的自主思維、創造力及人際互動。禁止使用生成式 AI 製作機密文書，包括國家機密文書及一般公務機密文書。業務承辦人不得向生成式 AI 提供涉及公務應保密、個人及未經機關同意公開之資訊。同時，也不得向其詢問可能涉及機密業務或個人資料之問題。例外若為封閉式地端部署的生成式 AI 模型，於確認系統環境安全性後，得依文書或資訊機密等級分級使用。

⑶ 產出內容的信任與決策依據

各機關不可完全信任生成式 AI 產出之資訊，故不得以未經確認之產出內容直接作成行政行為或作為公務決策的唯一依據。

⑷ 合法與揭露要求

使用生成式 AI 應遵守資通安全、個人資料保護、著作權及相關資訊使用規定，並且各機關使用生成式 AI 作為執行業務或提供服務輔助工具時，應適當揭露。

4. 司法院指引及函示與行政院指引之相異比較

⑴ 共同之處

二份指引及函示均確立「AI 僅為輔助工具」之核心定位，強調人的最終判斷權不可讓渡，AI 產出不得作為決策之唯一依據。在資安與個資保護方面，均要求遵守《資通安全管理法》及個人資料保護相關法令，並遵循智慧財產權規定。在透明度要求上，亦均要求對 AI 產出為適當揭露，並正視 AI 幻覺風險，要求使用者必須進行人工查核驗證。

⑵ 差異之處

司法院與行政院指引差異主要體現在規範範疇與禁止事項兩個面向。就規範範疇而言，司法院指引及函示分別針對 AI 系統之開發端（技術標準、模型管理、訓練資料偏見防範）與使用端（AI 工具之行為規範）進行規制，整體涵蓋範圍較廣；行政院指引則聚焦於公務員使用生成式 AI 工具之行為規範，以因應 ChatGPT 等商業工具普及所生之風險為主。就核心禁止事項而言，司法院指引之禁止紅線著重審判權核心，明定 AI 不得介入證據評價、事實認定、法律適用或量刑；行政院指引之禁止紅線則著重公

務機密保護，禁止以生成式 AI 撰寫機密文書，並明確區分一般 AI 服務與封閉式地端模型之分級使用條件，制度設計較司法院指引更為細緻。

三、台灣與世界各國運用 AI 規範之主要差異

(一) 規範層級與強制力之差異

1. 台灣

隨著人工智慧基本法於 115 年 1 月 14 日公布施行，我國 AI 規範層級已有實質提升。該法第 1 條確立其優先適用地位，並於第 16 條授權數位發展部推動風險分類框架，各目的事業主管機關應循此訂定以風險為基礎之管理規範；第 18 條更課予各機關應於施行後 2 年內完成法規調適義務。惟就司法領域而言，目前具體規範仍停留於行政規則層次，司法院現行參考指引均屬行政指導性質，尚未取得法律強制力。

2. 各國

美國透過 FRCP Rule 11 對虛構判例律師進行實質制裁，並正修訂《聯邦證據規則》擬增訂第 707 條；歐盟則以 EU AI Act 結合 GDPR 建立具備完整法律強制力之規範體系。整體而言，台灣雖已完成基本法之立法，司法專項規範之法律位階化仍落後於前述國家。

(二) 對於「律師與訴訟當事人」的監管密度

1. 台灣

人工智慧基本法及司法院、行政院現行指引均明定適用範圍為機關內部人員，對於外部使用者（律師、當事人）如何使用 AI 撰寫書狀或提出證據，迄今仍無具體規範或懲戒標準。

2. 各國

美國律師公會（ABA）已發布正式意見書明定律師查證義務；新加坡發布專門針對法院使用者（含律師、當事人）之指引，違反者得受支付訴訟費用或紀律處分；波蘭律師公會更發布涵蓋 AI 工具採購至客戶合約修改之三階段詳細指引。

(三) 證據法則與事實查核機制

1. 台灣

目前強調 AI 僅為輔助，不得取代法官認定事實，但對於「如何檢驗當事人提出的 AI 生成證據」或「AI 生成內容的證據能力」，尚無具體證據法則的修訂討論。

2. 各國

澳洲新南威爾斯州嚴格禁止以 AI 生成或潤飾證人陳述書及宣誓書；加拿大建議就 AI 證據舉行類似專家證人之可靠性聽證，並強制揭露 AI 系統錯誤率與訓練資料；美國擬議中之 FRE 第 707 條則要求機器生成證據須證明其輸入、運作流程及輸出之有效性。

（四）揭露義務的具體化程度

1. 台灣

人工智慧基本法第 4 條第 5 款僅原則性要求 AI 產出應做「適當資訊揭露或標記」，司法院現行發展指引雖亦要求適當揭露訓練及產出內容之依據，然具體操作標準（揭露位置、內容格式、強制與否）均付之闕如。

2. 各國

加拿大強制要求文件若含 AI 生成內容須於第一段明確聲明；巴西要求所有司法 AI 解決方案登錄國家平台並標明風險等級；澳洲規定專家報告若使用 AI，必須詳細揭露所使用之軟體版本及輸入提示詞內容。

四、我國現行 AI 運用規範之漏洞

綜合上述與世界各國比較，我國在司法 AI 規範上存在以下漏洞或不足：

（一）缺乏針對「外部使用者」（律師或當事人）的具體規範與懲戒機制

目前台灣的指引多是規範法院如何發展或行政機關所屬人員使用 AI。然而，國外案例顯示最大的風險往往來自「律師使用 AI 生成虛假判例」。若律師或當事人使用 ChatGPT 撰寫書狀並引用虛構判例，台灣目前缺乏類似美國 FRCP Rule 11 的明確制裁機制（如罰款、強制課程），僅能依賴法官一般的訴訟指揮權或建請法務部以違反律師倫理規範第 24 條懲戒，威嚇力不足。

（二）證據法則未與時俱進（缺乏 AI 證據的可採性標準）

隨著 Deepfake 與 AI 生成證據的普及，傳統證據法則可能不足以應對，缺乏針對「機器生成證據」的檢驗標準。法官面對 AI 生成的分析報告或影音證據，缺乏類似美國研擬中 FRE 第 707 條的「專家證人等級」審查架構，可能導致不可靠的 AI 證據進入法庭。亦未如澳洲明文禁止 AI 用於「證人證詞」轉譯可能導致證詞被 AI 修飾後失真，影響事實認定的真實性。

（三）缺乏風險分級與審查制度

台灣目前的規範較為籠統，僅劃分核心審判權不可由 AI 取代。未如巴西或歐盟建立「風險分級制度」（高風險 vs. 低風險）。例如，將 AI 用於「行政流程」與用於「犯罪預測」或「證據評估」應有不同密度的監管。缺乏類似阿根廷的「事前許可制」或巴西的「登錄制」，導致司法單位若引進外部 AI 工具，缺乏統一的審查與控管機制。

（四）透明度與揭露義務過於模糊

所謂使用 AI 適當揭露一詞過於空泛。未規定揭露的格式（如加拿大要求第一段聲明）、範圍（是否包含提示詞）及對象。這可能導致法官無法辨識哪些內容是由 AI 生成，進而無法進行必要的查核。

（五）缺乏對提示詞工程的規範與教育

南美國家（如阿根廷）的指引已細緻到教導法官如何下提示詞，以及禁止輸入哪些類型的提示詞。台灣指引僅強調資安與個資，對於如何正確操作 AI 以避免誘導式回答或偏見，缺乏具體操作層面的規範。

五、未來展望

針對未來台灣法院運用 AI 之管制，提出下列建議：

（一）引進醫療 AI 軟體風險矩陣概念進行風險分級化管理

醫療軟體中的風險矩陣概念原始用意為階層式風險管理，給予產生不同風險之醫療軟體核准適用不同之條件[117]，以醫療 AI 為例，因為運用可能直接影響病人生命、診

117 International Medical Device Regulators Forum (2014, “Software as a Medical Device”: Possible

斷、治療選擇，或者可能因資料偏差造成錯誤醫療行為，另部分為無法完全解釋之黑盒模型，因此，運用結果風險越高 → 要求越嚴 → 不得直接自動化決策，這種由低到高的階梯式管理就形成「風險矩陣」，舉凡美國 FDA [118]、歐盟醫療器材法 MDR [119]、日本 PMDA [120]，皆採風險分級（Risk-based）監管，醫療 AI 風險矩陣通常的兩大軸線：橫軸為使用情境（intended use）與縱軸影響程度（impact on clinical outcome），兩軸交錯每一格都要求不同強度的控管措施。而法院與醫療同樣具「高風險、不可逆後果」的特性，例如：錯誤判決會侵害人身自由、財產，甚至生命；錯誤採證可能影響事實認定，錯誤量刑可能造成重大權利侵害；錯誤引用法律損害司法信賴，法院運用 AI 之規範借用醫療 AI 的分級風險矩陣理念，應屬允當。

法院運用 AI 之風險矩陣細分為三個具體的監管區域（綠區、黃區、紅區）：

1. 低風險區（綠區）：行政輔助與形式審查

此區域涉及非審判核心的周邊行政事務，AI 作為「效率工具」，目標是節省人力時間成本。

⑴ 使用情境詳述

① 語言處理：法庭傳譯之即時語音轉文字（需人工校正）、涉外訴訟文書例如權利告知等初級翻譯。

② 文書行政：裁判書類格式自動排版、錯別字偵測、例行性公文生成（不含個資）。

③ 資訊檢索：法院內部行政規章檢索、單純法條查詢、庭務會議結論。

⑵ 影響程度分析

Framework for Risk Categorization and Corresponding Considerations, https://www.imdrf.org/sites/default/files/docs/imdrf/final/technical/imdrf-tech-140918-samd-framework-risk-categorization-141013.pdf, last visited on 2025/11/28.

118 FDA (2021.Jan), Artificial Intelligence/Machine Learning (AI/ML)-Based.: Jf/<X Software as a Medical Device (SaMD) Action Plan, https://www.fda.gov/media/145022/download, last visited on 2025/11/28.

119 REGULATION (EU) 2017/745 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 5 April 2017 on medical devices, https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32017R0745&qid=1779176737085, last visited on 2025/11/28.

120 REGULATION (EU) 2017/745 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 5 April 2017 on medical devices, https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32017R0745&qid=1779176737085, last visited on 2025/11/28.

此類應用不涉及對事實的實質認定或法律適用的價值判斷，錯誤通常顯而易見且易於更正（如錯字或排版跑掉），對當事人權益影響輕微，應屬於低度影響。

⑶ 必備控管機制

① 抽樣覆核機制（Random Audit）：雖無需逐件對照，但各法院應設立「品質管理員」，每月隨機抽取 5-10% 的 AI 產出文件進行準確度測試。若錯誤率超過 2%，應暫停該模型使用並進行校正。

② 標示與免責聲明：產出文件（如翻譯本）頁尾須註記：「本文件包含自動化翻譯轉寫內容，僅供參考，如有疑義以原文卷證為準」，明確劃分法律效力。

③ 資料安全：即便是行政資訊，亦須確保上傳至 AI 的過程符合資安規範。

⑷ 禁用點

禁止使用未經授權的公開雲端服務處理公務資料。

此區域之監管重點在於「準確度監控」與「個資防護」，在確保資安前提下，應鼓勵全面導入 AI 以提升司法行政效率。

2. 中風險區（黃區）：實質輔助

此區域為係 AI 在文字整理功能上最能發揮效能，但也因 AI 介入了資料的篩選與重組，具有潛在的引導性，亦成使用上灰色地帶。

⑴ 使用情境

① 卷證整理：筆錄摘要、證人及被告供述之整理分析、繁雜卷證之時間軸自動生成。

② 例稿草擬輔助：例稿式裁定（如補費或補正裁定）、支付命令或定型化簡易判決之初稿生成、類似事實判決之檢索。

③ 爭點整理：協助法官從大量卷證中歸納雙方攻防之不爭執事項。

⑵ 影響程度分析

AI 的摘要或卷證整理若有斷章取義或遺漏關鍵證據，可能對法官產生「第一印象」的錨定效應（Anchoring Effect），導致法官在後續審理中忽略了 AI 未提及的細節，間接造成誤判，因此運用 AI 造成中度影響之風險。

⑶ 必備控管機制

① 人機協作之實質審查

法官或助理不得僅進行形式瀏覽，必須執行「全案對照」。對於 AI 摘要的重點，必須親自閱讀原始卷證確認無誤，絕不可將 AI 摘要直接視為事實認定。

② 強制溯源連結（Citation & Linking）

系統功能必須強制要求：任何事實陳述或供述摘要，必須自動生成連回原始電子卷證的超連結或頁碼索引。無法提供來源的 AI 生成內容，視為無效資訊，不得採用。

③ 資料最小化與去識別（Data Minimization）

輸入模型之資料應僅限於案件爭點範圍，並且先經過去識別化處理後始得運算。若涉及性隱私、兒少保護等敏感案件，必須限制在法院內部封閉式內網環境運行。

⑷ 禁用點

禁止將 AI 生成之草稿直接「複製貼上」為最終判決理由。所有採用的段落，均須經過法官的邏輯修飾與語氣調整，轉化為法官之親自心證。

3. 高風險區（紅區）：決策輔助與核心認定

此區域觸碰司法權核心，涉及對人民生命、自由、財產的終局剝奪，必須採取最高標準的「防衛性監管」。

⑴ 使用情境

① 預測性分析：被告再犯風險評估（用於假釋審查）、量刑區間具體建議。

② 價值判斷：證據證明力之評價（例如判斷證言可信度）、法律適用之推論（例如判斷行為是否構成「猥褻」）。

⑵ 影響程度分析

若 AI 模型隱含訓練數據中的種族或性別偏見，將直接導致歧視性判決；且 AI 的「黑箱性」可能剝奪當事人的防禦權（無法反駁演算法如何得出結論），具有侵害人權風險之高度影響。

⑶ 必備控管機制

① 可解釋性要求（Explainability）

任何用於此區的 AI 系統，必須具備能說明其運算邏輯、參考因子權重（Feature Importance）。若系統僅能給出分數而無法解釋理由，則不符合正當法律程序（Due Process）。

② 強制人工介入與獨立心證

法官必須獨立撰寫心證理由。 AI 之輸出僅能列為參考資料之一，地位等同於學術論文或參考文獻，不得作為判決之唯一或主要依據。判決書中不得出現「經 AI 系統計算，被告再犯風險高…」等將責任轉嫁給系統之論述。

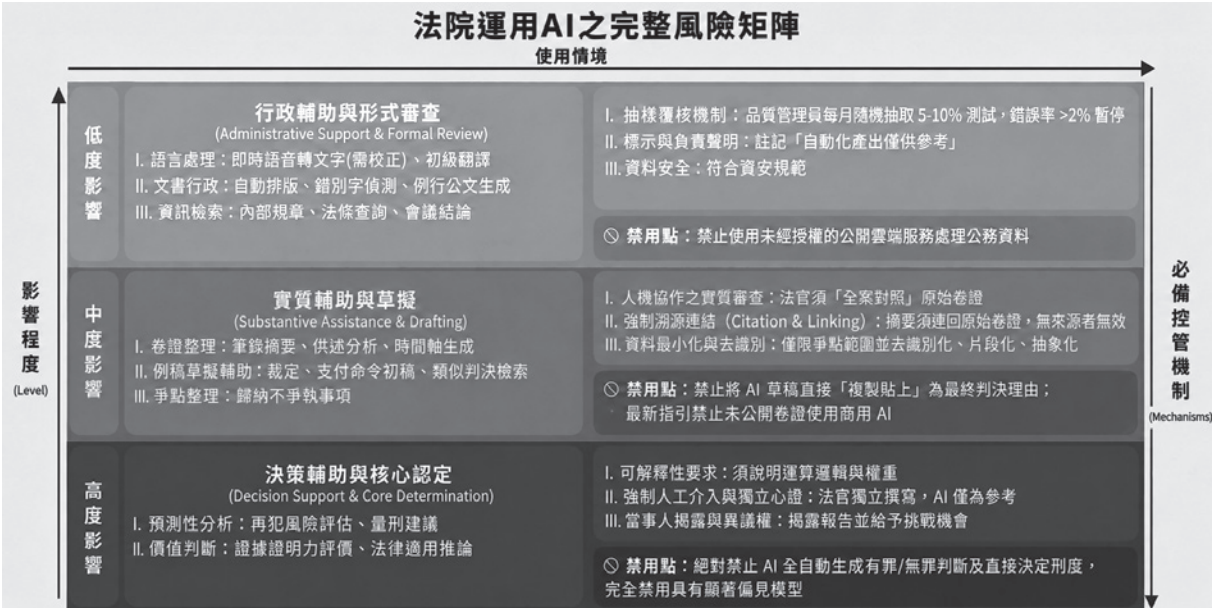
③ 當事人揭露與異議權

若法院在量刑審查中參考了 AI 評估報告，必須向當事人與辯護人揭露該報告內容，並給予辯方挑戰該演算法準確性與適用性的機會。

⑷ 禁用點

絕對禁止 AI 全自動生成有罪或無罪之判斷及直接決定具體刑度（如直接生成有期徒刑五年）。任何被證明具有顯著演算法偏見的模型，應立即列入負面清單，完全禁用。

附表一



(二) 建立 AI 使用紀錄表：數位履歷與程序正義之擔保

AI 紀錄表之設計目的，在於建立法院運用 AI 的「數位監管鏈（Digital Chain of Custody）」。面對未來可能的上訴或當事人異議，此紀錄將成為證明法官仍保有獨立審判意志且未受 AI 偏見誤導的關鍵程序證據。

1. 基礎資訊、風險辨識與必要性審查

此區塊要求使用者在啟動 AI 前，先進行「合法性自省」，確保使用行為符合比例原則與資安規範。

⑴ 案件識別與風險矩陣分級

使用者須強制勾選本次 AI 操作對應之風險等級（低風險行政、中風險摘要、高風險決策），此舉旨在建立「強制性停頓」，迫使法官或書記官在操作前自我警示，若勾選「高風險」，系統應自動跳出警語，提醒需進行高強度的邏輯檢驗與人工介入，避免法官或書記官在忙碌中無意識地過度依賴系統。

⑵ 輸入資料必要性審查

欄位設計應包含「是否為達成目的所必要」之檢核。例如：若僅是查詢法條，則無理由輸入詳細案情；若僅是排版，則不應輸入任何當事人個資。落實 GDPR 與歐盟 AI 法案強調的資料最小化原則，防止使用者為了貪圖方便，將整個卷證不相關的敏感個資（如醫療紀錄、前科表）均拋給 AI 處理，從源頭降低資安外洩風險。

⑶ 個資遮蔽與資料合法確認

針對輸入資料是否包含未公開卷證或敏感個資（如性侵、家事、少年案件），需具體記錄處理方式（例如：已執行去識別化程序、使用代號替換、或確認使用封閉式離線模型）。

2. 模型履歷與提示工程

此區塊記錄「人機互動」的過程，確保產出的邏輯並非源自於使用者的誘導或模型本身的缺陷。

⑴ 模型版本與參數歷程

詳實記錄使用的模型具體版本（例如 GPT-5.1, Gemini3 或法院自建模型 v2.1），由於不同版本的模型具有不同的幻覺率與偏見傾向。若未來發現某版本模型存在系統性錯誤，法院可依此紀錄快速回溯並檢視所有受影響的案件，進行必要的補救措施。

⑵ 關鍵提示語摘要與中立性檢核

記錄對 AI 下達的關鍵指示詞，重點在於檢視是否存在「誘導式提問」（Leading Questions），例如「請幫我寫一個判被告有罪的理由」與「請分析本案正反證據之強

弱」有本質區別。藉此防止人為偏見（Human Bias）透過 AI 的運算能力被放大（Bias Amplification）與合理化，確保 AI 是作為客觀的分析工具，而非特定心證的背書機器。

3. 輸出驗證、錯誤分類與參與程度

此區塊為紀錄表的核心，詳述人類如何審查 AI 產出，並量化 AI 對最終決策的影響力。

⑴ 標準化錯誤檢查類型

覆核時若發現 AI 產出有誤，不應只標記「錯誤」，應細分為：（A）事實錯誤、（B）邏輯推演錯誤、（C）幻覺（虛構判決或法條）、（D）偏見或情緒化用語、（E）引用來源錯誤。透過長期累積此類結構化數據，法院能建立「本土化 AI 錯誤資料庫」，分析哪類案件最容易導致模型犯錯，進而作為優化模型參數或加強法官教育訓練的實證基礎。

⑵ AI 參與程度量化指標

參考學術倫理 [121] 與歐盟規範，設定 0-4 級的參與度指標：

Level 0：無 AI 參與。

Level 1：僅作表面修飾（潤飾、除錯）。

Level 2：提供資料整理與摘要，未改變結論。

Level 3：提供部分推論或法律適用建議（須註明人工高度覆核）。

Level 4：採納 AI 之推論邏輯（此級別需詳細說明理由，且通常僅限於低風險事項）。

讓判決的「含 AI 量」透明化。在爭議案件中，上級審或外界可依此判斷原審法官是否實質履行審判職責，或是僅作為 AI 的橡皮圖章。

4. 揭露狀態與最終課責

此區塊處理 AI 產出與外部當事人的關係，確立程序透明度：

⑴ 對當事人揭露狀態

針對 AI 生成之內容（如摘要、時間軸、草擬書狀），若有提供給當事人閱覽或作

121 Suchikova, Y., & Tsybuliak, N. (2025). ChatGPT isn't an author, but a contribution taxonomy is needed. Accountability in Research, 32(8), 1590, https://doi.org/10.1080/08989621.2024.2405039.

為訴訟攻防基礎，需明確標示「已向當事人揭露」或「僅供法院內部參考」。避免當事人誤認 AI 生成之錯誤資訊為法院之正式認定。若未經揭露而使當事人產生信賴，可能構成突襲性裁判或程序瑕疵，引發國家賠償責任。

⑵ 引用來源與人工覆核簽署

系統生成內容所引用之法條、判決或學說，必須列出可驗證之清單。最後由承辦人員（法官 / 助理）簽名並押註時間，確認已針對關鍵事實與法律見解完成查核。確立「人是最終決策者」的倫理底線。無論 AI 技術如何先進，法律責任的主體始終是自然人法官，此簽署行為象徵著對判決品質的最終擔保。

附表二　AI 紀錄表

一、基礎資訊及風險辨識

欄位	內容
案號	
AI 使用目的（找法條／摘要證詞／文字修飾等）	
風險分級（低／中／高）	
是否必要使用	☐ 是（理由：）　☐ 否
輸入資料類型	☐ 公開資訊　☐ 未公開卷證 ☐ 敏感個資
是否已去識別化或使用封閉模型	☐ 已遮蔽　☐ 使用封閉模型 ☐ 否（須說明）

二、模型資訊與提示語內容

欄位	內容
使用模型與版本	
提示語摘要	
提示語是否可能造成誘導	☐ 是（需說明）　☐ 否
模型設定（若適用）	

三、**AI 輸出驗證與錯誤檢查**

欄位	內容
AI 輸出摘要	
錯誤檢查類型（可複選）	☐ 事實錯誤　☐ 邏輯錯誤 ☐ 虛構內容　☐ 偏見或情緒化語言 ☐ 引用來源錯誤
AI 參與程度（0–4）	Level：___（理由：）
引用來源列表	
人工覆核方式與結論	
覆核者與時間	簽名：_____　時間：_____

四、**對當事人揭露情形**

欄位	內容
是否提供 AI 產出內容予當事人	☐ 是（理由：）　☐ 否
是否附上 AI 使用限制聲明	☐ 是　☐ 否
是否可能影響當事人攻防（需備註）	

（三）法院內外部人員使用 AI 規範重點差異化

1. 當事人與律師（外部使用者規範）

針對外部使用者，重點在於將 AI 使用納入律師倫理規範第 8 條及第 24 條誠實原則及民事訴訟法第 249 條第 1 項第 8 款善意原則的範疇，並透過標準化格式與分級制裁，確保書狀品質。

⑴ 強制性揭露與標準化格式

凡書狀內容（含事實陳述、法律論證）有實質部分由生成式 AI 產出，均應主動揭露。為避免格式不一造成審判實務困擾，法院應提供制式範本，律師須於書狀首頁或該段落明確標示：「【AI 使用揭露】本書狀第〇〇段至第〇〇段經 AI 協助草擬，已由提出人逐項查證並負最終法律責任。」。統一格式的好處能大幅降低法官在審閱卷證時的認知負荷，並防止律師透過模糊用語（例如僅在末頁小字註記）試圖規避揭露義務，確保對造當事人與法院均能明確知悉書狀的生成來源。 AI 使用者揭露背後即有三大具

體保證事實：來源真實性、法源引用正確性及無誘導提示。日後若發生爭議，法院可據此認定律師或當事人可能具有「明知或可得而知」之故意或重大過失。

⑵ 程序法效力與違規制裁

在驗真手段未能追上偽造技術的推陳出新，唯一能仰仗即為當事人自我揭露，旨在維護司法程序的誠信與公正。若未揭露運用 AI 生成之文書，在民事上則應認為未盡書證提出之責，可能駁回該主張或認他造關於依該文書應證之事實為真實[122]；在刑事上則視 AI 生成部分是否已逸脫真實而構成偽證。此外，參酌美國 Mata v. Avianca 案[123]之法理，初次違規可能源於對技術不熟悉（過失）；但若同一律師或事務所發生第二次以上引用虛構案例或法條，則推定為「重大過失」或「蓄意隱瞞」，除移送律師懲戒委員會外，法院得裁定命其負擔因此所生之無益訴訟費用，或處以程序性罰鍰。

2. 法院內部人員（內部內控）

針對法院內部人員，重點在於將人的監管嵌入工作流程，並透過分類管理降低資安風險。

⑴ 卷證分類標示制度

建立嚴格的資料分級制度，嚴禁 AI 自動辨識處理所有卷證：

A 類（一般）：去識別化後可交付 AI 處理（如一般民事糾紛）。

B 類（需人工遮蔽）：需經人工重度遮蔽後始得輸入。

C 類（絕對禁止）：涉及國安、營業祕密、性侵、兒少、毒品線民、偵查中機密等高敏感資訊，物理上禁止輸入任何連網之 AI 系統。

透過源頭分類，解決「人為判斷失誤」的風險。 C 類資料應在法院封閉地端 AI 系統權限上直接設定阻擋，確保即便操作人員疏忽，系統也能發揮最後一道防線功能。

⑵ 即時課責與紀錄留存

法院使用 AI 系統應建置後端監控機制，自動記錄使用時間、模型版本及提示詞內容。若當事人對法官判決之 AI 參與程度之揭露提出具體質疑，承辦法官須於 24 小時內提出完整的「AI 使用紀錄表」。對於無法提出紀錄者，將被推定為違反內控規定，

122 民事訴訟法第 345 條第 1 項。

123 前揭註 48。

此舉能對法官產生強大的心理制約，避免便宜行事。

⑶ AI 風險事件回報系統

建議仿照醫療體系的「異常事件通報（Incident Reporting）」及歐盟人工智慧法制度，建立跨審級的風險回饋迴路。法院應設立「AI 風險通報平台」，允許法官、書記官、律師甚至當事人，針對 AI 的錯誤（幻覺）、偏見（歧視性輸出）或誤導性建議進行通報（可選擇匿名）。通報之資訊有助於：

① 早期預警

當多個案件回報同一模型出現特定法律見解錯誤時，司法院可即時發布警示，暫停該模型使用。

② 教育訓練素材

彙整通報案例建立「典型違規與錯誤案例庫」，作為年度教育訓練的實戰教材，讓使用者從他人的錯誤中學習，避免重蹈覆轍。

③ 系統優化依據

累積的錯誤數據將成為未來採購或微調司法專用模型的重要依據。

（四）AI 生成證據之可信性檢驗 ── 從舉證責任分配至準用鑑定法則

針對 Deepfake 與 AI 增強證據（如經 AI 降噪之錄音、畫質增強之影片）日益普及，法院須即早做好因應之道。本文認為，我國應借鏡美國聯邦證據規則（FRE）第 707 條修正草案之精神，並結合我國刑事訴訟法第 206 條之鑑定法理，建立 AI 證據的檢驗標準：

1. 舉證責任的動態分配：由「挑戰者舉證」轉向「提出者擔保」

傳統證據法則下，主張證據虛偽者負有舉證責任。然而，AI 生成內容具有不透明性與高擬真性，若仍要求對造證明該證據為假（例如證明某影片是 Deepfake），將使防禦權陷入極度不對等。因此，本文建議應參考美國 FRE 第 901 條修正討論中關於 Deepfake 的舉證責任倒置精神，確立加重提出者舉證責任之原則，也就是凡欲將 AI 生成或處理之資料作為證據進入法庭者，負有證明其「真實性」與「可靠性」之首要責任。提出者必須主動揭露是否使用了 AI 工具及其具體版本、參數。若未揭露而被發現，法院得直接推定該證據不具證據能力。

2. 實質準用刑事訴訟法第 206 條：AI 即「數位鑑定人」

AI 生成或處理之證據，已無從依傳統證據法則將其視為單純物證或文書或新型態之數位證據，以避免審查密度不足。本文建議援引美國聯邦證據規則擬議新增第 707 條之立法理由，並結合我國刑事訴訟法第 206 條鑑定報告之要求，重建 AI 證據的可信性檢驗標準：

⑴ 借鏡 FRE 第 707 條草案立法脈絡防止機器外觀規避專家審查

美國聯邦證據規則諮詢委員會在擬議 FRE 第 707 條時指出一個關鍵漏洞：現行實務上，證據提出者常直接提出機器輸出結果，卻未傳喚專家證人，試圖規避 FRE 第 702 條關於專家證詞的可靠性審查。委員會強調，如果某項證據是由人類提出會被視為專家意見（需具備專業知識與方法），那麼由機器生成同樣內容時，亦應受到同等標準的檢驗。簡言之，不能讓 AI 成為逃避科學檢驗的後門。

⑵ 定性 AI 為「數位鑑定人」：實質準用刑事訴訟法第 206 條

我國刑事訴訟法第 206 條第 1 項規定，鑑定之經過及其結果應以書面報告。此條文之精神在於確保鑑定過程的透明性與「可驗證性」。2023 年 12 月 15 日修正增訂第 206 條第 3 項規定於立法理由明定參考美國 FRE 第 702 條規定而來，美國檢驗所謂科學證據或專家證人所採科學鑑定方法，有採道伯法則（Daubert test）或佛萊法則（Frye test）二種標準 [124]，AI 模型（特別是涉及深度學習者）在處理證據時，並非單純的物理紀錄（如溫度計），而是進行了複雜的特徵提取、權重計算與機率推論。此過程在功能上完全等同於人類專家的鑑定行為。因此，論理上應將具備推論性質的 AI 輸出視為「準鑑定報告」，進而要求提出者必須履行第 206 條之義務。

⑶ 具體審查方法：以符合同儕接受具可解釋性之演算法過程作為證據能力門檻

基於上述脈絡，對於 AI 證據之可信性判斷，應由提出者證明該 AI 工具已滿足類似鑑定報告的標準：依據 FRE 第 707 條諮詢委員會意見，法院分析重點應包含「AI 輸入（Prompt / Input）」、「內部流程（Process）」及「輸出有效性」。對應至我國法制，提出者不能僅提出 AI 的結論（例如：照片中人像為被告），而必須詳細說明訓練資料的來源與偏差檢測、模型運作的科學原理及邏輯、以及該模型是否經過同行評審或具

124 廖建瑜，如何理解鑑定之證據方法 — 因應國民法官制度之挑戰（下），司法周刊第 2060 期，110 年 6 月 25 日，頁 2。

有公認的錯誤率。若提出者無法交代上述演算法，則等同於鑑定報告必要記載內容記載不全，並且如同鑑定人須接受詰問，AI 系統的開發者或操作專家，應有義務出庭說明，若未到庭，依實務見解應排除其證據能力。

透過引入 FRE 第 707 條草案「防規避」的思考脈絡，當 AI 執行了鑑定的功能，就必須符合刑事訴訟法第 206 條要求，以確保了司法審判不被無法解釋的演算法所綁架，方能避免演算法黑箱成為誤判的溫床。

（五）突破無封閉式地端 AI 模型前的卷證分析替代流程

在卷宗數位化與電子卷證普及的情況下，生成式 AI 可協助法院進行摘要、比對、關鍵字標示等作業，確實具有提高效率之潛力。然而，行政院就政府機關使用生成式 AI 所發布之相關指引，已明確提醒各機關不得將涉有個人資料、機密資訊或敏感案件內容，逕行輸入商業雲端 AI 服務，以免違反個人資料保護法、洩漏偵查與審判機密或侵害第三人權益。司法機關作為握有大量高度敏感資訊之機關，更應採取比一般行政機關更高標準之謹慎義務。

在尚未建置封閉式地端模型或政府專用雲端 AI 服務之前，若確有借助外部生成式 AI 進行文字整理、語言潤飾之必要，至少應同時採取「匿名化、最少資料、片段分割、抽象化」四項最低減少風險措施：

第一，匿名化處理：先依本文前述步驟，將電子卷證經 OCR 後，於 Word 檔內使用「取代」功能，將當事人、證人、醫師、醫院名稱等可辨識個人或機構資訊，以「A、B、C」或其他代碼全數替換，並刪除住址、身分證字號、病歷號碼等敏感標識，使單一段落文字無法直接指向特定自然人或案件。

第二，最少資料原則：不得為「方便」而將全部卷證全文上傳，而應依實際分析目的，僅擷取必要段落，例如僅上傳供述中涉及時間序列之部分，或僅上傳醫療紀錄中與爭點因果關係相關之片段，避免將與本次分析無關之隱私或機密一併外洩。

第三，片段分割抽象化處理：即使已作匿名化與擷取，仍應將文字內容切割為數個互相難以還原之片段，並予以抽象化，上傳時分批送出，避免單一服務提供者可能取得足以重構案件全貌之完整敘事。

綜上所述，在法院尚未擁有自主掌控的封閉式 AI 基礎設施前，對內應確立「不得

直接將原始卷證上傳商業雲端 AI」之禁止性原則；對外如有利用生成式 AI 之必要，則必須經過匿名化、資料最小化與片段分割抽象化等前處理程序，並留存操作紀錄，以便日後就可能發生之外洩或誤用事故追溯責任。如此方能在有限條件下兼顧審判效率與資訊安全，避免 AI 工具成為破壞訴訟秘密與隱私保障之新風險來源。

附表三　**AI** 卷證摘要合法流程



（六）115 年 4 月 22 日司法院及所屬各機關使用人工智慧系統參考指引

本文完稿後，司法院於 115 年 4 月 22 日以院台廳司二字第 1150500529 號函，正式訂定司法院及所屬各機關使用人工智慧系統參考指引（下稱最新指引），分為三章共 18 點，自即日起生效。此一最新指引其特點在於聚焦於使用端行為規範，適用對象為使用 AI 系統之司法院及所屬各機關人員，而非 AI 系統之開發部署，並同步發布司法院使用 AI 指引實務 Q&A（現階段過渡期版）」（下稱 Q&A），以 16 則問答填補指引之操作細節。以下介紹最新指引內容，並藉以印證本文建議方向之異同。

1. 最新指引內容逐點析論

⑴ 總則（第 1 至 3 點）── 目的、定義與適用範圍

指引訂定目的為建立「提升效率」與「防範風險」（國安、資安、人權及倫理）並

重的雙軸框架（第 1 點要旨）。另就指引中所使用 AI 系統定義為具備推理、學習、預測能力之運算系統；生成式 AI 定義為依提示詞生成文字、圖像或音訊之系統；而所謂「使用」則被定義為：執行公務時，透過軟體介面、應用程序或輸入提示詞運用 AI 之行為（第 2 點要旨）。而本指引適用對象則為司法院及所屬各機關人員（第 3 點）。

⑵ AI 素養要求（第 4 點）

第 4 點要求使用者在使用前具備「人工智慧基本知識、素養與能力，以充分瞭解系統之功能限制、倫理規範及潛在之法律責任」。Q&A 第 7 題對此進一步詮釋，指出素養並非會操作 AI 即可，而是應對 AI 三項限制有基本認識：其一，生成式 AI 非法律資料庫，訓練資料有截止日期，引用之判決見解、條文可能有誤、已過時或已遭廢棄；其二，AI 無法替代法律人的專業判斷，其結論僅為參考意見，非權威見解；其三，AI 系統具有幻覺現象，可能自行捏造案號、法條或引文。此一規範與本文對 AI 幻覺之系統性分析若合符節，也與本文所建議之「AI 風險通報與錯誤資料庫建置」之教育訓練面向一致。

⑶ 使用系統之白名單限制（第 5 點）

第 5 點明定使用者應使用「各機關內部建置或經本院資訊安全推行小組核准或核備之人工智慧系統」，即採取正面表列的白名單制。Q&A 第 2 題進一步說明，商業 AI（含企業版、商業版、付費版）均屬須經資安小組核准方可使用之類別，使用者不得自行認定其安全性；在資安小組正式評估並核備該特定企業版軟體前，使用者不得自行認定其安全性而將公務機密資料輸入。此一制度與本文所建議之「卷證分類標示制度」中，對 C 類（絕對禁止）資料適用物理阻擋機制的精神一致。

⑷ 資料輸入之絕對禁止（第 6 點）

第 6 點構成指引之核心禁止條款，明定涉及以下三類資料，禁止輸入、上傳或以任何形式提供予非經核准之 AI 系統：（甲）受法律保護之機密資訊、（乙）個人資料，以及（丙）未對公眾公開之訴訟資料（包括未公開之起訴書、書狀、筆錄、法庭錄音、調查訪視紀錄、裁判草稿等）。Q&A 第 1 題就最常見的實務問題作出明確回應：「去識別化後能否上傳商業 AI ？」回答是「目前不行」。理由有二：其一，商業 AI 多為雲端運算，資料會傳輸至外部伺服器，即使去除姓名，「案件獨特的犯罪脈絡仍可能導致洩密，損及司法信譽，並涉及資通安全及個人資料保護法規之違反」；其二，也是更根

本的原則性立場：未公開之訴訟資料具有保密性，不因去識別化而有不同[125]。

⑸ 人類最終判斷之不可替代性及終極責任（第 7 點及第 13 點）

第 7 點宣示：「使用者應瞭解人工智慧系統僅為輔助工具，其產出之結果，應由使用者本於知識與專業為最終判斷，不得取代使用者之自主思維、獨立判斷與核心決策」。Q&A 第 5 題及第 6 題對此加以引申：法官使用 AI 輔助分析公開判決見解後，仍負完全且最終之責任，核心職責不可讓渡，免責事由無效。依指引第 13 點「使用者不得以『系統產出如此』作為免責或減輕責任之事由」。此一三位一體原則（輔助工具定位、核心職責不可讓渡、最終責任不可轉移），正是本文主張人機協作 AI 運作框架，彼此在制度精神上完全吻合。

⑹ 幻覺查核義務（第 8 點）

第 8 點要求：對於生成式 AI 產出之「數據、事實整理及法律引用等」，使用者應與正式文件、卷證資料或法學資料搜索等可靠來源進行交叉比對與驗證後，始得運用。Q&A 第 9 題進一步說明，AI 可能引用已修正廢止之法律條文，或已遭廢棄之判決見解，且其引用格式看似正確，難以一眼辨識，因此「務必交叉比對現行法規與最新判決，不可逕行採用 AI 生成的法條或判決引用」。此一查核義務，與本文 AI 使用紀錄表之「強制溯源連結（Citation & Linking）」設計任何事實陳述或供述摘要必須自動生成連回原始電子卷證之超連結或頁碼索引，在制度目標上完全一致，均是以程序性機制確保 AI 產出不淩駕於原始文件之上。

⑺ 偏見與不透明風險之警示（第 9 點）

第 9 點要求使用者瞭解生成式 AI「具備運算過程不透明、難以驗證之侷限性，產出內容亦可能隱含偏見或過時資訊」，並應注意避免產生偏見、歧視或不公平之結果。Q&A 第 8 題進一步指出三類高風險場景：（甲）涉及特定族群、性別、宗教、國籍背

125 本文完稿後歐州司法效能委員會（CEPEJ）於 2025 年 12 月 19 日發布最新指引，在第 27 點至 29 點、明確指出使用商用 AI 的風險（境外資料傳輸、模型再訓練、缺乏法律專業訓練）建議優先採用客製化安全工具，而非通用現成解決方案；以及第 67 點指出使用生成式 AI 時，所有技術與資料都必須保留在司法機關的排他性控制之下，以防止未經授權的存取或外部干預，並維護機構對資訊的主權。上開建議與司法院最新 AI 指引方向相符，參見 European Commission for the Efficiency of Justice (CEPEJ). (2025). Guidelines on the use of generative artificial intelligence in courts (CEPEJ(2025)18Final). Council of Europe, https://rm.coe.int/cepej-2025-18final-en-draft-guidelines-on-the-use-of-generative-ai-for/48802a4ad1, last visited on 2026/05/19.

景之案件分析；（乙）使用 AI 生成量刑建議參考或風險評估；（丙）以 AI 分析當事人行為模式或動機。此三類高風險場景之點名，與本文介紹演算法偏差（Algorithmic Bias）及 COMPAS 系統種族偏見問題，以及本文紅區即高風險之決策輔助與核心認定管制對象，在識別邏輯上高度吻合。

⑻ 審判核心職責之保留（第 10 點）

第 10 點構成指引在審判業務中最重要的規範紅線：「於審判業務使用人工智慧系統時，應以資料整理、摘要分析、語言翻譯及校正核對等輔助作業為限，不得取代使用者對於證據評價、事實認定、法律適用或量刑等核心職責」。 Q&A 第 4 題對此操作界線作出說明：「已公開」的判決書可利用 AI 進行分析；但針對「尚未公開」的卷證、筆錄、草稿，則禁止進入外部商業 AI。此一對應關係，使本文所建議之 0 至 4 級量化指標，具有了更清晰的操作邊界錨點。

⑼ 錨定效應之預防（第 11 點）

第 11 點要求使用者對 AI 產出「應保持客觀中立，進行獨立且實質之評價，並應避免因受產出內容之影響，對案件事實、證據評價產生預斷」。 Q&A 第 10 題對此進行了頗具深度的補充說明，明確點名兩種心理機制：「錨定效應」AI 若先給出某一傾向性結論（如被告可能有罪），法官後續在評估證據時，可能在不自覺中受此錨點影響；以及自動化偏誤即對系統產出的過度信任，導致審查力度下降，並建議「在使用 AI 進行資料整理後，應先回歸原始卷證進行獨立閱卷，再形成對案件的初步評估，避免以 AI 結論作為閱卷的『預設框架』」。此一官方表述，是台灣司法規範文件中首次正式援引認知科學研究成果，與本文對錨定效應之深度學術分析相互呼應，也是本文禁止將 AI 生成之草稿直接作為最終判決理由這一黃區禁用點之更具體的心理學詮釋。

⑽ 揭露標記義務（第 12 點）

第 12 點明定「使用生成式 AI 作為執行公務之輔助工具所生成之資料，應為適當之揭露或標記」。 Q&A 第 11 題說明揭露範圍：行政文書（如簡報、簽呈）建議於備註欄或附記中說明「本文件部分內容曾借助生成式 AI 進行草擬，經人工審核修訂」；裁判書目前指引未明定強制揭露格式，但應遵循後續可能發布之具體規範。 Q&A 第 12 題則從實務可責性角度建議：雖指引未明文要求，但建議至少保留重要之提示詞與輸

出結果，以利事後稽核、品質控管，並在發生爭議時提供使用脈絡之說明。此一揭露義務與本文 AI 使用紀錄表之設計目標一致，然就適當揭露之操作標準而言，指引仍屬原則性宣示，尚未提供量化、可比較之揭露格式。本文所設計之 0 至 4 級 AI 參與度量化指標，恰可填補此一缺口，它為「適當揭露」提供了具體的表達媒介，在爭議案件中，此一指標將成為事後稽核原審是否實質履行審判職責的關鍵依據。

⑾ AI 素養訓練之機關義務（第 17 點）與定期陳報（第 18 點）

第 17 點明定本院及所屬各機關「應提供人工智慧基礎知識、倫理規範與風險管理之教育訓練或相關學習資源」，Q&A 第 16 題確認此為機關之「積極義務，而非選擇性辦理」，訓練內容涵蓋 AI 系統基礎知識（含功能限制與幻覺現象）、使用倫理規範（含偏見識別與人格權保護）及風險管理（含資安要求與通報機制）。第 18 點要求各機關定期向資訊安全推行小組陳報 AI 使用情形。此二點與本文所建議之「AI 風險事件回報系統」精神一致。

⑿ 機關自訂規範（第 16 點）

第 16 點規定：「本院及所屬各機關得視其業務屬性與需求，依據本參考指引，訂定使用規範或內控管理措施」。 Q&A 第 15 題進一步說明：本參考指引所定為「全院最低標準，各機關之自訂規範可能在特定業務面向設有更嚴格之限制」；使用者切勿以「本指引未禁止」作為自行擴大使用範圍之依據。故指引授權機關訂定更嚴格內部規範。

2. 指引與本文建議差異分析 ── 未公開之卷證商業 AI 絕對禁用

本文前開所提出之過渡性操作方案，即法院尚無部署封閉式地端 AI 模型前，透過匿名化、最少資料、片段分割、抽象化四項措施降低風險，借用外部商業 AI 處理卷證，實與指引 Q&A 第 1 題之明確立場存在實質歧異，且此一歧異無法透過第 16 點機關自訂規範加以調和，蓋機關自訂規範僅得更嚴格，而非放寬指引之禁止。最新指引之立場係商業 AI 對於未公開卷證之禁用，係絕對性的，不以技術去識別化程度為條件，未公開之訴訟資料具有保密性，不因去識別化而有不同。就制度原因而言，此一立場有三層根據：（甲）技術層面，去識別化並不能消滅案件脈絡的可識別性，尤其是媒體曾報導之重大案件；（乙）法律層面，訴訟資料之保密義務係基於法官法第 18 條

及法官倫理規範第 16 條，此等義務之性質並不因技術處理而改變；（丙）政策層面，司法機關承擔著公眾對司法信任之責任，任何因便宜行事而導致的資訊外洩，對司法公信力之損害均難以彌補。就本文的學術立場而言，過渡性方案之提出，係基於封閉式地端 AI 模型系統短期內無法到位，而審判效率需求迫切的現實性預設。然而，最新指引之嚴格立場即寧可暫時損失效率，亦不可承擔卷證外洩的制度性風險。本文原提出之過渡性操作方案，在最新指引正式確立絕對禁止立場後，其適用範圍應重新定位為：僅適用於已公開資料之輔助整理流程（如公開判決之摘要與比較），不得作為處理未公開卷證之任何授權依據。

3. 指引未觸及之面向

除上述差異之外，以下面向在指引中完全付之闕如，本文之分析具有獨立的制度補充意義：

其一，外部律師 AI 書狀揭露義務及虛構判例懲戒機制。律師向法院提出含有 AI 生成之虛假引用，現行指引完全未加以規範，需透過修訂律師倫理規範及訴訟程序法（民事訴訟法第 249 條、刑事訴訟程序指揮權）另行建構。

其二，AI 生成或處理證據之可信性審查架構。對 Deepfake 影片、AI 增強證據及 AI 輔助專家意見，指引未設任何規範，本文所建議之「準鑑定法則架構」（實質準用刑事訴訟法第 206 條及參照美國 FRE 第 707 條草案精神），係現行制度空白中最具操作性的填補方案。

其三，量化的 AI 參與度揭露標準。指引要求「適當揭露」但未定義「適當」之量化標準。本文所建議之 0 至 4 級指標，可在不增加任何法律強制義務的情況下，以自願性操作指引形式向各法院推廣，並配合第 16 點授權逐步轉化為機關內部之強制要求。

（七）結語：從指引原則到制度深化

本文完稿之際，台灣司法 AI 治理的制度地圖已發生根本性的改變。繼人工智慧基本法生效之後，司法院接續以函示及參考指引，宣告台灣司法 AI 治理正式進入有法可循的規範新紀元。然而，有法可循只是制度建構的起點，而非終點。人工智慧基本法所揭櫫七大原則中關於「人類自主」、「透明與可解釋」與「問責」三項原則，對司法

AI 而言具有特別深刻的意涵：它們不只是 AI 開發者的設計標準，更是法院 AI 使用者必須在每一個案件、每一份裁判中積極實踐的職業倫理。而人工智慧基本法明定「各目的事業主管機關」應循上位原則訂定配套法規，司法院指引之發布，正是司法院作為目的事業主管機關回應此一要求的第一步，但在法律意義上仍屬過渡期的行政指導，尚不具有法律規章的強制規範位階，但對司法院所屬人員仍具有一定行政內部拘束力。

由此可見，未來的制度發展方向至少包含五個層次的深化：

1. 從行政指導到法制化

司法院指引性質上係行政內部管理規範，欠缺對使用者的強制性問責機制，亦無對違反者的具體法律效果。在人工智慧基本法施行後，司法院作為目的事業主管機關，應依法在 2 年配套立法期限內，評估是否將 AI 使用規範提升至「法院組織法」、「法官法」或「法院辦案期限及防止遲延實施要點」等具有法律效力之規範位階，以確保規範之穩定性與可執行性，避免指引因行政便宜而朝令夕改。

2. 從統一規範到分級精緻化

人工智慧基本法第 5 條明定高風險 AI 應用須明確標示注意事項或警語；然而，何謂「司法領域之高風險 AI 應用」，迄今尚無法律定義。本文所提出的三級風險矩陣，在此框架下具有更為明確的法源依據，它可以成為司法院落實人工智慧基本法高風險 AI 管理義務的操作性工具。低風險行政輔助類應用（綠區），得在確保資安前提下彈性導入；中風險案情整理類應用（黃區），應建立強制溯源連結與人工實質審查機制；高風險決策輔助類應用（紅區），則應設置可解釋性要求、當事人揭露與異議權及絕對禁用點。此種分級精緻化，是指引走向真正制度深化的必要路徑。

3. 從內部規範到完整規範鏈

司法院指引目前僅覆蓋法院內部人員，但訴訟程序由內外多方共同構成。人工智慧基本法第 5 條第 1 項要求政府「應避免人工智慧之應用造成…… 資訊誤導或造假等違反相關法規之情事」，此一義務不應止於政府自身，更應透過訴訟法制延伸至外部律師之書狀提交責任。 AI 生成或處理之證據的真實性與可靠性審查，亦應在「準鑑定法

則」的框架下，透過修正刑事訴訟法、民事訴訟法或司法院「審判實務指引」等途徑明文化，完整的規範鏈，才能確保整個訴訟程序中 AI 介入的透明性與可究責性。

4. 揭露義務到可量化的透明標準

人工智慧基本法確立透明與可解釋為核心原則，但對於司法 AI 而言，「透明」不能僅止於文字聲明，而必須轉化為可供上訴審、監察機制及公眾審視的具體指標。本文所建議之 0 至 4 級 AI 參與度指標，提供了一套可量化的「含 AI 量」表達工具，在爭議案件中，此一指標將成為判斷使用者是否實質履行審判職責、是否淪為 AI 橡皮圖章的關鍵依據，其意義正是人工智慧基本法「問責」原則在司法場域的具體落地。

5. 過渡期資料使用之絕對界線

本文所提出之「匿名化、最少資料、片段分割、抽象化」過渡性操作方案，經指引 Q&A 之明確表態後，應予以修正定位：此類技術措施，僅得適用於已公開資料之輔助整理，絕對不得作為處理未公開卷證之授權依據。指引所揭示之核心原則是：訴訟卷證之保密性，不因技術性遮蔽而有所不同。在司法院落地型 AI 環境完備前，「公開資料外用，機密卷證內守」的內外有別原則，應是現階段最清晰的操作邊界。

綜上，人工智慧進入法院，從來不只是工具的革新，而是一項關乎司法權本質的制度選擇。人工智慧基本法的施行，為這項選擇確立了國家治理的上位框架，司法院指引的發布，則為司法場域確立了第一條可操作的規範界限。本文所建議之風險矩陣、AI 參與度指標、結構化使用紀錄、外部律師規範，以及 AI 證據的準鑑定審查機制，不是要否定這些已有的制度進步，而是要在此基礎上，補充更精細、更有制度及深度的操作 AI 工具，使揭露義務不流於形式宣示，使人類最終責任不只是原則重申，而是在每一份裁判的生成過程中，都留有可供檢驗的程序痕跡。司法的正當性，建立在程序的可信賴性之上，當科技進入審判，這份可信賴性不會自動保存。它需要制度的精心設計，需要使用者的主動自律，也需要學術研究持續發現漏洞並提出解方。真正必須被守護的，不只是審判效率，更是每一份裁判背後，人類法官仍為真正且可問責之判斷主體的那份不可讓渡的尊嚴。

【AI 使用聲明】

- 本文外文文獻部分：西班牙文（阿根廷）、葡萄牙文（巴西）、韓文（韓國）、德文（德國）、波蘭文（波蘭）等，若無相對應英文文件時，由 ChatGPT-5.0、Gemini2.5 協助翻譯並進行交叉驗證。
- 本文部分 ChatGPT-5.1、Gemini3 協助各國制度初步比較、校對，並經人工審查核對，附表三制作、附表一則由 claude 製作，提示詞為本文各該章節內容。
- 本文附表二表格部分由 notebookLM 協助生成，提示詞部分即以本文內容。

司法周刊

地址：臺北市重慶南路一段124號　　發行部電話：02-23826502　　編輯部電話：02-23618798