

AI 在刑事證據法之研究： 以美國聯邦證據規則為核心*

曾小榕

國立陽明交通大學科技法律學院
E-mail: hsiaojung.96@gmail.com

金孟華

國立陽明交通大學科技法律學院
E-mail: mhchin@nycu.edu.tw

摘要

現代社會中，人工智慧 (AI) 的判斷或預測結果逐漸成為人類決策的重要參考。雖然在刑事訴訟的論罪階段，尚無將 AI 之判斷結果，使用來證明被告有罪或無罪的案例；但隨著技術廣泛應用，訴訟一造主張以 AI 的判斷為論罪證據，恐是司法審判須面對的新型態資料。本文聚焦於具備機器學習過程並能獨立提出判斷結果的 AI，以美國聯邦證據規則為基礎，試以現行法律

© 中央研究院歐美研究所

投稿日期：112.7.20；接受刊登日期：113.10.25；最後修訂日期：113.10.4

責任校對：蔡旻芳、廖玉仙、陳昱之

* 曾小榕為第一作者，負責資料之蒐集與撰寫，合著貢獻比例 90%；金孟華為第二作者，協助第一作者撰寫內容並進行指導，合著貢獻比例 10%。本文部分內容為第一作者之碩士畢業論文。感謝二位匿名審查人對初稿的諸多建言，使本文對此議題的思考更加縝密，得以審視自身盲點。本文亦感謝《歐美研究》編輯委員會於論述架構與文意表達之諸多回饋與建議，於行文上方能更加通順。最後，特別感謝國立陽明交通大學科技法律研究所何欣蓉學姐協助潤飾文稿。惟作者當自負文責。

2 歐美研究

框架探討可能的適用問題。本文認為，此類 AI 證據屬於數位證據之一種，但無論視為物證或人證，均有適用法則之扞格。因此，需調整證據法的適用模式，以因應此新型態之證據。

關鍵詞：人工智慧 (AI)、機器學習、聯邦證據規則、數位證據

壹、前言

隨著人工智慧 (artificial intelligence; AI) 與大數據分析的發展，AI 已形成一類可執行預測或提出判定的應用方式。這些 AI 的判斷或預測結果，可能會越來越多地被應用於生活中，用來提醒、叮嚀人們，或作為人類決策前的參考資料。此類 AI 的判定或預測結果，如果成為刑事案件一造當事人用來證明某犯罪事實之構成要件該當與否的證據 (有學者稱為「AI 生成之機器證據」[AI-generated machine evidence])，其適用於美國聯邦證據規則 (Federal Rules of Evidence; FRE) 時，可能產生疑義 (Gless, 2020)。惟現階段尚無將此類證據使用在刑事審判程序論罪階段的案例，是以本文擬進行的是一種前瞻性探索。又由於國內尚無針對此類 AI 之證據資料於證據法研究中進行完整介紹的文獻，本文將以 FRE 為標的，並以外國文獻資料為主要的論證基礎，探索「AI 生成之機器證據」的證據法問題。

由於組成 AI 的方式多元，本文聚焦在具備機器學習及機器訓練過程並可獨立提出判定或預測結果的技術類型。以技術面觀察，此種 AI 具備二階段的特徵：在取得知識、形成判斷規則時是技術的前階段；開始執行預測或判定任務為技術的後階段。二階段之合作使 AI 得提出預測或做出判定。技術前階段倚賴技術人員輸入資料或校正機器的過程，可能有人證的內涵；惟技術後階段由機器產生的結果或報告，又不脫離物證的本質。此種具有人證內涵的物證，應該如何適用於 FRE，是一值得探究的問題。以傳統區分物證及人證的證據方法看待此種 AI 生成之證據，如將其判定結果視為物證，會傾向將此類證據視為一種不具備人類陳述的機器生成資料，因而不考慮適用以人類主張為前提的傳聞法則及質問權。但是在不適用前述二法則的結果，技術內的人證性質會更不易被發掘、

被檢討。而若將 AI 生成的結果視為具備人類主張的法庭外陳述，或是應該經過對質詰問的證詞型陳述時，如何對 AI 系統行使詰問，亦成問題。因此，此類證據適用在各種證據法則都有扞格。除了無法將其歸類至人證或物證任一種既有的證據方法，目前的證據法則可能也無法充分審查此種證據資料。本文以學說討論為基礎，經過整理與比較分析以後，主張以拆解技術階段的方式，將系爭資料視為一種兼具人證及物證的新型態數位證據。

以下為本文論證架構：第壹部分為前言。第貳部分簡述 AI、機器學習及 AI 未解決的隱憂問題，並指出系爭證據可能呈現在哪些具體技術上，末段聚焦在技術前階段的設計細節。第參部分將 AI 提出的判定結果與電腦資訊此類「數位證據」(digital evidence) 相比較，探討兩者在生成資料過程上的差異，以及將 AI 提出的判定結果直接歸類在電腦資訊之數位證據類型下，有何不相容之處。第肆部分，本文將此類 AI 之證據置於 FRE 下，進行全面性的適用分析，輔以技術過程、隱憂及特徵的觀察，從「科學證據可信與否」為切入點，再循序適用 FRE 之「關聯性」(relevance)、「驗真」(authentication)、「傳聞法則」(hearsay rule) 及「對質詰問權」(right to confrontation) 等各項證據法則，逐一提出適用問題。第伍部分為結論。

貳、AI 技術簡介

一、AI 之定義

AI 一詞於 1955 年由 John McCarthy 等人提出。當時對於 AI 研究的假設是，除了期待其模仿人類的學習及智能特徵，更欲試圖找出如何使機器得使用語言，形成抽象概念，解決人類無法解決的

問題，並達到機器能自我改進的成果 (McCarthy et al., 2006)。組成 AI 技術的方法中，「機器學習」的應用相對普遍 (張志勇等人，2022: 70)，在 AI 技術的教科書籍內多半可見其以獨立章節講述，屬 AI 技術組成之核心。且「機器學習」的特色，也體現其與傳統數位型法庭工具的差異。以下簡述機器學習技術。

二、機器學習技術

現今的 AI 系統，大多建立在機器學習的演算方法 (Seng & Mason, 2021: 244)。機器學習時常運用類神經網路技術 (龍建宇、莊弘鈺，2019)，¹ 且就使用目的不同，以及訓練資料在輸入之前是否需要進行標註，可大致分為三種類型。「監督式機器學習」是對訓練資料標註正確知識以後才對電腦輸入該筆訓練資料集，藉由人為標註使電腦學會正確的輸出規則。「非監督式機器學習」是只對電腦輸入資料，由電腦分析隱藏在資料中的結構，² 但是 AI 分析資料結構時，仍仰賴人類針對資料主題特徵的設計。例如 AI 的任務是找出資料主題為車子的特徵，但是車子應該具備哪些「特徵值」，³ 卻只能依賴人類設計者來思考與設定。而「深度學習」具備自行擷取學習資料集中特徵值的能力，並根據自動擷取的特徵值形成分類規則 (松尾豐，2015/2016: 156-175, 189, n. 36)。

¹ 類神經網路技術係模仿自生物神經網路處理能力，可理解為一種人為設計的訊息處理數學模型或資訊處理系統，各個人為神經元相連，並可相互輸入自外界取得或經過修正之實例、經驗、數據資料。相關內容可參閱李榮耕 (2018: 123)、葉怡成 (2009: 1-2, 1-3)、劉昌誠等人 (2010: 49)。

² 例如資料具有什麼模式、組別或異常點 (松尾豐，2015/2016: 128-129)。

³ 此處「特徵值」，指特徵分析工程 (feature engineering) 中「特徵量」的設計。特徵量是機器學習在輸入訓練資料時會使用的「變數」，該變數作為「某一特徵之定量值」 (松尾豐，2015/2016: 147-150)。

三、AI 系統之隱憂

AI 除了需要注意其機器學習過程的特點，還有一些技術上的問題尚未解決。以下分列三點說明。

(一) 相關性與因果關係不相等

無論是監督式或非監督式機器學習，都是藉由訓練資料讓 AI 學會與其任務相關的分類方法或潛在規則。當 AI 習得這些分類模式以後，即可針對新資料的輸入進行相關性的分析或預測 (Grimm et al., 2021: 25)。由此而言，機器學習的任務是傾向建立統計學式的相關性預測，並非模仿人類社會的因果關係推理 (Domingos, 2012: 87)。然而，統計學上的相關性不代表具有因果關係，即使兩組數據高度相關，也不能因此推論兩者之間具有因果關係，有時這可能僅是巧合 (Mayer-Schönberger & Cukier, 2013)。⁴

(二) 欠缺透明性與可解釋性

AI 之黑盒子 (black box，亦稱黑箱) 問題，指 AI 在資料輸入至輸出判定或預測結果的理由，無法由人類理解或解釋。法學討論中更有區分為法律黑盒子 (legal black box) 與技術黑盒子 (technical black box) (林勤富，2022)：前者指私營企業以受有營業秘密規範保護之地位，對於其 AI 技術之內容主張保密；後者指 AI 技術內部的未知狀態。

受限於黑盒子問題，當 AI 自身無法提出可檢視之推理依據，又無其他客觀的審查途徑時，若要聲稱 AI 是準確、有效的，這恐怕會令人質疑其具體驗證的方法為何？因此，在法律與技術黑盒子雙

⁴ 可參考著名的冰淇淋殺人案，詳見 Abu-Elyounes (2020b: 16-17)、Peters (2013)、Trafimow (2017: 743-744)。

盲之下，本文後續討論的重點，即是如何在法律上檢證 AI 提出之判定結果是可信的。

(三) 欠缺針對系統有效與可信的審查機制

有效性 (validity) 與系統計算、運行的「準確性」相關，其考究 AI 是否足以正確地進行分類或預測；可信性 (reliability) 與系統運作「維持一致性」有關，指其面對相同情狀時，可否維持計算或預測結果上的相同或相似性 (Grimm et al., 2021: 48; Heale & Twycross, 2015: 66)。

所謂「準確」，除了具備正確地執行系統任務的能力，還須考慮 AI 系統應用在目標領域時，是否會產生「非公正」的演算法判斷，這涉及 AI 偏見與歧視的議題。從技術建立的過程而言，偏見與歧視可能發生於兩個層面。第一層面是針對建構 AI 知識的訓練資料集而言，如果 AI 使用已存在偏見與歧視問題的訓練資料作為知識基礎，該偏見與歧視會永遠存在於 AI 的學習與訓練過程中 (Grimm et al., 2021: 42) (有學者稱此現象為「固化」)，⁵ 甚至因為 AI 反覆學習，使得偏見與歧視在系統演算法裡被強化。⁶ 資料集也可能以其他形式間接地呈現偏見與歧視的效果，比如已有許多研究顯示，黑人因為持有或使用毒品而被逮捕的機率比白人要高，所以如果採用毒品犯罪逮捕資料，會使其成為黑人群體的替代特徵，自然容易產生具有偏見的結果。⁷ 在第二層面，則要注意偏見與歧視

⁵ 「固化」一詞參見林勤富 (2022: 95)。

⁶ 典型案例如亞馬遜 (Amazon) 公司使用的實驗性工具——人工智慧徵人系統，詳見 Dastin (2018)。

⁷ 相關論證可參見 Grimm et al. (2021: 42-44)。有多則報導指出，黑人群體因持有或使用毒品被起訴或逮捕的比例，為白人群體之兩倍。可參見 Semmi (2024)、The Hamilton Project (2016)、Walker (2013)。另於研究「演算法公平性」之文章也指出，居住區域

可能來自技術人員在系統建立過程中的設定或取捨，例如技術人員對於非監督式機器學習設定資料的特徵值時、監督式機器學習在技術人員標註資料知識內容上，可能因為技術人員選取、教導的內容，輸入了偏見與歧視的價值觀 (Grimm et al., 2021: 44)。

關於 AI 判定結果欠缺標準化測試的主張，可以參考一則經典案件——*State v. Loomis*。⁸ 本案涉及法院在刑事審判程序的量刑階段，參考替代性制裁犯罪矯正管理剖析軟體 (Correctional Offender Management Profiling for Alternative Sanctions; COMPAS) 之再犯風險預測結果來決定被告之刑度，是否違反被告受正當法律程序保障之權利。本案被告 Eric L. Loomis 提出專家證人意見，主張法院並不清楚 COMPAS 進行再犯風險評估時，是依據何項人口基礎的母資料與 Loomis 個人的犯罪歷史、再犯風險進行預測比較，侵犯其受正當法律程序保護之權利。⁹ 法院最終參考專家意見指出，基於憲法上「正當法律程序之要求，應保障被告所受判決係以正確資料為基礎之權利」，¹⁰ 該權利包括可審查與驗證巡迴法院判決時所依據之量刑前調查報告 (The Presentence Investigation Report; PSI) 所含資訊。¹¹ 法院並要求，往後任何案件有使用 COMPAS，其量刑前調查報告中皆須提醒量刑法庭關於 COMPAS 準確性之注意事項及其局限性。該注意事項中提及，COMPAS 之風險評估係將被告與全國的樣本進行比較，但尚未完成與威斯康辛州 (本案事涉

代碼可能成為黑人群體之替代特徵。該文強調，特意忽略「種族」進行演算法，在追求公平上是徒勞的，並引註量化研究結果佐證。參見 Abu-Elyounes (2020b: 9)。

⁸ 371 Wis. 2d 235 (2016).

⁹ *Id.* at 250.

¹⁰ *Id.* at 257.

¹¹ *Id.*

州) 人口為交叉驗證之研究。¹² 藉由本案法院對於預測工具準確性在正當程序保障的要求，可認為最高法院已注意到風險預測工具有準確性的疑慮，惟本案係將 COMPAS 使用在刑事審判程序之量刑階段，並非使用在論罪階段，也因此，本案僅能說明如有其他案件在量刑階段使用準確性不明的工具時，可參考本案提出之正當法律程序要件。

附帶一提，有學者將藥品上市前管制與 AI 演算結果作為證據使用的情況互為比較。比較之基礎在於，藥品安全與 AI 之判定結果作為證據使用時，都可能對人類之生命、身體權產生重大影響；惟 AI 之演算法，卻沒有如同美國藥品上市前必須先通過事前測試及批准的程序。並且，即使 AI 之演算法經過審查，通常也不是由獨立機構、同儕審查或是足夠透明的程序進行，目前也沒有任何 AI 產品的標準化測試程序 (Grimm et al., 2021: 48)。因此，在欠缺標準化測試的情況下，更須注意隱藏於技術中的偏見與歧視問題，以及我們對於技術內部複雜運作的無知程度。

四、AI 於刑事證據之具體應用可能

本文以影像處理及專家型預測系統為例，說明 AI 之判定結果可能作為論罪證據使用的情況，並初步觀察出 AI 由機器學習到任務執行的兩階段特徵。

(一) 影像處理

影像處理是以攝影機和電腦代替人眼，對目標進行辨識、跟蹤和測量，並進一步做圖像處理的機器視覺 (machine vision)。惟電腦

¹² *Id.* at 264.

對於圖像的認識是像素的點集合，僅能辨識某一圖片內有紅色、綠色不同像素，無法自行判斷該紅色物體之內容，例如其是否為蘋果，或為其他物體。因此，關於圖像資訊的知識必須仰賴人類的教導，例如使用監督式或非監督式機器學習的演算法使電腦學會辨識圖像的實質內容 (山口達輝、松田洋之，2019/2020)。目前，影像處理技術廣泛應用在車牌辨識、人臉辨識、情緒辨識 (疲勞偵測) 等 (張志勇等人，2022)。

人臉辨識結果及車牌辨識結果，可能成為刑事案件中證明特定犯罪行為人在場與否之重要證據。又如疲勞駕駛偵測系統，透過對於人類情緒特徵的分析，在偵測到汽車駕駛人有疲勞的表情或身體狀態時，該系統會立刻發出警告訊息 (張志勇等人，2022: 33)，該警告訊息可能在交通意外之過失傷害或致死案件作為證明行為人有過失之證據。

(二) 專家型預測系統

專家型預測系統同樣是機器學習的應用，例如再犯風險評估系統採用機器學習中「隨機森林」(random forests) 的演算法 (Berk & Hyatt, 2015: 222)。隨機森林是一種進階的決策樹演算法，決策樹是由「是或否」(yes or no) 作為答覆的各節點，組成不斷二分的樹枝圖，最終取得目標問題的最適解答區間；而隨機森林則是透過多個樹枝圖同時對同一問題進行樹狀圖的節點分類，最終將多個樹枝圖的預測結果進行多數決或取平均數值，從而得到比單一決策樹更精準的判斷 (山口達輝、松田洋之，2019/2020)。

更進階的專家型預測系統是以深度學習為演算法，例如在疾病之預測上，以電腦快速閱讀、分析醫療數據與資料，加上深度學習自動擷取特徵的能力，AI 在醫療應用的加成，使診斷的預測或精準

度提升 (張志勇等人, 2022)。儘管 AI 在醫療領域之應用有助於提升診斷的預測或精準度, 但是在醫療過失傷害案件, AI 的預測與診斷結果紀錄, 可能成為醫事人員未來必須面對的新型態證據資料。

(三) 技術運作過程的二階段特徵

根據前述對於技術內容的觀察, 影像處理與專家型預測系統, 都需要先經過人為教育機器圖像識別、選擇專業知識資料集並設定演算法規則的環節, 此為建立技術之第一階段 (技術前階段)。當 AI 習得任務領域的知識並形成判斷規則以後, 才進入技術執行目標任務的第二階段 (技術後階段), 亦即經由黑盒子運算來生成判定結果。

五、AI 獨立決策能力養成：以再犯風險評估系統為例

基於上述二階段的特徵, 本文進一步探究第一階段中機器學習的技術細節, 發現技術人員可能左右 AI 之認知內容及判定結果, 詳述如下。

(一) 機器學習之技術建立過程可能隱藏技術人員之價值選擇 (AI 被決定之決定)

本文提取了建立機器學習技術時普遍通用的五項步驟, 分別是定義問題、資料蒐集、資料清理、選擇最佳化統計模型與模型訓練, 以及模型應用。然而機器學習過程並不是線性地遵從上述步驟, 而是會在各個步驟間前後、來回適用, 方可達成學習的成效。前述五項步驟也並非完全適用在所有的機器學習演算法 (尤其深度學習的多層神經網路是由多個非線性之計算組成)。儘管如此, 上述的五項步驟, 可以幫助法學研究者初步認識機器學習更細節的組成, 並透

過解構技術的方式，觀察隱藏於技術過程中可能涉及的法律問題 (Lehr & Ohm, 2017)。

本文以再犯風險評估系統為具體說明的媒介，¹³ 指出技術人員在系統的設定上，可能無意識地輸入個人主觀的價值。學者也指出，雖然再犯風險評估系統計算出的結果，乍看之下是由客觀的資料、演算法及電腦程式運算而來，但實際上這些運算結果並非全然客觀 (李榮耕，2022)。

1. 定義問題

再犯風險評估系統針對執行目標的定義方式，可能會影響風險值計算的觀察基準。例如，當再犯風險評估的是「某位被告未來再次犯罪的可能性」時，「未來再次犯罪」究指為何？在屏除法律明文定義為前提下，其定義可能因為不同人對於公共安全減損的標準不同而存在不一樣的詮釋 (Eaglin, 2017: 75-76; Weisberg, 2014: 786-787)。例如有認為「再犯」可定義為一項新的犯罪；也有認為受假釋之人如果違反假釋或緩刑條件，或因不支付罰金而再次入獄，已對於「再犯」評估目的所為保護的「公共安全重要性」產生減損 (Eaglin, 2017: 75-76)。此外，根據 COMPAS 的操作指引及文獻資料發現，COMPAS 並沒有特別說明為何以「兩年內」(無論涉犯重罪 [felony] 或輕罪 [misdemeanor]) 受逮捕作為再犯的風險預測基準。¹⁴

¹³ 後述於風險預測系統建立步驟的說明，主要參見 Lehr & Ohm (2017: 670-702)、Nishi (2019: 1678-1681)。有學者也提到，於審判中我們比以往更時常使用風險預測系統，其一原因可歸責於機器學習技術新模型的發展；COMPAS 進行的再犯風險預測就是採用機器學習的技術。詳見 Abu-Elyounes (2020a: 418)、Nishi (2019: 1677)。

¹⁴ 相關內容可參見 Eaglin (2017: 77-78)、Northpointe Inc. (2015)。

設計者預設建立系統所費的時間與金錢條件，也會影響任務目標的定義。例如俄亥俄州風險評估系統 (Ohio Risk Assessment System; ORAS) 計劃在三年內開發完成，因此設計者將觀察再犯發生之期間，限縮在一年內 (Eaglin, 2017: 101)。學者指出，設計者設定系統評估再次犯罪的時間長短，原因不一，很多時候與學理或價值判斷毫無關係，單純是因為研發設計與技術成本的考量 (李榮耕，2022: 139)。

2. 資料之選擇與蒐集

技術人員會需要一個足夠龐大的資料集，使預測系統足以建構執行預測任務必須具備的模式 (patterns)、相關性之規則 (Lehr & Ohm, 2017: 671)。此時必須注意兩個問題：第一，技術人員擷取或蒐集之資料，是否不具代表性，導致僅能彰顯某區域性之現象。以美國地域為例，如果僅僅使用某一州內部分區域之逮捕紀錄、起訴書或判決書，作為訓練再犯風險評估系統之資料，若該地區之犯罪率通常較高 (例如是毒品犯罪之聚集地)，則該部分區域內之資料，恐怕不能完全反映整個州內所有地區之實際犯罪情況，因而失準地高估或低估全州之再犯風險。第二，設計者選擇機器學習的訓練資料時，是否考慮該筆資料在製作當時的初始使用目的及其涵蓋事項之範圍？學者指出，被選擇使用的訓練資料集內容，除了實際上已被限縮於資料製作當時設定的使用目的範圍，訓練資料可能也早已經過篩選或刪減出需要記錄的項目。倘將此筆資料直接用於目前 AI 任務的知識使用，可能已逾越該筆資料可信性之範圍。因此，為其他目的而蒐集的數據或資料，是否真的符合或足以提供現在 AI 執行預測任務的學習資料，也是技術建立過程需要注意的可信性問題 (李榮耕，2022: 136-137)。

3. 資料清理

在資料清理時，設計者需要檢查訓練資料是否有錯誤或缺漏，並決定移除資料或是以整個資料集之中位數或眾數取代 (Lehr & Ohm, 2017: 673)。若設計者選擇刪除資料，必須注意整體數據的減少，是否會影響系統運算之顯著性；如果係以替代數值補充，則須考量替代數值是否有害於運算目標之準確性。學者指出，無論設計者決定採行何種資料清理方式，都將影響風險評估系統最終運算結果之正確性。此外，資料之取捨、修改，時常不為外界所知，而不同的系統設計者，也可能有不同的資料清理標準 (李榮耕，2022: 142)。

4. 選擇最佳化統計模型與模型訓練

在模型選擇與模型訓練的階段，設計者會選出最佳化預測目標之統計模型。此外，設計者也需要進行資料的轉換，例如，將一段文字敘述由非數值的資料轉變成適合演算法運算的 0、1、2 等數字編碼，以取代原資料的表達方式，從而便於模型使用。設計者並且在模型訓練時調整系統判讀的錯誤 (山口達輝、松田洋之，2019/2020: 66-69)，同時在這個階段決定影響再犯率之風險參酌因子 (factors)，並確認各參酌因子均表現統計上之顯著性 (statistical significance) (李榮耕，2022: 142; Nishi, 2019: 1678)。

5. 模型應用

最終，在模型應用階段，設計者會將系統計算出的風險結果分級，通常以集群分析劃分為高、中、低度三個等級，以便使用者以

風險等級理解判讀結果。¹⁵ 學者指出，什麼樣的風險值應該歸類在高、中或低的哪一個集群，(若無法律強制規定風險值如何分類) 實際上純粹是人為的判斷，或是反映自系統設計者對於風險的容忍程度，¹⁶ 並不是純粹的電腦程式、科學或數學的運算，甚至是無關數學的。更需加以注意的是，人類社會對風險的容忍程度，實際上又隨著不同社會、文化或族群，在不同時代，會提出不盡相同的答案(李榮耕，2022: 149-150)。

(二) 小結

從問題定義、資料選擇與蒐集、資料清理，到模型選擇、訓練及應用，各個步驟都需要專業技術人員對於數據內容進行調整、確認資料判讀正確，並為風險程度分級等。這些人為介入的操作，本質上無法脫離設計人員主觀決定與價值判斷所帶來之風險(李榮耕，2022: 150)。

參、AI 系統提出的判定結果與電腦產出的數位資訊之差異

本第參節將介紹在美國證據法體系中，由司法實務發展出的電腦數位資訊的證據分類模式。本文並將電腦產生數位資訊的過程，與 AI 系統產生判定結果的過程進行比較。本文認為，由電腦紀錄

¹⁵ 以 COMPAS 為例，其中多個預測模型，皆以最終風險得分 1-4 級分為低風險 (low risk); 5-7 級分為中度風險 (medium risk); 8-10 級分為高度風險 (high risk)。可參見 Abu-Elyounes (2020a: 418-422)。

¹⁶ 例如 20% 的再犯可能性，有些人會認為已經是無法忍受的高風險，但也有認為只是低度風險。抑或是 50% 是中度風險，但 51% 過半的機率，是否已經成為高度風險？這些都涉及個人或整體社會對於風險的容忍程度(李榮耕，2022: 149-150)。

的技術內涵建構出的證據規則，可能不適合直接適用在此類新型態的判定、預測型 AI 證據。

一、何謂數位證據？

數位證據係指任何可以支持或推翻某一犯罪事實的發生，或是可以證明犯罪之主觀故意、意圖或不在場證明等重要的犯罪事實要素，並且由電腦設備或電磁紀錄載體等數位方式儲存或傳送的「資料」(Casey, 2011)。¹⁷ 亦即，數位證據是指功能上用來證明某犯罪事實成立與否之法律上構成要件，並以電腦設備或電磁紀錄等載體儲存或傳送之數位型態資料。

本文特定於探討電腦紀錄類型的數位證據，因為數位證據之指涉範圍廣泛。FRE 並沒有針對數位證據進行直接的定義，但可找到「(包含) 以電子方式儲存之紀錄」或是「由電子程序或系統產生，並經過書面認證之紀錄」等數位形式儲存或產生之資訊的相關描述。¹⁸ 由於 AI 為電腦發展歷史過程下的產物，本文將比較的範疇限定在電腦紀錄的數位資訊，係認為二者之間具有可比較的關聯基礎。¹⁹

二、電腦產生資料的數位證據類型

(一) 三種電腦紀錄

電腦產生之數位資訊或數位資料，在法律實務的發展中提出了以下的三分法：「電腦存儲紀錄」(computer-stored data)、「電腦

¹⁷ 可比較參閱李榮耕 (2014: 173)。

¹⁸ See FED. R. EVID. 101(b)(6) & 902(13)。另外參考我國法律實務工作者之見解，數位形式之證據，可以技術為基礎有多種分類方法，詳見朱帥俊 (2011: 42)。

¹⁹ 亦可參見孫宏民、呂沐錡 (2015: 330)，書中提到，人工智慧源自於研究如何使電腦能夠擁有智慧及推理能力，並根據這些能力做出相對應的行為的研究領域。

生成紀錄」(computer-generated data) 與「混合型電腦紀錄」。此分類方法可追溯到 1983 年，由路易斯安納州最高法院在 *State v. Armstead* 案²⁰ 首先提出區別電腦生成數據與電腦存儲之數據。本案涉及 Gregory Armstead 涉嫌撥打多通猥褻電話，檢方提出儲存於電腦並影印出來的電話紀錄，以證明該通話紀錄與電話來源與被告之犯行相關。法院認為，倘一項證據完全由電腦自行記錄與產出(不經由人類之手)，則該證據為電腦生成之數據，而非電腦儲存了人類陳述的電腦存儲數據。²¹ 美國司法部(The United States Department of Justice) 於 2001 年在此二分法的基礎上，延伸出第三種「混合型電腦紀錄」(Kerr, 2001)。²²

司法部指出，電腦存儲紀錄包含人類著述且以電子形式呈現的紀錄，例如電子郵件、文字處理的檔案資料、網路聊天訊息。電腦生成紀錄是電腦程式的輸出，且紀錄產生過程不經由人類之手，例如網路服務提供者的使用者登入與登出紀錄、電話通聯紀錄及 ATM 自動櫃員機影印產生的收據。混合型電腦紀錄兼具前兩者的性質，例如試算表的數學運算結果本為電腦生成紀錄，但若該試算表中的財務數據(屬於電腦存儲紀錄)經過詐欺之犯罪行為人的竄改，則該試算表之運算結果為混合型電腦紀錄(Kerr, 2001)。

(二) 三種電腦紀錄適用之證據法則

前述三種數位證據依其資料產生的過程不同，適用不同的證據法則。對於電腦生成紀錄，*Armstead* 案指出，被告撥打電話之電信

²⁰ 432 So. 2d 837, 840 (La. 1983).

²¹ *Id.* 本案受美國伊利諾州最高法院於 1985 年的 *People v. Holowko* 案中再次引用。參見 *People v. Holowko*, 109 Ill. 2d 187, 188-189 (1985).

²² 亦可參見 The United States Department of Justice (1997)。

通聯紀錄是在通話資訊撥入後由電腦即時、不經過人類之手製作成與之相關的資訊紀錄，由於不具有人類供述的內容，故不適用證據法上為檢驗人類證述可信與否的傳聞法則；惟應考慮電腦的計算結果，是否屬於證據法定義的科學證據，故須經過證據法上科學技術可信性的審查，²³ 也需要審查電腦輸出程序是否正常運行，以通過證據法上驗真證據真實性的要求。就電腦存儲紀錄，*Armstead* 案及司法部指出，當電腦輸出之資料是人類證述內容的直接紀錄時，該紀錄只是人類證述內容轉換至電腦儲存載體而已，因而具有人類於法庭外陳述之性質，適用傳聞法則。²⁴ 至於混合型電腦紀錄，仍根據各筆電腦紀錄的性質，分別適用至前述兩種紀錄對應的證據法則 (Kerr, 2001)。

三、AI 提出之判定結果屬於哪一種電腦紀錄？

(一) AI 之判定結果非電腦存儲紀錄

AI 係藉由資料庫之知識，透過機器運算產生判定結果。這一過程，與電腦存儲紀錄係忠實地將人類思想內容之文字或陳述，轉換至數位形式載體上的技術過程，有所不同。再回顧 McCarthy 等人對於 AI 研究設定的初始目標，是為達成 AI 能表現人類的智能特徵，亦見其技術發展的目的不是追求 AI 作為一項資訊儲存的載體。故本文認為，AI 提出之判定結果的資料，並不能符合實務對於如實轉載人類之陳述內容作為歸類標準的電腦存儲紀錄。因此，此類 AI 之證據無法劃分至電腦存儲紀錄。

²³ *Armstead*, *supra* note 20, at 839-841; *Holowko*, *supra* note 21, at 193.

²⁴ *Armstead*, *supra* note 20, at 839-840. 司法部採用 *Armstead* 案見解可參見 Kerr (2001: 2)。

但如果是訓練 AI 的資料集或文字資料內容（例如在自然語言學習訓練上使用的多筆文本範例）本身具有人類陳述的性質，且資料輸入 AI 時只是轉換文本的載體，那麼各筆訓練資料本身仍符合電腦存儲紀錄的證據內涵。

(二) AI 之判定結果與電腦生成紀錄之內容不同

由前述 *Armstead* 案件事實可知，電話通訊紀錄是電腦直接紀錄當時通訊之電話號碼、通訊時間等與已發生之現實狀況相關的資料。與此相比，AI 技術則是經過學習知識的過程，形成一種判斷規則以後，再執行類似於人類對某一情狀、現況進行「思考、斟酌」的運算，方提出其判斷或決策。並且此項判斷或決策，具有本文特別指稱經由黑盒子演算法「獨立」決策的性質。這不同於電話通訊紀錄僅僅是直接記錄現狀。

因此，雖然由形式上的觀察，電腦生成紀錄與 AI 產出判定結果的過程，可能難以區分有何不同之處，但是如果我們更深入地瞭解機器學習與判定結果的前後關係以及資料輸出的過程，仍然可以從中比較出其與電腦生成紀錄的過程不同。

(三) AI 之判定結果與混合型電腦紀錄之內容不同

因為混合型電腦紀錄具備存儲紀錄與生成紀錄的雙重性質，單由技術程序的形式面觀察，可能認為 AI 的知識資料集可歸為電腦存儲紀錄，而 AI 的判定結果可以歸為電腦生成紀錄。惟 AI 之判定結果涉及對現狀進行詮釋，或對未來進行預測，亦即是一種關聯性的決定，而非對現實狀況的直接記錄，已如前述。再者，AI 針對輸入資料的使用方式與混合型電腦紀錄不同。詳言之，混合型電腦紀錄是以存儲紀錄的內容，直接作為運算的前提（例如人為輸入股票

代號、張數及價格後，電腦直接計算出應成交的標的、成交金額、成交時間)(李榮耕，2014: 177)，與 AI 係透過人為輸入之資料集形成判斷規則，並以此針對其他標的執行判斷任務的資料運用方式，存在差異。

(四) 小結

在 AI 生成數位資訊的過程中，電腦作成決策的自主性與獨立性較高，²⁵ 並且更接近於人腦決策的思考和決定過程。相較之下，傳統類型的電腦紀錄僅是機械化地儲存人類思想內容之文字表述，或是執行簡單的數學運算、數據資料整理、時間戳記等，其產生的紀錄內容主要是對人類社會資訊的直接記載。因此，AI 作成的判定結果作為一種具備分析過程的數位資訊，並不適合直接適用於傳統三種電腦紀錄發展出的證據法則。再者，前述提到 AI 所存在的黑盒子問題、系統有效性與可信性未經驗證的問題，使得 AI 技術相較於傳統電腦技術，展現出更多技術過程上的不可知性。

四、本文觀點：應重新檢視 AI 提出判定結果資料之證據性質

綜前所述，AI 技術所生成的判讀資料之所以與電腦紀錄之數位證據不同，不僅在於 AI 技術進行「決策」的特徵與其隱憂問題，更獨具一格之處是其「機器學習」的過程。在技術的前階段，AI 經由技術人員框架知識，利用設定相關性預測或採用類神經網路學習模

²⁵ 此處所謂 AI 決策的獨立性與自主性較高，特指在 AI 建置完成以後，任其獨立執行委派任務時，AI 的黑盒子決策過程不是人類可以插手或是理解的。但 AI 取得判斷之能力，仍然不可脫離技術人員的設計過程，而這其中可能摻雜技術人員主觀的價值觀或認知偏誤。

式，使 AI 足以對於標的進行特徵之分解、重組，進而產生演算法的規則。而在技術的後階段，AI 則根據前階段所習得之判斷規則，透過黑盒子運算提出預測或判定。本文認為，AI 的判斷與預測結果，在紀錄生成的技術過程上與電腦紀錄之傳統數位證據有明顯的區別，進而在證據法則的適用上，應有獨立出來另外討論其適用模式的必要。

但需要先釐清的問題是，技術人員在技術建立的過程可能嵌入不公平之價值選擇。而這些選擇可能體現在系統任務之定義方式、資料的選擇及取捨方式、資料的編碼方式等方面，再結合前述提到的三項隱憂，這是否會直接導致此種 AI 證據應該在刑事程序中禁絕使用，是值得思考的問題。本文認為，從技術層面觀察，因黑盒子問題、欠缺有效性與可信性審查機制，加上如果無從確認技術人員是否正確地使用訓練資料、對問題的定義是否存在偏見價值等情況下，AI 之判定結果確實存在可信性的疑慮。目前的司法實務，也尚未見到將 AI 判定作為論罪證據的案例，至多是在非論罪階段的緩刑、假釋及量刑程序參考再犯風險評估系統的預測，輔助法院進行決定（李榮耕，2022）。

然而，科技工具產出的證據是否欠缺可信性的問題，並非 AI 之判定結果獨有。目前已經使用於司法程序的酒精吹氣檢測儀、DNA 檢測儀、GPS 追蹤等數位工具，在過去也曾經面臨技術可信性的質疑，而此質疑也不是一經司法實務決定信賴某一項工具之後，其可信性就可以永久地不受推翻。美國實務判決及學者皆有提到「證據的生命週期」(the evidentiary cycle) 的概念 (Gless, 2020: 207, 215-216)。此概念係建立一套面對「新證據」時，可概略判斷該證據應該放在可信程度哪一個階段的判斷規則。生命週期分為四個階段，包括：(1) 證據類型太新而不可靠；(2) 證據類型新穎，但須接受檢

驗；(3) 證據通常認為是可靠的，但偶爾會被不當地使用；(4) 證據被盲目地使用。上述的生命週期並非不可逆轉，DNA 檢測及酒精吹氣測試都曾被盲目地使用，然而目前有認為其使用上應該經過更嚴謹的審查 (Kaestle et al., 2006: 85-87)。²⁶

承前所述，即便此類 AI 之證據目前可能被認為是太新而不可靠的 (Gless, 2020)，也不能排除 AI 形成的判定結果可能在未來會被普遍地認為可靠；因此，針對此類新興證據的證據法問題仍有討論實益。依據證據生命週期的概念，AI 的判定結果不應該在刑事程序中被完全禁絕。目前只是因為司法實務及大眾對於 AI 形成的判定結果尚未賦予高度可信性，因此被暫時排除而已。

此外，儘管數位工具存在可信性的質疑，但仍可能有助於發現更精確的犯罪事實 (Gless, 2020)。例如以再犯風險評估系統而言，因為系統認知方式不如人類有彈性，或因為技術設計層面的局限性，有認為其有犧牲個案正義的危險，但亦有認為它的判斷可以比人為量刑的方法更客觀、公平，更能達成量刑一致性。²⁷ 又如同樣具有可信性風險之酒測結果、DNA 比對結果在實務上仍普遍被使用，此現象似乎也顯示大眾與法院為發現真實或期待貼近事實，而傾向於接受一定可信性風險的選擇。再隨著技術革新，未來 AI 的大數據分析能力如能更準確或更靈活地被運用，人們可能隨之更信任 AI 的判定或預測，不排除屆時已被證立或被普遍地視為客觀、準確、更公平並維持一致性等正面價值。此外，亦有其他法學研究，

²⁶ 相關見解可參見 *United States v. Beasley* 案，本案肯認使用 DNA 測試結果的可信性，並認為其可信程度達到允許往後的司法實務皆可考慮參考 DNA 檢測結果。惟本案法院提到，DNA 檢測結果的可信性仍可以受挑戰推翻，例如當樣本處理流程馬虎、執行檢測人員未經適當培訓、未遵循適當的檢驗準則等削弱健全科學方法的情況。*United States v. Beasley*, 102 F.3d 1440, 1448 (8th Cir. 1996).

²⁷ 可對照李榮耕 (2022: 127-128)、薛智仁 (2023: 107) 與 *Metallo* (2020: 2059)。

在探討訴訟當事人與提供法律意見的 AI 工具之間的詢答內容，是否能受到證據法上秘密溝通特權 (legal advice privilege) 保護的議題。²⁸ 因此本文認為，AI 產生的判定資料固然有預先進行證據法研究的實益。

肆、適用 FRE 檢驗 AI 提出之判定結果資料將產生之問題：以刑事訴訟程序為核心

對於 AI 的技術及特徵有基礎認識以後，本文進一步分析 AI 判定結果證據於 FRE 之適用。在此之前，本文藉由以下假設案例，幫助讀者具體想像將 AI 判定結果主張為證據的情境：假設醫師因診斷失誤導致病患死亡，病患家屬於刑事審判之論罪程序中，提出專家型疾病預測與診斷的醫療影像辨識系統之判讀結果，指稱連 AI 系統都可以辨識出的早期病灶，醫師卻因為自身的疏失而未及早發現。病患家屬並以此判定結果，主張證明醫師的醫療行為存在過失 (該當主觀犯罪構成要件)。此時，法院必須先評價系爭 AI 判定結果資料是否具備證據資格。

在 FRE 中，一項證據要取得證據資格，無論其為人證或物證，必須通過至少以下五個證據法則關卡：原始證據 (originality)、拒絕證言特權 (privileges)、關聯性、驗真及傳聞。本文聚焦在關聯性、驗真及傳聞，檢視 AI 之判定結果適用上述三種證據法則時，可能存在哪些排除證據資格的情況，以及可能面臨的適用問題。

再者，無論是因為缺乏有效與可信性的審查機制、技術黑盒子問題，或是人為介入技術設計，這些因素都可能引發對 AI 技術可信性的疑慮，也是此類 AI 證據的核心問題之一。而為了應對科學

²⁸ 詳見 Stockdale & Mitchell (2019: 422)。

技術普遍具有可信性的疑慮，證據法則中實際上已有建立了「科學證據可信性」的參考標準，本文將首先探討此一核心問題。在證據資格的基礎門檻——證據「關聯性」審查中，本文將論及 AI 之判定結果是否構成 FRE 第 403 條的排除條件，以及其是否可能成為一種品格 (character) 證據或習慣證據。接著討論證據「驗真」的問題，涉及如何驗真 AI 判定結果為出證者所宣稱之證據。最終則探討此類證據適用「傳聞法則」的適合性，以及對於此種證據行使「對質詰問權」的方法，存在哪些不同意見。

一、AI 提出之判定結果資料於科學知識可信性的審查

(一) 道伯法則及 FRE 第 702 條

美國聯邦最高法院在 *Daubert v. Merrell Dow Pharmaceuticals, Inc.* 案中提出法院在判斷科學知識「可信」與否時可以參考的五項核心要素，包括：(1) 理論或技術必須要能夠經過測試、證實；(2) 該理論或技術必須經過同儕審查及出版；(3) 法院應考量該科學技術之錯誤率；(4) 是否存在並維持著該技術操作的標準作業程序；(5) 是否為相關科學社群所普遍接受。²⁹ 隨後於 *Kumho Tire Co. v. Carmichael* 案件表示，即使是非科學性之技術或其他專門知識之專家證言，也須適用 *Daubert* 案的五項技術可信性要素。³⁰ 於 2000 年時，國會審議通過修訂 FRE 第 702 條，其修法重點即是納入道伯法則 (Daubert test)，除了明文要求專家證人出庭作證之資格，也明文要求專家證人之證詞應基於「可信的原理或方法」且可信地應

²⁹ 509 U.S. 579, 593-594 (1993). 相關內容亦可參見金孟華、陳又寧 (2015: 451-452)、Park et al. (2022: 358)。

³⁰ 526 U.S. 137, 141 (1999).

用於本案事實。³¹ 此外，由於在 FRE 中，需要專家證人提供法律以外專業證言之證據資料多數具有科學性質，因此，涉及「專家證人證言」的證據能力議題，常被稱為科學證據之證據能力議題（蘇凱平，2021a: 1958-1959, n. 109）。

（二）AI 提出之判定結果資料與科學知識可信的核心要素

AI 所做成之判斷、預測結果或報告，不論視為專家證人提出之證詞，或僅被視為單純的非供述證據之物證，³² 無疑都是與科學或技術有關的證據。因此，這項判定報告必須經過專家證人出庭作證說明後方能使用。然而在適用 *Daubert* 案件的核心要素時，AI 技術可能無法通過要素 2，即技術經過同儕審查及出版的要求。原因在於，目前的 AI 技術幾乎是由私營企業為商業用途而研發，企業通常不會選擇公開技術內容或是交給同儕審查，更何況其等擁有主張法律上營業秘密保護的權利（即所謂法律黑盒子）（Gless, 2020）。再者，如果美國政府亦未設定 AI 技術上市前需通過有效性、可信性檢驗的客觀標準或審查機制，綜前所述，這將導致亦無其他中立的、可以間接確認 AI 輸出答案準確性與一致性的替代驗證途徑。再加上技術黑盒子問題，法院在審查 AI 輸出判讀結果過程的可信性，恐怕面臨困難。

因此，倘若系爭 AI 技術還未能通過道伯法則的科學知識可信性要素，專家證人（無論是一位人類的專家證人或是 AI 本身作為專

³¹ FED. R. EVID. 702. 參見 FRE 第 702 條編纂委員會註解 (FED. R. EVID. 702 Committee Notes on Rules—2000 Amendment) 指出，修訂本條文，係為使道伯法則明文化。但立法者無意將 *Daubert* 案各項要素逐一規定，其目的在於讓法院適用道伯法則時，能夠保有彈性，以因應個案不同的情況。

³² AI 證據資料之性質係非供述證據或有供述證據之內涵，及個別存在的證據法問題，悉於後述。

家證人)³³ 將無法符合 FRE 第 702 條第 (c) 項及第 (d) 項中，證詞應基於「可信之原理或方法」，並且將其「可信地應用在本案事實」之要求。

(三) 小結

由於營業秘密抗辯、欠缺有效性與可信性的審查標準，以及局限在技術黑盒子困境等問題，法院恐怕認為 AI 之判定結果，無法通過目前的科學證據可信性審查要素。因此，以前揭醫療過失假設為例，若一造當事人以醫療影像辨識技術之判定結果主張他造當事人具備過失之主觀構成要件，法院可能會首先衡量 AI 辨識技術是否符合道伯法則。但如前所述，新證據的生命週期並非不可前進或逆轉，仍可期待此類新興科學證據之可信性轉變為正面評價，或是其受質疑之處能夠得到解決。

二、AI 提出之判定結果資料與關聯性

(一) FRE 第 401 條、第 402 條、第 403 條

依據美國證據法，證據資料應先具備與本案事實及爭點之「關聯性」(Lilly et al., 2015)，與此相關的條文是 FRE 第 401 條、第 402 條與第 403 條。第 401 條定義「相關」之證據的兩個要件；第 402 條揭示具關聯性之證據原則上具有證據資格 (admissible, 又稱為有證據能力)，³⁴ 而無關聯性之證據絕對無證據資格；³⁵ 第 403 條則揭示一項權衡機制，即一項證據之提出產生「不利影響」之風

³³ AI 作為專家證人之問題還包括無法命令 AI 具結、無法對 AI 質問等，本文將於質問權的標題下詳細說明。

³⁴ 除非有其他憲法或法律規定排除其證據容許性。參見 FED. R. EVID. 402。

³⁵ *Id.*

險若會大幅度地超越該證據之證明價值 (probative value)，儘管該證據具有關聯性，法院仍得予排除。³⁶ 學者提到，法院實務通常將 FRE 第 403 條放在所有其他證據法則 (例如關聯性、科學證據可信性、驗真、傳聞與否、品格證據與否等) 審查之後，作為最後一項資格審查關卡 (Anderson, 2012)。

(二) FRE 第 403 條與 AI 判定結果資料之適用關聯

AI 系統之可信性質疑及黑盒子問題，可能與 FRE 第 403 條例示「不利影響」中的「不公平之偏見」及「誤導陪審團」有關。

不公平偏見的不利影響，指的是 AI 可能繼受了含有偏見與歧視問題的訓練資料，或是技術人員無意中置入了偏見、歧視的價值觀。隨著機器反覆學習、訓練與修正決策的過程中，固化或強化了偏見歧視，其判定結果必然會存在不公平之偏見。而誤導陪審團的情況，則因 AI 給出結果的過程存在黑盒子問題，且欠缺上市前審查技術有效性、可信性的強制規定，因此難以確認本案系爭 AI 技術判定結果之可信程度。進一步而言，如果 AI 在機器建立過程中設定了不當的任務定義，或使用了代表性不足的訓練資料，這可能導致 AI 判斷或預測錯誤，進而使陪審團或事實的審判者受到其判定錯誤的誤導。

與誤導陪審團類似的情況，還有陪審團可能因為 AI 技術聲稱準確或聲稱客觀，而賦予系爭 AI 判定結果過高的證據價值。這反映了心理學上「自動化偏誤」(automation bias) 的概念，即指人類傾向去相信自動化決策系統提供的答案，而忽視或不去尋找其他矛盾的資訊 (Cummings, 2004)，因為他們相信自動化決策系統是更值得信賴且更為客觀的。常見的現實案例是人們誤信 Google 導航，

³⁶ FED. R. EVID. 403. 亦可參見 Grimm et al. (2021: 86)、Lilly et al. (2015: 33-34, 43)。

反而將汽車駛入河堤。人類過度依賴自動化決策系統的結果，除了可能繼承其偏見與歧視，還可能因為太過信賴 AI 系統假性客觀的答案，忽略其他單獨於系統之外的客觀證據。

(三) 小結

在前述醫療影像辨識的例子中，法院必須去審視影像辨識技術在設計或建立的過程中是否形成偏見與歧視，並且檢視機器訓練時的定義與其訓練資料的使用之間是否存在不當或錯誤，以避免誤導陪審團。更重要的是，法院應注意陪審團是否會過度提高系爭判定結果的證據價值。如果法院最終認為系爭證據可以在論罪階段使用，法院應再次提醒陪審團，須綜合一切證據資料，依據經驗及論理法則來判斷證據之證明價值是高或低，以防止陪審團落入自動化偏誤。

三、AI 提出之判定結果資料與品格證據及習慣證據

(一) 品格證據與習慣證據之概念

依據 FRE 第 404 條第 (a) 項第 (1) 款，關於某一人之性格或性格特徵之證據，原則上不能用來證明其在某特定場合之行為符合該性格或特徵。³⁷ 惟在極少數的案件中，如果案件的核心爭點涉及品格，則無禁止使用品格證據的問題（金孟華，2021: 144）。³⁸ 品格證據的問題點不在於其缺乏證據關聯性，而是陪審團會過度看重這種證據，導致預斷的危險，進而阻礙公平審判。因此，法規原則

³⁷ FED. R. EVID. 404.

³⁸ 也就是說，品格證據作為直接證據時，原則上可以使用。但其作為情況證據時，原則上不可使用。

禁止使用此類證據，但根據 FRE 第 404 條第 (a) 項第 (2) 款及同條第 (b) 項，仍有允許使用品格證據的明文例外。³⁹

再者，FRE 第 404 條並未定義「品格」一詞，導致美國的學說及實務中呈現不同的解釋。有認為「品格」係描述一個人的性格或性格特徵，若某一人具有「照護／小心」之性格，我們會預設他在生活的不同情況中傾向於謹慎行事。⁴⁰ 另有認為，品格體現之傾向，來自於個人的「道德選擇」，並且觀察者在知悉該個人之行為選擇後，會對其有更多或更少的好感。⁴¹ 也有認為「品格」是個人特定作為或不作為的性格或行為傾向 (Anderson, 2012: 1924)。⁴²

品格必須與習慣及例行性作為相區分。「習慣」是在面對重複之特定情況會表現出的通常反應，例如左轉時習慣打手勢、上樓梯時習慣一次踩兩階。習慣行為時常是半自動或未經思考的，是在相對狹隘且特定情境下之反應，並形成劃一的模式。⁴³ 依據 FRE 第 406 條，習慣證據原則上可以用來證明某人或組織的行為與其習慣及例行性作為相一致，而其之所以原則上允許在法庭中使用，是因為法律實務及學界多數認為，它對於預測一個人的行為具有較高的準確性 (金孟華，2021: 147)。

³⁹ 關於品格證據之明文例外可詳見金孟華 (2021: 145-146)、Lilly et al. (2015: 81-86)。

⁴⁰ FED. R. EVID. 406 Notes of Advisory Committee on Proposed Rules.

⁴¹ *Id.* 相關論證亦可參見 Keith v. State, 152 S.W. 1029, 1030 (1913) 及 Anderson (2012: 1945)。

⁴² 此說與 2021 年美國奧勒岡州最高法院判決於品格之定義相同，參見 State v. Skillicorn, 367 Or. 464, 479 P.3d 254 (2021); 我國學者亦從此說，參見金孟華 (2021: 143-144)。

⁴³ FED. R. EVID. 406 Notes of Advisory Committee on Proposed Rules; Lilly et al. (2015: 93)。相關內容亦可詳見 Park et al. (2022: 307)。

(二) 品格證據與習慣證據在 AI 判定結果資料之適用關聯

有認為在性犯罪之特殊案件，演算法推導出的行為模式與傳統意義上的品格證據或習慣證據具有相似性，因此可以考慮參考相關規則並賦予其一定的證明價值（江湖，2020: 153）；但在此觀點中並未詳細說明為何只限定在性犯罪案件有其適用可能性。因此，本文嘗試探討 AI 之判定結果資料是否可能與習慣或品格有關，並應於審判中普遍納入審查環節。

涉及 AI 技術之證據可能在法庭中呈現眾多型態，例如普遍使用之智慧手錶、智慧家電等皆屬之。這些智能設備持續蒐集裝置使用人的每日活動資料，藉此也可以量化形成使用者的常規模式、行為和生理反應等生活圖像。AI 甚至因此可以預測或是比使用者意識到之前，更早地提供、提醒使用者即將出現的需求（Ferguson, 2023）。而智慧型裝置內部儲存的人類行為紀錄、其紀錄量化後的趨勢圖，或是 AI 的提醒紀錄，很可能作為習慣證據在訴訟程序中被提出，以主張證明某待證事實。

又例如汽車配備疲勞駕駛偵測系統，其持續監測並記錄駕駛者手握方向盤時的身體狀態、體溫、臉部活動以及車輛行駛於道路上的位置。當系統發現駕駛人的典型姿勢、頭部位置、眨眼頻率、面部表情、方向盤轉動模式出現不規則時，它將警告駕駛人停車休息。⁴⁴ 此類透過 AI 影像辨識產生的警訊紀錄，也可能被用作證明某駕駛人於發生交通事故時具有過失的證據（Gless, 2020）。此外，長期的疲勞警訊紀錄可以量化呈現出某人駕駛行為的疲勞圖像，此疲勞常態的紀錄，可以視本案待證事實內容而成立 AI 之習慣證據或品格證據。詳言之，疲勞本身是一種身體狀態，並非習慣；但如

⁴⁴ 有關疲勞駕駛之各項生理反應可詳見 Dong et al. (2011: 598-599, 604)。

果是在前往某處途中都有 AI 判斷為疲勞之紀錄，那麼這些紀錄可能可以被視為駕駛者前往特定地點有開車習慣的證據。⁴⁵ 然而，如果是某人常在去完酒吧後出現疲勞駕駛紀錄，那麼當他某日去完酒吧後開車時發生車禍，因此被起訴過失傷害罪時，該疲勞常態紀錄可能成為 AI 之品格證據。因為該疲勞紀錄表彰行為人具有輕率的性格，亦即在已知自己每次去完酒吧後都處在疲態駕駛的狀態，行為人卻仍然選擇於此時駕駛汽車，並確信其駕駛行為不至於造成危險。

再者，前述提到再犯風險評估系統，根據該系統判讀風險的參酌因子而言，其預測結果構成 FRE 第 404 條第 (a) 項第 (1) 款與同條第 (b) 項禁止使用之品格證據。例如 COMPAS 內一般再犯風險量表 (The General Recidivism Risk Scale) 的參酌因子包括曾經被逮捕或判刑入監之紀錄、職業與教育問題、使用毒品紀錄、第一次受逮捕之年齡等；又如暴力行為再犯風險量表 (The Violent Recidivism Risk Scale) 還加入受預測者的暴力與不服從行為紀錄 (Abu-Elyounes, 2020a)。這些參酌因子反映出受測者可能具有不良性格，例如曾使用毒品、曾因刑事犯罪而受逮捕或判刑，或具有暴力行為之高度可能性。依照品格的定義，這些不良行為是受測者實質選擇下的作為，否則不良行為通常不會發生。且當任何他人知悉受測者曾有這些性格與行為時，很難避免對受測者形成好感度上升或下降的評價，符合品格的定義。如果進一步將這些由曾被逮捕、

⁴⁵ 舉例而言，某甲每天下班之後都會開三個小時的車前往某處，但某甲每次開車達到兩個小時的時候，都會因為疲勞的神情及姿態，觸發疲勞駕駛偵測警告某甲應休息。故當案件之爭點是「某甲前往前述所指的處所時，是否有開車的習慣？」，每天下班後某甲疲勞警訊的紀錄，就可以成為證明某甲前往系爭處所有選擇駕駛汽車習慣的證據。

判刑或使用毒品等紀錄匯聚形成的再犯風險預測結果，主張以證明被告在本案中之犯罪行為符合該性格時，此證據就落入 FRE 第 404 條第 (b) 項第 (1) 款禁止使用的品格證據。也因此，美國法院目前僅在量刑階段參考再犯風險評估結果。

(三) 小結

總的而言，AI 作成之判定結果，根據待證事實之內容，可能同時或分別成立 AI 的習慣證據或 AI 的品格證據。這也意味 AI 作成之判定結果不能只考慮到它是一項數位形式之證據而已。此種證據也可能符合其他特殊的證據種類，並應注意涉及禁止使用的狀況。當然，某些 AI 技術輸出的判斷內容可能完全與品格或習慣無關，例如醫療影像辨識的案例。

(四) 補充說明：成立 AI 品格證據或習慣證據後的其他問題

在 AI 的習慣證據方面，有另一個值得一提的爭議：智慧型裝置產生的數位型習慣證據，是否真的表示人類主動的行為？還是實質上，人類是被 AI 系統的通知所驅動？例如，使用者可能是因為智慧手錶提醒要定時活動、呼吸調節才驅動其生活習慣。如果將智慧手錶的身體活動資料，直接視為個人之習慣證據，有認為非正確且非公平 (Ferguson, 2023)。因此，AI 生成的習慣證據可能不適合被視為原則可以使用的習慣證據類型。

再者，由於 AI 的判斷往往是基於受測個人與其他相似情況的比較結果，在尚未解決 AI 的可信性質疑與排除偏見歧視問題以前，有認為「預測結果」不應該成為證據之替代品 (Ferguson, 2023)。例如，再犯風險評估結果僅彰顯再犯之趨勢與傾向，但「趨勢與傾向」不等於再犯的「必然」。學者指出，採用大數據預測系統等於

是在懲罰行為人之傾向，而非懲罰其行為，而「懲罰行為人之傾向」等於是在否認人之自由意志並侵蝕人性尊嚴 (Mayer-Schönberger & Cukier, 2013: 170)。然而，此問題實則並非僅限於大數據預測的判斷獨有。在刑事訴訟程序之緩刑宣告、假釋的裁定，或預防性羈押的裁定，實際上也是法院以人為對被告進行再犯風險的「預測」，並依據「趨勢與傾向」作為決定的基礎，同樣存在侵害人性尊嚴的問題。

然而本文認為，人為的推論過程與大數據預測之間仍存在差異。其一，在緩刑宣告、假釋裁定及預防性羈押程序中，法院是針對個案的刑事被告進行獨立、個別化地再犯風險評價；但是大數據預測系統則需區分情形。例如智慧手錶此裝置，藉由平日蒐集使用者個人的運動頻率、運動習慣及身體數值資料，進一步利用這些個人資料形成裝置使用者的身體資料圖像，並供日後評估使用者身體復原、體能回復的頻率，其資料圖像的運算基礎，固然是以個案化因子進行評估。但是例如疲勞駕駛偵測的技術，是基於技術人員輸入之預設特徵，經由機器學習建構出大數據式、通案式的疲勞圖像之後，再以這些大數據建構出來通案式的疲勞圖像作為判斷標準，與現實駕駛者是否符合統計上的疲勞特徵，進行相關性的比較 (Dong et al., 2011)。亦即，倘若大數據的預測或評估過程傾向於統計式 (統化) 的相關性衡量時，這與法院針對犯罪行為人進行個化因子審酌的方式有所不同。⁴⁶

⁴⁶ 本文提到「個化判斷」與「統化判斷」之旨，除了為表達個案特徵與統計相關性之不同以外，亦有參考刑事鑑定程序針對證物區分「個化特徵」與「類化特徵」不同分析方法，個化特徵具有「獨一無二」的特性，二者在偵查程序呈現不同的功能。可參見曾春僑、莊忠進 (2020: 9-11)。

其二，本文已提到「自動化偏誤」的問題，這是使用大數據預測系統時可能獨有的情況。例如，智慧手錶可能根據大數據預測出目前的氣溫變化會增加裝置使用者突發心肌梗塞的風險。若裝置穿戴者知悉此事後仍執意開車，結果因心肌梗塞發生車禍。此時，若法院接觸了穿戴裝置於肇事前之警訊，可能因為相信 AI 對於心肌梗塞風險的預測是客觀正確的，進而更確信心肌梗塞與車禍發生之間具有因果關係，且依客觀歸責理論可歸責於行為人。然而，若法院未接觸到此 AI 證據，可能會更加注意心肌梗塞與車禍發生之間，是否有其他中斷客觀構成要件的情況，並考量將行為人執意開車的時點，作為認定行為人已具備主觀過失，可能已將過失行為判斷的時點，太過前移。亦即，行為人決定開車之時點，與車禍案件發生的肇事行為時點，已相距太遠，二時點之間關於肇事行為與肇事行為責任的認定已過於分離。亦即，當屏除大數據預測結果造成的自動化偏誤時，人類的預測決定多半更融合社會事實裡經驗式和直覺式的思維。相比之下，AI 之判定過程是客觀和統計式的判斷規則，而人類的預測更具有彈性。且人類對單獨個案的觀察上不易忽略個別細緻化的差異，在審酌因子的考量上，可能也比預測系統更完整、全面。⁴⁷

趨勢與傾向固不等於必然，惟 AI 大數據預測結果與人類的預測行為相比，更有可能發生偏誤或過度狹隘化因子審酌的範圍。更何況，這些系爭判定結果尚欠缺可供大眾檢視的說理過程。學者提到，人類很可能受到大數據分析能力之誘騙，採用了非真正契合的技術，或是對技術分析結果過度自信 (Mayer-Schönberger & Cukier, 2013: 170)。因此，面對 AI 預測系統，我們應該更加謹慎

⁴⁷ 類似之觀點可參見 Metallo (2020: 2059-2061)。

地考量這些預測、判定型的 AI 是否如聲稱那樣客觀、周全，並注意避免自動化偏誤的發生。

四、AI 提出之判定結果資料與驗真

(一) 驗真之概念

依據 FRE 第 901 條第 (a) 項，「驗真」係指出證者必須提出外部證據，以證明他所宣稱與案件相關的某項證據，與他提出於法庭上的證據資料是相同的。驗真檢驗的是證據資料「形式的真實性」。⁴⁸

無論是「人證」或「物證」，皆需進行驗真。「人證」的驗真，是在確認到庭之人是否確為檢察官聲請傳喚之人，亦即身分的確認。此過程可藉由查看到場人附有照片之身分證、駕駛執照等證件，驗真人別的真实性 (李榮耕，2014: 183)，且通常在主詰問的程序完成。「物證」的驗真，常見的標的物是文書證據及其他實物證據，但需要驗真之項目不限於有形事物。驗真程序亦無固定規則，沒有必須遵守的證據程序或慣例，其彈性的設計使法庭兩造可以創意地思考如何呈現某一項證據符合驗真的方法 (Lilly et al., 2015: 53)。

(二) AI 提出之判定結果資料在執行驗真之問題

對於 AI 判定結果證據的驗真程度，存在不同見解。有認為驗真並不處理證據實際內容是否可信、準確等「實質的真實性」之判斷 (蘇凱平，2021b)。但有認為，依據 FRE 第 901 條第 (a) 項，此類 AI 證據的驗真，應端視出證者對於證據真實的宣稱程度。⁴⁹ 如

⁴⁸ FED. R. EVID. 901(a). 概念理解參見蘇凱平 (2021b: 1038)。

⁴⁹ 參見 Grimm et al. (2021: 90-92)，該文作者提到：「倘於審判中提出來自 BlueDot [係一加拿大公司] 開發之演算法生成之證據，則提出該證據之一方，必須證明該演算

果出證者宣稱「系爭 AI 及生成資訊為有效且可信」，則驗真時應驗證出證者聲稱「AI 及其運算為有效且可信」一事是否真實。若出證者僅宣稱「系爭判定結果確實由案發當時位在現場的 AI 所產生」，則驗真僅需證明「提出於法庭的 AI，的確是案發現場的 AI，且系爭判定結果確實係該 AI 產生的」。

再者，如果主張應驗真「AI 及其運算為有效且可信」時，可能存在兩種狀況：第一種是 AI 作為物證的驗真。若出證者宣稱他提出的 AI 及其判定是有效且可信的，依據 FRE 第 901 條第 (a) 項，應驗真出證者之主張是否真實。第二種狀況是假設 AI 作為一位專家證人的驗真。⁵⁰ 首先需要比對出證者提出於法庭的 AI 是否為其主張的系爭系統。例如該 AI 必須是案發現場的 AI，且是生成系爭判定的 AI。此外，由於 AI 是被主張為一位「專家」證人，故還需驗證該 AI 技術，是否確實具備判斷或運算的「專家」能力。而 AI 是否具備「專家程度的」任務判斷能力，應檢視技術的實質內容，以判斷其生成之結果是否可被認為是有效且可信賴的。

由於 AI 技術建立程序之複雜性且涉及多種專業，加上獨立運算過程未知，AI 判定結果作為證據的驗真上，具有以下爭議。

1. 受傳喚為驗真者

FRE 第 901 條第 (b) 項例示 10 種驗真方法中，有認為最適合

法如何得以實現所聲稱之準確及可信賴等功能」(If evidence derived from use of the BlueDot algorithm was being offered into evidence at trial, the party seeking to introduce it would be required to show how it could accurately and reliably accomplish what its developers claimed it could) (92)。該文作者對於驗真程度之論述較為隱晦，並係由本文觀察該篇文章之前後文，推論作者對於驗真之見解，因而本處特別呈獻原文關鍵段落，供讀者參閱。

⁵⁰ 有關 AI 作為一般證人或專家證人的假設，將於「傳聞法則」子題中更詳細地說明。

用來驗真 AI 判定證據之方式為第 (1) 款「(提出另一項) 證詞去證明某事物為其所宣稱」與第 (9) 款「(提出) 一項描述運作過程或系統之證據，且顯示其產生準確結果」(Grimm et al., 2021)。⁵¹

為了驗真而提出之「證詞」同樣必須符合其他證據規則之要求。例如依據 FRE 第 602 條，證人必須先提出證據，證明其對於待證事項有個人之親身見聞或相關知識，才能為特定事項作證。⁵² 也就是說，在以證詞描述技術過程來驗真 AI 判定結果上，提出證詞之證人必須對於系爭 AI 技術具有相關知識或個人見聞。

透過瞭解機器學習的五個步驟，可見技術建立的過程繁雜。除了需要 AI 專家以外，諸如醫療影像辨識技術或是再犯風險預測系統等，可能還需要醫療、心理、法律、犯罪學等具備特定專業知識之人參與技術前階段的概念定義。因此，若要透過證人證詞驗真 AI 判定結果為有效且可信時，法庭上可能需要傳喚包括機器學習各個技術階段中，對該 AI 有個人見聞或知識的專業技術人員；惟受傳喚證人之數目與司法資源有限性及效率有關 (蘇凱平，2021b; Grimm et al., 2021)。此外，使用 AI 之人，也必須到庭說明：在 AI 產生判定結果當時，他是如何操作系統及得到什麼運作結果 (Grimm et al., 2021: 91)。因此，從研發團隊到操作者之間，可能需要傳喚為數不少的證人到庭。

2. 驗真非能實質有效地審查 AI 判定結果之證據資格

透過「一項描述技術運作的系統或流程的證據」來驗真，則對應到專家證人於科學知識可信性審查的問題。詳言之，由於 AI 存在黑盒子問題且目前尚無客觀的可信性測試標準，法院在審查技術

⁵¹ 亦可參見 FED. R. EVID. 901(b)。

⁵² FED. R. EVID. 602. 法條解釋參見自張瑋心 (2016)。

可信性時受制於法律黑盒子。又針對系統當次是否「產生準確結果」，受限於技術黑盒子，即使是 AI 專家也無法瞭解 AI 判定過程的論理，他們至多藉由提出訓練資料內容來間接證明系統可能會產生準確的結果。

再者，無論是將 AI 提出之判定結果認為是物證或人證，倘若出證者僅主張驗真至系爭證據「形式的真實性」，即僅限於人別或物證的同一性，那麼這也導致 AI 技術建立過程中的人為設定、系統存在的瑕疵、機器訓練階段形成偏見與歧視等爭議問題，將無法藉由驗真的程序進行審查或提出來進行攻防辯論。此時，單就驗真程序恐怕無法實質發揮證據資格審查制度之機能。

3. 輔以證據開示驗真

以 AI 判定結果為物證的前提，有認為可透過證據開示 (disclosure) 制度來輔助執行驗真。證據開示係使法庭中兩造藉由揭露證據資料，因此得預先、充分為審判做準備，包括為抗辯而傳喚證人或提出新論點 (Seng & Mason, 2021)。

要驗真 AI 之判定結果資料是否如出證者聲稱之有效和可信，勢必須參考原程式編碼、機器學習演算法，或研發人員曾於系統建立過程中，為提升 AI 之效率、有效性所採取的任何權衡措施。針對取得智慧機器內部資訊的方法，有主張應制定「『數位布雷迪』規則」(“digital Brady” rule) (Gless, 2020: 230-232)，此係布雷迪規則 (Brady rule) 的變型。布雷迪規則來自 *Brady v. Maryland* 案，本案美國聯邦最高法院要求檢方向辯方提供任何可能改變審判結果的重要證據。若檢方未提供證據，無論出於善意或惡意，皆屬違反被告正當法律程序之保障。⁵³ 但 AI 系統的證據開示，與 AI 技

⁵³ 373 U.S. 83, 87 (1963).

術所有者的營業秘密、智慧財產權之保護產生衝突，此即對應科學證據可信性要素 2 (未能經同儕審查亦無出版) 的情況。⁵⁴ 因此，證據開示可取得之資訊恐怕仍受營業秘密等權利保護的限制 (Gless, 2020; Seng & Mason, 2021)。但有學者指出，法院可藉由保護命令來填補對於智慧財產、營業秘密等權利的損害。例如參考美國民事訴訟程序的作法，由法院命令僅由特定專家取得軟體程式編碼，或由法院命令訴訟程序不公開進行、命令要求遮蔽與智慧財產權有關的技術內容，或是命令應銷毀所有與軟體智慧財產權相關的證據、筆記等。法院可裁定不同內容的保護命令，因此不需要一概阻擋軟體的證據開示 (Imwinkelried, 2016; Seng & Mason, 2021)。

在酒精吹氣測試等自動取證技術 (automated forensic technique) 上，也已存在是否應揭露原程式碼的爭議。其中有認為，亦可透過揭露取證技術在有效性研究 (validation studies) 測試中的書面資料，來取代直接揭露技術原程式碼。在有效性研究的主張程序上，檢方先提出系爭技術有效性研究之證詞或書面證據之後，辯方須反證該測試範圍未能充分解決本案的事實應用，例如未包括本案存在的特定條件或變量。法院並且要求辯方之反駁須由專家說

⁵⁴ 因此，當出證者主張技術相當有效可信，驗真的程度，會涉及技術「可信」與否，並依據 FRE 第 901 條第 (a) 項、第 901 條第 (b) 項第 (9) 款的規定，驗真的程序因此涉及與 FRE 第 702 條科學證據可信性要素一樣的問題。本文仍重複提到可信性的問題，除了 AI 技術的本質使然，本文也在強調，在證據法的審查上，科學證據之證據資格門檻 (道伯法則) 與驗真的證據資格關卡，應是兩立的證據法程序權利及攻防方法。此部分從美國憲法第 6 修正案 (Confrontation Clause of the Sixth Amendment to the United States Constitution) 保障刑事被告質問權來看，亦是當然解釋。本文認為，儘管存在道伯法則的要素門檻，這並不當然使得兩造失去於驗真程序挑戰證據可信性的程序權。即使系爭證據通過道伯法則，也不當然意味著任何一造失去了在驗真程序提出動搖技術可信性之攻防方法的權利，來影響事實的審判者對於系爭證據資格之評價或其證明力的心證高低。參見 Imwinkelried (2016: 117-120)。

明，以說服法院這些特定條件或變量的缺漏，足以影響有效性研究之可信性。當法院認為辯方已充分證明有效性研究存在缺漏時，法院將給予系統製造者兩個選擇：一是命令製造者允許辯方測試系爭技術在本案事實之應用，包括有效性研究中缺少的重要條件或變量；二是命令製造者提供軟體程式碼給辯方 (Imwinkelried, 2016)。然而有認為，有效性研究的證明方式只足夠適用在系統運作過程定義明確並於科學上可供檢驗的簡單運作，例如法庭中常見的鑑識化學 (forensic chemistry) 檢驗。⁵⁵ 對於 AI 系統的複雜性，有效性研究測試可能衍生更多問題。例如「有效性研究」應執行幾次測試？進行測試之程序？執行上應該費時多長？再者，即使已提供系統有效性研究的所有檢驗歷史，也並未必能夠說明 AI 正確地處理了「本案」的判定結果 (Seng & Mason, 2021: 276)。況且，即使技術所有者公開程式碼，被告可能被成千上萬的編碼、資料淹沒，不僅檢閱上耗費時日 (Imwinkelried, 2016: 130, n. 235)，更遑論被告恐怕不具備足夠的知識、能力或財力，可以實質地去審視、挑戰這些軟體工程及統計資料。

(三) AI 提出之判定結果資料在自我驗真之問題

FRE 第 902 條列舉 14 種可推定已經過驗真的證據資料，是豁免出證者須提出外部證據驗真的例外規定，惟他造當事人仍可爭執及推翻經自我驗真 (self-authentication) 之證據 (蘇凱平, 2021b; Lilly et al., 2015)。與 AI 技術相關自我驗真方式，應為 FRE 第 902 條第 (13) 項。當系爭紀錄由電子流程或系統產出，並由符合資格之人書面證稱該電子流程或系統確實可以產出準確結果時，該

⁵⁵ 學者亦指出，系統之有效性測試也只是「對於系統本身檢驗」的「替代」(proxy)。參見 Seng & Mason (2021: 276)。

系爭紀錄即可符合自我驗真的條件。⁵⁶ 但如前述，即使是自我驗真，符合資格之證人也須描述 AI 技術是如何達成準確結果。而在未解決黑盒子問題及欠缺可信性審查測試以前，無論是驗真或自我驗真，都無法直接驗證系爭該筆判定結果是否有效、準確，至多僅能由相關證人表示 AI 技術判斷準確率高等之非直接對於 AI 決策論理過程的描述。

(四) 小結

自我驗真的功能，只是讓兩造可在審判前先行決定本案涉及之數位證據是否確實有真實性的疑慮。如果兩造不爭執其真實性，則可適用 FRE 第 902 條第 (13) 項自我驗真程序，因此不必以證人到庭作證的方式，從而簡化數位證據的驗真程序，這是司法在費用與效率考量下的法規增訂 (蘇凱平，2021b)。

綜上所述，AI 之判定證據若適用 FRE 第 902 條第 (13) 項的自我驗真，可對應解決傳喚證人數眾多的問題。而若採取證據開示制度輔助驗真，可對應到驗真無法實質審查 AI 技術內容可信性的問題。然而，證據開示制度與營業秘密抗辯存在權利保障的衝突，因此有認為有效性研究測試可作為替代方案。但總結而言，無論是自我驗真或有效性研究測試，實際上都不是直接檢視 AI 涉及本案系爭判斷結果之論理過程。此外若採公開程式碼的方式，對於缺乏專業知識之被告而言實亦無理解、挑戰這些程式碼之能力。

故而，謹守形式真實性的驗真方式，對於 AI 判定結果比較不會發生審查方式的疑義，惟其無法評價 AI 技術內部存在之人證性

⁵⁶ FED. R. EVID. 902(13). 本條項是銜接在第 901 條第 (b) 項第 (9) 款而來，出證者可自由擇以「驗真」或「自我驗真」滿足程序要件。詳見 FED. R. EVID. 902(13) Committee Notes on Rules—2017 Amendment 及蘇凱平 (2021b: 1048)。

質。而若需要驗真 AI 判定結果至有效且可信的程度，無論採取哪一種驗真方式，都無法針對「系爭判定結果」本身進行直接的證據資格審查。由前述結論顯示，現行驗真制度對於 AI 判定結果此類證據，明顯未能審查其人證性質，也未能實質審查其判定結果本身的真實性，此制度在 AI 判定證據的審查功能有所不足。

五、AI 提出之判定結果資料與傳聞法則

(一) AI 提出之判定結果資料在傳聞法則之適用問題

美國證據法中，傳聞法則係採「主張取向」(assertion-oriented)為原理。FRE 第 801 條第 (a) 項提到，「陳述」係由人類做成，且其內容帶有「主張」(assertion) 的性質。「傳聞」則指某人之「陳述」非於本案法庭內作成，且該法庭外陳述係用於說服事實之審判者，使其相信該陳述內容為真實之目的。⁵⁷ 於法庭內提出傳聞證據，原則上不具證據容許性，除非符合同法第 802 條至第 807 條揭示例外允許之情況。⁵⁸

依據先前所述，電腦紀錄以其內容是否具備人類陳述的性質區分為二。由於電腦存儲紀錄僅是忠實地將人類陳述內容的資料進行儲存載體的轉換，因此區別電腦紀錄是否具備人類陳述較為容易，也能輕易地從形式上判斷電腦存儲紀錄是否為傳聞證據。相反地，AI 的判定結果乍看之下是電腦生成紀錄，容易使人未考慮到 AI 判定結果有無可能具備傳聞的主張，也不易區辨具有人類陳述之具體內容為何。但是若直接將 AI 的判定視為陳述，並適用傳聞法則，實際上會與現行法則的適用框架有所扞格，以下詳言之。

⁵⁷ FED. R. EVID. 801(c). 相關內容亦可詳見 Lilly et al. (2015: 153)。

⁵⁸ FED. R. EVID. 802. 相關內容亦可詳見 Lilly et al. (2015: 153)。

AI 做成之判斷、預測結果確實類似於一般證人或專家證人在法庭外作成證詞。例如在影像辨識中，AI 判定認為早期圖像已經呈現病變。惟依據現行 FRE 第 801 條法規之文義而言，傳聞法則只適用於「人類」做成之陳述，而 AI 本質上只是一項技術。在判斷或預測當下，AI 也不會意識到它正在做出一項證明醫師「有醫療過失」等與某個案系爭事實有關的證詞。因此，縱然可認為 AI 係作成法庭外陳述，也無法直接適用在現行傳聞證據的規範框架中。

但是若從技術取得知識的設計程序來觀察，由於 AI 所具備的知識，來自於帶有人類價值選擇的資料且再經由人為進行標註與教導內容，因此，將 AI 的判定結果適用傳聞法則，可能比以往想像的更合適 (Seng & Mason, 2021: 241)。學者指出，如果能將 AI 的輸出結果置於傳聞規則審查，有助於揀選出 (tease out) AI 生成證據資料中嵌入的人類主張 (human assertions)，並可進一步判斷不同的 AI 技術是否需適用傳聞法則 (261-263)。例如監督式機器學習的人為決定，係來自訓練演算法使用的資料，或是人員在標記或教導過程中輸入了過於主觀的價值觀；非監督式機器學習雖然不需要人類教導正確答案的過程，但在特徵值設定的環節，仍可能受到人類主觀價值的影響。

綜觀前述，有關 AI 提出之判定結果是否應適用傳聞法則，有反對與支持兩種見解。本文採「修正式的」支持 AI 之判定結果資料適用傳聞法則。理由在於，雖然 AI 的判定結果或判定報告乍看之下是機器產生，呈現為非供述性質之數位形式證據。但誠如本文從技術建立過程的觀察指出，AI 在訓練階段中，無論是訓練資料本身，或是技術人員於建立過程中的設定，都可能摻雜人類主張的內容，進而有傳聞證據存在的可能。惟與學者肯認 AI 判定結果適用傳聞

法則的不同之處是，本文認為應該將傳聞的評價重點，前移到技術建立的程序，而非只評價後階段 AI 輸出的判讀結果而已。

(二) 小結

本文的初步想法是，若 AI 的「判讀結果本身」直接適用在現行的傳聞法則，除非已經過修法或透過詮釋，否則機器陳述與人類陳述是否可一體適用傳聞法則仍有疑問，這也連帶影響同以「人類陳述」為規範的傳聞例外規定。如不考慮更動 FRE 第 801 條第 (a)、(c) 項文義明示的規範對象，既然技術建立階段是存在傳聞之處，似可考慮將評價重點前移至該階段。以一開始的例子而言，醫療影像技術的建立過程中，所使用的醫療資料、定義或選擇演算法的方式、人員的校正等都可能存在傳聞的成分，故可考慮將傳聞法則及傳聞例外的審查前移至存在人類主張的技術建立階段。且就該傳聞內容的可信程度如何，想必也是因法院信賴 AI 判定結果，因而將受法院在事實認定上不利益之一方，希望可以對之為爭執的。

六、AI 提出之判定結果資料與對質詰問權

(一) AI 提出之判定結果資料在對質詰問權行使之問題

Crawford v. Washington 案是關於對質詰問權之經典案例。本案最高法院認為，證人於法庭外做成之「證詞型陳述」(testimonial statements)，即使具有特別可信性，仍須經由對陳述者之對質詰問後，該證詞方得具有證據能力。⁵⁹

⁵⁹ 承傳聞法則對於「陳述」之定義，「證詞型陳述」即指具有主張的性質，且其主張內容是在作出證明或指控他人某事實的人類供述。參見 *Crawford v. Washington*, 541 U.S. 36, 65, 68-69 (2004)。

在電腦存儲紀錄或混合型電腦紀錄作為證據的情況中，如前所述，該紀錄內容是否為指控他人的陳述，以及其陳述者為何人，通常可直觀地文字內容去判斷是否構成證詞型陳述，並且可以進一步對陳述者質問。但是對於 AI 之判斷或預測結果本身，作為一項機器運算的結果，是否類似於在作證或指控他人，則是較具疑問的。如能認為 AI 之判斷或預測結果是證詞型陳述，依據 *Crawford* 案件提出必須經對質詰問之要求，須在質問 AI 以後，該 AI 提出之證詞型陳述，方可取得證據能力；但是如何對一項機器行使質問是另一個難題。如認為 AI 之判斷或預測結果並非證詞型陳述，則不須經過對質詰問。

藉由文獻回顧，目前多數研究直接切入討論對機器行使質問的方法，但是對 AI 判定結果是否確實屬於證詞型陳述的前提問題，卻未有提出討論。本文認為，例如 AI 係於案發現場當下即時做出駕駛者目前為疲勞狀態之判定結果，該判定結果實與證人處於案發現場之目擊證詞類似。又例如 AI 進行人臉辨識認為某 B 出現在案發現場，且更進行影像強化與修勻後認為某 B 有扣下板機的動作，則針對 AI 指稱的內容亦與專家證人證詞類似；其理由在於，AI 實以數萬筆資料的知識為分析後，提出人臉辨識及圖像推理的專業判斷。本文認為，如 AI 判斷內容之指稱，與本案實體法上爭議或犯罪事實存否具備合理關連時，可認為是 AI 做出「證詞型陳述」。

對於 AI 行使質問的方法，有不同見解。有學者曾主張，電腦軟體的運作實質上來自人類工程師編寫的軟體程式碼及指令，即便電腦得自行編寫程式，其編寫能力亦被框架在人類工程師曾經教導、設計的指令範圍，並吸收工程師設定之意志、指令與主張，因此，電腦軟體本質上是提供電腦科學家陳述用的「工具」(Chessman, 2017: 220)。本文延伸前述學者之見解認為，就發展自電腦軟體的

AI，技術人員在技術建立時，也是透過定義任務、對之下指令、輸入資料等行為進行陳述，可傳喚至法庭行使質問的對象應是 AI 技術建立人員。

另有認為，對於非人類資訊 (a nonhuman source) 行使實質的彈劾，可能需要採用其他的法庭檢驗機制。有學者即試圖打破目前的證據方法，以建立外部科學委員會的方式來實現。詳言之，其建議要求程式編寫者於科學委員會中做成證詞，且每當軟體有所更改，應再次於委員會中做成新證詞，學者並主張以此機制取代對程式編寫者在法庭中的質問程序 (Roth, 2017)。再考究該文註腳，此外部科學委員會，應係指可利用美國國家標準和科技機構 (National Institute of Standards and Technology; NIST) 中，法庭科學技術標準制定委員會 (Forensic Science Standards Board; FSSB) 在法庭科學標準之促進與頒布上所定期舉行之既存會議 (National Institute of Standards and Technology [NIST], n.d.)。⁶⁰

也有認為，為能實質彈劾非人類陳述之資訊，與其對軟體程式設計人員行交互詰問，不如在審判前讓兩造取得關於該 AI 技術之程式編碼、標準化後的數據資料、作為訓練之數據資料內容等，並且有機會對於該機器與演算法進行實驗與測試分析。此見解雖然類似前述的證據開示及有效性研究測試，惟學者更強調，前述對機器之測試，可認為實質上相當於一種「法庭外的對質詰問」(Gless, 2020: 239)；但法庭外質問的測試結果需由參與測試之專家證人引入法庭，並進行說明 (239)。

⁶⁰ NIST 網站所示 FSSB 係隸屬於法庭科學領域委員會組織 (The Organization of Scientific Area Committees for Forensic Science; OSAC) 的職掌委員會。OSAC 轄下有與不同法庭科學相關之小組委員會，其中有一專門處理數位證據之小組。針對 OSAC 的相關資料，可參見 NIST 網站：<https://www.nist.gov/osac/osac-subcommittees>

(二) 小結

雖然 AI 機器不會意識到它的決定或判斷是在主張證明某一事實，但是若人們嚴守著 *Crawford* 案件於證詞型陳述的定義，係指證明某一事實存在或成立的「嚴謹的主張」(solemn declaration)，則藉由提出黑盒子演算法產生的證據，勢必可成為一種躲避對質詰問審查的手段。⁶¹

關於無法質問機器的問題，學說見解不一。有認為以質問技術人員的方式取代，也有認為透過外部科學委員會發表技術內容解決。另有認為應使兩造取得技術資料，並容許對於機器進行測試等替代方案。本文則認為，後兩種方案都可能過度侵害企業之營業秘密保障或其智慧財產權。以既有的證據調查方式來說，即時且可行的做法或能考慮採行質問技術人員之方法。由於技術人員最清楚餵養訓練資料、問題定義及標註過程，法院亦可透過質問技術人員的過程，瞭解設計者的設定內容、AI 的判讀結果有無繼受偏見與歧視的不當影響。同時，在質問的過程中，法院亦能詳加瞭解 AI 判定結果所主張的內容，到底是僅僅計算相關程度，或該結果還表彰了因果關係的推理過程。此外，以質問技術人員的方式，也毋庸直接、全部地公開技術內容，既能取得更細節的資訊，卻又能限縮在與證據資格相關的部分。本文並且認為，採行此方法，可理解為藉著質問技術人員，間接地質問 AI 的證述。

⁶¹ *Crawford, supra* note 59, at 51. 另外參見 Roth (2017: 2047-2048)。

伍、結論：AI 提出之判定結果資料在證據種類與審查模式的建議

回到以醫療影像辨識結果證明醫療過失的假設例子，如果法院有注意到 AI 的黑盒子問題、偏見歧視的固化問題、缺乏事前驗證可信性的機制，以及對技術建立過程有基本的瞭解，那麼法院在考量此類 AI 證據與證據資格審查的方式，必然會受上述幾個特性的影響。

AI 之判定證據在適用個別證據法則的分析中，有一更根本的問題是此類 AI 之證據實際上同時具備了人證跟物證的內涵。除了可以認為是機器產生的資料，又可被認為具有經由機器分析判斷下的指證或主張之陳述性質，亦可認為陳述內容實則包裝在技術中；然而，儘管如此，無論採取人證或物證哪一種證據類型審查，目前的證據法則都可能無法妥善的適用。如擇定證據類型為人證，這將逸脫原本於傳聞法則及傳聞例外是針對「人類」主張的適用範圍，因此需要透過詮釋或打破法則的框架來解決，甚且對機器交互詰問的方式也會遭遇諸多困難。如擇定為單純的 AI 產生之物證，則無法實質審查 AI 是否存在偏見與歧視，以及 AI 在人為價值的介入與判斷結果 (AI 被決定的決定) 之間是否形成不良的影響。因此，一個可能的結論是，既然現行證據法則就此類證據實際上無法有效地發揮審查的機能，在未能實質通過證據法則的篩選程序的情況下，似應認為法庭中不能使用相關之證據。

然而，若直接因 AI 之判讀結果證據在目前的 FRE 存在適用上的困難，就一概否定此類證據在法庭內的論罪階段中使用，在 AI 快速發展的時代，可能不是對未來法制規劃上最完整的建議。例如在醫療過失的例子裡，若早期病灶的資訊已經出現在醫療文獻中，且

影像辨識技術學習了目前已經存在的所有專業醫療知識才做出判定，那麼該判定結果或可用來證明人類醫師有疏未注意的情況。

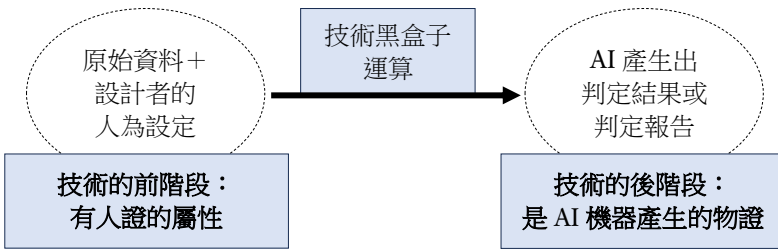
本文的看法是，AI 獨立提出判讀結果（例如 AI 在醫療影像辨識的判斷），雖然此時 AI 就像是一名證人或專家證人作出證詞，但 AI 的判定結果或判讀報告本質上仍是機器產生、經過黑盒子的數據運算的結果。又從 AI 不具有人格的見解而言，AI 在產生資料時也不會意識到其決定或判斷是在主張證明某一待證事實。因此，本文不認同推定 AI 之判定結果為證詞型陳述的見解。再加上即便認定 AI 之判定報告為證詞型陳述，如何對機器質問仍是問題。基於最終 AI 提出判定結果或報告的計算過程，純然是由機器運算出來的數據結果或資料，本文傾向以非供述證據的物證視之。⁶²

但若擇定屬機器「物證」作為唯一的證據方法，誠如前述，在適用驗真程序發生以下問題：(1) 因黑盒子計算過程，當驗真內容主張判定結果為有效及可信時，面臨各驗證方法妥適性的疑義；(2) 採取自我驗真或有效性研究測試，實際上都不是對於 AI 技術「提出系爭判斷結果過程」的直接檢視；(3) 若僅適用「形式的」驗真程序，則無從辯證技術內部是否存在因原始訓練資料的先前揀選，或設計者的系統設定等所導致的偏見問題或價值選擇。

藉由對於 AI 技術在機器學習程序的瞭解，並針對 AI 之判定結果逐一進行證據法則的適用分析以後，本文主張應將 AI 之判定結果資料，視為一種兼具人證與物證性質之第四種數位證據。採取兼具雙重性質的見解，可解決擇定一種證據方法無法妥適發揮審查機能的問題。

⁶² 再加上學者對於 AI 之判定報告為證詞型陳述的見解亦無明確表示，僅是本文以邏輯論證的推理可得出的推定。

AI 之判定結果「兼具人證及物證之雙重性質」，係以回溯觀察 AI 技術過程的方式，將產出判定結果的「技術建立程序」加入證據資格審查的範圍，本文稱之為技術的前階段；而 AI 提出的判定結果或判定報告本身，稱為技術的後階段（參見圖 1）。



資料來源：作者自行繪製。

圖 1 AI 技術建立至 AI 獨立提出判斷報告之前、後階段示意圖

AI 可以提出判定結果，係仰賴技術建立前階段的機器訓練與人員設定，這其中可能包含人證的性質。首先，作為 AI 知識基礎的原始資料本身可能已經融入了人為價值。其次，於技術建立的過程中，技術人員的教導內容、資料蒐集的內容、資料使用的方式、資料變形的方式及演算法設定等機器學習，會形成人類主導的知識框架，並影響 AI 爾後判定的價值觀及知識範疇，使人類間接地介入 AI 的決定。因此，基於支撐 AI 的決定，係形塑自技術前階段的資料學習及具備人為價值的設計，可考慮將傳聞法則的評價重點放在支撐 AI 決定的原始資料及技術設計者之設定行為上，並將傳聞例外的審查重點放在原始資料的可信性。這樣的論證結果，也可以避免 AI 判定之結果並非目前傳聞規定的適用範圍，即武斷地認為無從審查 AI 判定證據中具備傳聞的內容。

在對質詰問權的行使上，因為 AI 目前不是可質問的主體，現階段可行的做法是，如果在技術前階段中（例如在判定目標的定義、訓練資料的標記、數值的設定等技術建立過程）具備證詞型陳述的內涵，則以技術設計人員為行使質問的主體。透過質問設計者的過程，使法院可實質審查技術內是否嵌入不當偏見或歧視的價值觀，並可藉由質問過程更有效率地瞭解其相關性的計算內容。法院也可藉由質問設計者的過程，以更聚焦的方式，瞭解與證據資格有關的技術細節。例如技術應用在什麼情狀下，可能導致較高的判讀錯誤率、技術的運算上是否採取了構成品格證據的判讀因子。以此做法，法院可特定在調查與證據資格取得有關聯且為訴訟上兩造有爭執的部分，也無需直接公開技術內容或其程式碼。另一個好處是，在質問技術人員的過程，由於法院易於理解技術內容，因此可決定是否存在 FRE 第 403 條揭示之不利影響的情況。

在技術的後階段，本文認為，AI 最後產生出來之判讀結果，仍是一項機器經過黑盒子數據分析所產生的物證。因此，AI 之判定結果仍應適用驗真程序，以確保系統產生出來之內容是提出 AI 證據之人所宣稱之內容，或者需要驗真至宣稱 AI 運作有效且可信的程度。當然，如需驗真至 AI 為有效且可信的程度，勢必因黑盒子問題產生困難。但若藉由傳聞法則與質問程序加入 AI 判定結果的審查方式，應可某程度解決黑盒子問題產生之疑慮。最後，在科學證據可信性的審查同樣面臨無法驗證 AI 技術可信與否的問題，亦即 AI 技術擁有者可以法律黑盒子為抗辯，而技術黑盒子也導致審查可信性有事實上的困難，法院恐怕認為此類證據無法通過道伯法則。但同如前述，藉由提前在技術前階段適用傳聞法則及質問程序，以法院質問技術人員來限縮營業秘密揭露的範圍，應可某程度削弱法律

黑盒子原本以營業秘密抗辯權為反駁的高度可採性，也可某程度解決技術黑盒子導致法院驗證判定結果可信性的困難。

可能有讀者會認為，區分前後階段的見解與混合型電腦紀錄的證據法則適用方式相同，但本文認為實則不然。在混合型電腦紀錄中，電腦存儲紀錄適用傳聞法則，是由於該紀錄內容本來就表彰著人類輸入的文字；但在 AI 的情況，傳聞的內容是包裝在程序及判定結果之內。這使得我們必須撥開技術的外殼，探知技術之內裡是否存在使 AI 判定結果應喪失證據資格的情況。又，AI 產生判定結果與電腦生成紀錄的資料內容並不相同，已如前述。因此本文認為，基於技術建立過程拆分適用人證及物證不同證據方法的見解，與混合型電腦紀錄存在兩種不同性質的電腦紀錄並分別適用不同法則的情況，並不相同。

AI 的發展日新月異，AI 可能與人類的日常生活逐漸密不可分。本文以具備機器學習技術的 AI 判定結果資料，探討其於美國證據法的適用及可能的法制問題，期望對於新興科技在刑事程序證據法則的適用上，提出一些新的看法。以目前而言，雖然社會大眾及司法實務對於 AI 技術的可信性還存在質疑，但如同過往的 DNA 檢測儀、GPS 追蹤等科技工具，這些工具的可信性曾備受質疑，現今受司法實務廣泛採用。隨著 AI 技術不斷發展與精進，當有朝一日社會對於 AI 的信賴，超越面對黑盒子過程未知的恐懼時，法院在技術可信與否的審查，勢必需要為之因應。此時，針對涉及 AI 的證據在資格審查保持嚴謹，方能避免法庭受科技斷案所凌駕，或發生技術人員間接主導的法院生態。

參考文獻

- 山口達輝、松田洋之 (2020)。《圖解AI：機器學習和深度學習的技術與原理》(衛宮紘譯)。碁峰資訊。(原著2019年出版)(Yamaguchi, T., & Matsuda, H. [2020]. *Illustrated AI: Techniques and principles of machine learning and deep learning* [H. Emiya, Trans.]. Gotop.) (Original work published 2019)
- 朱帥俊 (2011)。〈論電子證據之分類與傳聞法則〉，《司法新聲》，99: 37-52。(Chu, S.-C. [2011]. Classification of electronic evidence and the hearsay rule. *Judicial Aspirations*, 99: 37-52.)
- 江溯 (2020)。〈大數據在刑事司法體系中的應用及其問題〉，《月旦法學雜誌》，304: 135-154。(Chiang, S. [2020]. The application of big data in the criminal justice system and its problems. *The Taiwan Law Review*, 304: 135-154.) <https://doi.org/10.3966/1025593130408>
- 李榮耕 (2014)。〈刑事審判程序中數位證據的證據能力——以傳聞法則及驗真程序為主〉，《臺北大學法學論叢》，91: 169-211。(Li, R.-G. [2014]. The admissibility of digital evidence at criminal trial—Focusing on the hearsay rule and authentication. *Taipei University Law Review*, 91: 169-211.)
- 李榮耕 (2018)。〈初探刑事程序法的人工智慧應用：以犯罪熱區為例〉，劉靜怡 (編)，《人工智慧相關法律議題芻議》，頁120-152。元照。(Li, R.-G. [2018]. Initial inquiries about the application of AI in criminal procedural law. In J.-Y. Liu [Ed.], *On the legal issues of AI: A primer* [pp. 120-152]. Angle.)
- 李榮耕 (2022)。〈刑事程序中人工智慧於風險評估上的應用〉，《政大法學評論》，168: 117-186。(Li, R.-G. [2022]. AI risk assessment in criminal justice. *Chengchi Law Review*, 168: 117-186.) <https://doi.org/10.53106/102398202022030168003>
- 松尾豐 (2016)。《了解人工智慧的第一本書》(江裕真譯)。經濟新潮社。(原著2015年出版)(Yutaka, M. [2016]. *Illustrated AI: Techniques and principles of machine learning and deep learning* [Y.-C. Chiang, Trans.]. EcoTrend.) (Original work published 2015)

- 金孟華 (2021)。〈國民法官法第46條之解釋與適用〉，《檢察新論》，29: 136-149。(Chin, M.-H. [2021]. Interpretation and application of article 46 of the Citizen Judges Act. *Taiwan Prosecutor Review*, 29: 136-149.)
- 金孟華、陳又寧 (2015)。〈論圖案與印記證據之證據能力〉，《中研院法學期刊》，17: 423-476。(Chin, M.-H., & Chen, Y.-N. [2015]. The admissibility of pattern and impression evidence. *Academia Sinica Law Journal*, 17: 423-476.)
- 林勤富 (2022)。〈智慧法院之發展與界限 (上)——演算法、科技治理與司法韌性〉，《月旦法學雜誌》，323: 72-98。(Lin, C.-F. [2022]. The development and boundaries of smart courts [Part I] — Algorithms, technological governance and judicial resilience. *The Taiwan Law Review*, 323: 72-98.) <https://doi.org/10.53106/1025593132305>
- 孫宏民、呂沐錡 (2015)。《計算機概論》。高立。(Sun, H.-M., & Lu, M.-C. [2015]. *An overview of computer science*. Gau Lih.)
- 張志勇、廖文華、石貴平、王勝石、游國忠 (編) (2022)。《人工智慧》。全華圖書。(Chang, C.-Y., Liao, W.-H., Shih, K.-P., Wang, S.-S., & Yu, K.-C. [Eds.]. [2022]. *Artificial intelligence*. Chuan Hwa.)
- 張瑋心 (2016)。〈專家證人——司法語言學家〉，《軍法專刊》，60, 5: 153-174。(Chang, W.-S. [2016]. Expert witnesses—Forensic linguists. *The Military Law Journal*, 60, 5: 153-174.)
- 曾春僑、莊忠進 (2020)。《刑案現場處理與採證》。元照。(Tzeng, C.-C., & Chuang, C.-C. [2020]. *Practical crime scene investigation*. Angle.)
- 葉怡成 (2009年)。《類神經網路模式應用與實作》。儒林。(Yeh, I.-C. [2009]. *Application and practice of neural-like network patterns*. Scholars Books.)
- 劉昌誠、徐建業、陳炯旭、蕭百勝 (2010)。〈應用類神經網路建構妨害性自主罪再犯預測模型之初步嚐試〉，《亞洲家庭暴力與性侵害期刊》，6, 1: 43-64。(Liu, C.-C., Hsu, C.-Y., Chen, C.-H., & Hsiao, B.-S. [2010]. The preliminary attempt to use artificial neural network model for the sexual offender

- recidivism prediction. *Asian Journal of Domestic Violence and Sexual Offense*, 6, 1: 43-64.) <https://doi.org/10.29804/AJDVSO.201007.0003>
- 龍建宇、莊弘鈺 (2019)。「人工智慧於司法實務之可能運用與挑戰」，《國立中正大學法學集刊》，62: 43-108。(Long, C.-Y., & Chung, H.-Y. [2019]. Possible application and challenge of artificial intelligence in judicial practice. *National Chung Cheng University Law Journal*, 62: 43-108.)
- 薛智仁 (2023)。「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》，27: 104-133。(Hsueh, C.-J. [2023]. An initial exploration of the challenges of artificial intelligence in criminal law—A case study on autonomous driving. *Formosan Jurist*, 27: 104-133.)
- 蘇凱平 (2021a)。「國民參與刑事審判程序的證據能力判斷：兼論鑑定證據之證據能力判斷標準」，《臺大法學論叢》，50, 4: 1923-1989。(Su, K.-P. [2021a]. Admissibility of evidence in the lay participation system of criminal courts: With an attention to scientific evidence. *National Taiwan University Law Journal*, 50, 4: 1923-1989.) [https://doi.org/10.6199/NTULJ.202112_50\(4\).0005](https://doi.org/10.6199/NTULJ.202112_50(4).0005)
- 蘇凱平 (2021b)。「當證據「上鏈」：論區塊鏈科技應用於法庭證據」，《臺大法學論叢》，50, 3: 993-1071。(Su, K.-P. [2021b]. Blockchain technology and courtroom evidence. *National Taiwan University Law Journal*, 50, 3: 993-1071.) [https://doi.org/10.6199/NTULJ.202109_50\(3\).0005](https://doi.org/10.6199/NTULJ.202109_50(3).0005)
- Abu-Elyounes, D. (2020a). Bail or jail? Judicial versus algorithmic decision-making in the pretrial system. *Columbia Science and Technology Law Review*, 21, 2: 376-446. <https://doi.org/10.7916/dzbj-ff79>
- Abu-Elyounes, D. (2020b). Contextual fairness: A legal and policy analysis of algorithmic fairness. *University of Illinois Journal of Law, Technology & Policy*, 2020, 1: 1-54.
- Anderson, B. J. (2012). Recognizing character: A new perspective on character evidence. *The Yale Law Journal*, 121, 7: 1912-1968.

- Berk, R., & Hyatt, J. (2015). Machine learning forecasts of risk to inform sentencing decisions. *Federal Sentencing Reporter*, 27, 4: 222-228. <https://doi.org/10.1525/fsr.2015.27.4.222>
- Casey, E. (2011). *Digital evidence and computer crime: Forensic science, computers and the internet*. Waltham.
- Chessman, C. (2017). A “source” of error: Computer code, criminal defendants, and the Constitution. *California Law Review*, 105, 1: 179-228. <https://doi.org/10.2139/ssrn.2707101>
- Cummings, M. L. (2004, September). *Automation bias in intelligent time critical decision support systems* [Paper presentation]. AIAA 1st Intelligent Systems Technical Conference, Chicago, US. <https://doi.org/10.2514/6.2004-6313>
- Dastin, J. (2018, October 11). Insight—Amazon scraps secret AI recruiting tool that showed bias against women. *Reuters*. <https://www.reuters.com/article/us-amazon-com-jobs-automation-insight-idUSKCN1MK08G>
- Domingos, P. (2012). A few useful things to know about machine learning. *Communications of the ACM*, 55, 10: 78-87. <https://doi.org/10.1145/2347736.2347755>
- Dong, Y., Hu, Z., Uchimura, K., & Murayama, N. (2011). Driver inattention monitoring system for intelligent vehicles: A review. *IEEE Transactions on Intelligent Transportation Systems*, 12, 2: 596-614. <https://doi.org/10.1109/TITS.2010.2092770>
- Eaglin, J. M. (2017). Constructing recidivism risk. *Emory Law Journal*, 67, 1: 59-122. <https://doi.org/10.2139/ssrn.2821136>
- Ferguson, A. G. (2023). Digital habit evidence. *Duke Law Journal*, 72, 4: 723-796. <https://doi.org/10.2139/ssrn.4051801>
- Gless, S. (2020). AI in the courtroom: A comparative analysis of machine evidence in criminal trials. *Georgetown Journal of International Law*, 51, 2: 195-253.
- 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.
- Heale, R., & Twycross, A. (2015). Validity and reliability in

- quantitative studies. *Evidence-Based Nursing*, 18, 3: 66-67. <https://doi.org/10.1136/eb-2015-102129>
- Imwinkelried, E. J. (2016). Computer source code: A source of the growing controversy over the reliability of automated forensic techniques. *DePaul Law Review*, 66, 1: 97-132.
- Kaestle, F. A., Kittles, R. A., Roth, A. L., & Ungvarsky, E. J. (2006). Database limitations on the evidentiary value of forensic mitochondrial DNA evidence. *The American Criminal Law Review*, 43, 1: 53-88.
- Kerr, O. S. (2001). Computer records and the Federal Rules of Evidence. *USA Bulletin*, 49, 2: 1-9.
- Lehr, D., & Ohm, P. (2017). Playing with the data: What legal scholars should learn about machine learning. *U.C. Davis Law Review*, 51, 2: 653-717.
- Lilly, G. C., Carpa, D. J., & Saltzburg, S. A. (2015). *Principles of evidence*. West Academic.
- Mayer-Schönberger, V., & Cukier, K. (2013). *Big data: A revolution that will transform how we live, work, and think*. Boston.
- McCarthy, J., Minsky, M. L., Rochester, N., & Shannon, C. E. (2006). A proposal for the Dartmouth Summer Research Project on artificial intelligence, August 31, 1955. *AI Magazine*, 27, 4: 12-14. <https://doi.org/10.1609/aimag.v27i4.1904>
- Metallo, V. N. A. (2020). The impact of artificial intelligence on forensic accounting and testimony—Congress should amend “The Daubert Rule” to include a new standard. *Emory Law Journal Online*, 69: 2039-2063.
- National Institute of Standards and Technology. (n.d.). *Forensic science standards board*. <https://www.nist.gov/organization-scientific-area-committees-forensic-science/forensic-science-standards-board>
- Nishi, A. (2019). Privatizing sentencing: A delegation framework for recidivism risk assessment. *Columbia Law Review*, 119, 6: 1671-1710. <https://doi.org/10.2139/ssrn.3335946>
- Northpointe Inc. (2015, March 13). *Practitioner’s guide to COMPAS core*. <https://archive.epic.org/algorithmic-transparency/crim->

- justice/EPIC-16-06-23-WI-FOIA-201600805-COMPASPractitioner Guide.pdf
- Park, R. C., Orenstein, A. A., & Nance, D. A. (2022). *Evidence Law: A student's guide to the law of evidence as applied in American trials*. West Academic.
- Peters, J. (2013, July 9). When ice cream sales rise, so do homicides. Coincidence, or will your next cone murder you? *Slate*. <https://slate.com/news-and-politics/2013/07/warm-weather-homicide-rates-when-ice-cream-sales-rise-homicides-rise-coincidence.html>
- Roth, A. (2017). Machine testimony. *The Yale Law Journal*, 126, 7: 1972-2053.
- Semmi, W. (2024, February 8). Fact check: Are black people the majority of drug users and distributors? *Revolt*. <https://www.revolt.tv/article/2024-02-08/350905/fact-check-are-black-people-the-majority-of-drug-users>
- Seng, D., & Mason, S. (2021). Artificial intelligence and evidence [Special issue]. *Singapore Academy of Law Journal*, 33: 241-279.
- Stockdale, M., & Mitchell, R. (2019). Legal advice privilege and artificial legal intelligence: Can robots give privileged legal advice? *The International Journal of Evidence & Proof*, 23, 4: 422-439. <https://doi.org/10.1177/1365712719862296>
- The Hamilton Project. (2016, October 20). *Twelve facts about incarceration and prisoner reentry*. Retrieved September 26, 2024, from <https://www.hamiltonproject.org/data/rates-of-drug-use-and-sales-by-race-rates-of-drug-related-criminal-justice-measures-by-race/>
- The United States Department of Justice. (1997). *Searching and seizing computers and obtaining electronic evidence in criminal investigations*. <https://www.justice.gov/file/442111/download>
- Trafimow, D. (2017). The probability of simple versus complex causal models in causal analyses. *Behavior Research Methods*, 49, 2: 739-746. <https://doi.org/10.3758/s13428-016-0731-3>

- Walker, P. (2013, August 21). Black people twice as likely to be charged with drugs possession—Report. *The Guardian*. <https://www.theguardian.com/world/2013/aug/21/ethnic-minorities-likely-charged-drug-possession>
- Weisberg, R. (2014). Meanings and measures of recidivism. *Southern California Law Review*, 87, 3: 785-804.

A Study on the Use of Evidence Generated by Artificial Intelligence in Criminal Proceeding: U.S. Federal Rules of Evidence

Hsiao-Jung Tseng

School of Law, National Yang Ming Chiao Tung University
E-mail: hsiaojung.96@gmail.com

Mong-Hwa Chin

School of Law, National Yang Ming Chiao Tung University
E-mail: mhchin@nycu.edu.tw

Abstract

In the wake of rapid developments in artificial intelligence (AI), judgments or predictions made by AI are being used in place of human judgment or as a reference source. Although there have been, as yet, no cases in which judgments or predictive data generated by AI have been used as the basis for criminal charges in the fact-finding stage, it is anticipated that, given current trends and the already extensive application of AI, there may come a time when the judiciary must confront evidence of this new type in criminal proceedings. This article focuses on AI equipped with machine learning processes and capable of independently rendering judgmental results as the primary subject of discussion. It aims to examine applicability issues within the existing legal framework, using the United States Federal Rules of Evidence as the subject for the application of evidentiary rules. This article contends that evidence derived from AI falls under the category of digital evidence. However, the current framework governing evidence law will encounter challenges treating artifacts of AI as real or testimonial evidence, necessitating adjustments addressing this form of evidence.

Key Words: artificial intelligence (AI), machine learning, Federal Rules of Evidence, digital evidence