

以人工智慧輔佐法院之刑罰裁量： 破譯「黑盒子」的可能與顧慮*

蘇凱平**

<摘要>

大眾文化中常見以人工智慧或其他科技手段，取代人類擔任審判者的描寫，反映出社會對於司法系統的期待。然而，今日人工智慧的發展進程，是否已經使這種期待有實現的可能？本文從最有可能引入人工智慧和大數據分析等新興科技，輔助刑事法院判斷的證據類型：「統計證據」之觀點切入，說明當代的行為科學研究已經充分證明「人類智慧」在判斷上難以避免的種種缺陷，因此有求助於「人工智慧」的必要性。繼而說明人工智慧的發展歷程，探討「以人工智慧取代或輔助人類審判者」的可能性。

本文分別以使用於美國多個法域的刑事司法系統輔助軟體 COMPAS，及我國司法院建置的第一代與第二代量刑輔助系統為例，說明以人工智慧和大數據分析技術輔助法院進行量刑決策時，其原理、具體方式、優缺點以及

* 作者感謝編輯委員與2位匿名審查委員提出寶貴修正意見，讓本文能夠進一步完善論點。本文之完成，有賴以下國科會專題研究計畫的補助：「法庭上的數字難題：統計證據的法理與應用研究」（113-2410-H-002-157-MY2）、「法律、訊息與人類行為：跨學科的探問（1/3）」（113-2423-H-001-008-）、「數位智能法院、法律科技與接近正義—數位智能法院、法律科技與接近正義(3/3)」（112-2423-H-002-001-）、「以行為科學概念分析刑事法院決策機制：以定罪與量刑為中心」（111-2410-H-002-029-MY2），作者感謝國科會的補助。

** 臺灣大學法律學系副教授，美國柏克萊加州大學法學博士。

E-mail: sukaiping@ntu.edu.tw。

• 投稿日：08/30/2023；接受刊登日：03/27/2024。

• 責任校對：龔與正、蔡雅棋、陳怡君。

• DOI:10.6199/NTULJ.202509_54(3).0004

可能的憲法爭議。除了分別說明 2 種不同的科技手段如何輔助刑事法院進行量刑決策外，本文也比較了人類社會對於這 2 種科技手段產出判斷結果的接受程度，以及「人類智慧」與「人工智慧」間的互動與信任關係。

本文主張，至少以今日的技術發展而論，仍不可能以人工智慧「取代」人類法院對於量刑的判斷。一方面，司法當局已經意識到：若強制要求審判者以科技的判斷結果決定量刑，固然更可能達成一致的量刑，卻會有侵害審判權核心的疑慮；另一方面，社會大眾也不易逕行接受人工智慧判斷的結果，因為人工智慧的技術發展至今，仍難提供人類願意接受司法裁判的必備要素：充分的可解釋性與透明性。

關鍵詞：量刑、統計證據、行為科學、捷思法、機器學習、COMPAS

◆目次◆

- 壹、科技輔助司法決策的熱區：統計證據
 - 一、統計證據的概念
 - 二、行為科學的概念：人類智慧的限制
- 貳、人工智慧的發展進程
 - 一、人工智慧的概念轉變
 - 二、人工智慧的三波發展
 - 三、人工智慧「取代」人類法院？
- 參、以人工智慧輔助量刑之一：美國 COMPAS 系統
 - 一、技術原理與應用情況
 - 二、Loomis 案判決理由分析
- 肆、以大數據輔助量刑：司法院第一代量刑輔助系統
 - 一、技術原理與應用顧慮
 - 二、法院的適用擴張現象
 - 三、適用擴張的違憲疑慮？
- 伍、以人工智慧輔助量刑之二：司法院第二代量刑輔助系統
 - 一、技術原理與社會顧慮
 - 二、機器學習方法的共通本質問題
 - 三、我國法制脈絡下必須面對的問題
- 陸、結論：揭開不同的「黑盒子」

壹、科技輔助司法決策的熱區：統計證據

在流行文化中，經常可見 2 種與法律系統有關的描寫：第一種是絕對公正且可迅速實現「正義」的審判者（通常還兼任執行者）形象，例如「包青天」；第二種是永遠保持客觀、不受情緒干擾或任何影響、具有強大能力、只知執行命令的智慧機器形象，例如「魔鬼終結者」（The Terminator）。將以上 2 種形象揉合在一起，就得到了一個能夠客觀、中立、公平、正確判斷

各種紛爭中誰是誰非的「智慧機器」形象。例如在 2002 年改編自小說的好萊塢電影「關鍵報告」(Minority Report)中，就描述了一個將犯罪幾乎根絕的烏托邦。因為這個世界建立了一套可以早在犯罪事件發生「之前」，即鎖定犯罪者的預防系統。儘管在電影中，可以產出犯罪預見知識 (foreknowledge) 的來源，是 3 個大腦頻率波動狀態異常活躍的先知靈媒 (precogs)，並非人工智慧或機器人等科技手段，但是究其本質顯然無異於一種智慧機器¹。

流行文化中以人工智慧或其他科技手段，取代人類判斷的描繪，反映出社會對於現行由人類執掌的司法系統、特別是負責犯罪判斷與執行的刑事司法系統的不滿意或不信任。職司審判的人類，不論是職業法官，或是非法律專業而擔任審判者的陪審員、國民法官，都有可能做出主觀、有偏見、不公平的決定，難以滿足社會大眾對於司法系統應能客觀、公正、準確、無偏誤地實現正義的期待。然而，儘管當代人工智慧技術發展迅速，但這種「以人工智慧取代刑事司法系統判斷」的想像，仍然停留在想像或創作的階段，在近期內尚無法實現²。

本文希望探討的是，人工智慧或其他的科技手段（例如大數據分析）縱然無法「取代」人類法院的判斷，但這些科技手段是否可能「輔助」司法系統的判斷過程，使其更客觀、更公正、更準確、更不發生偏誤呢？為了聚焦討論，本文以下將這些問題置於刑事審判的框架下進行探討。因為刑事審判者負責判斷的對象是犯罪與犯罪者，正是最容易引起社會大眾矚目的案件類型；也正因為刑事案件易受矚目的性質，刑事法院也往往是最積極引入科技以輔助法院進行判斷的場域³。而刑事法院在審判時要進行的各種判斷中，

¹ See Adam Edwards, *Big Data, Predictive Machines and Security: The Minority Report*, in THE ROUTLEDGE HANDBOOK OF TECHNOLOGY, CRIME AND JUSTICE 451 (M. R. McGuire & Thomas. J. Holt eds., 2017).

² 當代人工智慧技術的發展進程，以及「以人工智慧取代人類審判者」的技術尚未能實現的理由，請詳見本文「貳、人工智慧的發展進程」章說明。

³ 請詳見本文「參、以人工智慧輔助量刑之一：美國COMPAS系統」以及「肆、以大數據輔助量刑：司法院第一代量刑輔助系統」等章的說明。

又以「統計證據」(statistical evidence)此種證據類型，最常成為引入科技輔助法院判斷的「熱區」。

一、統計證據的概念

必須先說明的是，在證據法上，「統計證據」一詞的意義，不僅僅是如同其字面所表達的「使用統計數字作為證據」的情況，而是還包括了指涉各種「使用『本案以外』之資訊，作為判斷『本案中』爭點的證據」而言⁴。例如當法院在審理被告吸食毒品後駕車的公共危險案件時，如果在判斷被告是否因吸毒致不能安全駕駛時，不僅考慮本案被告吸食毒品後是否肇生事故等本案資訊，還考慮了「通常一般人體內驗出此種毒品含量時，是否仍可能安全駕車」為本案判決的證據之一，此時法院即係使用「本案以外」之資訊，作為判斷本案爭點的證據，屬於使用統計證據的情況。又如法院在量刑時，若不僅考慮到本案被告的犯行、犯罪人特質和其犯罪造成之影響等本案情況，而是還考慮到在過去的其他類似案件中，法院通常的判刑區間如何，縱然並不是以平均數、中位數、眾數等統計數字的方式來考慮過去的判刑情況，也屬於使用統計證據的情況。一言以蔽之，統計證據處理的問題，即是如何正確地綜合考量系爭案件內、外的多重因素，以判斷出系爭事項發生之「可能性」(probability，亦稱「機率」或「概率」)高低。其中雖然經常涉及統計數字，但並不以此為限⁵。

統計證據是證據法中的重要主題，其重要性在於兩方面：

1. 審判者在個案的審判中，是否能夠正確的理解和運用統計證據？

⁴ 因此，與statistical evidence相對的證據型態，文獻上稱為individualized evidence(個別化證據)，意即僅以出自本案的資訊為證據的情況。See, e.g., Richard Schmalbeck, *The Trouble with Statistical Evidence*, 49 LAW & CONTEMP. PROBS. 221, 221 (1986); Amit Pundik, *The Epistemology of Statistical Evidence*, 15 LAW & PHIL. 117, 126 (2011).

⁵ Amit Pundik, *Rethinking the Use of Statistical Evidence to Prove Causation in Criminal Cases: A Tale of (Im)Probability and Free Will*, 40 LAW & PHIL. 97, 98-99 (2021); Georgi Gardiner, *Legal Burdens of Proof and Statistical Evidence*, in THE ROUTLEDGE HANDBOOK OF APPLIED EPISTEMOLOGY 179, 180-85 (David Coady & James Chase eds., 2019).

2. 統計證據使用的資訊，並非來自法院審理的個案本身，那麼法院以統計證據作為判決基礎的法理上正當性何在？

（一）正確理解和運用統計證據的困難

首先，統計證據雖然不限於以統計數字為證據的情況，但確實經常涉及統計數字的詮釋和評價。無論是職業法官或是因參與審判而擔任審判者的國民（例如美國的陪審員，或我國的國民法官），都不一定具備有理解統計數字的基本能力，因此審判者在運用其自由心證，評估指涉機率或可能性的證據價值時，其理解是否正確？實在令人有所疑問。誠然，當審判者發現案件證據涉及複雜的數字概念時，法院通常會請專業的鑑定人進行說明（例如在吸食毒品導致不能安全駕駛的情況，「一般人」究竟吸食多少量的某種毒品，始可認為已達不能安全駕駛之程度？）⁶；但是在許多案件中，當事人、證人或鑑定人的陳述雖然涉及統計證據，該陳述也成為當事人間爭執的焦點，但法院卻不會求助於專業協助來判斷這些統計證據之意義、甚至根本沒有意識

⁶ 例如最高法院107年度台上字第205號刑事判決記載的情況：「況所謂不能安全駕駛，非係以瀕死亡、休克為判斷標準，法醫研究所經參考各國統計分析結果，認定血中甲基安非他命在500ng/ml以上，即構成不能安全駕駛，係依一般客觀情形判斷，常人若施用毒品達上開尿液濃度，已達不能安全駕駛之狀態，上訴人既已符合上開要件，應認已有不能安全駕駛之情」；最高法院110年度台上字第4670號刑事判決亦稱：「參諸法務部法醫研究所105年12月2日法醫理字第10500056540號函略謂：『臺灣經修法後再參考歐美先進國家，經過大數據中，綜合各類藥物、各個國家各種統計分析結果，已導出一般性致不能安全駕駛中包括甲基安非他命類使用後，其相對影響精神致不能安全駕駛能力，一般血中甲基安非他命以500ng/ml以上為認定標準』，而一般認定甲基安非他命於血、尿之比率，在分佈均勻狀況約為1：14至1：20，有該所104醫文字第1041102335號函可稽，本件案發後約隔四小時所採集上訴人尿液之上開鑑定結果，甲基安非他命之濃度為5983ng/ml，依對其最有利之比例1：20換算其血液中甲基安非他命濃度為4047.05ng/ml，確逾上開500ng/ml之標準甚多」。在此類案件中，判決所依據的「各國統計分析結果」或「歐美先進國家，經過大數據中，綜合各類藥物、各個國家各種統計分析結果」究竟能否適用於個案，均成為當事人爭執的重點。

到這是一個需要專業協助的問題(例如在判斷死亡結果是否為「嬰兒猝死症」所導致時)⁷。

再者，由於人類在進化過程中發展出的「捷思法」(heuristics)和各種「偏誤」(biases)，使得審判者在面對統計證據時，特別容易發生理解上的錯誤。大量的行為科學研究指出了這樣的現象：即使是意義完全相同的統計證據，只要呈現的方式不同，就足以讓審判者的判斷結果截然不同。經典的例子是行為經濟學家、也是諾貝爾經濟學獎得主的 Daniel Kahneman 在其名著《快思慢想》(Thinking, Fast and Slow)中所論述者⁸：某城中只有 2 家

⁷ 例如最高法院104年度台上字第1222號刑事判決稱：「檢察官上訴意旨略稱：(一)、原判決就被害人陳○是否因由側睡變成趴睡，以致窒息而死一點，未將吳育維之偵訊證詞……作為判決之基礎……且原判決僅引述吳育維之部分供述，而疏漏吳育維於偵訊中證稱：『他(指陳○)的代謝性酸中毒、高血鉀、跟凝血功能異常應該是窒息造成的。』、『**趴睡是嬰兒猝死的高危險群，比起仰睡，趴睡容易窒息而造成嬰兒猝死。**』等足以佐證陳○之死因是窒息，非嬰兒猝死，或至少是因窒息而引起嬰兒猝死之證言。」、「原判決於理由內已依憑台灣兒科醫學會曾於他案函覆原審法院載稱：目前通認的嬰兒猝死症候群的定義是：『一歲以下嬰兒突然死亡，且經過完整病理解剖、解析死亡過程並檢視臨床病史等詳細調查後仍未能找到死因者。』，在發展中國家，據統計報告，每一千名活產兒中約○.八名可能發生，台灣估計為一千名活產兒中約有一.五名，好發於二個月至四個月大嬰兒，亦為該年齡層兒童的最常見死因之一。……另證人即鑑定人陳明宏於原審法院第二次更審審理時證稱……本件雖不能完全排除睡眠中自然停止呼吸而死亡，但此種可能性比較低，製作法務部法醫研究所鑑定書時，沒有考慮這樣低的機會。依伊過去累積解剖案例經驗與本件解剖結果發現來說，沒有發現睡眠中自然呼吸停止的表徵，外力造成窒息死亡的機會要比睡眠呼吸自然停止可能性要高很多，相關卷證呈現被害人死亡時，周圍環境有外力發生呼吸道阻塞窒息風險之可能，**因為上開二個原因，而認定外力造成窒息死亡，而不採用可能較低之睡眠自然停止為死亡原因等語，亦無法絕對排除其他死因(如睡眠自動停止呼吸)的可能性，僅係以可能性較高之外力或外物阻塞呼吸而窒息作為判定被害人死亡原因。**」上開判決中鑑定人或證人提及的嬰兒猝死症好發率、某種睡姿是否為造成嬰兒猝死的「高危險群」、某種死因的「可能性比較低(或比較高)」等語，同樣涉及本案之外的資訊，亦即統計證據，但事實審法院並沒有另請鑑定人(統計學家或醫療統計專家)釐清這些資訊的意涵，即引用為證據；即使當事人有所爭執而上訴至最高法院，最高法院也向來認為此屬於事實審法院認事用法之職權，法院逕行判斷並無違法可言。

⁸ 以下案例取自康納曼(Daniel Kahneman)(著)，洪蘭(譯)(2022)，《快思

計程車行，分別是綠色車和藍色車。現在有一輛計程車在晚上肇事逃逸，法院獲知的全部證據如下（「情境一」）：

證據 A. 城中 85%的計程車是綠色的，15%是藍色的；

證據 B. 目擊肇事的證人說，他看到肇事車輛是藍色的，但經過檢驗，發現證人可以正確指出顏色的機率是 80%，錯誤的機率是 20%。

請問，肇事車輛確實是藍色的機率有多少？

Kahneman 透過實驗發現，大部分的人在面對上述資訊時，會忽略證據 A 的統計證據而採信目擊證人的話，因此普遍的答案是 80%。但是如果將證據的內容略做改變，變成以下的情境（「情境二」）：

證據 A. 2 家計程車行有相同數量的計程車，但是綠色車的肇事率是 85%，藍色車是 15%；

證據 B. 目擊證人的證詞維持不變。

請問，肇事車輛確實是藍色的機率有多少？

此時，大多數的受試者會注意到證據 A，因此會納入統計證據來進行計算，最後得出 41%的正確答案⁹。

事實上，前後 2 個版本的證據內容，在數學上是完全相同的，但是在受試者（審判者）的心理上卻截然不同¹⁰。「情境一」的證據 A，是關於城裡

慢想》，頁225-227，遠見天下文化。

⁹ 具體的計算方式是依據「貝氏定理」(Bayes' theorem)計算「條件機率」(conditional probability)，條件機率的定義是 $P(A|B) = P(A \cap B) / P(B)$ 。亦即在已知事件B發生的情況下，事件A發生的機率。

貝式定理與條件機率的基礎概念，可以參考：陳昱成(2013)，〈貝氏定理的應用〉，《科學教育月刊》，357期，頁19-28。本案例詳細的計算則可以參考 Greg Ross, *The Taxicab Problem*, FUTILITY CLOSET (June 8, 2019), <https://www.futilitycloset.com/2019/06/08/the-taxicab-problem/>.

¹⁰ 康納曼 (Daniel Kahneman)，前揭註8，頁226。

有多少綠色車和藍色車的統計事實，無法觸動受試者的心理，大多數受試者無意識地、直覺地傾向認為此一資訊與本案無關，因此僅依據證據 B 做出了判斷（肇事車輛是藍色的機率為 80%）；「情境二」的證據 A，則在受試者心中描繪出一個「綠色車肇事率可怕的高」的印象，大多數受試者會將其納入考量，與證據 B 合併思考後得出判斷（肇事車輛是藍色的機率為 41%）。由此可見，涉及統計事實的統計證據，由於人類直覺的作用，容易在審判者的視野中「視而不見」，很難正確評估。白話的說，人類的直覺很容易被統計證據所欺騙，審判者也不例外。

（二）使用統計證據為判決基礎的正當性

統計證據使用的資訊，既然並非來自法院審理的個案本身，那麼法院以統計證據作為判決基礎的法理上正當性何在？這是國外文獻討論統計證據議題的焦點所在¹¹。可惜的是，我國法律學文獻對於統計證據少有討論，因此統計證據在我國法院（尤其是刑事法院）能否使用？正當性基礎何在？有無限制？是否於判斷犯罪成立或科處刑罰時皆可使用統計證據？使用統計證據是否侵害憲法對被告權利的保障？是否符合正當法律程序？等重要議題，在學理上均尚未能釐清¹²。儘管如此，我國法院判決中卻經常可見使用統計證據做為論罪和科刑基礎的情況，可見我國法院在實務運作上確實有使用統計證據的需要。

¹¹ See, e.g., Stephen E. Fienberg & Miron L. Straf, *Statistical Evidence in the US Courts: An Appraisal*, 154 J. ROYAL STAT. SOC'Y A: STAT. & SOC'Y 49, 53-54 (1991); Reva B. Siegel, *Blind Justice: Why the Court Refused to Accept Statistical Evidence of Discriminatory Purpose in McCleskey v. Kemp—And Some Pathways for Change*, 112 NW. U. L. REV. 1269, 1280-82 (2018); Lewis Ross, *Rehabilitating Statistical Evidence*, 102 PHIL. & PHENOMENOLOGICAL RES. 3, 3-5 (2021).

¹² 由於本文乃以人工智慧和其他數據科技對於法院決策的輔助為討論主旨，此處只是說明科技輔助法院決策最有可能涉及的證據類型，乃為統計證據。關於統計證據的正當性、使用界限、常見謬誤等概念，基於本文主旨和篇幅考量，作者將另行撰文說明。

將統計證據運用在判斷被告是否成立犯罪的典型，例如在涉及以 DNA 比對嫌犯身分的案件中，我國法院判決指出：「實務上常針對 DNA-STR 型別實施檢測，將檢體之細胞核 DNA 進行分析，取得 15 組 STR 數值與性別染色體，再以統計推論所得之特定人口中 DNA 型別重複出現頻率為基礎，計算 15 組 STR 均相同之機率，如該機率數值甚微，代表該特定人口中幾無可能出現另一相同 DNA-STR 型別之人……足徵被告確曾至遭竊現場飲用上開寶特瓶所盛裝之飲料，是被告前開辯解，並不足採。¹³」而在判斷被告是否因為吸食毒品而致不能安全駕駛之程度時，則有判決指出：「況所謂不能安全駕駛，非係以瀕死亡、休克為判斷標準，**法醫研究所經參考各國統計分析結果，認定血中甲基安非他命在 500ng/ml 以上，即構成不能安全駕駛，係依一般客觀情形判斷，常人若施用毒品達上開尿液濃度，已達不能安全駕駛之狀態**，上訴人既已符合上開要件，應認已有不能安全駕駛之情。¹⁴」

又如在醫療過失案件中，最高法院要求事實審法院應調查各項醫療統計資料，以判斷被告醫師的醫療行為是否有疏失而造成病患死亡：「又脂肪栓塞除了造成肺部血管之栓塞外，全身其他器官血管是否都會有可能發生栓塞？腦部發生機率是否最高？……**統計上，在沒有被考慮『脂肪性肺栓塞』診斷的病人中其死亡率多少？……敗血症、脂肪栓塞、血管內凝血溶血症（DIC）、肺部感染後導致之敗血症，是否均為引起急性呼吸窘迫症候群的原因？統計上，其死亡率若干？¹⁵**」而在證券內線交易案件中，在判斷證券交易法上規定的「有重大影響其股票價格之消息」（即所謂「內線消息」）要件時，最高法院則要求事實審法院在判斷犯罪構成要件時「**在重大性認定上，均應將其發生之『或然性』列入考慮**，亦即應就內部人所獲悉之資訊足以推斷該事項發生機率之『可能性』，以及此事項若確實發生，其對於投資人投資決定等可能產生之『影響程度』做綜合判斷¹⁶。」

¹³ 臺灣高等法院102年度上易字第2280號刑事判決。

¹⁴ 最高法院107年度台上字第205號刑事判決。並請參見前揭註6。

¹⁵ 最高法院102年度台上字第809號刑事判決。

¹⁶ 最高法院106年度台上字第86號刑事判決。本文引用判決文字內的（雙）引號為判決原文中即使用。

除了上述以統計證據作為「定罪」（判決犯罪是否成立）證據的情況，也有許多判決是以統計證據作為「科刑」（判決有罪時，決定刑罰之輕重）的基礎。例如有事實審法院在量處刑罰時認為：「參酌司法院量刑資訊及量刑趨勢建議，原審就被告所犯傷害致死罪量處有期徒刑 10 年 6 月，**核與同類型案件平均及建議刑度顯有過重之情，其罪刑並不相當，自非允當。**¹⁷」既然第二審法院科刑時不僅考慮到本案情節的輕重，還考慮到了「同類型案件」（亦即他案）的「平均刑度」如何，以作為本案第一審法院之量刑是否允當的判斷標準，亦屬於使用統計證據的情況。

針對量刑時使用統計證據的情況，一方面，實務上認為法院進行「科刑資料之調查」時，對「非屬犯罪事實而單純作為科刑應審酌情狀之事實者」並不適用嚴格證明法則，而是屬於自由證明之事項，因此證據方法和調查程序並無特別設限，可以容許使用各種各樣的證據資料來認定科刑事實¹⁸。如此說來，似乎在審酌單純科刑事實時，有可能接受使用統計證據；但另一方面，刑事量刑由於強調「罪責原則」和「罪刑相當原則」，對於犯罪人所科之刑罰，必須維持刑罰與犯罪人罪責之間的相當性，刑罰不應超過犯罪人自

¹⁷ 臺灣高等法院108年度上訴字第1460號刑事判決。

¹⁸ 例如最高法院111年度台上字第5138號刑事判決：「又其所謂『科刑資料』，參諸刑事訴訟法第310條第3款『有罪判決書應於理由內記載『科刑時就刑法第57條或第58條規定事項所審酌之情形』之規定，應係指與刑法第57條或第58條所列科刑輕重標準有關之具體事實而言。其中除與犯罪事實有關者（如犯罪之手段、違反義務之程度、所生危險或損害等），因屬嚴格證明之事項，故須於論罪階段依各項證據之性質踐行法定調查程序，此外，其餘非屬犯罪事實而單純作為科刑應審酌情狀之事實者（如犯罪行為人之生活狀況、犯罪後之態度等），均為自由證明事項，此部分事實用以證明之證據，其證據能力或證據調查程序法無限制之明文，目前刑事訴訟法，僅規定科刑資料之調查時期應於罪責資料調查後為之（刑事訴訟法第288條第4項參照），及增列規定審判長於命檢察官、被告、辯護人依序就事實及法律分別辯論後，應命依同一次序，就科刑範圍辯論之，並應予到場之告訴人、被害人或其家屬或其他依法得陳述意見之人就科刑範圍表示意見之機會（刑事訴訟法第289條第2項），然對於科刑資料應如何進行調查仍未明文規定。」最高法院111年度台上字第5140號刑事判決、最高法院109年度台上字第3938號刑事判決等均同此意旨。另，關於自由證明與嚴格證明的異同，請參見林鈺雄（2023），《刑事訴訟法（上冊）》，12版，頁511-512，新學林。

身犯行之罪責。如此一來，法院在考慮「個案中」犯罪人的行為處罰時，如果使用了統計證據作為判決基礎，亦即考慮到了「個案以外」的因素，其正當性即有疑問¹⁹。

二、行為科學的概念：人類智慧的限制

如前所述，法院面對統計證據議題時，經常涉及對於案件內、外多項不同資訊（變項）的綜合考量，但人類在應對這種複雜情況時，卻很難對於各種變項逐一做出精準的估算，因為這本來就不是「人類智慧」認知事物的方式。

人類腦容量有限的先天限制，使得人類在演化過程中形成了「修剪」（prune）所需認知、記憶事物的數量，以便有效處理面前的狀況。更具體而言，根據迄今行為科學（Behavioral Sciences）領域的研究成果發現，人類在進行判斷時，經常要面對所需資訊不足、狀況不確定與決策時間有限等情況，因而在進化的過程中發展出各種「捷思法」，亦即在人類思維中具有普遍性的直覺反應，幫助人類簡化思考的模式或內容，以便得出一套據以行動的策略²⁰。人類思維在上述「修剪」和「捷思法」運作的影響下，不免會產生各種認知和判斷上的瑕疵或偏誤。而這樣的「弱點」一旦展現在法院審判的過程之中，就不免令人擔心法院是否可能誤判、甚至造成冤獄²¹。

¹⁹ 關於事實審法院使用司法院製作的量刑系統（包括量刑資訊系統和量刑趨勢建議系統）提供的「他案」資訊，作為「本案」量刑判斷依據的實務運作方式，其正當性問題的探討，請詳本文「肆、以大數據輔助量刑：司法院第一代量刑輔助系統」章。

²⁰ Heuristics的原意是指「發現事物或發現知識的方法」，中文一般翻譯為「捷思法」，現在多用以描述人在進行決策和選擇時，如何在客觀資訊和主觀認知能力有限的情況下，簡化其思考的模式或內容，以得出行動的策略。換言之，採取捷思法的思考模式時，決策者會以忽略部分的資訊為代價，換取使決策過程更加快速、準確、節省腦力消耗等好處。See Amos Tversky & Daniel Kahneman, *Judgment under Uncertainty: Heuristics and Biases*, 185 SCIENCE 1124 (1974); DANIEL KAHNEMAN, *THINKING, FAST AND SLOW* 45-58 (2012).

²¹ 行為科學領域（以認知心理學與行為經濟學為主軸）就人類智慧運作方式如何影響法院——尤其是職司定罪科刑的刑事法院——判斷個案結果的研究成果極多，絕大

換言之，「人類智慧」的運行方式，是減少需要認知、記憶和判斷的事物數量，在有限的考量中，做出當下看來最妥適的選擇。反之，「人工智慧」由於電腦擁有幾乎無窮盡的數學能力和運算速度，在面對複雜多重的選項時，可以選擇計算出各種可能性及其組合，再進行最適合的選擇²²。

因此，基於當人類智慧用於處理複雜多重的情況時，由每個個體選擇認知和記憶的選項不同、詮釋方式不同、甚至產生偏誤和運用捷思法的方式也可能不同，即使面對相同或很類似的事實情況，個體的判斷結果很容易大相逕庭。這樣的情況雖然是人類社會的常態，但出現在法院時，其結果就是在事實情況很類似的案件中，即使是同一（組）審判者的判斷結果也經常不一致，更遑論不同審判者之間的判斷了。法院常因此招致批評，認為對於類似案件的判斷結果經常不同，有違反平等原則、欠缺正當依據、標準流於抽象等缺點。典型的例子是刑事法院的量刑決定，因為依據刑法規定，法院量刑

部分為英文文獻，例如：William C. Thompson & Edward L. Schumann, *Interpretation of Statistical Evidence in Criminal Trials: The Prosecutor's Fallacy and The Defense Attorney's Fallacy*, 11 LAW & HUM. BEHAV. 167, 167-87 (1987); NEIL VIDMAR & VALERIE P. HANS, *AMERICAN JURIES: THE VERDICT* (2007); Neil Vidmar, *The Psychology of Trial Judging*, 20 CURRENT DIRECTIONS IN PSYCHOL. SCI. 58, 58-62 (2011); Pamela A. Wilkins, *Confronting the Invisible Witness: The Use of Narrative to Neutralize Capital Jurors' Implicit Racial Biases*, 115 W. VA. L. REV. 305, 306-12 (2012). 本土脈絡的行為科學與刑事法院決策初步討論，可參考：蘇凱平（2023），〈第二審量刑與定錨效應〉，《月旦法學教室》，249期，頁26-29。

²² 埃米爾·侯賽因（Amir Husain）（著），溫力秦（譯）（2018），《AI創世紀：即將來臨的超級人工智慧時代》，頁54-61，日月文化。必須說明的是，並不是全部的人工智慧都使用被稱為「蠻力」（brute force）、窮舉一切可能性的方式運作；但是從認知方式的角度觀察，人工智慧與人類智慧在面對同一問題時，人工智慧以其強大的數學能力和運算速度，能夠考慮的因素和面向組合顯然遠多於人類。

時必須「審酌一切情狀」²³，此種必須考量多重因素進行的綜合判斷，就常使法院被批評為「黑盒子」²⁴。

上述人類思考判斷方式的難處，造成了刑事法院在處理統計證據議題時，有更強的動機和需要引入科技手段以輔助法院判斷，以實現同時考慮多個變項及準確的計算評估。然而，所謂「使用科技手段輔助法院進行判斷」，這裡所說的「輔助」實際上究竟如何實現？在技術上或法理上，有沒有使用上的限制？法院在得到「輔助」之後判斷的結果，是否即能符合社會期待？在法院的脈絡下，人類又是否能全然信任人工智慧產出的判斷結果？這些正是本文希望探討的問題。

本文以下將先介紹人工智慧的發展進程，以第一波、第二波、第三波人工智慧的內涵與特性，說明當代人工智慧發展的進境，以及為什麼人工智慧尚不能有效取代人類審判者進行法律判斷（第貳章）。為了具體說明人工智慧如何「輔助」法院進行判斷，本文將以美國多個法域持續使用的刑事司法系統輔助軟體——「矯正犯罪者替代性處遇管理剖繪」系統（Correctional Offender Management Profiling for Alternative Sanctions，以下簡稱為

²³ 刑法57條：「科刑時應以行為人之責任為基礎，並審酌一切情狀，尤應注意下列事項，為科刑輕重之標準：

- 一、犯罪之動機、目的。
- 二、犯罪時所受之刺激。
- 三、犯罪之手段。
- 四、犯罪行為人之生活狀況。
- 五、犯罪行為人之品行。
- 六、犯罪行為人之智識程度。
- 七、犯罪行為人與被害人之關係。
- 八、犯罪行為人違反義務之程度。
- 九、犯罪所生之危險或損害。
- 十、犯罪後之態度。」

²⁴ 胡宜如（2016），〈量刑公開透明：司法院量刑系統介紹〉，《法律扶助期刊》，50期，頁43。（「加以過去實務判決對於形成量刑心證之論述往往流於抽象，例如判決理由中僅形式上概括論述刑法第57條之量刑審酌事項，就具體之量刑內涵未善盡說理之義務，法官所裁量的刑罰種類及刑度運用，是否符合刑罰目的要求及罪刑相當原則，往往無從加以審查，致有『量刑黑盒子』的批評」。）

COMPAS 系統)為例,探討人工智慧輔助法院進行判斷時,其優勢、風險和涉及的憲法爭議所在(第參章)²⁵。繼而,本文延續 COMPAS 系統輔助法院量刑的同一脈絡,探討我國使用科技輔助法院決策的發展與現況,包括司法院以「非人工智慧」的大數據分析方法,所建置的第一代量刑輔助系統(包括「量刑資訊系統」與「量刑趨勢建議系統」)(第肆章),以及以機器學習的人工智慧方法建置的第二代量刑輔助系統(包括「事實型」和「評價型」兩種量刑輔助系統)(第伍章)²⁶。本文將詳細說明這兩代系統的建置原理、應用時造成的法理問題與社會顧慮,分析這些問題及顧慮在憲法與法律上的意義,嘗試就可能解決的問題提出解決方案,也就本文認為難以解決的問題,說明其原因所在。最後,本文將綜合全文觀點和討論的案例,探討人類智慧與人工智慧間的互動與信任關係(第陸章)。

²⁵ COMPAS本身即為系統的縮寫詞(acronym),本無須贅稱「COMPAS系統」,惟為配合中文論述和閱讀的習慣,本文仍以「COMPAS系統」稱之。

²⁶ 如同本文的審查委員與編輯委員準確指出者:本文所分析的3種系統,並非基於完全相同的原理所建置。我國司法院的第一代量刑輔助系統並沒有採取人工智慧方法,而司法院的第二代量刑輔助系統雖然與美國COMPAS系統均採取人工智慧方法,但其具體原理仍不相同(請詳本文第參、肆、伍章說明)。本文將這3種系統合併觀察、分析的原因,在於這3種系統要解決的問題是相同的:「達成更妥適的量刑」。或者更精準的說,是「利用本案以外的資訊(統計證據),幫助本案達成更妥適的量刑」。換言之,在以「問題解決」(problem-solving)為導向的科技應用討論中,這3種系統具有相同的出發點,因此得以作為本文綜合探討和比較之標的。而這種「問題解決」導向的發展與討論,也正是今日科技飛速推進的重要特徵。關於當代的數位科技環境下,以「問題解決」為導向發展科技的特色和價值, see Francois Coetzee, *Creativity and Problem-Solving in the Digital Age: Navigating Challenges with Innovation*, MEDIUM (Oct. 4, 2023), <https://nlpwithpurpose.medium.com/creativity-and-problem-solving-in-the-digital-age-navigating-challenges-with-innovation-37385f87272>.

貳、人工智慧的發展進程

一、人工智慧的概念轉變

人工智慧 (artificial intelligence) 的概念，一般認為最早係由美國電腦科學家、認知科學家 John McCarthy 等 4 位學者於 1955 年合著的論文“A Proposal for The Dartmouth Summer Research Project on Artificial Intelligence”中提出²⁷。在這篇關於人工智慧最早期的論文中，人工智慧被定義為一種「自動電腦／計算器」(automatic computers / calculators)，其作用是「模擬準確描述出的(人類)智慧特徵」²⁸。換言之，McCarthy 等學者最初是在「人類的智慧特徵原則上可以被準確描述出來」之假設基礎上，探討如何以電腦模擬出這些智慧特徵。而這種由電腦模擬人類智慧的成果，即稱為「人工智慧」，這是人工智慧最初的概念定義。

歷經數十年的發展後，隨著投入的研究者與資源增加，包括電腦科學、機械工程、哲學、倫理學乃至於人文社會學科等，都對於人工智慧有不同的詮釋、理解和應用方式²⁹。不過，被稱為「人工智慧之父」的 McCarthy，在其 2004 年的論文中，則將人工智慧的概念重新定義為「製造『智慧機器』(intelligent machines)、尤其是『智慧電腦程式』的科學與工程」。McCarthy 並特別強調：人工智慧雖然與運用電腦來瞭解人類的智慧運作方式有相似之處，但是並不僅限於模仿人類或其他生物展現出來的智慧運作形式 (biologically observable intelligence)；反之，McCarthy 認為，凡是可以使用「運算」(computational) 能力以達成目標者，都可以被稱為一種「智慧」

²⁷ John McCarthy et al., *A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence*, 27 AI MAG. 12, 12-14 (2006). 這篇論文促成了史上第一次以人工智慧作為專門研究領域而組織的學術研討：1956年的「達特茅斯暑期人工智慧研究計畫」(Dartmouth Summer Research Project on Artificial Intelligence)。

²⁸ *Id.* at 12.

²⁹ 李建良 (2020)，〈人工智慧與法學變遷：法律人面對科技的反思〉，李建良 (主編)，《法律思維與制度的智慧轉型》，頁4-11，元照。

³⁰。至此，以 McCarthy 為首的初代人工智慧研究者，已經清楚認識到人工智慧不僅限於對人類或其他生物行為的模擬（例如人類或動物的學習行為，如小孩如何學習識字、蜜蜂如何學習採蜜回巢的軌跡等），而是更包括機器或其他非生物形式的智慧模式。一言以蔽之，只要展現出透過運算、達成任務的能力，都是「智慧」；而以創造發明智慧機器（電腦）來模擬、展現此種能力之特徵者，均屬於人工智慧的範疇。

在歷經了 70 餘年的科學技術發展後，今日描述人工智慧發展進程的文獻同樣繁多且紛雜。例如從智慧機器是否具有自我意識的角度，可以區分為「強人工智慧」（strong AI）與「弱人工智慧」（weak AI）；從人工智慧應用範圍的觀點，則有區分為「通用人工智慧」（artificial general intelligence）和「專用人工智慧」（applied intelligence，亦有稱為 narrow AI 者），不一而足。本文主旨在於探討人工智慧輔助法院進行判斷的可能性，而法院在進行判斷時，極重視對於「判斷資料之來源」和「判斷理由何在」的說明。因此，本文以下以強調人工智慧的知識來源與決策理由解釋能力的美國國防部國防高等研究計畫署（Defense Advanced Research Projects Agency，以下簡稱為 DARPA）於 2017 年提出的「人工智慧三波論」（Three Waves of AI）觀點，作為討論分析的框架³¹。

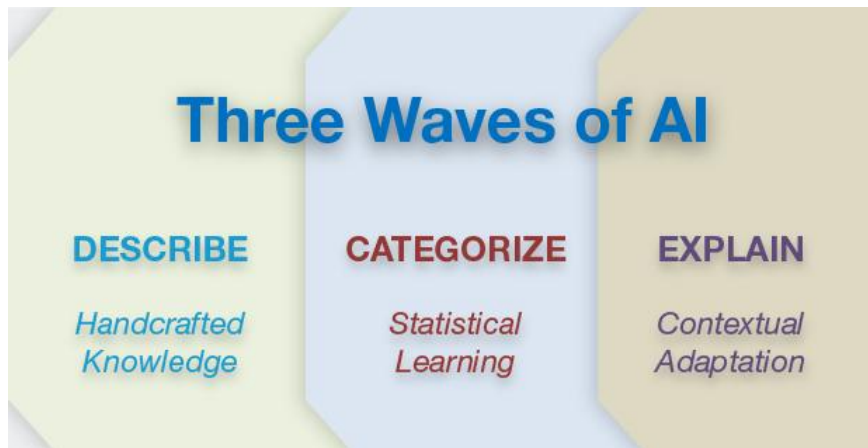
二、人工智慧的三波發展

DARPA 將人工智慧科學技術發展七十餘年來典範轉移的現象分為 3 個階段，分別稱為「手工（打造）知識」（handcrafted knowledge）、「統計學

³⁰ JOHN MCCARTHY, WHAT IS ARTIFICIAL INTELLIGENCE? (Nov. 12, 2007), <https://www-formal.stanford.edu/jmc/whatisai.pdf>.

³¹ The Association for the Advancement of Artificial Intelligence, *DARPA's Three Waves of AI Research — A Special Issue of AI Magazine*, INTERACTIVE AI MAGAZINE (Jun. 29, 2020), <https://interactiveaimag.org/ai-magazine-previews/ai-mag-previews-vol41-no2/>. 介紹DARPA關於人工智慧觀點的中文文獻，請參見：邱文聰（2020），〈第二波人工智慧知識學習與生產對法學的挑戰：資訊、科技與社會研究及法學的對話〉，李建良（主編），《法律思維與制度的智慧轉型》，頁137，元照。

習」(statistical learning)和「脈絡適應」(contextual adaptation)，三階段各自的內涵與特色如以下【圖一】所示：



【圖一】Three Waves of AI (DARPA Perspective on AI)

※ 資料來源：John Everett, *DARPA and the Exploration of Artificial Intelligence*, DEFENSEMEDIANETWORK (Feb. 28, 2024), <https://www.defensemedianetwork.com/stories/darpa-and-the-exploration-of-artificial-intelligence/2/>

根據 DARPA 的定義，第一波的「手工（打造）知識」是指智慧機器的人類開發者，以創造「規則」(rules)的方式，來表達特定領域的知識，並令智慧機器依據這些規則來運行。在此階段，智慧機器本身並沒有如同人類般可以透過解決問題的過程，而習得新知識的這種「學習」能力。因此，其人工智慧程度只適用於處理相對狹窄的領域，或用於處理可以清楚界定規則的問題。一旦出現無法清楚界定、相對模糊的問題，或過於寬廣的領域，此階段的智慧機器就難以運用既定的規則，來解決這些問題中的不確定性。由於第一波的人工智慧以預先設定的「若……，則……」(if..., then...)規則為本，因此又稱為「以規則為基礎的人工智慧」(Rule-Based AI)；又因為

處理的任務限縮在相對狹窄、特定的領域，又被稱為「專家系統」(Expert Systems)³²。

第一波人工智慧以人類開發者制定的清晰規則為基礎，雖然是可以用來解決問題的智慧機器，但可以適用的範圍狹窄，僅能用於事先明確界定清晰規則的領域。然而，人類社會的知識複雜多端，許多領域的知識與問題很難僅用「若……，則……」的條件式來充分表達，由此可以看出第一波人工智慧的侷限性。這種侷限性一方面來自於研究者、開發者對於人工智慧的想像和創新論述需要時間推進，另一方面也由於欠缺硬體的配合。自 1956 年人工智慧的概念提出，一直到 1980 年代的三十餘年間，電腦的硬體技術發展未盡成熟，尤其是受到儲存空間和記憶體效能的限制，使得電腦的運算功能受到相當的限制。而在 1990 至 2000 年間，由於電腦硬體研發的突破性發展，終於使第二波的人工智慧得以實現³³。

第二波的人工智慧以「統計學習」為基礎，亦即（智慧機器的）開發者針對欲解決的特定領域問題，建立統計模型 (statistical models)，並以「大數據」(big data) 的方式訓練智慧機器。DARPA 指出，第二波的人工智慧機器，具有將資訊進行精微分類與預測的能力，也已經具有最基本的推理 (reasoning) 能力，但是尚不具備脈絡化 (contextualization) 的能力。因此，第二波人工智慧的主要缺陷（或者用 DARPA 的話說，是「面臨的挑戰」），主要在於可能出現「統計上表現突出，但落實到個案上卻不足以信賴」的情況，而且這些錯誤從人類的觀點，可能認為是極為明顯、以人類智慧而言通常不會犯的錯誤。如下【圖二】所示，第二波人工智慧可能將下圖的「嬰兒拿牙刷」圖片，辨認為「小男孩拿著球棒」：

³² See, e.g., William J. Clancey, *Notes on "Epistemology of A Rule-based Expert System"*, 59 ARTIFICIAL INTELL. 197, 199-200 (1993).

³³ Scott Fouse, Stephen Cross & Zachary J. Lapin, *DARPA's Impact on Artificial Intelligence*, <https://ojs.aaai.org/aimagazine/index.php/aimagazine/article/view/5294> (last visited August 15, 2025).



【圖二】Challenges with second wave (DARPA Perspective on AI)

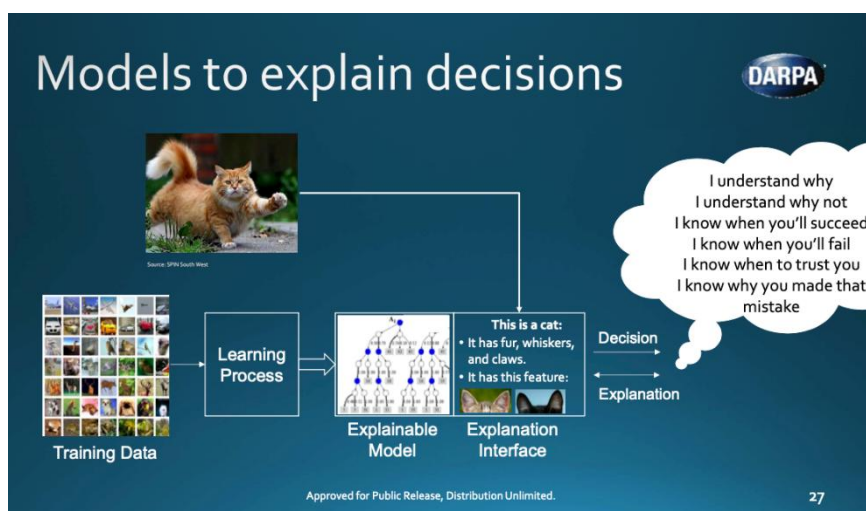
※ 資料來源： *Artificial Intelligence Exploration/Explanation*, DARPA (PDF), archived at Wayback Machine Oct. 23, 2020,

<https://web.archive.org/web/20201023055513/https://www.darpa.mil/attachments/AIFull.pdf>.

人工智慧機器對於上圖的錯誤理解，很可能是來自於對社會脈絡的認知不足；而增進對於脈絡的認知和理解，正是第三波人工智慧的目標：「脈絡適應」。因為訓練人工智慧的資料，在人類社會中經常具有多重或分歧的意涵，這些資料背後真正的意涵，必須視具體的脈絡而定³⁴。

³⁴ 邱文聰，前揭註31，頁137。

因此，第三波人工智慧的重點，在於智慧機器要有能力針對真實世界中不同類型的社會現象，建構具有不同脈絡解釋性的模型（contextual explanatory models）³⁵。換言之，第三波人工智慧希望建立「決策結果」和「決策理由」之間的關聯性系統，亦即「具備可解釋性的人工智慧」（explainable artificial intelligence）³⁶。以圖像識別（pattern recognition）中的經典案例「如何分辨圖中的動物是貓或狗」為例，第三波人工智慧要求不僅能夠正確識別出圖中是貓是狗，更要能說明達成此種決策結果的理由，其過程如下【圖三】所示：



【圖三】Models to explain decisions (DARPA Perspective on AI)

※ 資料來源：Artificial Intelligence Exploration/Explanation, DARPA (PDF), archived at Wayback Machine Oct. 23, 2020,

³⁵ Defense Advanced Research Projects Agency, *supra* note 31.

³⁶ Wojciech Samek & Klaus-Robert Müller, *Towards Explainable Artificial Intelligence*, in EXPLAINABLE AI: INTERPRETING, EXPLAINING AND VISUALIZING DEEP LEARNING 5 (WOJCIECH SAMEK et al. eds., 2019). Jonathan Johnson, *Interpretability vs Explainability: The Black Box of Machine Learning*, BMC BLOG (July. 16, 2020), <https://www.bmc.com/blogs/machine-learning-interpretability-vs-explainability/>

<https://web.archive.org/web/20201023055513/https://www.darpa.mil/attachments/AIFull.pdf>

三、人工智慧「取代」人類法院？

截至 2017 年 DARPA 發表人工智慧三階段論的報告時，第三波的「具備可解釋性的人工智慧，技術上尚無明顯的突破，DARPA 因此將之稱為「未來的人工智慧」（the future AI）。不過這個「未來」，即使到了本文截稿的 2023 年 11 月底，仍尚未見有明顯的進展³⁷。而本文第壹章中指出，大眾文化對於人工智慧得以「取代」司法系統進行判斷的描繪，正是奠基於第三波「脈絡調適」人工智慧的一種想像。因為司法系統的運作與判斷，本質上就是建立在因應人類社會脈絡而做出決策的一種典型情況³⁸。

以檢察官決定是否起訴某案件為例，舉凡判斷嫌疑人的行為是否已達到起訴標準³⁹？達到起訴標準的嫌疑人是否應使用緩起訴、相對不起訴或認罪協商的方式處理？檢察官對於已達起訴標準的嫌疑人，若選擇不起訴時，應命嫌疑人做出何種行為？是道歉、立悔過書、向被害人支付賠償？向國庫或公益團體支付金額？賠償或支付金額的額度如何決定？向公益團體提供義務勞務？義務勞務的種類為何？期間多長？⁴⁰凡此均涉及檢察官對於個案

³⁷ 關於「具備可解釋性的人工智慧」最新的發展，See Waddah Saeed & Christian Omlin, *Explainable AI (XAI): A Systematic Meta-Survey of Current Challenges and Future Opportunities*, 263 KNOWLEDGE-BASED SYSTEMS 1, 1-24 (2023).

³⁸ See Ryan Calo, *Artificial Intelligence Policy: A Primer and Roadmap*, 51 U.C.D. L. REV. 399, 413-15 (2017).

³⁹ 刑事訴訟法規定的起訴標準是「足認有犯罪嫌疑」，請參見刑事訴訟法第251條第1項規定：「檢察官依偵查所得之證據，足認被告有犯罪嫌疑者，應提起公訴。」然而，什麼時候「足認被告有犯罪嫌疑」？判斷標準何在？法律卻沒有更明確的定義。

⁴⁰ 刑事訴訟法第253條之2第1項：

「檢察官為緩起訴處分者，得命被告於一定期間內遵守或履行下列各款事項：

一、向被害人道歉。

二、立悔過書。

三、向被害人支付相當數額之財產或非財產上之損害賠償。

中嫌疑人的犯罪情節與生活情況、被害人及其家屬的生活情況和受犯罪造成之影響、社會公益與正義的需求、該地區近期與歷來犯罪趨勢等時空因素的複雜考量⁴¹。

而在法院判斷的層次，也有類似的社會脈絡適應性。不論是審前或審理中決定是否要羈押被告的判斷；審理案件時認為被告是否達到有罪判決的判斷；對於各項證據是否真實可信的判斷；對於雙方當事人是否已盡舉證責任的判斷；科刑或免刑判決的選擇；在諭知免刑判決時如何「斟酌情形」命被告向被害人道歉、立悔過書、向被害人支付慰輔金的判斷；諭知科刑判決時如何決定刑度的判斷等，這些判斷事項雖然具有不同的心證程度要求，但無一不涉及各種社會脈絡的理解、切換與調適⁴²。由此可知，如電影「關鍵報告」等大眾文化中所描繪「以人工智慧或其他智慧方式『取代』人類的司法

四、向公庫支付一定金額，並得由該管檢察署依規定提撥一定比率補助相關公益團體或地方自治團體。

五、向該管檢察署指定之政府機關、政府機構、行政法人、社區或其他符合公益目的之機構或團體提供四十小時以上二百四十小時以下之義務勞務。

六、完成戒癮治療、精神治療、心理輔導或其他適當之處遇措施。

七、保護被害人安全之必要命令。

八、預防再犯所為之必要命令。」

⁴¹ 依據法務部所訂定的「檢察機關辦理緩起訴處分作業要點」（以下簡稱為「作業要點」），這些考量五花八門，例如檢察官指定緩起訴之被告履行義務勞務時，除了考量所犯罪名之外，還必須考量「其性別、家庭、身分、職業、經歷、特殊專長、體能狀況及素行紀錄與參加意願等個人因素」（作業要點第3點第9項）；檢察官若命被告遵守「保護被害人安全之必要命令」或「預防再犯所為之必要命令」時（刑事訴訟法第253條之2第1項第7款、第8款），則必須依循「目的性、明確性、具體性、可行性、對應性、管考性、標籤性」等7大原則（作業要點第3點第11項）；而在決定受緩起訴被告之處分方式時，還必須考量「城鄉差距」之屬性，依據「機動調整落實實施原則、現有資源優先利用原則、公家機關優先合作原則、社區文化連結整合原則、運用傳媒密集宣導原則」等5大原則（作業要點第6點第2項），始能做出最終的決定，具有明顯的社會脈絡適應要求。

⁴² 關於量刑的社會脈絡適應性議題，請參見：范耕維（2022），〈建構量刑階段中罪刑相當原則的第一哩路：自應報觀點形塑刑罰量定之理論嘗試〉，《中研院法學期刊》，30期，頁148-157。

系統」，以做出更好、更令社會大眾滿意之司法判斷的想法，以現今技術而論，在第三波人工智慧「脈絡適應」的發展有所突破之前，暫時仍難以實現。

以上舉例的法律問題判斷，涉及的因素往往多端而複雜，不僅現有的人工智慧技術發展無法解決脈絡適應的問題，對於運用人類智慧的法院而言，也相當困難。具體而言，法院處理上述問題的難處主要有 2 點：第一，上述各項因素的「綜合考量」，往往極為抽象，很難說明做成一個具體法律決定的實質理由何在；第二，也由於「綜合考量」的因素多重且複雜，即使是相同的法院、完全相同的審判者，在不同的時間面對情節類似的案件，做出的決定也很可能不同；若由不同的法院、審判者來判斷這些情節類似的案件，其差距通常更大⁴³。對於類似情節的案件，法院若做出明顯不同的判斷結果，不僅難以說服社會大眾法院的判斷是正當合理的，甚至可能違反憲法平等原則的保障⁴⁴。

面對這些人類法院的判斷困難，儘管無法以人工智慧「取代」人類法院進行判斷，但是既有的人工智慧技術，以及大數據分析等科學技術，卻確實有可能提供法院有價值的參考資訊，輔助法院進行判斷。更重要的是，這些科技手段有明確的標準（儘管不一定是「妥適」的標準），可以幫助法院在不同案件中，對於相同事項維持相同的判斷標準。以人工智慧的手段而言，

⁴³ 審判者的組成方式對此也有影響。我國既有的刑事審判程序，僅以職業法官擔任審判者。職業法官雖有個體的差異，但是教育、考試、訓練過程是相對一致的，即使如此，都常因類似案件的判決落差過大現象，而遭致社會大眾非議，可見「綜合考量」之難。相對的，在納入了非法律專業人士為審判者的陪審團審判（jury trial），或是由專業法官和非法律專業人士共同組成審判者（例如我國已於 2023 年上路的國民參與刑事審判制度中的「國民法官法庭」）中，由於審判者的組成多元，背景多種多樣，在自然形成判決、亦即沒有同一的事實認定或量刑判斷基準介入的情況下，理論上判斷結果會有更大的差距。對此，無論是學者或實務界人士都表示了憂心。請參見：謝煜偉（2020），〈從量刑目的論形構量刑框架及量刑理由之判決架構〉，《法官協會雜誌》，22 卷，頁 103。文家倩（2023），〈從量刑工具探討國民法官的量刑評議〉，《司法周刊》，2156 期（別冊），頁 1。

⁴⁴ 關於平等原則與量刑間的關係，請參見：蘇凱平（2020），〈以平等原則建立量刑原則的意義與價值：臺灣高等法院 105 年度交上易字第 117 號刑事判決評析〉，《台灣法學雜誌》，393 期，頁 31-42。

最顯著的輔助法院判斷案例，應屬美國的 COMPAS 系統。美國有數個州的法院在量刑時，參考了 COMPAS 系統以機器學習原理運算出的再犯風險評估分數，作為量刑輕重的參考資訊，但卻引起了「是否侵犯憲法保障刑事被告之正當程序權利（right to due process）」的重大爭議與訴訟。COMPAS 系統的運用和憲法爭議，有助於吾人瞭解人工智慧輔助法院判決的實際作用和界限所在，本文以下詳細說明 COMPAS 系統和相關判決之內容⁴⁵。

參、以人工智慧輔助量刑之一：美國 COMPAS 系統

一、技術原理與應用情況

COMPAS 系統是由美國的私人企業 Northpointe 公司自 1998 年開始開發，設計目的是要輔助法院評估犯罪者潛在之再犯危險性⁴⁶。COMPAS 系統中包括多個風險評估模型，例如「一般風險評估」(General Recidivism Risk)、「暴力風險評估」(Violent Recidivism Risk)、「再犯風險篩選」(Recidivism Risk Screen) 和「審前釋放風險等級」(Pretrial Release Risk Scale) 等，均建立在統計學的基礎上⁴⁷。由於美國刑事案件數量眾多，刑事司法系統資源嚴重不足，COMPAS 透過統計模型的方式輔助法院進行決策，可以大量減

⁴⁵ 本文以下關於 COMPAS 系統與 Loomis 案判決的說明，著重於其運用人工智慧的方式，與既有的人類智慧判斷方式比較，揭示出哪些人工智慧的特性。至於其他國家或美國其他的風險評估系統，其運作方式與相關法院訴訟案件，請參見：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期，頁129-160。

⁴⁶ See generally Northpointe Inc., *Practitioner's Guide to COMPAS Core* (Mar. 19, 2015), <https://s3.documentcloud.org/documents/2840784/Practitioner-s-Guide-to-COMPAS-Core.pdf>. Northpointe 公司後來改稱 equivalent 公司（公司名稱 e 為小寫字母），COMPAS 系統併入 equivalent 公司推出的綜合風險需求評估套裝程式 Northpointe Suite Risk Need Assessments 之中。equivalent, *Our Products-Northpointe Suite Risk Need Assessments*, <https://www.equivalent.com/northpointe-risk-need-assessments/> (last visited August 15, 2025).

⁴⁷ See Northpointe Inc., *supra* note 46, at 1-2.

輕刑事司法系統的人力財力負擔，因此受到許多州法院的青睞。因此儘管有所爭議（詳下述），但截至 2020 年 2 月為止，美國繼續使用 COMPAS 協助評估被告或犯罪者風險的州法院，仍有佛羅里達（Florida）、密西根（Michigan）、威斯康辛（Wisconsin）、懷俄明（Wyoming）等州⁴⁸。

在 COMPAS 系統開發商釋出的 2015 年版使用手冊中，由於商業保密的考量，並沒有明確指出使用的技術方法，僅暗示其 2 個主要的模型「一般風險評估」和「暴力風險評估」，與「迴歸建模」（regression modeling）和「機器學習」（machine learning）的方法有關⁴⁹。舉例而言，COMPAS 的「暴力風險評估」的模型使用的公式為⁵⁰：

$$\text{Violent Recidivism Risk Score} = (\text{age} \times -w) + (\text{age-at-first-arrest} \times -w) + (\text{history of violence} \times w) + (\text{vocation education} \times w) + (\text{history of noncompliance} \times w)$$

其中的 w 代表的是某一個「權重」（weight），其所乘以的數值越大，對於最終風險分數的增加就越多； $-w$ 則代表「負向權重」，其所乘以的數值越大，對於最終風險分數的減低就越多。因此，以上的公式以中文表示即為：

$$\text{暴力再犯風險分數} = (\text{年齡} \times \text{負向權重}) + (\text{初次遭逮捕年齡} \times \text{負向權重}) + (\text{暴力史} \times \text{權重}) + (\text{職業教育程度} \times \text{權重}) + (\text{不服從歷史} \times \text{權重})$$

⁴⁸ Electronic Privacy Information Center, *AI and Human Rights: AI in the Criminal Justice System*, ARTIFICIAL INTELLIGENCE AND HUMAN RIGHTS, <https://epic.org/issues/ai/ai-in-the-criminal-justice-system/> (last visited August 15, 2025).

⁴⁹ COMPAS系統使用手冊中的說法，其大意是「本系統所使用的方法，在許多有關迴歸建模與機器學習的書籍中有所敘述」。Northpointe Inc., *supra* note 46, at 12-13. (“Our methods for developing and validating the General Recidivism Risk Scale were strongly influenced by the research of John Copas and colleagues who have developed an outcomes-based recidivism scale for England and Wales (Copas & Marshall, 1998). The methods used to develop both risk scales are described in various books on regression modeling and machine learning (*see e.g.*, Harrell, 2001; Hastie, Tibshirani, & Friedman, 2008; Kuhn & Johnson, 2013).”)

⁵⁰ Northpointe Inc., *supra* note 46, at 29.

上述用以評估暴力再犯風險的各種變項，據 COMPAS 使用手冊記載，是來自於歷來刑事司法研究對於暴力犯罪的整合分析（meta-analysis），具體方法是先特定出與暴力犯罪有關的特徵，包括年齡、初次遭逮捕年齡、暴力史、職業教育程度、不服從歷史等 5 個項目，並決定各個項目的權重⁵¹。然而，在上述公式中，儘管年齡、初次遭逮捕年齡、暴力史、職業教育程度、不服從歷史等項目，確實都以客觀存在的數據為本，但對於這些項目所乘上的「正向權重」 w 或「負向權重」 $-w$ 如何決定？資料來源和具體數值為何？COMPAS 系統的開發商並沒有說明。COMPAS 系統使用手冊僅指出：「權重數值乃依據本公司在研究資料中觀察到各項目對於個人再犯罪的關係強度而定」，並未提供明確的權重算法或計算過程⁵²。不過，研究 COMPAS 系統議題的文獻，均認為 COMPAS 使用的是機器學習方法，亦即屬於「統計學習」模式的第二波人工智慧⁵³。

由於 COMPAS 的開發商沒有揭露完整的演算法，因此無論是系統的使用者或是外部研究者，都難以測試其再犯風險評估公式中各項權重的計算是否正確或合理⁵⁴。這種情況造成了「法院使用 COMPAS 風險評估分數作為被告量刑參考，是否會侵害被告憲法保障的正當程序權利」的憲法爭議。美國威斯康辛州（以下簡稱「威州」）最高法院於 2016 年審理的 *Loomis v.*

⁵¹ *Id.*

⁵² *Id.* (“The size of the weight is determined by the strength of the item’s relationship to person offense recidivism that we observed in our study data.”)

⁵³ See e.g., Chris DeBrusk, *The Risk of Machine Learning Bias (And How to Prevent It)*, MIT SLOAN MANAGEMENT REVIEW: FRONTIERS (Mar. 26, 2018), <https://sloanreview.mit.edu/article/the-risk-of-machine-learning-bias-and-how-to-prevent-it/>; Ben Green & Lily Hu, *The Myth in the Methodology: Towards a Recontextualization of Fairness in Machine Learning*, presented in Machine Learning: The Debates Workshop, 35th International Conference on Machine Learning (ICML) (2018), <https://scholar.harvard.edu/files/bggreen/files/18-icmldebates.pdf>.

⁵⁴ Ashley Nicole Shadowen, *Ethics and Bias in Machine Learning: A Technical Study of What Makes Us “Good”*, in THE TRANSHUMANISM HANDBOOK 247, 250-53 (Newton Lee ed., 2019).

Wisconsin 案件（以下簡稱「Loomis 案」）⁵⁵，是處理此一憲法爭議的最知名案件。威州最高法院最後判決：量刑法院使用 COMPAS 系統提供的評估結果，作為量刑之參考資訊，並不違背被告受憲法保障的正當程序權利。威州最高法院所持的判決理由，對於瞭解人工智慧對於司法系統決策造成的影響和界線，有重大的啟示意義，本文以下詳細分析⁵⁶。

二、Loomis 案判決理由分析

本案的事實為被告、即上訴人 Loomis 因犯罪遭判刑，而事實審法院在決定其刑期時所使用的「量刑前調查報告」（Presentence Investigation Report）中，除了參考了 Loomis 過往的犯罪記錄外，也參考了經 COMPAS 系統運算出的再犯危險評估結果。在 COMPAS 系統的分類中，Loomis 在審前再犯風險、一般再犯風險，或暴力犯罪風險等 3 個風險評估項目中，均屬於具有再犯之高危險性的犯罪者，並最終遭法院判刑 6 年。Loomis 對此量刑判決上訴，主要理由是法官參考了 COMPAS 系統產出的高再犯危險判斷資訊，但是 COMPAS 系統究竟是如何運作而得出此種結論的過程並不透明，被告也因此無法對於系統運作的結果，為己提出辯護或解釋，因此 Loomis 認為其受憲法保障的正當程序權利受到侵犯，因而先上訴至威州上訴法院，被駁回後又上訴到威州最高法院。

本案的主要爭點在於：被告 Loomis 認為本案的事實審法院使用 COMPAS 系統的再犯風險評估程式來決定其刑罰，違反其受憲法保障的正當程序權利。因為根據專家證人提供的意見：第一，COMPAS 這整套風險評估機制，並不是為了輔助量刑決定而設計；第二，法院並不瞭解 COMPAS 系統實際上如何運作，包括不瞭解 COMPAS 是如何將本案被告的具體情況和前述「和本案被告有類似情況的資料群組」（統計證據）進行比較、法院

⁵⁵ Loomis v. Wisconsin, 881 N.W.2d 749 (Wis. 2016).

⁵⁶ 本案上訴人 Loomis 對於原審威州最高法院之判決不滿，遂於 2016 年上訴至美國聯邦最高法院，惟聯邦最高法院於 2017 年駁回其上訴聲請（petition denied），亦即本案最終以威州最高法院的判決確定。本文以下即根據威州最高法院的判決內容分析。

甚至不知道所謂「和本案被告有類似情況的資料群組」是來自威斯康辛州本地的統計數據、還是來自外地等。據此，被告 Loomis 主張：法院使用 COMPAS 系統對其量刑，共有以下 3 種情況違反了憲法保障刑事被告的正當程序權利⁵⁷：

- (1) 違反被告所受刑罰應根據「準確資訊」(accurate information) 判斷之權利。因為 COMPAS 開發商對其計算風險評估分數之演算法的詳情保密，被告和法院皆無從得知關於其演算法設計的準確資訊；
- (2) COMPAS 系統使用「群體數據」(group data) 評估風險高低，違反量刑時應該僅根據被告個人情況，量處個人化刑罰(individualized sentence) 的憲法保障；
- (3) COMPAS 系統將被告之「性別」列為考量因素，亦違反其受憲法保障之正當程序權利。

威州最高法院在審理 Loomis 的上訴時，首先說明其審查的重點在於「量刑法院在量刑時，是否使用了錯誤的裁量標準」。而根據歷來判決，這主要是指法院在量刑時是否考慮了「明顯無關或不適合的因素」(clearly irrelevant or improper factors)。被告對此有舉證責任，且必須證明至「清楚明確」的程度(clear and convincing evidence)⁵⁸。威州最高法院也承認此類風險評估工具的開發，原先確實只用在評估被定罪之被告是否適合於服刑前交付保護管束(probation)，或評估受刑人服刑一段期間後是否適合假釋(parole)，並不是設計來輔助量刑。只是由於美國社會面對與日俱增的犯罪威脅，各地都有希望法院量刑時能夠將「減少再犯」列為重要考量的強烈呼聲，遂使各地法院將此類的風險評估工具逐漸延伸至量刑領域⁵⁹。法院也指出，相較於

⁵⁷ Loomis, 881 N.W.2d at 756-57.

⁵⁸ Id. 關於證據清楚明確程度的舉證責任，可見王兆鵬、張明偉、李榮耕（2023），《刑事訴訟法（下）》，6版，頁177-178，新學林。

⁵⁹ 在判決中，威州最高法院只是敘述了「法院量刑時需要風險評估工具」的實際需求，但是卻沒有探討「將原先用於刑前交付保護管束或假釋的風險評估方法與資料，用於量刑時的再犯風險評估」的作法，本身存在評估方法與資料蒐集錯置

交付保護管束和假釋的決定主要就是考量被告的風險程度，在量刑議題中被告的風險程度僅是法院必須考量的因素之一，因此於量刑時，使用 COMPAS 系統這種風險評估工具的考量更為複雜⁶⁰。

（一）主張一：COMPAS 未能提供準確資訊

針對 Loomis 的第一點主張，威州最高法院同意 Loomis 所言，COMPAS 系統及其開發者確實沒有解釋該系統是如何使用相關資訊，以致能得出最終的風險評估分數⁶¹。然而法院認為 COMPAS 的 2015 年版使用手冊中，解釋了風險分數的評估主要來自於被告犯罪歷史等「靜態資訊」（static information），包括：「當被告獲得假釋後，有多少次曾經必須重新監禁？（Loomis 為 5 次以上）」、「當被告處於交付保管束期間，有多少次曾因新犯行而被逮捕或起訴？（Loomis 為 4 次）」、「被告曾有多少次刑事逮捕記錄？（Loomis 有 12 次）」等包含問題與答案的靜態資訊。威州最高法院認為，COMPAS 系統在針對 Loomis 進行風險評估時，使用的主要是公開資訊和 Loomis 本人對問題的回答，而且這些資訊都載明在 Loomis 也可閱讀到的量刑前調查報告中，因此 Loomis 實際上是有機會驗證這些 COMPAS 所使用的資訊是否為正確，進而挑戰 COMPAS 產出的風險分數。因此，

（displaced）的問題，以及不同決策所需的評估機制容許的錯誤率（error rate）不同的問題。評估方法與資料蒐集的錯置，係指原本為了評估交付保護管束或假釋而開發的評估系統，其所運用的方法與蒐集的資料，若用於不同事項（例如量刑）的評估，很可能會造成不準確、代表性不足和歧視特定族群的結果。關此，詳細論述請參見李榮耕，前揭註45，頁135-138。而容錯率不同，則是指相較於量刑出錯而言，交付保護管束或假釋的決定通常可以容許更高的錯誤率。例如70%準確率的風險評估工具，用在判斷是否交付保護管束或假釋的決定上，通常是可以接受的，但是在量刑時則無法接受有30%判斷錯誤的可能性。See John Lightbourne, *Damned Lies & Criminal Sentencing Using Evidence-Based Tools*, 15 DUKE L. & TECH. REV. 327, 335-37 (2017).

⁶⁰ Loomis, 881 N.W.2d at 752-54.

⁶¹ Id. at 761.

Loomis 主張其對於 COMPAS 系統所使用的資訊一無所知，因此無從挑戰，並非事實⁶²。

另一方面，威州最高法院也指出，COMPAS 系統是否提供了準確資訊的議題，涉及到系統本身的程式設計是否能提供準確有效的再犯危險性評估。由於威斯康辛州本身沒有對 COMPAS 系統評估再犯風險的有效性和預測準確性進行驗證研究，威州最高法院綜覽了紐約州、加州、佛羅里達州對 COMPAS 系統所進行的驗證研究後，認為由於各州對於此系統的有效性、準確性有不同的觀點。威州最高法院指出：由於本案爭點在於將 COMPAS 系統的風險評估用於量刑時，其準確性問題是否會影響到被告受憲法保障的正當程序權利。因此法院在此採取的立場是「只要在量刑前調查報告中，明確記載 COMPAS 系統風險評估準確性的使用限制和相關警語，則使用 COMPAS 系統於量刑事項即不違反被告受憲法保障的正當程序權利」。⁶³

（二）主張二：COMPAS 使用非被告個人的資料

被告第二點主張其所受科處的刑罰，並非基於其個人情況而量定，因為 COMPAS 系統使用了「據說」與其具有類似特徵的「群體數據」⁶⁴。威州最高法院也肯認了這個事實，在判決中指出：根據 Loomis 的量刑前調查報告，COMPAS 進行的預測是「『與本案被告有類似犯罪歷史情況之人』，在未來獲釋之後，較有可能或較不可能再次犯罪」，而非「『本案被告』再犯之可能性」。

不過，威州最高法院認為，依據該州歷來判決的估點，如果法院是將 COMPAS 系統的風險評估分數作為量刑的「決定性因素」，此時確實會有「量刑並非依據被告個人情況」的憲法疑慮。然而，本案中 COMPAS 系統雖然使用了群體數據，但其作用正如同保險業的精算評估系統，於評估個別

⁶² *Id.*

⁶³ *Id.* at 763-64.

⁶⁴ 亦即「使用本案以外之資訊，作為判斷本案爭點」的「統計證據」，請詳見本文「壹、科技輔助司法決策的熱區：統計證據」章之說明。

駕駛的風險時，會考慮與該駕駛有類似情況的群體數據一般。況且，在 COMPAS 的使用說明書中，也明確指出「風險評估乃用於預測群體行為（group behavior）……而非預測個人行為。」亦即 COMPAS 系統中的「風險」評估，原本就不是用來識別出某一特定的「個別犯罪者」是否具有再犯的高風險（a particular high-risk individual），而是用以識別出具有再犯高風險的「犯罪群體」（groups of high-risk offenders）⁶⁵。

威州最高法院認為，根據使用手冊的上述說明，COMPAS 的使用者本來就知道應該依其專業和個案情況，判斷裁量出最適合的處遇方式。例如威州的矯正機關在使用 COMPAS 時評估受刑人風險時，在約有 10% 的案件中，矯正機關會因為另外考慮一些在 COMPAS 演算法中考慮較少的加重或減輕事由，而做出與 COMPAS 系統不同的判斷⁶⁶。威州最高法院由此立論：既然量刑法院有能力、也有義務裁量各種量刑相關因素，則當法院發現 COMPAS 的風險評估結果，與其他量刑考慮因素呈現出來的再犯風險並不一致時，應可期待量刑法院運用其裁量權，依據個案中被告的個人情況量處適當之刑罰。因此，威州最高法院認為，不能僅僅因為法院量刑時參考了 COMPAS 的風險評估資料，即認為其量刑決定並非基於被告個人的情況。在 COMPAS 系統已經明載其用途與限制的情況下，量刑法院將 COMPAS 的風險評估結果列為其參考資料之一、而非作為決定性的單一因素，並未侵犯被告受憲法正當程序保障的權利⁶⁷。

（三）主張三：COMPAS 將性別列為風險評估因素

Loomis 的第三項主張，是法院量刑時將其性別列為考量因素，亦屬違反其受憲法保障之正當程序權利。在此，Loomis 並非主張 COMPAS 系統認

⁶⁵ Loomis, 881 N.W.2d at 764.

⁶⁶ See also WISCONSIN DEPARTMENT OF CORRECTIONS, ELECTRONIC CASE REFERENCE MANUAL, *COMPAS Assessment Frequently Asked Questions*, 2-5 (2018), <https://www.manuallib.com/download/2023-10-18/Electronic%20Case%20Reference%20Manual.pdf> (last visited August 15, 2025).

⁶⁷ Loomis, 881 N.W.2d at 765.

為其性別（男性）有較女性更高的再犯風險，有違平等原則；而是主張 COMPAS 系統將「性別」列為再犯風險評估因素的作法，不論其如何評估，都違反威州法院過往判決（普通法）認為量刑時不得考慮被告性別之見解，因此有違刑事被告正當程序權利的保護⁶⁸。

威州最高法院則認為，Loomis 誤解了歷來判決的觀點。法院指出：量刑時並非絕對不得考慮與性別有關的因素，重點在於法院量刑時若考量性別因素，是否「以事實為基礎」（a factual basis）。由於歷來多有文獻指出，男性的再犯率與暴力犯罪率確實都較女性為高，威州最高法院認為 COMPAS 系統將性別列入評估被告再犯風險時的考量因素，確實具有事實基礎，可以增進系統運算結果的準確度，對於公共利益、甚至是被告的利益均有助益。此外，儘管 COMPAS 的系統運作中考慮了性別，但是如前所述，COMPAS 的評估結果只是量刑法院參考的因素之一，量刑法院是否確實考量 Loomis 的性別，以致於最後做出此種量刑判斷，需要由被告負起舉證責任，且須證明至清楚而明確的程度，而 Loomis 未能達成此種舉證責任。基於以上理由，威州最高法院認為「COMPAS 系統的演算法中，考慮性別為風險評估因素之一」雖是事實，但並未侵害被告受憲法保障的正當程序權利⁶⁹。

綜合以上理由，威州最高法院認為，在 COMPAS 系統被正常使用的前提下，基於量刑前調查報告中明確記載的 COMPAS 系統使用限制和相關注意事項（包括 COMPAS 的風險評估不應用於決定刑罰輕重，亦不應用於決定被告是否應被判處自由刑等警語），則使用該系統作為量刑參考事由，並不違反憲法對被告的正當程序保障⁷⁰。

為了方便讀者理解和比較，本文將 Loomis 在訴訟中提出的主張，與威州最高法院的回應，整理如【表一】：

⁶⁸ *Id.* at 766.

⁶⁹ *Id.* at 766-67.

⁷⁰ *Id.* at 771-72.

【表一】威州最高法院 State v. Loomis 案判決整理

	上訴人即被告 Loomis 主張	威州最高法院的回應
主張 1	被告無從得知 COMPAS 用以進行風險評估的「準確資訊」之內容。	COMPAS 使用的部分資訊，是來自被告回答自身犯罪歷史的問卷，被告有機會就此部分資訊進行驗證，並挑戰 COMPAS 系統的運算結果。就 COMPAS 系統本身運算的準確度，由於不同的研究有不同的觀點，因此只要 COMPAS 系統有載明其使用的限制等警語，法院量刑時參考 COMPAS 產出的資訊，即不違背被告的憲法正當程序保障。
主張 2	COMPAS 使用的是群體數據，而非被告個人的實際情況作為風險評估依據。	COMPAS 系統使用說明書中有明確記載其使用的是群體數據。包括量刑法院在內的 COMPAS 系統使用者，應可知悉 COMPAS 的運算結果僅供參考，使用者仍必須自行裁量判斷，以做出最終的決定。
主張 3	COMPAS 使用性別作為評估風險的要素之一，違反憲法禁止法院以性別為刑罰考量因素的規範。	「COMPAS 系統」考量性別做為被告再犯風險的評估因素，有合理的事實依據。 被告無法證明「量刑法院」是因為考量了其性別，才做出本案中的量刑決定。

※ 資料來源：作者自製。

三、Loomis 案揭示的人工智慧特性

雖然 Loomis 案件判決的法律上爭點，是在於「刑事被告受威州憲法保障的正當程序權利，是否因為量刑法院參考了 COMPAS 提供的風險評估資訊而受到侵害」，並非直接挑戰「法院能否使用人工智慧產出的資訊以輔助其判斷」的根本問題。惟細譯本案判決理由，可以發現上訴人、即被告 Loomis 的主張和質疑，其實正是圍繞著第二波人工智慧的特性「統計學習」展開。

如本文前述，第二波的「統計學習」人工智慧中的主流「機器學習」，是建立統計模型並以大量數據訓練電腦，從而得出與訓練資料符合的判斷結果。然而，機器學習等第二波人工智慧，並不具備能夠解釋其決策理由的能力，亦即欠缺「可解釋性」（explainability）。COMPAS 固然因為是私人企業開發的程式，有商業保密的考量，而不能將其所使用的運算資料完全公開；但即使沒有這一層商業考量，純粹從人工智慧進展的觀點而言，COMPAS 所使用的機器學習方法，也無法清楚說明其公式中各項數值乘以的正負向「權重」是如何得出的。換言之，機器學習方法本身具有相當程度不可解釋的「黑盒子」性質⁷¹。

回顧 Loomis 的第一項主張：「被告無從得知 COMPAS 用以進行風險評估的『準確資訊』之內容」，其實正是一種對於 COMPAS「黑盒子」性質的挑戰。不過，從人工智慧的技術觀點而言，Loomis 的這一項挑戰是相對表面、淺層的，僅從商業保密的觀點質疑 COMPAS 的開發商沒有釋出足夠的準確資訊，以致於「被告」無從挑戰 COMPAS 的風險評估運算結果。事實上，COMPAS 不只存在著「被告無從得知其運作的準確資訊」這樣的「主觀不能」（被告本人無從得知）情況，更根本的問題在於「任何人（即使開發商自身）都無從得知 COMPAS 是如何通過訓練資料而得出各項權重數值」的「客觀不能」（任何人都無從得知）困境⁷²。

⁷¹ 劉靜怡（2020），〈人工智慧時代的法學研究路徑初探〉，李建良（主編），《法律思維與制度的智慧轉型》，頁126-129，元照。

⁷² 關於機器產出的資訊能否有效驗證的觀點，請參見：劉靜怡，前揭註71，頁127。
See also Andrea Roth, *Machine Testimony*, 126 YALE L.J. 1972, 2034 (2017).

而 Loomis 的第二項主張「COMPAS 使用的是群體數據，而非被告個人的實際情況作為風險評估依據」，同樣涉及第二波人工智慧的技術本質。因為統計學習模式的本质，正是在依據相關學科領域的觀點，建立了一套相對合理的統計模型後，投入大量資料加以訓練。而這裡所稱的「大量資料」，顯然並非全部來自於要接受刑罰的被告個人之資訊，而是來自大量被認為在某種框架或定義下，與被告具有同一處境、特徵或性質的人士，例如因同一罪名遭定罪的其他刑事被告、具有某些相同犯罪歷史的其他刑事被告等。換言之，COMPAS 系統是以採納「統計證據」為基礎建立，而當法院使用 COMPAS 系統時，等同於採納了「以統計證據（本案以外之資訊）作為判斷依據」的立場⁷³。

這種第二波人工智慧的技術特性，引起了被告對於「使用 COMPAS 是否違反刑罰應『個人化』」的質疑。因為依據一般的量刑理論和規範，對犯罪者的刑罰應該圍繞著「犯罪者」及其「所犯之罪」具體考量⁷⁴。如果在對犯罪者 A 量刑時，所考慮的卻是其他類似情況的犯罪者 X、Y、Z……等人的刑罰情況，無異於背離了量刑理論。縱然如判決中所指出：COMPAS 原本的設計目的並非是要輔助量刑決策，而是為了在刑事司法系統判斷被告是否適合交付保護管束或准予假釋時，提供評估被告再犯風險的資訊。然而，即使是在保護管束或假釋的判斷上，向來的學理也認為應該圍繞著系爭被告或受刑人的個別情況加以評估。換言之，即令是在 COMPAS 系統開發出來的主要目的上，使用第二波人工智慧統計學習模式的 COMPAS 系統，都仍無法免於「使用與被告具有類似特徵的群體資料，替代特定被告的個人情況

⁷³ 本文此處稱為「統計證據」，是從證據法的觀點來看。若從統計學的觀點，COMPAS 使用了「代理變項」（proxy variable），亦即對於無法測量或不易測量的對象（例如個案中被告未來的再犯可能性），使用可以測量或容易測量的變項來進行測量（例如過去具有類似特徵之被告的再犯情況）。See e.g., David Enoch & Talia Fisher, *Sense and "Sensitivity": Epistemic and Instrumental Approaches to Statistical Evidence*, 67 STAN. L. REV. 557, 595 (2015).

⁷⁴ 例如我國刑法第 57 條的規定。我國刑法學理上通常以「罪責原則」中的「罪刑相當原則」（行為人所受的刑罰，應與其罪責相當）來說明此一概念。請參見：王皇玉（2023），《刑法總則》，9 版，頁 58-59、636，新學林。

進行風險評估」的質疑，因為使用「統計證據」（或說「代理變項」，proxy variable）正是統計學習模型進行運算的技術特性之本質⁷⁵。

由於威州最高法院以單純「法律審」的方式處理本案，上述關於使用群體數據的重要問題，並沒有在本案判決中呈現出來。因為威州最高法院並沒有深究 COMPAS 系統所使用的統計學習人工智慧特性，而是以「COMPAS 運算出的資訊僅是供量刑法院『參考』，並非決定性的因素，且使用手冊中已經載明了其資訊的限制和警語」為基礎，展開了對於被告全部 3 項主張的回應。

換言之，威州最高法院認為本案的爭點並非量刑是否妥適，而是 Loomis 的正當程序憲法權利是否被侵害；而判斷憲法權利是否被侵害，則取決於量刑法院運用其裁量權做出決定的方式是否適當。一旦問題的焦點集中在量刑法院的裁量權運用，因為量刑法院本來就有權評估被告未來的再犯可能性，再與被告個人特質和所犯之罪的特質等因素「綜合考量」適合的刑罰種類與輕重程度，在此脈絡下，COMPAS 系統運算出的風險評估結果作為法院量刑時的參考因素之一、而非決定性的因素，自然難以認為量刑法院參考了這一項特定的資訊，就使被告的正當程序憲法權利受到侵害。

至於被告的第三項主張「COMPAS 使用性別作為評估風險的要素之一，違反憲法禁止法院以性別為刑罰考量因素的規範」，同樣涉及了前述第二波人工智慧的各項技術特性，而法院的回應也同樣是針對法律面、避開了人工智慧的科技本質問題。例如因為現階段人工智慧的「不可解釋性」，法院僅能說明 COMPAS 在評估風險時考慮性別因素，是有事實依據為基礎的，但是法院並未處理「COMPAS『如何』考慮性別及其理由何在？」這個第二波人工智慧仍難解釋其決策理由所在的痛點問題；法院也同樣用了「被告無法證明『量刑法院』是因為考量了其性別，才做出本案中的量刑決定」這種揉合了舉證責任分配和量刑法院有權裁量參考各種資訊的「法律」說法來回應。

本文認為，威州最高法院採取這種可說是純粹法律的取徑，來回應被告主張，固然可以認為是「就法論法」，但是避開了對於人工智慧運作原理的

⁷⁵ 請參見前揭註73的說明及所列參考文獻。

探討，卻因此滋生更多法律、甚至是憲法層次的問題。例如調查報導媒體 ProPublica 於 2016 年揭露 COMPAS 系統的運作結果帶有種族偏見，原因正在於 COMPAS 系統將過去犯罪者的「種族」列為一項評估風險的因素⁷⁶。而這種作法，若比照威州最高法院對於 COMPAS 將「性別」列入評估風險因素的觀點，也可以說有其事實的基礎。因為在過去的某段時間內，某些種族人士遭到逮捕或判決有罪、入獄的比例，可能確實比其他種族人士來得更高。然而，這種「事實基礎」卻很可能是來自於過往刑事司法系統運作過程中，在執法實務上對於某些族群本有的歧視作法⁷⁷。本文認為，在審查用於輔助司法判斷的人工智慧模型時，如果忽略了對於人工智慧技術本質的探討，將第二波人工智慧的「不可解釋性」和「使用群體資料訓練模型」等技術特點排除在討論的範圍之外，審判很容易就會變成一場「容許過往的歧視作法透過人工智慧模型的『黑盒子』演算法不斷複製」的悲劇，這種作法恰恰侵害了 Loomis 案判決中威州最高法院所設定的審理主題：「被告受憲法保障的正當程序權利」。

同樣為了解決法院量刑時的困難，我國司法院於 2011 年起逐步建置了 2 套量刑輔助系統，分別是「量刑資訊系統」與「量刑趨勢建議系統」。由於當時人工智慧的系統開發技術在我國尚未普及，司法院並沒有採取如同美國 COMPAS 系統使用的機器學習人工智慧方法，而是以法律數據分析(legal data analysis)的取徑建置上述系統，希望幫助法院解決量刑時經常遭遇的各種困難，包括法院需要考量的因素複雜多端、法院考量的具體因素不透明、類似案件的量刑結果不一致等問題⁷⁸。在上述量刑輔助系統問世大約 10 年

⁷⁶ Julia Angwin et al., *Machine Bias: There's Software Used Across the Country to Predict Future Criminals. And It's Biased Against Blacks*, PROPUBLICA (May. 23, 2016) <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>.

⁷⁷ See, e.g., Anthony W. Flores et al., *False Positives, False Negatives, and False Analyses: A Rejoinder to "Machine Bias: There's Software Used across the Country to Predict Future Criminals. And It's Biased against Blacks."*, 80 FED. PROBATION 38, 38 (2016).

⁷⁸ 司法院官方網站（刑事量刑專區），<https://www.judicial.gov.tw/tw/cp-83-57186->

後，隨著法律科技的進步與人工智慧技術的普及，司法院又於 2020 年至 2021 年間，基於機器學習的人工智慧方法，開發了「AI 輔助量刑資訊系統」（或稱「量刑智慧分析系統」）並於 2023 年啟用⁷⁹；更進一步在 AI 量刑系統的基礎上，發展出以人工智慧撰寫裁判書草稿的系統⁸⁰。

為了清楚指稱司法院在不同時期開發的量刑輔助系統，本文以下將 2011 年起逐步建置的「量刑資訊系統」與「量刑趨勢建議系統」合稱為司法院的「第一代量刑輔助系統」；而將 2023 年起啟用的「AI 輔助量刑資訊系統」（包括「事實型」和「評價型」的 AI 系統），以及在此系統基礎上建立的人工智慧草擬判決書系統，併稱為司法院的「第二代量刑輔助系統」。由於司法院的「第一代量刑輔助系統」並未使用人工智慧技術，其意義、技術原理、效果與面臨的問題，都會與美國 COMPAS 系統有明顯的不同。而司法院「第二代量刑輔助系統」，其技術原理固然更接近同樣使用機器學習方法的美國 COMPAS 系統，但因為司法院係以政府機關角色，推動以人工智慧技術建立的量刑輔助系統，司法院也會面對與美國 COMPAS 系統（以私人、商業模式建置系統）不完全相同的問題。

必須說明的是，雖然司法院的第二代量刑輔助系統已經問世，但是第一代量刑輔助系統並沒有「退役」，而是在量刑輔助系統的使用中佔有主要的地位。究其原因，一方面是因為第一代量刑輔助系統發表已有十餘年，在司法實務上運用的實例頗多；另一面則是因為第二代量刑輔助系統在 2023 年 2 月才開放法官使用、同年 5 月和 9 月才先後將其中的「事實型」AI 系統與「評價型」AI 系統開放給檢察官、律師使用，而迄於本文完稿的 2023 年 11

1ef46-1.html（最後瀏覽日：05/15/2024）。

⁷⁹ 司法院官方網站（刑事量刑專區），<https://www.judicial.gov.tw/tw/cp-83-57186-1ef46-1.html>（最後瀏覽日：05/15/2024）。

⁸⁰ 司法院（08/27/2023），〈司法院審慎發展生成式AI應用，以撰寫刑事裁判草稿初試啼聲；期望減輕法官工作負荷，審判核心仍由法官自行決定〉，<https://www.judicial.gov.tw/tw/cp-1887-929494-8a9fb-1.html>（最後瀏覽日：05/15/2024）。

月底，一般刑事被告與社會大眾都尚未能使用第二代量刑輔助系統，因此判決實務上目前對於第二代量刑輔助系統還鮮有討論。

由於司法院建置的第一代量刑輔助系統採用的技術方法，與美國 COMPAS 系統明顯不同，可以兩相對照，探討刑事量刑實務需要人工智慧或數據分析技術奧援的需求面何在。本文以下第肆章，首先探討第一代量刑輔助系統之技術原理與實際運用方式，與美國 COMPAS 系統運用的人工智慧方法相比較，藉此進一步映射出「人工智慧」與「人類智慧」的特徵差異。其後，則於第伍章介紹司法院最近建置的第二代量刑輔助系統，包括此系統與 COMPAS 系統同樣以機器學習的人工智慧方法建置，因此面臨的共同質疑；以及第二代量刑輔助系統在我國脈絡下，所面對不同於 COMPAS 系統的獨有問題。本文希望透過這些比較，提取出各種輔佐法院量刑心證決定的系統之特徵，並呈現各種量刑輔助系統所涉及的憲法和法律問題，以及社會對於其應用的不同顧慮。

肆、以大數據輔助量刑：司法院第一代量刑輔助系統

一、技術原理與應用顧慮

（一）技術原理

司法院最早是在 2011 年時，為了回應引起「白玫瑰運動」的法院量刑欠缺公平、透明和妥適現象，利用法律數據分析方法建置了 2 種基於不同理念的量刑輔助系統，分別是「量刑資訊系統」與「量刑趨勢建議系統」，此即為司法院建置的第一代量刑輔助系統，也是我國使用電腦數據方法進行量刑改革的開始⁸¹。依據司法院在官方網站上的說明，量刑資訊系統是「結合

⁸¹ 司法院官方網站（刑事量刑專區），<https://www.judicial.gov.tw/tw/cp-83-57186-1ef46-1.html>（最後瀏覽日：05/15/2024）。「量刑向來是各國刑事審判實務之難

刑事法官、統計、資訊專業及編碼人員，從各類犯罪之法院判決書中，蒐集判決書上記載之刑法第 57 條各款量刑審酌事由」而建置。自 2011 年 7 月起，陸續啟用了妨害性自主罪、不能安全駕駛罪、槍砲案件、幫助詐欺罪、毒品相關犯罪、竊盜罪、搶奪罪暨強盜罪、殺人罪、普通傷害罪、肇事逃逸罪等量刑資訊系統。此外，並針對刑法上放火罪、偽造有價證券罪、妨害風化罪、妨害婚姻及家庭罪、賭博罪、妨害自由罪、妨害名譽及信用罪，以及懲治走私條例、毒品危害防制條例、商標法、著作權法、證券交易法、貪污治罪條例等多種特別刑法規定，另行建置了「類似判決刑度資訊檢索系統」⁸²。

同樣依據司法院官方網站的記載，「量刑趨勢建議系統」是以專家「焦點團體」（focus group）的研究方法所建置，具體方法是「蒐集各類犯罪之法院判決，以統計迴歸方法，分析法官量刑時考量之量刑因子，與其對量刑影響力大小，並召開焦點團體會議，匯集審、檢、辯、學及相關民間團體，對於分析實務判決所得之量刑因子及其影響力大小討論所形成之意見，據以提供各類犯罪之量刑趨勢，供法官參考」。量刑趨勢建議系統最早自 2014 年 9 月啟用，包括的犯罪類型有不能安全駕駛罪、對未成年人性交罪、強制性交罪與加重強制性交罪、殺人罪、傷害致死罪、幫助詐欺罪、違反槍砲彈藥刀械管制條例案件之罪、竊盜罪及加重竊盜罪、強奪罪及強盜罪、電信詐欺罪等⁸³。

根據上述的系統建置原理可知，「量刑資訊系統」是根據「實然面」資訊建立的系統，藉由篩選出過往實際上發生案件中的特定「量刑因子」，對

題，追求妥適量刑亦是各國司法改革之重點。民國99年9月間，高雄地院一件性侵幼童的判決，引發『白玫瑰運動』，本院基於司法為民的理念，回應各界對公平、透明、妥適量刑的要求，自100年2月中旬起，首先針對民眾所關注的妨害性自主罪，著手建置符合我國民情、法制的『妨害性自主罪量刑資訊系統』，揭開我國司法量刑改革的序幕。」

⁸² 司法院官方網站（刑事量刑專區），<https://www.judicial.gov.tw/tw/cp-83-57186-1ef46-1.html>（最後瀏覽日：05/15/2024）。

⁸³ 司法院官方網站（刑事量刑專區），<https://www.judicial.gov.tw/tw/cp-83-57186-1ef46-1.html>（最後瀏覽日：05/15/2024）。

比出與當下法院所審理、需要量刑之案件具有類似性（具有類似量刑因子）的過往案件，藉由這些過往真實案件的量刑結果，對於要進行量刑的法院發揮提示（可能的量刑區間何在？）或提醒（科刑法院所欲量處之刑罰，相較於過去法院量刑，是否太重或太輕？）的作用。因此，量刑資訊系統運作後，輸出的結果是過往類似案件的「量刑分布統計狀況圖」、「最低刑度」、「最高刑度」與「平均刑度」等資訊。相對地，「量刑趨勢建議系統」反映的則是「應然面」的某種價值判斷，目的在於系統性地反映出司法系統中審、檢、辯、學，以及婦幼、被害人保護等民間團體等不同立場的個人專家或專業團體，對於各類犯罪應該如何量刑的意見，希望彌補法官的生活經驗、提供法官量刑時的參考建議，目標是做出更符合社會期待的量刑判斷結果。因此量刑趨勢建議系統運作後給出的結論，是「建議刑度」和「建議量刑區間」⁸⁴。

雖然司法院基於不同的理念分別建置了「量刑資訊系統」與「量刑趨勢建議系統」，但就技術方法而論，2 個量刑輔助系統都是由大量的過往判決中，分析出影響量刑決定的因素（稱為「量刑因子」），再進行比對或加權計算的分析，均屬於一種利用法實證研究原理的（大）數據分析方法⁸⁵。

著重「實然面」的「量刑資訊系統」，透過蒐集過往判決中的量刑結果，建立量刑相關因子的變項，以大數據的方式呈現出具有類似因子的過往判決之量刑情況，以供今日法院對個案量刑的參考⁸⁶。量刑資訊系統的原理是對於大量判決中量刑因子的建立與篩選，雖然看來相對基礎，但是如果法院普

⁸⁴ 第一代量刑輔助系統中，2 個子系統分別是以「應然」與「實然」的方法論所建立的觀察，最早可能是由本文作者在另一篇論文中提出，請參見：蘇凱平（2020），〈以司法院量刑資訊作為量刑之內部性界限？：評最高法院108年度台上字第3728號刑事判決〉，《月旦裁判時報》，98期，頁90-91。

⁸⁵ 關於法律數據分析之意義、原理、技術與數據分析方法如何輔助法律決策，請參見：蘇凱平（2019），〈法律數據分析的意義、理論與應用：以探索刑事法院對證據的裁量與評價為例〉，《月旦法學雜誌》，294期，頁101-124；邵軒磊、吳國清、黃詩淳（2018），〈大數據與法資訊學：機器提取裁判內容要素之實踐〉，《月旦裁判時報》，71期，頁46-52。

⁸⁶ 司法院（2015），〈司法院量刑準據八大系統建置完成〉，《司法周刊》，1771期，頁1。

遍願意使用，確實可以促成量刑因子的透明化，並藉此達成「類似案件，類似量刑」的期待⁸⁷。

因為在「量刑資訊系統」創建之前，多數法官為了避免自己的量刑出現畸輕畸重的情況，而欲參酌其他法官同儕的量刑時，僅能透過翻閱司法院或各法院製作的裁判彙編，或至司法院法學資料檢索系統等判決蒐集網站上，自行設定關鍵字搜尋過往判決⁸⁸。然而這些作法都不具有系統性，獲得的結果高度取決於個別法官所請教的對象，或是法官自行認為應影響量刑的因子為何，很難達到促成類似情節的不同案件間，量刑應具有一致性或相類似的目標。再者，即使法院在量刑時有意達成與過往類似案件量刑一致的目標，也不會在判決中具體說明是請教了哪些法官的經驗、或參考了哪些具有類似量刑因子的過往判決後做成了本案的量刑決定。因此，如果認為量刑的政策目標，應包括與不同、但是具有類似情節之案件間的量刑平等在內，或是包括能對於量刑因子有清楚的說明在內，則量刑資訊系統確實在這些目標的達成上，提供了單純的人類智慧運用所難以企及的效果⁸⁹。

而「量刑趨勢建議系統」則關注「應然面」，不僅著眼於今日的判決結果與過去類似的案件具有一致性，而是更希望法院量刑「能融入社會公平正義之情感，反應社會各方之價值與期待」⁹⁰。根據參與建置此系統的司法院刑事廳法官說明，司法院建置「量刑趨勢建議系統」主要分為2步驟：第一，先以迴歸分析方法找出各類型實務判決中影響量刑結果的顯著量刑因子，並

⁸⁷ 關於量刑資訊系統的具體操作方式以及量刑因子的設定，請參見：蘇凱平（2022），〈對法院量刑心證的追索：司法院量刑資訊系統之運作與展望〉，李建良、林文源（主編），《人文社會的跨領域AI探索》，頁312-314，國立清華大學出版社。

⁸⁸ 胡宜如（2015），〈司法院焦點團體建議量刑參考基準介紹〉，《司法周刊》，1766期，頁2。

⁸⁹ 此處的政策目標，是指司法院「刑事量刑專區」官方網站上所記載的「本院基於司法為民的理念，回應各界對公平、透明、妥適量刑的要求」的「公平、透明、妥適」而言。司法院官方網站（刑事量刑專區），<https://www.judicial.gov.tw/tw/cp-83-57186-1ef46-1.html>（最後瀏覽日：05/15/2024）。。

⁹⁰ 司法院官方網站（刑事量刑專區），<https://www.judicial.gov.tw/tw/cp-83-57186-1ef46-1.html>（最後瀏覽日：05/15/2024）。。

分析各項因子交互影響的情況，以及影響力的大小；第二，使用焦點團體（focus group）方法，集合審、檢、辯等法律實務工作者；法律學、統計學、犯罪學等學科之專家學者；及婦幼團體、社工團體、被害人保護團體等公益團體代表之意見，對於參數估計不穩定或影響權重被認為不合理的量刑因子進行調整。透過以上 2 步驟，最終建立一個「量刑估計模型」。當使用者鍵入量刑因子，透過量刑估計模型的運算，系統即可提供焦點團體建議之刑度預測與量刑區間預測，作為法院對個案量刑之參考資訊⁹¹。

（二）應用顧慮

從上述說明可以發現，指向「應然面」的量刑趨勢建議系統，在建立其量刑區間預測與建議之模型時，雖然並沒有使用人工智慧的方法，但是其建立模型的基本原理，很類似於美國 COMPAS 系統以「風險因子乘以特定權重之總和」之公式得出風險評估分數的作法。只是在關鍵的「權重」設定，COMPAS 系統使用了「人工智慧」，我國的量刑趨勢建議系統使用的則是「人類（專家）智慧」⁹²。

不論是使用人工或人類智慧來決定權重，COMPAS 系統或量刑趨勢建議系統顯然都不是僅僅從法院審理的特定個案中取得的訊息，而是利用成千上萬個過往的類似判決中的資訊，所得出的判斷結果。換言之，都是「使用本案以外之資訊，作為判斷本案爭點的證據」，亦即使用「統計證據」的情況。那麼，如同 Loomis 判決中發生的憲法爭議，在我國法院使用量刑趨勢建議系統作為輔助時，是否也同樣會發生呢？

本文認為，Loomis 案判決中反覆出現的議題：「COMPAS 系統所提供的量刑輔助參考資訊，究竟在個案被告的量刑中，扮演何種角色？」依據目前司法院對第一代量刑輔助系統界定的功能觀之，在我國並不會引起爭議。

⁹¹ 胡宜如，前揭註88，頁2-3。

⁹² 雖說如此，但是美國的COMPAS系統、甚至其他的風險評估系統，在決定其公式的權重時，仍無法完全避免人類專家的介入。請參見：李榮耕，前揭註45，頁133-135。

因為司法院非常明確的表示：不論是建構在實然面上的「量刑資訊系統」，還是指向應然面的「量刑趨勢建議系統」，都僅供法院「參考」，並不「拘束」法院在個案中的量刑⁹³。對照 Loomis 判決中，威州最高法院對於上訴人 Loomis 就 COMPAS 系統提出的各種質疑，幾乎都以「COMPAS 系統產出的資訊僅供法院參考，並無拘束力」為基礎加以駁斥，可以發現目前不論是美國或我國，對於人工智慧或其他科技手段在司法決策中的定位都很清楚，亦即僅具有「輔助」和「參考」的作用⁹⁴。最終的決定，仍然必須由人類的審判者做成。換言之，在今日的司法決策中，仍然認為應以人類智慧為主，人工智慧或數據分析等科技手段則被界定為只能提供參考資訊的輔助角色。

二、法院的適用擴張現象

不過，人類法官對於科技手段產出資訊的「參考」，是否有其他的要求或限制呢？具體而言，因為人工智慧或其他的科技系統，可能具有「人不能」的優勢，例如司法院所建置的量刑資訊系統，有透過大數據以顯示「一般」案件如何量刑的能力，那麼是否在某些情況或條件成就時，可以要求司法系統決策時，「應參考」科技手段所提供的資訊呢？舉例而言，當個案中法官的量刑，明顯高於或低於某個科技系統顯示出的既有（或應有）量刑區間時，上訴審法院是否可以要求原審量刑法官，針對其量刑決定與科技手段顯示出的資訊明顯不一致這一點，特別提出說明？上訴審在衡量原審法院的量刑是否「妥適」時，又應以何種標準看待或使用科技手段產出的量刑參考資訊？無論原審是否在量刑時參考了量刑輔助系統所提供的資訊，上訴審是

⁹³ 司法院刑事案件量刑及定執行刑參考要點第27點：「本要點供法院裁量刑罰時參考，不影響審判權之獨立行使。」

⁹⁴ 儘管如此，文獻上仍有認為，如果法院實際上是「依靠」（而非只是「參考」）COMPAS系統的運作來進行判斷，即屬違反正當法律程序。請參見：鄭明政（2020），〈從State v. Loomis案件看AI應用於司法審判上的若干問題〉，《台日法政研究》，4期，頁165-178。

否可以以「根據量刑輔助系統提供的資訊，原審量刑屬於畸輕（或畸重）」為由，據以撤銷原審法院的量刑？

以上一連串的問題並不是單純的假設，而是我國法院已經遭遇的真實問題。在最高法院 108 年度台上字第 3728 號刑事判決中，被告和檢察官均對第一審法院的量刑感到不滿（被告認為過重，檢察官認為過輕），分別提起上訴；第二審法院即以「參酌司法院量刑資訊及量刑趨勢建議，原審就被告所犯傷害致死罪量處有期徒刑 10 年 6 月，核與同類型案件平均及建議刑度顯有過重之情，其罪刑並不相當，自非允當。」為由，撤銷了原判決，改判為較輕的有期徒刑 9 年 4 月。孰料被告主張：第二審法院的量刑，「仍重於上開系統建議之刑度區間即有期徒刑 7 年 1 月至有期徒刑 8 年 8 月」，因此上訴至第三審最高法院。而最高法院最終的判決，則是回歸到前述司法院的基本立場：量刑資訊系統僅有輔助量刑的作用，因此認為「司法院量刑資訊暨趨勢系統，僅供個案刑罰裁量之參考，並無拘束相關個案之法律上效力。」最終駁回了上訴。

本案中，最高法院的立場雖然與司法院對於人工智慧系統所持的立場看似一致，但卻沒有解決最基礎而重要的問題，即本文上述提出的「是否在某些情況或條件成就時，可以要求司法系統決策時，『應參考』人工智慧系統所提供的資訊呢？」畢竟本案中的第二審法院，正是以第一審法院的量刑判斷，與司法院的 2 個量刑輔助系統所產出的資訊不吻合為由，撤銷了第一審判決。既然第二審法院在判決中「明示」其參考了量刑輔助系統，如此一來，此項論述是否構成了對第二審法院本身的拘束、導致第二審法院「應」（如同本案被告所主張的）在量刑輔助系統提供的資訊區間內進行量刑呢？最高法院對此並沒有加以說明⁹⁵。

而在另一起案件中，我國的上訴審法院則明確的利用了量刑輔助系統，作為審查原審法院量刑是否妥適的標準。在臺灣高等法院 105 年度交上易字第 117 號刑事判決中，第一審法院對於第 4 次犯公共危險罪（酒後駕車）的被告，量處了較過去幾次犯罪時更重的刑罰。被告上訴至第二審法院請求

⁹⁵ 關於本案的詳細評釋，請參見：蘇凱平，前揭註 84，頁 85-94。

輕判。第二審法院認為：「〔量刑資訊系統所指出的〕相類似案件之量刑分布卻足可供為量刑時用以檢驗個案裁量結果是否確有畸重、畸輕；如有顯著乖離，法院則應盡其說理義務，並排除裁量結果係因非理性之主觀因素所致之量刑歧異，以此濟量刑裁量之審查標準抽象、審查密度偏低之窮，並昭折服。」簡言之，第二審法院認為，當原審法院對個案的量刑判斷，與量刑資訊系統所產出的「相類似案件之量刑分布資訊」間存在有顯著的差異時（本案判決稱之為「畸重、畸輕」之量刑），原審法院即負有一項特別的義務：必須說明為什麼做出了此種量刑判斷結果，以排除可能是因為「非理性之主觀因素」所造成的畸重或畸輕量刑。

本判決被選為 109 年臺灣高等法院的「具參考價值裁判」，對於事實審法院有相當的參考價值⁹⁶。而根據本判決的意旨，第一審法院在量刑時，必須要參考量刑輔助系統產出的資訊。審理法院就個案的量刑裁量結果，一旦與量刑輔助系統提供的資訊出現了明顯齟齬，而審理法院仍欲堅持自己對於個案量刑的判斷時，法院必須於判決中進行特別清楚、足以排除其裁量結果是來源於「非理性之主觀因素」的說理，否則其判決即有遭上訴審撤銷的可能。換言之，在這種情況下，**量刑資訊系統產出的資訊，就不再僅僅是單純作為參考、輔助的用途，而是實質上造成了事實審法院量刑時的某種限制。**

從以上 2 則判決中所顯示的法院觀點，不難看出所謂「量刑輔助系統提供的資訊僅供法院參考」的觀點，實際上有更多細緻的面向值得討論。**特別值得注意的是，不論是美國的 Loomis 判決、我國司法院的政策立場或上述最高法院的法律審判決，都僅進展到「量刑輔助系統提供的資訊僅供法院參考」的地步，但是我國的事實審法院（主要是有權直接改變量刑結果、且對於第一審量刑結果有審查權限的第二審法院）則傾向於給予量刑輔助系統更多的「用武之地」。**

本文將所觀察到的這種「事實審法院使用量刑輔助系統的方式，超出了司法院、最高法院所預定的系統任務範疇，成為實質上某種量刑判斷標準」的情況，稱之為「**量刑輔助系統的適用擴張現象**」。本文認為，會發生適用

⁹⁶ 關於本案的詳細評釋，請參見：蘇凱平，前揭註44，頁31-42。

擴張的現象，是因為當法院面對「如何量刑」的難題時，立法者固然為了因應個案的不同情況，在刑法等實體法的法定刑中賦予法院相當程度的裁量空間；但是這樣的裁量空間，卻也帶給法院在判決中具體說明量刑決定的困難。畢竟我國刑法對於量刑的要求，是要求法院必須「審酌一切情狀」⁹⁷。儘管刑法明定了 10 種法院「尤應注意」的事項，但是在這些一字排開的注意事項中，並沒有「權重」的概念，以致於不同法院之間對於類似案件의 量刑、甚至同一法院在不同時期對類似案件의 量刑，都可能有相當的出入。對於個別人類審判者運用其智慧裁量所做出的量刑決定而言，無論再怎麼謹慎、認真、注意每一個量刑因子，都無法避免因為時間（不同時期）和空間（不同法院、審判者）的隔閡，就情節類似的案件得出前後不一致的判斷結果，或做出與其他同樣也謹慎、認真、注意每一個量刑因子的法院不一致的判斷結果。

對於這種判斷不一致的結果，在過去有一種常見的解釋方式是「沒有 2 個案件是完全相同的」，由此得出「看起來極為相似的案件，量刑未必就要相似才屬正確」的結論。本文認為，雖然這種解釋方式並不是沒有道理，但是並不能解決司法系統實際上面對的問題。因為不可否認的，在大量的法院判決中，經常能觀察到一些主要事實看來相當近似、但判決結果卻差異甚大的情況⁹⁸。而當社會大眾注意到這種量刑結果的不一致時，刑事法院就容易招致廣泛的批評。此時法院仍然需要一個清楚、明確、一致的標準，得以向社會大眾說明量刑的具體標準所在，同時維持（各）法院對於類似案件量刑的最基本一致性⁹⁹。而**量刑輔助系統等科技手段的強項所在，正是可以呈現**

⁹⁷ 刑法第 57 條規定：「科刑時應以行為人之責任為基礎，並審酌一切情狀，尤應注意下列事項，為科刑輕重之標準：……。」

⁹⁸ 例如在案情通常相對單純的施用毒品和酒醉駕車等犯罪中，主要犯罪事實均高度相似，但正因為這些犯罪的案情相對單純、法定刑偏低且裁量空間也相對狹窄，法院在判決中通常不會對於刑度如何裁量多做說明，更容易造成社會大眾的不滿。請參見：呂寧莉（2020），〈臺灣高等法院 105 年度交上易字第 117 號刑事判決評析：兼論使用「量刑資訊系統」之相關問題〉，《台灣法學雜誌》，393 期，頁 55。

⁹⁹ 從國民參與審判的角度切入而結論相同的觀點，請參見：謝煜偉，前揭註 43，頁

出一個更大的圖像(過往的整體量刑情況),並維持這個圖像的各個元素(各個具體個案的量刑情況)之間的一致性。事實審法院因此願意給予使用科技手段產出的資訊,有更大的揮灑空間,以形成一致化的標準,就是自然而然、順應實際需要而必然發生的情況。在我國的刑事審判系統中,第二審法院職司第二次的事實審(覆審制),對於第一審法院的量刑妥適與否,本有審查的義務,當然更需要形成明確的審查標準,以應每日審判工作的需要。因為這種實際需要的存在,事實審法院願意賦予量刑輔助系統等科技手段更大的實質上影響力,是自然且可以理解的現象。

三、適用擴張的違憲疑慮？

那麼,這種適用擴張現象,導致原本應「僅供參考」的量刑輔助系統提供資訊,實質上對於事實審量刑結果有一定的影響力,也就是在某些狀況下,事實審法院必須以量刑輔助系統提供的資訊,衡量自己對於所審理個案的量刑決定,是否符合量刑輔助系統所提供的過往類似案件之量刑區間。此種適用擴張現象是否可能違反憲法對於法院量刑的相關要求或人民權利保護呢?具體而言,此處主要涉及的是 2 項憲法規範:第一,上級審法院以過往類似案件之量刑結果,亦即「本案以外資訊」此種統計證據,判斷原審量刑是否適法,是否即屬違反憲法第 80 條要求法官應「依據法律獨立審判,不受任何干涉」之規定?第二,事實審法院在量刑時,若考慮了本案以外的量刑資訊,是否即屬有違「憲法上罪刑相當原則」?本文認為,就目前刑事法院運用第一代量刑輔助系統而形成判決的實際運作方式觀之,以上 2 個「是否違憲」問題的答案,都是否定的。其原因如下:

第一,事實審法院上述擴張使用量刑輔助系統提供資訊的作法,仍在法律規定容許的範疇之內。對於事實審法院應如何量刑的法律規定,主要是刑法第 57 條的「科刑時應以行為人之責任為基礎,並審酌一切情狀,尤應注意下列事項,為科刑輕重之標準:……。」雖然本條列出的 10 款事項,如犯罪之動機、目的;犯罪時所受之刺激;犯罪之手段;犯罪行為人之生活情

況等等，全部都是「本案」犯罪的事由，並不涉及過往類似案件的資訊（亦即統計證據），但是本條也明確指出這 10 款事由僅是法院量刑時「尤應注意」的事項，而非「僅能注意」的事項。反之，上開法條明確指出法院科刑時「應以行為人之責任為基礎，並審酌一切情狀」。

既然我國法律容許法院「審酌一切情狀」而為量刑，那麼事實審法院若選擇將量刑輔助系統提供的類似案件量刑結果（統計證據）考量在內，仍屬於符合法律規定的範疇。而若上級審法院因為考慮到了過往類似案件量刑結果（例如考慮量刑輔助系統提供的資訊），因而認為原審法院的量刑結果出現了畸輕或畸重的不合理情況，但原審判決中卻沒有相應的合理說明，這也屬於上級審法院依據刑法第 57 條要求原審法院應「審酌一切情狀」而來的合理要求，並不違背法官應「依據法律獨立審判」的憲法規範¹⁰⁰。

第二，在「法院、檢察官與被告三方均有使用相同量刑輔助系統之權限」的前提下，法院不會因為在科刑時，將量刑輔助系統提供的他案資訊納入考量，即違背憲法上的罪刑相當原則。按，國家所施加之刑罰須與行為人之罪責相當，刑罰不得超過罪責，此即我國憲法解釋所承認之「憲法上罪刑相當原則」¹⁰¹。而司法院釋字第 775 號解釋指出，依據憲法上罪刑相當原則，法院必須使當事人有機會對於科刑資料指出證明方法、對該科刑資料進行周詳調查，並使雙方當事人就科刑資料進行充分辯論，此外法院還必須具體說明據以量定個案中刑罰之理由，法院所為科刑判決始滿足憲法上的罪刑相當原

¹⁰⁰ 相同結論，但從憲法平等原則與最高法院「量刑內部性界限」觀點論述者，可參照蘇凱平，前揭註 84，頁 88-92。

¹⁰¹ 例如司法院釋字第 775 號解釋理由書指出：「又有關刑罰法律，基於無責任無處罰之憲法原則，人民僅因自己之刑事違法且有責行為而受刑事處罰（本院釋字第 687 號解釋參照）。**刑罰須以罪責為基礎，並受罪責原則之拘束，無罪責即無刑罰，刑罰須與罪責相對應（本院釋字第 551 號及第 669 號解釋參照）。亦即國家所施加之刑罰須與行為人之罪責相當，刑罰不得超過罪責。基於憲法罪刑相當原則（本院釋字第 602 號、第 630 號、第 662 號、第 669 號及第 679 號解釋參照），立法機關衡量其所欲維護法益之重要性、防止侵害之可能性及事後矯正行為人之必要性，綜合斟酌各項情狀，以法律規定法官所得科處之刑罰種類及其上下限，應與該犯罪行為所生之危害、行為人責任之輕重相符，始與憲法罪刑相當原則及憲法第 23 條比例原則無違。」**

則之要求¹⁰²。我國刑事訴訟法因此於 2020 年修正，將第 289 條原先僅給予「當事人就科刑範圍表示意見之機會」而無辯論之規定，修正為檢察官、被告及辯護人應就科刑範圍進行辯論，以符合司法院釋字第 775 號解釋的要求¹⁰³。

根據上述大法官解釋對於憲法上罪刑相當原則的闡釋，本文認為，法院為科刑判決時，是否選擇考慮量刑輔助系統提供的他案資訊，固然由各審理法院自行決定；可是一旦審理法院選擇納入量刑輔助系統提供的他案資訊，作為本案科刑考慮的因素，則法院有義務使雙方當事人有合理的途徑，知悉法院可能考慮的量刑因子，以及這些量刑因子在法院所使用的量刑輔助系統上可能產出的結果。尤其是刑事被告，身為國家施加刑罰的對象，如果對於法院科刑有實質影響的資訊，並非被告可合理得知者，將使被告無從對此項資訊表達意見，更遑論進行有意義的科刑辯論。如此一來，顯然將違背上述司法院釋字第 775 號解釋賦予法院應使當事人就科刑資料「指出證明方法，進行周詳調查與充分辯論」之義務，法院也無從於判決中「具體說明據以量定刑罰之理由」，而將使刑事訴訟法第 289 條增訂科刑範圍辯論之規定所欲達成的「精緻、妥適」量刑，成為空談¹⁰⁴。

根據此一標準，本文認為，司法院建置的第一代量刑輔助系統，包括量刑資訊系統和量刑趨勢建議系統的呈現結果，以目前實務上運用的情況觀之，尚足以使當事人對於法院所可能考慮之量刑因子有所瞭解，並據以進行

¹⁰² 司法院釋字第 775 號解釋理由書：「為使法院科刑判決符合憲法上罪刑相當原則，法院審判時應先由當事人就加重、減輕或免除其刑等事實（刑法第 47 條第 1 項及第 59 條至第 62 條參照）及其他科刑資料（刑法第 57 條及第 58 條參照），指出證明方法，進行周詳調查與充分辯論，最後由法院依法詳加斟酌取舍，並具體說明據以量定刑罰之理由，俾作出符合憲法罪刑相當原則之科刑判決。」

¹⁰³ 2020 年 1 月 15 日刑事訴訟法第 289 條修正理由（摘錄）：「按犯罪事實有無之認定，與應如何科刑，均同等重要，其影響被告之權益甚鉅，原條文第 3 項僅給予當事人就科刑範圍表示意見之機會，而未經辯論，尚有未足，爰依司法院釋字第 775 號解釋意旨，將原條文第 3 項移列第 2 項，明定當事人、辯護人就事實及法律辯論後，應依第 1 項所定次序，就科刑範圍辯論之，俾使量刑更加精緻、妥適」。

¹⁰⁴ 司法院釋字第 775 號解釋理由書，前揭註 102。2020 年 1 月 15 日刑事訴訟法第 289 條修正理由，前揭註 103。

有意義的科刑範圍辯論，因此不致於違反憲法上的罪刑相當原則。因為如同本文以上對於第一代量刑輔助系統的介紹，司法院清楚的公布其系統建置原理，並且使包括訴訟當事人在內的社會大眾，均有相同的暢通途徑可以使用量刑輔助系統，從而知道選擇採納量刑輔助系統提供之資訊為科刑判決基礎的法院可能考慮的量刑因子為何。

例如在下圖【圖四】的「司法院殺人案件量刑資訊系統」中，使用者可以勾選各種的「犯罪方法」，例如毒殺、溺水、放火、使墜落、以熱水燙、電擊、漏逸氣體等等，多達 15 種。既然知道審理殺人案的法院科刑時，可能參考本系統提供之資訊，訴訟當事人在科刑範圍辯論的階段，可以針對本案事實中出現的量刑因子進行辯論，也可以在輸入本案的量刑因子後，瞭解法院可能參考的量刑區間為何。而在下圖【圖五】的「量刑趨勢建議系統」對於刑法第 328 條第 1 項強盜罪案件建置的使用介面中，使用者可以瞭解就「犯罪之動機」、「犯罪之手段」、「犯罪行為人之生活狀況」、「犯罪行為人之品行」等刑法第 57 條規定的法院科刑時應注意事項，法院如何具體考慮的方式，進而就這些事項的內涵表示意見，或進行科刑範圍辯論。凡此，均可以使訴訟當事人盡可能與法院站在資訊平等的立場上，針對科刑判決涉及的事實與詮釋方式，進行有意義的意見表達和辯論，從而達成上開憲法解釋與刑事訴訟法規定所要求的目標：符合憲法上罪刑相當原則，並達成精緻、妥適的量刑¹⁰⁵。

¹⁰⁵ 不過，本文以上的觀點，只是認為第一代量刑輔助系統的建置原理與目前實務上的運用方式，並無違憲之虞，並不代表本文認為第一代量刑輔助系統提供的資訊是正確無誤的。因為當涉及事實因子的擷取和詮釋時，系統建置者本應以符合實證研究原理的方法作為基礎，其中包括：要能確認負責從判決中擷取、詮釋和歸類事實因子的工作人員經過充分的訓練，從而能夠在閱讀判決與擷取、詮釋和歸類（亦即進行標註）其中的事實因子時，採取相同的標準；也要能夠有排除錯誤（包括排除單純出於工作人員疏失的錯誤，以及排除實際上不同工作人員採取了不同標註標準的錯誤）的系統化機制。而由於第一代量刑輔助系統的量刑因子標註，是建立在大量司法行政替代役人員在服役期間的人力支援上。考慮到經年累月下服役人員的不斷更替，是否能確認參與量刑因子標註作業的人員均經過相當程度的實證研究訓練、具有以相同標準進行標註的判斷能力，恐怕值得商榷。

【圖四】司法院殺人案件量刑資訊系統使用介面

※ 資料來源：司法院網站，

https://sen.judicial.gov.tw/pub_kill_sbin/kill_chkid_Project3.cgi

（最後瀏覽日：05/15/2024）。

量刑趨勢建議系統

首頁 妨害性自主案件 殺人與傷害致死案件 搶奪案件 強盜案件 電信詐欺案件 搶劫案件 不能安全駕駛案件 竊盜案件

【刑法第328條第1項 強盜罪】
法定刑：處5年以上有期徒刑

犯罪所得金額：1萬元以下

犯罪之動機
為其他不法目的（搶奪犯它罪工具）
☒ 否 ☐ 是

犯罪之手段
強盜時間一夜間
☒ 否 ☐ 是
有計畫犯罪（例如：冒充刑警、製造假車禍、遮掩面容或破壞監視器）
☒ 否 ☐ 是
使用兇器以外之器械（例如：以繩索纏綁被害人、持不具殺傷力之

建議刑度
5年 4月

建議量刑區間
5年 2月 至 5年 6月

以上量刑結果乃依據您所輸入的選項：

- 犯罪之動機
 - 為其他不法目的（搶奪犯它罪工具）：☒ 否
- 犯罪之手段
 - 強盜時間一夜間：☒ 否
 - 有計畫犯罪（例如：冒充刑警、製造假車禍、遮掩面容或破壞監視器）：☒ 否
 - 使用兇器以外之器械（例如：以繩索纏

【圖五】司法院量刑趨勢建議系統，刑法第 328 條第 1 項強盜罪使用介面

※ 資料來源：司法院網站，

https://sen.judicial.gov.tw/pub_platform/sugg/328I.html

（最後瀏覽日：05/15/2024）。

伍、以人工智慧輔助量刑之二：司法院第二代量刑輔助系統

一、技術原理與社會顧慮

（一）技術原理

司法院的第一代量刑輔助系統，使用的是（大）數據分析的方法，並建立了明確的各項量刑因子標籤。然而，一方面基於科技的進步，人工智慧技術，特別是如同美國 COMPAS 系統的機器學習方法，在 21 世紀的第二個 10 年中逐漸成為顯學¹⁰⁶；另一方面則是因為現實的考量：原本建置第一代量刑輔助系統時，司法院有大量的司法行政替代役人員可協助閱讀既有判決的量刑，並透過標註方式提取指定的量刑因子。然因兵役制度改革，司法行政替代役停止招募，因此司法院欠缺足夠的人力可以更新第一代的量刑輔助系統。在科技供給與現實需求的雙重作用下，司法院產生了以「電腦算力」取代「人力」的工智慧技術，建立第二代輔助量刑系統的想法¹⁰⁷。

根據司法院官方網站中的「量刑專區」專頁，以及發布的新聞稿指出：第二代量刑輔助系統是自 2020 年 9 月起開發，並於 2023 年 2 月啟用。具體包括 2 種類型的系統，分別是「事實型 AI 輔助量刑資訊系統」與「評價型 AI 輔助量刑資訊系統」（以下簡稱為「事實型 AI 系統」與「評價型 AI 系統」）。事實型 AI 系統包括不能安全駕駛、肇事逃逸、傷害、竊盜與詐欺等 5 種案件類型，評價型 AI 系統則包括槍砲彈藥、妨害性自主與毒品等 3 種案件類型。雖然司法院在「量刑專區」專頁中，稱這 2 類的 AI 輔助量刑系統已於 2023 年 2 月 6 日「全面啟用」，但根據 2023 年 2 月 10 日出刊的司法周刊報導：「AI 量刑資訊系統目前僅供法官使用，未來將陸續開放予

¹⁰⁶ 李開復、王詠剛（2017），《人工智慧來了》，頁92-102，遠見天下文化。

¹⁰⁷ 司法院官方網站（刑事量刑專區），<https://www.judicial.gov.tw/tw/cp-83-57186-1ef46-1.html>（最後瀏覽日：05/15/2024）。。

檢察官、律師及一般民眾使用」¹⁰⁸。而司法院分別於 2023 年 5 月和 9 月，先後開放了事實型 AI 系統與評價型 AI 系統予檢察官和律師使用¹⁰⁹。律師且必須經由「司法院線上起訴暨律師單一登入」系統，以註冊於「電子筆錄調閱服務」的帳號登入使用¹¹⁰。換言之，至本文截稿（2023 年 11 月底）為止，第二代的量刑輔助系統尚未開放刑事被告和一般民眾使用¹¹¹。

而在 2023 年 8 月，司法院進一步發布新聞稿稱：在事實型 AI 輔助量刑資訊系統中的「不能安全駕駛」、「幫助詐欺」2 類犯罪量刑資訊的基礎上，司法院於 2022 年 4 月開始發展生成式 AI 應用，目標是「以 AI 來撰寫上開兩類犯罪的裁判草稿，以供法官製作裁判時參考」¹¹²。此一 AI 裁判書草擬系統正由部分法院試辦評估中，不過相較於上述司法院新聞稿中對於第二代量刑資訊輔助系統的說明中，提及未來將陸續開放予檢察官、律師與一般大眾使用，AI 裁判書草擬系統似乎專屬於法官，並未預計開放給檢察官、律師或社會大眾使用。

¹⁰⁸ 司法院（2023），〈因應國民法官新制施行 司法院啟用AI量刑系統〉，《司法周刊》，2143期，頁4。

¹⁰⁹ 全國律師聯合會網站，〈司法院開放「事實型AI輔助量刑資訊系統」（含不能安全駕駛、肇事逃逸、傷害、竊盜及詐欺案件）予律師使用，敬請會員善加利用〉，<https://www.twba.org.tw/news/2e477381-4823-4bb0-acc1-0cde8aed9e9d>（最後瀏覽日：05/15/2024）；全國律師聯合會網站，〈司法院開放「評價型AI輔助量刑資訊系統」（含槍砲彈藥、妨害性自主及毒品案件）予律師使用，敬請會員多加利用〉，<https://www.twba.org.tw/news/82331621-4f37-406a-b220-635adef7b54>（最後瀏覽日：05/17/2024）。

¹¹⁰ 司法院，線上起訴暨律師單一登入網站，<https://portal.ezlawyer.com.tw/LOGIN.jsp>（最後瀏覽日：05/15/2024）。

¹¹¹ 本文於2023年11月底完成並投稿本期刊進行審查，至2024年3月接獲論文經雙重匿名審查通過之通知。而司法院於2024年11月27日發布新聞稿指出，於該日開始開放事實型量刑資訊系統供民眾使用。司法院（2024），〈司法院開放事實型量刑資訊系統供民眾使用 回應各界對量刑妥適性之期待〉，<https://www.judicial.gov.tw/tw/cp-1887-1207000-1faaf-1.html>（最後瀏覽日：11/27/2024）。而至2025年8月底本文編輯作業時，「評價型量刑資訊系統」仍未開放予民眾使用。因此本文以下「伍、三、我國法制脈絡下必須面對的問題、（一）由政府機關推動人工智慧量刑系統的疑慮」節所述的問題，在我國依然存在。

¹¹² 司法院，前揭註80。

根據司法院的說明，運用人工智慧技術的第二代量刑輔助系統，整體而言是針對既有的（第一代）量刑資訊系統進行改善，藉由人工智慧的自然語言處理技術，在判決文字中標註相關量刑事由，並自動擷取關鍵字，搜尋及判讀加重、減輕法條的適用；並以此標註過程作為機器學習的訓練資料，取代原本量刑資訊系統以人工進行標註的方式，達到即時更新資料庫的效果¹¹³。

司法院指出，「事實型 AI 輔助量刑資訊系統」的技術原理沿襲了第一代量刑輔助系統中「量刑資訊系統」之選項設計，要求使用者自行輸入個案之事實內容（即「量刑因子」），以搜尋過往法院類似判決之量刑情形，並在此基礎上加入 AI 標註判決書之技術。司法院認為，此一系統因為使用者輸入之條件較多，有助於搜尋出與個案更為相近之過往判決作為量刑參考¹¹⁴。在具體操作上，「事實型 AI 輔助量刑資訊系統」要求使用者在不能安全駕駛、肇事逃逸、傷害、竊盜與詐欺等 5 種案件類型中，勾選該案件類型所欲查詢的具體法律條文（以肇事逃逸案件為例，包括刑法第 185 條之 4 第 1 項前段的致人傷害而逃逸，以及同條項後段的致人於死或重傷而逃逸等法律條文），並輸入判決年度、判決法院、緩刑、加重減輕、犯罪情節、犯罪動機、前案紀錄與犯後態度等「量刑因子」後，系統將輸出一個「統計結果」頁面，呈現出符合使用者輸入之量刑因子的各種類別刑罰（如有期徒刑、拘役、罰金、有期徒刑併科罰金、拘役併科罰金等）之件數、這些件數的平均刑度與最高、最低刑度，以及各種類別刑罰的比例圓餅圖。使用者還可以進一步點選系統列出的「件數」，進入「裁判書清單」頁面，檢視過往具有類似量刑因子的判決內容。

而「評價型 AI 輔助量刑資訊系統」的技術原理，則與第一代的量刑資訊系統有較大區別。使用者在選取槍砲彈藥、妨害性自主與毒品案件中的具

¹¹³ 司法院（02/06/2023），〈因應國民法官新制，司法院啟用AI量刑資訊系統—具備二種模式、擁有四大優點〉，<https://www.judicial.gov.tw/tw/cp-1887-806741-d6471-1.html>（最後瀏覽日：05/15/2024）。

¹¹⁴ 司法院官方網站（刑事量刑專區），<https://www.judicial.gov.tw/tw/cp-83-57186-1ef46-1.html>（最後瀏覽日：05/15/2024）。

體犯罪條文後，依序勾選是否考慮普通刑法和特別刑法上的法定刑加重與減輕事由（例如刑法 47 條累犯、刑法第 59 條情堪憫恕、刑法第 62 條自首、槍砲彈藥刀械管制條例第 18 條的自首或自白等規定），並於刑法第 57 條所列的量刑審酌事項（犯罪後之態度、犯罪所生之危險或損害、犯罪行為人之品行等）中，分別勾選使用者認為個案中是對被告有利、不利或「中性或不明確」，或者暫時不考慮某項量刑審酌事項，由此檢索過往實務相關判決之量刑情形。根據司法院指出，「評價型 AI 輔助量刑資訊系統」有「輸入少數條件，即可取得有效統計分布之優勢」。¹¹⁵

（二）社會顧慮

司法院的第二代量刑輔助系統自 2023 年 2 月啟用後，並沒有立即引起社會大眾的關注。原因可能在於社會大眾根本無法使用這個系統。如同前文所介紹，一直到了 2023 年 9 月，第二代的量刑輔助系統才完整開放給法官、檢察官和律師使用，而截至本文完稿的 2023 年 11 月底，都尚未開放給一般刑事被告與社會大眾使用¹¹⁶。

不過，司法院在推動另一項與第二代量刑輔助系統相關的數位政策時，卻引起了民間團體的重視。司法院在 2023 年 8 月發布的新聞稿中指出：在事實型 AI 系統中的「不能安全駕駛」、「幫助詐欺」2 類犯罪量刑資訊的基礎上，司法院於 2022 年 4 月開始發展「生成式 AI 應用」，希望以 AI 來撰寫上開 2 類犯罪的裁判草稿，上述 2 種犯罪類型的 AI 裁判書草稿生成系統即將由數家法院進行試辦；並預計於 2023 年 12 月試辦以 AI 撰寫毒品犯罪的裁判書草稿、2024 年起規劃開發民事交通損害賠償、消費者債務清理更生與清算等民事事件的裁定草稿¹¹⁷。

¹¹⁵ 司法院官方網站（刑事量刑專區），<https://www.judicial.gov.tw/tw/cp-83-57186-1ef46-1.html>（最後瀏覽日：05/15/2024）。

¹¹⁶ 請參見本文「伍、一、技術原理與社會顧慮」章節的說明。

¹¹⁷ 司法院，前揭註 80。

本則新聞稿引起了包括台灣人權促進會、民間司法改革基金會等司法人權團體，以及全國律師聯合會等律師公會團體的關注。基於在系統資訊揭露和風險評估等面向上極大的疑慮，這些團體共同發表聲明並召開記者會，指出對於 AI 草擬判決的「三大疑問與三大風險」¹¹⁸。在共同聲明稿中，民間團體指出，司法院應達成「資訊充分揭露」、「加強風險評估」和「強化社會溝通」等 3 項目標。其中，就「資訊充分揭露」，要求司法院應說明所採用的 AI 生成系統，其建置與監督訓練的範圍與方式；「加強風險評估」則是要求司法院應審慎評估 AI 生成系統使用方式，避免判斷錯誤、複製偏誤及司法信任風險；「強化社會溝通」則是要求司法院應先建立符合國際標準的 AI 使用規範後，才可以開放給法官使用，不可以「先上線後補票」¹¹⁹。

司法政策機關在推動涉及人工智慧的司法數位化政策時，所涉及的議題極多，例如上述民間團體提到的「符合國際標準的 AI 使用規範」其內容與意涵為何；政府機關在推動數位化政策時，如何兼顧政府資訊安全與個人資料保護；技術上如何辨識出訓練資料集中具有歧視性的資料並予以排除等等。本文則希望專注在所設定的主題「以人工智慧輔佐法院心證」範疇內進行探討。因迄於本文完成時，司法院推動的第二代量刑輔助系統尚未開放社會大眾使用，本文僅能就司法院對外發布的歷次新聞稿等官方訊息作為評論的依據。而台灣人權促進會等團體所發布的上述聲明稿，雖然是就司法院 AI 裁判書草稿生成系統而發，但細讀其聲明內容可知，幾乎都是針對比 AI 裁判書草稿生成系統更基礎的數位化政策、亦即 AI 量刑系統而發¹²⁰。因此本

¹¹⁸ 台灣人權促進會等團體共同聲明新聞稿（09/26/2023），〈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〉，<https://www.tahr.org.tw/news/3449>（最後瀏覽日：05/15/2024）。本聲明稿由全國律師聯合會、台北律師公會、臺中律師公會、開放文化基金會（OCF）、台灣人權促進會與民間司法改革基金會等團體共同發出。

¹¹⁹ 台灣人權促進會等團體共同聲明新聞稿，前揭註118。

¹²⁰ 例如聲明中指出司法院開發的系統，將產生「判斷錯誤」、「複製系統偏誤」與「司法信任」等風險，顯然並非針對判決書草稿如何製作而言，而是針對案件內涵的實際判斷是否受AI系統輸出結果之影響而發。此觀諸聲明中亦舉美國COMPAS系統為例，即屬自明。請參見台灣人權促進會等團體共同聲明稿，前揭

文以下的評析，也將引用上述聲明稿中民間團體對於司法院 AI 量刑系統的質疑與呼籲內容。

根據司法院發布的官方訊息，本文認為，司法院的第二代量刑輔助系統，由於與美國 COMPAS 系統同樣採取機器學習的人工智慧方式建置，除了勢必要面對與 COMPAS 系統相同的質疑之外，司法院還必須另外處理兩項在我國法制脈絡下存在的重要問題，分別是「由政府機關推動人工智慧量刑系統的疑慮」與「被告無法使用系統造成的憲法權利侵害」，而後者問題直接關係到憲法保障一般人民與刑事被告之基本權利，有亟待解決的急迫性。以下分別詳述之。

二、機器學習方法的共通本質問題

根據司法院指出，第二代量刑輔助系統「係藉由人工智慧的自然語言處理技術，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，作為機器學習的訓練資料，取代人工標註的方式，達到即時更新資料庫的效果。¹²¹」由技術特性觀之，此一系統與美國 COMPAS 系統相同，均屬於以「統計學習」為基礎的第二波人工智慧成果。在運用大數據與機器學習的技術下，能夠將資訊進行精微分類，也具有基本的推理能力，但是不具有「脈絡適應」的能力¹²²。由於基礎的技術原理相同，本文第貳章所分析的美國 COMPAS 系統在 Loomis 案判決中遭到質疑的面向，包括系統未能提供準確資訊、系統使用的並非被告個人資料等，均可能成為對於司法院第二代量刑輔助系統的質疑。而迄今為止，根據司法院所公開的資訊，確實沒有解釋第二代量刑輔助系統是「如何使用」相關資訊，而得出最後的結果。具體而言，系統如何在過往判決文字中標註量刑事由？如何將標註結果進行分類？採行何種機器學習方法？在此種方法下，人工智慧系統具體如何運作？這些重要的資訊，

註118。

¹²¹ 司法院，前揭註108。

¹²² 請參見本文第貳章「貳、人工智慧的發展進程」章節之說明。

新聞稿中並沒有提供。無論是實際使用此一系統的專業法律人、刑事被告或社會大眾，均不得而知。

以「系統如何在過往判決文字中標註量刑事由」此一基礎重要問題為例，司法院新聞稿指出，「評價型 AI 輔助量刑資訊系統」乃是「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、對被告不利或中性的量刑因子，並進行歸納整理，使用者可以用個案的量刑情狀進行查詢。¹²³」具體的畫面操作圖示如下圖【圖六】：

■ 刑法第57條之量刑審酌事項

	暫不考慮	對被告有利	中性或不明確	對被告不利
犯罪後之態度	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
犯罪手段與所生之損害	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
犯罪行為人之品行	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
被害人的態度	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
其他審酌事項 ①	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

【圖六】評價型 AI 輔助量刑資訊系統中「刑法第 57 條之量刑審酌事項」之操作介面

※ 資料來源：司法院，前揭註 113。

然而，刑法第 57 條共列出 10 款法院科刑時應作為科刑輕重標準之注意事項，但根據上圖，評價型 AI 系統僅列出了 4 種具體的注意事項讓使用者勾選，其餘各種情況均合併為「其他審酌事項」。換言之，此一 AI 系統在建置與運作時，係將法律規定之條款「自行歸納」為各種類別，但其理由與正當性何在？為什麼 AI 系統選擇將刑法第 57 條中明列為法院科刑時「尤應注意」的「犯罪之動機、目的」（第 1 款）、「犯罪時所受之刺激」（第 2 款）、「犯罪行為人之生活狀況」（第 4 款）、「犯罪行為人之智識程度」（第 6 款）、「犯罪行為人與被害人之關係」（第 7 款）與「犯罪行為人違反義務之程度」（第 8 款）等事項，全部「打包」合併為一項「其他審酌事項」來讓使用者勾選？上述各款事由對於量刑的影響力，顯然將因為被「打

¹²³ 司法院，前揭註 113。

包」而遭到稀釋，系統此種運作方式是否仍屬符合法律規範？上述被「打包」的各款事由對於被告有利、不利的判斷，顯然也不會完全一致，使用者又應如何勾選？

而若細查 AI 系統列出的 4 種具體可勾選的注意事項，也有令人不解之處。例如為何系統將刑法第 57 條法律規定中分列的「犯罪之手段」（第 3 款）與「犯罪所生之危險或損害」（第 9 款），合併稱為「犯罪手段與所生之損害」，要求使用者進行判斷？如果這 2 款對被告的有利、不利判斷不同（例如犯罪手段並不特別兇殘，但是所生損害極大），使用者又應如何勾選？更遑論上圖中列出的「被害人之態度」，根本未列在刑法第 57 條規定之中。因此似乎是系統自行創設、並無法律依據的選項，又是從何而來？

再就機器學習的人工智慧方法所必須進行的訓練資料標註而言，評價型 AI 系統建置時，必須先將作為訓練資料的同類型過往判決文字進行標註，標註類別即為上圖【圖六】所示的「暫不考慮」、「對被告有利」、「中性或不明確」與「對被告不利」等 4 類評價。**問題是，在司法院目前已公開的資料中，並未說明系統建置過程中，究竟是如何將過往判決文字分類標註為以上 4 類評價者。**例如在槍砲彈藥案件中，常見的科刑判決文字可能為「兼衡被告自陳大專畢業，目前無業，經濟狀況普通，有未成年子女需扶養，並提出 111 年度綜合所得稅各類所得資料清單及戶籍謄本各 1 份為證等一切情狀，量處如主文所示之刑¹²⁴」；或「**併考量被告本案非法持有槍枝及子彈之數量**，兼衡被告自述於案發時間前罹有身心疾病之情形，暨被告於本院審理程序中自述國中肄業之智識程度，現無業、須扶養女兒及母親之家庭經濟情況等一切情狀，分別量處如主文所示之刑¹²⁵」；又或「**考量其持有槍彈之數量及期間**、前科素行，兼衡其犯罪動機、目的及手段，暨其自述國中畢業、

¹²⁴ 臺灣臺北地方法院 112 年度訴字第 478 號刑事判決，其中引用法院卷宗頁數部分從略。

¹²⁵ 臺灣臺北地方法院 111 年度訴字第 745 號刑事判決，其中引用法院卷宗頁數部分從略。

從事酒店工作、家庭經濟狀況屬低收入戶但未申請、需扶養父親等一切情狀，量處如主文所示之刑¹²⁶。」

上述 3 則判決皆在科刑理由中提及被告的學歷（分別為國中肄業、國中畢業和大專畢業），顯見學歷對於科刑有所影響，但究竟是何種方向的影響呢？是被告學歷越高，應該判刑越重？還是被告學歷越低，應該判刑越重？歷來法院判決中，經常僅指出被告學歷為科刑考量事由之一，但並未具體「表態」學歷對於量刑造成的影響方向（加重或減輕），此時 AI 系統應如何標註哪一種學歷屬於「對被告有利」、「對被告不利」還是「中性或不明確」？又如「被告持有槍彈之數量」越多時，判刑固然應該越重，但是「多」與「少」是相對的概念，如上述科刑判決內容均僅籠統指出：法院科刑時會「考量被告持有槍彈之數量」，但未具體指出法院考量被告持有特定槍彈數量的結果，究竟是對於被告有利或不利，系統又要如何進行標註呢？

本文以上指出的問題，有部分固屬 AI 系統建置時對於輸出項（output）考量的問題，例如將刑法第 57 條明定之量刑審酌事項，片面變更其類型或內容讓使用者勾選；有部分則為我國法院科刑判決中向來存在的問題，例如僅提及事實（學歷如何、持有槍彈若干），卻沒有說明該項事實對於法院量刑的影響（某種學歷、或持有某種數量的槍彈，究竟會加重或減輕刑罰）。然不論以上何者，都會對機器學習的人工智慧方法所必經的「標註」程序造成實作上的困難，也因此必將帶來使用者與社會大眾對於系統準確性、正當性甚至適法性的質疑。建置 AI 系統的司法院，如果難以說明系統的標註原理以及系統運作的演算法，很容易被認為標註有所缺失，不符合法律規定或實務運作方式。如此一來，又如何能期待建立在此標註基礎上的 AI 系統能夠運算出準確的結果？

說穿了，無法對於機器學習的 AI 系統運作原理「說清楚、講明白」，是司法院第二代量刑輔助系統與美國 COMPAS 系統所共同面對的「痛點」：未能提供準確資訊。差別僅在於美國 COMPAS 系統公布了風險評估的公式，

¹²⁶ 臺灣臺北地方法院 112 年度訴字第 975 號刑事判決，其中引用法院卷宗頁數部分從略。

因此在無法清楚說明公式的各項「指標」（即犯罪者特徵）與這些指標乘上的「權重」分數由何而來時，受到嚴厲的質疑¹²⁷。我國則是因為法律中明定了法院科刑時「尤應注意」的審酌事項，因此在 AI 系統如何從判決文字中對這些法定事項進行標註、歸納與分類上，無論是在理論或實務運作上，都必然會以法律規定為標準，受到嚴格的檢視。

三、我國法制脈絡下必須面對的問題

（一）由政府機關推動人工智慧量刑系統的疑慮

雖然同樣因為採取機器學習方法而受到質疑和挑戰，司法建置的第二代量刑輔助系統也有其獨有的問題需要面對。**首先就是司法以國家政策機關的角色，建置並推動 AI 量刑系統，其負有的說明技術原理的責任，理應比私人企業更大、更完整。**美國 COMPAS 系統的開發商，以「商業保密」為由，僅在最小程度上揭示其主要模型使用了「迴歸建模」和「機器學習」的方法，並未揭露完整的演算法。也因此，在本文前述的美國威州最高法院 Loomis 案中，被告質疑科刑法院使用 COMPAS 系統違反憲法對於刑事被告的正當程序保障時，威州最高法院基本上認可了 COMPAS 開發商的「商業保密」說法，認為只要開發商提供的使用手冊解釋了風險分數評估的項目包括哪些，就不算是違反憲法正當程序保障¹²⁸。

然而，司法的情況則不同。**司法本身是政府機關，開發建置 AI 量刑輔助系統完全是為了幫助法院形成更理想的科刑判決，全然屬於公益考量。**在此意義下，很難想像司法有對於其用於建置 AI 系統的演算法進行保密之需要或正當理由。相反的，本文認為，司法在 AI 量刑系統中所使用的演算法，應屬於政府資訊公開法第 7 條第 1 項第 2 款所規定之「政府機關為協助下級機關或屬官統一解釋法令、認定事實、及行使裁量權，而訂

¹²⁷ 詳見本文「參、一、技術原理與應用情況」章節之說明。

¹²⁸ 詳見本文「參、二、Loomis案判決理由分析」章節之說明。

頒之解釋性規定及裁量基準。」而應主動公開，以增進人民對於法院行使刑罰裁量權所可能參考之基準的瞭解、信賴及監督¹²⁹。

司法院固然強調：AI 量刑系統（以及以其為基礎的 AI 判決書草稿生成系統）的產出結果均「僅供參考」，居於審判核心地位的認定事實、適用法律與決定量刑等事項，仍由法官自行決定¹³⁰。惟既然此一 AI 量刑系統乃由最高司法政策機關費心建置；且司法院也認為此系統的產出結果，對於法官如何決定審判核心事項有參考之價值；再加以目前實務上也確實有法院運用司法院建置的（第一代）量刑輔助系統之產出結果，作為判斷法院量刑是否適當之依據的實例，顯見司法院建置的量刑輔助系統對於法院決定如何量刑或判斷量刑是否得當，有相當程度的影響力¹³¹。在此事實基礎上，很難認為司法院對於 AI 量刑系統之技術原理，沒有充分、完整說明的義務。

直言之，在對於 AI 系統的演算法內涵難以表達清楚、或甚至開發者因故有意避談時，私人公司開發的商業系統，尚有可能以「商業保密」的擋箭牌，將其說明義務限制在一定範圍內；但是當 AI 量刑輔助系統的開發者是政府機關，例如我國司法院時，並沒有這一塊擋箭牌可用。司法院必須能夠對所建置的我國 AI 量刑輔助系統之演算法（包含標註方式）等技術原理，全面性地做出充分解釋和說明，方屬滿足政府機關向人民公開資訊的法定義務。而根據目前 AI 量刑系統已經供法官、檢察官與律師參考，司法院卻仍未見有對其演算法等技術原理有充分說明的現象觀之，上述政府機關向人民公開政府資訊之義務，恐尚未善盡¹³²。

¹²⁹ 政府資訊公開法第1條、第6條參照。

¹³⁰ 司法院，前揭註80。

¹³¹ 詳見本文「肆、二、法院的適用擴張現象」章節之說明。

¹³² 司法院針對AI量刑系統所發的歷次新聞稿，以及發函予全國律師聯合會時，公文後附之事實型AI系統使用手冊、評價型AI量刑系統使用說明中，均未具體說明建置第二代量刑輔助系統所使用機器學習方法的演算法內涵。請參見前揭註80、108、113；全國律師聯合會網站，前揭註109。此外，人權團體發布的聲明中，也清楚指出：「根據9月21日司法院新聞採訪，此系統在訓練階段的主要素材是大量的起訴書及對應的裁判書資料，所謂大量，指涉的資料時間區段為何？又起訴書部分，是由法務部協助提供『去識別化』的檢察書類。然而，如僅提供裁判

（二）被告無法使用系統造成的憲法權利侵害

相較於司法院目前並未主動向社會公開 AI 量刑系統的技術原理，從憲法的觀點，更嚴重的問題在於即使一般人民、甚至刑事被告希望使用 AI 量刑系統查找資訊，在現階段亦不可得。如同本文前述，截至 2023 年 11 月底，司法院建置的第二代量刑輔助系統仍僅開放給法官、檢察官和律師使用。換言之，若刑事被告無委任律師（辯護人），即無權使用第二代量刑輔助系統（包括事實型和評價型系統在內）。本文認為，司法院現行並未向一般人民開放使用 AI 量刑系統的作法，侵害了憲法第 8 條正當法律程序與第 16 條訴訟權對於刑事被告之保障，同時也違反憲法上的罪刑相當原則與平等原則，亟待修正。

首先，根據我國歷來憲法解釋，本於憲法第 8 條及第 16 條所保障之人身自由與訴訟權，刑事被告享有依正當法律程序之原則，在當事人對等的基礎上，受法院公平審判之權利，並特別強調保障刑事被告在訴訟上有充分之防禦權¹³³。刑事被告雖然與檢察官均為審判中的當事人，但現況下，檢察官有權使用司法院建置的 AI 量刑系統，藉以瞭解法院判斷刑罰輕重時可能參考的量刑因子為何，以及在諸多量刑因子疊加之下，AI 系統可能提供給法院的參考量刑區間為何。檢察官得根據這些資訊，準備進行科刑範圍之辯論¹³⁴。而刑事被告卻被拒於 AI 量刑系統之外，欠缺對於系統內資訊的可近用性（accessibility），無從得知法院可能參考何等量刑因子來判斷刑罰輕重，也無法知悉案內各項因子疊加之後，法院可能審酌參考的判刑區間何在，因

書及起訴書，而未使用案件的卷宗資料，如何達成司法院所稱，該系統得以『即時分析犯罪事實、證據等段落資料』，從而生產出如新聞截圖的判決書草稿？」等關於基本系統建置方法的疑問。請參見台灣人權促進會等團體共同聲明新聞稿，前揭註 118。可見目前就司法院 AI 量刑系統的基礎建置原理，無論是社會大眾或專業人士都仍不甚瞭解。

¹³³ 司法院釋字第 582 號、司法院釋字第 636 號、司法院釋字第 654 號、司法院釋字第 762 號、司法院釋字第 789 號等解釋參照。

¹³⁴ 刑事訴訟法第 289 條第 2 項：「前項辯論後，應命依同一次序，就科刑範圍辯論之。於科刑辯論前，並應予到場之告訴人、被害人或其家屬或其他依法得陳述意見之人就科刑範圍表示意見之機會。」

此難以充分準備科刑範圍而為自己防禦。如此一來，司法院建置的 AI 量刑系統現階段在雙方當事人間，僅開放予檢察官使用，刑事被告卻不得其門而入，自難謂刑事被告依據當事人對等、公平審判等原則而享有充分的防禦權¹³⁵。

再就憲法罪刑相當原則而言，司法院釋字第 775 號理由書指出：「為使法院科刑判決符合憲法上罪刑相當原則，法院審判時應先由當事人就加重、減輕或免除其刑等事實（刑法第 47 條第 1 項及第 59 條至第 62 條參照）及其他科刑資料（刑法第 57 條及第 58 條參照），指出證明方法，進行周詳調查與充分辯論，最後由法院依法詳加斟酌取捨，並具體說明據以量定刑罰之理由，俾作出符合憲法罪刑相當原則之科刑判決。」在憲法罪刑相當原則之要求下，當事人必須能就各項科刑資料指出證明方法，以進行周詳調查與充分辯論；法院也要能夠具體說明量定刑罰之理由，方符合此項憲法原則。檢察官身為一方當事人，可以依據 AI 量刑系統提供的各項量刑相關參考資訊，指出證明方法，使法院進行周詳調查，並據以進行科刑辯論；法院也能夠靠著參考 AI 量刑系統，細緻衡量各項量刑因子並參考量刑建議區間，從而能夠具體說明量定刑罰輕重之依據何在。惟有刑事被告因無操作 AI 量刑系統之權限，難以就法院和檢察官參考 AI 系統所得對於量刑因子的觀照，針對具體量刑因子逐一指出證明方法、使法院詳細調查並進行充分辯論。如此將導致刑事被告可能蒙受不符合其所犯罪之刑，憲法上罪刑相當原則亦遭破壞。

最後，就憲法上平等原則而言，刑事訴訟的審判結果如何，不應繫諸於被告之財富，否則即為非正義之司法制度¹³⁶。然而，司法院建置的 AI 量刑系統目前僅有法官、檢察官與律師可以使用，刑事被告不與焉。現況下，若非強制辯護案件，除非刑事被告有資力聘僱律師為辯護人，否則即無法使用

¹³⁵ 司法院釋字第 582 號理由書：「憲法第 16 條規定人民有訴訟之權，就刑事審判上之被告而言，應使其在對審制度下，依當事人對等原則，享有充分之防禦權，俾受公平審判之保障。」

¹³⁶ 王兆鵬、張明偉、李榮耕（2022），《刑事訴訟法（上）》，6 版，頁 592，新學林。

AI 量刑系統，瞭解法官、檢察官可能參考審酌的量刑因子與量刑區間，無辯護人之被告因此難以充分準備科刑辯論。換言之，在現況下，刑事被告能否使用 AI 量刑系統這一項對於訴訟結果可能發生影響的國家公器，竟取決於被告有無聘僱律師；而被告是否聘僱律師，又經常取決於被告之財富實力。如此一來，刑事被告之有無財富，即可能對於訴訟之科刑結果發生實質影響，造成有實力與無實力刑事被告之間的不平等，亦違反憲法上平等原則之保障。

綜上所述，本文認為，目前司法院僅開放法官、檢察官與律師使用第二代量刑輔助系統的規範，造成無律師的刑事被告無法使用該系統，以取得法官可能參考審酌的量刑因子與量刑區間等訊息，已經違反憲法對於刑事被告的人身自由與訴訟權保障，也違反憲法罪刑相當與平等原則，此種規範方式實為不當。司法院應盡快開放一般刑事被告亦得使用該 AI 量刑系統，取得量刑相關參考資訊。雖然即使開放刑事被告得以使用 AI 量刑系統，仍無法解決本文前述因該系統使用機器學習方法，伴隨而生的技術本質問題，也無解於司法院以政府機關角色主動建置並推動 AI 量刑系統，對人民所應善盡的政府資訊公開義務，但是在很可能已經有法官、檢察官正在參考、使用 AI 量刑系統的當下，讓刑事被告盡快得以近用同一系統資訊，仍是保障被告憲法權利所必須的亡羊補牢之舉。

陸、結論：揭開不同的「黑盒子」

本文在統計證據的框架下，介紹了美國與我國各種不同的輔助法院形成量刑心證的系統，包括以人工智慧方法建置的美國 COMPAS 系統、我國第二代量刑輔助系統，以及以（大）數據分析方法建立的我國第一代量刑輔助系統，並在這 3 個系統的相互比較中，呈現出「人類智慧」與「人工智慧」的不同特徵，以及各自遭遇的挑戰。這 3 個系統雖然並非基於相同的原理所建置，但是它們的目標卻是一致的：「利用本案中不存在的資訊，幫助達成

更妥適的本案量刑」¹³⁷。可是，既然用以輔助量刑的資訊，都是「本案中不存在」者，那麼系統如何取得、分類和應用這些資訊，也就不是審判者或社會大眾可以輕易、直覺觀察到的。這就難怪使用這些資訊的量刑輔助系統，容易被批評為「黑盒子」。

以美國 COMPAS 系統與我國第二代量刑輔助系統相較，兩者同樣面對機器學習方法本質上的「黑盒子」質疑，但是我國的系統由於是政府機關主動建置，完全出於公益考量，且基於政府資訊公開法的要求，司法院因此有相較於 COMPAS 系統開發商更大、更完整的說明義務。而又因為現況下，刑事被告若無律師，無法使用第二代的量刑輔助系統，因此造成被告受憲法第 8 條、第 16 條保障基本權利的侵害，以及憲法上罪刑相當原則、平等原則的破壞，本文認為司法院應儘速開放系統予刑事被告使用，以符憲法保障人民基本權的意旨。

而以我國第一代與第二代量刑輔助系統相較，第一代系統由於不涉及「機器（學習）的黑盒子」，系統建置原理相對透明；加上刑法第 57 條就法院量刑審酌事項之規定本有相當的解釋空間；而且無論是刑事被告或一般社會大眾，均可自由使用第一代的量刑輔助系統，瞭解法院裁判時可能參考的量刑因子，與系統可能建議法院就本案裁判的量刑區間，並沒有系統資訊可近用性的阻礙，因此第一代量刑系統的運作，目前並無違憲之虞。在法院充分告知當事人可能參考量刑輔助系統的前提下，運用第一代量刑系統甚至有助於落實科刑辯論的爭點集中，達成更為精緻、妥適量刑的目標。

若以我國第一代量刑輔助系統與美國 COMPAS 系統比較，則會呈現出 2 種不同的「黑盒子」。因為 COMPAS 系統使用機器學習方法決定量刑因子「權重」的透明性難題，其實也存在於我國第一代量刑輔助系統中的量刑趨勢建議系統。由人類專家共同決定或調整過的量刑因子項目與權重，固然是一種具有專業支持的群體決定，但是若要恪守量刑透明性的要求，如「為什麼某項量刑因子的權重是另一項的 0.3 倍」這類問題，仍然難有直接明確的答案。換言之，在任何的量刑估計模型中，只要使用了「權重」的調整概

¹³⁷ 請參見前揭註 26。

念，不可避免的會產生難以說明清楚其來源與正當性的模糊地帶。如果要追根究底，透過焦點團體、專家會議方式形成的權重數值，如果存在著此種說理上的模糊地帶，似乎也可以認為是某種意義上的「黑盒子」，但在現有的量刑趨勢建議系統中，這個透明性的問題卻少見遭遇挑戰¹³⁸。

本文認為，這是因為這些模糊地帶，仍然是由「人類智慧」所直接創造的。而在人類社會中，對於服膺於其他同類、特別是所謂「專家」智慧的觀點，是極常見的現象，因此不以為怪¹³⁹。因此，對於「本系統的量刑評估公式中的各項權重是由專家學者所共同討論得出」的說明，人類社會是可以接受的。可是，對於以機器學習方式產生的人工智慧，人類是否仍能順服如斯呢？儘管在機器學習的過程中，標籤（在監督式學習的情況下）和訓練資料都是由人類提供的，但人工智慧的意義，就在於機器是在利用這些標籤與資料，自行分析出了各項量刑因子應有的權重結果。至於機器究竟如何得出這樣的權重結果，吾人目前尚難得知。

對於人工智慧，人類通常仍抱持著懷疑或觀望的態度，至少在第三波「脈絡適應」的人工智慧發展成熟、機器能夠解釋其做成某種權衡判斷的理由之前，都會是如此。無論這種懷疑、觀望的態度，是來自於人類對於人工智慧或機器學習原理的誤解，或是出於「雖然瞭解但是仍難以信任機器」的人類本位心態，都會構成實際應用第二波人工智慧產出知識的阻礙。因此本文對於我國採取機器學習等人工智慧方法建置新一代量刑輔助系統的作法，在技術上是樂觀的，而且也認為有深入探索的價值與必要性；但對於要立即具體應用在實際量刑上的可能性，卻採取相對保守的態度。

因為我國量刑改革的目標，根據司法院指出，是在於「基於司法為民的理念，回應各界對公平、透明、妥適量刑的要求」¹⁴⁰。而在「公平、透明、

¹³⁸ 儘管有實務界人士指出量刑趨勢建議系統有「未揭露計算量刑因子權重公式」的缺點，但重點在於「未揭露」，而不是如同COMPAS的「無法揭露」。請參見文家倩，前揭註43，頁17。

¹³⁹ See generally Cathrine Holst & Anders Molander, *Epistemic Democracy and the Role of Experts*, 18 CONTEMP. POL. THEORY 541 (2019).

¹⁴⁰ 司法院官方網站（刑事量刑專區），<https://www.judicial.gov.tw/tw/cp-83-57186->

妥適」3項具體目標中，如果將「公平」和「妥適」理解為「類似的判決能得到類似的量刑結果」，那麼此種「一致性」的要求，透過第一代量刑輔助系統提供的資訊，可說已經獲得相當程度的實現。然而，若要進一步回應量刑應「透明」的要求，卻正是第二波人工智慧的弱點所在¹⁴¹。因為目前的人工智慧技術，尚不具有解釋其做成某種決定的理由何在的能力，因而很難認為透過人工智慧系統所產出的資訊，可以滿足量刑「透明」的要求¹⁴²。只要這一點不能克服，人工智慧技術產生的「黑盒子」，就成為了量刑過程中的「黑盒子」。而「打破量刑的黑盒子」，卻正是我國量刑改革最大的呼聲¹⁴³。因此，目標決定了路徑，如果量刑的透明化（或說可解釋性）是一個必須被滿足的需求，那麼在「可解釋的人工智慧」有相當的技術突破以前，以機器學習為主流的第二波人工智慧技術還無法滿足這個需求¹⁴⁴。

1ef46-1.html（最後瀏覽日：05/15/2024）。

¹⁴¹ 根據聯合國教科文組織（UNESCO）最近於2023年7月發布的新聞稿，人工智慧應用於司法裁判領域時，遭遇到的主要困難有四，分別是可靠性（reliability）、透明性（transparency）、可解釋性（interpretability）和偏誤性（bias）。UNESCO, *How to determine the admissibility of AI-generated evidence in courts?*, UNESCO:NEWS (July 26, 2023), <https://www.unesco.org/en/articles/how-determine-admissibility-ai-generated-evidence-courts>

在這個分類中，我國司法院改革量刑程序所要追求的「公平、透明、妥適量刑」中的「透明」，其實更接近UNESCO所說的「可解釋性」。亦即，為何如此量刑的理由，審判者必須能夠充分的向當事人與社會大眾進行說明。本段依循司法院提出的「公平、透明、妥適量刑」主張，仍使用「透明」之語彙表達此一概念。

¹⁴² 劉靜怡，前揭註72，頁100-104。

¹⁴³ 胡宜如，前揭註24，頁43。

¹⁴⁴ 人工智慧的透明性議題，除了本文討論的「審判者能夠充分向當事人與社會大眾說明如此量刑之理由」外，也涉及憲法上刑事被告防禦權、特別是上訴權的面向。亦即刑事被告必須清楚知道法院量刑所根據的事由，才能夠有意義地實現憲法所保障的救濟權利。請參見：李榮耕，前揭註45，頁160-162。

參考文獻

一、中文部分

- 文家倩（2023），〈從量刑工具探討國民法官的量刑評議〉，《司法周刊》，2156 期（別冊），頁 1-41。
- 王兆鵬、張明偉、李榮耕（2022），《刑事訴訟法（上）》，6 版，新學林。
- （2023），《刑事訴訟法（下）》，6 版，新學林。
- 王皇玉（2023），《刑法總則》，9 版，新學林。
- 司法院（2023），〈因應國民法官新制施行 司法院啟用 AI 量刑系統〉，《司法周刊》，2143 期，頁 4。
- （2015），〈司法院量刑準據八大系統建置完成〉，《司法周刊》，1771 期，頁 1。
- 呂寧莉（2020），〈臺灣高等法院 105 年度交上易字第 117 號刑事判決評析：兼論使用「量刑資訊系統」之相關問題〉，《台灣法學雜誌》，393 期，頁 53-61。
- 李建良（2020），〈人工智慧與法學變遷：法律人面對科技的反思〉，收於：李建良（主編），《法律思維與制度的智慧轉型》，頁 3-91，元照。
- 李開復、王詠剛（2020），《人工智慧來了》，遠見天下文化。
- 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 期，頁 117-186。
<https://doi.org/10.53106/102398202022030168003>
- 林鈺雄（2023），《刑事訴訟法（上冊）》，12 版，新學林。
- 邱文聰（2020），〈第二波人工智慧知識學習與生產對法學的挑戰：資訊、科技與社會研究及法學的對話〉，收於：李建良（主編），《法律思維與制度的智慧轉型》，頁 137-167，元照。

- 邵軒磊、吳國清、黃詩淳（2018），〈大數據與法資訊學：機器提取裁判內容要素之實踐〉，《月旦裁判時報》，71期，頁46-52。
<https://doi.org/10.3966/207798362018050071006>
- 胡宜如（2015），〈司法院焦點團體建議量刑參考基準介紹〉，《司法周刊》，1766期，頁2-3。
- （2016），〈量刑公開透明：司法院量刑系統介紹〉，《法律扶助期刊》，50期，頁43-44。
- 范耕維（2022），〈建構量刑階段中罪刑相當原則的第一哩路：自應報觀點形塑刑罰量定之理論嘗試〉，《中研院法學期刊》，30期，頁79-170。
- 埃米爾·侯賽因（Amir Husain）（著），溫力秦（譯）（2018），《AI創世紀：即將來臨的超級人工智慧時代》，日月文化。
- 康納曼（Daniel Kahneman）（著），洪蘭（譯）（2022），《快思慢想》，遠見天下文化。
- 陳昱成（2013），〈貝氏定理的應用〉，《科學教育月刊》，357期，頁19-28。
[https://doi.org/10.6216/SEM.201304_\(357\).0002](https://doi.org/10.6216/SEM.201304_(357).0002)
- 劉靜怡（2020），〈人工智慧時代的法學研究路徑初探〉，收於：李建良（主編），《法律思維與制度的智慧轉型》，頁91-134，元照。
- 鄭明政（2020），〈從 State v. Loomis 案件看 AI 應用於司法審判上的若干問題〉，《台日法政研究》，4期，頁165-178。
- 總統府（2017），〈總統府司法改革國是會議成果報告〉，載於：
<https://www.president.gov.tw/File/Doc/1754f2f0-c60d-4de1-a2e3-4c967610bcaa>。
- 總統府司法改革國是會議（2017），〈第五分組第五次會議〉，載於：
<https://justice.sayit.mysociety.org/總統府司法改革國是會議第五分組/第五分組第五次會議>。
- 謝煜偉（2020），〈從量刑目的論形構量刑框架及量刑理由之判決架構〉，《法官協會雜誌》，22卷，頁86-104。

- 蘇凱平（2019），〈法律數據分析的意義、理論與應用：以探索刑事法院對證據的裁量與評價為例〉，《月旦法學雜誌》，294 期，頁 101-124。
<https://doi.org/10.3966/102559312019110294007>
- （2020），〈以司法院量刑資訊作為量刑之內部性界限？：評最高法院 108 年度台上字第 3728 號刑事判決〉，《月旦裁判時報》，98 期，頁 85-94。
<https://doi.org/10.3966/207798362020080098009>
- （2020），〈以平等原則建立量刑原則的意義與價值：臺灣高等法院 105 年度交上易字第 117 號刑事判決評析〉，《台灣法學雜誌》，393 期，頁 31-42。
- （2022），〈對法院量刑心證的追索：司法院量刑資訊系統之運作與展望〉，收於：李建良、林文源（主編），《人文社會的跨領域 AI 探索》，頁 305-322，國立清華大學出版社。
- （2023），〈第二審量刑與定錨效應〉，《月旦法學教室》，249 期，頁 26-29。
<https://doi.org/10.53106/1684739324906>

二、英文部分

- 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>
- Association for the Advancement of Artificial Intelligence. (2020, June 29). *DARPA's three waves of AI research — A special issue of AI Magazine. Interactive AI Magazine*. Retrieved May 15, 2024, from <https://interactiveaimag.org/ai-magazine-previews/ai-mag-previews-vol41-no2/>
- Coetzee, F. (2023, October 4). *Creativity and Problem-Solving in the Digital Age: Navigating Challenges with Innovation*. MEDIUM.

<https://nlpwithpurpose.medium.com/creativity-and-problem-solving-in-the-digital-age-navigating-challenges-with-innovation-37385f87272>

Calo, R. (2017). Artificial Intelligence Policy: A Primer and Roadmap. *UC Davis Law Review*, 51, 399-435. <http://dx.doi.org/10.2139/ssrn.3015350>

Chan, B. K. (2024, June 13). *First, second, third wave of artificial intelligence (AI)* [Blog post]. Mind Data Intelligence Consulting Services. Archived at Wayback Machine. Retrieved from <https://web.archive.org/web/20240613211739/https://minddata.org/3-waves-Brian-Ka-Chan-AI>

Clancey, W. J. (1993). Notes on “Epistemology of A Rule-based Expert System”. *Artificial Intelligence*, 59, 197-204. [https://doi.org/10.1016/0004-3702\(93\)90186-F](https://doi.org/10.1016/0004-3702(93)90186-F)

DeBrusk, C. (2018, March 26). The Risk of Machine Learning Bias (and How to Prevent It). *MIT Sloan Management Review*. <https://sloanreview.mit.edu/article/the-risk-of-machine-learning-bias-and-how-to-prevent-it/>.

Edwards, A. (2017). Big Data, Predictive Machines and Security: The Minority Report. In M. R. McGuire & T. Holt (Eds.), *The Routledge Handbook of Technology, Crime and Justice* (pp. 451-461). Routledge.

Electronic Privacy Information Center. (n.d.). *AI and Human Rights: AI in the Criminal Justice System*. <https://epic.org/issues/ai/ai-in-the-criminal-justice-system/>

Enoch, D., & Fisher, T. (2015). Sense and “Sensitivity”: Epistemic and Instrumental Approaches to Statistical Evidence. *Stanford Law Review*, 67, 557-611.

equivant. *Our Products-Northpointe Suite Risk Need Assessments*. <https://www.equivant.com/northpointe-risk-need-assessments/>

- Flores, A. W., Bechtel, K., & Lowenkamp, C. T. (2016). False Positives, False Negatives, and False Analyses: A Rejoinder to “Machine Bias: There's Software Used across the Country to Predict Future Criminals and It's Biased against Blacks.” *Federal Probation*, 80, 38-46.
- Fienberg, S. E. & Straf, M. L. (1991). Statistical Evidence in the US Courts: an Appraisal. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 154(1), 49-59. <https://doi.org/10.2307/2982694>
- Fouse, S., Cross, S., & Lapin, Z. J. (2020). *DARPA's impact on artificial intelligence. Interactive AI Magazine.* <https://ojs.aaai.org/aimagazine/index.php/aimagazine/article/view/5294>
- Gardiner, G. (2019). Legal Burdens of Proof and Statistical Evidence. In D. Coady & J. Chase (Eds.), *The Routledge Handbook of Applied Epistemology* (pp. 179-195). Routledge.
- Green, B. & Hu, L. (2018). The Myth in the Methodology: Towards a Recontextualization of Fairness in Machine Learning. *Machine Learning: The Debates Workshop at the 35th International Conference on Machine Learning (ICML)*. <https://scholar.harvard.edu/files/bgreen/files/18-icmldebates.pdf>
- Holst, C. & Molander, A. (2019). Epistemic Democracy and the Role of Experts. *Contemporary Political Theory*, 18, 541-561. <https://doi.org/10.1057/s41296-018-00299-4>
- Johnson, J. (2020, July 16). *Interpretability vs Explainability: The Black Box of Machine Learning.* *BMC Blogs*. <https://www.bmc.com/blogs/machine-learning-interpretability-vs-explainability>
- Kahneman, D. (2012). *Thinking, Fast and Slow*. Penguin Group.
- Lightbourne, J. (2017). Damned Lies & Criminal Sentencing Using Evidence-Based Tools. *Duke Law and Technology Review*, 15, 327-343.

- McCarthy, J. (2007, November 12). *What is Artificial Intelligence?*. <https://www-formal.stanford.edu/jmc/whatisai.pdf>
- McCarthy, J., Minsky, M. L., Rochester, N. & Shannon, C. E. (2006). A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence. *AI Magazine*, 27(4), 12-14.
- McDermid, J. A., Jia, Y., Porter, Z. & Habli, I. (2021). Artificial Intelligence Explainability: The Technical and Ethical Dimensions. *The Royal Society*, 379(2207), 1-18. <https://doi.org/10.1098/rsta.2020.0363>
- Northpointe Inc. (2015, March 19). *Practitioner's Guide to COMPAS Core*. <https://s3.documentcloud.org/documents/2840784/Practitioner-s-Guide-to-COMPAS-Core.pdf>
- Pundik, A. (2011). The Epistemology of Statistical Evidence. *The International Journal of Evidence & Proof*, 15(2), 117-143. <https://doi.org/10.1350/ijep.2011.15.2.373>
- (2021). Rethinking the Use of Statistical Evidence to Prove Causation in Criminal Cases: A Tale of (Im)Probability and Free Will. *Law and Philosophy*, 40, 97-128. <https://doi.org/10.1007/s10982-020-09389-0>
- Ross, L. (2021). Rehabilitating Statistical Evidence. *Philosophy and Phenomenological Research*, 102(1), 3-23. <https://doi.org/10.1111/phpr.12622>
- Ross, G. (2019, June 8). *The taxicab problem*. *Futility Closet*. <https://www.futilitycloset.com/2019/06/08/the-taxicab-problem/>
- Roth, A. (2017). Machine Testimony. *Yale Law Journal*, 126, 1972-2053.
- Samek, W. & Müller, K. R. (2019). Towards Explainable Artificial Intelligence. In W. Samek, G. Montavon, A. Vedaldi, L. K. Hansen & K.-R. Müller (Eds.), *Explainable AI: Interpreting, Explaining and Visualizing Deep Learning*. (pp. 5-22). Springer. https://doi.org/10.1007/978-3-030-28954-6_1

- Saeed, W. & Omlin, C. (2023). Explainable AI (XAI): A Systematic Meta-Survey of Current Challenges and Future Opportunities. *Knowledge-Based Systems*, 263, 110273. <https://doi.org/10.1016/j.knosys.2023.110273>
- Schmalbeck, R. (1986). The Trouble with Statistical Evidence. *Law and Contemporary Problems*, 49(3), 221-236. <https://doi.org/10.2307/1191634>
- Shadowen, A. N. (2019). Ethics and Bias in Machine Learning: A Technical Study of What Makes Us “Good”. In N. Lee (Ed.), *The Transhumanism Handbook*. (pp. 247-261). Springer. https://doi.org/10.1007/978-3-030-16920-6_12
- Shin, D. (2001). The Effects of Explainability and Causability on Perception, Trust, and Acceptance: Implications for Explainable AI. *International Journal of Human-Computer Studies*, 146, 1-10. <https://doi.org/10.1016/j.ijhcs.2020.102551>
- Siegel, R. B. (2018). Blind Justice: Why the Court Refused to Accept Statistical Evidence of Discriminatory Purpose in *McCleskey v. Kemp* - And Some Pathways for Charge. *Northwestern University Law Review*, 112(6), 1269-1292.
- State of Wisconsin, Department of Corrections. (2018, October 30). *Electronic Case Reference Manual, COMPAS Assessment Frequently Asked Questions*. <https://www.manuallib.com/download/2023-10-18/Electronic%20Case%20Reference%20Manual.pdf>
- Thompson W. C. & Schumann, E. L. (1987). Interpretation of Statistical Evidence in Criminal Trials: The Prosecutor's Fallacy and The Defense Attorney's Fallacy. *Law and Human Behavior*, 11(3), 167-187. <https://doi.org/10.1007/BF01044641>
- Tversky, A. & Kahneman, D. (1974). Judgment under Uncertainty: Heuristics and Biases. *Science*, 185(4157), 1124-1131.
- Vidmar, N. & Hans, V. P. (2007). *American Juries: The Verdict*. Prometheus.

- Vidmar, N. (2011). The Psychology of Trial Judging. *Current Directions in Psychological Science*, 20 (1), 58–62.
<https://doi.org/10.1177/0963721410397283>
- Wilkins, P. A. (2012). Confronting the Invisible Witness: The Use of Narrative to Neutralize Capital Jurors' Implicit Racial Biases. *West Virginia Law Review*, 115(1), 305-362.

Using Artificial Intelligence to Assist Sentencing in Courts? Perspectives on Statistical Evidence and Big Data

*Kai-Ping Su**

Abstract

In popular culture, the depiction of artificial intelligence or other technological means replacing human judges reflects society's expectations of the judicial system. However, has the current development of artificial intelligence made these expectations potentially, or practically, attainable? This article approaches the question from the standpoint of "statistical evidence," a type of evidence that could most possibly be introduced with emerging technologies like artificial intelligence and data analysis to aid criminal court judgments. By quoting literature, this article explains that contemporary behavioral science research has adequately demonstrated the unavoidable limitations of "human intelligence" in judgment, thus necessitating the involvement of "artificial intelligence" to aid humans in decision-making, in particular in court decision-making.

This article further traces the three waves of development of artificial intelligence and explores the possibility of "replacing human judges with artificial intelligence" in terms of the technology development today. Using the widely employed U.S. criminal justice system assistance software, the COMPAS, as well as the first and the second generation of "Sentencing Assistance System", established by the Judicial Yuan (Taiwan's highest judicial authority), this article elaborates on the specific theories, methodologies, advantages, disadvantages, and potential constitutional disputes when using artificial intelligence and big data analysis technology to assist courts in sentencing decisions. In addition to detailing how these two different technological approaches aid criminal courts in decision-making, the article also compares society's acceptance levels of judgment results

* Associate Professor, College of Law, National Taiwan University.
E-mail: sukaiping@ntu.edu.tw

produced by these technologies and delves into the interaction and trust relationship between human intelligence and artificial intelligence.

This article holds that artificial intelligence is not able to “replace” the human decision-making in court sentencing, because of two reasons. First, the judicial authority would not allow this practice, because it seems to inevitably infringe the core of the judiciary. Second, the public would not allow that the sentencing decisions are made by AI machines which lacks transparency and explainability.

Keywords: sentencing, statistical evidence, behavioral science, heuristics, machine learning, COMPAS

